NEW DELHI METALLO-BETA-LACTAMASE (NDM) PLASMID PREDICTIONS

USING THE RESISTANCE GENE IDENTIFIER (RGI) AND MOB-SUITE SOFTWARE TO PREDICT NEWDELHI METALLO-BETA-LACTAMASE (NDM) ASSOCIATED PLASMIDS

BY
MARTINS OLONI, B.Sc.

A THESIS
SUBMITTED TO THE TO THE DEPARTMENT OF BIOCHEMISTRY & BIOMEDICAL SCIENCES
AND THE SCHOOL OF GRADUATE STUDIES
OF MCMASTER UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF MASTER OF SCIENCE

McMaster University MASTER OF SCIENCE (2020) Hamilton, Ontario (Biochemical & Biomedical Sciences)

TITLE:	Using the Resistance Gene Identifier (RGI) and MOB-Suite software to predict New-Delhi Metallo-beta-lactamase (NDM) associated plasmids
AUTHOR:	Martins Oloni, B.Sc. (Biochemistry), York University, Toronto, Canada
SUPERVISOR:	Dr. Andrew G. McArthur
NUMBER OF PAGES:	x, 75

Lay Abstract

The spread of antimicrobial resistance often arises from the exchange of resistance plasmids between bacterial species in their immediate environment. One resistance gene carried on plasmids is the New Delhi Metallo-beta-lactamase (NDM) gene, which confers resistance to almost all beta-lactam antibiotics. To control the spread of NDM plasmids, we need to detect them completely. Using genome sequencing, we analyzed clinical isolates from patients in hospitals within the Hamilton-Niagara community for NDM plasmids. From our results, we detected one complete match to a known NDM plasmid. We also detected seven NDM-associated plasmids that are close variants of known NDM plasmids, plus another seven NDM-associated plasmid that we classified as 'novel' since they are widely different from known NDM plasmids. This thesis reveals the diverse NDM plasmids within the clinics of the Hamilton-Niagara community, which should guide public health workers to develop policies to mitigate the spread of resistance NDM plasmids.

Abstract

The complete detection and genome sequencing of resistance plasmids is required to understand the co-occurrence of resistance genes on plasmids and their transmission, which is necessary to mitigate the spread of drug resistant bacterial infections. One group of multi-drug resistance plasmids of interest are the **New Delhi Metallo-β-lactamase (NDM) associated** plasmids. Bacteria pathogens with NDM genes are resistant to a broad range of beta-lactam antibiotics, including carbapenems, a mainstay for treating bacterial infections. We sequenced clinical isolates collected from patients that failed to respond to antibiotic treatments in the Hamilton-Niagara community using Next Generation Sequencing (NGS) technology, which is the standard technique for sequencing bacterial genomes. I analyzed their assembled genomes via the development of a new resistance plasmid prediction bioinformatics pipeline: RGI:Mobilome. RGI:Mobilome predicted NDM bearing plasmids in 15 isolates, along with other resistance genes on plasmids. However, the NDM plasmid predictions of all 15 isolates were fragmented due to incomplete genome assemblies, caused by repetitive sequences within bacterial genomes and use of Illumina sequencing technology. However, plasmid predictions greatly improved when we leveraged the high nucleotide accuracy of Illumina reads and the structural resolving power of long reads generated with the Oxford Nanopore Technology (ONT) to produce 'hybrid' genome assemblies. From the 15 putative hybrid-assembly NDM plasmids, one plasmid was a complete match to a known pKP-NDM1 plasmid, seven plasmids were "variants" of known and well-characterized plasmids, and the remaining seven plasmids were "distant homologs" of known and well-characterized plasmids, suggesting previously undescribed NDM-associated plasmids in our community. This thesis project reveals the diversity of NDM plasmid types within the Hamilton-Niagara community, which is valuable to epidemiologists and public health practitioners to devise actionable plans required to mitigate the spread of multi-drug resistance NDM plasmids and for clinicians to treat infections caused by NDM positive strains.

Acknowledgements

First, I would like to thank God almighty for making it possible to complete and write this thesis, especially during this unprecedented pandemic period. The past two years of my graduate school career is filled with memories that I would cherish forever. There were trying times, but looking back now, it shaped me to be a better researcher, and above all, a better person.

I want to express my deepest and sincere gratitude to my principal investigator, Dr. Andrew G. McArthur, for allowing me to transition into Bioinformatics, for his supervision, direction, support, and above all, creating a conducive and inclusive environment for learning. I would also like to thank my committee members: Dr. Sara Andres, Dr. John Whitney, and Dr. John Nash for your leadership, support, guidance, and encouragement. I would also like to thank Dr. Kyrylo Bessonov from the Public Health Agency of Canada (PHAC), Guelph, for his mentorship, commitment, ideas, and helping with the completion of the RGI:Mobilome pipeline. I would also like to thank Allison Guitor from the Wright lab for the extraction, purification, and Nanopore sequencing of the NDM clinical isolates. I would also like to express my deepest gratitude to the current and past McArthur lab members. I don't think I would be able to get through my thesis without your assistance, support, encouragement, and camaraderie. Thank you, Jalees Nasir, Kara Tsang, Amos Raphenya, Brian Alcock, David Speicher, Arman Edalatmand, Rachael Tran, Sohaib Syed, Anatoly Mirosnichenko, Rafik El Werfalli, Sally Min, Anna-Lisa Nguyen, Hafsa Omer, Sarah Yaqoob, Corie Niu, Hamna Imtiaz, and Megan Kwong.

Finally, and more importantly, I would like to thank my family and friends, for being my support system. To my mom, Florence Oloni, thank you for your unconditional love, sacrifice, encouragement, and always believing in me. To my siblings, Princeton Oloni and Tunolase Oloni, thanks for cheering me on and praying for me. Thank you, Dominique Brown, for riding with me on this one. My utmost appreciation to my friends and family home and abroad who supported me during this incredible journey. I love you all!

Table of Contents

List of Tables

List of Figures

Declaration of Academic Achievement

This research project was partly funded by the Genome Canada project known as ARETE. ARETE stands for the Antimicrobial Resistance: Emergence, Transmission, and Ecology, aiming to understand the transmission dynamics of AMR. The extraction of clinical bacterial isolates, bacterial genome isolation, purification, and genome sequencing analyzed for this project work, was funded by the Canadian Institute of Health Research and the Ontario Research Fund.

I would like to thank Alison Guitor from Dr. Gerry Wright's lab for the purification and Nanopore sequencing of the genomic DNA of our 15 NDM isolates of interest. A big thank you to our colleagues at PHAC, Guelph: James Robertson, Dr. Kirill Bessonov, and Dr. John Nash for developing and maintaining MOB-Suite, and for allowing its integration with Resistance Gene Identifier (RGI). I would also like to thank Amos Raphenya and Brian Alcock for maintaining RGI software and the Comprehensive Antibiotic Resistance Database (CARD). The computing resources used for this project work was provided by the McMaster Service Lab and Repository computing clusters.

Chapter 1

Introduction

Antimicrobial resistance (AMR) is an increasing threat to efficient therapeutics for a wide range of infections caused by bacteria[1,2].Most antibiotic resistant infections occur in hospitals and clinics, leading to additional risks with medical treatments and procedures like organ transplantation and chemotherapy[3]. As a result, AMR poses a challenge regarding a hospital's ability to provide refuge to heal and perform medical procedures without the risk of secondary infections[4]. Currently, there is a dearth in the discovery of new antibiotics, requiring careful management of our existing antibiotics[5,6]. Our two potent 'drug classes of last resort' – carbapenems and colistin[7] – are under threat with the arrival of multi-drug resistance NDM-1 in 2006[8] and MCR-1 in 2015[9]. Bacteria develop AMR intrinsically or acquire AMR through various mechanisms. The innate ability of bacteria to resist certain antibiotics results from the structural components and functional features of certain bacteria species[10]. For example, Gram-negative bacteria are inherently resistant to vancomycin antibiotics since they can't cross the outer membrane of Gram-negative bacteria[11]. Acquired resistance results from the spontaneous mutation of bacterial DNA or due to the acquisition of new genetic material (e.g., naked DNA, transposons, integrons, insertion elements, plasmids, integrative and conjugative elements) by a susceptible bacterium, which then becomes resistant[12]. Most bacteria pathogens continually acquire genetic elements that encode a diverse range of AMR genes and become multi-drug resistant to a broad range of antibiotics. Bacteria inherits these genetic materials differently depending on whether they are encoded by the bacterial chromosome (i.e., vertically) or acquired through extra-chromosomal DNA (i.e., horizontally)[13]. To deal with the spread of infections caused by multi-drug resistant bacteria, we need to be able to fully characterize the mobile genetic elements that transmit these multi-drug resistance genes. Thus, the goal of this thesis is to detect multi-drug resistance plasmids that transmits across bacterial species.

1.1 Mutational resistance

Bacteria can spontaneously develop mutation for AMR. The mutation frequency for AMR is usually within the order of 10^{-10} to 10^{-9}, which means one in 10^{10} to 10^{9} bacteria will spontaneously mutate and develop AMR[14]. The rapid division of bacterial cells allows for bacteria to evolve resistance quickly, which explains why infections caused by bacterial pathogens like *Neisseria gonorrhea* and *Mycobacterium tuberculosis* are difficult to treat[15]. Although, most mutational resistance arises in a bacteria population primarily due to antibiotic pressure, but the frequency of mutation depends on antibiotic concentration. Antibiotics are developed to target different gene products, leading to the destruction or slow growth of bacteria. For example, rifamycin generally targets mRNA transcription by binding RNA polymerase, which reduces the expression level of essential genes[16]. Some antibiotics target the conserved bacterial machinery like the topoisomerases. For example, quinolones target DNA gyrase and inhibit the unwinding of bacteria DNA, thus preventing replication and hinders bacterial growth[17]. Even in the presence of antibiotics, the mutant cells replicate and pass on their resistant genes vertically to all the bacteria progenies. Antibiotics will eliminate bacterial cells without the newly evolved drug capacity.[18]. Over time, the entire bacterial population becomes resistant to that particular class of antibiotics.

1.2 Horizontal transfer of AMR genes

In contrast to vertically inherited mutation in response to antibiotic pressure, bacteria also use horizontal gene transfer (HGT) to acquire new genetic material independent of their clonal lineage, allowing rapid change of their antibiotic resistance phenotype. For example, when bacteria are faced with persistent selection pressure due to antibiotics, acquisition of multi-drug resistance genes from external sources enhances bacteria fitness, leading to increased resistance to a broad range of antibiotics[19]. Evolution of AMR thus becomes a community instead of population problem in which unrelated bacteria trade AMR determinants. For example, the clinical multi-drug resistance genes NDM-1 (carbapenems) and MCR-1 (colistin) are thought to have environmental origins

from which they rapidly disseminated into farms and clinics through HGT. HGT also plays a significant role in the outbreak of infections due to the transfer of pathogenic traits, such as virulence genes, and the natural tendency of bacteria to form biofilms[20]. Understanding of HGT and its impact on AMR transmission dynamics is thus important for elucidating the mobile genetic material that makes bacteria multi-drug resistant.

HGT in bacteria occurs through one of three primary mechanisms: conjugation, transformation, and transduction[19,21]. In transformation, foreign DNA is taken up by bacteria from the surrounding environment by binding non-covalently to sites present on the cell surface. The uptake of DNA varies among bacteria species and is yet to be fully understood. In general, double-stranded DNA gets converted to single-stranded DNA as it translocates across the membrane into the bacterial cytoplasm[22]. The horizontally transferred DNA is then integrated into the host bacterial genome using homologous recombination. However, the transferred DNA must contain 25 to 200 bp sequences highly similar to the host genome to allow for recombination[23]. In the event of an unsuccessful recombination event, the transferred DNA gets fragmented by restriction enzymes. In transduction, bacteriophages (viruses that infect bacteria) mediate gene transfer[24] When bacteriophages infect bacteria, they can carry parts of its genome and then transfer it to a new infected bacterial cell. Gene transfer via transduction can either be generalized or specialized[25]. In generalized transduction, bacteriophages infect a donor bacteria cell and begin the lytic pathway. It takes over the replication machinery to synthesize virus particles like capsids, enzymes, and tail fibers. The viral enzyme breaks down the bacteria chromosome into fragments. During the assembly of progeny phage particles, the fragment of donor bacterial DNA gets incorporated into the phage capsid. When the bacteriophage carrying bacterial DNA infects a new cell, it transfers the donor bacteria DNA into the new bacterial genome through homologous recombination. Conversely, specialized transduction requires the transfer of a restricted set of genes from donor to recipient bacteria through a lysogenic pathway. It usually occurs during the life cycle of temperate bacteriophages. First, a temperate bacteriophage integrates its genome into a host bacterial cell and remains inactive in the host cell. The phage's genome's exposure to ultraviolet (UV)

light or chemical prompts its excision from the host cell. During this process, the bacterial genes lying adjacent to the phage genome are also excised and packaged as part of a new phage particle. Once the new phage infects a new bacterial cell, it transfers the excised bacterial genes into the recipient bacterial genome. In conjugation, DNA molecules, such as plasmids or chromosomally integrated conjugative elements (ICEs)[26], are transferred from a donor to a recipient bacterium by direct contact. The donor bacterium carries a conjugative plasmid, and the recipient bacterium does not, in most cases. Conjugative plasmids are self-transmissible and carry all the necessary genes required for transfer[27,28]. One gene class carried by conjugative plasmids are the tra genes, which allows for mate-pair formation between bacterial strains via a pilus. They also possess the origin of transfer (oriT) sequences, which serve as start sites to initiate plasmid DNA replication and transfer. Mobilizable plasmids are not self-transmissible since they do not encode the tra genes needed for conjugation but possess the oriT sequences to initiate plasmid transfer, and with the help of conjugative plasmids. The tra genes of conjuative plasmids are essentially responsible for mate-pair formation between plasmids, while oriT of mobilizable plasmids enable plasmid DNA movement through the pilus. Integrative and conjugative elements (ICEs) reside in bacterial chromosomes but can be excised for transfer since they encode the genes required for integration and excision[28]. Like conjugative plasmids, ICEs can thus facilitate their own transfer. Unlike conjugative plasmids, they cannot replicate autonomously outside of bacterial chromosomes and are typically larger in size, varying from 18 to 500 kbp. ICEs exist in different forms ranging from conjugative transposons to non-mobile genomic islands (GI)[29], making them difficult to detect. The latter is the emphasis of the ongoing IslandViewer project. Yet, while AMR plasmids are well described due to their prevalence in clinical infections, the landscape of plasmid-borne AMR is still relatively unknown. This is exacerbated by the ongoing challenges for sequencing and detecting plasmids from bacterial genome projects. As such, this thesis focuses on detecting resistance plasmids, accurately determining their DNA sequence, and understanding their role in the transfer of multi-drug resistance genes amongst clinical pathogens.

1.3 Next-generation sequencing of bacterial pathogens and tools for AMR surveillance

The diminishing cost and increasing throughput of next-generation DNA sequencing enable AMR surveillance in clinical, agricultural, and environmental settings[30]. First, the rise in the use of genomics in microbial ecology and public health has led to the sequencing of an enormous number of relevant clinical bacterial genomes. For example, the NCBI Pathogen Detection Isolates Browser (https://www.ncbi.nlm.nih.gov/pathogens/isolates#/search/) contains more than 430,00 isolates with whole-genome shotgun sequence data. Of this, the GenomeTrakr network has more than 167,000 isolates and is now regularly sequencing over 5,000 isolates each month (https://www.fda.gov/food/whole-genome-sequencing-wgs-program/genometrakr- network)[31]. Second, the explosion in the number of datasets has led to the advent of data management tools like the Comprehensive Antibiotic Resistance Database (CARD; card.mcmaster.ca). Yet, the present obstacle with the genomic surveillance of AMR is no longer data acquisition but data analysis, where our current bioinformatics tools do not provide a complete evaluation of our genomic data. A broad knowledge of AMR genotype, and its relation to AMR phenotype, is required for choosing the correct antibiotics to treat patients in hospitals and for antimicrobial stewardship.

CARD[32–34] is an actively curated database of reference AMR sequence data for prediction of AMR genotype from genomic sequences and is one of a few bioinformatics tools developed for studying antibiotic resistance using genome sequencing. The basis of CARD is the comprehensive biocuration of molecular AMR sequences, which includes both chromosomally and extra-chromosomally encoded AMR genes and mutations covering the entire scope of resistance mechanisms and bacterial pathogens. The CARD curation team thoroughly investigates and verifies new AMR terms for complete DNA sequence and checks for high minimum inhibitory concentration (MIC) levels, relative to controls described in a scientific publication, before curation into CARD[33]. Central to CARD is the Antibiotic Resistance Ontology (ARO), which provides thorough vocabularies for AMR genes and their products, thereby enabling the comprehensive study

of molecular sequence data. The ARO classifies AMR terms with four primary tags (AMR Gene Family, Drug Class, Resistance Mechanisms, Antibiotics) and three secondary tags (Adjuvants, Efflux Component, Efflux Regulator)[33]. In addition to ARO, the CARD includes the Resistance Gene Identifier (RGI) software that incorporates the ARO, bioinformatics models in CARD's Model Ontology (MO), and reference AMR sequence data to predict resistance genes from complete or partial genome sequences.

RGI uses four detection model types to predict AMR determinants: protein homolog model, protein variant model, protein over-expression model, rRNA gene variant model. Each reference AMR gene curated into CARD has a detection model ascribed to it. RGI also analyzes genome sequences under three detection criteria: Perfect, Strict, and Loose. The perfect algorithm detects a complete match to curated AMR reference sequences. The strict algorithm detects functional variants of reference AMR sequences using model cut-offs, and the loose algorithm detects distant homologs and potential novel AMR genes. The perfect and strict algorithms are used to monitor AMR genes in clinics, while the loose algorithm monitors any looming AMR gene(s) that poses a high threat. As of August 2020, CARD contains 4550 ontology terms, 3057 reference sequences, 1357 single nucleotide polymorphic sequences, 2735 publications, and 3103 AMR detection models. CARD has additional used RGI to predict AMR determinants from 9560 chromosomes, 21362 plasmids, 102181 whole-genome shotgun (WGS) assemblies, 2220111 alleles, and the genomes of 88 bacterial pathogens, providing a breadth of information regarding AMR gene prevalence. The CARD's robust detection models and algorithm provide researchers, clinicians, public health agencies, and industries with a useful and practical database for AMR analysis. Hence, the CARD is a dependable and established database of reference AMR sequences and variants.

1.4 MOB-Suite: A software tool for plasmid prediction from genome assembly data

Detection of plasmid-borne AMR is critical for understanding AMR's transmission dynamics in different habitats (e.g., clinic, agriculture, environment, community). The use

of a high-throughput short-read sequencing platform enables the assembly of bacterial genomes[35], yet these genomes usually contain repetitive sequences that hinder complete plasmid assembly from short-read sequencing data[36]. Consequently, plasmids are often mis-predicted from WGS assemblies, with only fragments of plasmid DNA sequences obtained, inappropriately placed contiguous to chromosomal sequences, or placed in the assembly 'trash bin' altogether. There are few *in silico* tools that detect plasmid sequences from genome assembly data with high accuracy, sensitivity, and specificity. One such tool is MOB-Suite[37], which identifies plasmid sequences from WGS assemblies, making this a breakthrough for studying plasmid transmission dynamics. To overcome plasmid assembly challenges, MOB-Suite first reconstructs the plasmid sequences from genome assembly data, using a reference plasmid database to identify plasmid contigs, and then uses a reference gene database to type plasmids as conjugative, mobilizable, or non-mobilizable.

MOB-Suite is comprised of three linked modules for plasmid prediction: MOB-Cluster, MOB-Typer, and MOB-Recon. MOB-cluster groups reference plasmids based on sequence similarities, using MASH software to estimate the genomic distance between reference plasmids and identified plasmid contigs from draft assembly contigs[38]. MOB-cluster then assigns the identified plasmid to a cluster of related plasmids with a similar plasmid backbone. Hence, MOB-cluster is responsible for choosing the closest reference plasmid to the predicted plasmid contigs. MOB-recon reconstructs and aggregates plasmid contigs using the reference plasmids of MOB-cluster. This module is responsible for assigning each contig of a genome assembly to a plasmid or chromosomal unit. Contigs with a replicon and/or relaxase gene are grouped as plasmid contigs. MOB-recon assigns the remaining contigs as belonging to a chromosome or unresolvable repetitive sequences. Lastly, MOB-typer is responsible for replicon and relaxase typing of plasmid genes, plus typing for the origin of transfer (oriT) and mate-pair formation (mpf) sequences, relative to a reference database. The predicted genes from the identified plasmid contigs determine the potential mobility of the plasmid: prediction of only a replicon classifies the plasmid as non-mobilizable, prediction of either a relaxase or oriT sequence classifies the plasmid as

mobilizable, prediction of a mate-pair formation sequence classifies the plasmid as conjugative.

1.5 Third-generation sequencing (TGS) of bacterial pathogens

While next-generation sequencing (NGS) technologies can analyze bacterial genomes on a large-scale, it performs poorly when resolving bacteria genome and plasmid structure due to assembly challenges[39]. The genome assemblies of next-generation sequencing reads create genomes with missing regions that sometimes encode essential genes, making it difficult to analyze bacterial genomes[40]. The arrival of third-generation sequencing (TGS) technology enables the complete assembly of bacterial genomes[41]. TGS technologies produce read lengths of 100,000 or more, capable of spanning and resolving the repetitive sequences of bacterial genomes, leading to a complete genome and plasmid structure. The two most commercially available TGS technologies are Pacific Biosciences (PacBio) Single-Molecule Real-Time (SMRT)[42] and the Oxford Nanopore Technologies (ONT) sequencing platforms[43]. The PacBio sequencing platform was introduced in 2010 and uses a similar sequencing-by-synthesis strategy as a NGS platform. Unlike NGS, which senses amplified signals from a population of fluorescent DNA fragments, PacBio captures a single DNA molecule. Normally, PacBio technology gives a high error rate of around 13-15%, due to the low signal to noise ratio of single bacteria DNA, but this error can be reduced with higher read coverage. The Oxford Nanopore Technologies (ONT) is the more recent TGS and was introduced in 2014. This technology uses the Oxford Nanopore MinION instrument to sequence native single-stranded DNA (ssDNA) by measuring the change in electric current as nucleotide bases pass through the nanopore[44]. Like PacBio, the nanopore sequencing of the bacterial genome is limited by high error-rate. However, its small size and relatively low-cost compared to PacBio makes it easily deployable in laboratories and clinics[45]. Both sequencing platforms have helped improve the completeness, correctness, and contiguity with sequencing and assembly of bacterial genomes. The emerging paradigm for accurate sequencing of bacterial genomes and plasmids is to combine both NGS and TGS data in a ‘hybrid’ assembly approach[46]. This

hybrid approach requires supplementing the accurate, short-read NGS data with less accurate, long TGS data to resolve the complex repetitive sequences of bacterial genomes. This approach is useful since NGS only results in incomplete genome assembly, and TGS only results in low assembly accuracy due to inherent error of the sequencing data.

1.6 Research Question and Hypothesis

The New Delhi Metallo-β-lactamase (NDM) belongs to the family of Metallo-β-lactamase (MBL) that hydrolyzes most beta-lactams including carbapenems, a drug of last resort.,[47]. Most beta-lactams are inhibitors of bacterial peptidases, which blocks the cross-linking of peptidoglycan molecules during cell wall biosynthesis. In the past decades, the overuse and misuse of beta-lactam antibiotics has allowed for the evolution and dissemination of beta-lactamases like NDM. The first NDM, NDM-1, was first recognized in a *Klebsiella pneumoniae* strain isolated from a Swedish patient of Indian descent, who was admitted for urinary tract infection in New Delhi, India[47]. Since this first discovery, NDM genes have been identified in diverse bacteria species, with 24 NDM variants already characterized[47]. The NDM enzyme is a class B -β-lactamase with two zinc ions present at the active site, where the hydrolysis of β-lactams occurs. The zinc ions initiate hydrolysis by activating a water molecule and leading to the formation of a proton and active hydroxide ion. The activated hydroxide ion then attacks the β-lactam carbonyl group, which leads to hydrolysis[47]. The NDM enzyme secondary structure comprises 9 α-helices, 17 β-strands, and 3 turns responsible for its stability. NDM is a lipoprotein and binds to the outer lipid membrane of most gram-negative bacteria due to hydrophobic interactions. It's anchoring to the membranes ensures its stability under conditions of Zinc deprivation[48]. The mechanism of resistance of NDM and other MBL to beta-lactams include target modification (mutation or expression of alternative penicillin binding protein, PBP), disruption of cell permeability through reduction of porin formation required for beta-lactam entry, and overexpression of efflux pumps

The 24 NDM variants differ from another due to substitution with one or more of the 270 primary amino acids. The NDM variants NDM-5, NDM-17, and NDM-20 have a

common Valine (V) to Leucine (L) substitution in position 88 amidst other changes, which increases their carbapenemase activity[49]. The NDM-1 gene is commonly identified in *Klebsiella pneumonia* isolates, *Acinetobacter* species, and *Enterobacter* species, whereas NDM-5 is predominant in *Escherichia coli* isolates. These NDM isolates are widely known to cause hospital-acquired infections like bloodstream infections, abscesses, urinary tract infections, and so on. Since the first discovery of NDM-1 in India, NDM positive strains have spread globally[50]. While the role that travel plays in the number of observed clinical cases of NDM-associated bacterial infection is well documented, we have less data on how NDM positive strains become pervasive within a community. The NDM-1 gene has been identified on bacterial chromosomes, however, mobile elements like plasmids play a major role in the spread of NDM-1 gene. For this project, we aim to study the diversity of resistance NDM plasmids within the Hamilton-Niagara community, which is part of a larger Genome Canada funded effort to understand the transmission dynamics of AMR in not just clinical but agricultural and environmental settings. CARD in its current form is relatively uninformative regarding resistance mobile elements, so this study requires the use of sequencing technologies, CARD, and MOB-Suite to get a stronger detection of resistance NDM plasmids. Thus, if complete multi-drug resistance plasmids can be predicted accurately from genome assembly data, we can better understand AMR transmission and inform public policy. There is a strong need to understand the distribution and diversity of multi-drug resistance genes within bacteria pathogens, how often the AMR gene transfer between pathogens, their mechanism of transmission, and their epidemiological effect within communities and the overall population

Chapter 2

Materials and Methods

2.1 RGI:Mobilome algorithm for resistance plasmid predictions

A RGI:Mobilome plasmid prediction algorithm was written in Python 3.7, built upon the MOB-Suite (v3.0.0) and RGI (v5.1.0) code bases. This algorithm was designed as an object-oriented class in the rgi_mobilome.py python module. The class name is RGI_Mobilome and can only be initialized with an input file, a folder for the pipeline output, and the number of central processing units (CPU) threads required. Aside from the initializing function for the RGI_Mobilome class, it currently has 8 different python functions within the class. The tentative names of the functions are validate_inputs, mob_suite, run_rgi_on_mob_result run_rgi_tool, load_contig_report, plasmid_report, per_sample_collapse_contig_report, chromosome_report, write_output2table. The algorithm also accepts command-line arguments available in the mob_arg.py python module. The user arguments include –input_file (-i), --run_mob_suite, --num_threads (-n), --plasmid_rgi_report (-p), --chromosome_rgi_report (-c), --output_directory (-o). The –input_file accepts an input assembly contigs file (accepts only FASTA file format). The –-run_mob_suite is a required command for running the MOB-Suite algorithm upon the assembly data to differentiate chromosomal and plasmid contigs. The –num_threads accepts the number of CPU thread to run the algorithm. The –-plasmid_rgi_report prompts the algorithm to run RGI's *main* subroutines on the putative plasmid(s) contigs to predict plasmid encoded AMR genes. The –-chromosome_rgi_report prompts the algorithm to run RGI's *main* subroutines on the chromosomal contigs. The –output_directory defines an output folder for the results. The results of the RGI:Mobilome algorithm are reports of the predicted AMR genes on both the chromosomal and plasmid contigs (file format: .txt), RGI *main* report of either the chromosomal or plasmid contigs (file formats: .txt and .json).

RGI:Mobilome is currently available at https://gitlab.com/Oloni/rgi-mobilome-in-card. This algorithm works on a Linux terminal and uses command-line arguments.

2.2 Clinical isolate collection

The GDW clinical surveillance project, supervised by Dr. Gerry Wright, involved collecting clinical multi-drug resistant bacterial isolates from patients diagnosed with bacterial infections at Hamilton Health Science hospitals from 2015 to 2018. The isolates were collected from different sources like urine, blood, skin, mucus, depending on the type and nature of bacterial infections. The bacterial isolates chosen for DNA extraction and purification were the isolates that failed to respond to at least three classes of antibiotics following Antibiotics Susceptibility Testing (AST). Isolate culturing and AST were performed at Hamilton General Microbiology Laboratory under the Clinical and Laboratory Standards Institute (CLSI) guidelines. Subsequent laboratory work was performed by members of the Wright laboratory. Briefly, each isolate was cultured to on agar plate and a single colony of each isolate was inoculated in Luria Bertani (LB) broth media and allowed to incubate overnight. Bacterial cultures were lysed using a lysis buffer comprised of EDTA chelator and sodium dodecyl sulfate (SDS). The lysed bacteria cell cultures were centrifuged at room temperature. The PureLink genomic DNA mini kit (Invitrogen, Carlsbad, CA) was used to isolate and purify genomic DNA from cell pellets, with the remaining genomic DNA stored in a -80°C freezer.

2.3 Whole Genome Sequencing (WGS), assembly, and RGI analysis of clinical bacteria genomes

A total of 478 genomic DNA Library preparations were made using the Illumina Next XT DNA Library Preparation Kit (Illumina, San Diego, CA) and sequencing performed using 2 x 150 bp paired-end sequencing on an Illumina HiSeq 1500 platform or 2 x 250 bp paired-end sequencing on an Illumina MiSeq v3 platform at the Farncombe Metagenomics Facility at McMaster University. Sequencing results were inspected using

FASTQC[51] (v0.11.9, https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and trimmed using Trimmomatic (v0.38, https://github.com/timflutre/trimmomatic) to remove barcodes, adapters, and low-quality bases[52]. The trimmed reads were assembled *de novo* into draft genome assembly contigs using SPAdes (v3.12, https://github.com/ablab/spades). The 478 genome assemblies were subsequently analyzed using the Resistance Gene Identifier (v4.2.2, https://github.com/arpcard/rgi).

2.4 MOB-Suite and PlasmidProfiler analysis of clinical NDM isolates

MOB-Suite (v2.1.0, https://github.com/phac-nml/mob-suite) was used to analyze the 478 genome assemblies, with a focus on genome assemblies associated with NDM ß-lactamases. The databases used by MOB-Suite for prediction included the reference plasmid dataset, replicon gene sequences, relaxase sequences, origin of transfer (oriT) sequences, and mate-pair formation sequences all gathered from the NCBI repository. The software PlasmidProfiler (v0.1.6, https://plasmid-profiler.readthedocs.io/en/latest/), which examines raw sequencing reads instead of assembly contigs, was additionally used for comparative plasmid analysis of the 15 NDM-associated isolates, using MOB-Suite's replicon gene database as the reference plasmid gene database for the comparative plasmid analysis.

2.5 Nanopore Sequencing and Hybrid genome assembly of the NDM clinical bacterial genomes

The 15 isolates with predicted NDM genes were selected for Nanopore sequencing to allow hybrid Illumina + Nanopore assembly for resolution of complete chromosome and plasmid sequences. This work was performed by Ms. Allison Guitor of the Wright laboratory. Prior to sequencing, each genomic DNA was assessed for quality via spectrophotometer and fluorometer. Only high-molecular-weight DNA with no RNA or protein contamination was processed for library preparation. Approximately 1000ng of genomic DNA was used per isolate for library preparation, with the Nanopore libraries prepared using the Rapid Barcoding Kit (Oxford Nanopore Technologies, Oxford, UK).

The flow cell was primed (FLO-MIN106; Oxford Nanopore Technologies, Oxford, UK) by following the Ligation Sequencing Kit and Native Barcoding manual and using materials from the flow cell priming kit (EXP-FLP002; Oxford Nanopore Technologies, Oxford, UK). Before loading, the flow cell was checked to ensure the number of pores exceeded the minimum number of active pores covered by warranty (at least 800 pores for the MinION flow cell). All reagents were prepared according to the manual regarding thawing, mixing, and storage. Barcoded samples were pooled in an equal ratio, each library was mixed prior to loading by pipetting, and the sample port was loaded on the flow cell. The flow cell ran for 48 hours.

The Nanopore reads were generated as multi FAST5 files but were converted to single FAST5 files for binning using multi_to_single_fast5 (https://github.com/nanoporetech/ont_fast5_api). Reads were classified before base calling using "deepbinner classify –rapid singlefast5 > classifications" (https://github.com/rrwick/Deepbinner). The classification file output was used after base-calling to bin reads into their respective barcodes. Base-calling was performed using "guppy_basecaller -I input_fast5_file -s output_basecalled_reads config dna_r9.4.1_450bps_hac.cfg device cuda:0" (https://github.com/nanoporetech/pyguppyclient). The chosen configuration was dependent on the kit and flow cell used. This base caller was chosen for high accuracy. The resulting files were binned using "deepbinner bin - classes classification – reads base_called_reads out_dir output_directory" (https://github.com/rrwick/Deepbinner). After barcoding and binning, the reads were trimmed using "porechop -I input_file.fastq.gz. -o output_reads.fastq.gz –discard_middle" (https://github.com/rrwick/Porechop). NanoStat was used to generate read stats and quality plots. The trimmed Nanopore reads of the 15 NDM-associated isolates were hybrid assembled their associated trimmed WGS reads using the Unicycler (v0.4.8, https://github.com/rrwick/Unicycler) assembly pipeline and command "unicycler -1 short_read1.fastq.gz -2 short_reads2.fastq.gz -l long_reads.fastq.gz -o output_dir –-keep 1 -t 32". The resulting assemblies were analyzed using the new RGI:Mobilome algorithms to predict plasmids and plasmid-associated AMR genes

2.6 Read mapping and plasmid annotation

The whole genome sequencing (WGS) paired-end reads of the 15 NDM-associated isolates were aligned to their respective putative NDM plasmids and MOB-Suite predicted most related reference plasmid using the Bowtie2 (v2.2.4 http://bowtie- bio.sourceforge.net/bowtie2/manual.shtml) aligning software. Samtools (https://github.com/samtools/samtools) was used to convert the bowtie2 output (.sam file format) to binary output (.bam file format and the Geneious software (v 2020.0.3) was used to visualize the read mapping and annotate the NDM-associated plasmids. For plasmid annotation, Geneious provided a BLAST search of the NDM plasmid sequences against the nucleotide gene database in NCBI repository to annotate the plasmid genes.

Chapter 3

Results

3.1 Development of a RGI:Mobilome pipeline to predict resistance plasmids from genome assembly data

To predict plasmid associated mobility of AMR genes predicted by RGI, I developed the RGI:Mobilome pipeline within RGI, combining both the plasmid-typing models of MOB-Suite and the existing RGI algorithms to predict plasmid-borne AMR genes as shown in Figure 3.1. Like RGI, RGI:Mobilome is written in Python version 3.7.2 and builds upon the existing RGI version 5.1.0 and the new MOB-Suite version 3.0.0 code bases. RGI:Mobilome is an object-oriented Python class that uses the MOB-Suite plasmid gene database and plasmid reference database, plus the CARD reference data and detection models. RGI:Mobilome can only be used in the command line interface (Figure 3.2) and is currently available at https://gitlab.com/Oloni/rgi-mobilome-in-card, but will be incorporated into the official RGI code once vetted by CARD's software engineers. First, the algorithm checks and validates the input assembly contigs using the --input_file. Then it uses the –-mob_suite flag to run the MOB-Suite algorithm, first initializing the reference plasmid database. Subsequently, it runs the MOB-Suite command (mob_recon --infile assembly_contig.fasta –-outdir output_dir --run_typer) to run the MOB-recon and MOB-typer modules. The command allows MOB-Suite to re-construct the pre-assembled contigs using the reference plasmid database via its MOB-Cluster algorithms and types for replicons, relaxases, oriT, and mpf genes using MOB-Suite's pre-compiled plasmid gene database. The plasmid gene typing step separates the chromosomal contigs from the plasmid contigs. Depending on the desired outcome, RGI:Mobilome runs RGI's AMR annotation pipeline (i.e. RGI main) on either the chromosomal or plasmid contigs using the –-chromosomal_rgi_report or –plasmid_rgi_report to output either the chromosomal or plasmid AMR genes prediction reports. Both flags can also be specified to generate both

chromosomal and plasmid AMR gene prediction reports. To achieve this AMR gene prediction, RGI:Mobilome runs the RGI main command (rgi main –input_sequence plasmid/chromosomal_contig –output_file output_directory –input_type contig –local –clean) on chromosomal and/or plasmid contigs. Overall, RGI:Mobilome cross-references the output of MOB-Suite and RGI main to output contig files and report files for both chromosomal and plasmid encoded AMR genes.

3.2 Whole Genome Sequencing (WGS) of clinical bacterial isolates and RGI analysis of New Delhi β-lactamase (NDM) encoding bacterial genomes

In total, 478 multidrug-resistant bacteria were isolated from patients in Hamilton Health Sciences hospitals (Hamilton, Ontario, Canada) as part of a clinical surveillance project led by Dr. Gerry Wright (McMaster University) to collect diverse, multi-antimicrobial resistant strains. The strains were submitted for Illumina genome sequencing and antibiotic susceptibility testing (AST) in four different rounds. AST was used to classify isolates as either 'resistant' or 'susceptible' to 18 antibiotics under the Clinical and Laboratory Standards Institute (CLSI) guidelines.

For each isolate, Illumina reads were trimmed with Trimmomatic to remove adapters and low-quality bases[52], FASTQC was used to confirm the quality of the reads[51], and the genomes assembled using SPAdes[53]. The 478 genomes were analyzed using RGI to find isolates with multi-drug resistance NDM genes under RGI's Perfect or Strict criteria. The 478 assembled genomes consisted of 124 *Escherichia coli* strains, 94 *Pseudomonas aeruginosa strains*, 78 *Klebsiella pneumoniae strains*, 54 *Proteus mirabilis strains*, 41 *Enterobacter cloacae* strains, 20 *Citrobacter freundii* strains, 16 *Staphylococcus aureus* strains, 20 *Morganella morganii* strains, 15 *Acinetobacter baumanii* strains, 12 *Klebsiella oxytoca* strains, 2 *Providencia stuartii* strains, and 2 *Providencia rettgeri* strains (Figure 3.3), many of which are known to cause nosocomial infections . RGI predicted NDM genes as Perfect hits for 15 isolates (Table 3.1). 13 of the 15 NDM strains belonged to the gram negative Enterobactericeae (6 *Klebsiella pneumoniae*, 3 *Escherichia coli*, 2 *Enterobacter cloacae*, 1 *Citrobacter fruendii*, 1 *Klebsiella oxytoca*) and the remaining two belonged to

the Morganellaceae (1 *Morganella morganii*, 1 *Providencia stuartii*). For each of these NDM-associated assemblies (Table 3.2), the total number of raw Illumina reads ranged from 918,031 to 13,666,670, the total number of assembled contigs ranged from 62 to 214, the assembly length ranged from 4,018,549 bp to 6,288,238 bp, the GC content ranged from 41.7% to 52.25%, and the assembly N50 ranged from 95,996 bp to 398,997 bp. All 15 of the NDM bearing isolates were selected for further investigation

3.3 Prediction of New Delhi β-lactamase (NDM) gene bearing plasmids from whole genome sequencing (WGS) data

Using PlasmidProfiler, we analyzed the WGS reads of our 15 NDM genomes of interest for plasmid incompatibility groups and similar plasmid content amongst samples. PlasmidProfiler binned plasmids from the 15 isolates using the kmers analysis toolkit (KAT), short-read sequence typing, and BLAST. Among the predicted plasmids, the detected incompatibility groups found were Inc A/C2, Inc Col, Inc ColKP3, Inc ColpVC, Inc ColRNAI, Inc FIA, Inc FIB, Inc FII, Inc L/M, Inc N, and Inc R, after filtering initial result (sureness = 0.75) as shown in Figure 3.4. All the isolates had at least one predicted replicon gene as part of their plasmid contents and the dendrogram reveals similarities in plasmid content, with GDW2288 *Enterobacter cloacae* and C0612 *Klebsiella pneumoniae* isolates both having the Inc HI1B replicon genes, isolates GDW2286 *Citrobacter fruendii* and GDW2284 *Providencia stuartii* both having the Inc A/C2 replicon genes, isolates GDW2290 *Klebsiella oxytoca* and C0650 *Klebsiella pnuemoniae* both having the Inc FIB and Inc FII(S) replicon genes, GDW2287 and GDW2216 *Klebsiella pneumoniae* isolates both having the Inc A/C2 and Inc FII replicon genes, and GDW2285 and C0648 *Escherichia coli* isolates both having the Inc ColRNAI genes. To supplement these results, RGI:Mobilome was used to analyze the genome assemblies of our 15 NDM isolates of interest. The MOB-Suite version 2.1.0 was used for the analysis of NDM plasmid, which differs from the updated version 3.0.0 currently incorporated into RGI:Mobilome. A total of 84 plasmids were predicted from the WGS assembled data, of which 46 were predicted to be non-mobilizable, 24 predicted as mobilizable, and 16 predicted as conjugative (Figure 3.5A). RGI:Mobilome predicted that the NDM gene in every isolate was associated with a

plasmid. Overall, MOB-Suite predicted a broad range of possible plasmids in these isolates, with an average of 6 plasmids per isolate. However, none of the predicted plasmid contigs from the Illumina assembly data was complete (i.e. circular), undermining confidence in our ability to fully predict each plasmid's biology. Like PlasmidProfiler, MOB-Suite predicted Inc A/C2, Inc FIA, Inc FIB, IncFII, and Inc N replicon incompatibility genes as well as MOBF, MOBH and MOBP relaxases from the Illumina NDM plasmids (Table 3.3).

3.4 Complete prediction of New Delhi β-lactamase (NDM) gene associated plasmids prediction from Illumina plus Nanopore hybrid assemblies

To obtain a better prediction of isolate chromosome and plasmid sequences, Nanopore long-read data was obtained for each isolate (Table 3.4). Hybrid assembly of the Illumina plus Nanopore reads for the NDM isolate genomes using Unicycler resulted in 70 plasmids, 66 of which were completely assembled, i.e. circular, and 4 plasmids predicted as incomplete, i.e. non-circular. Based on mobility predictions, 31 plasmids were predicted as conjugative, 18 plasmids as mobilizable, and 31 plasmids as non-mobilizable (Figure 3.5B). From both the Illumina assemblies alone and Illumina plus Nanopore hybrid assemblies, the MOB-Suite plasmid novel_0 had the highest number of observations (10 for Illumina alone, 8 for hybrid assemblies). These plasmids are labeled 'novel' because their genomic distance is greater than the minimum MASH distance (0.05) required to assign them to a MOB-Suite cluster. From the Illumina plus Nanopore hybrid assemblies, plasmid clusters of 955 and 1220 had high numbers of observations (6 and 5, respectively). The total number of predicted NDM hybrid assembly plasmids was 15, one for each NDM-associated isolate, 14 of which were complete plasmid assemblies and 1 incomplete (Table 3.5). Ten of the NDM plasmids were predicted as conjugative, three predicted as mobilizable, and two predicted as non-mobilizable (Table 3.5). The predicted NDM plasmid in GDW2218 *Klebsiella pneumoniae* isolate was 100% identical to known plasmid pKP-NDM1, while predicted plasmids in seven other isolates appeared to be variants of known plasmids and 7 additional predicted plasmids had low similarity to any known

plasmids (Table 3.5). Based on MOB-Suite clustering, six plasmids may have common ancestry (plasmid_955 cluster), with two isolates with MASH neighbour pMS6198A and another two with MASH neighbour P1002-NDM1. An additional two plasmids may share common ancestry within the plasmd_1220 cluster, both with closest MASH neighbour blaNDM1 plasmid p2, and all others may have independent origins (Table 3.5).

3.5 Read mapping and sequence analysis to validate NDM plasmid predictions

To validate all NDM hybrid assembly plasmids predictions, we compared how similar they are to their MASH nearest neighbor plasmid provided by MOB-Suite (Table 3.5). Predicted plasmids were completely annotated with Geneious software and the Illumina reads for each isolate were mapped to both the predicted plasmid sequence and their MOB-Suite mash neighbor plasmid for comparison. Four of the predicted plasmids (1 identical, 2 variant, 1 distant) are detailed below to illustrate the validation of known, variant, and novel NDM plasmids discovered in these clinical isolates. See the Appendix for the results for all predicted NDM plasmids.

3.5.1 Validating detection of known plasmid pKP-NDM-1

With RGI-Mobilome, six different plasmids were predicted from GDW2218 *Klebsiella pneumoniae* Illumina assembly data. AMR genes were predicted on two of these plasmids. Hybrid assembly predicted two complete AMR encoding plasmids, plus two additional complete plasmids (Table 3.6). The predicted NDM- 1 bearing plasmid in this isolate has 100% sequence identity with pKP-NDM1 (KF992018)[54], a conjugative plasmid isolated from *Klebsiella pneumoniae* strain KP1. Both plasmids have identical annotation, AMR gene content (Table 3.6), and read mapping coverage (Figure 3.6).

3.5.2 Validating variant NDM plasmids

RGI Mobilome predicted eleven different plasmids from C0648 *Escherichia coli* isolate based on Illumina data, but following hybrid assembly the plasmid prediction was reduced to eight circular plasmids. The predicted hybrid assembly plasmids consist of two complete AMR encoding plasmids and six additional complete plasmids (Table 3.7). The predicted NDM-5 bearing plasmid in this isolate referenced nearest mash neighbor plasmid pECY55, known from *Escherichia coli* strain Y5. Both plasmids are relatively similar in sizes: 131,146 bp for NDM-5 bearing plasmid and 124,378 bp for mash neighbor plasmid pECY55. Notably, plasmid pECY55 lacks any NDM gene, suggesting acquisition of NDM-5 by the C0648 plasmid. Both plasmids have similar backbones and mobility genes. The difference between them lies between their resistance island. The NDM-5 bearing plasmid differs from the plasmid pECY55 by missing aminoglycoside AAC (6') Ib-cr gene, blaOXA-1, and catB3 genes on a 21Kbp resistance island. The unmapped 2409 bp region of plasmid pECY55 encodes these missing genes. Also, the NDM-5 plasmid does not encode the IS 1380 transposase, the CTX-M-15 gene and Tn3 transposase seen in the second unmapped region (1986 bp) of plasmid pECY55. The resistance genes aadA12, rmtB, TEM-1 and NDM-5 genes were predicted on the resistance island of the NDM-5 plasmid but not present in plasmid pECY55. Overall, both plasmids have some regions with similar plasmid contents and identical AMR genes, but also regions of AMR gene differences. Plasmid annotation and read mapping illustrates the differences between these two plasmids (Figure 3.7).

RGI mobilome was also used to analyze GDW2285 *Escherichia coli* isolate to illustrate another instance of complete NDM plasmid variant prediction. RGI mobilome predicted five different plasmids from the Illumina assembly data for this isolate but following hybrid assembly the number of predicted plasmids was reduced to four (Table 3.8). The predicted hybrid assembly plasmids consist of two complete AMR encoding plasmids and two addition plasmids (Table 3.8). The predicted NDM-1 bearing plasmid referenced mash neighbor plasmid pMS6198A, isolated from *Escherichia coli* strain

MS1698. Both plasmids have identical AMR genes around the same regions (Figure 3.8). The only difference between the plasmids is a missing 18 kb region in mash neighbor plasmid pMS1698A that encodes a few mobility genes but not AMR genes. This missing region was validated through read mapping (Figure 3.8).

3.5.3 Validating putative novel plasmids

RGI Mobilome was used to analyze GDW2287 *Klebsiella pneumoniae* Illumina data. Three plasmids were predicted, including an NDM-1 plasmid, but were incomplete. Hybrid assembly predicted two complete AMR encoding plasmids, plus one additional complete plasmid (Table 3.9). The predicted NDM-1 bearing plasmid was predicted to be conjugative from the hybrid assembly data and referenced mash nearest neighbor p1002-NDM1, isolated form *Escherichia coli* strain Z1002. Notably, this predicted NDM-1 plasmid was 100% identical in size and sequence to the NDM-1 plasmid of isolate *Klebsiella pneumoniae* GDW2216 (Table 3.5). The NDM1- bearing plasmid and mash neighbor plasmid p1002-NDM1 encode five identical AMR genes, including NDM-1, in different regions of both plasmids. The size of the GDW2287 NDM-1 plasmid is 82,327 bp whereas p1002-NDM1 is 111,688 bp and read mapping illustrates significant differences between the two plasmids (Figure 3.9). The unmapped regions of p1002-NDM1 plasmid encode four AMR genes not encoded in the GDW2287 plasmid: BRP(MBL), mphA, TEM-1, and CTX-M-14. Overall, the predicted NDM-1 bearing plasmid and mash neighbor plasmid p1002-NDM1 are different in size, AMR gene organization, and read mapping coverage (Figure 3.9).

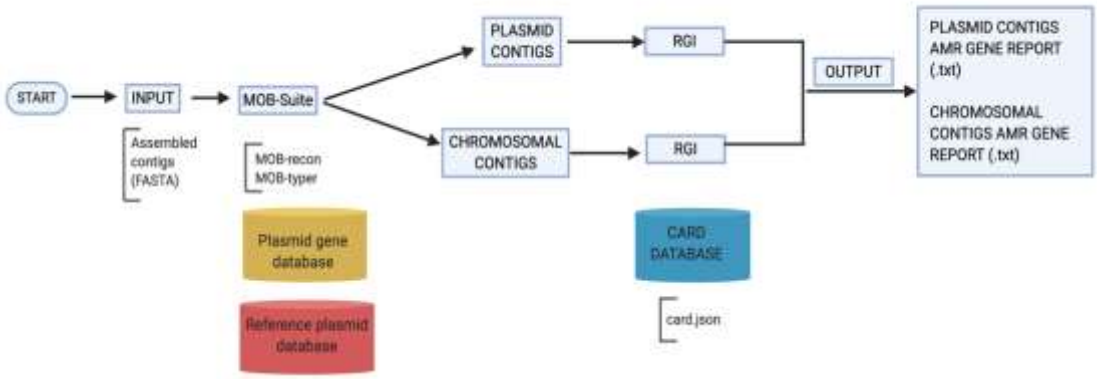


Figure 3.1: **RGI:Mobilome pipeline for plasmid and chromosomal resistance gene predictions**. RGI:Mobilome begins with assembled contigs (.fasta format), then separates plasmid from chromosomal contigs by MOB-Suite algorithms (Version 3.0.0), using the plasmid gene database and reference plasmid database. The RGI algorithm predicts AMR genes from both chromosomal and plasmid contigs using CARD reference data and models (card.json). The outputs of this pipeline are contigs and summary reports.

```
usage: rgi mobilome [-h] -i INPUT_FILE -o OUTPUT_DIRECTORY --run_mob_suite
                    [-n NUM_THREADS] [-p] [-c]

RGI Mobilome to predict plasmid-borne AMR from genome assembly data

optional arguments:
  -h, --help            show this help message and exit
  -i INPUT_FILE, --input_file INPUT_FILE
                        Input assembly fasta file to process
  -o OUTPUT_DIRECTORY, --output_directory OUTPUT_DIRECTORY
                        Output directory
  --run_mob_suite       runs mobsuite on input_fasta file
  -n NUM_THREADS, --num_threads NUM_THREADS
                        Number of threads to be used
  -p, --plasmid_rgi_report
                        runs rgi on mobsuite plasmid fasta files
  -c, --chromosome_rgi_report
                        gives chromosome rgi report in addition

Plasmid AMR surveillance tool
```

Figure 3.2: **RGI:Mobilome algorithm with command line arguments (screenshot)**. The –input_file command is used for validation of the input assembled contigs. The –run_mob_suite command separates chromosomal from plasmid contigs, while –num_threads specifies the number of CPU core/threads to be used. The –plasmid_rgi_report command specifies only plasmid contigs and plasmid AMR gene report to be created in the output directory. The –chromosome_rgi_report command specifies only chromosomal contigs and chromosomal AMR gene report to be created in the output directory. The –output_directory command specifies the name of the output folder or directory for the RGI:Mobilome pipeline.

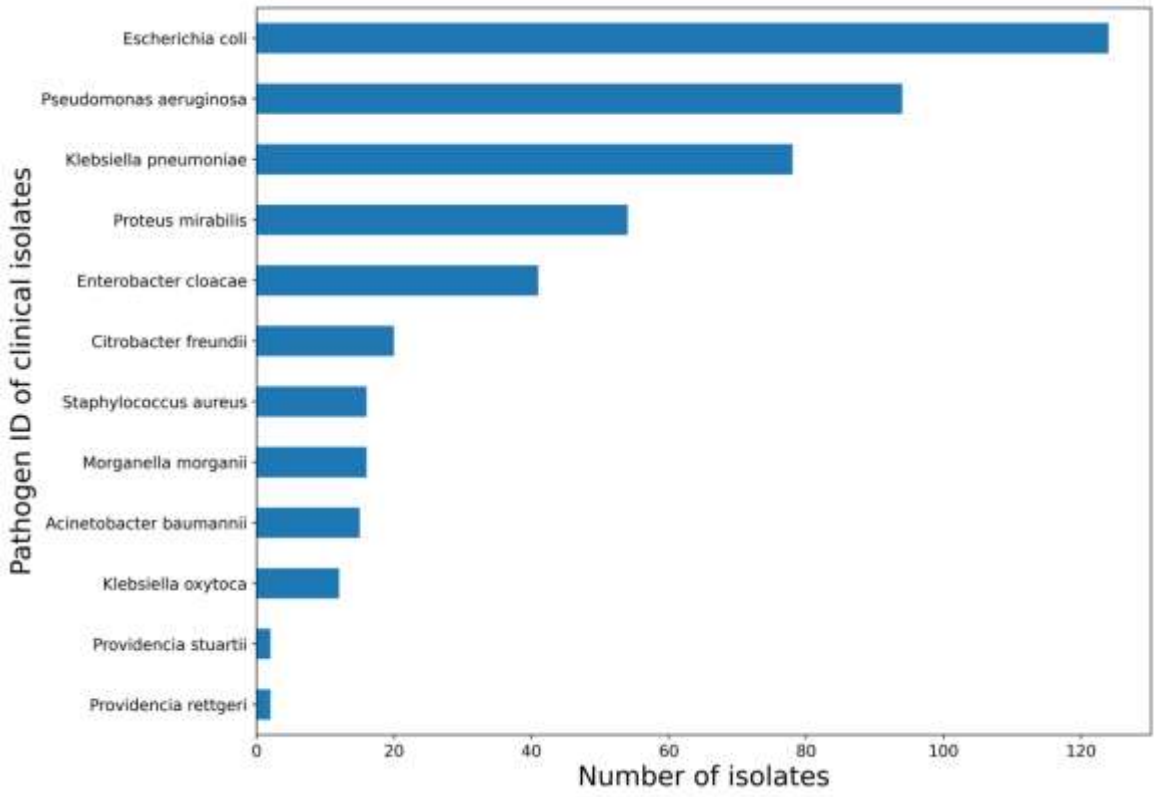


Figure 3.3: **Pathogen ID of clinical isolates analyzed with the Resistance Gene Identifier (RGI).** The isolates were identified with clinical microbiology techniques and by genome sequencing, using the Livermore Metagenomics Analysis Toolkit (LMAT, v1.2.6). The RGI identified isolates bearing the multi-drug resistance NDM genes.

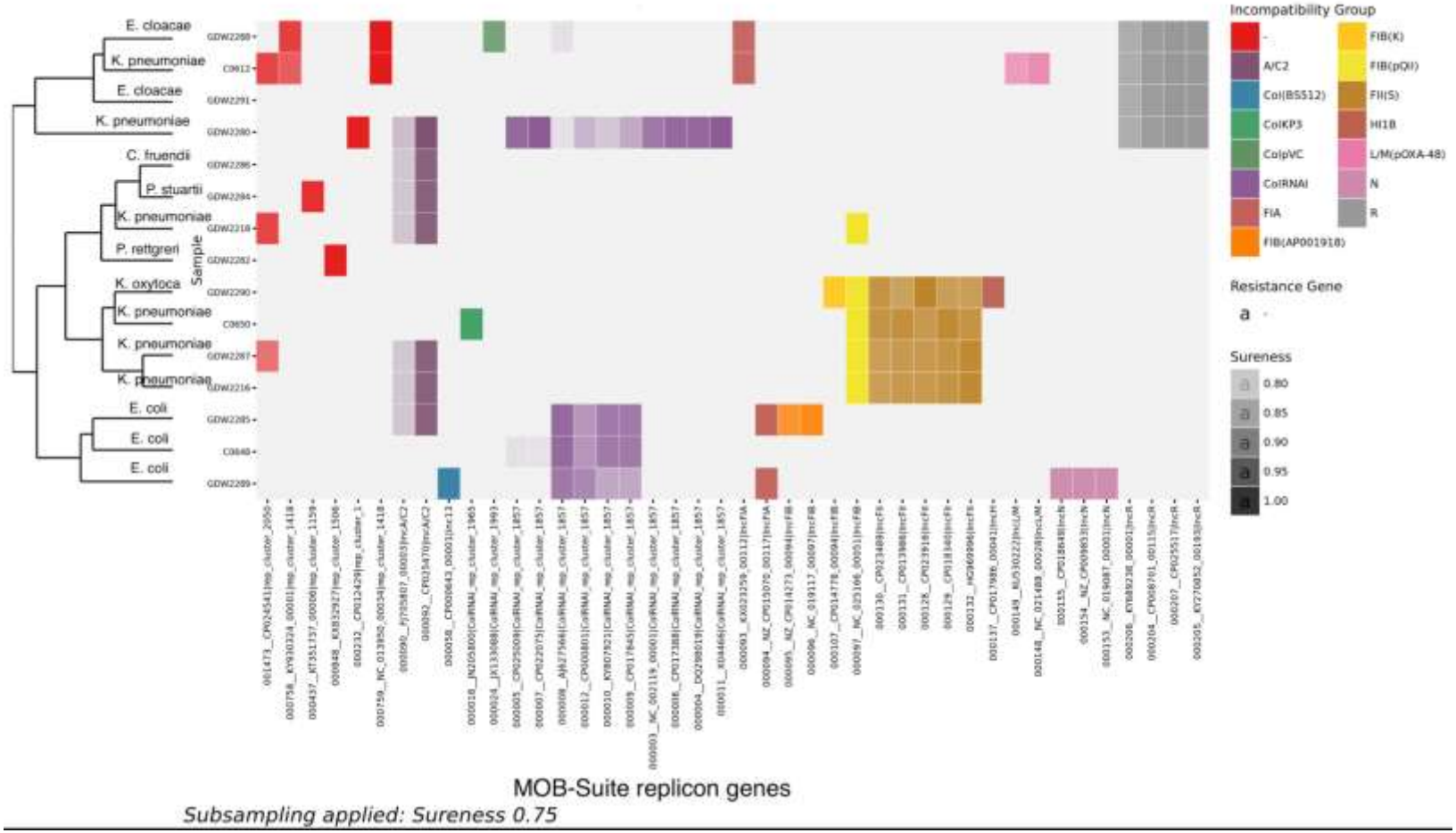
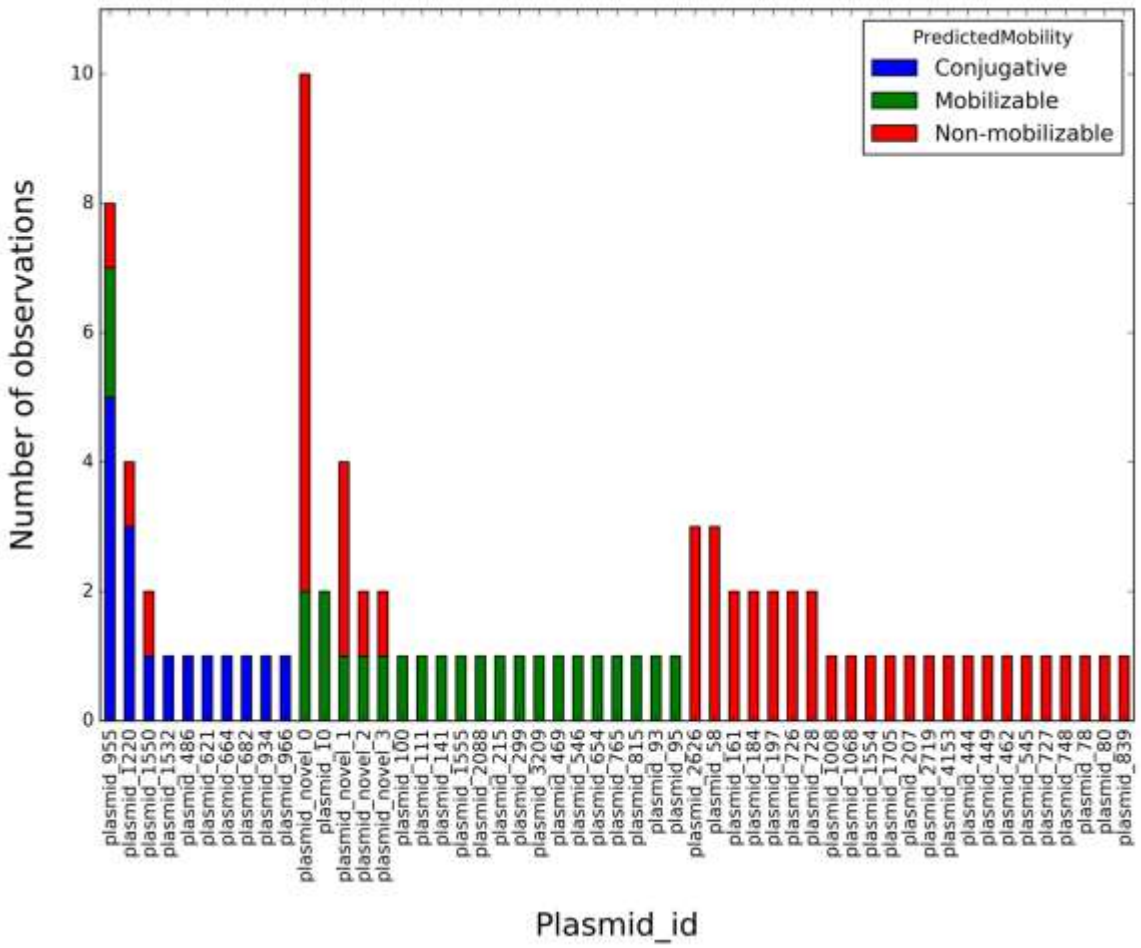

E. cloacae
K. pneumoniae
E. cloacae
K. pneumoniae
C. fruendii
P. stuartii
K. pneumoniae
P. rettgeri
K. oxytoca
K. pneumoniae
K. pneumoniae
K. pneumoniae
E. coli
E. coli
E. coli
Sample
MOB-Suite replicon genes
Subsampling applied: Sureness 0.75
Incompatibility Group
-
A/C2
Col(BS512)
ColKP3
ColpVC
ColRNAI
FIA
FIB(AP001918)
FIB(K)
FIB(pQil)
FII(S)
HI1B
L/M(pOXA-48)
N
R
Resistance Gene
a
Sureness
0.80
0.85
0.90
0.95
1.00

Figure 3.4: **Plasmid gene content similarities amongst NDM isolates by Illumina read-alignment to MOB-Suite reference replicon incompatibility genes, as generated using PlasmidProfiler.** The dendrogram shows the isolates and how they are grouped based on similarities of their plasmid contents. The heatmap shows the incompatibility groups used for grouping the isolates. The incompatibility groups within the isolates are Inc A/C2, Inc Col (BS512), Inc ColKP3, Inc ColpVC, Inc ColRNAI, Inc FIA, IncFIB, Inc FII, InH, IncL/M, Inc N, Inc R. The default 0.75 sureness value is applied for filtering reference and provides confidence on the incompatibility gene prediction.

A)


PredictedMobility
Conjugative
Mobilizable
Non-mobilizable
Number of observations
0
2
4
6
8
10
plasmid_955
plasmid_1220
plasmid_1550
plasmid_1532
plasmid_486
plasmid_621
plasmid_664
plasmid_682
plasmid_934
plasmid_966
plasmid_novel_0
plasmid_10
plasmid_novel_1
plasmid_novel_2
plasmid_novel_3
plasmid_100
plasmid_111
plasmid_141
plasmid_1555
plasmid_2088
plasmid_215
plasmid_299
plasmid_3209
plasmid_469
plasmid_546
plasmid_654
plasmid_765
plasmid_815
plasmid_93
plasmid_95
plasmid_2626
plasmid_58
plasmid_161
plasmid_184
plasmid_197
plasmid_726
plasmid_728
plasmid_1008
plasmid_1068
plasmid_1554
plasmid_1705
plasmid_207
plasmid_2719
plasmid_4153
plasmid_444
plasmid_449
plasmid_462
plasmid_545
plasmid_727
plasmid_748
plasmid_78
plasmid_80
plasmid_839
Plasmid_id

B)

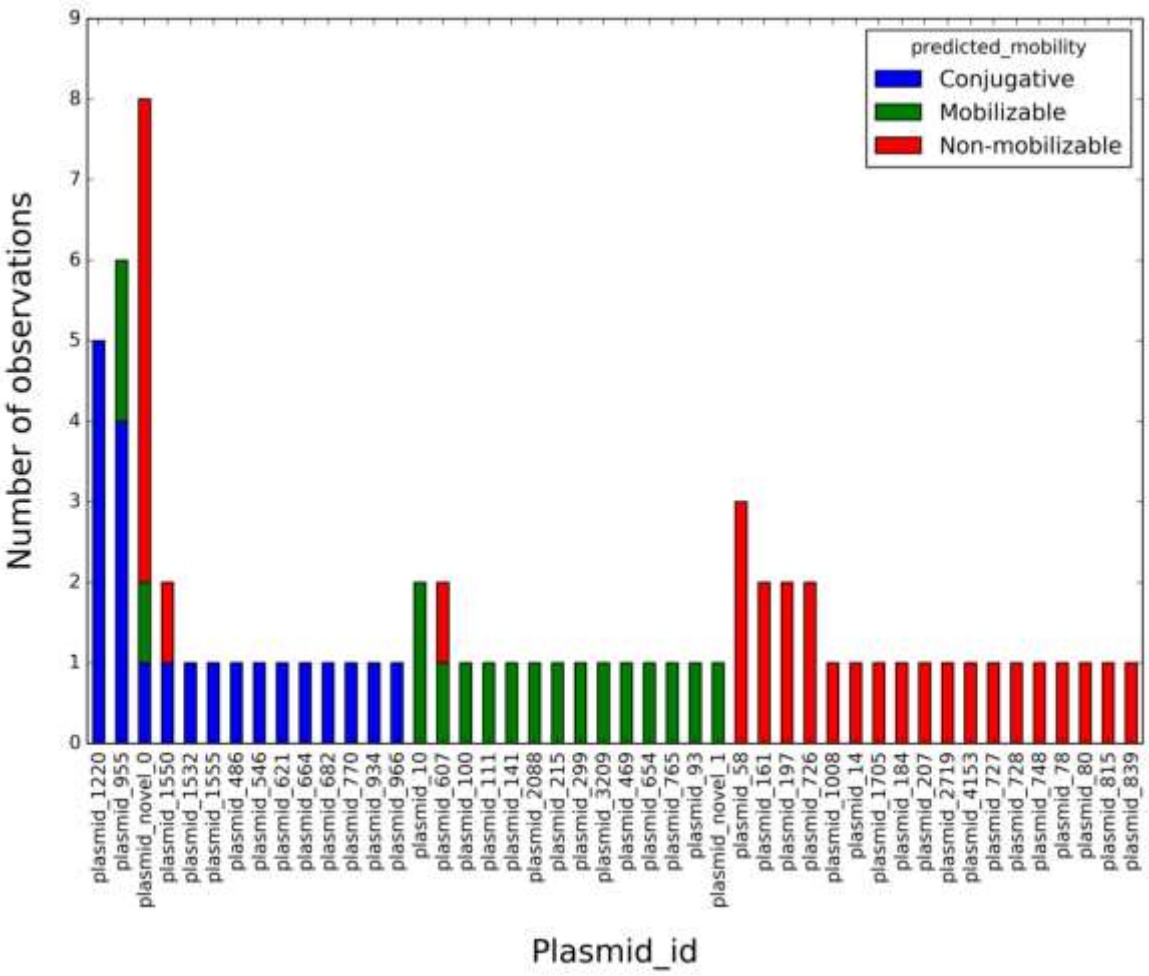
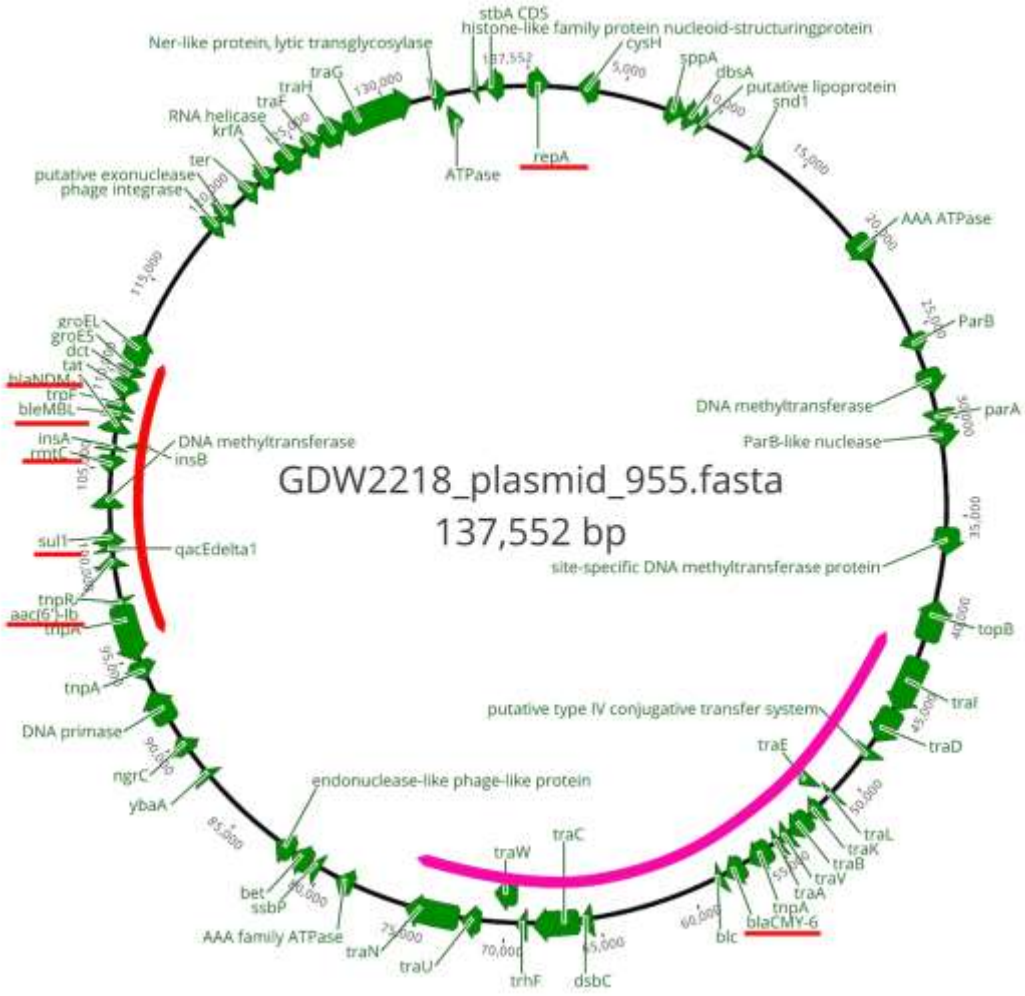
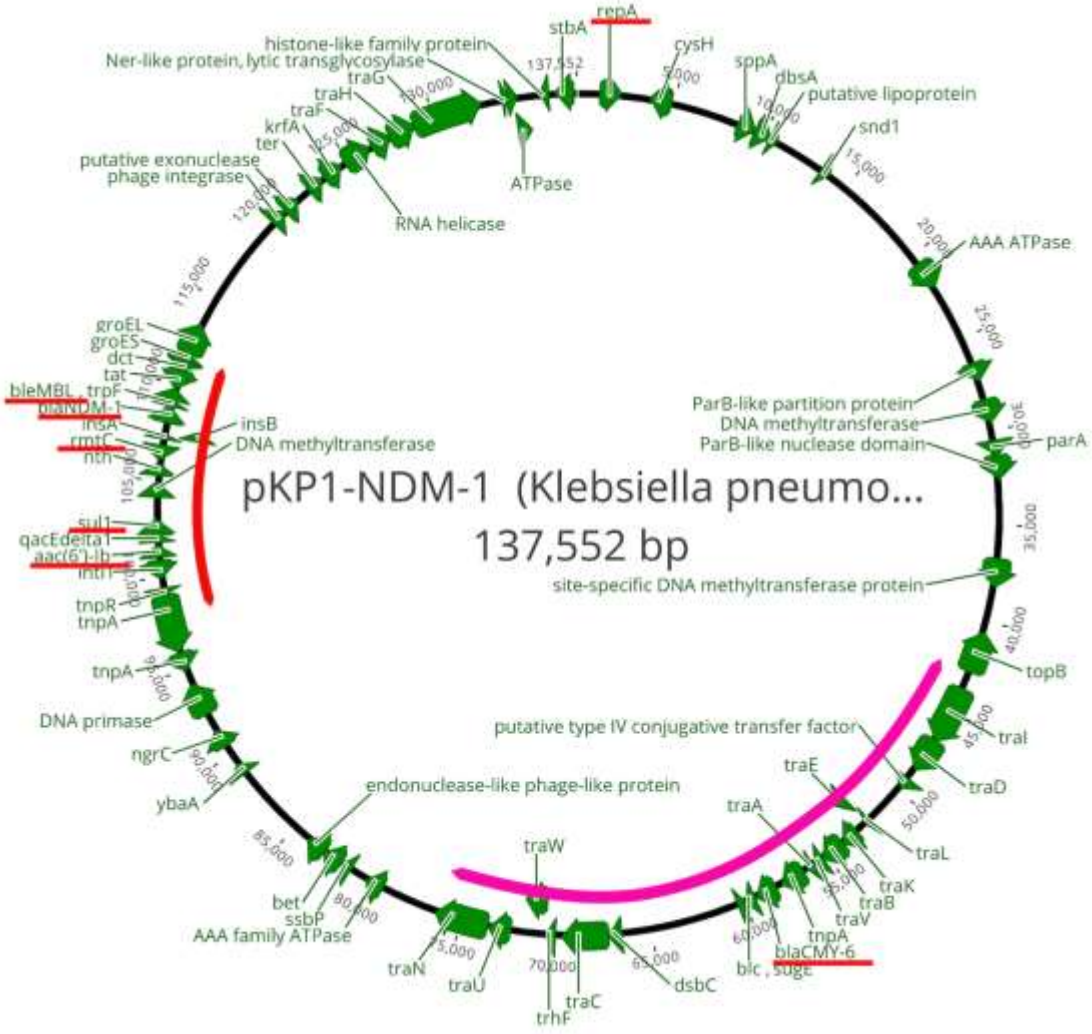


Figure 3.5: **Predicted MOB-Suite plasmid cluster_id from NDM isolates.** The identity of each plasmid is provided by MOB-Suite mash clusters and contains the mash neighbor (reference) plasmids used for predictions. A) plasmids predicted from Illumina assembly data only: 48 plasmids are non-mobilizable, 24 plasmids are mobilizable, and 16 plasmids are conjugative. B) plasmids predicted from Illumina + Nanopore hybrid assembly data: 31 plasmids are non-mobilizable, 18 plasmids are mobilizable, and 21plasmids are conjugative.

A)


GDW2218_plasmid_955.fasta
137,552 bp
stbA CDS
histone-like family protein nucleoid-structuringprotein
Ner-like protein, lytic transglycosylase
cysH
sppA
dbsA
putative lipoprotein
snd1
traG
traH
traF
RNA helicase
krfA
ter
putative exonuclease
phage integrase
ATPase
repA
AAA ATPase
ParB
DNA methyltransferase
parA
ParB-like nuclease
site-specific DNA methyltransferase protein
topB
traI
traD
putative type IV conjugative transfer system
traE
traL
traK
traB
traV
traA
tnpA
blaCMY-6
blc
traC
traW
endonuclease-like phage-like protein
bet
ssbP
AAA family ATPase
traN
traU
trhF
dsbC
ybaA
ngrC
DNA primase
tnpA
tnpR
aac(6')-Ib
tnpA
sul1
qacEdelta1
DNA methyltransferase
insB
insA
rmtC
bleMBL
tnpR
groEL
groES
dct
tat
blaNDM-1

B)


pKP1-NDM-1 (Klebsiella pneumo...
137,552 bp
repA
stbA
cysH
sppA
dbsA
putative lipoprotein
snd1
AAA ATPase
parA
topB
traI
traD
traL
traK
traB
traV
tnpA
blaCMY-6
dsbC
traC
traF
traU
traN
ssbP
AAA family ATPase
bet
endonuclease-like phage-like protein
ybaA
ngrC
DNA primase
tnpA
tnpA
tnpR
intI1
qacEdelta1
sul1
nth
rmtC
insA
blaNDM-1
bleMBL
trpF
tat
dct
groES
groEL
insB
DNA methyltransferase
putative exonuclease
phage integrase
ter
krfA
traF
traH
traG
Ner-like protein, lytic transglycosylase
histone-like family protein
ATPase
RNA helicase
ParB-like partition protein
DNA methyltransferase
ParB-like nuclease domain
site-specific DNA methyltransferase protein
putative type IV conjugative transfer factor
traE
traA
traW
trhF

C)

D)

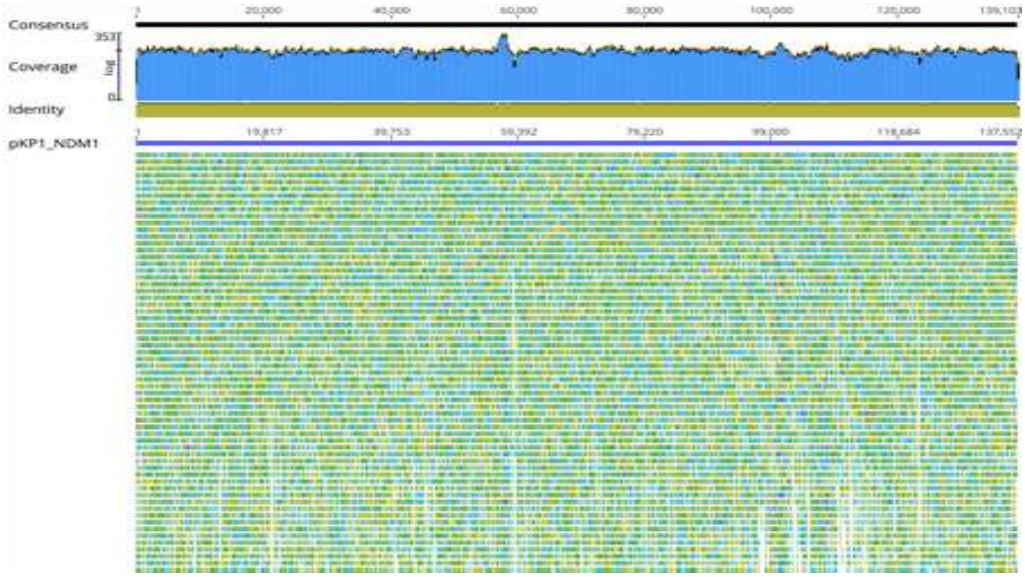
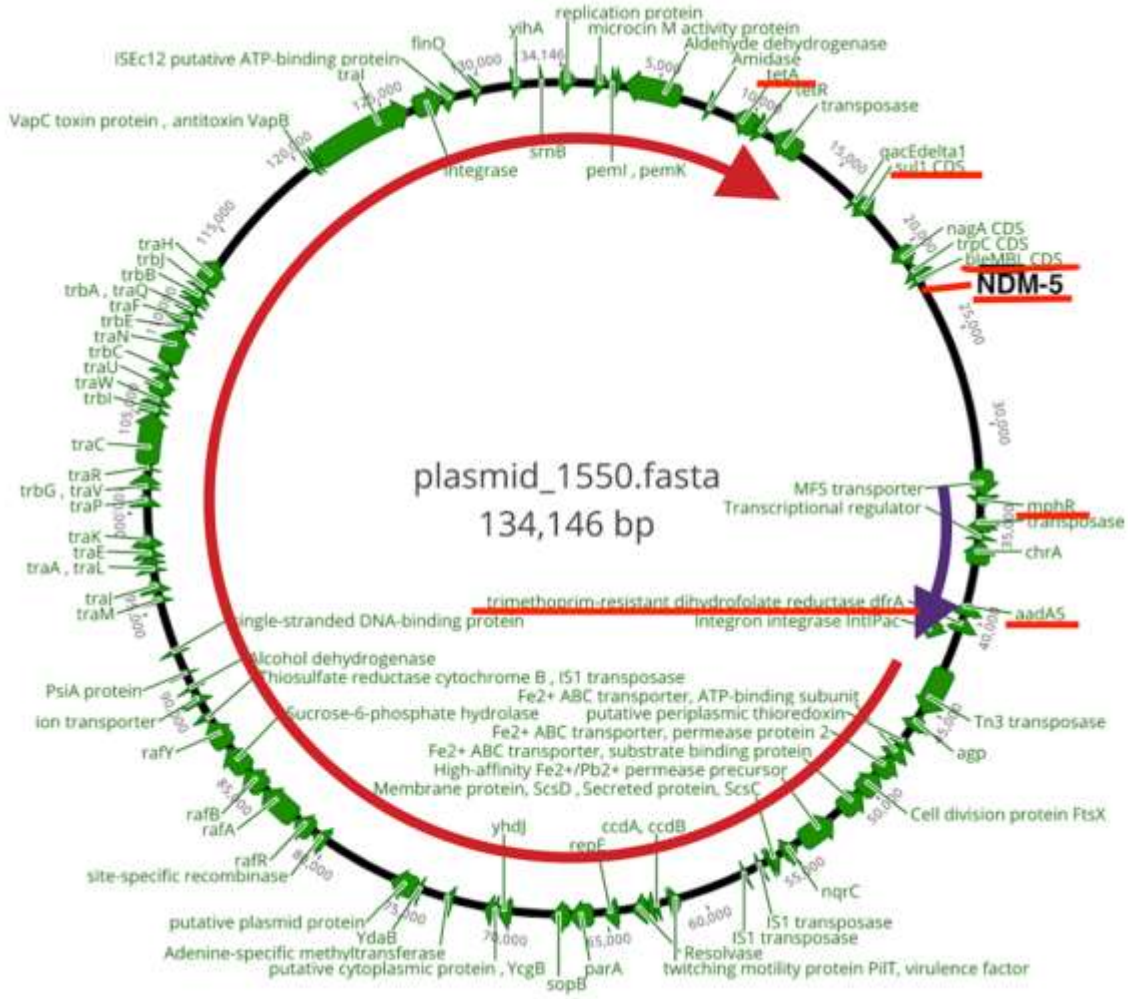
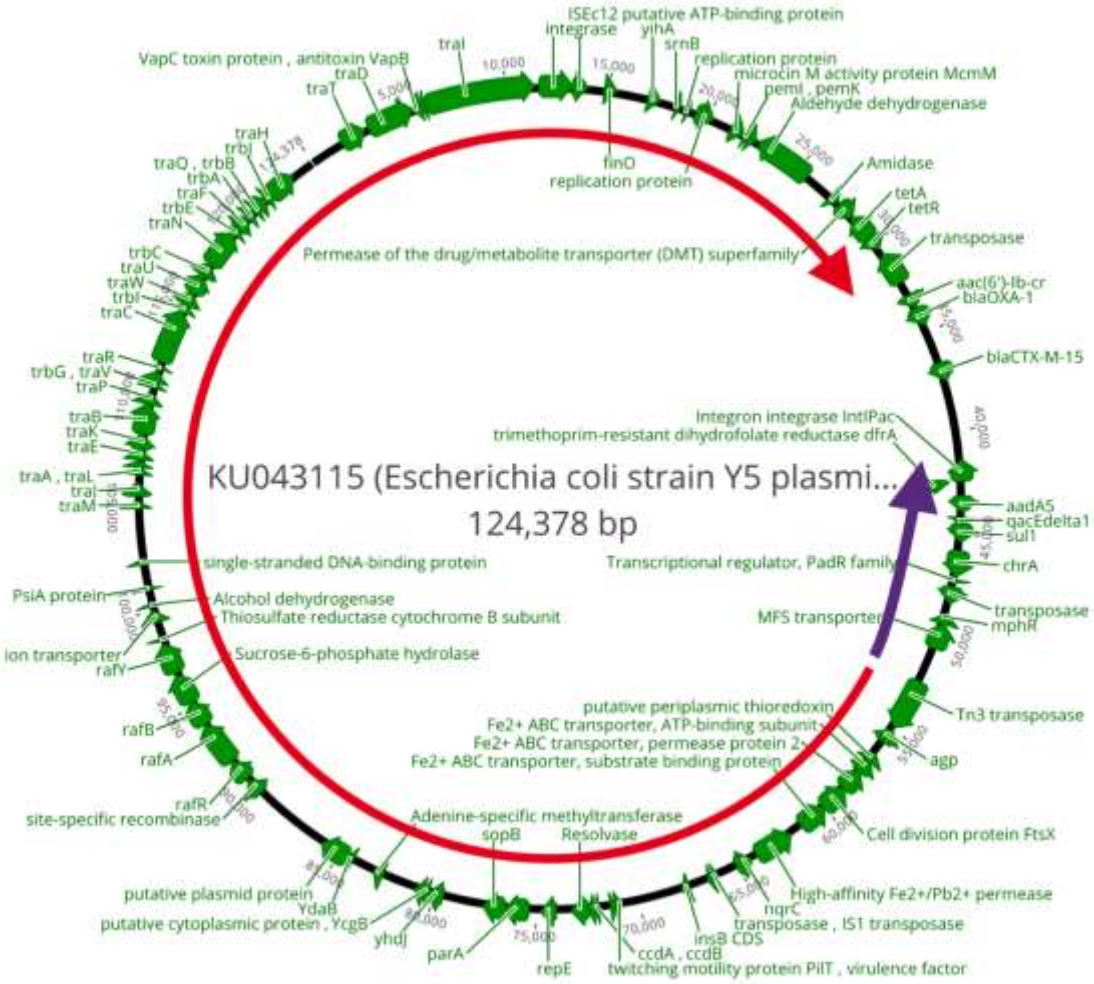


Figure 3.6: **Annotation and Illumina read mapping for the predicted NDM-associated plasmid from GDW2218 *Klebsiella pneumoniae* isolate and reference plasmid pKP-NDM1**. The two plasmids are identical in overall nucleotide sequence, AMR gene content, and Illumina read mapping coverage. A) and B) annotated GDW2218 plasmid and pKP-NDM1. C) and D) Illumina read mapping to the GDW2218 plasmid and pKP-NDM1, with 43,949 reads mapped to the GDW2218 plasmid and 43,946 reads mapped to pKP-NDM1.

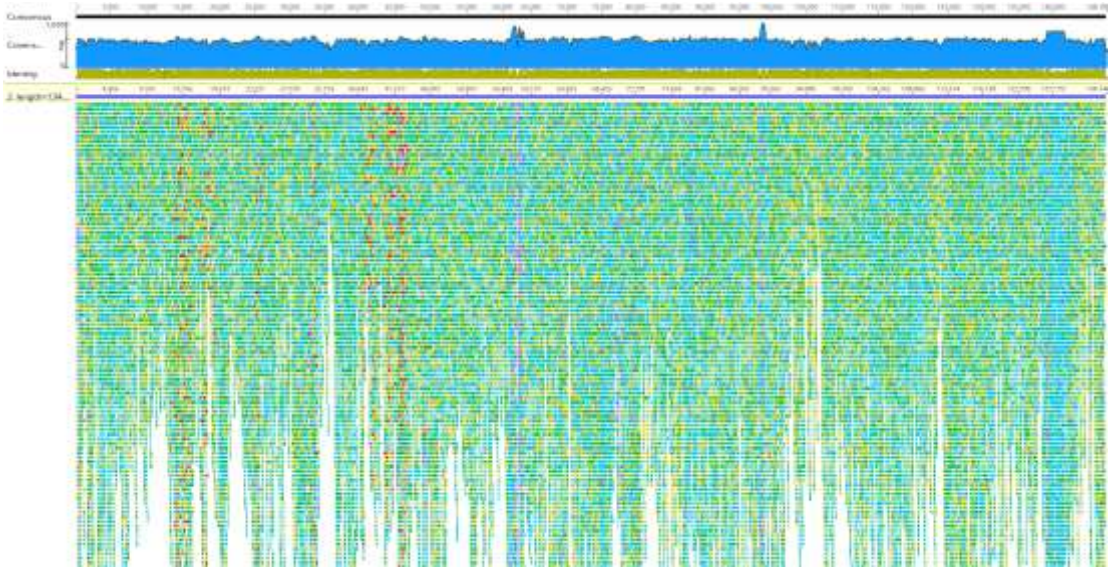
A)


plasmid_1550.fasta
134,146 bp
replication protein
microcin M activity protein
yihA
finO
ISEc12 putative ATP-binding protein
traI
VapC toxin protein , antitoxin VapB
integrase
srnB
pemI , pemK
Aldehyde dehydrogenase
Amidase
tetA
tetR
transposase
qacEdelta1
nagA CDS
trpC CDS
NDM-5
traH
trbJ
trbB
trbA , traQ
traF
trbE
traN
trbC
traU
traW
trbI
traC
traR
trbG , traV
traP
traK
traE
traA , traL
traI
traM
MFS transporter
Transcriptional regulator
mphR
transposase
chrA
trimethoprim-resistant dihydrofolate reductase dfrA
Integron integrase IntI1Pac
aadA5
ingle-stranded DNA-binding protein
Alcohol dehydrogenase
Thiosulfate reductase cytochrome B , IS1 transposase
PsiA protein
Fe2+ ABC transporter, ATP-binding subunit
ion transporter
Sucrose-6-phosphate hydrolase
putative periplasmic thioredoxin
Tn3 transposase
Fe2+ ABC transporter, permease protein 2
rafY
Fe2+ ABC transporter, substrate binding protein
agp
High-affinity Fe2+/Pb2+ permease precursor
Membrane protein, ScsD , Secreted protein, ScsC
Cell division protein FtsX
rafB
rafA
yhdJ
ccdA, ccdB
repE
rafR
site-specific recombinase
nqrC
putative plasmid protein
YdaB
IS1 transposase
Adenine-specific methyltransferase
IS1 transposase
Resolvase
putative cytoplasmic protein , YcgB
parA
twitching motility protein PilT, virulence factor
sopB

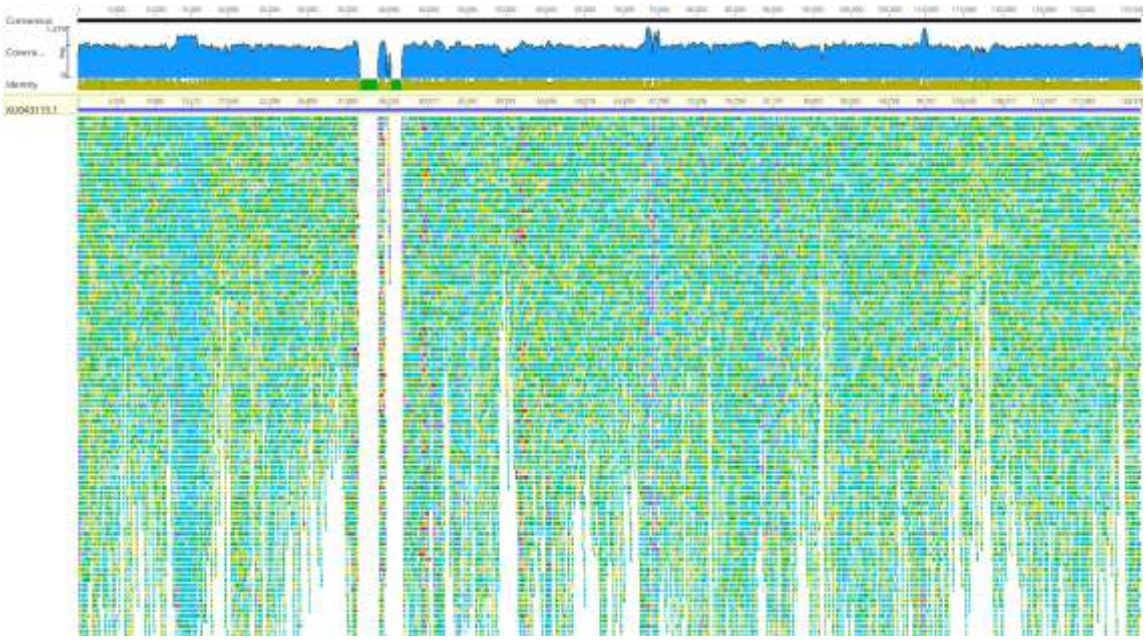
B)


KU043115 (Escherichia coli strain Y5 plasmi...
124,378 bp
ISEc12 putative ATP-binding protein
integrase
yihA
srnB
replication protein
microcin M activity protein McmM
pemI , pemK
Aldehyde dehydrogenase
Amidase
tetA
tetR
transposase
aac(6')-Ib-cr
blaOXA-1
blaCTX-M-15
Integron integrase IntIPac
trimethoprim-resistant dihydrofolate reductase dfrA
aadA5
qacEdelta1
sul1
chrA
transposase
mphR
MFS transporter
Tn3 transposase
agp
Cell division protein FtsX
High-affinity Fe2+/Pb2+ permease
nqrC
transposase , IS1 transposase
insB CDS
ccdA , ccdB
twitching motility protein PilT , virulence factor
repE
parA
putative cytoplasmic protein , YcgB
putative plasmid protein
YdaB
yhdJ
site-specific recombinase
rafR
rafB
rafA
rafY
ion transporter
PsiA protein
single-stranded DNA-binding protein
Alcohol dehydrogenase
Thiosulfate reductase cytochrome B subunit
Sucrose-6-phosphate hydrolase
Adenine-specific methyltransferase
sopB
Resolvase
putative periplasmic thioredoxin
Fe2+ ABC transporter, ATP-binding subunit
Fe2+ ABC transporter, permease protein 2
Fe2+ ABC transporter, substrate binding protein
Transcriptional regulator, PadR family
Permease of the drug/metabolite transporter (DMT) superfamily
finO
replication protein
traI
VapC toxin protein , antitoxin VapB
traD
traT
traH
trbI
traQ , trbB
trbA
traF
trbE
traN
trbC
traU
traW
trbI
traC
traR
trbG , traV
traP
traB
traK
traE
traA , traL
traI
traM

C)



D)



E)

F)

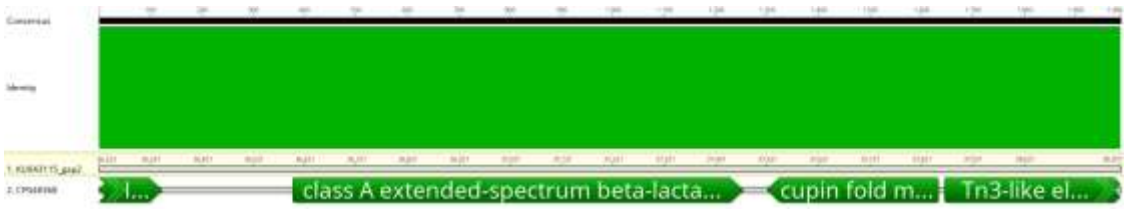
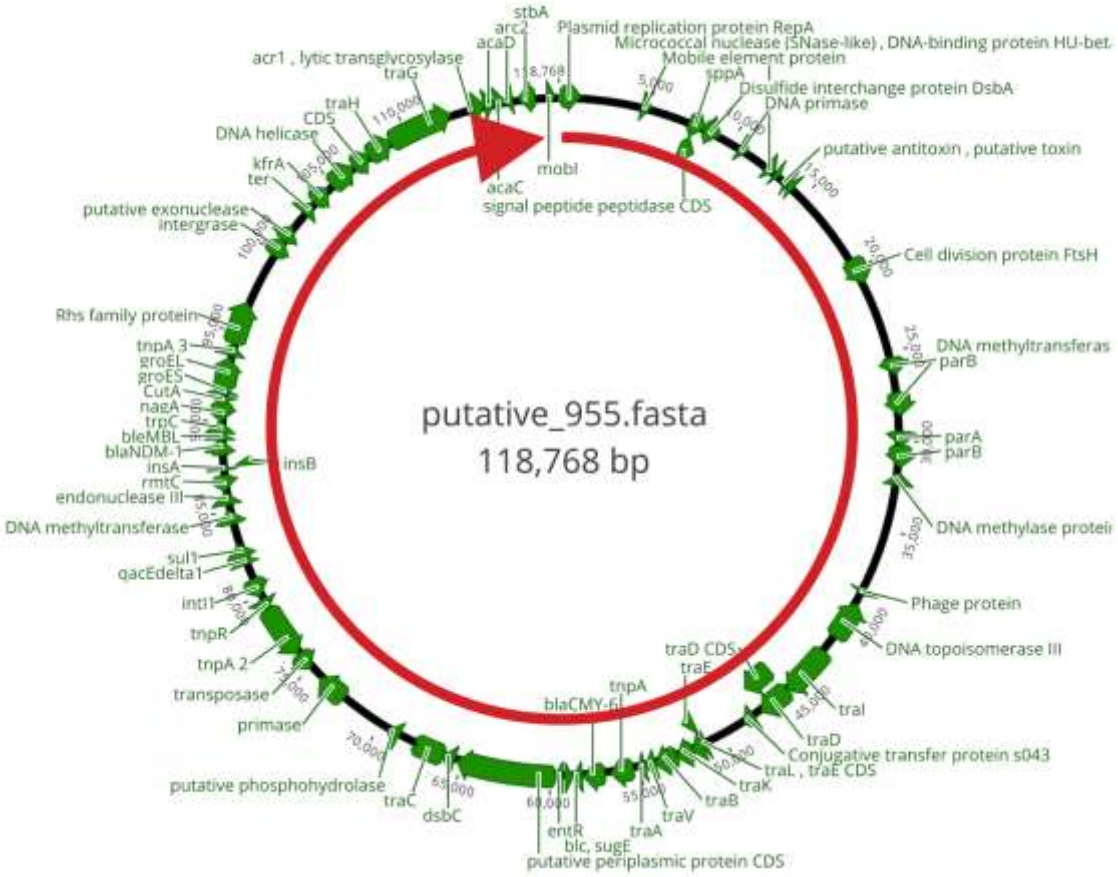
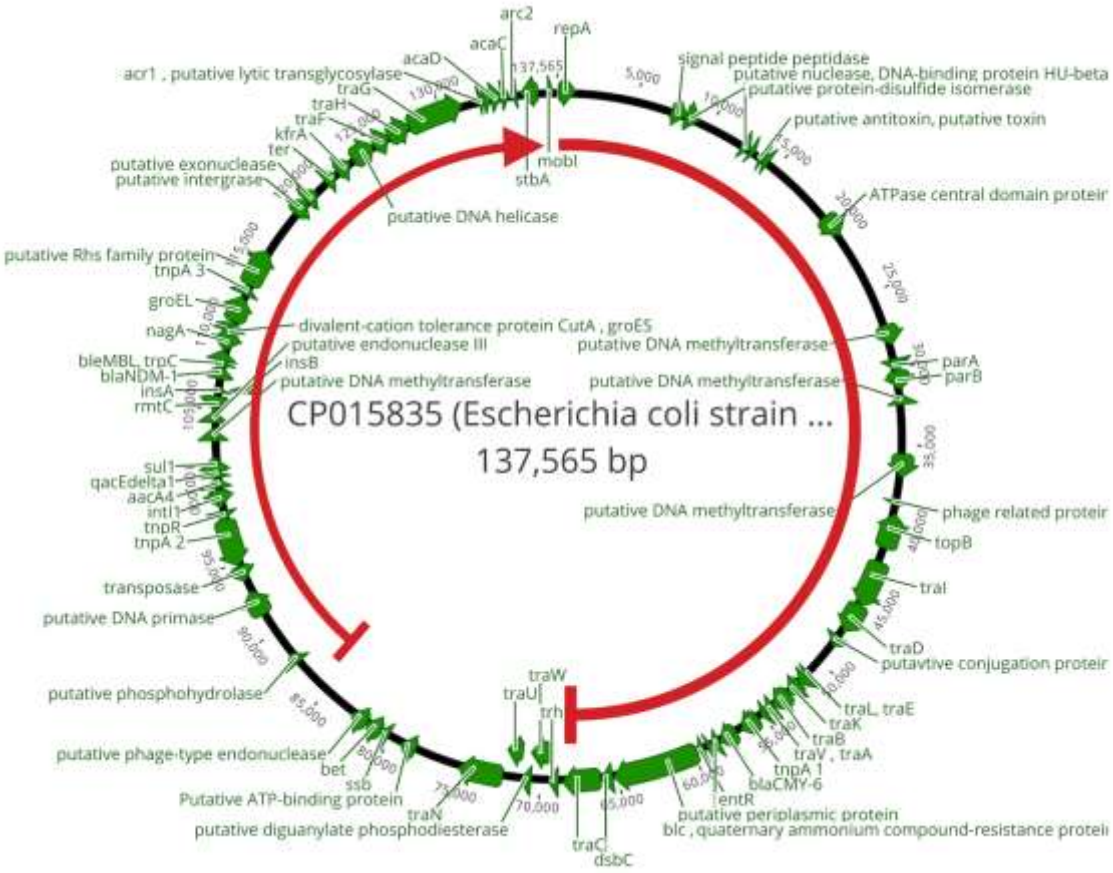


Figure 3.7: **Annotation and Illumina read mapping for the predicted NDM-associated plasmid from C0648 *Escherichia coli* isolate and reference plasmid pECY55**. The multi-drug resistance NDM5 gene is not present in plasmid pECY55. A) and B) Annotated C0648 plasmid and plasmid pECY55 C). 57,614 reads map uniformly to the **C0648** plasmid. D). 53,789 reads map to plasmid pECY55, where there are two regions where the reads do not map. E) The first unmapped region (2,409 bp) of plasmid pECY55 which encodes aminoglycoside AAC (6') Ib-cr gene, blaOXA-1, and catB3 all located on a 21Kbp resistance island. F) The second unmapped region (1906 bp) of plasmid pECY55 encodes an IS 1380 transposase, AMR gene CTX-M-15, cupin fold metalloprotein, and Tn3 like transposase. pECY55 additionally does not encode an NDM gene.

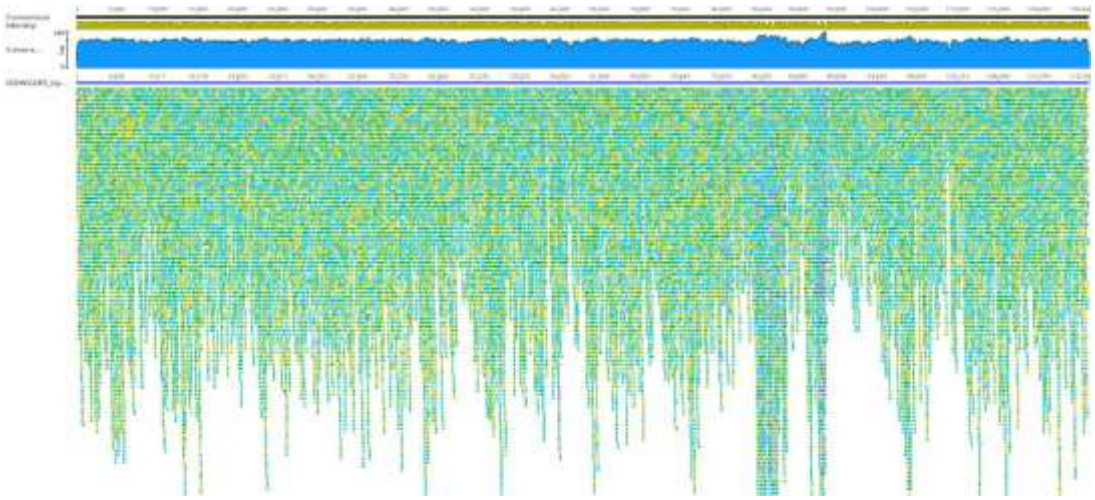
A)


putative_955.fasta
118,768 bp
stbA
arc2
acaD
Plasmid replication protein RepA
Micrococcal nuclease (SNase-like), DNA-binding protein HU-bet
Mobile element protein
sppA
Disulfide interchange protein DsbA
DNA primase
putative antitoxin, putative toxin
Cell division protein FtsH
DNA methyltransferas
parB
parA
parB
DNA methylase protei
Phage protein
DNA topoisomerase III
traD CDS
traF
tral
traD
Conjugative transfer protein s043
traL, traE CDS
traK
traB
traV
traA
tnpA
blaCMY-6
entR
blc, sugE
putative periplasmic protein CDS
traC
dsbC
putative phosphohydrolase
tnpA 2
transposase
primase
tnpR
intI1
qacEdelta1
sul1
endonuclease III
DNA methyltransferase
rmtC
insA
blaNDM-1
bleMBL
trpC
nagA
CutA
groES
groEL
tnpA 3
Rhs family protein
putative exonuclease
intergrase
kfrA
ter
DNA helicase
CDS
traH
traG
acr1, lytic transglycosylase
acaC
mobI
signal peptide peptidase CDS
insB

B)


CP015835 (Escherichia coli strain ...
137,565 bp
arc2
repA
acaC
acaD
137,565
acr1 , putative lytic transglycosylase
traG
traH
traF
kfrA
ter
putative exonuclease
putative intergrase
putative DNA helicase
mobI
stbA
signal peptide peptidase
putative nuclease, DNA-binding protein HU-beta
putative protein-disulfide isomerase
putative antitoxin, putative toxin
ATPase central domain protein
putative Rhs family protein
tnpA 3
groEL
nagA
divalent-cation tolerance protein CutA , groES
putative endonuclease III
insB
putative DNA methyltransferase
bleMBL, trpC
blaNDM-1
insA
rmtC
putative DNA methyltransferase
putative DNA methyltransferase
parA
parB
sul1
qacEdelta1
aacA4
intI1
tnpR
tnpA 2
putative DNA methyltransferase
phage related proteir
topB
transposase
putative DNA primase
traI
traD
putavtive conjugation proteir
putative phosphohydrolase
traW
traU
trh
traL, traE
traK
traB
traV , traA
tnpA 1
blaCMY-6
entR
putative periplasmic protein
blc , quaternary ammonium compound-resistance protei
putative phage-type endonuclease
bet
ssb
Putative ATP-binding protein
traN
putative diguanylate phosphodiesterase
traC
dsbC

C)



D)

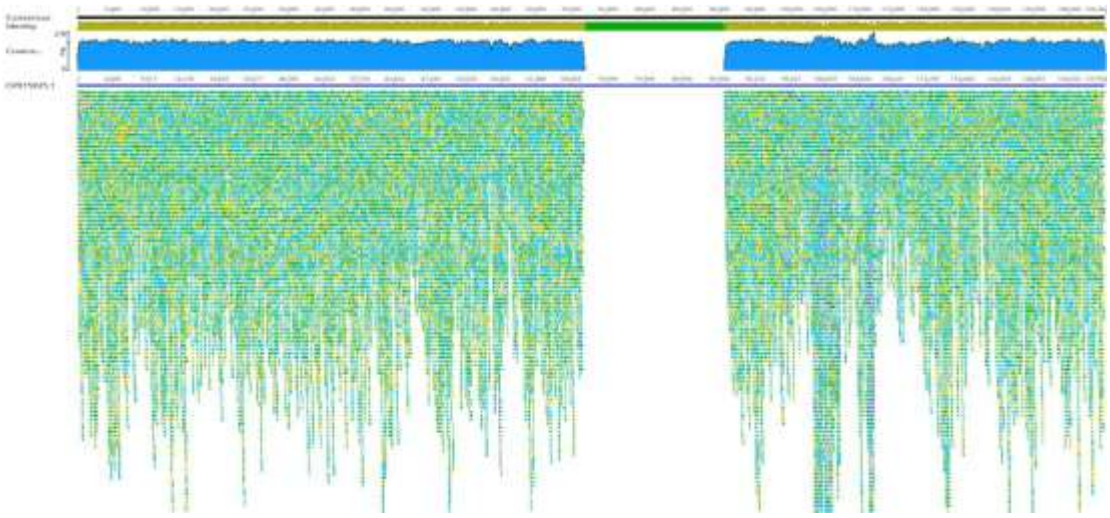
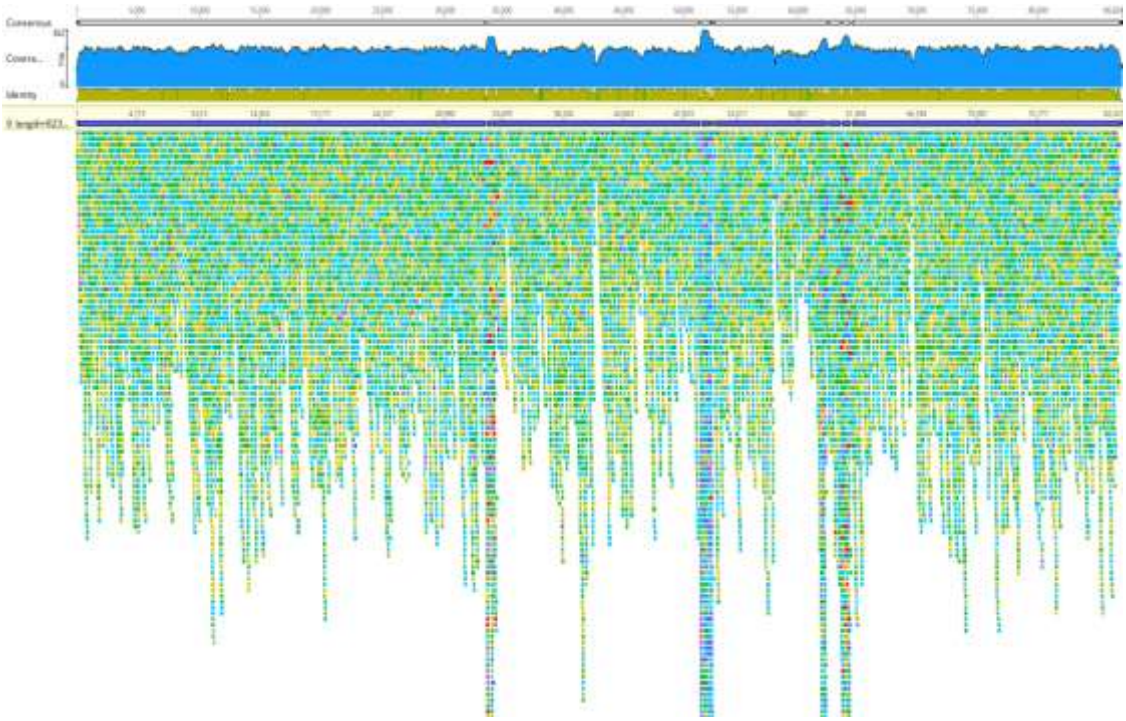
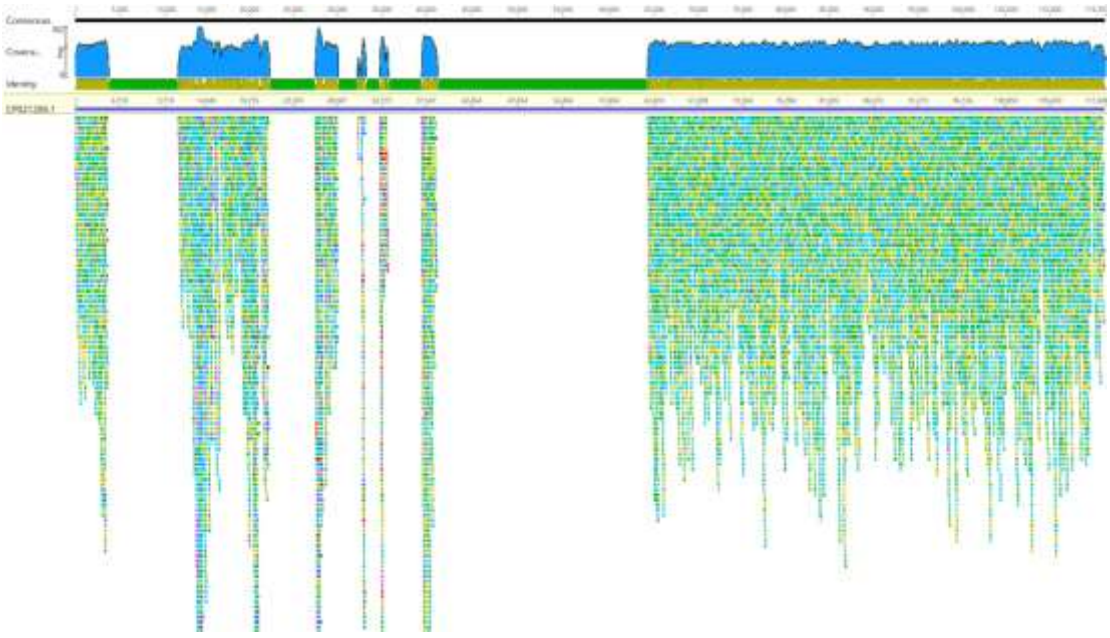


Figure 3.8: **Annotation and Illumina read mapping for the predicted NDM-associated plasmid from GDW2285 *Escherichia coli* isolate and reference plasmid pMS6198A**. A) and B) annotated GDW2285 plasmid and plasmid pMS6198A. C) and D) Illumina read mapping to the GDW2285 plasmid and pMS6198A, with 32,438 reads mapped to the GDW2285 plasmid and 32,401 reads mapped to pMS6198A. The unmapped region in pMS6198A is around 18 kbp in size and it encodes mobility genes but not AMR genes.

A)



B)



C)

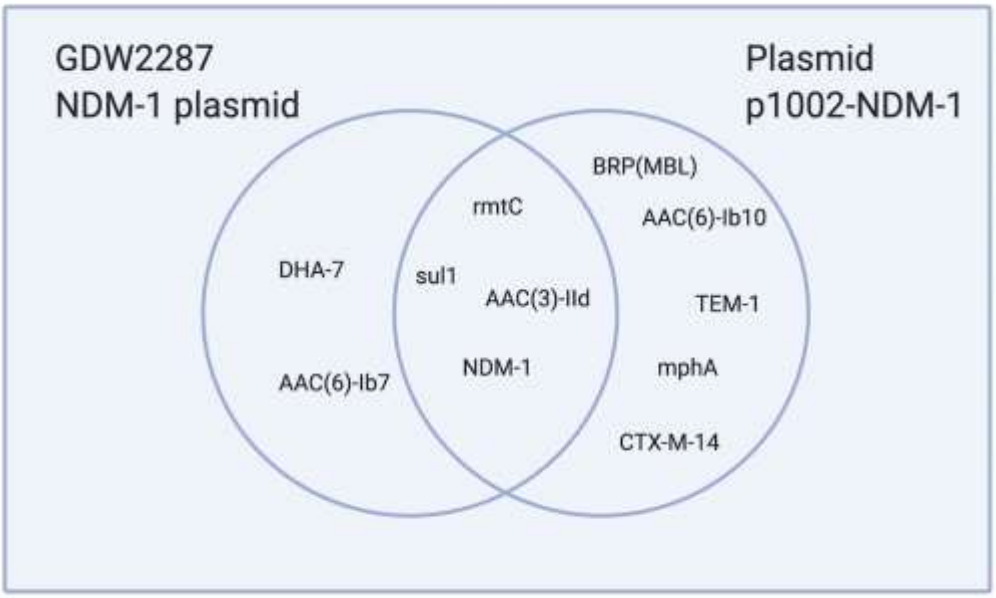


Figure 3.9: **Illumina read mapping and relationship of predicted AMR genes between NDM-associated plasmid from GDW2287 *Klebsiella pneumoniae* isolate and reference plasmid p1002-NDM1.** A) and B) read mapping for GDW2218 and p1002-NDM1, respectively. C). Venn diagram of the shared and unshared AMR genes between the GDW2287 NDM-1 plasmid and p1002-NDM1.

Table 3.1**: WGS and RGI summary of the NDM-associated clinical genome isolates.** Each isolate has a unique ID and all isolates were collected and sequenced in four rounds. The site of infection is known for the C0612, C0650 *Klebsiella pneumoniae* strains, and C0648 *Escherichia coli* strain, but not for the GDW strains. The isolates are classified into two different taxon families and six different taxon species based on clinical identification. All identified NDM genes (bolded) were RGI Perfect hits.

Isolate ID	Sites of Infection	Clinical Family Identification	Clinical Species Identification	RGI Perfect hits	RGI Strict hits
C0612	Blood	Enterobacteriaceae	*Klebsiella pneumoniae*	**NDM-1**, SHV-1, oqxA, CTX-M-14, BRP(MBL), APH(3')-VI, QnrS1, TEM-1, OXA-9, tet(D), mphA, CTX-M-15	acrA. ampH, msbA, OmpK37, KpnE, KpnF, marA, CRP, adEF, emrR, KpnG, KnpH, FosA6, H-NS, adeF, baeR, APH(6)-Id, dfrA14, AAC(6')-Ib10, AAC(3)-IIe, APH(6)-Id, , APH(3'')-Ib, PBP3, ParC, gyrA, UhpT, EF-Tu, marR
C0648	Rectal	Enterobacteriaceae	*Escherichia coli*	msbA, marA, PmrF, emrB, emrA, mdtP, acrD, ampC1, H-NS, acrA, acrB, kdpE, bacA, mdtG, mdtH, cpxA, emrY, emrK, evgA, ampH, CMY-42, AcrF, rmtB, TEM-1, BRP(MBL), **NDM-5**, dfrA12, aadA5, sul1	mdfA, KpnE, KpnF, emrR, ampC, eptA, mdtN, mdtO, mphB, CRP, baeR, baeS, mdtC, evgS, mdtM, AcrE, tet(A), aadA, dfrA17, gyrA, PBP3, ParC, marR, soxR, soxS, acrR

C0650	Rectal	Enterobacteriaceae	*Klebsiella pneumoniae*	SHV-27, oqxA, KpnF, sul1, DHA-1, QnrB4, **NDM-1**, APH(3')-VI, QnrS1, CTX-M-15, OXA-232	marA, H-NS, adeF, ampH, KpnH, KpnG, emrR, adeF, OmpK37, KpnE, CRP, msbA, FosA6, baeR, OXA-9, aadA, AAC(6')-Ib10, TEM-55, CMY-33, PBP3, marR
GDW2216	N/A	Enterobacteriaceae	*Klebsiella pneumoniae*	oqxA, SHV-11, KpnF, AAC(3)-IId, rmtC, **NDM-1**, CTX-M-15, OXA-1, sul1	adeF, emrR, KpnG, H-NS, baeR, FosA6, acrA, ampH, msbA, CRP, marA, OmpK37, KpnE, BRP(MBL), DHA-1, catB3, AAC(6')-Ib10, parC, gyrA, PBP3, UhpT, marR, NfxB
GDW2218	N/A	Enterobacteriaceae	*Klebsiella pneumoniae*	oqxA, oqxB, KpnF, CTX-M-15, BRP(MBL), **NDM-1**, CMY-6, rmtC, sul1	acrA, H-NS, adeF, emrR, KpnG, KpnH, KpnE, OmpK37, baeR, marA, SHV-89, FosA6, msbA, dfrA14, AAC(6')-Ib10, QnrB17, SHV-1, UhpT, marR
GDW2280	N/A	Enterobacteriaceae	*Klebsiella pneumoniae*	oqxA, sul2, KpnF, BRP(MBL), **NDM-1**, armA, msrE, mphE, QnrS1, TEM-1, OXA-1, AAC(6')-Ib-cr, CTX-M-15, dfrA12, sul1	msbA, OmpK37, ampH, adeF, emrR, KpnG, H-NS, baeR, FosA, marA, KpnE, tet(A), APH(6')-Id, APH(3'')-Ib, catB3, aadA2, SHV-64, dfrA1, SHV-89, AAC(6')-Ib7, PBP3, parC, UhpT

GDW2281	N/A	Morganellaceae	*Morganella morganii*	DHA-1, CTX-M-15, AAC (3)-IId, dfrA12, aadA2, sul1, **NDM-1**, sul1, armA, msrE, mphE	KpnH, catII, K-12, tet(B), mexM, MexC, PBP3, tetR, nalD
GDW2284	N/A	Morganellaceae	*Providencia stuartii*	**NDM-1**, BRP(MBL), CMY-6, sul1, rmtC, AAC (3)-IId, APH(3')-VI	KpnF, tet(D), AAC(2')-Ia, CRP, AAC(6')-Ib10, KpnH, adeF, PBP3,
GDW2285	N/A	Enterobacteriaceae	*Escherichia coli*	emrB, emrR, evgA, baeS, mdtE, mdtH, mdtG, ampH, acrB, acrA, msbA, AcrE, TolC, **NDM-1**, BRP(MBL), marA, cpxA, CMY-6, catI, rmtC, OXA-2, TEM-166, OXA-1, sul1, CTX-M-15	emrA, emrY, emrK, evgS, acrD, mphB, KpnE, mdtO, PmrF, mdtM, AAC(3')-IIe, APH(3'')-Ib, APH(6)-Id, catB3, AAC(6')-Ib-cr, AAC(6')-Ib10, gyrA, marR
GDW2286	N/A	Enterobacteriaceae	*Citrobacter freundii*	CMY-46, BRP(MBL), **NDM-1**, APH(3')-VI, mphA, AAC(3)-IId, sul2, rmtC, dfrA12, OXA-1, sul1, CTX-M-15, CMY-6, TEM-1	emrR, emrB, H-NS, QnrB6, tet(A), acrA, ampH, AAC(3)-IIe, aadA2, catB3, CMY -74, AAC(6')-Ib10, APH(3')-IIb,

GDW2287	N/A	Enterobacteriaceae	*Klebsiella pnemoniae*	oqxA, SHV-11, KpnF, AAC(3)-IId, rmtC, **NDM-1**, CTX-M-15, OXA-1, sul1	adeF, emrR, KpnG, KpnH, H-NS, FosA6, acrA, ampH, msbA, marA, OmpK37, BRP(MBL), DHA-1, catB3, AA(6')-Ib10, MuxC, MuxA, AAC(6')-Ib-cr, parC, gyrA, PBP3, UhpT, marR
GDW2288	N/A	Enterobacteriaceae	*Enterobacter cloacae*	CTX-M-15, ACT-25, BRP(MBL), **NDM-1**, OXA-9, rmtC, TEM-1	oqxA, adeF, ramA, ampH, H-NS, marA, FosA2, CRP, adeF, bacA, KpnE, sul1, AAC(3)-IIe, AAC(6')-Ib10, UhpT, marR
GDW2289	N/A	Enterobacteriaceae	*Escherichia coli*	emrY, emrK, evgA, TolC, msbA, baeR, emrB, emrR, mdtH, mdtG, marA, H-NS, **NDM-1**, mphA, aadA5, sul1, CTX-M-15, AAC(6')-Ib-cr, OXA-1, TEM-1, dfrA5	yojI, PmrF, bacA, AcrS, AcrE, KpnF, ampC, eptA, mdtP, baeS, ugd, BRP(MBL), mdtE, tet(B), parC, PBP3, soxR, AAC(3)-IIe, dfrA17
GDW2290	N/A	Enterobacteriaceae	*Klebsiella oxytoca*	TEM-1, CTX-M-15, QnrS1, APH(3')-VI, **NDM-1**, mphA	KpnE, KpnF, marA, adeF, oqxA, Oxy-2-8, adeF, msbA, CRP, OXA-9, aadA, AAC(6')-Ib10, BRP(MBL), gyrB, UhpT, PBP3

GDW2291	N/A	Enterobacteriaceae	*Enterobacter cloacae*	ACT-25, APH(3')-VI, **NDM-1**, sul2, armA, sul1, aadA2, dfrA12, CTX-M-15, AAC(3)-IId, QnrB1, AAC(6')-Ib-cr, OXA-1, TEM-1	H-NS, KpnE, msbA, ramA, ampH, baeR, adeF, FosA2, PBP3, UhpT

Table 3.2**: Whole genome assembly statistics for NDM-associated isolates**.

Isolate ID	**Clinical species ID**	**Number of raw reads**	**Number of assembled contigs**	**Total assembly length (bp)**	**GC content**	**Assembly N50 (bp)**
C0612	*Klebsiella pneumoniae*	918,031	100	5,613,729	0.5699	330,832
C0648	*Escherichia coli*	13,666,670	174	5,140,713	0.5073	123,605
C0650	*Klebsiella pneumoniae*	1,330,379	116	5,545,533	0.5697	372,277
GDW2216	*Klebsiella pneumoniae*	1,846,772	79	5,513,876	0.5725	262,587
GDW2218	*Klebsiella pneumoniae*	2,118,962	92	5,595,136	0.5721	266,834
GDW2280	*Klebsiella pneumoniae*	3,116,160	81	5,788,657	0.5687	309,438
GDW2281	*Morganella morganii*	1,888,462	62	4,018,549	0.5095	192,419
GDW2284	*Providencia stuartii*	5,367,558	72	4,388,572	0.4107	269,524
GDW2285	*Escherichia coli*	1,460,414	165	5,165,710	0.5064	134,636
GDW2286	*Citrobacter freundii*	1,844,888	204	5,169,088	0.5213	320,922
GDW2287	*Klebsiella pneumoniae*	1,638,437	98	5,522,538	0.5725	251,151
GDW2288	*Enterobacter cloacae*	1,578,205	139	4,955,765	0.5491	170,350
GDW2289	*Escherichia coli*	1,501,266	214	5,560,603	0.5062	95,996
GDW2290	*Klebsiella oxytoca*	1,598,213	79	6,288,238	0.5479	398,977
GDW2291	*Enterobacter cloacae*	1,662,158	84	4,899,849	0.5501	305,605

Table 3.3**: Predicted replicon and relaxase gene from Illumina assembly NDM-associated plasmids, plus NCBI host-range predictions generated using RGI:Mobilome**.

.

Illumina NDM plasmid cluster_id (isolate_id)	**Replicon_types**	**Relaxase_types**	**Predicted NCBI-HR-Name**
Plasmid 726 (C0612)	IncFIA,IncR,rep_cluster_1418	-	Enterobacteriaceae
Plasmid 1550 (C0648)	IncFIA,IncFII,IncFIIA	MOBF	Enterobacterales
Plasmid 1220 (C0650)	IncFIB,IncFII	MOBF	Enterobacterales
Plasmid 955 (GDW2216)	IncA/C2	-	Gammaproteobacteria
Plasmid 955 (GDW2218)	IncA/C2	MOBH	Gammaproteobacteria
Plasmid 966 (GDW2280)	IncA/C2,rep_cluster_1254	MOBH	Gammaproteobacteria
Plasmid 2719 (GDW2281)	n/a	n/a	n/a
Plasmid 955 (GDW2284)	IncA/C2	MOBH	Gammaproteobacteria
Plasmid 955 (GDW2285)	IncA/C2	MOBH	Gammaproteobacteria
Plasmid 955 (GDW2286)	IncA/C2	MOBH	Gammaproteobacteria
Plasmid 955 (GDW2287)	IncA/C2	-	Gammaproteobacteria
Plasmid 621 (GDW2288)	IncFIB,IncFII,IncFII	MOBP	Enterobacterales
Plasmid 934 (GDW2289)	IncN	MOBF	Gammaproteobacteria
Plasmid 1220 (GDW2290)	IncFIB,IncFII,IncFII	MOBF	Enterobacterales
Plasmid 654 (GDW2291)	rep_cluster_1444	n/a	Enterobacteriaceae

Table 3.4**: Nanopore sequencing statistics for the NDM-associated isolates**.

Isolate ID	**Clinical species ID**	**Number of reads**	**Mean read length**	**Mean read quality**	**Total bases (bp)**	**Read length N50**
C0612	*Klebsiella pneumoniae*	18,458	6,086	11.2	112,341,501	13,015
C0648	*Escherichia coli*	29596	8,409.5	12.1	248,887,693	13,364
C0650	*Klebsiella pneumoniae*	24263	6,646.5	11.7	161,265,242	11,351
GDW2216	*Klebsiealla pneumoniae*	301,286	2,889	10.5	1,497,729,366	9,360
GDW2218	*Klebsiella pneumoniae*	26,587	8,554.8	11.5	227,447,521	10,376
GDW2280	*Klebsiella pneumoniae*	7,122	12,753	11.4	90,833,472	22,792
GDW2281	*Morganella morganii*	421,363	5,808.1	10.1	2,447,319,296	11,339
GDW2284	*Providencia stuartii*	35,231	3,997	11.9	269,413,278	15.823
GDW2285	*Escherichia coli*	35,947	2,384	9.7	85,711,743	3,259
GDW2286	*Citrobacter freundii*	36,923	3,204.1	11.4	118,306,457	10,032
GDW2287	*Klebsiella pneumoniae*	3,745	2,841	11.7	26,696,234	17,774
GDW2288	*Enterobacter cloacae*	19,313	4,648	11.9	89,774,842	9,088
GDW2289	*Escherichia coli*	13,243	3,783	12.1	95,395,836	14,658
GDW2290	*Klebsiella oxytoca*	24,413	2,228	11.4	112,925, 135	9,867
GDW2291	*Enterobacter cloacae*	7,273	84	11.5	555,919,98	15,673

Table: 3.5**: Comparison of hybrid assembly NDM plasmids with MOB-Suite mash neighbor (reference) plasmids**. The term "Variant of reference plasmid" is used for comparisons where the genomic distance is less than 0.01 but not zero and the term "Distant homolog of reference plasmid" is used for comparison where the genomic distance is greater than 0.01. Incomplete, non-circular plasmid predictions are labelled as such. RGI Perfect hits are shown in bold.

Isolate ID with putative plasmid size	**MOB-Suite mash neighbor plasmid (Plasmid cluster_id)**	**Mash genomic distance**	**Hybrid NDM plasmid sequence comparison to MOB-Suite mash neighbor plasmid**	**Mobility prediction**	**Predicted AMR genes**
GDW2218 137,552 bp	pKP-NDM1 (Plasmid 955)	2.382×10^{-5}	Complete match to reference plasmid (100%)	Conjugative	**CMY-6**, AAC (6')-Ib7, **sul1, rmtC**, **NDM-1**, **BRP(MBL)**
C0612 94,836 bp	Pkp_Goel2164 (Plasmid 726)	0.0081	Variant of reference plasmid	Non-Mobilizable	**tet(D)**, AAC (3)-IIe, **mphA** dfrA14, **TEM-1**, **OXA-9**, aadA AAC (6')-Ib*7*, **CTX-M-15**, **QnrS1, APH (3')-VI**, **NDM-1, BRP(MBL)**
C0648 134,146 bp	pECY55 (Plasmid 1550)	0.00305	Variant of reference plasmid	Conjugative	tet(A), **dfrA12**, **aadA12**, **sul1**, **NDM-5**, **rmtB**, **TEM-1**, **mphA**, **sul1**, **aadA5**, dfrA17
C0650 113,323 bp	blaNDM1 plasmid p2 (Plasmid 1220)	0.00149	Variant of reference plasmid	Conjugative	**BRP(MBL)**, **NDM-1**, **APH (3')-VI**, **QnrS1**, **CTX-M-15**, AAC (6')-Ib7, aadA, OXA-9, TEM-183
GDW2284 147,866 bp	pstGN576 (Plasmid 955)	None	Variant of reference plasmid	Conjugative	**CMY-6**, AAC (6')-Ib10, **sul1**, **rmtC**, **AAC (3)-IId**, APH (3')-VI, **NDM-1**, BRP(MBL)
GDW2285 118,768 bp	pMS6198A (Plasmid 955)	0.003169	Variant of reference plasmid	Conjugative	**CMY-6**, AAC (6')-Ib-cr, **sul1**, **rmtC**, **NDM-1**, BRP(MBL)

GDW2286 179,697 bp *incomplete*	pMS6198A (Plasmid 955)	0.00659	Variant of reference plasmid	Conjugative	CTX-M-15, **CMY-6**, AAC (6)-Ib7, **sul1**, **rmtC**, **APH (3)-VI**, **NDM-1**, **BRP(MBL)**, **TEM-1**, **sul1**
GDW2288 108,467 bp	pECL3-NDM1 (Plasmid 621)	0.00238	Variant of reference plasmid	Conjugative	OXA 9, ANT (3''), AAC (6')-Ib7, BRP(MBL), **NDM-1**, **APH (3')-VI**, **QnrS1**, **CTX-M-15**
GDW2290 140,507 bp	blaNDM-1 plasmid p2 (Plasmid 1220)	0.0113973	Distant homolog of reference plasmid	Conjugative	**TEM-1**, **OXA-9**, aadA, AAC (6')-Ib10, **CTX-M-15**, **QnrS1**, **APH (3')-VI**, **NDM-1, BRP(MBL)**, **mphA**
GDW2216 82,327 bp	P1002-NDM1 (Plasmid 955)	0.0183	Distant homolog of reference plasmid	Mobilizable	**AAC (3')-IId**, **sul1**, DHA-7, **NDM-1**, **rmtC**, sul1, AAC (6')-Ib7
GDW2287 82,327 bp	P1002-NDM1 (Plasmid 955)	0.0183	Distant homolog of reference plasmid	Mobilizable	**AAC (3)-IId**, **sul1**, DHA-7, **NDM-1**, **rmtC**, **sul1**, AAC (6')-Ib7
GDW2280 159,260 bp	pPG010208 (Plasmid 966)	0.0209	Distant homolog of reference plasmid	Conjugative	**dfrA12**, **aadA2**, **sul1**, **armA**, **msrE, mphE, sul2**, **NDM-1,** BRP(MBL), SHV-66
GDW2281 6,541 bp	Pkp_Goe_579 (Plasmid 2719)	0.047688	Distant homolog of reference plasmid	Non-mobilizable	**NDM-1**, **sul1**
GDW2289 60,362 bp	pHHA45 (Plasmid 934)	0.0205	Distant homolog of reference plasmid	Conjugative	**dfrA5**, **BRP(MBL)**, **NDM-1**
GDW2291 113,558 bp	P112298-KPC (Plasmid 654)	0.0310	Distant homolog of reference plasmid	Mobilizable	**APH (3)-VI**, **NDM-1**, **TEM-1**

Table: 3.6**: Hybrid assembly plasmid predictions from GDW2218 *Klebsiella pneumoniae* isolate.** Four complete, circular plasmids were predicted from the hybrid assembly isolate GDW2218. Incomplete, non-circular plasmid predictions are labelled as such. Predicted replicon genes that could not be assigned to known replicon genes are labelled as rep_cluster. RGI Perfect hits are shown in bold.

Size of putative plasmids from GDW2218 isolate	**Putative plasmid replicon type**	**MOB-Suite mash neighbor plasmid (Plasmid cluster_id)**	**Mobility predictions**	**Predicted AMR genes**
16,320 bp	rep_cluster	pCAV1374 (Plasmid 607)	Non-mobilizable	n/a
4,709 bp	rep_cluster	Plasmid pKPN6 (Plasmid novel_0)	Non-mobilizable	n/a
109,749 bp	IncFIB, IncFII	Plasmid p2 (Plasmid 1220)	Conjugative	**SHV-1, QnrB1**, dfrA14
137,552 bp	IncA/C2	Plasmid pKP-NDM-1 (Plasmid 955)	Conjugative	**CMY-6**, AAC (6')-Ib7 (strict), **sul1**, **rmtC**, **NDM-1 (perfect)**, **BRP(MBL)**

Table: 3.7**: Hybrid assembly plasmid predictions from C0648 *Escherichia coli* isolate**. Eight plasmids were predicted from the hybrid assembly isolate C0648. Incomplete, non-circular plasmid predictions are labelled as such. Predicted replicon genes that could not be assigned to known replicon genes are labelled as rep_cluster. RGI Perfect hits are shown in bold.

Size of putative plasmids from C0648 isolate	**Putative plasmid replicon types**	**MOB-Suite mash neighbor plasmid (Plasmid cluster_id)**	**Mobility predictions**	**Predicted AMR genes**
3,648 bp	rep_cluster	Plasmid pSZECL_e (Plasmid novel_0)	Non-mobilizable	n/a
6,965 bp	rep_cluster	Plasmid pEcoFMU073332a (Plasmid 299)	Mobilizable	n/a
4,018 bp	rep_cluster	PSAN2-06-0624 (Plasmid 141)	Mobilizable	n/a
3,174 bp	rep_cluster	Plasmid C (Plasmid 10)	Mobilizable	n/a
4,085 bp	rep_cluster	Plasmid pEC545_4 (Plasmid 469)	Mobilizable	n/a
36,889 bp	IncX3	Plasmid pUCLAOXA23 (Plasmid 664)	Conjugative	n/a
54,557 bp	IncI1	pCMY-42 (Plasmid 486)	Conjugative	**CMY-42**
134,146 bp	IncFIA, IncFII, IncFIIA	Plasmid 1550 (Plasmid pECY55)	Conjugative	tet(A), **dfrA12, aadA12**, **sul1**, **NDM-5, rmtB**, **TEM-1**, **mphA**, **aadA5**, dfrA17

Table: 3.8**: Hybrid assembly plasmid predictions from GDW2285 *Escherichia coli* isolate.** Four plasmids are predicted from the hybrid assembly isolate. Predicted replicon genes that could not be assigned to known replicon genes are labelled as rep_cluster Incomplete, non-circular plasmid predictions are labelled as such. RGI Perfect hits are shown in bold.

Size of putative plasmids from GDW2285 isolate	**Putative plasmid replicon types**	**Plasmid cluster_id**	**Mobility predictions**	**Predicted AMR genes**
1,934 bp	rep_cluster	Plasmid pKp_Goe_579 (Plasmid 58)	Non-mobilizable	n/a
3,174 bp	rep_cluster	Plasmid pB1020 (Plasmid 10)	Non-mobilizable	n/a
112,137 bp	IncFIB, IncFII,IncFIA	Plasmid pEC224 Plasmid 1550	Non-mobilizable	**CTX-M-15**, catB3, OXA-1, AAC (6')-Ib-cr
118,768 bp	IncA/C2	Plasmid pMS6198A (Plasmid 955)	Conjugative	**CMY-6,** AAC (6)'-Ib-cr, **sul1**, **rmtC**, **NDM-1**, BRP(MBL)

Table: 3.9: **Hybrid assembly plasmid predictions from GDW2287 *Klebsiella pneumoniae* isolate.** Three plasmids are predicted from this hybrid assembly isolate. Incomplete, non-circular plasmid predictions are labelled as such. Predicted replicon genes that could not be assigned to known replicon genes are labelled as rep_cluster. RGI Perfect hits are shown in bold.

Size of putative plasmids from GDW2287 isolate	**Putative plasmid replicon types**	**Plasmid cluster_id**	**Mobility predictions**	**Predicted AMR genes**
95,619 bp	IncFIB, IncFII	Plasmid pK66-45-3 (Plasmid 1220)	Conjugative	**CTX-M-15,** AAC (6')-Ib-cr, **OXA-1**
4,474 bp	rep_cluster	Plasmid pCAV1344 (Plasmid novel_0)	Mobilizable	n/a
82,327 bp	IncA/C2	Plasmid p1002-NDM1 (Plasmid 955)	Conjugative	**AAC (3)-IId**, **sul1**, DHA-7 (strict), **NDM-1, rmtC**, **sul1**, AAC (6')-Ib7

Chapter 4

Discussion

Research on multi-drug resistance NDM plasmids is essential due to the prevalence of resistant bacteria in different habitats, including the clinic, farm, and environment. Bacteria are known to share genes amongst one another through a process known as horizontal gene transfer (HGT), which means antibiotic resistance can disperse amongst different pathogens, the non-pathogenetic microbiome, and the environmental microbiome. Despite this knowledge, there are a limited range of bioinformatics tools to reliably detect resistance plasmids, thereby making AMR surveillance incomplete.

4.1 Whole Genome shotgun assembly does not result in complete plasmid prediction

Bacteria are known to hold zero, one, or multiple copies of plasmids with varying sizes that may or may not be beneficial to the bacteria[55]. Often, the accessory genes carried on plasmids, such as antimicrobial resistance genes, provide bacteria with selective advantages[55]. Plasmid DNA purification is labor-intensive, lacks standard protocols, and is usually accompanied by contamination with chromosomal DNA[56]. The diminishing cost and increasing throughput of next-generation DNA sequencing and WGS workflows have allowed for the acquisition of genomic data. However, there are still challenges when it comes to the analysis of these data. We still face a bottleneck in developing tools that correctly analyze resistance mobile elements, like resistance plasmids or integrative and conjugative elements, from genome assembly data. Traditionally, the sequencing of just plasmids requires purification of plasmid DNA, followed by shotgun sequencing, and requires primer walking to close all gaps[57]. Yet, this process is tedious with many bacterial isolates and the process may be different across all the bacterial strains.

In recent years, the analysis of plasmids from WGS sequencing reads requires the assembly of the reads, followed by plasmid detection from the assembly data using plasmid prediction tools. There are known bioinformatics tools that detect plasmids from WGS assemblies, but none predict resistance elements on plasmids. For this reason, I developed the RGI:Mobilome pipeline within RGI to identify resistance plasmids from genomic data. This tool separates resistance chromosomal contigs from resistance plasmid contigs by taking advantage of MOB-Suite's ability to reconstruct and type plasmids from assembled contigs, using a plasmid gene database and a reference plasmid database. RGI:Mobilome identified 84 plasmids among 15 NDM-associated Illumina assembled genomes from patients from Hamilton area hospitals. In each assembled genome, RGI:Mobilome predicted a plasmid that encodes a broad-spectrum New Delhi Metallo-beta-lactamase gene and other families of resistance genes. Hence, RGI:Mobilome serves as a pipeline tool for initial surveillance of multi-drug resistance plasmids that exist in a bacteria population. Since RGI:Mobilome results from the harmonization of known and validated bioinformatics tools, there is confidence with its prediction of resistance plasmids. The integration of RGI:Mobilome into RGI framework will expand RGI predictions beyond AMR genes, improve bacterial genome analysis, and augment AMR surveillance especially during an outbreak.

Plasmids are commonly classified based on replicon incompatibility groups since closely related plasmids cannot coexist stably in the same cell due to competition for similar replication initiation systems required for replication[58]. This method of classification reveals the evolutionary relatedness among bacterial strains and for comparing plasmid contents. For instance, I used the PlasmidProfiler tool to determine how similar the NDM isolates were, based on their plasmid incompatibility groups, and the dendrogram revealed the clustering of the same bacterial species within a clade (Figure 3.4). However, in WGS data, the drawback to this type of classification is that it does not reveal important details about the genomic structure of plasmids, the arrangement of plasmid genes, particularly mobility genes and antimicrobial resistance genes, which is an essential factor in deciphering the epidemiology of plasmid-mediated antibiotic resistance[59]. Some plasmids

are known to have multiple replicons which complicate plasmid classifications and usually leads to overestimating the number of plasmids and plasmid types[60]. For this reason, I developed RGI:Mobilome as a resistance plasmid prediction tool that types not only plasmid replicons, but also relaxases and mate-pair formation genes using the MOB-Suite algorithms. The MOB-Suite typing method provides better plasmid classification since most plasmids usually encode just one relaxase family and mate-pair formation gene, which are useful genes to predict the potential mobility of plasmids[37,61]. In addition to replicon (Inc) and relaxase typing, RGI:Mobilome predicted more than 15 resistance gene families from the 84 putative plasmids.

RGI:Mobilome can be deployed as an initial AMR surveillance tool to detect and monitor the spread of resistance plasmids responsible for AMR transmission. However, the pipeline does not always output complete resistance plasmids, primarily due to assembly algorithm performance with Illumina data. The *de novo* assembly of Illumina reads usually results in the overlap of chromosomal and plasmid reads. Therefore, no matter how good a plasmid prediction tool is with identifying plasmid contigs, the challenge will always be the complete separation of chromosomal sequences from plasmid sequences[62]. Additionally, plasmid assemblies are complicated by various repeats that are usually difficult to resolve. These repeats could exist within plasmids or be shared by multiples plasmids or shared between a plasmid and a chromosome[63,64]. The short-read length of shotgun sequencing (< 300 bp) compared to the large size of most bacterial repeats leads to incomplete genome assembly since most assembly algorithms cannot resolve most of the repeats. From our Illumina assembly plasmid prediction (Figure 3.5A), the relatively high plasmid predictions from our NDM isolates of interest compared to hybrid assemblies (Figure 3.5B) is due to genome fragmentation and unresolved repeats. Few tools predict plasmids from *de Bruijn* graphs, which are constructed with k-mer(s) from Illumina reads, as this approach relies on significant differences in chromosomal to plasmid read coverage[65]. However, the repetitive sequences of bacterial genomes collapse these differences. Even in cases where the *de Bruijn* graph helped to identify plasmid sequences, there was no structural accuracy[65]. Thus, no matter how sophisticated an assembly

algorithm is, if there are no reads that span the repeat-containing genomic regions, it will never be able to resolve complete plasmid sequences without addition of longer reads.

4.2 Hybrid assembly resolves gaps and repeats of bacterial genomes

As stated earlier, WGS of bacterial genomes is rapid, cheap, and accurate, but short reads do not result in complete genome assembly and therefore impede the detection of complete plasmid structure. Long read sequencing data resolves bacterial genome structures which makes it useful for plasmid predictions. However, its combination with short-read sequencing data generates accurate plasmid predictions due to the accuracy of short-read sequencing data[46,66]. We sequenced all 15 NDM isolates of interest with the Oxford Nanopore Technologies (ONT) MinION long-read sequencing and combined these data with Illumina reads generated with the MiSeq platform. Alone, long read assemblies are not suitable for plasmid typing or resistance allele typing due to high nucleotide errors[67,68], but in combination with high-accuracy Illumina data are powerful for complete plasmid prediction. Hybrid assembly algorithms like Unicycler uses three-step approaches to generate 'hybrid' assemblies from short and long read sequencing data: contig construction with short read data, the scaffolding of contigs with long Nanopore reads, and polishing of contigs for error-correction and to fill in gaps[69].

Following the hybrid assembly of NDM isolates, the number of putative plasmids reduced to 70, as shown in Figure 3.5B, and nearly all plasmids were complete, i.e. circular. The reduction in the number of plasmids from Illumina to hybrid assembly predictions results from the resolution of plasmid repeats since each isolate's long reads enable the read-through of these repeats during assembly. Overall, hybrid assemblies ensure that repeats are assigned correctly to their respective DNA (chromosome/plasmid) sequences[70]. With regards to circularity, 13 of the 15 isolates had closed and circular chromosomes. The number of closed and circular plasmids was 66 out of 70 putative plasmids. All the NDM associated plasmids were closed and circular, except for the NDM plasmid from the GDW2286 isolate. The detection of a closed, circular plasmid for most of the NDM plasmids can be explained by the end to end joining of their Nanopore reads, which is also

responsible for their complete plasmid structure. Complete chromosome and plasmid sequences allows for complete differentiation and annotation of plasmid genes, particularly plasmid borne antibiotic resistance elements[46,71]. The sequenced NDM-associated plasmids each encoded at least three antibiotic resistance genes, with most of the encoded AMR genes conferring resistance to fluoroquinolone, aminoglycosides, sulfonamides, diaminopyrimidine, and beta-lactam antibiotics (Table 3.5). The hybrid assembly of our isolates also resolved the location of the antibiotic resistance islands on our NDM plasmids (Figures 3.7 and 3.8).

Despite the efficiency of hybrid assemblies in resolving plasmids, there can be setbacks in its usage since it involves the combination of two different sequencing technologies. For example, Nanopore long-read sequencing produces chimeric reads, resulting from the joining of two separate pieces of DNA in a single read, providing misleading information on bacterial genomic structures[72,73]. Therefore, I cannot discount its probable effect on detecting our 'novel' NDM plasmids. However, we used the bioinformatics long-reads trimming tool porechop to minimize this effect, since it is known to remove chimeric reads prior to assembly[74]. Furthermore, we analyzed the hybrid assembly of our NDM isolates by using the default settings of the Unicycler assembly algorithm, which includes polishing of long reads with short reads during genome assembly. However, there are reported cases where the assembly of long reads is first required, then subsequent polishing with short reads to achieve a complete plasmid assembly[75]. We did not attempt to use this method to find out if it would result in a complete plasmid structure, but we take confidence in our results given that our method was able to assemble an NDM plasmid with 100% match to a known plasmid pKP-NDM1 (Figure 3.6). Finally, hybrid assembly helps plasmid prediction tools detect complete plasmid, and suggest any identical, similar, or novel plasmids, using algorithms like mash nearest neighbors. Yet, most plasmid prediction tools like MOB-Suite and PlasmidFinder uses a reference database of Enterobacteriaceae plasmids for predictions, which is why these tools did not perform well for other bacterial groups in our study, such as *Pseudomonas aeruginosa*, *Proteus mirabilis*, and *Acinetobacter baumannii*.

4.3 Evolution, Diversity and Epidemiology of NDM plasmids from clinical isolates

The RGI predicted NDM genes in 15 isolates from the Wright clinical collection, and the predictions were on plasmids based on RGI:Mobillome analysis. The 15 NDM-bearing plasmids were grouped into 7 replicon types: IncFIA, IncR, IncFII, IncFIIA, IncFIB, IncA/C2, and an IncN (Table 3.3). This suggests that multiple plasmid types acquired the NDM gene through horizontal gene transfer (HGT) on its migration from its initial report in India to Hamilton, Ontario. From Table 3.3, IncA/C2 plasmids have the highest number with 7, described as one of the common NDM carrying plasmids worldwide. This plasmid type is present in most gram-negative species and is usually larger in size, with low-copy numbers and broad host ranges[54,76]. They are also known to disseminate the CMY and NDM genes[77–79]. The CMY gene was predicted in 4 out of the 7 IncA/C2 predicted NDM plasmids. These also possess the tra genes responsible for mobility, and the parA/B genes enables their correct positioning, especially during cell division [80]. The incA/C2 plasmids replicate using the repA gene, which is a binding site for RepA protein. The NDM gene in the Inc A/C2 plasmids is commonly situated on an antibiotic resistance island between two *tra* regions[54]. We observed the same feature in the annotated NDM-1 plasmids of the GDW2218 and GDW2285 isolates (Figures 3.6 and 3.8). Besides plasmids, NDM genes exist in various mobile genetic elements such as insertion sequences, transposons, and integrons, and each plays a role in its mobilization[81,82]. The two features of mobile genetic elements that encode the NDM genes are the insertion sequence (intact/truncated) upstream of the NDM gene and a bleomycin resistance gene, BRP(MBL) downstream[83,84]. We observed these two features in the annotated NDM-1 bearing plasmids of the GDW2218 *Klebsiella pneumoniae* and GDW2285 *Escherichia coli* isolates (Figures 3.6 and 3.8). For the NDM-5 bearing plasmid of C0648 *Escherichia coli* isolate, it only encodes the bleomycin resistance gene (Figure 3.7). In general, the 15 predicted NDM plasmids co-harbor other resistance genes to different antibiotics such as ESBLs, aminoglycosides, quinolones, sulfonamides, which are antibiotics commonly used in clinical settings[85]. Like the NDM gene, the co-harbored resistance genes are acquired by

plasmids via transposition, mostly facilitated by transposons and insertion sequences[86–88]. In each NDM positive strain, the detected plasmids including the NDM plasmids have different replicon incompatibility groups (Table 3.6 – 3.9), which prevents competition for replication machineries, and allows for the stability of all plasmids[89,90].

There is a need to know how similar the predicted NDM plasmids are to known and well-characterized plasmids. For this reason, I grouped the predicted NDM plasmids from the 15 isolates into three groups: complete match to a reference plasmid, variant of reference plasmid, and distant homolog of reference plasmid. The reference plasmids used for comparison are plasmids within MOB-Suite's MOB-cluster reference data, originally obtained from the NCBI repository. The GDW2218 NDM plasmid is a complete match to its reference pKP-NDM-1 plasmid, with the smallest mash genomic distance of 2.382×10^{-5}. This plasmid was originally described from a *Klebsiella pneumoniae* strain isolated from a patient at a clinical research center in Brisbane, Australia[54]. The detection of a complete and identical NDM plasmid in the GDW2218 *Klebsiella pneumoniae* clinical strain in Hamilton, ON, suggests that vertical transmission occurred. The NDM-1 bearing plasmid is predicted as conjugative and encodes AMR genes resistant to other beta-lactams, aminoglycosides, and sulfonamides. We speculate that the presence of other AMR genes may have helped with its positive selection over time. In addition to the predicted NDM-1 bearing plasmid, three other plasmids were predicted from hybrid assembled GDW2218 *Klebsiella pneumoniae* strain (Table 3.6). The 109,749 bp plasmid is a conjugative plasmid, has different replicon genes and resistance genes from the NDM bearing plasmid. However, the 109,749 was not reported in the pKP-NDM-1 *Klebsiella pneumoniae* strain suggesting that it was acquired via horizontal transmission in its new environment.

The default MASH genomic distance used by MOB-Suite to assign putative and reference plasmids belonging to the same cluster falls between 0.001 and 0.05. The seven plasmids designated as variants of their mash neighbor reference plasmid have a mash genomic distance that falls between 0.001 to 0.01, except for the NDM-1 plasmid of the GDW2284 *Providencia stuartii* isolate with no MASH genomic distance (Table 3.5). The resistance genes across all 7 NDM variant plasmids are clustered together along their

resistance island. The NDM plasmids from GDW2284 *Providencia stuartii*, GDW2285 *Escherichia coli*, and GDW2286 *Citrobacter freundii* isolates are variants of the complete NDM plasmid from GDW2218 *Klebsiella pneumoniae* isolate, as they share similar incompatibility groups (IncA/C2), conjugal transfer relaxase gene, but slightly different resistance genes on the resistance island. Like NDM plasmid from GDW2218 *Klebsiella pneumoniae* isolate, the resistance islands of these NDM variant plasmids are flanked by insertion sequences and transposases, which allows for acquisition and transposition of AMR genes. The NDM plasmids from GDW2284 *Providencia stuartii* and GDW2286 *Citrobacter freundii* isolates differ from the NDM plasmid from GDW2218 *Klebsiella pneumoniae* isolate by acquiring aminoglycoside resistance genes in the resistance island, but only the NDM plasmid from GDW2286 *Citrobacter freundii* isolate acquired two additional beta-lactamases: CTX-M-15 and TEM-1 (Table 3.5). In contrast, the resistance genes on the NDM plasmid from GDW2285 *Escherichia coli* isolate is identical to the resistance genes on the NDM plasmid from GDW2218 *Klebsiella pneumoniae* isolate. However, there is a missing 18 kb region in the NDM plasmid backbone from GDW2285 *Escherichia coli* isolate. The missing 18kb region encodes mobility genes, endonucleases, and phosphodiesterase genes, which are not essential genes for plasmid replication. This 18kb region was perhaps excised from the GDW2285 *Escherichia coli* NDM plasmid by the nearby transposase to reduce the size of the plasmid, which in turn reduces the fitness cost for the host bacterium (GDW2285)[91]. Another explanation for its excision was perhaps to prevent the transmission or loss of the NDM-1 bearing plasmid due to the selective advantage that it provides the GDW2285 *Escherichia coli* strain. The remaining four NDM plasmids in this group of variant plasmids belong to the family of Inc F plasmids: NDM plasmids from C0612 *Klebsiella pneumoniae*, C0648 *Escherichia coli* and C0650 *Klebsiella pneumoniae*, and GDW2288 *Enterobacter cloacae* isolates. Like the IncA/C2 plasmids, the IncF plasmids are mostly found in gram-negative family, Enterobacteriaceae and are known to carry the extended-spectrum ß-lactamases (ESBLs), especially CTX-M's, quinolone resistance genes, and aminoglycoside resistance genes[92,93]. These genes were predicted on the C0612 *Klebsiella pneumoniae*, C0650 *Klebsiella pneumoniae*, and

GDW2288 *Enterobacter cloacae* NDM plasmids, whereas beta-lactamase TEM-1 and aminoglycoside resistance gene were predicted on C0648 *Escherichia coli* NDM plasmid. These four IncF NDM plasmids are not only variants of their mash neighbor reference plasmid, but also variants of one another due to the differences in both their resistance genes predictions and relaxase gene types (Table 3.3). The IncF NDM plasmids are conjugative plasmids except for the NDM plasmid from C0612 *Klebsiella pneumoniae* isolate, which is a non-mobilizable plasmid but coexists with a 72 kb conjugative plasmid that perhaps facilitated its transmission.

The seven remaining NDM plasmids are assigned as 'distant homologs" of their mash neighbor reference plasmids with large differences in plasmid sequences and MASH genomic distance greater than 0.01. Also, these NDM plasmids do not have close similarities to known plasmids in the NCBI repositories. For instance, the NDM plasmid of GDW2216 *Klebsiella pneumoniae* isolate differs widely from mash neighbor p1002-NDM1 plasmid based on read-mapping results in Figure 3.9. However, this NDM plasmid shares a similar incompatibility group (IncA/C2) and AMR genes with the GDW2218 *Klebsiella pneumoniae* NDM plasmid but 40 kb lesser in size. The missing 40kb region does not affect the detection of replicon incompatibility genes, AMR genes, mobility genes, and still predicted as a circular plasmid. I hypothesize that the 40 kb region was not required to detect a complete plasmid, and perhaps lost during transmission. However, I cannot disregard the notion that the missing region was due to gaps in genome assembly. The NDM plasmids from the GDW2290 *Klebsiella oxytoca* isolate share similar incompatibility group (IncF) and AMR gene prediction with reference blaNDM-1 p2 plasmid and C0648 *Escherichia coli* NDM plasmid, but there is an additional 20 kb sequence as part of the GDW2290 *Klebsiella oxytoca* NDM plasmid backbone. In general, the NDM plasmids for this 'distant homolog plasmids' group have missing regions or acquired additional regions due to transposition. However, they are generally circular, have backbones seen in Enterobacteriaceae plasmids, and encode AMR genes that include NDM. The only exception is the NDM plasmid from GDW2281 *Morganella morganii* and GDW2291 *Enterobacter cloacae* with no defined replicon incompatibility group, conjugal transfer

gene, but carry the NDM gene and other resistance gene. For this group, the detection of complete and unknown NDM plasmids could be a discovery of previously unreported NDM plasmids in clinical settings and suggests that there is a lot more to know about NDM plasmids.

Chapter 5

Conclusions and Future Directions

5.1 Conclusions

The complete detection of multi-drug resistance NDM plasmids is required to address the epidemiology of NDM positive bacterial pathogens worldwide. While there is evidence that travel plays an important role in the spread of NDM positive strains internationally, there are fewer reports on the prevalence and diversity of NDM positive strains within communities. First, this study validates RGI:Mobilome pipeline as a surveillance tool for initial detection of resistance plasmids. Yet, this study also proves that Illumina plus Nanopore sequencing of bacteria genome is required for the accurate and complete detection of clinical resistance plasmids (e.g., NDM plasmids), and should be the standard sequencing paradigm for analyzing bacterial genomes. All the 15 NDM positive strains in this study have their NDM gene carried on plasmids, and the NDM gene co-harbors with resistance genes to commonly used antibiotics in clinics. Based on the sample size of the detected NDM plasmids in Hamilton-Niagara community, the NDM plasmids are typically large (>100 kb), transmissible, with varying incompatibility backbones, and coexist only with neighboring plasmids of different incompatibility groups. The detected NDM plasmids were put into three groups (i.e., ‘complete match’, ‘variant’, and ‘distant homolog’) depending on how similar they are to known and well-characterized NDM plasmids. Overall, this study provides insight on the diversity of known clinical NDM plasmids within the Hamilton-Niagara community. This diversity makes bacterial infection caused by NDM positive strains difficult to treat, which places extra burden on our healthcare system. However, the result of this study will guide public health practitioners in developing optimal treatments plan for different types of resistance NDM bacterial infections.

5.2 Future directions

The result of this study shows that we have known and diverse NDM plasmids within clinics of the Hamilton-Niagara community. Based on the result of the 'distant homolog' plasmid predictions, future studies could require improving our current plasmid prediction tools to detect transposable elements that excise from or insert to a plasmid. The detection of transposable elements will result in more confident plasmid predictions. The bioinformatics analysis of the clinical NDM isolates for this thesis project is of small sample size, so it does not fully epitomize the extent on how widespread the NDM genes and only represents a fraction of NDM bacterial isolates within the Hamilton-Niagara community. Therefore, there is a need to analyze isolates from environmental and agricultural settings for NDM genes, co-harboring genes, and plasmid backbones, which are useful for deciphering the NDM plasmid type. Also, the 15 NDM isolates from this study need to be experimentally validated to have the predicted NDM plasmids, and isolates should also be tested for a higher level of minimum inhibitory concentration (MIC), due to the presence of the NDM plasmid, particularly the NDM gene. Additional genomic analyses can be done to check if the interactions between the NDM gene, plasmid backbone, mobility genes, and bacteria species determine what resistance genes are acquired and coexist with the NDM gene.

Bibliography

1. Frieri, M., Kumar, K. & Boutin, A. Antibiotic resistance. *Journal of Infection and Public Health*, 10(4):369–378, 2017.
2. Périchon, B., Courvalin, P. & Stratton, C. W. Antibiotic resistance. *Encyclopedia of Microbiology*, 12(3):127–139, 2019.
3. Cervera, C. *et al.* Multidrug-resistant bacteria in solid organ transplant recipients. *Clinical Microbiology and Infection*, 20(7):49–73, 2014.
4. Langford, B. J. *et al.* Bacterial co-infection and secondary infection in patients with COVID-19: a living rapid review and meta-analysis. *Clinical Microbiology and Infection*, 743(20):30423-20427, 2020.
5. Brown, E. D. & Wright, G. D. Antibacterial drug discovery in the resistance era. *Nature*, 529:336–343, 2016.
6. Mobarki, N., Almerabi, B. & Hattan, A. Antibiotic Resistance Crisis. *International Journal of Medicine in Developing Countries*, 3(6):561–564, 2019.
7. McKenna, M. Antibiotic resistance: The last resort. *Nature*, 499(7459), 394–396 (2013).
8. Yong, D. *et al.* Characterization of a new metallo-β-lactamase gene, bla NDM-1, and a novel erythromycin esterase gene carried on a unique genetic structure in *Klebsiella pneumoniae* sequence type 14 from India. *Antimicrobial Agents and Chemotherapy*, 53(12):5046–5054, 2009.
9. 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. *The Lancet Infectious Diseases*, 16(2):161–168, 2016.
10. Anderson, G. G. & O'Toole, G. A. Innate and induced resistance mechanisms of bacterial biofilms. *Current Topics in Microbiology and Immunology*, 322:85–105, 2008.
11. Miller, S. I. Antibiotic resistance and regulation of the Gram-negative bacterial outer membrane barrier by host innate immune molecules. *mBio*, 7(5):1541-1546, 2016.
12. van Hoek, A. H. A. M. *et al.* Acquired antibiotic resistance genes: An overview. *Frontiers in Microbiology*, 2:203-230, 2011.

13. Boerlin, P. & Reid-Smith, R. J. Antimicrobial resistance: its emergence and transmission. *Animal health research reviews / Conference of Research Workers in Animal Diseases*, 2(9):115–126, 2008.
14. Schroeder, J. W., Yeesin, P., Simmons, L. A. & Wang, J. D. Sources of spontaneous mutagenesis in bacteria. *Critical Reviews in Biochemistry and Molecular Biology*, 53(1):29–48, 2018.
15. Woodford, N. & Ellington, M. J. The emergence of antibiotic resistance by mutation. *Clinical Microbiology and Infection*, 13(1):5–18, 2007.
16. Goldstein, B. P. Resistance to rifampicin: A review. *Journal of Antibiotics*, 67(7425):625–630, 2014.
17. Hooper, D. C. & Jacoby, G. A. Mechanisms of drug resistance: Quinolone resistance. *Annals of the New York Academy of Sciences*, 1354(1):12–31, 2015.
18. Manaia, C. M. Assessing the risk of antibiotic resistance transmission from the environment to humans: Non-Direct Proportionality between Abundance and Risk. *Trends in Microbiology*, 25(3):173–181, 2017.
19. Lerminiaux, N. A. & Cameron, A. D. S. Horizontal transfer of antibiotic resistance genes in clinical environments. *Canadian Journal of Microbiology*, 65(1):34–44, 2019.
20. Madsen, J. S., Burmølle, M., Hansen, L. H. & Sørensen, S. J. The interconnection between biofilm formation and horizontal gene transfer. *FEMS Immunology and Medical Microbiology*, 65(2):183–195, 2012.
21. Thomas, C. M. & Nielsen, K. M. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature Reviews Microbiology*, 3:711–721, 2005.
22. Chen, I. & Dubnau, D. DNA uptake during bacterial transformation. *Nature Reviews Microbiology*, 2:241–249, 2004.
23. Lovett, S. T., Hurley, R. L., Sutera, V. A., Aubuchon, R. H. & Lebedeva, M. A. Crossing over between regions of limited homology in *Escherichia coli*: RecA-dependent and RecA-independent pathways. *Genetics*, 160(3):851–859, 2002.
24. Olson, M. E. & Horswill, A. R. Bacteriophage transduction in *Staphylococcus epidermidis*. *Methods in Molecular Biology*, 1106:167–172, 2014.

25. Chiang, Y. N., Penadés, J. R. & Chen, J. Genetic transduction by phages and chromosomal islands: The new and noncanonical. *PLoS Pathogens*, 15(8):7878-7885, 2019.

26. Johnson, C. M. & Grossman, A. D. Integrative and Conjugative Elements (ICEs): what they do and how they work. *Annual Review of Genetics*, 49:577–601, 2015.

27. Grohmann, E., Muth, G. & Espinosa, M. Conjugative Plasmid transfer in gram-positive bacteria. *Microbiology and Molecular Biology Reviews*, 67(2):277–301, 2003.

28. Smillie, C., Garcillán-Barcia, M. P., Francia, M. V., Rocha, E. P. C. & de la Cruz, F. Mobility of plasmids. *Microbiology and Molecular Biology Reviews*, 74(3): 434–452, 2010.

29. Kaufman, J. *et al.* Integrative and Conjugative Elements (ICE) and associated cargo genes within and across hundreds of bacterial genera, *bioRxiv*, 20:1-35, 2020.

30. Shendure, J. & Ji, H. Next-generation DNA sequencing. *Nature Biotechnology*, 26(1):1135–1145, 2008.

31. Timme, R. E., Sanchez Leon, M. & Allard, M. W. Utilizing the Public GenomeTrakr Database for foodborne pathogen traceback. *Methods in Molecular Biology*, 1918:201–212, 2019.

32. Jia, B. *et al.* CARD 2017: Expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Research*, 45(1):566–573, 2017.

33. Alcock, B. P. *et al.* CARD 2020: Antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Research*, 48(1):517–525, 2020.

34. McArthur, A. G. *et al.* The comprehensive antibiotic resistance database. *Antimicrobial Agents and Chemotherapy*, 57(7):3348–3357, 2013.

35. Hernandez, D., François, P., Farinelli, L., Østerås, M. & Schrenzel, J. De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer. *Genome Research*, 18:802–809, 2008.

36. Pevzner, P. A., Tang, H. & Tesler, G. De novo repeats classification and fragment assembly. *Genome Research*, 14:1786–1796, 2004.

37. Robertson, J. & Nash, J. H. E. MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. *Microbial genomics*, 4(8):206-213, 2018.
38. Ondov, B. D. *et al.* Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biology*, 17:132-145, 2016.
39. Zhang, J., Chiodini, R., Badr, A. & Zhang, G. The impact of next-generation sequencing on genomics. *Journal of Genetics and Genomics*, 38(3):95–109, 2011.
40. Baker, M. De novo genome assembly: What every biologist should know. *Nature Methods*, **9**:333–337, 2012.
41. Bleidorn, C. Third generation sequencing: Technology and its potential impact on evolutionary biodiversity research. *Systematics and Biodiversity*, 14(1):1–8, 2016.
42. Rhoads, A. & Au, K. F. PacBio sequencing and Its applications. *Genomics, Proteomics and Bioinformatics*, 13(5):278–289, 2015.
43. Buck, D. *et al.* Comprehensive comparison of Pacific Biosciences and Oxford Nanopore technologies and their applications to transcriptome analysis. *F1000Research*, 6(2):100-131, 2017.
44. Lu, H., Giordano, F. & Ning, Z. Oxford Nanopore MinION sequencing and genome assembly. *Genomics, Proteomics and Bioinformatics*, 14(3):265–279, 2016.
45. Tyler, A. D. *et al.* Evaluation of Oxford Nanopore's MinION Sequencing Device for microbial whole genome sequencing applications. *Scientific Reports*, 8: 10931-10942, 2018.
46. Sydenham, T. v. *et al.* Complete hybrid genome assembly of clinical multidrug-resistant bacteroides fragilis isolates enables comprehensive identification of antimicrobial-resistance genes and plasmids. *Microbial Genomics*, 5(11):312–329. 2019.
47. Nordmann, P., Poirel, L., Walsh, T. R. & Livermore, D. M. The emerging NDM carbapenemases. *Trends in Microbiology*, 19(12):588–595, 2011.
48. González, L. J. *et al.* Membrane anchoring stabilizes and favors secretion of New Delhi metallo-β-lactamase. *Nature Chemical Biology*, 12:516–522, 2016.
49. Hornsey, M., Phee, L. & Wareham, D. W. A novel variant, NDM-5, of the New Delhi metallo-β-lactamase in a multidrug-resistant *Escherichia coli* ST648 isolate recovered

from a patient in the United Kingdom. *Antimicrobial Agents and Chemotherapy*, 55(12):5952–5954, 2011.

50. Kumarasamy, K. K. *et al.* Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: A molecular, biological, and epidemiological study. *The Lancet Infectious Diseases*, 10(9):597–602, 2010.
51. Andrews, S. Babraham Bioinformatics - FastQC A quality control tool for high throughput sequence Data. *Soil*, 5:47–81 2010.
52. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15):2114–2120, 2014.
53. Bankevich, A. *et al.* SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, 19(5):455–477, 2012.
54. Wailan, A. M. *et al.* Mechanisms involved in acquisition of blaNDM genes by IncA/C2 and IncFIIY plasmids. *Antimicrobial Agents and Chemotherapy*, 60(7):4082–4088, 2016.
55. Sano, E., Maisnier-Patin, S., Aboubechara, J. P., Quiñones-Soto, S. & Roth, J. R. Plasmid copy number underlies adaptive mutability in bacteria. *Genetics*, 198:919–933, 2014.
56. Heilig, J. S., Elbing, K. L. & Brent, R. Large-scale preparation of plasmid DNA. *Current Protocols in Molecular Biology*, 41:1-16, 1998.
57. Smalla, K., Jechalke, S. & Top, E. M. Plasmid detection, characterization, and ecology. *Microbiology Spectrum*, 3:445-448, 2015.
58. Couturier, M., Bex, F., Bergquist, P. L. & Maas, W. K. Identification and classification of bacterial plasmids. *Microbiological Reviews*, 52(3):375–395, 1988.
59. Orlek, A. *et al.* Plasmid classification in an era of whole-genome sequencing: Application in studies of antibiotic resistance epidemiology. *Frontiers in Microbiology*, 8:182-192, 2017.
60. Shintani, M., Sanchez, Z. K. & Kimbara, K. Genomics of microbial plasmids: Classification and identification based on replication and transfer systems and host taxonomy. *Frontiers in Microbiology*, 6:242-256, 2015.

61. Garcillán-Barcia, M. P., Alvarado, A. & de la Cruz, F. Identification of bacterial plasmids based on mobility and plasmid population biology. *FEMS Microbiology Reviews*, 35(5):936–956, 2011.

62. Arredondo-Alonso, S., Willems, R. J., van Schaik, W. & Schürch, A. C. On the (im)possibility of reconstructing plasmids from whole-genome short-read sequencing data. *Microbial Genomics*, 3(10):128-135, 2017.

63. Kuśmirek, W. & Nowak, R. De novo assembly of bacterial genomes with repetitive DNA regions by dnaasm application. *BMC Bioinformatics*, 19:273-282, 2018.

64. De Bustos, A., Cuadrado, A. & Jouve, N. Sequencing of long stretches of repetitive DNA. *Scientific Reports*, 6:36665-36671, 2016.

65. Antipov, D. *et al.* PlasmidSPAdes: Assembling plasmids from whole genome sequencing data. *Bioinformatics*, 32:3380–3387, 2016.

66. De Maio, N. *et al.* Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes. *Microbial Genomics*, 5(9):294-305, 2019.

67. Lemon, J. K., Khil, P. P., Frank, K. M. & Dekker, J. P. Rapid nanopore sequencing of plasmids and resistance gene detection in clinical isolates. *Journal of Clinical Microbiology*, 55(12):3530–3543, 2017.

68. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microbial Genomics*, 3(10):132-139, 2017.

69. Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Computational Biology*, 13(6):5955-5976, 2017.

70 Miller, J. R. *et al.* Hybrid assembly with long and short reads improves discovery of gene family expansions. *BMC Genomics*, 18:541-557, 2017.

71. Ruan, Z. *et al.* Hybrid genome assembly and annotation of a pandrug-resistant *Klebsiella pneumoniae* strain using nanopore and illumina sequencing. *Infection and Drug Resistance*, 13:199–206, 2020.

72. Kono, N. & Arakawa, K. Nanopore sequencing: Review of potential applications in functional genomics. *Development Growth and Differentiation*, 6(5):316–326, 2019.

73. Eccles, D., White, R., Pellefigues, C., Ronchese, F. & Lamiable, O. Investigation of chimeric reads using the MinION. *F1000Research*, 6(2):631-648, 2017.

74. Wick, R. R., Judd, L. M. & Holt, K. E. Deepbinner: Demultiplexing barcoded Oxford Nanopore reads with deep convolutional neural networks. *PLoS Computational Biology*, 14(11):6583-6594, 2018.

75. Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research*, 27:737–746, 2017.

76. Ambrose, S. J., Harmer, C. J. & Hall, R. M. Compatibility and entry exclusion of IncA and IncC plasmids revisited: IncA and IncC plasmids are compatible. *Plasmid*, 96:7–12, 2018.

77. Wu, W. *et al.* NDM metallo-β-lactamases and their bacterial producers in health care settings. *Clinical Microbiology Reviews*, 32(2):115-160, 2019.

78. Carattoli, A., Villa, L., Poirel, L., Bonnin, R. A. & Nordmann, P. Evolution of IncA/C blaCMY-2-carrying plasmids by acquisition of the blaNDM-1 carbapenemase gene. *Antimicrobial Agents and Chemotherapy*, 56:783–786, 2012.

79. Paskova, V. *et al.* Characterization of NDM-encoding plasmids from Enterobacteriaceae recovered from Czech hospitals. *Frontiers in Microbiology*, 9:1549-1561, (2018).

80. Hancock, S. J. *et al.* Identification of IncA/C plasmid replication and maintenance genes and development of a plasmid multilocus sequence typing scheme. *Antimicrobial Agents and Chemotherapy*, 61(2):1740-1757, 2017.

81. Partridge, S. R., Kwong, S. M., Firth, N. & Jensen, S. O. Mobile genetic elements associated with antimicrobial resistance. *Clinical Microbiology Reviews*, 31(4):8817-8888, 2018.

82. Reyes, J. A., Melano, R., Cárdenas, P. A. & Trueba, G. Mobile genetic elements associated with carbapenemase genes in South American Enterobacterales. *Brazilian Journal of Infectious Diseases*, 24(3):231–238, 2020.

83. Liu, L., Feng, Y., McNally, A. & Zong, Z. BlaNDM-21, a new variant of blaNDM in an *Escherichia coli* clinical isolate carrying blaCTX-M-55 and rmtB. *Journal of Antimicrobial* Chemotherap,y 73(9):2336–2339, 2018.

84. Wailan, A. M. *et al.* Genomic characteristics of NDM-producing Enterobacteriaceae isolates in Australia and their blaNDM genetic contexts. *Antimicrobial Agents and Chemotherapy*, 60(1):136–141, 2016.

85. Sarma, J., Bhattacharya, P., Kalita, D. & Rajbangshi, M. Multidrug-resistant Enterobacteriaceae including metallo-β-lactamase producers are predominant pathogens of healthcare-associated infections in an Indian teaching hospital. *Indian Journal of Medical Microbiology*, 29(1):22–27, 2011.

86. Toleman, M. A. & Walsh, T. R. Combinatorial events of insertion sequences and ICE in gram-negative bacteria. *FEMS Microbiology Reviews*, 35(5):912–935, 2011.

87. Partridge, S. R., Tsafnat, G., Coiera, E. & Iredell, J. R. Gene cassettes and cassette arrays in mobile resistance integrons: Review article. *FEMS Microbiology Reviews*, 33(4):757–784, 2009.

88. Partridge, S. R. Analysis of antibiotic resistance regions in Gram-negative bacteria. *FEMS Microbiology Reviews*, 35(5):820–855, 2011.

89. Novick, R. P. Plasmid incompatibility. *Microbiological Reviews*, 5(4):381–395, 1987.

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

91. San Millan, A. & MacLean, R. C. Fitness Costs of Plasmids: A Limit to plasmid transmission. in *Microbial Transmission*, 5(5):65–79, 2017.

92. Villa, L., García-Fernández, A., Fortini, D. & Carattoli, A. Replicon sequence typing of IncF plasmids carrying virulence and resistance determinants. *Journal of Antimicrobial Chemotherapy*, 65(12):2518–2529, 2010.

93. Yang, Q. E. *et al.* IncF plasmid diversity in multi-drug resistant *Escherichia coli* strains from animals in China. *Frontiers in Microbiology*, **6**:964-973, 2015.

Supplementary Data

***Klebsiella pneumoniae* C0612**

Following hybrid assembly, AMR genes were predicted on complete chromosome and four complete circular plasmids. These four complete circular plasmids and its AMR gene predictions are listed below.

- **plasmid_682 (Conjugative)**: OXA-48 (perfect), CTX-M-9 (perfect), APH (3")-Ib, APH(6)-Id, APH(3")-Ib (strict), APH(6)-Id, APH(3')-VIb
- **plasmid_207 (Non-Mobilizable)**: No AMR gene predictions
- **plasmid_93 (Mobilizable)**: No AMR gene predictions
- **plasmid_726 (Non-Mobilizable)**: tet(D) (perfect), AAC(3)-IIe (strict), mphA (perfect), dfrA14 (strict), TEM-1 (perfect), OXA-9 (perfect), aadA (strict), AAC(6')-Ib7, CTX-M-15 (perfect), QnrS1 (perfect), APH(3')-VI (perfect), **NDM-1 (perfect)**, BRP(MBL) (perfect), mphA (pefect).

One plasmid is conjugative, one is mobilizable and two are non-mobilizable. Based on the AST result shown in Tables S1a and S1b, I hypothesize that the presence of plasmid_682 or plasmid_726 in this isolate makes this isolate resistant to penam and cephalosporin antibiotics: ampicillin, amoxicillin clavulanic acid, piperacillin tazobactam, cefazolin, cefalotin, cefixime, ceftazidime and ceftriaxone caused by either OXA-48, OXA-9, CTX-M-9, CTX-M-15 or TEM-1. This could also be caused by class A beta-lactamase SHV-212 beta-lactamase predicted on the bacterial chromosome. The presence of QnrS1 in plasmid_726 makes this isolate resistant to ciprofloxacin (fluoroquinolone antibiotics) and nitrofurantoin or it could be due to the group of resistance-nodulation-cell division (RND) antibiotic efflux pump present in the chromosome. This isolate is also resistant to amikacin, tobramycin, and gentamicin which could be caused by any of the aminoglycosides resistant genes APH (3")-Ib, APH (6)-Id, APH(3")-Ib (strict), APH(6)-Id, APH(3')-VIb, ANT(3") in

plasmid_682 or plasmid_726. The presence of tet(D) in plasmid_726 makes this isolate resistant to Tetracycline antibiotics. Most importantly, the resistance of this isolate to broad-spectrum carbapenems such as meropenem and ertapenem could be due to the presence of NDM-1 in plasmid_726. On the other hand, mph(A) and dfrA14 are predicted as resistance genes on plasmid_726, but no known antibiotics that it confers resistance to are reported in the MIC result.Based on MOBsuite mash neighbor predictions, plasmid_726 references plasmid pKp_Goe (CP018737) as its identical plasmid. Plasmid pKp_Goe was isolated from *Klebsiella pneumoniae* strain Kp_Goe121641. The predicted AMR genes on this plasmid are listed below

- **pKp_Goe121641**: AAC (3)-IIe (Strict), catB3 (Strict), OXA-1 (Strict), AAC(6)-Ib-cr (Perfect), AAC(6)-Ib10 (Strict), aadA (Strict), OXA-9 (Perfect), TEM-1 (Perfect), CTX-M-15 (Perfect), dfrA14 (Strict).

Plasmid_726 vs plasmid pkP_Goe121641

The size of NDM-1 plasmid_726 is 94836 bp whereas the size of mash neighbor plasmid pkP_Goe121641 is 72, 952bp. 7 similar AMR genes are found on both plasmids. The gene of interest NDM-1 is not present in pKp_Goe121641. To compare these two plasmids, we mapped the Illumina read of isolate C0612 to both plasmids as shown in Figure S1A and 1B below. In Figure S1A, 13,710 reads maps to plasmid_726 uniformly. In Figure S1B, 10,241 reads map to plasmid pKp-Goe_641, and there is a little gap where the reads do not map. No AMR gene is present in this unmapped region.

***Escherichia coli* C0648**

Following hybrid assembly, AMR genes were predicted on complete chromosome and eight complete plasmids as listed below.

- **plasmid_novel_0 (Non-mobilizable)**: No AMR gene predictions
- **plasmid_299 (Mobilizable)**: No AMR gene predictions
- **plasmid_141 (Mobilizable)**: No AMR gene predictions
- **plasmid_10 (Mobilizable)**: No AMR gene predictions
- **plasmid_469 (Mobilizable)**: No AMR gene predictions
- **plasmid_664 (Conjugative)**: No AMR gene predictions
- **plasmid_486 (Conjugative)**: CMY-42 (perfect)
- **plasmid_1550 (Conjugative)**: tet(A) (strict), dfrA12 (perfect), aadA12 (perfect), sul1 (perfect), **NDM-5 (perfect)**, rmtB (perfect), TEM-1 (perfect), mphA (perfect), sul1 (perfect), aadA5 (perfect), dfrA17 (strict).

Three plasmids are conjugative, four are non-mobilizable, and one is mobilizable. Six of the eight plasmids have no AMR genes predicted on them. The NDM-5 bearing plasmid_1550 which was predicted to be non-mobilizable from the Illumina assembly data is predicted as conjugative from the hybrid assembly data. Based on the AST result shown in Table S2A and S2B, I hypothesize that the presence of plasmid_1550 and plasmid_486 makes this isolate resistant to beta-lactam antibiotics: ampicillin, amoxicillin clavulanic acid, piperacillin tazobactam, cefazolin, cefalotin, cefixime, ceftazidime and ceftriaxone caused by TEM-1, CMY-42 or NDM-5. Alternatively, the presence of multi-drug efflux genes in the chromosome like acrR, soxR, soxS, marR, ampH, ampC could be highly responsible for this isolate to not only be resistant to beta-lactam antibiotics, but also the fluoroquinolones, aminoglycosides and tetracycline antibiotics.

Based on MOBsuite mash neighbor predictions, plasmid_1550 for these isolate references plasmid pECY55 (KU043115) as its identical plasmid. This plasmid was isolated from *Escherichia coli* strain Y5. The predicted AMR genes on the plasmid are listed below

- **pECY55:** tet(A) (strict), AAC (6')-Ib-cr (Perfect), OXA-1 (Strict), cat3 (Strict), CTX-M-15 (Perfect), dfrA17 (Strict), aadA5 (Perfect), sul1 (Perfect), mphA (Perfect).

Plasmid_1550 vs plasmid pECY55

The size of NDM-5 plasmid_1550 is 134,146bp, whereas the size of pECY55 is 124,378bp. The predicted NDM-5 gene on plasmid_1550 is not present in pECY55. Five similar AMR genes are found on both plasmids. The two plasmids are compared by read mapping of Illumina reads to the plasmid sequence as shown in Figure S2A and S2B below. In Figure S2A, 57,614 reads maps to the plasmids_1550, and there is uniform distribution of reads. In Figure S2B, 53, 789 reads map to plasmid pECY55. There is also a little region where the reads do not map to, and it has no AMR gene.

***Klebsiella pneumoniae* C0650**

Following hybrid assembly, AMR gene were predicted on complete chromosome and three complete plasmids.

- **plasmid_2088 (Mobilizable)**: OXA-232 (Perfect)
- **plasmid_770 (Conjugative)**: QnrB4 (Perfect), DHA-17 (Strict), sul1 (Perfect)
- **plasmid_1220 (Conjugative)**: BRP(MBL) (perfect), **NDM-1 (perfect)**, APH(3')-VI (Perfect), QnrS1 (perfect), CTX-M-15 (perfect), AAC(6')-Ib7 (Strict), aadA (Strict), OXA-9 (Strict), TEM-183 (Strict).

Two plasmids are conjugative, and one is non-mobilizable. The NDM- plasmid which was predicted to be non-mobilizable from the Illumina assembly data is predicted as conjugative from the hybrid assembly data. Based on the AST results shown in Tables S3A and S3B, I hypothesize that presence of one or more beta-lactam resistant genes in any of these three plasmids makes this isolate resistant to penam and cephalosporin antibiotics: ampicillin,

amoxicillin clavulanic acid, piperacillin tazobactam, cefazolin, cefalotin, cefixime, ceftazidime and ceftriaxone. Alternatively, this could also be due to class A beta-lactamase SHV-212 beta-lactamase predicted on the isolate's chromosome. This isolate is also resistant to tobramycin which could be due to any of the aminoglycosides resistant genes APH (3')-VI or AAC(6')-Ib7 (Strict) in plasmid_1220. The presence of NDM-1 in plasmid_1220 makes this isolate resistant to ertapenem and meropenem.

Based on MOBsuite mash neighbor predictions, plasmid_1220 references blaNDM-1 plasmid p2 (CP009115) as its identical plasmid. Plasmid p2 was isolated from a *Klebsiella pneumoniae* strain. The predicted AMR genes on this plasmid are listed below.

- **blaNDM-1 plasmid p2 (CP009115)**: TEM-1 (Perfect), OXA-9 (Perfect), aadA (Strict), AAC (6')-Ib7 (Strict), CTX-M-15 (Perfect), QnrS1 (Perfect), APH (3')-VI (Perfect), **NDM-1 (Perfect)**, BRP(MBL) (Perfect).

Plasmid_1220 vs blaNDM-1 plasmid p2

Both plasmids are conjugative and have similar plasmid backbone as shown in Figure S3A and S3B. The difference in the two plasmids are the AMR gene region. They are different in both OXA-9 and TEM-19 gene regions. The OXA-9 gene in both plasmids differ by a SNP. For the TEM gene, it is TEM-19 in putative plasmid_1220, but TEM-1 in NDM-1 plasmid_p2. The percentage identity of TEM-19 to TEM-1 in NDM-1_plasmid_p2 is 71.4 percent. The hypothesis here is that TEM-19 in reference plasmid lost some of its sequence since it is next to a transposable element. To compare how similar the two plasmids are, both plasmids were mapped by C0650 Illumina reads as shown in Figure S3C and S3D below. In Figure S3C, 77,557 reads maps to the plasmid_1220 and there is uniform distribution of reads. In Figure S3D, 77,549 reads map to plasmid blaNDM-1 plasmid p2.

***Klebsiella pneumoniae* GDW2216**

Following hybrid assembly, AMR genes were predicted on a complete circular chromosome and three complete plasmid.

- **plasmid_novel_0 (Non-mobilizable)**: No AMR gene predictions
- **plasmid_1220 (Conjugative)**: CTX-M-15 (perfect), AAC (6')-Ib-cr (strict), OXA-1 (perfect), catB3 (strict)
- **plasmid_955 (Mobilizable)**: AAC (3')-IId (perfect), sul1 (perfect), DHA-7 (strict), **NDM-1 (perfect)**, rmtC (perfect), sul1 (perfect), AAC(6')-Ib7

The NDM-1 plasmid is predicted to be mobilizable from the hybrid assembly data. Based on MOBsuite mash neighbor predictions, plasmid_955 references p1002-NDM1 (CP021206) as its identical plasmid, which was isolated from *Escherichia coli* strain Z1002. The predicted AMR genes on plasmid p1002-NDM1 are listed below

- **Plasmid p1002-NDM1 (CP021206):** BRP(MBL) (perfect), **NDM-1 (perfect),** rmtC (perfect), sul1 (perfect), AAC (6')-Ib10 (strict), mphA (perfect), AAC(3)-IId (perfect), TEM-1 (perfect), CTX-M-14 (perfect).

Plasmid_955 vs Plasmid p1002-NDM1

The size of NDM-1 plasmid_955 is 82,327bp whereas the size of plasmid p1002-NDM1 is 111,688bp. The number of similar AMR genes predicted on both plasmids is four. The two plasmids are compared by read mapping in Figure S4A and S4bB below. In Figure S4, 15,519 reads are assembled to plasmid_955 and 13,113 reads are assembled to plasmid p1002-NDM1.

***Klebsiella pneumoniae* GDW2218**

Following hybrid assembly, AMR genes were completely predicted on a circular chromosome and four complete plasmids.

- **plasmid_607 (Non-Mobilizable)**: No AMR gene predictions
- **plasmid_novel_0 (Non-Mobilizable)**: no AMR gene prediction
- **plasmid_1220 (conjugative)**: SHV-11 (perfect), QnrB1 (perfect), dfrA14 (strict)
- **plasmid_955 (Conjugative)**: CMY-6 (perfect), AAC (6')-Ib7 (strict), sul1 (perfect), rmtC (perfect), **NDM-1 (perfect)**, BRP(MBL) (perfect).

Two plasmids are conjugative, and two are non-mobilizable. Plasmid_955 remains conjugative after hybrid assembly. Based on MOBsuite mash neighbor prediction, plasmid_955 of GDW2218 references pKP-NDM-1 (KF002018) as its identical plasmid. Plasmid pKP-NDM-1 is conjugative and isolated from *Klebsiella pneumoniae* strain KP1. The predicted AMR genes on this plasmid are likeplasmid_955.

- **Plasmid pKP-NDM-1 (KF002018)**: CMY-6 (perfect), AAC (6')-Ib7 (strict), sul1 (perfect), rmtC (perfect), **NDM-1 (perfect)**, BRP(MBL) (perfect)

Plasmid_955 vs pKP-NDM-1

Both plasmids are identical with the same size and predicted AMR genes shown in Figure S5A and S5B. The two plasmids are compared by read mapping in Figure S5C and S5C. In Figure S5, 43, 949 reads are mapped to putative plasmid_955 compared to 43,946 reads mapped to pKP-NDM1 plasmid.

***Klebsiella pneumoniae* GDW2280**

Following hybrid assembly, AMR genes were predicted on a complete circular chromosome and six plasmids.

- **plasmid_novel_0 (Non-mobilizable)**: No AMR gene predictions
- **plasmid_765 (Mobilizable)**: No AMR gene predictions
- **plasmid_1705 (Non-mobilizable)**: No AMR gene predictions

- **plasmid_727 (Non-mobilizable)**: APH (3")-Ib (strict), APH(6)-Id (strict), AAC(6')-Ib-cr (perfect), OXA-1 (strict), catB3 (strict), CTX-M-15 (perfect), TEM-1 (perfect), QnrS1 (strict), sul1 (perfect), dfrA1 (strict), tetA (strict).
- **plasmid_80 (Non-mobilizable)**: No AMR gene predictions.
- **plasmid_966 (Conjugative)**: dfrA12 (perfect), aadA2 (perfect), sul1 (perfect), armA (perfect), msrE (perfect), mphE (perfect), sul2 (perfect), **NDM-1 (perfect)**, BRP(MBL) (strict), SHV-66 (strict)

One plasmid is conjugative, one is mobilizable and four plasmids are non-mobilizable. Based on MOBsuite mash neighbor predictions, plasmid_966 references plasmid pPG010208 (HQ023861) isolated from *Escherichia coli* strain PG010208 as its identical plasmid. The predicted AMR genes on plasmid pPG010208 are listed below

- **Plasmid pPG010208 (HQ023861):** mphE (strict), msrE (perfect), floR (strict), tet(A), APH (6)-Id (strict), APH (3'')-Ib (strict), sul2 (perfect).

Plasmid_966 vs Plasmid pPG010208

The size of plasmid_966 is 159,260 bp compared to the size of plasmid pPG010208 which is 135,803bp, and isolated from *Escherichia coli* strain PG010208. The two plasmids are compared by read mapping in Figure S6A and S6B below. In Figure S6, 109, 437 reads are assembled to plasmid_966 while 82,068 reads are assembled to plasmid pPG010208. For plasmid pPG010208, there are few gaps where the reads do not map to.

***Morganella morganii* GDW2281**

Following hybrid assembly, AMR gene were predicted on a chromosome and three complete plasmids.

- **plasmid_58 (Non-mobilizable)**: No AMR gene predictions
- **plasmid_161 (Non-mobilizable)**: No AMR gene predictions
- **plasmid_2719 (Non-mobilizable)**: **NDM-1 (Perfect)**, sul1 (Perfect)

All three plasmids are predicted to be complete and circular here. All three plasmids are non-mobilizable including the NDM-1 plasmid. Based on MOBsuite mash neighbor predictions, plasmid_2719 references plasmid_pkp_Goe_579 (CP018317) from *Klebsiella pneumoniae* strain Kp_Goe_822579 as its identical plasmid

- **plasmid_pkp_Goe_579 (CP018317):** No AMR gene predictions.

Plasmid_2719 vs Plasmid_pkp_Goe_579

The size of plasmid_2719 is 6541 bp whereas the size of plasmid pkp_Goe_579 is 4249 bp, isolated from *Klebsiella pneumoniae* strain Kp_Goe_822579. The two plasmids are compared by read mapping in Figure S7A and S7B. In Figure S7, 4,586 reads map uniformly to plasmid_2719 and 12,885 reads map to plasmid_Kp_Goe_579

***Providencia stuartii* GDW2284**

Following hybrid assembly, AMR genes were predicted on chromosome and two plasmids.

- **plasmid_815 (Mobilizable)**: No AMR gene predictions
- **plasmid_955 (Conjugative)**: CMY-6 (Perfect), AAC (6')-Ib10 (Strict), sul1 (Perfect), rmtC (Perfect), AAC (3)-IId (Perfect), APH(3')-VI (Perfect), **NDM-1 (Perfect)**, BRP(MBL) (Strict)

The two plasmids are predicted to be complete and circular here. The NDM plasmid is predicted to be conjugative following hybrid assembly of the genome. Based on MOBsuite mash neighbor predictions, plasmid_955 references plasmid pNDM-PstGN576 (KJ802405) from *Providencia stuartii* isolate GN576. The predicted AMR genes on plasmid pNDM-PstGN576 are listed below

- **Plasmid pNDM-PstGN576 (KJ802405):** CMY-6 (perfect), AAC (6')-Ib10 (strict), sul1 (perfect), rmtC (perfect), AAC (3)-IId (perfect), APH(3')-VI (perfect), NDM-1 (perfect), BRP(MBL) (strict).

Plasmid _955 vs Plasmid pNDM-PstGN576

The size of plasmid_955 with NDM-1 is 147866bp whereas the size of plasmid pNDM-PstGN576 is 147886bp. Both plasmids have the same AMR genes and share similar taxon. The two plasmid sequences are compared by read mapping in Figure S8A and S8bB respectively. The same number of reads (152,020) map to both plasmids.

***Escherichia coli* GDW2285**

Following hybrid assembly, AMR genes were predicted on a chromosome and on four plasmids.

- **plasmid_58 (Non-mobilizable)**: No AMR gene prediction
- **plasmid_10 (Mobilizable)**: No AMR gene prediction
- **plasmid_1550 (Mobilizable)**: CTX-M-15 (perfect), catB3 (strict), OXA-1 (strict), AAC (6')-Ib-cr
- **plasmid_955 (conjugative)**: CMY-6 (perfect), AAC (6)'-Ib-cr (strict), sul1 (Perfect), rmtC (perfect), **NDM-1 (perfect)**, BRP(MBL) (strict).

One plasmid is conjugative, two are mobilizable, and the other is non-mobilizable. The NDM plasmid is predicted to be conjugative from the hybrid assembly data. Based on MOBsuite mash neighbor predictions, plasmid_955 references plasmid pMS6198A (CP015835) isolated from *Escherichia coli* strain MS1698 as its identical plasmid. The predicted AMR genes on plasmid_955 are listed below.

- **Plasmid pMS6198A (CP015835):** CMY-6 (perfect), AAC (6)'-Ib10 (strict), sul1 (Perfect), rmtC (perfect), **NDM-1 (perfect)**, BRP(MBL) (strict).

Plasmid_955 vs Plasmid pMS6198A

The two plasmids are of the same plasmid types and have similar backbone as shown in Figure S9A and S9B. The size of NDM-1 plasmid_955 is 118768 bp whereas the size of plasmid pMS6198A is 137, 565 bp. Both plasmids have similar AMR genes and the

missing region in pMS619A does not encode any AMR genes. In Figure S9C, 32,438 reads map to plasmid_955 while 32, 401 reads map to plasmid pMS6198A in Figure S9D. The gap where the reads do not map explains the difference between the two plasmid structures.

***Citrobacter freundii* GDW2286**

Following hybrid assembly, AMR genes were predicted on chromosome and two plasmid predictions.

- **plasmid_748 (Non-mobilizable):** aadA2 (Perfect), sul1 (Perfect), mphA (Perfect), AAC (3)-IId, TEM-1 (Perfect), catII (Strict), sul2 (Perfect)
- **plasmid_955 (Conjugative):** CTX-M-15 (Strict), CMY-6 (Perfect), AAC (6')-Ib7 (Strict), sul1 (Perfect), rmtC (Perfect), APH (3')-VI (Perfect), **NDM-1 (Perfect)**, BRP(MBL) (Perfect), TEM-1 (Perfect), sul1 (Perfect).

One plasmid is conjugative, and the other is non-mobilizable. Plasmid_955 remains conjugative following hybrid assembly, but less AMR genes are predicted on it. Just like GDW2285. plasmid_955 of GDW2286 references pMS6198A (CP015835) as its identical plasmid.

- **Plasmid pMS6198A (CP015835):** CMY-6 (perfect), AAC (6)'-Ib10 (strict), sul1 (Perfect), rmtC (perfect), **NDM-1 (perfect)**, BRP(MBL) (strict).

Plasmid_955 vs Plasmid pMS6198A

The size of NDM-1 plasmid_955 is 184749, whereas the size of plasmid pMS6198A is 137, 565 bp. The number of similar AMR genes predicted on both plasmids is five. The two plasmids are compared by read mapping in Figure S10A and S10B. In Figure S10 167, 842 reads map to plasmid_955 while 681 reads map to plasmid pMS6198A.

***Klebsiella pneumoniae* GDW2287**

Following hybrid assembly, AMR genes were predicted on chromosome and three complete plasmids.

- **plasmid_1220 (Conjugative)**: CTX-M-15 (perfect), AAC (6')-Ib-cr (strict), OXA-1 (perfect)
- **plasmid_novel_0 (Non-mobilizable)**: No AMR gene predictions
- **plasmid_955 (Conjugative)**: AAC (3)-IId (perfect), sul1 (perfect), DHA-7 (strict), **NDM-1 (perfect)**, rmtC (perfect), sul1 (perfect), AAC (6')-Ib7

Two plasmids are predicted to be conjugative and one is non-mobilizable. The NDM-1 plasmid is predicted to be conjugative from the hybrid assembly data. Based on MOBsuite mash neighbor predictions, NDM-1 borne plasmid_955 references plasmid p1002-NDM (CP021206) isolated from *Escherichia coli* strain Z1002 as its identical plasmid. The predicted AMR gene on plasmid p1002-NDM1 are listed below

- **Plasmid p1002-NDM1 (CP021206):** BRP(MBL) (perfect), **NDM-1 (perfect),** rmtC (perfect), sul1 (perfect), AAC (6')-Ib10 (strict), mphA (perfect), AAC (3)-IId (perfect), TEM-1 (perfect), CTX-M-14 (perfect).

Plasmid_955 vs Plasmid p1002-NDM1

The size of NDM-1 plasmid_955 is 82,327bp whereas the size of plasmid p1002-NDM1 is 111,688bp. The number of similar AMR genes predicted on both plasmids is four. The two plasmids are compared by read mapping in Figure 11a and 11b below. In Figure S11A, 18,824 reads map to plasmid_955, and in Figure S11B, 15784 reads map to plasmid p1002-NDM1. There are gaps where the reads do not map. The plasmid_955 of GDW2216 is identical to plasmid_955 of this isolate so the isolates carrying this plasmid is clones.

***Enterobacter cloacae* GDW2288**

Following hybrid assembly, AMR genes were predicted on a chromosome and four complete plasmids.

- **plasmid_621 (Conjugative)**: OXA-9(strict), ANT (3'')-IIa (strict), AAC (6')-Ib7 (strict), BRB(MBL) (strict), **NDM-1 (perfect)**, APH(3')-VI, (perfect), QnrS1 (perfect), CTX-M-15 (perfect).
- **plasmid_2088 (Mobilizable)**: OXA-232 (Perfect)
- **plasmid_novel_0 (Mobilizable)**: No AMR gene predictions
- **plasmid_novel_1 (Non-Mobilizable)**: No AMR gene predictions

One plasmid is conjugative, two plasmids are mobilizable and one plasmid is non-mobilizable. The NDM-1 plasmid_621 remains conjugative following hybrid assembly, and there are more AMR genes predicted on it. Based on MOBsuite mash neighbor predictions, plasmid_621 references pECL3-NDM-1 (KC887917) isolated from *Enterobacter cloacae* as its identical plasmid. The predicted AMR gene on plasmid pECL3-NDM-1 (KC887917) is listed below.

- **Plasmid pECL3-NDM-1 (KC887917)**: BRP(MBL) (perfect), **NDM-1 (perfect)**, sul1 (strict), rmtC (perfect).

Plasmid_621 vs Plasmid pECL3-NDM1

The size of plasmid_621 is 108467bp whereas the size of plasmid pECL3-NDM1 is 99, 435bp. The two plasmids are compared by read mapping in Figure S12A and S12B In Figure S12, 49869 reads map to plasmid_621 while 45,031 reads map to plasmid pECL3-NDM1

***Escherichia coli* GDW2289**

Following hybrid assembly, AMR genes were predicted on chromosome and eight complete plasmids.

- **plasmid_1555 (Conjugative)**: tetB (strict), mphA (perfect), sul1 (perfect), aadA5 (perfect), CTX-M-15 (perfect), AAC (3)-IIe (strict), catB3 (strict), OXA-1 (strict), AAC (6')-Ib-cr (perfect), tetR (strict) - Size: 169806
- **plasmid_1008 (Non-mobilizable)**: No AMR gene prediction - Size: 159607
- **plasmid_546 (Conjugative)**: TEM-1 (perfect), mphA (strict) - size: 66701
- **plasmid_607 (Mobilizable)**: No AMR gene prediction
- **plasmid_839 (Non-mobilizable)**: No AMR gene prediction
- **plasmid_novel_0 (Mobilizable)**: No AMR gene prediction
- **plasmid_novel_1 (Mobilizable)** No AMR gene predictions.
- **plasmid_934 (Conjugative)**: dfrA5 (perfect), BRP(MBL) (perfect), **NDM-1 (perfect)**

Three plasmids are conjugative, three are mobilizable and two are non-mobilizable. The NDM-1 plasmid remains conjugative following hybrid assembly. Based on MOBsuite nearest neighbor, plasmid_943 references plasmid pHHA45 (NC_019098) isolated from *Escherichia coli* strain HHA45 as its identical plasmid. The predicted AMR genes on plasmid pHHA45 (NC_019098).

- **plasmid pHHA45 (NC_019098):** CTX-M-1 (perfect)

Plasmid_934 vs Plasmid pHHA45

The size of NDM-1 plasmid_934 is 60,362bp whereas the size of plasmid pHHA45 (NC_019098) is 39,510 bp. The AMR gene predictions on both plasmids are different. The two plasmids are compared by read mapping in Figure S13A and B In Figure S13, 29,597 reads maps to plasmid_934 and 17, 837 reads are assembled to plasmid pHHA45.

***Klebsiella oxytoca* GDW2290**

Following hybrid assembly, AMR genes were predicted on chromosome and two plasmid predictions.

- **Plasmid_novel_0 (Conjugative)**: No AMR gene predictions
- **Plasmid_1220 (Conjugative)**: TEM-1 (Perfect), OXA-9 (Strict), aadA (Strict), AAC (6')-Ib10 (Strict), CTX-M-15 (Perfect), QnrS1 (Perfect), APH(3')-VI (Perfect), **NDM-1 (Perfect)**, BRP(MBL) (Perfect), mphA (Perfect).

The two plasmids are complete and conjugative. The number of AMR gene predictions from Illumina assembly to hybrid assembly plasmid_1220 increased. Based on MOBsuite mash neighbor predictions, plasmid_955 references blaNDM-1 plasmid p2 (CP009115) isolated from *Klebsiella pneumoniae* strain as its identical plasmid. The predicted AMR genes on blaNDM-1 plasmid p2 (CP009115) are listed below

- **blaNDM-1 plasmid p2 (CP009115)**: TEM-1 (Perfect), OXA-9 (Perfect), aadA (Strict), AAC (6')-Ib7 (Strict), CTX-M-15 (Perfect), QnrS1 (Perfect), APH (3')-VI (Perfect), NDM-1 (Perfect), BRP(MBL) (Perfect).

Plasmid_1220 vs blaNDM-1 plasmid P2 (CP009115)

The size of NDM-1 plasmid_1220 is 140,507bp whereas the size of blaNDM-1 plasmid P2 is 118061 bp. Both plasmids have similar AMR gene predictions except for mphA in plasmid_1220. The two plasmids are compared by read mapping in Figure S14A and 14B. In Figure S14, 76585 reads map to plasmid_1220 and 58,100 reads map to blaNDM-1 plasmid p2, including few unmapped regions.

***Enterobacter cloacae* GDW2291**

Following hybrid assembly, AMR genes were predicted on chromosomes and six plasmids.

- **plasmid_111 (Mobilizable)**: No AMR gene predictions
- **plasmid_197 (Non_mobilizable)**: No AMR gene predictions
- **plasmid_215 (Mobilizable)**: No AMR gene predictions
- **plasmid_58 (Non-mobilizable)**: No AMR gene predictions

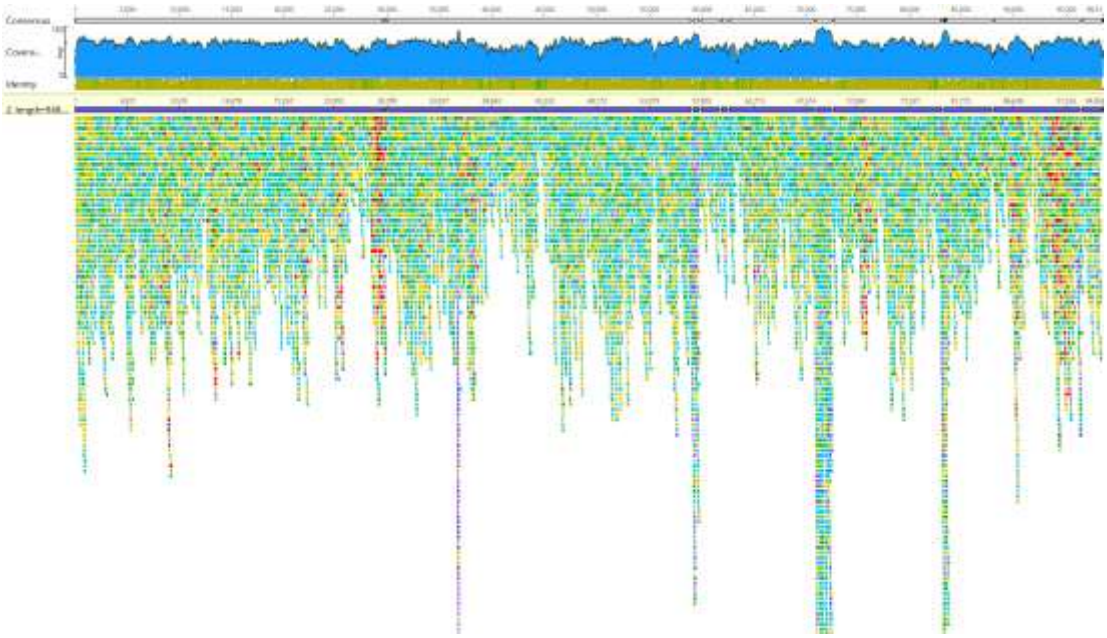
- **plasmid_728 (Non-mobilizable)**: sul2 (perfect), APH(3")-Ib (strict), APH(6)-Id (strict), TEM-1 (perfect), CTX-M-15 (perfect), AAC(3)-IId (perfect), dfrA12 (perfect), aadA2 (strict), sul1 (perfect), armA(perfect), QnrB1 (perfect), AAC(6')-Ib-cr (Perfect), OXA-1 (strict), catB3 (strict), dfrA14 (strict).
- **plasmid_654 (Mobilizable)**: APH (3')-VI (Perfect), NDM-1 (perfect), TEM-1 (perfect)

Three plasmids are mobilizable and three plasmids are non-mobilizable. Based on MOBsuite mash nearest neighbor, plasmid_654 references plasmid p112298-KPC (KP987215) as its identical plasmid which was isolated from Citrobacter freundii.

- **plasmid p112298-KPC**: KPC-2 (perfect), mphA (perfect), sul1 (perfect), arr-3 (perfect), catB3 (perfect), OXA-1 (perfect), AAC (6')-Ib-cr (perfect), FosA3 (perfect), CTX-M-14 (perfect).

The size of NDM-1 plasmid_654 is 113558bp whereas the size of plasmid p112298-KPC is 117288bp. No similar AMR gene predictions in both plasmids. The two plasmids are compared by read mapping in Figure S15A and S15b below. In Figure S15, 59,367 reads map to plasmid_654 with NDM1 gene and 43,404 reads map to plasmid p112298-KPC.

A)



B)

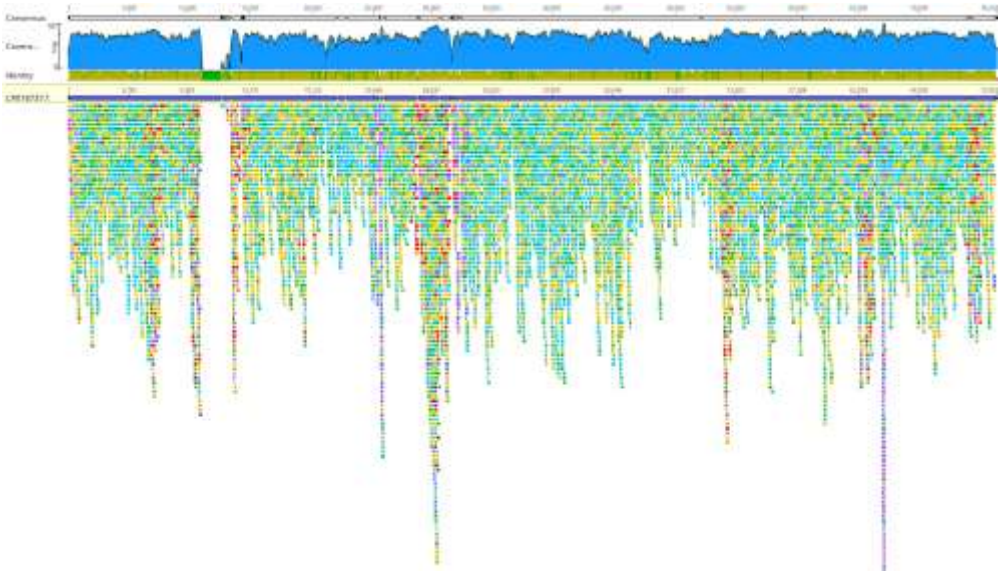
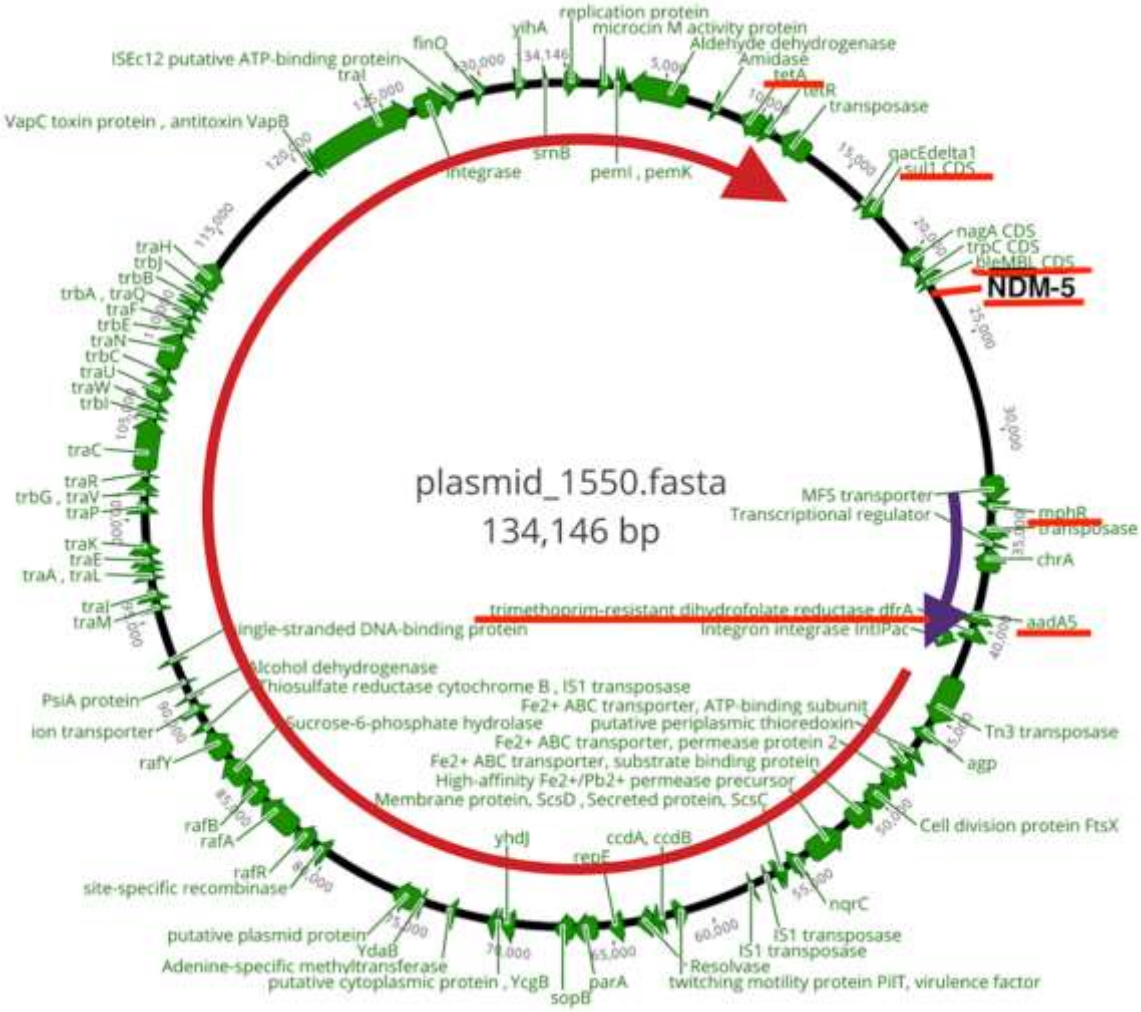
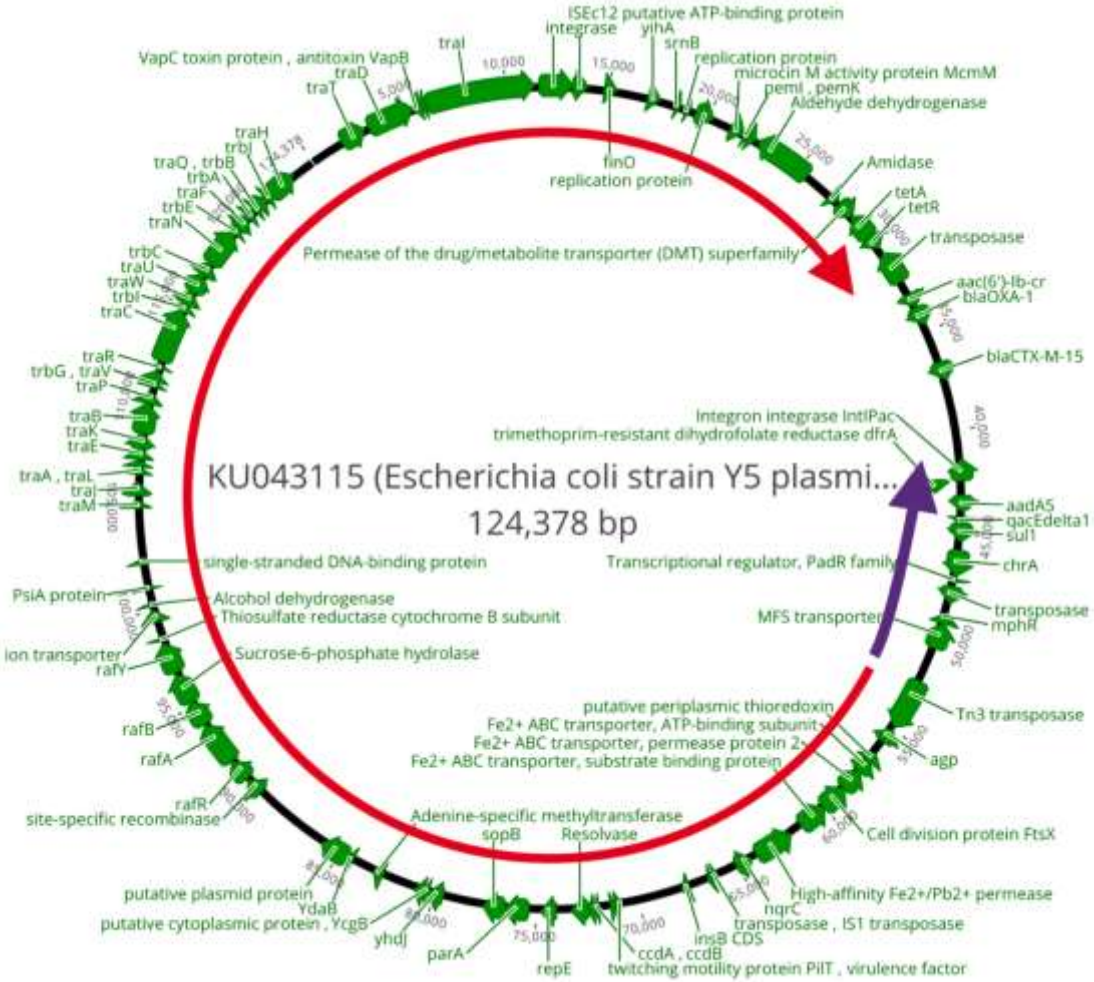


Figure S1: Read mapping of *Klebsiella pneumoniae* C0612 Illumina reads to; A) plasmid_726. B) mash neighbor plasmid pkP_Goe121641.

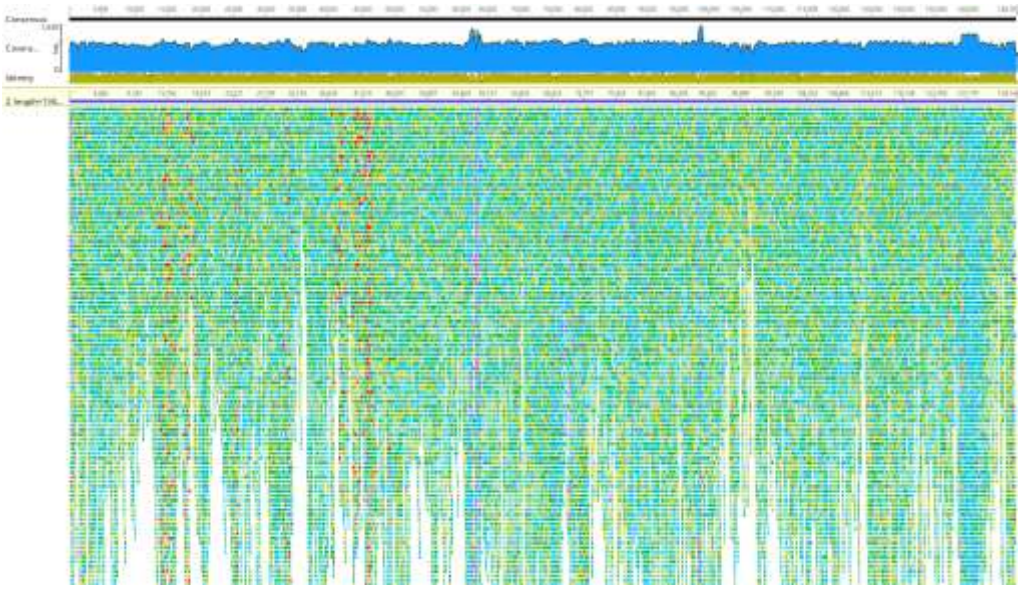
A)


plasmid_1550.fasta
134,146 bp
replication protein
microcin M activity protein
Aldehyde dehydrogenase
Amidase
tetA
tetR
transposase
qacEdelta1
sul1 CDS
nagA CDS
trpC CDS
NDM-5
MFS transporter
Transcriptional regulator
mphR
transposase
chrA
trimethoprim-resistant dihydrofolate reductase dfrA
Integron integrase IntI1Pac
aadA5
Tn3 transposase
agp
Cell division protein FtsX
nqrC
IS1 transposase
IS1 transposase
Resolvase
twitching motility protein PilT, virulence factor
parA
sopB
putative cytoplasmic protein , YcgB
Adenine-specific methyltransferase
putative plasmid protein
YdaB
site-specific recombinase
rafR
rafA
rafB
rafY
ion transporter
PsiA protein
Alcohol dehydrogenase
Thiosulfate reductase cytochrome B , IS1 transposase
Fe2+ ABC transporter, ATP-binding subunit
putative periplasmic thioredoxin
Fe2+ ABC transporter, permease protein 2
Fe2+ ABC transporter, substrate binding protein
High-affinity Fe2+/Pb2+ permease precursor
Membrane protein, ScsD , Secreted protein, ScsC
Sucrose-6-phosphate hydrolase
ingle-stranded DNA-binding protein
yhdJ
ccdA, ccdB
repE
traI
traM
traA , traL
traK
traE
traR
trbG , traV
traP
traC
traH
trbJ
trbB
trbA , traQ
traF
trbE
traN
trbC
traU
traW
trbI
VapC toxin protein , antitoxin VapB
ISEc12 putative ATP-binding protein
traI
finO
yihA
integrase
srnB
pemI , pemK

B)


KU043115 (Escherichia coli strain Y5 plasmi...
124,378 bp
ISEc12 putative ATP-binding protein
integrase
yihA
srnB
replication protein
microcin M activity protein McmM
pemI , pemK
Aldehyde dehydrogenase
Amidase
tetA
tetR
transposase
aac(6')-Ib-cr
blaOXA-1
blaCTX-M-15
Integron integrase IntIPac
trimethoprim-resistant dihydrofolate reductase dfrA
aadA5
qacEdelta1
sul1
chrA
transposase
mphR
Tn3 transposase
agp
Cell division protein FtsX
High-affinity Fe2+/Pb2+ permease
nqrC
transposase , IS1 transposase
insB CDS
ccdA , ccdB
twitching motility protein PilT , virulence factor
repE
parA
yhdJ
putative cytoplasmic protein , YcgB
YdaB
putative plasmid protein
site-specific recombinase
rafR
rafA
rafB
rafY
ion transporter
PsiA protein
single-stranded DNA-binding protein
Alcohol dehydrogenase
Thiosulfate reductase cytochrome B subunit
Sucrose-6-phosphate hydrolase
Transcriptional regulator, PadR family
MFS transporter
putative periplasmic thioredoxin
Fe2+ ABC transporter, ATP-binding subunit
Fe2+ ABC transporter, permease protein 2
Fe2+ ABC transporter, substrate binding protein
Adenine-specific methyltransferase
sopB
Resolvase
finO
replication protein
Permease of the drug/metabolite transporter (DMT) superfamily
traI
VapC toxin protein , antitoxin VapB
traD
traT
traH
trbJ
traQ , trbB
trbA
traF
trbE
traN
trbC
traU
traW
trbI
traC
traR
trbG , traV
traP
traB
traK
traE
traA , traL
traI
traM

C)



D)

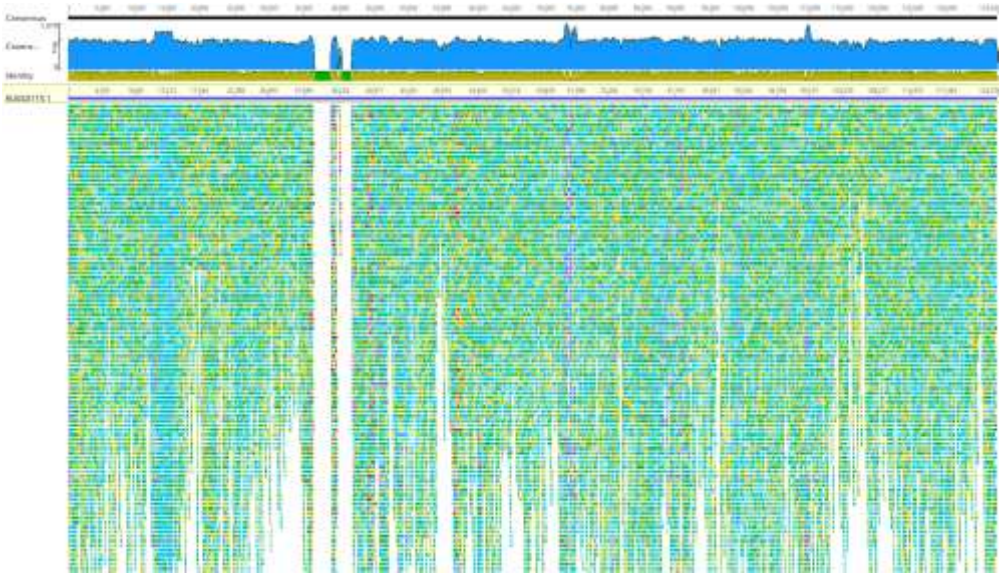
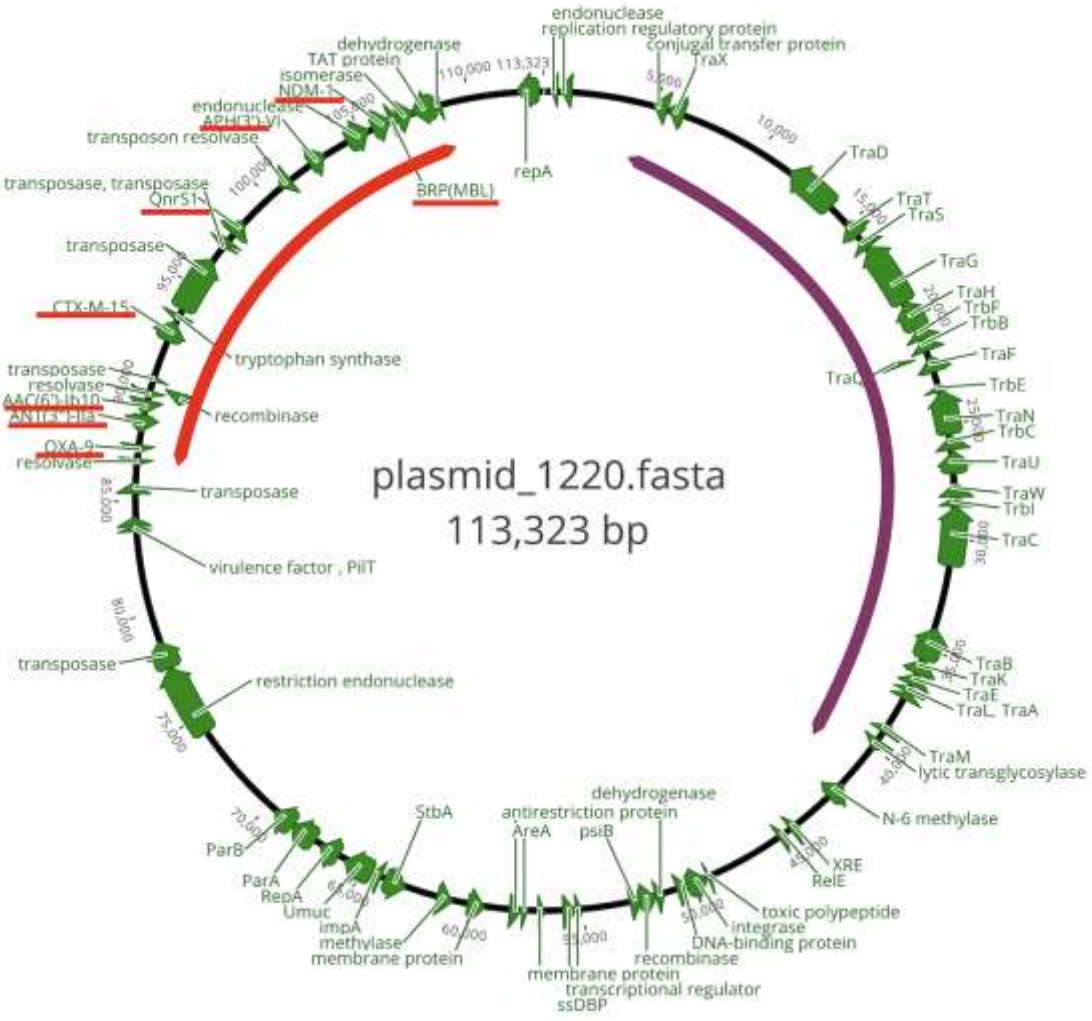
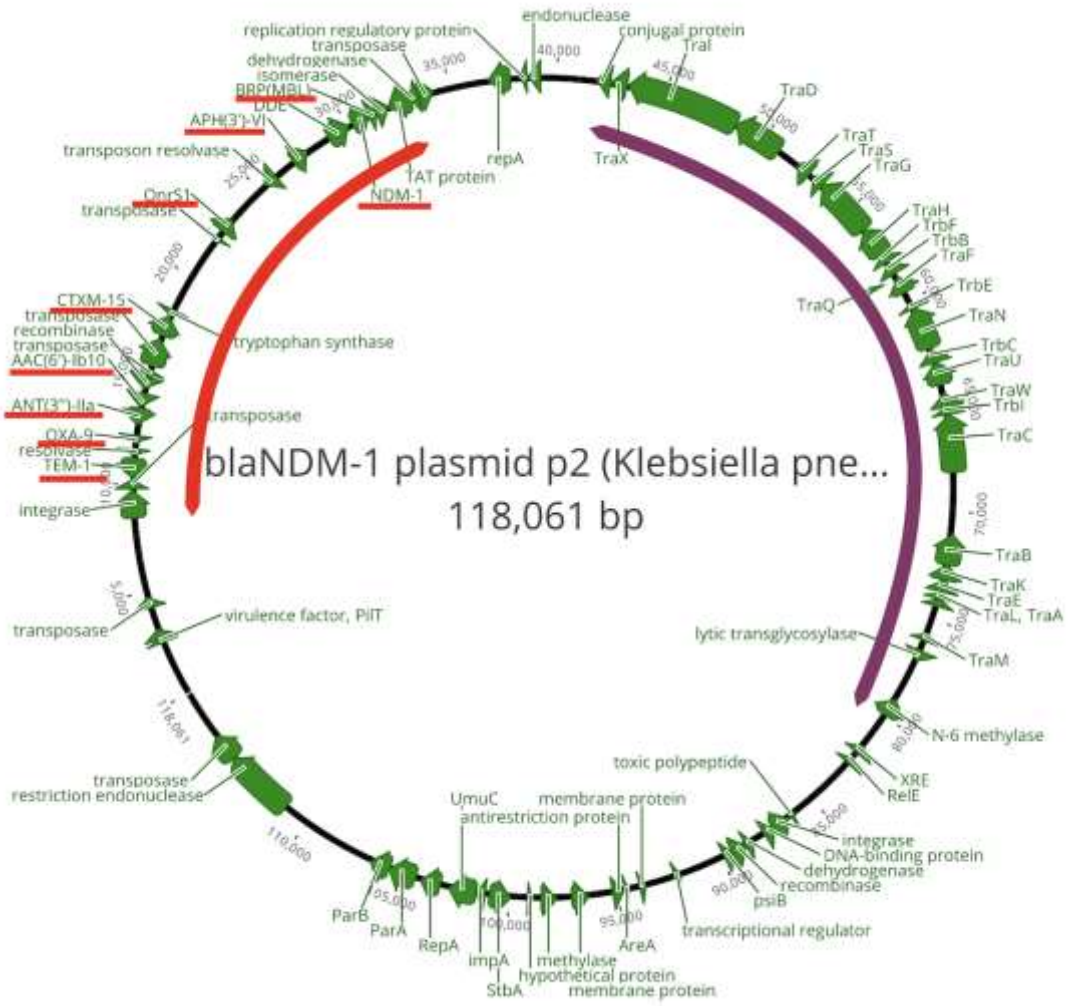


Figure S2: Annotated plasmid image and read mapping of NDM plasmid from C0648 *Escherichia coli* and reference plasmid pECY55 A) and B) annotated hybrid assembly NDM plasmid and mash neighbor reference plasmid pECY55. C) and D) read mapping to NDM plasmid and reference plasmid.

A)


plasmid_1220.fasta
113,323 bp
endonuclease
replication regulatory protein
conjugal transfer protein
TraX
repA
TraD
TraT
TraS
TraG
TraH
TrbF
TrbB
TraF
TraU
TrbE
TraN
TrbC
TraU
TraW
TrbI
TraC
TraB
TraK
TraE
TraL, TraA
TraM
lytic transglycosylase
N-6 methylase
XRE
RelE
toxic polypeptide
integrase
DNA-binding protein
recombinase
membrane protein
transcriptional regulator
ssDBP
dehydrogenase
antirestriction protein
AreA
psiB
StbA
ParB
ParA
RepA
Umuc
impA
methylase
membrane protein
restriction endonuclease
transposase
virulence factor , PilT
transposase
resolvase
OXA-9
transposase
recombinase
transposase
resolvase
tryptophan synthase
CTX-M-15
transposase
Qnr51
transposase, transposase
transposon resolvase
endonuclease
APH(3")-VI
isomerase
NDM-1
dehydrogenase
TAT protein
BRP(MBL)

B)


blaNDM-1 plasmid p2 (Klebsiella pne...
118,061 bp
endonuclease
conjugal protein
TraI
TraD
TraT
TraS
TraG
TraH
TrbF
TrbB
TraF
TrbE
TraN
TrbC
TraU
TraW
TrbI
TraC
TraB
TraK
TraE
TraL, TraA
TraM
N-6 methylase
XRE
RelE
integrase
DNA-binding protein
dehydrogenase
recombinase
psiB
transcriptional regulator
AreA
methylase
hypothetical protein
membrane protein
impA
StbA
RepA
ParA
ParB
UmuC
antirestriction protein
membrane protein
toxic polypeptide
lytic transglycosylase
virulence factor, PilT
transposase
restriction endonuclease
integrase
transposase
ANT(3'')-IIa
OXA-9
resolvase
TEM-1
CTXM-15
transposase
recombinase
transposase
AAC(6')-Ib10
tryptophan synthase
transposase
transposase
QnrS1
transposase
transposon resolvase
APH(3')-VI
BRP(MBL)
DDE
dehydrogenase
isomerase
transposase
replication regulatory protein
repA
TAT protein
NDM-1
TraX

C)

D)

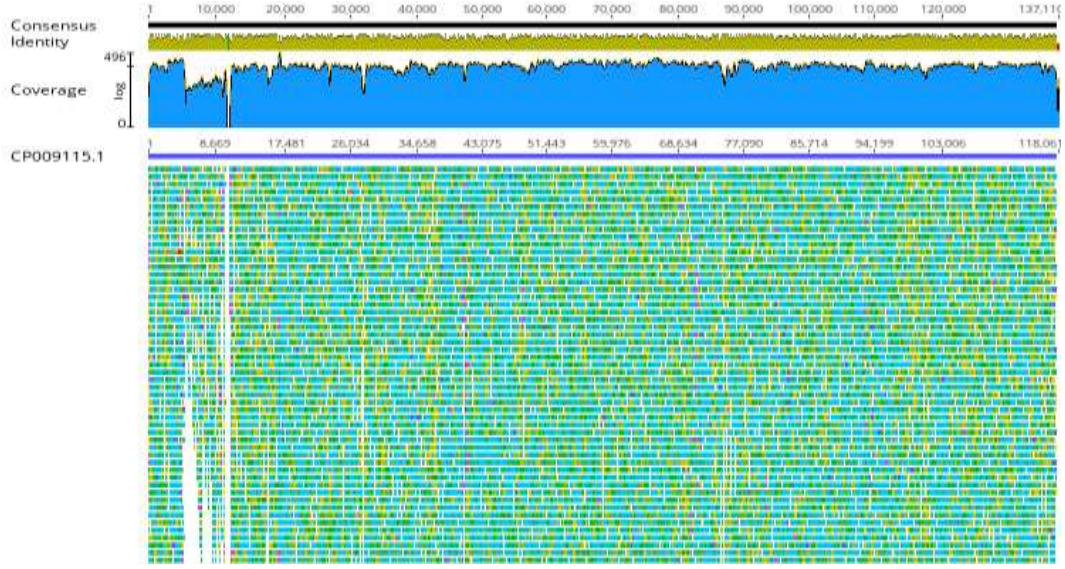
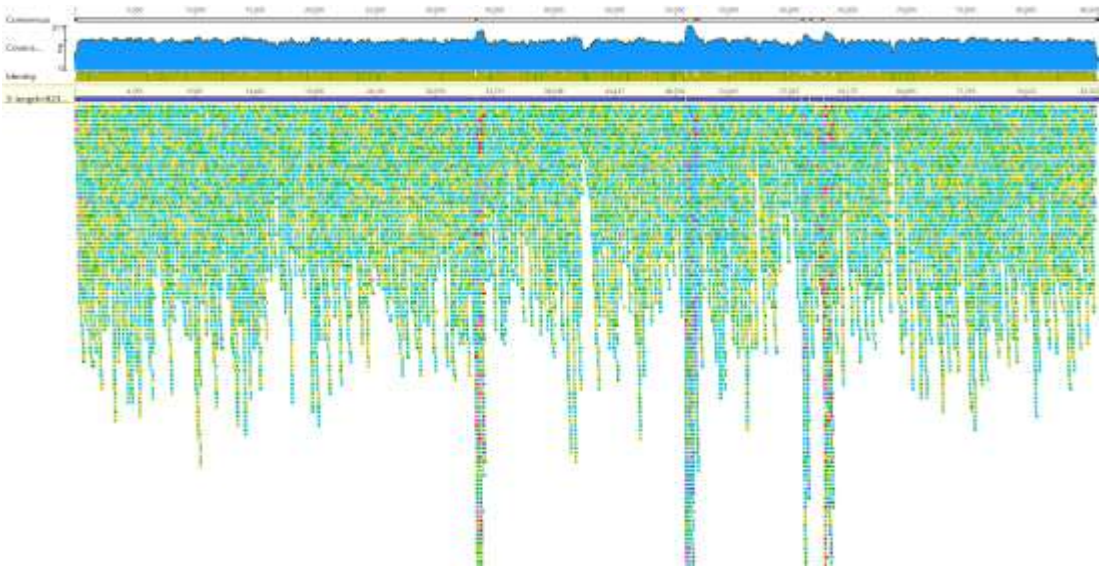


Figure S3: Annotated plasmid image and read mapping of NDM plasmid from C0650 *Klebsiella pneumoniae* and reference plasmid blaNDM-1 A) and B) annotated hybrid assembly NDM plasmid and mash neighbor reference plasmid blaNDM-1C) and D) read mapping to NDM plasmid and reference plasmid.

A)



B)

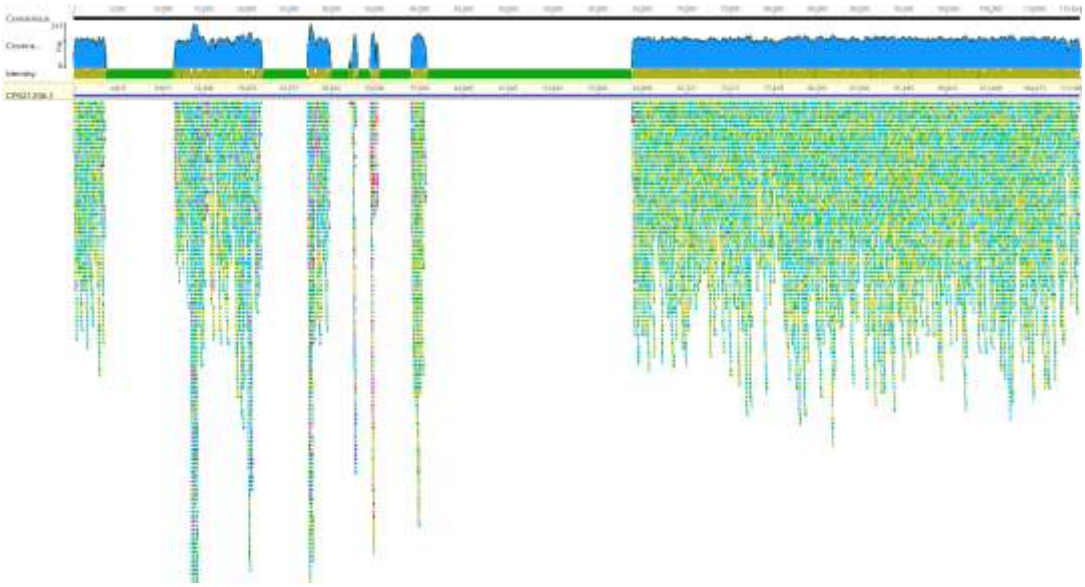
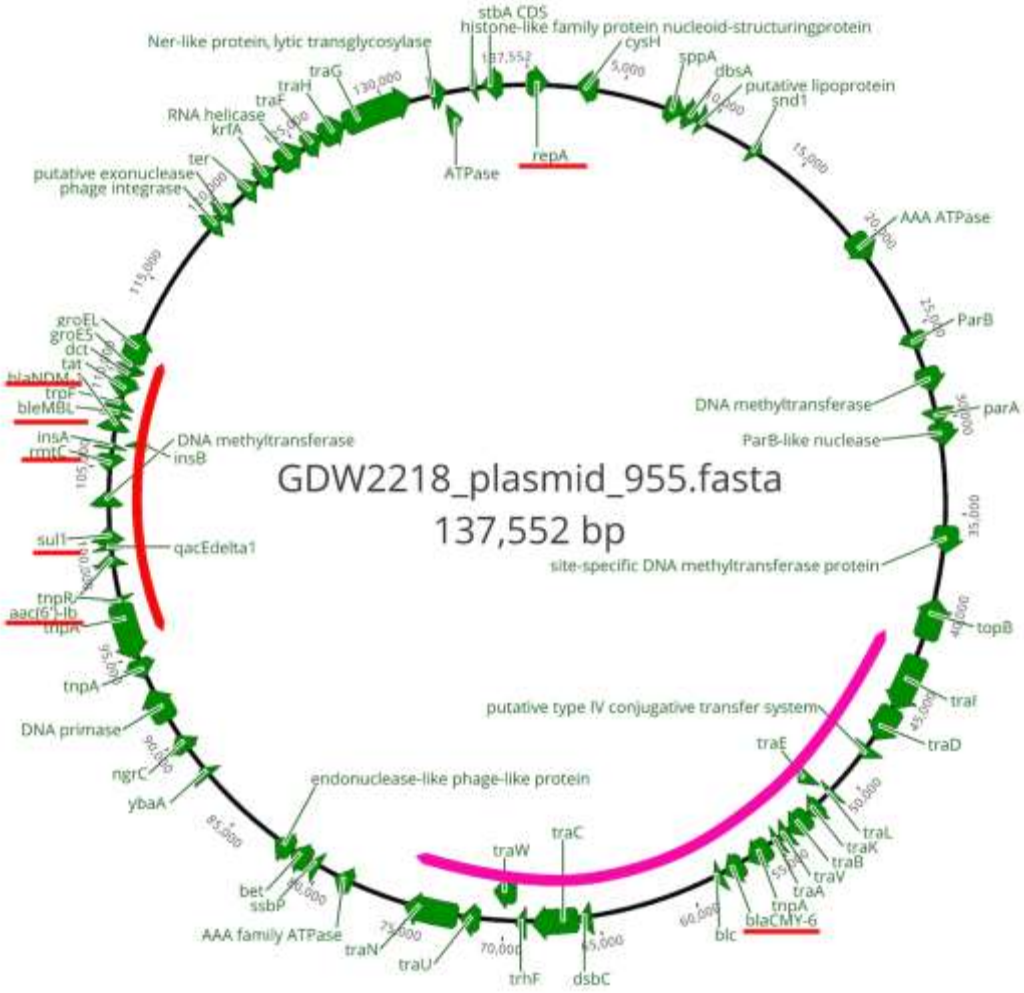
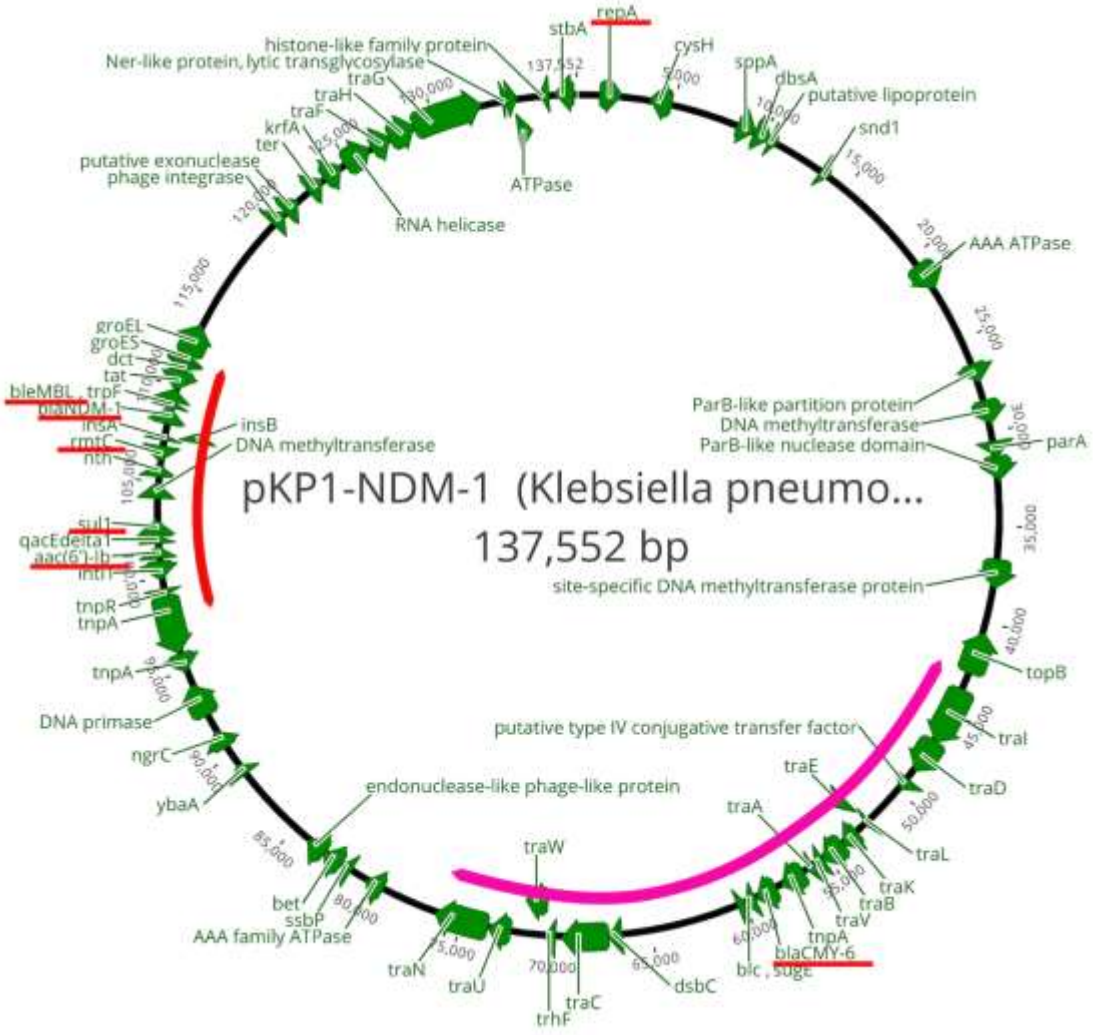
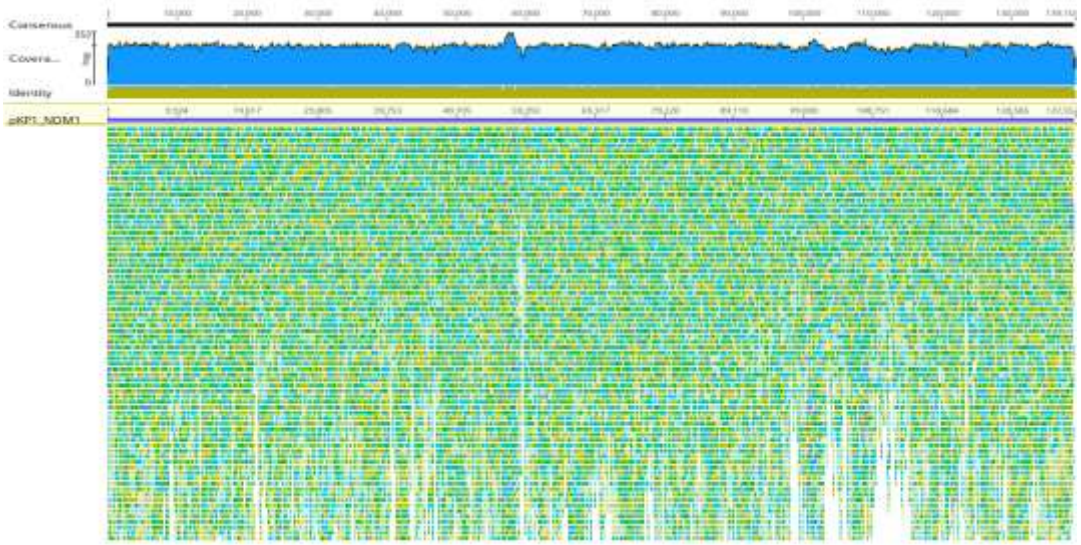


Figure S4: Read mapping of *Klebsiella pneumoniae* GDW2216 Illumina reads to; A) hybrid assembly NDM plasmid from GDW2216, and B) mash neighbor plasmid p1002-NDM1.


GDW2218_plasmid_955.fasta
137,552 bp
stbA CDS
histone-like family protein nucleoid-structuringprotein
cysH
sppA
dbsA
putative lipoprotein
snd1
AAA ATPase
ParB
parA
topB
traI
traD
traL
traK
traB
traV
traA
tnpA
blaCMY-6
blc
dsbC
trhF
traU
traN
AAA family ATPase
bet
ssbP
endonuclease-like phage-like protein
ybaA
ngrC
DNA primase
tnpA
tnpR
aacI6'-Ib
tnpA
sul1
qacEdelta1
insA
tnpC
DNA methyltransferase
insB
groEL
groES
dct
tat
blaNDM-1
trpF
bleMBL
putative exonuclease
phage integrase
ter
krfA
RNA helicase
traH
traG
traF
Ner-like protein, lytic transglycosylase
ATPase
repA
DNA methyltransferase
ParB-like nuclease
site-specific DNA methyltransferase protein
putative type IV conjugative transfer system
traE
traC
traW


pKP1-NDM-1 (Klebsiella pneumo...
137,552 bp
repA
stbA
cysH
sppA
dbsA
putative lipoprotein
snd1
AAA ATPase
parA
topB
traI
traD
traL
traK
traB
traV
tnpA
blaCMY-6
dsbC
traC
trhF
traU
traN
AAA family ATPase
ssbP
bet
endonuclease-like phage-like protein
ybaA
ngrC
DNA primase
tnpA
tnpR
qacEdelta1
sul1
insA
rmtC
bleMBL
trpF
dct
groES
groEL
ter
krfA
traF
traH
traG
Ner-like protein, lytic transglycosylase
histone-like family protein
ATPase
RNA helicase
putative exonuclease
phage integrase
insB
DNA methyltransferase
ParB-like partition protein
DNA methyltransferase
ParB-like nuclease domain
site-specific DNA methyltransferase protein
putative type IV conjugative transfer factor
traE
traA
traW

C)



D)

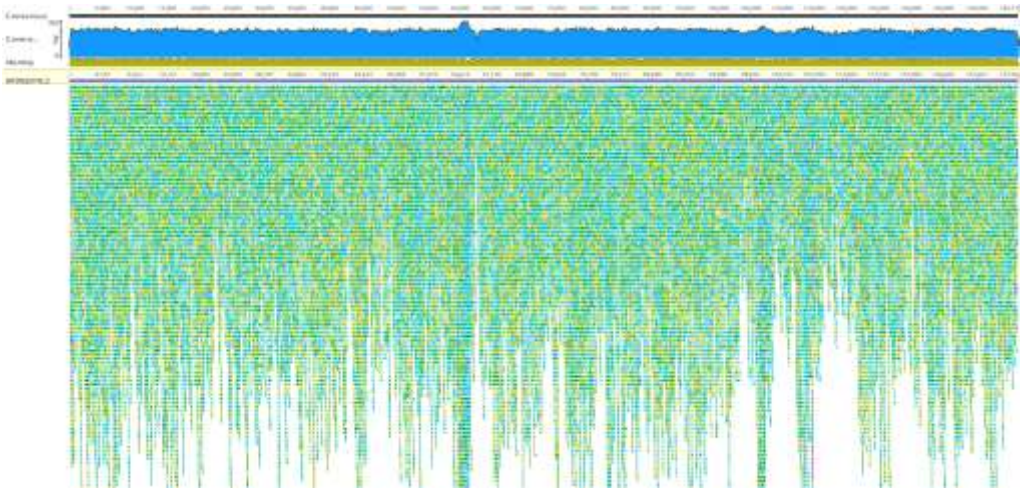
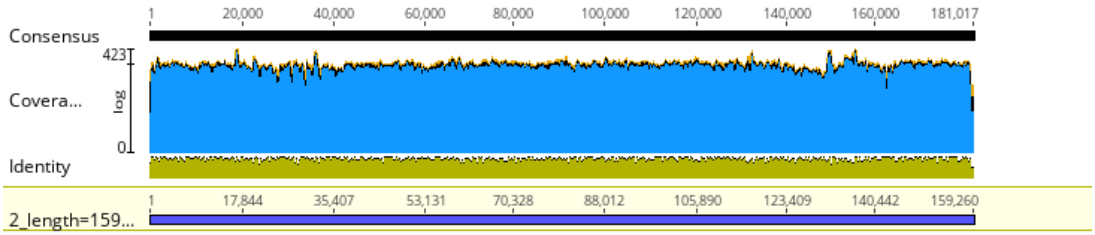


Figure S5: Annotated plasmid image and read mapping of NDM plasmid from GDW2218 *Klebsiella pneumoniae* and reference plasmid pKP1-NDM-1 A) and B) annotated hybrid assembly NDM plasmid and mash neighbor reference plasmid pKP1-NDM-1. C) and D) read mapping to NDM plasmid and reference plasmid.

A)



B)

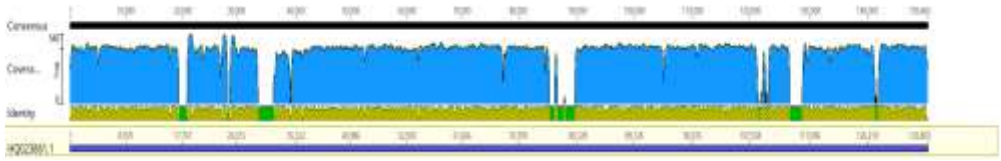
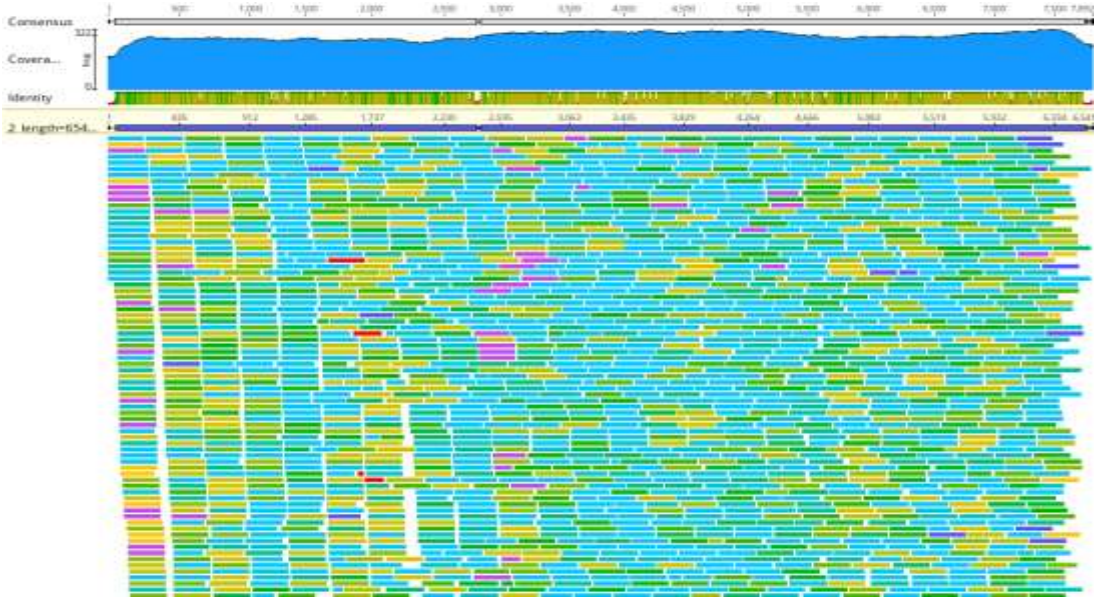


Figure S6: Read mapping of *Klebsiella pneumoniae* GW2280 Illumina reads to; A) hybrid assembly NDM plasmid from GDW2280, and B) mash neighbor plasmid pPG010208.

A)



B)

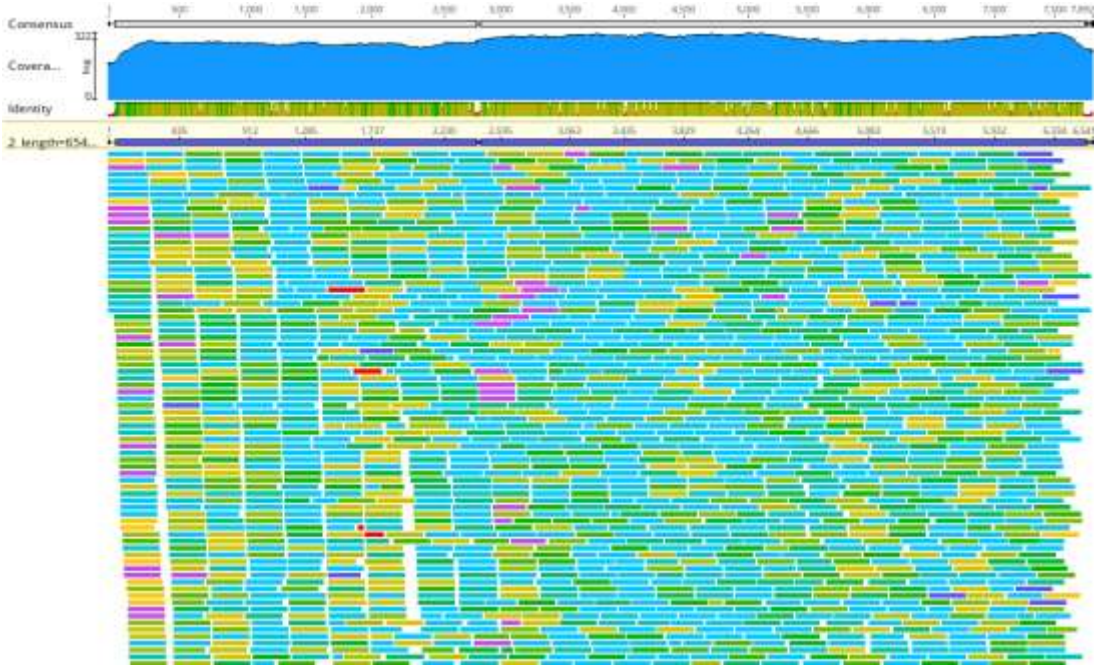
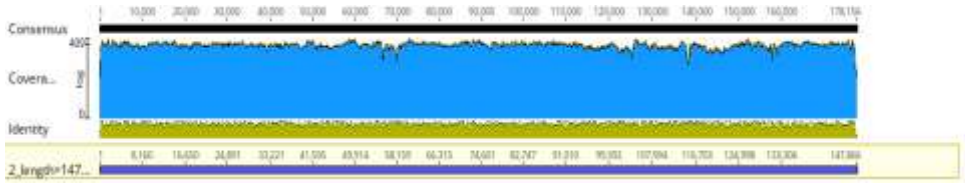


Figure S7: Read mapping of *Morganella morganii* GDW2281 Illumina reads to; A) hybrid assembly NDM plasmid from GDW2281 and. B) mash neighbor plasmid pkp_Goe_579.

A)



B)

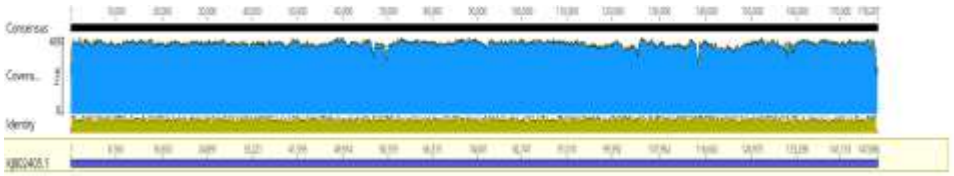
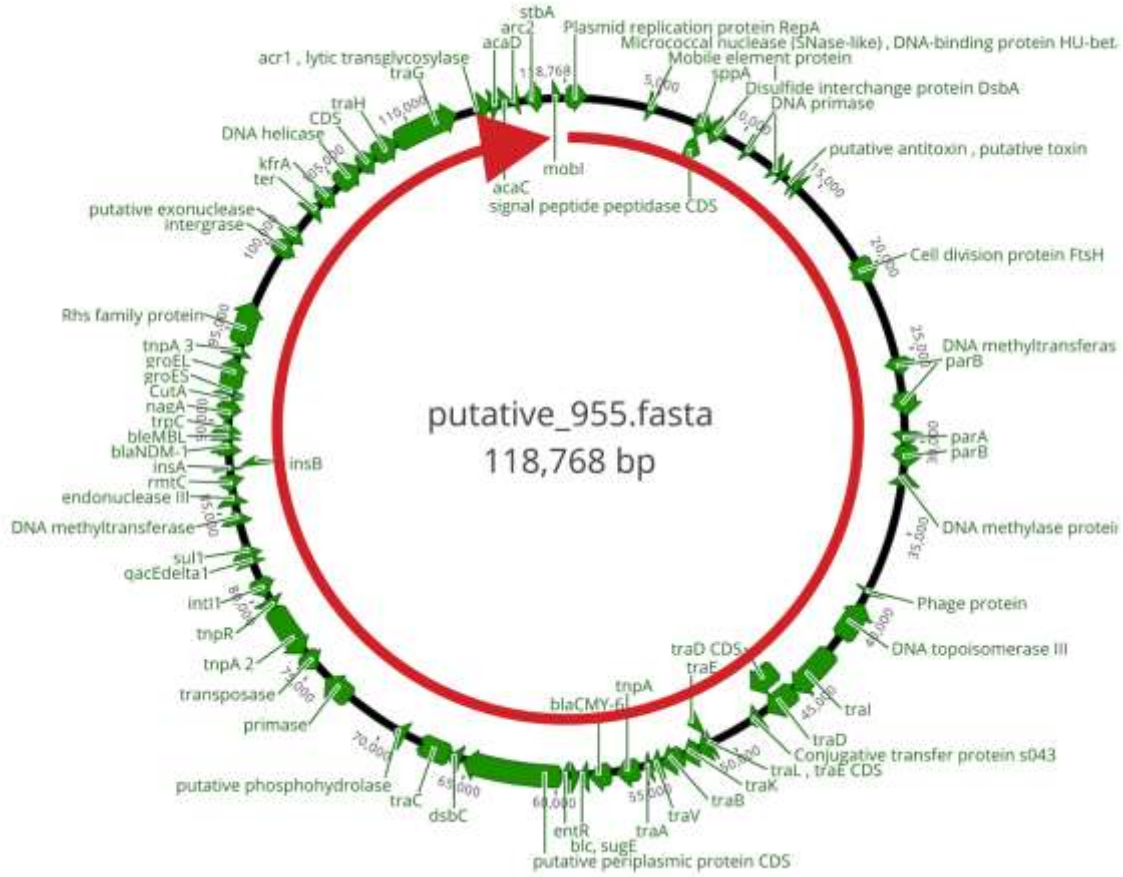
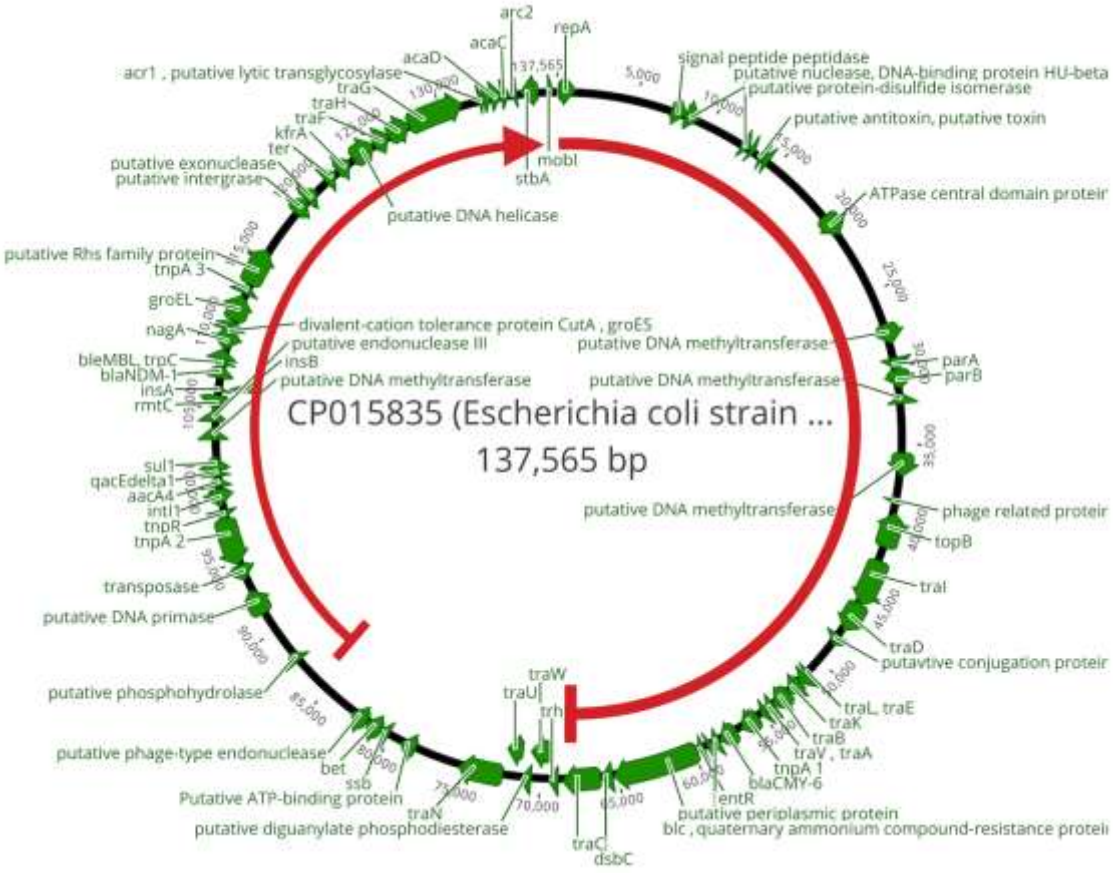


Figure S8: Read mapping of *Providencia stuartii* GDW2284 Illumina reads to; A) NDM plasmid from GDW2284, and. B) mash neighbor plasmid pNDM-PstGN576.

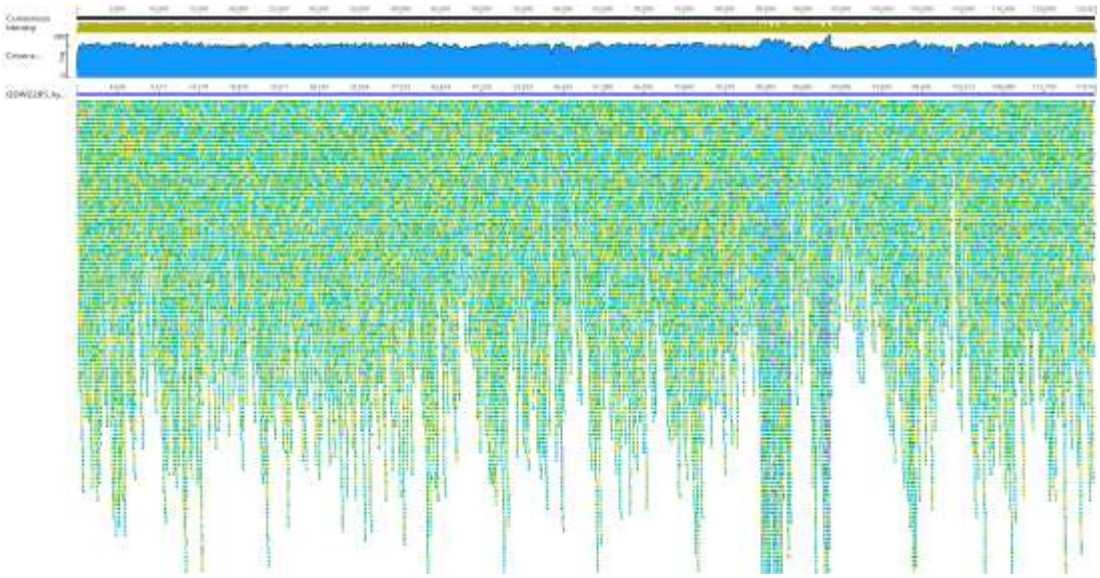
A)


putative_955.fasta
118,768 bp
stbA
arc2
acaD
Plasmid replication protein RepA
Micrococcal nuclease (SNase-like), DNA-binding protein HU-bet
Mobile element protein
sppA
Disulfide interchange protein DsbA
DNA primase
putative antitoxin, putative toxin
Cell division protein FtsH
DNA methyltransferas
parB
parA
parB
DNA methylase protei
Phage protein
DNA topoisomerase III
traD CDS
traE
traI
traD
Conjugative transfer protein s043
traL, traE CDS
traK
traB
traV
traA
blc, sugE
putative periplasmic protein CDS
entR
tnpA
blaCMY-6
dsbC
traC
putative phosphohydrolase
primase
transposase
tnpA 2
tnpR
int11
qacEdelta1
sul1
DNA methyltransferase
endonuclease III
rmtC
insA
blaNDM-1
bleMBL
trpC
nagA
CutA
groES
groEL
tnpA 3
Rhs family protein
putative exonuclease
intergrase
kfrA
ter
DNA helicase
CDS
traH
traG
acr1, lytic transglycosylase
acaC
mobI
signal peptide peptidase CDS
insB

B)


CP015835 (Escherichia coli strain ...
137,565 bp
arc2
repA
acaC
acaD
acr1 , putative lytic transglycosylase
traG
traH
traF
kfrA
ter
putative exonuclease
putative intergrase
putative DNA helicase
mobI
stbA
signal peptide peptidase
putative nuclease, DNA-binding protein HU-beta
putative protein-disulfide isomerase
putative antitoxin, putative toxin
ATPase central domain protein
putative Rhs family protein
tnpA 3
groEL
nagA
divalent-cation tolerance protein CutA , groES
putative endonuclease III
bleMBL, trpC
blaNDM-1
insA
rmtC
insB
putative DNA methyltransferase
putative DNA methyltransferase
putative DNA methyltransferase
putative DNA methyltransferase
parA
parB
sul1
qacEdelta1
aacA4
intI1
tnpR
tnpA 2
transposase
putative DNA primase
phage related protein
topB
traI
traD
putavtive conjugation protein
traW
traU
trh
traL, traE
traK
traB
traV , traA
tnpA 1
blaCMY-6
entR
putative periplasmic protein
blc , quaternary ammonium compound-resistance protei
putative phosphohydrolase
putative phage-type endonuclease
bet
ssb
Putative ATP-binding protein
traN
putative diguanylate phosphodiesterase
traC
dsbC

C)



D)

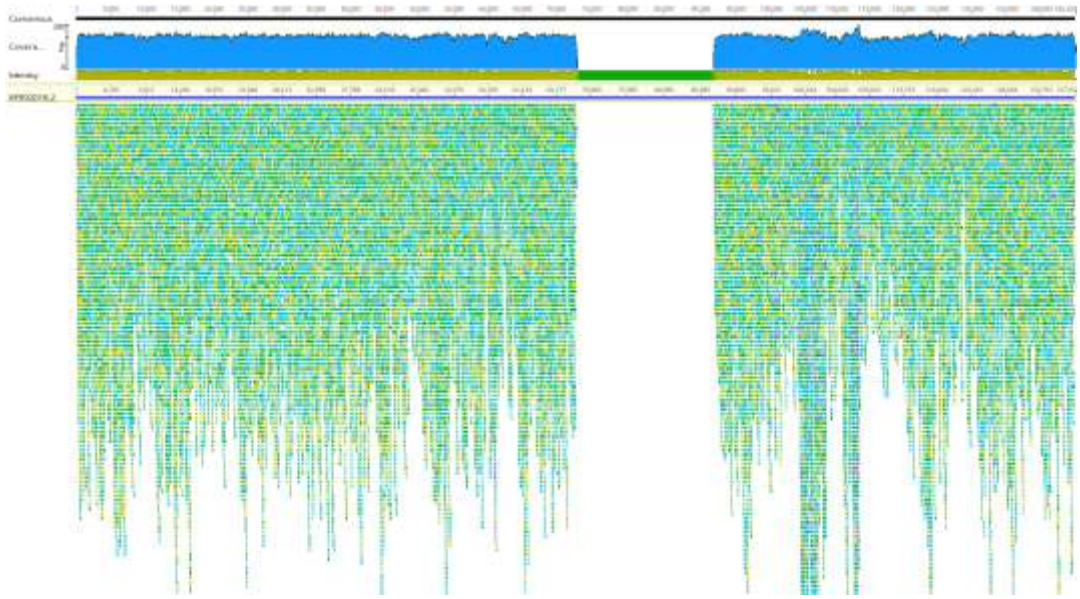


Figure S9: Annotated plasmid image and read mapping of NDM plasmid from GDW2285 *Escherichia coli* and reference plasmid pMS6198A A) and B) annotated hybrid assembly NDM plasmid and mash neighbor reference plasmid pMS6198A. C) and D) read mapping to NDM plasmid and reference plasmid.

A)

B

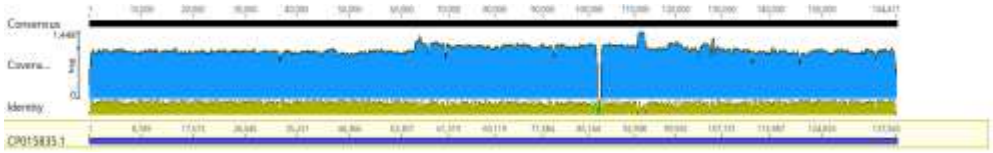
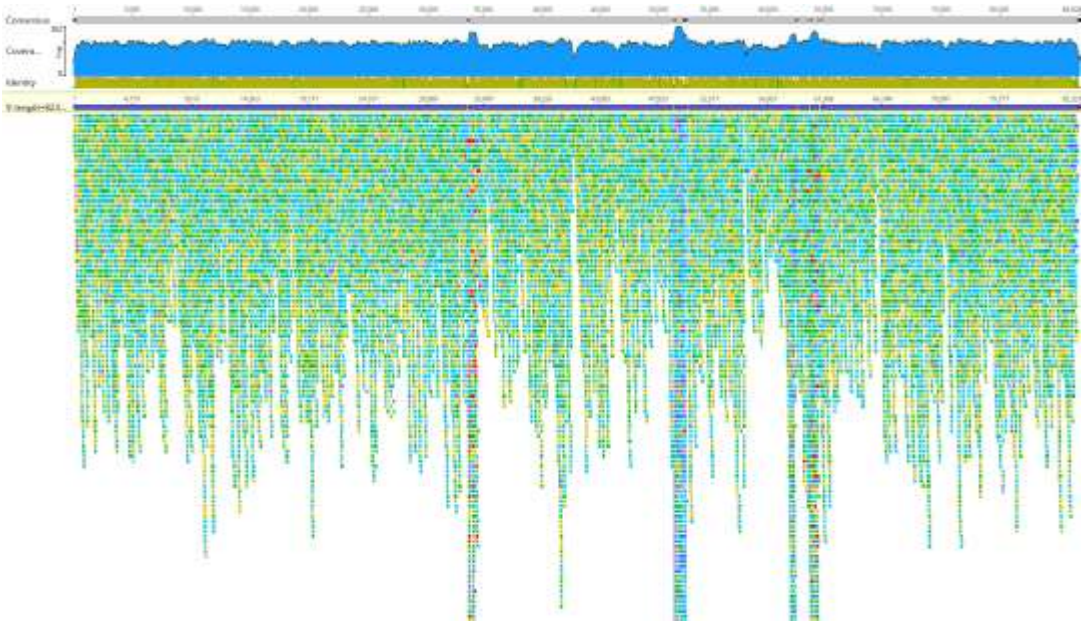


Figure S10: Read mapping of *Citrobacter freundii* GDW2286 Illumina reads to; A) hybrid assembly NDM plasmid from GDW2286, and B) mash neighbor plasmid pMS6198A.

A)



B)

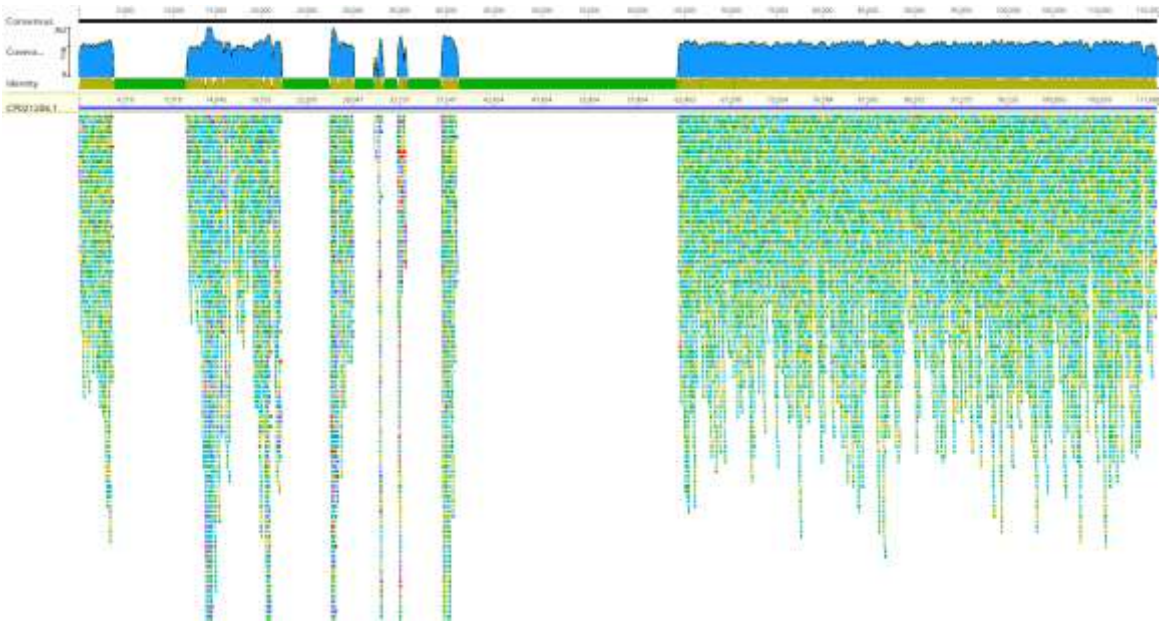
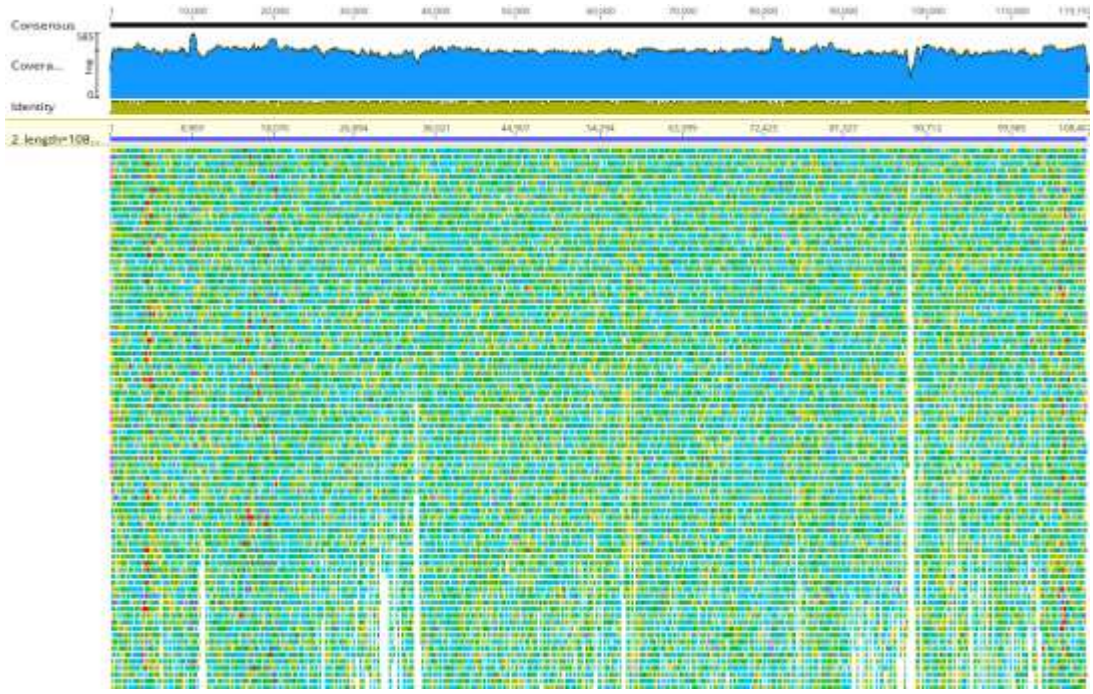


Figure S11: Read mapping of *Klebsiella pneumoniae* GDW2287 Illumina reads to; A) hybrid assembly NDM plasmid from GDW2287, and B) mash neighbor plasmid p1002-NDM-1.

A)



B)

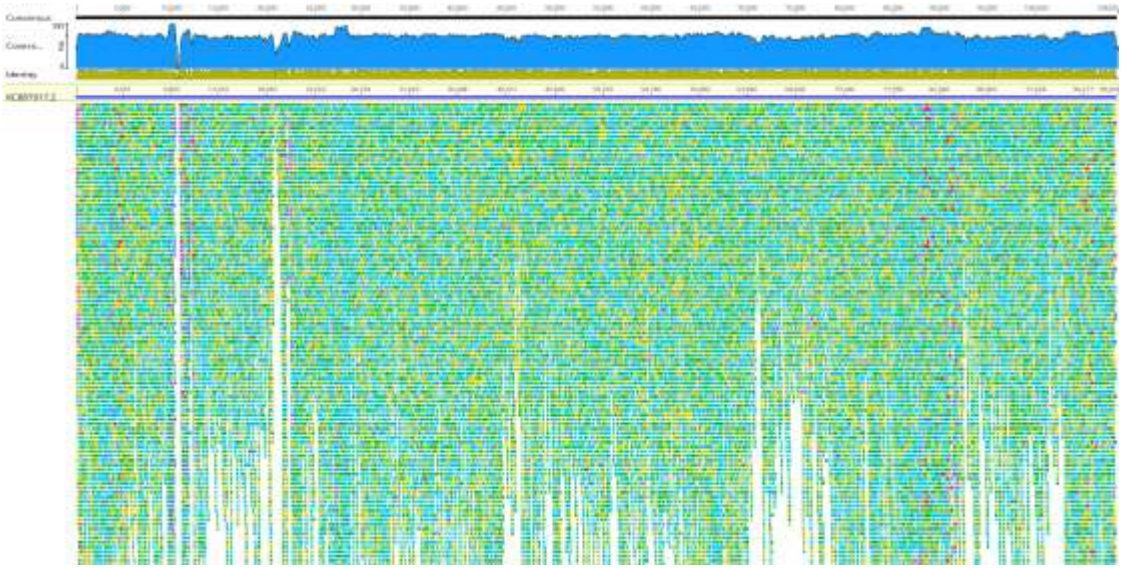
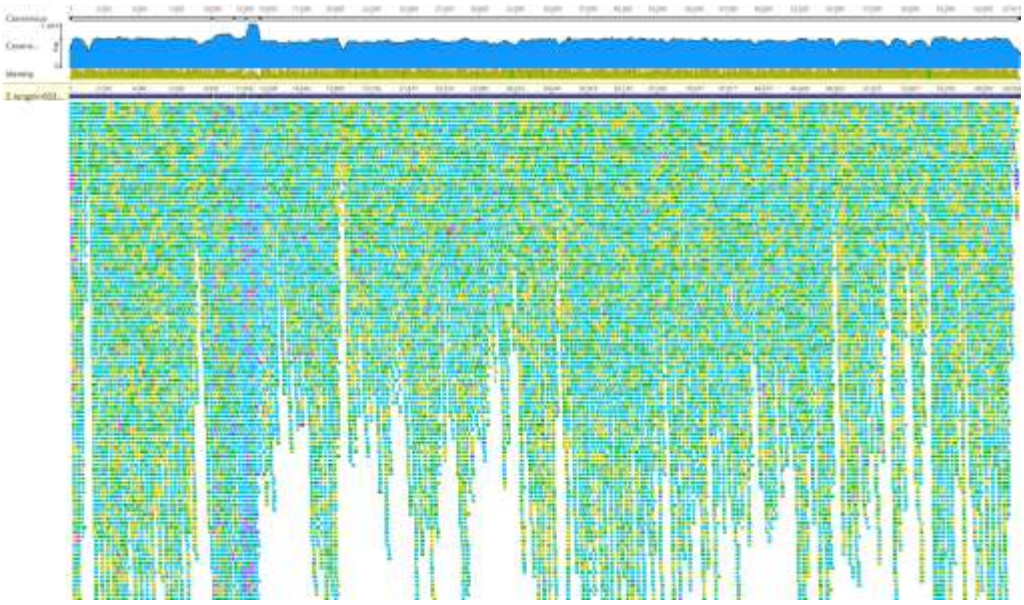


Figure S12: Read mapping of *Enterobacter cloacae* GDW2288 Illumina reads to; A) hybrid assembly NDM plasmid from GDW2288, and. B) mash neighbor plasmid pECL3-NDM1.

A)



B)

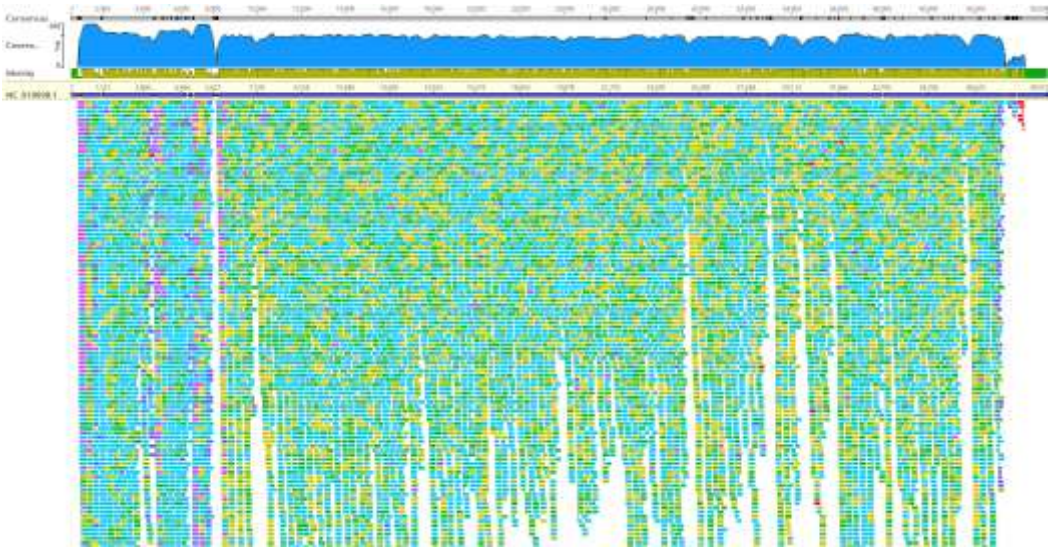
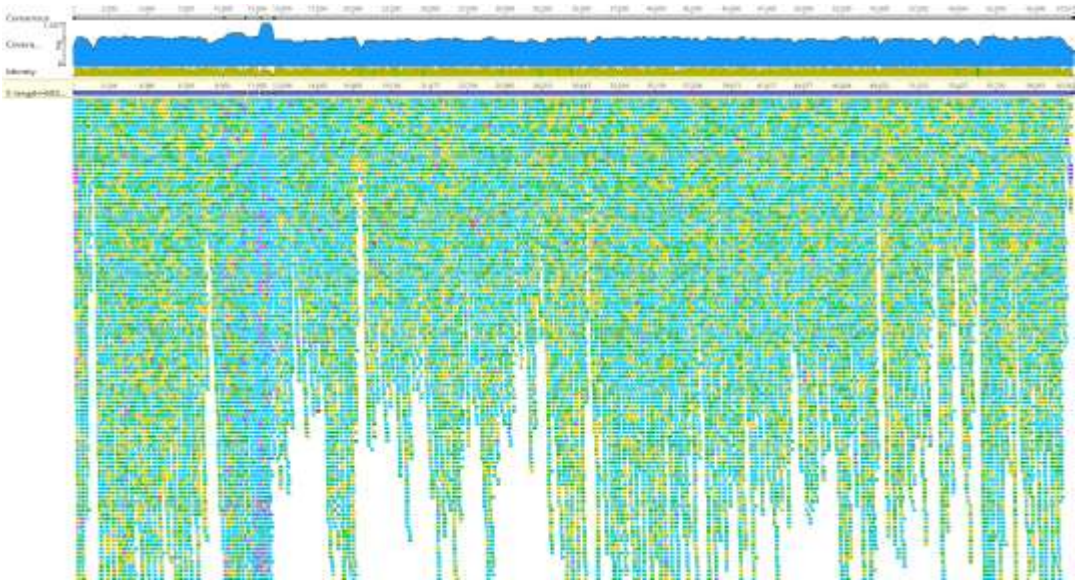


Figure S13: Read mapping of *Escherichia coli* GDW2289 Illumina reads to; A) hybrid assembly NDM plasmid from GDW2289, and B) mash neighbor plasmid pHHAA45(NC_019098).

A)



B)

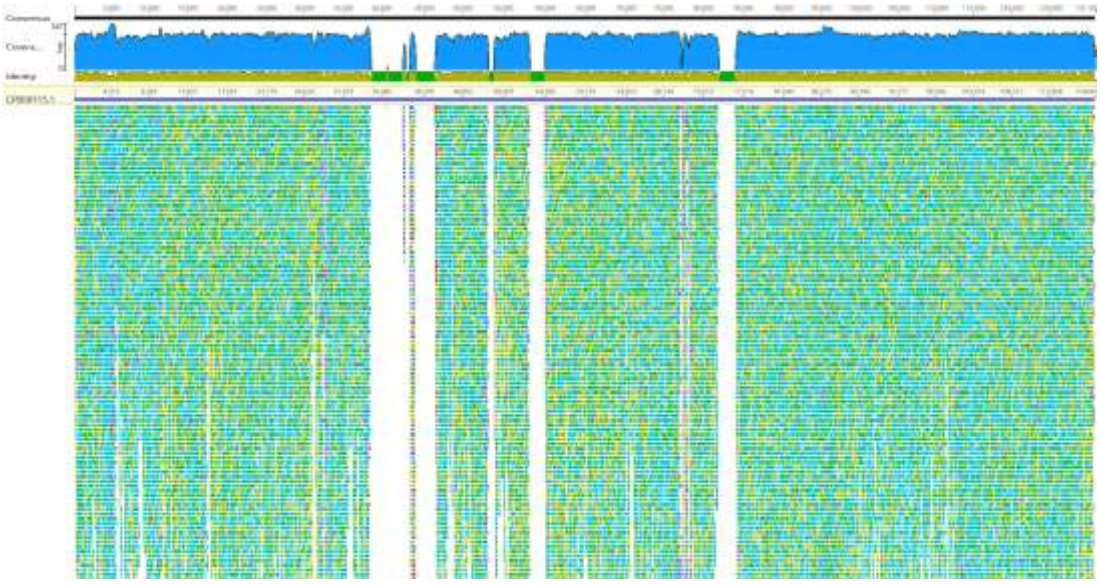
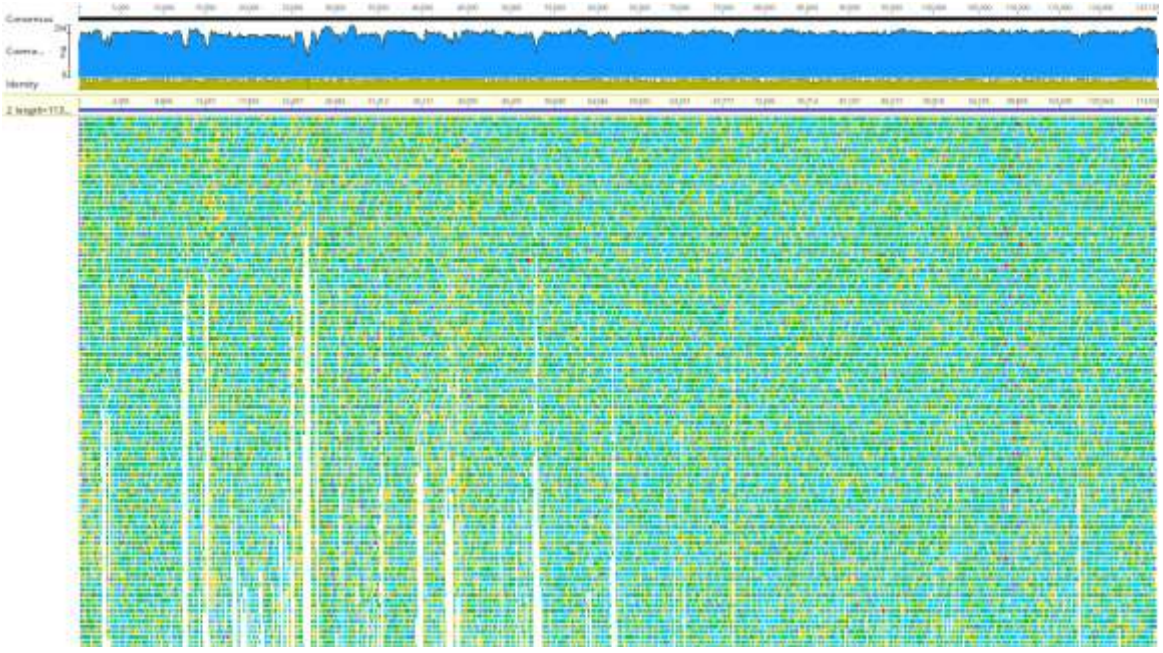


Figure S14: Read mapping of *Klebsiella oxytoca* GDW2290 Illumina reads to; A) hybrid assembly NDM plasmid from GDW2290, and B) mash neighbor plasmid blaNDM-1 plasmid P2.

A)



B)

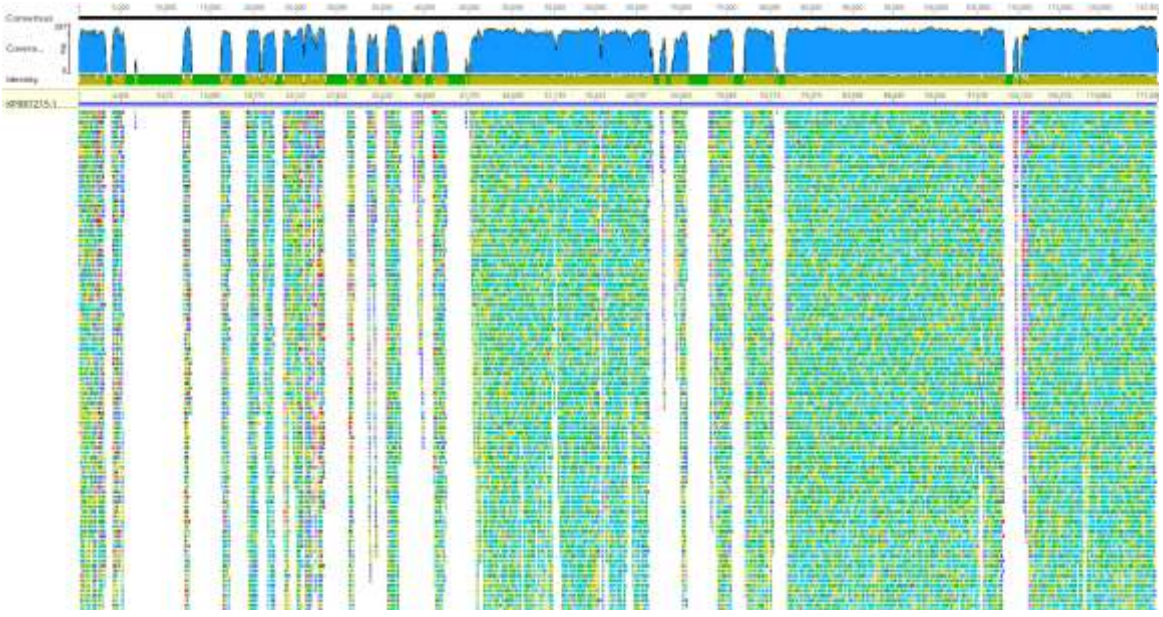


Figure S15: Read mapping of *Enterobacter cloacae* GDW2291 Illumina reads to; A) hybrid assembly NDM plasmid from GDW2291, and. B) mash neighbor plasmid p112298-KPC.

Table S1A: Antibiotic Suscetibility Test (AST) results for Isolate C0612.

Antibiotics	**C0612**
am ampicillin interp	R
am amoxicillin clavulanic acid interp	R
am amikacin interp	R
am cefazolin interp	R
am cefalotin interp	R
am ciprofloxacin interp	R
am cefixime interp	R
am ceftazidime interp	R
am gentamicin interp	R
am meropenem interp	R
am nitrofuratoin interp	R
am piperacillin tazobactam interp	R
am tetracycline interp	R
am tobramycin interp	R
am cefoxitin interp	R
am ceftriaxone interp	R
am ertapenem interp	R

Table S1B: Predicted AMR gene product from C0612 isolate.

	PRODUCT	RESISTANCE
chromosome	multidrug efflux RND transporter permease subunit OqxB17	PHENICOL;QUINOLONE
chromosome	multidrug efflux RND transporter periplasmic adaptor subunit OqxA10	PHENICOL;QUINOLONE
chromosome	multidrug efflux RND transporter permease subunit OqxB17	PHENICOL;QUINOLONE
chromosome	class A beta-lactamase SHV-212	BETA-LACTAM
chromosome	FosA family fosfomycin resistance glutathione transferase	FOSFOMYCIN
chromosome	FosA family fosfomycin resistance glutathione transferase	FOSFOMYCIN
chromosome	multidrug efflux RND transporter permease subunit OqxB17	PHENICOL;QUINOLONE
chromosome	multidrug efflux RND transporter periplasmic adaptor subunit OqxA10	PHENICOL;QUINOLONE
chromosome	FosA family fosfomycin resistance glutathione transferase	FOSFOMYCIN
chromosome	multidrug efflux RND transporter periplasmic adaptor subunit OqxA10	PHENICOL;QUINOLONE
plasmid_682	carbapenem-hydrolyzing class D beta-lactamase OXA-48	CARBAPENEM

plasmid_682	class A extended-spectrum beta-lactamase CTX-M-9	CEPHALOSPORIN
plasmid_682	aminoglycoside O-phosphotransferase APH(3")-Ib	STREPTOMYCIN
plasmid_682	aminoglycoside O-phosphotransferase APH(3')-VIb	AMIKACIN;KANAMYCIN
plasmid_682	aminoglycoside O-phosphotransferase APH(3")-Ib	STREPTOMYCIN
plasmid_682	aminoglycoside O-phosphotransferase APH(6)-Id	STREPTOMYCIN
plasmid_726	Mph(A) family macrolide 2'-phosphotransferase	MACROLIDE
plasmid_726	bleomycin binding protein Ble-MBL	BLEOMYCIN
plasmid_726	subclass B1 metallo-beta-lactamase NDM-1	CARBAPENEM
plasmid_726	APH(3')-VI family aminoglycoside O-phosphotransferase	AMIKACIN;KANAMYCIN
plasmid_726	quinolone resistance pentapeptide repeat protein QnrS1	QUINOLONE
plasmid_726	oxacillin-hydrolyzing class D beta-lactamase OXA-9	BETA-LACTAM
plasmid_726	AAC(6')-Ib family aminoglycoside 6'-N-acetyltransferase	AMIKACIN;KANAMYCIN;TOBRAMYCIN
plasmid_726	ANT(3")-Ia family aminoglycoside nucleotidyltransferase AadA1	STREPTOMYCIN

plasmid_726	class A beta-lactamase TEM-1	BETA-LACTAM
plasmid_726	trimethoprim-resistant dihydrofolate reductase DfrA14	TRIMETHOPRIM
plasmid_726	Mph(A) family macrolide 2'-phosphotransferase	MACROLIDE
plasmid_726	tetracycline efflux MFS transporter, Tet(D)	TETRACYCLINE
plasmid_726.	class A extended-spectrum beta-lactamase CTX-M-15	CEPHALOSPORIN
plasmid_726	aminoglycoside N-acetyltransferase AAC(3)-IIe	GENTAMICIN

Table 2A: Antibiotic Suscetibility Test (AST) results for Isolate C0648.

Antibiotics	**C0648**
am ampicillin interp	R
am amoxicillin clavulanic acid interp	R
am amikacin interp	R
am cefazolin interp	R
am cefalotin interp	R
am ciprofloxacin interp	R
am cefixime interp	R
am ceftazidime interp	R
am gentamicin interp	R
am metropenem interp	R
am nitrofuratoin interp	R
am piperacillin tazobactam interp	R
am tetracycline interp	R
am tobramycin interp	R
am trimethoprim sulfamethoxazole interp	R
am cefoxitin interp	R
am ceftriaxone interp	R
am ertapenem interp	R

Table S2B: Predicted AMR gene product from C0648 isolate.

	PRODUCT	RESISTANCE
Chromosome	class C extended-spectrum beta-lactamase EC-15	CEPHALOSPORIN
Chromosome	Escherichia coli EF-Tu mutants	CEPHALOSPORIN
Chromosome	Escherichia coli parC	QUINOLONE
Chromosome	Escherichia coli gyrA	QUINOLONE
Chromosome	Haemophilus influenzae PBP3	BETA-LACTAM
Chromosome	Escherichia coli ampH beta-lactamase	BETA-LACTAM
Chromosome	Escherichia coli ampC beta-lactamase	BETA-LACTAM
Chromosome	Escherichia coli EF-Tu mutants	PULVOMYCIN
Chromosome	Escherichia coli marR mutant	MULTIPLE DRUG CLASSES
Chromosome	Escherichia coli acrR	MULTIPLE DRUG CLASSES
Chromosome	Escherichia coli soxR	MULTIPLE DRUG CLASSES
Chromosome	Escherichia coli soxS	MULTIPLE DRUG CLASSES
plasmid_1550	tetracycline efflux MFS transporter Tet(A)	TETRACYCLINE
plasmid_1550	trimethoprim-resistant dihydrofolate reductase DfrA12	TRIMETHOPRIM
plasmid_1550	ANT(3")-Ia family aminoglycoside nucleotidyltransferase AadA2	STREPTOMYCIN
plasmid_1550	sulfonamide-resistant dihydropteroate synthase Sul1	SULFONAMIDE
plasmid_1550	bleomycin binding protein Ble-MBL	BLEOMYCIN
plasmid_1550	subclass B1 metallo-beta-lactamase NDM-5	CARBAPENEM
plasmid_1550	16S rRNA (guanine(1405)-N(7))-methyltransferase RmtB1	AMINOGLYCOSIDE
plasmid_1550	class A broad-spectrum beta-lactamase TEM-1	BETA-LACTAM
plasmid_1550	Mph(A) family macrolide 2'-phosphotransferase	MACROLIDE
plasmid_1550	sulfonamide-resistant dihydropteroate synthase Sul1	SULFONAMIDE
plasmid_1550	ANT(3")-Ia family aminoglycoside nucleotidyltransferase AadA5	STREPTOMYCIN
plasmid_1550	trimethoprim-resistant dihydrofolate reductase DfrA17	TRIMETHOPRIM
plasmid_486	class C extended-spectrum beta-lactamase CMY-42	CEPHALOSPORIN

Table S3A: Antibiotic Suscetibility Test (AST) results for Isolate C0650.

Antibiotics	C0650
am ampicillin interp	R
am amoxicillin clavulanic acid interp	R
am cefazolin interp	R
am cefalotin interp	R
am cefixime interp	R
am ceftazidime interp	R
am metropenem interp	R
am piperacillin tazobactam interp	R
am tobramycin interp	R
am cefoxitin interp	R
am ceftriaxone interp	R
am ertapenem interp	R

Table S3B: Predicted AMR gene product from C0650 isolate.

	PRODUCT	RESISTANCE
chromosome	multidrug efflux RND transporter permease subunit OqxB26	PHENICOL;QUINOLONE
chromosome	multidrug efflux RND transporter periplasmic adaptor subunit OqxA6	PHENICOL;QUINOLONE
chromosome	class A extended-spectrum beta-lactamase SHV-27	CEPHALOSPORIN
chromosome	FosA family fosfomycin resistance glutathione transferase	FOSFOMYCIN
plasmid_1220	bleomycin binding protein Ble-MBL	BLEOMYCIN
plasmid_1220	subclass B1 metallo-beta-lactamase NDM-1	CARBAPENEM

plasmid_1220	APH(3')-VI family aminoglycoside O-phosphotransferase	AMIKACIN;KANAMYCIN
plasmid_1220	quinolone resistance pentapeptide repeat protein QnrS1	QUINOLONE
plasmid_1220	class A extended-spectrum beta-lactamase CTX-M-15	CEPHALOSPORIN
plasmid_1220	AAC(6')-Ib family aminoglycoside 6'-N-acetyltransferase	AMIKACIN;KANAMYCIN;TOBRAMYCIN
plasmid_1220	ANT(3'')-Ia family aminoglycoside nucleotidyltransferase AadA1	STREPTOMYCIN
plasmid_1220	oxacillin-hydrolyzing class D beta-lactamase OXA-9	BETA-LACTAM
plasmid_2088	OXA-48 family carbapenem-hydrolyzing class D beta-lactamase OXA-232	CARBAPENEM
plasmid_770	quinolone resistance pentapeptide repeat protein QnrB4	QUINOLONE
plasmid_770	class C beta-lactamase DHA-1	CEPHALOSPORIN
plasmid_770	sulfonamide-resistant dihydropteroate synthase Sul1	SULFONAMIDE