ABSTRACT

Title of Dissertation: CHARACTERIZING THE FULL COMPLEMENT OF ANTIMICROBIAL RESISTANCE GENES AND LINKING THE RESISTANCE GENES AND THE PLASMID TO SOURCE BACTERIA

Saul Herney Sarria

Doctor of Philosophy 2023

Dissertation directed by: Dr. Jiuzhou Song
Biological Sciences Program (BISI)
Dr. Daniel A. Tadesse
Food and Drug Administration (FDA)

Animals' gut is inhabited and shaped by diverse and complex microbial communities yet to be characterized and to fully understand their crucial role in antimicrobial resistance's widespread. The development of antimicrobial resistance (AMR) to medically important classes of drugs and the resulting loss of effectiveness remains a major public health concern in the U.S. Surveillance of AMR is currently based on the isolation of indicator microorganisms followed by phenotypic and genotypic characterization of these organisms. However, these approaches rely on culture-based detection and characterization of pathogens and may ignore potential antimicrobial resistance reservoirs present in

organisms or that are not routinely examined and greatly impede our understanding of the role played by resistance genes present in these organisms. Characterization of the total microbiota including those that are not culturable in artificial medium and anaerobic organisms will help understand the genes that make up the environmental resistome and the role played by non-culturable bacteria in antimicrobial resistance dissemination. On the other hand, antimicrobials play an important role in the fight against infectious diseases in both humans and animals. Nevertheless, selective pressure exerted by antimicrobial use has been the primary driving force leading to the emergence, prevalence, and spread of resistant traits among pathogenic and commensal bacterial species where mobile genetic elements (MGE) have a critical role in the transmission of antimicrobial resistance genes within the microbial community.

The recent advances in next-generation sequencing technologies have allowed culture-independent approaches to analyze community DNA extracted directly from the sample rather than from individually cultured microbes. Metagenomics offers insights into the resistomes' potential of the microbial communities. However, one of the challenges of the metagenomics approach is the difficulty in attributing the identified antimicrobial resistance genes (AMR genes) to specific bacterial strains present in the sample. When traditional metagenomics is used to detect AMR genes, it is difficult to predict if they belong to a specific pathogen or to the background microbiota. Therefore, the utilization of non-culture-based techniques will help to better understand the resistome composition, and the potential of that resistome to be transferred to pathogens and to help explain more accurately the changes and the genetic makeup of bacterial populations in any given

sample. Meanwhile, employing proximity ligation (Hi-C) technology to map antimicrobial resistance genes and mobile genetic elements to source bacteria helps to explain the reservoirs of AMR genes accessible to pathogens and their potential for widespread.

We employed deep shotgun metagenomic sequencing and proximity ligation (Hi-C) library sequencing to characterize the resistome, assemble genomes from metagenomic samples, and accurately attribute antibiotic resistance genes and plasmids to host bacteria cells. We randomly selected 21 cecal samples from food animal sources (cattle n= 6, swine n= 6, chicken n= 3, and turkey n= 6). We generated more than 75 million reads/sample for Hi-C and more than 100 million reads/sample for shotgun metagenomic sequence reads.

Bioinformatics analysis revealed over 200 bins containing metagenome-assembled genomes (MAGs) with different levels of completeness, novelty scores, and contamination based on CheckM. A total of 245 previously uncharacterized genomes were reconstructed with a high level of confidence (>90% Completeness, >90% Novelty, < 5% contamination). Of the 245 newly reconstructed MAGs, 24 were at bacteria taxonomic rank level, 5 at phyla (Actinobacteria; 11 genomes. Firmicutes; 3 genomes. Bacteroidetes; 17 genomes. Euryarchaeota; 2 genome), 4 at class (Bacilli; 1 genome. Clostridia; 9 genomes. Deltaproteobacteria; 3 genomes. Gammaproteobacteria; 1 genome), 5 at order level (Actinomycetales; 2 genomes. Bacteroidales; 25 genomes. Clostridiales; 114 genomes. Lactobacillales; 2 genomes. Selenomonadales; 2 genomes), and 3 at family level (Lachnospiraceae, 24 genomes. Spirochaetaceae; 1 genome. Spirochaetaceae; 2 genomes).

We identified over 400 antimicrobial resistance genes representing 22 antimicrobial classes including: aminoglycoside (40 gene variants), beta-lactams (37 gene variants), bleomycin (2

gene variants), colistin (3 gene variants), fosfomycin (4 gene variants), glycopeptide (6 gene variants), lincosamide (9 gene variants), lincosamide/streptogramin (2 gene variants), macrolide (16 gene variants), macrolide/lincosamide/streptogramin (4 gene variants), nitroimidazole (1 gene variant), phenicol (9 gene variants), phenicol/oxazolidinone (1 gene variant), phenicol/quinolone (2 gene variants), pleuromutilin (1 gene variant), quinolone (5 gene variants), streptogramin (1 gene variant), streptothricin (3 gene variants), sulfonamide (3 gene variants), tetracycline (25 gene variants), and trimethoprim (8 gene variants).

Plasmid characterization using Hi-C proximity ligation and shotgun metagenomics allowed the identification of 146 plasmids (>= 85% completeness, >= 90% reference sequence similarity), and over 13000 plasmid-contigs (<85% completeness, < 90% reference sequence similarity).

Shotgun metagenomics provides valuable insights into the diversity and identity of the resistome present in a microbiome, while Hi-C generates millions of paired-end reads linking DNA fragments in close proximity. When shotgun metagenomics is coupled with the Hi-C proximity ligation approach it shows a great capability in genome binning and simultaneous retrieval of high-quality MAGs from a single sample, thusly enabling the link of resistance genes and plasmids to host bacterial cells and facilitating the public health management decisions aimed at reducing the source and exposure routes of AMR to humans.

CHARACTERIZING THE FULL COMPLEMENT OF ANTIMICROBIAL RESISTANCE GENES AND LINKING THE RESISTANCE GENES AND THE PLASMID TO SOURCE BACTERIA

by

Saul Herney Sarria

Dissertation submitted to the Faculty of the Graduate School of the University of Maryland, College Park, in partial fulfillment of the requirements for the degree of Doctor of Philosophy
2023

Advisory Committee:
Dr. Jiuzhou Song, Chair/Advisor
Dr. Daniel A. Tadesse, Co-Advisor
Dr. Najib El-Sayed, Dean's Representative
Dr. Li Ma
Dr. Louisa Wu

Preface

The results in this dissertation have either been published in peer-reviewed journals, have been submitted and are under review, or are currently under preparation for submission. I am thankful for the help of all co-authors whose dedication and knowledge resulted in greatly improved works.

Other publications to which I have contributed during my graduate training.

Liu, J.; Bai, Y.; Liu, F.; Kohn, R.A.; Tadesse, D.A.; Sarria, S.; Li, R.W.; Song, J. Rumen Microbial Predictors for Short-Chain Fatty Acid Levels and the Grass-Fed Regimen in Angus Cattle. *Animals* 2022, *12*, 2995. https://doi.org/10.3390/ani12212995.

Javkar K, Rand H, Hoffmann M, Luo Y, Sarria S, Thirunavukkarasu N, Pillai CA, McGann P, Johnson JK, Strain E, Pop M. Whole-Genome Assessment of Clinical Acinetobacter baumannii Isolates Uncovers Potentially Novel Factors Influencing Carbapenem Resistance. Front Microbiol. 2021 Oct 1;12:714284. doi: 10.3389/fmicb.2021.714284. PMID: 34659144; PMCID: PMC8518998.

Sakowski E, Uritskiy G, Cooper R, Gomes M, McLaren MR, Meisel JS, Mickol RL, Mintz CD, Mongodin EF, Pop M, Rahman MA, Sanchez A, Timp W, Vela JD, Wolz CM, Zackular JP, Chopyk J, Commichaux S, Davis M, Dluzen D, Ganesan SM, Haruna M, Nasko D, Regan MJ, **Sarria S**,

Shah N, Stacy B, Taylor D, DiRuggiero J, Preheim SP. Current State of and Future Opportunities for Prediction in Microbiome Research: Report from the Mid-Atlantic Microbiome Meet-up in Baltimore on 9 January 2019. mSystems. 2019 Oct 8;4(5):e00392-19. doi: 10.1128/mSystems.00392-19. PMID: 31594828; PMCID: PMC6787564.

Tadesse DA, Hoffmann M, **Sarria S**, Lam C, Brown E, Allard M, McDermott PF. Complete Genome Sequences of 14 Salmonella enterica Serovar Enteritidis Strains Recovered from Human Clinical Cases between 1949 and 1995 in the United States. Genome Announc. 2018 Jan 4;6(1):e01406-17. doi: 10.1128/genomeA.01406-17. PMID: 29301885; PMCID: PMC5754494.

Dedication

To all my family, friends, mentors, teachers, facilitators, and collaborators for all their support and wisdom. Special dedication to my mother Ana Ruiz+, to my father Hernan Sarria +, to my beloved brothers Evert Lot + and Esaud Rodrigo+, and to Cookies, Alburris and Pirris.

Acknowledgments

I have had the chance to meet and be supported by so many brilliant people leading up to and during the graduate research training. I cannot remember all their names or all their contributions, but I am truly grateful and hope I can pay forward a fraction of what I have received.

First, I sincerely appreciate my advisors Jiuzhou Song and Daniel Tadesse, for being amazing and patient mentors, for training me, allowing me to pursue my research interests, and helping me transform those interests into published works. The many coffee meetings and gatherings I have had throughout the development of my professional career were all important and essential to my continued growth not only as a researcher but as a person.

I also want to thank my committee members. Najib El-Sayed, Louisa Wu, and Li Ma for being a source of creative energy, and scientific curiosity, for helping me improve my technical skills and to progress in scientific rigor and professionalism. I also appreciate all the constructive feedback and the insightful questions that helped me grow on my graduate research path.

I have also had many amazing schoolmates friends, and coworkers whom I could always rely upon for brainstorming, collaborating, networking or simply to enjoy a good meal, have fun and a good laugh: Seth Commichaux, Mary Chey, Ashley Della Fera, Zabdiel Alvarado, Jamie Kelly, Oladipupo Ridwan Bello, and Kyle Brumfield, Domenick Braccia, Kiran Javkar, Jacquelyn Stephanie Meisel, Dan Nasko, Jeremy Selengut, Nidhi Shah, Brook Stacy, and Todd Treangen. I also want to thank all my coworkers at the FDA, Center for Veterinary Medicine, and Office of Applied Science for all their invaluable support. Special thanks to John Graham, Maureen Davidson, Regina Tan, Chris Whitehouse, Stuart Gaines, Claudia Lam, Gordon Martin, Mike

Mikailov, Jason Abbott, Sonya Bodeis-Jones, Paul Heidenreich, Alberto Chiesa, Sherry Ayers, Shenia Young, Chih-Hao Hsu, Thu- Thuy Tran, Cong Li, Sampa Mukherjee, Gregory Tyson, Beilei Ge, Shaohua Zhao, Sanchez S. Fleurant, Kelly Domesle, Crystal Rice-Trujillo, and Lucas Harrison.

A very special thanks to Caitlin A. Welsh (CAW) for all the support, time, and wonderful contribution to my professional and personal growth - Your thoughtfulness is a very special gift I will always treasure -

It also would have been impossible to succeed without the CBBG, BISI, and ANSC communities, including fellow students, faculty, and staff. Special thanks to Dr. Steve Mount, for providing mentorship and support during my graduate research training, especially when I was still trying to find my research direction.

Lastly, I would like to thank all my family and friends for all their love and support. And a very special thanks to my wife Maria and my daughters Alba and Nicole. They have been my rock through all the storms, and I would not have made it to the finish line without their unconditional support and love.

Table of Contents

List of Tables

List of Figures

PART I

1. INTRODUCTION

Culture-based methodologies are used for the assessment of antibiotic resistance in samples collected in different environments. One traditional approach involves culturing microorganisms on specific artificial media and then extracting the DNA from isolated colonies for sequencing and analysis, but over 90% of microorganisms in the environment cannot be cultured using artificial media [1]. This has the significant disadvantage of requiring a relatively large number of viable organisms for analysis and is time-consuming, and the output only characterizes an isolated organism [2]. It has been considered a standard key to understanding the phenotypic characteristics of isolates and their resistance properties. Nevertheless, it overlooks the biological importance of bacteria that cannot grown in a laboratory setting due to the lack of knowledge of their unique media requirements and growth conditions, including fastidious and anaerobic microorganisms, or the inability to recreate the environment required for growth for extremophiles. Metagenomics is a field of molecular biology and genomics that involves the study of genetic material directly obtained from environmental samples without artificial media isolation. It goes beyond traditional genomics, which focuses on the study of individual microorganisms, by allowing the analysis of the collective genetic material of an entire community of organisms inhabiting a specific environment as a whole.

The term "metagenomics" was first published in 1998 in a study of soil microbes using random cloning of environmental DNA is a combination of two words: "meta" (meaning "beyond" or "transcending") and "genomics" (the study of genomes) [3]. Metagenomics aims to understand

the structure, function, and dynamics of microbial communities in different environments, including oceans, soil, the human gut, the animal gut, and other ecosystems. It's important to remark that the field of metagenomics is constantly evolving, and new developments can be expected with the decreasing cost of sequencing technology, access to unique sampling environments, computational and bioinformatic tools advancement, and improved annotation databases for characterization.

The basic process of metagenomics involves the following steps; however, it is complex and prone to bias at all steps: A. Sample collection and preservation: Environmental samples, such as soil, water, wastewater, or feces, are collected from the target environment and preserved at preferable at -80C for long term or until extraction. Collection methods should be appropriate to the environmental sample type, as well as to minimize the potential for biodegradation between sample processing and analysis; keeping the sample as cool as possible without freezing is recommended. Collection methods demonstrated bias in microbiome composition [4].Preservation methods can be customized using pH control, chemical addition, amber or opaque bottles, filtration, refrigeration, and freezing. B. DNA Extraction: Involves isolating DNA from a sample by various methods of cell lysis, removal of cellular debris, removal of residual chemical impurities, and preservation. Thus, it is considered a critical factor for the success of DNA sequencing and unbiased bacteria composition [4-6]. DNA is extracted from collected samples, preserving the genetic material from all organisms present in the environment. Enzymatic cell lysis based on lysozyme lysis and mechanical cell lysis based on bead beating are the most commonly use methods, oftentimes in combination, but could potentially introduce bias into the results if some bacterial species are lysed less

efficiently than others [6-8]. Spin filter columns with membranes that enable DNA binding and washing are commonly used as they can boost efficiency and purity. However, there is no single DNA extraction method that will efficiency work for all environmental sample types. Therefore, it is very critical to have an appropriate extraction method based on the metagenomic sample type [7, 8]. C. Library preparation and Sequencing: Library preparation plays a critical role for sequencing, and factors such as the quality and quantity of the DNA source material [9] will go along with the sequencing approach chosen as every sequencing method will have its own library preparation specifications. Basic steps of short read metagenomics library preparation using the Illumina platform include 1) Input DNA quantification by fluorometric-based method, such as Qubit assays, 2) Assessing the DNA purity by Tape Station or bioanalyzer, 3) Tagmentation of genomic DNA using Nextera transposome for genomic DNA fragment and tags DNA with adapter sequences, 4) Library amplification which amplifies the tagmented DNA using a limited-cycle PCR program, the PCR step adds the Index 1 (i7) adapters, Index 2 (i5) adapters, and sequences required for sequencing cluster generation, 5) Library clean up uses single-sided bead purification to purify amplified libraries, 6) Library normalization and pooling. This last step of the library preparation process normalizes the quantity of each metagenomics library to ensure equal representation and diversity in the pooled library for high sequencing performance. Finally, a metagenomic library pool is sequenced using high-throughput techniques to obtain the vast majority of genetic data. It can be carried by short read sequencing technology like Illumina or long read sequencing technology like PabBio. Illumina high throughput sequencing-by-synthesis technology provides over 99% accuracy (https://www.illumina.com/science/technology/next-generation-sequencing/sequencing-

technology.html), uses a proprietary platform to amplify fragments of the genome being sequenced and then reads which base is added as the fragment is synthesized using fluorescently tagged bases. PacBio, on the other hand, has made adjustments to the sequencing-by-synthesis method using this HiFi sequencing protocol to produce large, circular DNA molecules which can then be sequenced continuously and also has an accuracy rate of over 99% (https://www.pacb.com/technology/hifi-sequencing/how-it-works/).

D. Metagenomic sequencing data analysis: Once obtained sequencing data raw output, there is a need to convert these raw data image files into a fastQ files, remove low complexity, low quality read, and removed host DNA to obtain a clean metagenomic data output. The metagenomic data output produced by these sequencing platforms has enormous potential to provide its biological value to better understand compositional structures, functional connections within the microbiome, and the role of bacteria ranks and resistome composition in the biological sample, as the Next Generation Sequencing advance rapidly as well as curated database for better characterization. Extracting this valuable information requires significant downstream processing and analysis. Once the sequencing is done, the data generated includes multiple metagenomics samples with millions of sequences reads each. Short-read sequencing approaches based on the Illumina technology will output highly fragmented sequencing reads output, while long-read sequencing based on PacBio outputs longer stretches of sequences[10, 11]. Computational tools and pipeline for shotgun and target metagenomic sequencing analysis are evolving constantly and have become more diverse, fully automate, and available as open-source software. This involves taxonomic application to characterize the phylogenetic relationships of the sequenced data with taxonomic groups of microorganisms known in the

specific database and functional application to interrogate the presence of functional genes having a particular role in the microbial community, for example antimicrobial resistance genes.[11].

Therefore, a computational approach as well as curated database is necessary for more accurate data analysis and interpretation. Metagenomics has numerous applications, including ecology, environmental conservation, infectious disease diagnosis, environmental remediation, biotechnology, agriculture, animal resistome and taxonomic characterization. It mostly focusses on but is not limited to microbial diversity, community composition, genetic, and microbiome interactions and relationships within the environment or host. Metagenomics allows assessment of the diversity of microorganisms and resistomes present in each environment and identify species that were previously unknown or unculturable. It helps scientists study and understand how this biodiversity is affected by continuously changing conditions in the environment and how microbial community interactions within the microbiome play an important role on the health of the host [12]. Metagenomics can be used to monitor changes in microbial communities over time, enabling the study of ecosystem dynamics and responses to environmental changes such as acquired antimicrobial resistance genes on a particular microbiome. Thus, by analyzing the genetic potential of the microbial community, one can gain insights into the various functional capabilities of the microorganisms in a particular environment. Metagenomics can lead to the discovery of novel enzymes, bioactive compounds, and other useful biomolecules with potential applications in medicine, agriculture, and industry. One of the most common applications of metagenomics technology is to interrogate and monitor microbiomes and their role in antimicrobial resistance traits [13, 14] that could be

involved in the creation of new antimicrobial resistance gene reservoirs that could potentially shape the microbial community composition.

1.2 Metagenomics and Antimicrobial Resistance

Research has shown that the use of antimicrobials will lead to an increase in evolution of antimicrobial resistance (AMR), an abundance and distribution of the resistome, and an exchange of antibiotic resistance genes among bacteria [15]. Less appreciated are the concomitant changes in microbiome in response to antimicrobial exposure and their contribution to clinical resistance [16, 17]. Understanding the impact that the use of antimicrobials and environmental resistomes may have on natural microbial population is important for a more comprehensive understanding of the dynamic of antimicrobial resistance [18]. Commensal bacteria may function as a reservoir of resistance genes and increase the accessibility of resistance determinants to potentially pathogenic organisms [19]. The intestinal microbiome is an important reservoir of AMR genes [20], yet the full composition and role of intestinal resistomes in food producing animals is poorly understood. Since antibiotic resistance determinants are readily exchanged among bacterial species, there is an increasing interest in investigating reservoirs of antibiotic resistance accessible to pathogens along the 'farm to fork' continuum. Antimicrobial resistance surveillance is critical for tracking changes in AMR genes [21], identifying new and emerging resistance traits, and monitoring the impact and effectiveness of public health interventions designed to limit the spread of resistance and preserve the effectiveness of antibiotics for humans and animals [22-24]. AMR surveillance is solely founded on the isolation of a few indicator microorganisms followed by phenotypic and

genotypic characterization of these organisms. However, this methodology has limitation as it relies on culture-based detection and characterization, ignoring the potential antimicrobial resistance reservoirs present in microorganisms that cannot be cultured on artificial media or are not routinely examined [25].

Advances in next generation sequencing technologies offer novel non-culture-based approaches for AMR monitoring. Sequencing of whole microbial communities provides comprehensive genomic snapshots of the community including bacteria that are not culturable [26, 27]. Metagenomics in the antimicrobial resistance field allows the detection of AMR genes, which provides information on their prevalence, distribution, and possible routes of transmission and may facilitate the design of interventions aimed at reducing antimicrobial resistance risk [27]. It offers insights into the resistomes, metabolic, and virulome potential of the microbial communities [28, 29]. Although metagenomics approach has promises, interpretation of the massive amounts of data produced poses challenges for public health decision makers. Defining the significance of a given genomic signal in the context of risk is necessary to translate metagenomic data into public health decisionmaking. Another challenge of the metagenomics approach is the difficulty in attributing the identified AMR genes to specific bacterial strains present in the sample [30, 31]. When traditional metagenomics is used to detect AMR genes, it is difficult to predict if they belong to a specific pathogen or to the background microbiota [31, 32]. Recent advances demonstrated the possibility of carrying out assembly of microbial genomes from metagenomics data using the intra-cellular proximity signal captured by Hi-C as a direct

indicator of which sequences originated from the same cell could allow the attribution of AMR genes and mobile genetic elements to specific bacteria host [30, 31, 33-35].

1.3 Antimicrobial Resistance Impacts Global Health

Antimicrobial resistance has increased at an alarming rate not only in a clinical setting but also in the environment and poses a threat to human and animal health worldwide. The alarming resistance levels have been reported in many countries, and treatment for common diseases has become increasingly challenging [19]. Estimates suggest that drug resistant infections cause at least 700,000 deaths per year worldwide, a figure that could increase to 10 million deaths per year by 2050 [36]. One way to overcome the increasing rate of microbial resistance is to monitor the use of antimicrobial drugs in humans, animals, and the environment [16]. Public health organizations have created complete lists of critically important antimicrobial drugs ranked by commonality and effectiveness. This ensures that the most critical drugs are used judiciously in both human and veterinary medicine so that microbes do not become resistant to available antimicrobial drug treatments. Although current policies for antimicrobial use in humans and animals are in place to prevent the rise of antimicrobial resistance such usage does not resolve the increasing rate of antimicrobial resistance. Metagenomics could help strengthen antimicrobial surveillance programs currently in place to more effectively control and minimize future drug resistance.

To develop more robust surveillance programs there needs to be a better understanding of the spread of antimicrobial resistance worldwide. Each bacterial community possesses the ability to evolve over time and become resistant to any harmful environmental stimuli such as

antimicrobial drugs. These antimicrobial resistance genes are contained in reservoirs held by other pathogens or commensal bacteria that combine to form an entire bacterial community [15, 16]. The sole presence of these resistance genes in a bacterial community does not account for the spread of resistance, but their mobilization to new hosts and expression under different environmental settings does. Mobile genetic elements (MGE) such as plasmids and integrative conjugative elements can transfer these antimicrobial resistance genes between bacterial communities. A public health concern [16]. Therefore, metagenomics with help understand when, where, and how these genes spread within different microbial communities and can provide ways predict the emergence of antimicrobial resistance as well as decrease the rates of infectious diseases seen in humans and animals.

1.4 Antimicrobial Use on Food Production Animals

Antimicrobial treatments are typically used during high-risk periods to treat infectious diseases in humans, but they are also used as therapeutics to treat sick animals. Food production animals in the United States are often exposed to antimicrobial treatments to control and prevent disease outbreaks [37]. National regulatory authorities, including the Food and Drug Administration (FDA), monitor antimicrobial treatment on animals and animal food products because of the potential harm it poses to the animal and to the people who consume them. The FDA annually reports the use of approved antimicrobial drugs on food producing animals. In 2018, six million kilograms of antimicrobial were used, (52%) of such drugs were significantly important in human medical therapy where tetracycline accounted for approximately 34% of the total number of antimicrobials used that year and about 66% of medical drugs available for treatment[38]. The use of medically essential antimicrobials, such as

third and fourth generation cephalosporins, fluoroquinolones, and macrolides in veterinary medicine, are the most significant concern to global health [39]. While not all antimicrobials used in veterinary medicine are used in human medicine, most of them have very similar biochemical structures.

Because of the biochemical similar structures, exposure to humans/animals that have antibiotic use could cause antimicrobial resistance. It could be challenging to treat humans with the same or related antimicrobial drugs [40]. Therefore, it is critical to monitor the use of antimicrobials in food-producing animals.

Unfortunately, in food-animal production, treatment of individual animals is not always practical; thereby making it more challenging to monitor and develop suitable standards for the use of antimicrobials on food-producing animals. It is more feasible to treat groups of sick animals by medicating their feed or water source prophylactically for diseases [21, 41, 42] but this can also lead to the subsequent transfer of bacterial resistance genes between bacterial communities present within food animals, animal products, the environment, and humans. Antimicrobial treatment in food animals and feed contributes to the rise in microbial resistance that threatens human and animal health. Such effects can lead to the increase in morbidity and mortality rates and the cost of medical care in both human and animals[40]. Although knowing the quantities and distribution of the types of antibiotics used in animal agriculture can monitor for the overuse of antimicrobials, it fails to provide a suitable means for tracking all antimicrobial treatments used in the agricultural industry. Practical standards can prevent overuse of antimicrobials and promote awareness of the hazards they pose to animal and human health [43]. The FDA reports an annual increase of 9% from 2017-2018 in domestic sales

and the distribution of medically important antimicrobials used in food-producing animals [43]. In 2018, the total annual drug sales in the US equaled more than 11 million. Tetracycline (66%) was reported as the most commonly sold medically important drug, followed by penicillin (12%), and macrolides (8%). Sulfas, aminoglycosides, lincosamides, cephalosporins, and fluoroquinolones accounted for approximately 5% of medically important drugs sold [38]. Among the antimicrobial drugs used in food-producing animals, an estimated 42% was intended for use in cattle, 39% for swine, 11% for turkey, 4% for chicken, and the remaining 4% for use in other species. The distribution of drug classes also varied between food animals. An estimated 81% of cephalosporins, 67% of sulfas, 47% of aminoglycocides, and 44% of tetracyclines were intended for use in cattle; 83% of lincosamides and 41% of macrolides were intended for swine; and 63% of penicillins were intended for turkey[38] . Notably, the antimicrobials most commonly used in human medicine, such as cephalosporins, lincosamides, penicillin, and tetracyclines, comprised about 42% of the antimicrobial classes used in food-producing animals. This is a major concern for public health organizations because of the impacts antimicrobial treatment in food animals poses on the treatments used for human and animal health. The potential increase of medically important antimicrobial drugs that could be in use by farmers and the continuous selective pressure from routine antibiotic use is an essential precondition for the rise in resistant bacterial strains. Therefore, the most important cause of antimicrobial selective pressure is the widespread and indiscriminate usage of medically important drugs, which can result in the dominance of resistant bacteria strains among animal and human pathogens [1].

1.5 Mobile Genetic Elements

Mobile genetic elements are involved in the transfer of antimicrobial resistance genes, enabling a bacterial community to evolve and adapt to the changes in the environment. Mobile genes can confer different phenotypes for different strains of the same species [44]. One of the major forces that shape the genomic landscape of microbial communities is horizontal gene transfer (HGT), the transfer of genes asexually between organisms [45-47]. Plasmids and integrons, are important to the animal gut microbiome because they facilitate the acquisition and expression of AMR genes for the bacteria, and are highly involved in the emergence of antibiotic resistant bacterial strains and mobilization of virulent factors in different pathogenic bacteria [39]. Plasmids are extrachromosomal DNA that can be horizontally transferred between different bacterial cells by conjugation. Horizontal gene transfer of plasmids promotes rapid evolution and adaptation of bacteria by passing various traits involved in antibiotic resistance, virulence, and metabolism to their hosts; they can be classified into narrow-host-range (NHR) and broad-host range (BHR) [48, 49]. The host range of plasmids is an important feature for understanding how they spread in gut microbial communities [50, 51]. Therefore, the alarming rate at which bacteria adapt to antibiotics is partly due to their ability to acquire AMR genes through horizontal transfer of MGE [52]. For example, BHR and NHR plasmids with linkage to *IncQ* and *IncU* plasmids and a colE-type plasmid within Aeromonadaceae, is consistent with reports based on cultured bacteria [44]. Plasmid-mediated resistance has emerged against quinolones [53], carbapenems [37], and colistin [38]. Furthermore, an *ant3'* resistance gene and a class 3 integron marker were linked with a cluster related to an *Acinetobacter johnsonnii*, a pathogen usually found in the environment and animals [54] which

can colonize human skin and cause clinical infections [55], as well as *IncQ* plasmids which had the broadest range of putative hosts, followed by the *IncP-1β* plasmids [56]. Cultivation-independent gut metagenomics approaches have become an essential tool to decipher microbial community interactions, and while this technique can identify MGE's in a sample it cannot assign a specific bacterial host to the AMR genes and the MGE they are encoded to [43]. Host-plasmid association is nevertheless critical to understand the effects of antibiotic resistance and the trajectories that bring resistance genes across bacterial communities [57, 58].

1.6 Recent Advances in Sequencing Technologies

Decreasing costs for Next Generation Sequencing (NGS) technology has made it easier to use metagenomic approaches to sequence bacterial community DNA from a single sample *en masse*. This provides a comprehensive understanding of an entire microbial population, including those that cannot be cultured on artificial media [59]. This advance in sequencing technology has allowed scientists to identify and interrogate the evolving resistance of microbes at an unprecedented scale. Sequencing of pure culture isolates in clinical microbiology laboratories has revolutionized the fields of food safety and public health. Thus, whole-genome sequencing (WGS) has become an essential tool for numerous applications in the field from the development of novel antibiotics to clinical diagnostics, infection control, and antimicrobial-resistant surveillance [60-62]. Next-Generation Sequencing technologies like shotgun metagenomics offer non-culture-based sequencing approaches that provide comprehensive genomic snapshots of bacterial communities and resistome compositions [59, 63]. Comprehensive research can then be employed with metagenomic sequencing to study the

diversity and abundance of AMR genes [61, 62, 64]. For example, shotgun metagenomics can be performed on the commensal bacteria of food-producing animals to show how dairy cattle resistomes during early life serve as a reservoir for over 300 AMR genes that confer resistance to 17 classes of antibiotics [60]. This increases the ability to detect and study antimicrobial resistance gene-bacteria host interactions on the animal gut microbiome and to better characterize uncultured bacteria [44]. We can use metagenomics to assess whether AMR genes are present in pathogens or commensals, track their spread in a community, and assess the potential hazards on animal and human health. However, it is challenging to link AMR genes to their associated MGE and source bacteria from a metagenomic dataset. To overcome such limitations, a combination of deep metagenomics sequencing and Hi-C (high- throughput chromosome conformation capture -3C) proximity ligation can identify long-range interactions in an unbiased, genome-wide fashion [45, 65-67]. Hi-C produces high-quality metagenome-assembled genomes (HQ MAGs) and has been applied in microbial genomics studies to help answer the question of attributing genomic material to individual isolates of bacterial species [44, 67, 68]. It could be used as a direct indicator of which sequences originated in the same cell, enabling of mixed genomes without any reliance on priori information [65]. Thus Hi-C links can be used in a map clustering context to deconvolute contigs into their original cellular groupings and plasmids [44, 69].

Most research on antimicrobial resistance studies relevant to human and animal health are based on the ability to culture and isolate a bacteria colony in a laboratory setting. Because of this, bacteria that are incapable of growing on artificial media are unculturable and disregarded in the majority of antimicrobial resistance studies. Studying unculturable bacteria

can provide invaluable biological information about microbial communities and their potential as new antimicrobial resistance reservoirs [70]. Various culture-based studies have characterized the gut microbiota composition of food-producing animals reporting Pseudomonas and Aeromonas as the most represented in the microbial community in addition to their highly resistant phenotype to oxytetracycline [20]. These culture-based techniques lack the ability to analyze whole microbial populations and could only provide information that is limited to few marker organisms, representing a potential towards culturable bacteria. On the contrary, culture free metagenomics studies will reveal all of the bacteria present in a microbiome that do not grow in artificial medium [44, 63, 71]. Thus, metagenomics will provide a complete picture of bacterial taxonomic and resistome composition in a complex gut microbiome system.

Metagenomics has been implicated as a means for studying the composition of unculturable bacteria. In this technique, DNA is extracted from a microbiome sample and sequenced in a shotgun manner [42, 72]. The sequenced DNA can be pieced together using assembly algorithms or can be analyzed individually using BLAST tools. The resulting assembled contigs can be binned into Metagenome-Assembled Genomes (MAG). Binning can occur across a pool of samples based on nucleotide frequency, abundance, and/or co-variation of abundance [73]. In contrast to binning, a variant-aware metagenomic scaffolder can be used on the meta file to detect repeats by graph analytics [74]. This will identify variants that can be used to precisely reconstruct genomic segments from high complexity microbial mixtures. It can also be used to accurately classify and characterize several classes of common genomic variants [75]. Thus, binning can validate metagenomic approaches as potential means to analyze unculturable

bacteria. A recent metagenomics study re-analyzed the unculturable bacteria within the human gut microbiome to identify unknown species [73]. The study showed a reconstruction of over 60,000 draft prokaryotic genomes out of 3,810 fecal metagenomes samples. This provides reference points for over 2,000 newly identified species-level operational taxonomic units (OTUs), representing a 50% increase of previously unknown phylogenetic diversity in the gut bacteria [73]. This research supports metagenomics as an approach to analyze unculturable bacteria. Therefore, the characterization of previously unculturable microorganisms can provide insight into the unidentified antimicrobial reservoirs responsible for spreading microbial resistance.

1.7 Omics Approaches

There are different omics approaches to fully interrogate an enviromental microbiome composition: shotgun metagenomics, 16S rRNA, Hi-C metagenomics, metatranscriptomics, metaproteomics, metabolomics, and single cell metagenomics. Metagenomics data provides the opportunity to explore two aspects of a microbial community: who is there (microbiome composition) and what are they doing (metatranscriptomics) to interrogate the biological functions associated with the bacterial community and identify novel genes. Metagenomics sequence data presents several challenges due to the complexity of data outputs and the amount of seqeucning data generated, thus it requires a large amount of processing computational power and data storage. To provide insight into the taxonomic macrobiome composition and identify novel genomes as well as functions interactions witing the microbiome, bioinformatic tools for metagenomic analisys could be based on 1) marker gene analisys: comparing each sequencing read to a reference database, thus determining if the read

is a homolog of a marker gene, 2) binning: to cluster the sequencing reads base on similarity, fragment recurrenment, and 3) assembly: where seqeuncing reads are aligned to create contigs, which can subsequently be assembled into scalfolds and complete genomes. To analyse metagenomic sequencing data, computational and bioinformatics tools like: nf-core/mag [76], UMGAP [77], SqueezeMeta [78], Metagenomics Bioinformatic Pipeline [79], HOME-BIO [80], QIIME [81], UPARSE [82], MOTHUR [83], DADA2 [84], MED [85], HiCBin [86], Bin3C [45], qc3C [87], meta3C [88], metaTOR [89], are avaialable on line as open source.

1.7.1 Shotgun metagenomics:

Shotgun metagenomic sequencing allows to comprehensively interrogate all genes in all organisms present in a complex enviromental sample. The method enables to evaluate bacterial diversity, detect the abundance of microbes in specific environments as well as characterized it's gene composition, including 16S rRNA present in the metagenomic sample. Shotgun metagenomics also provides a means to discover and characterize unculturable microorganisms that are otherwise difficult or impossible to analyze. This method is PCR independent leaving little to no bias due to primer binding, enabling novel or less dominant microbes to be sequenced from an environment metagenomic sample [11, 90].

A typical shotgun metagenomics study comprises seven major steps, after the initial study design. 1) sample collection, storage and preservation for quality and accuracy of metagenomics data, 2) community DNA library preparation, quantification, and quality , 3) sequencing platform, depth coverage and read length, 4) sequencing reads data processing and bioinformatic pipeline, 5) computational sequence analysis to profile bacteria taxonomic composition, resistome features of the microbiome, 6) statistical analysis, and 7) data

validation. As the metagenomics field evolves rapidly, there are many experimental and computational approaches available to carry out each step. Nevertheless, shotgun metagenomics has limitations, due to experimental biases, the complexity of the metagenomics sample, sequencing depth of coverage, computational approaches, curated bacteria and resistome databases, and data interpretation. nf-core/mag [76], UMGAP [77], Metagenomics Bioinformatic Pipeline [79], HOME-BIO [80], SqueezeMeta [78] are among the latest bioinformatic tools currently used for shotgun metagenomics analysis.

1.7.2 nf-core/mag is a pipeline for metagenome assembly, binning, and taxonomic classification. It provides a capability to work for short read and long reads, as well as a combination of both. Written in Nextflow to enable scalable and reproducible scientific workflows using software containers, it enables a modularized pipeline structure. The pipeline strongly benefits from the nf-core framework, which enforces a set of strict best-practice guidelines to ensure high-quality pipeline development and maintenance. It uses Docker and Singularity, container technology, to enable reproducibility and portability across different computer systems. nf-core/mag input can be either FASTQ files containing the short reads or a sample sheet in CSV format containing the paths to short and, optionally, long read files. Assembly and binning process includes de novo assembled with MEGAHIT [91] or SPAdes [92]. If both short and long reads are provided, a hybrid assembly can be performed using hybridSPAdes [93] while taxonomic profiling is performed by GTDB-TK [94] or CAT/BAT [94]. Preprocessed short reads are classified using Kraken2 [95] or Centrifuge [96], and visualized in Krona charts [97].

1.7.3 Unipept MetaGenomics Analysis Pipeline (UMGAP), In general, UMGAP performs profiling on individual reads or read pairs, taking into account that protein sequences are more conserved than the genes encoding them, to profile based on protein coding regions translated from the shotgun metagenomics reads, instead of directly processing at the DNA level. The resulting protein fragments are mapped onto a reference protein sequence database. Therefore, it speeds up computational processing, and enables characterization of low abundance microorganisms that cannot be assembled de novo. Gene prediction is carried out relatively fast and accurately by Rust called FragGeneScanRs (FGSrs) [98], but false negatives and false positives due to missed coding regions may lead to information loss and reduced precision of the pipeline [77].

1.7.4 Metagenomics Bioinformatic Pipeline is considered a straightforward computational pipeline that takes shotgun metagenomics raw reads through a quality control process, metagenomic assembly, binning, taxonomic assignment, and taxonomic diversity analysis and visualization. Assembled read are carried out by using metaSPAdes [99], while assembly is done using BinMix [100] which automates the binning of assembled metagenomic scaffolds using an expectation-maximization algorithm after the assembly of metagenomic sequencing reads, CheckM [101] is used for completeness and contamination assessment, and Kraken2 [95] for taxonomic profiling.

1.7.5 HOME-BIO (sHOtgun MEtagenomic analysis of BIOlogical entities) is a modular and flexible pipeline that performs taxonomic characterization and abundance estimation based on exact k-mer matches and de novo assembly modules, to achieve high accuracy and fast classification speeds. One of the features of HOME-BIO is an additional protein validation step

available in both modules for a more robust the taxonomic characterization as well as the use of quality control, metagenomic shotgun and assembly de novo modules are used to perform sequence read quality checks and include FASTQC [102] and MultiQC [103]. It uses Docker container technology. Raw data input includes DNA or RNA, single or paired-end sequencing reads.

1.7.6 SqueezeMeta is a fully automatic bioinformatic pipeline for metagenomics and metatranscriptomics analysis. It utilizes a co-assembly procedure, co-assembling an unlimited number of metagenomes via merging of individual assembled metagenomes, combined with read mapping for estimation of the abundances of genes in each metagenome. It provides internal checks for the assembly and binning steps by CheckM. SqueezeMeta's merged mode permits co-assembly of a large number of samples. Samples are first assembled individually then the resulting sets of contigs are merged by combining contigs with ≥99% semi-global identity, using CD-HIT [104] and re-assembled using Minimus2 [105]. SqueezeMata uses Megahit [106] for assembly but also support SPAdes [92] and Canu [107] and different methods, including Maxbin [100], Metabat2 [108] binning, and DAS Tool [109] to merge the multiple binning results in just one set. Data outputs are merged in several tables and stored in a MySQL database.

1.7.7 Targeted 16S rRNA

The number of unique taxonomic assignments in reference databases limits diversity descriptions. Molecular databases generally lack discrete name assignments for the majority of the microbiome composition, like uncultured microbes. In contrast, taxonomy-independent

Operational Taxonomic Unit (OTU) clustering divides datasets into OTUs without requiring a curated taxonomic database [85]. rRNA universality, activity, and extremely conserved structure and nucleotide sequence allows researchers to study prokaryotics via rRNA target sequences. According to the sedimentation rate, bacterial rRNA are classified by 23S, 16S, and 5S rRNA. 16S ribosomal RNA gene is becoming standard in bacteria taxonomic classification because it is more easily sequenced and contains enough phylogenetic information [110]. The degree of sequence conservation varies widely where more conserved regions correlate to high level taxonomic ranks and less conserved regions to lower levels such as genus and species. Therefore, it is used as gold standard for taxonomic identification at species level [11]. Some of the major advantages and the main reason for 16S rRNA to be used for genome classification is the 16S rRNA gene is universally distributed, the abundance of 16S rRNA gene sequence is higher than another bacterial gene, it could be used to measure phylogenetic relationship across different taxa rank. 16S rRNA has also some disadvantages such as copy number per genome variation, thus output variation among strains can be possible, PCR amplification can introduce biases making relative abundance measurements unreliable, and resolution of the 16S rRNA gene is often too low to differentiate closely related bacteria species. 16S provides limited information compared to shotgun metagenomics [111]. A typical 16S rRNA workflow includes sample preparation, library preparation, sequencing, and data analysis. Sequencing can be conducted by large scale clonal Sanger sequencing, PacBio SMRT sequencing and Illumina MiSeq using a specific set primer for amplification or Illumina NextSeq 2000. When sequencing is done with Illumina MiSeq, V3 and V4 regions are commonly amplified, PCR products are purified and quantified, and Illumina dual index adaptors are added to the

amplicon target. Once the sequencing is done, high quality PCR amplification products are clustered on OUTs based on 97% identity of the reads. Species level annotation and phylogenetic OTU analysis can be performed by using Greengenes, a dedicated full-length 16S rRNA gene database that provides users with a curated taxonomy based on de novo tree inference [112], NCBI taxonomic database [113], SILVA [114] or ribosomal database project RDPII [115]. A Study comparing shotgun metagenomics and 16S rRNA to establish reliability for bacteria profiling, showed that 16S rRNA gene sequencing detects only part of the gut microbiota community revealed by shotgun sequencing [116]. Therefore, shotgun sequencing has more power to identify less abundant taxa than 16S rRNA sequencing. The two major approaches for partitioning large datasets include: (i) taxonomic classification of sequences through comparison with curated databases, for example, GreenGenes [112] or SILVA [114] and (ii) de novo clustering by sequence similarity to define operational taxonomic units (OTUs). There are several bioinformatic tools for 16S rRNA analysis: QIIME [81], UPARSE [82], MOTHUR [83], DADA2 [84], Minimum entropy decomposition (MED) [85].

1.7.8 QIIME is a pipeline for performing microbiome analysis from raw sequencing data through publication statistics and visualization. It was developed on the basis of a plugin architecture that allows third parties to contribute functionality, it supports the latest generation tools for sequence quality control from different sequencing platforms. The plugins also support qualitative new functionality, including microbiome paired sample and machine learning. QIIME is a downstream data analyses program that uses the PyCogent toolkit to address the problem of interpretation and database deposition using raw sequencing data. QIIME is not specific to 16S rRNA genes; as a plugin architecture platform many users

commonly use other coding and non-coding marker genes, e.g., COI, ITS, even some functional genes like nifH. QIIME 2 provides many new interactive tools for taxonomic assignment, phylogenetic reconstruction, visualizations, facilitating exploratory analyses and result reporting. It also allows users to securely share and interact with results without installing QIIME 2. Availability exists as open source, tutorial, basic command lines, and compressed packages for friendly use.

1.7.9 UPARSE is a bioinformatic pipeline to construct OTUs de novo, achieves high accuracy in biological sequence recovery, and improves richness estimates on mock communities. Clustering is performed by a novel 'greedy' algorithm that performs chimera filtering and OTU clustering simultaneously (UPARSE-OUT). UPARSE could be successfully applied to a wide range of marker genes and sequencing technologies. UPARSE workflow includes merging of paired reads, if applicable; read quality filtering; length trimming; merging of identical reads (dereplication); optionally discarding singleton reads; and OTU clustering using the UPARSE-OUT and UPARSE-REF algorithm as follows: UPARSE-OTU is a greedy clustering method that uses a single representative sequence to define each cluster (OTU), if the read matches an existing OTU within the identify threshold (default 97%), the OTU abundance is updated but the database is otherwise unchanged. Otherwise, a model of the read is constructed by the UPARSE-REF algorithm using the current OTU database as a reference. If the model is chimeric, the read is discarded; otherwise, the read is added to the database and thus becomes the representative sequence for a new OTU. Tutorials are available.

1.7.10 MOTHUR is one of the most comprehensive analyses software. It implements algorithms from DOTUR [117], ∫-LIBSHUFF [118], SONS (shared OTUs and similarity)[119], TreeClimber [120], and UniFrac [121]. Mothur has become one of the most cited bioinformatics tools for analyzing 16S rRNA gene sequences. Its principal is to address how to assign sequences to OTUs and how to use that data to estimate the richness and diversity of a microbial sample. DOTUR (distance based OTUs and richness) uses a matrix quantifying the genetic distance between pairs of sequences to cluster those sequences into OTUs for any distance threshold to define the OTUs then using the frequency of each OTU to calculate a variety of alpha diversity metrics. ∫-LIBSHUFF, SONS (shared OTUs and similarity), and TreeClimber will help estimate richness and diversity, with greater flexibility to specify pairwise comparisons, generate dendrograms, heat maps, and Venn diagrams, and the ability to call individual commands either from within the Mothur, using files with lists of commands like batch files, or directly from the command line, providing for more flexibility for users.

1.7.11 The Divisive Amplicon Denoising Algorithm (DADA) introduced a model-based approach for correcting amplicon errors without constructing OUT, implementing the full amplicon workflow for analysis such as filtering, dereplication, sample inference, chimera identification, merging of paired-end reads. DADA2 is an open-source R package that improves the DADA algorithm, it uses a parametric model to infer true biological sequences from sequencing reads. The model relies on input read abundances and distances. DADA2 uses a probability threshold to decide whether to assign counts from a less abundant, “error-derived” read to a more abundant, “true” sequence, infers exact amplicon sequence variants from high throughput amplicon sequencing data, replacing the coarser and less accurate OTU clustering

approach. The DADA2 pipeline inputs demultiplexed fastq files, and outputs the sequence variants and their sample wise abundances after removing substitution and chimera errors. It produces a single nucleotide resolution with a lower false positive rate. The precision of DADA2 improves downstream measures of diversity and dissimilarity, and potentially allows amplicon methods to probe strain level variation.

1.7.12 Minimum entropy decomposition (MED) is a pipeline to distinguish fine-scale diversity in amplicon data. Its algorithm extends the principles of oligotyping and uses Shannon entropy to identify information-rich positions within an internal node on the entire marker gene data sets. The MED algorithm operates on high-throughput sequencing datasets without requiring an initial DNA sequence alignment step. It provides a computationally efficient means to partition marker gene datasets into MED-nodes, which represent homogeneous operational taxonomic units. MED partitions high-throughput sequencing data sets into ecologically meaningful and phylogenetically homogeneous units. Thus, iteratively partitions amplicon sequences into homogenous OTUs that serve as input to alpha- diversity and beta-diversity. MED uses GAST [122] to assign taxonomy to our reads individually while for OTUs clustering QIIME [123].

There are other metagenomics tool analysis like MANTA (Microbiota And pheNoType correlation Analysis platform) [124]. The framework consists of a database and a web application, including a suite of analytical and visualization tools. Manta, instead of focusing on taxonomic profiling and comparative analysis among groups, correlate microbiota composition with the host phenotypic parameters. It is a web-based integrative database and analysis platform that provides (i) a smooth and interactive user interface to analyze the data quickly

and efficiently with no programming efforts on the part of the user, (ii) the ability to save the data in a readily accessible format for all the microbiome and phenotypic data using PostgreSQL [125], and (iii) flexibility and easily installed on individual workstations or servers to ensure quick access and secure data storage. Microbiota composition, pre-processing is performed to provide additional information for identification of the dominate taxonomic rank for each metagenomics sample for data visualization. This annotation is achieved by merging the low abundance taxonomies in the "others" category set below a specific threshold. Shannon, Simpson, and Chao1 alpha diversity indices are included for a more robust data analysis as well as phylogenetic distances used for hierarchical clustering and principal coordinate analysis, such as Jaccard distance, weighted and unweighted UniFrac distance.

1.7.13 Hi-C metagenomics, as a culture-independent genomic approach, metagenomics avoids the limitation on isolation or cultivation of microorganisms and provides a snapshot of the community structure and the functional capabilities present complex microbiomes. Thus, it offers insights into the resistomes, metabolic, and virulome potential of the microbial communities. However, one of the challenges of the metagenomics approach is the difficulty in attributing the identified genes to specific bacterial strains present in the sample. When traditional metagenomics is used to detect specific genes, it is difficult to predict if they belong to a particular bacterium or to the background microbiota. High-throughput proximity ligation (Hi-C), a DNA proximity ligation technique has the capacity to retrieve metagenomic assembly genomes (MAGs) accurately from a single sample, it could generate millions of paired-end reads based off sequence depth coverage and linking DNA fragments in close proximity in 3D space within cells. [126, 127]. When the Hi-C technique is coupled with traditional shotgun

sequencing, it shows great capabilities for genome binning and retrieval of high quality MAGs from complex microbiome samples [30, 32, 35, 128]. Thus, the workflow for Hi-C metagenomics can be summarized on five steps 1) metagenomic sample is processed for Hi-C metagenomics libraries: (i) Formaldehyde treatment to induce covalent bonds between DNAs that are close in three-dimensional space and therefore within the same cell. (ii) Hi-C library preparation via restriction enzyme digestion, followed by ligation and junction enrichment, to preparation of DNA fragments for paired-end sequencing. Sequencing reads from fragments generated by Hi-C provide linkage information of non-contiguous DNA originated from the same cell. The proximity ligation step is where blunt ends are ligated under dilute conditions to permit ligation to occur preferentially among DNA strands bound in the same protein complex. 3) high throughput deep sequencing that depends on the complexity of the metagenomics samples could require >100 million reads for shotgun metagenomics and > 75 million reads for Hi-C metagenomics. 4) paired-end Hi-C reads are mapped to the assembled contigs to generate raw contact maps, which are then normalized to correct experimental biases. 5) Normalized contact maps are clustered to construct draft genome bins.

Binning approaches are crucial to deconvolute contigs into their original cellular groupings, including genes, chromosomes, and plasmids. Due to the excessive amount of data generated from the metagenomics sequencing, Hi-C metagenomic approaches can demand significant resolving power and high data storage requirements when combined with conventional shotgun sequencing. ProxiMeta, a commercial metagenomic genome binning platform, utilizes contig abundance to normalize the raw Hi-C contacts and then clusters contigs into genome bins using a proprietary MCMC-based algorithm based on their Hi-C proximity linkages. It aims

to lessen the experimental challenges in library preparation for non-specialized laboratories while also raising the quality of data generated. There are a few open source bioinformatic tools with the capability to take Hi-C metagenomics sequencing data to generated MAGs and Species-level deconvolution contact maps. MetaTOR [89], HiCBin [86], bin3C [45].

Proximity ligation methods such as Hi-C have been used to detect intra and extra chromosomal interactions between DNA molecules originating in the same cell within microbial communities. Hi-C is able to reconstruct strain and species level genomes from complex bacteria communities and map plasmids and genes to the bacteria host [30, 44, 129, 130]. Nevertheless, normalization methods and clustering algorithms for the Hi-C metagenomic binning are still rare.

The richness of biological information captured in Hi-C proximity ligation experiments is such that the technique has subsequently been applied to a wide range of problems in genomics, like such as genome reassembly [131] , assembly clustering [132, 133], and centromere prediction [134].

1.7.13.1 Metagenomic Tridimensional Organisation–based Reassembly (MetaTOR) is an open-source computational pipeline that bins DNA contigs into individual genomes according to their 3D contact frequencies then, those contacts are quantified by proximity-ligation approaches. MetaTOR is a comprehensive, end-to-end computational pipelines to process Hi-C and meta3C datasets. Sequencing read are assembled by Megahit [106] and analyzed by MG-RAST pipelines [135] to automate annotation of complete or draft microbial genomes and provides information on phylogenetic and functional classification of the contigs as well as alpha diversity measurement of the assembly. MetaTOR divided the raw Hi-C contacts by the

product of coverage of contig pairs and then applied the Louvain algorithm with the classical Newman-Girvan criterion to cluster contigs. Binning products are assessed for completeness and contamination using CheckM [101]. Gene prediction is performed using [136], and genes of interest are detected using HMM models [137]. HiCBin, bin3C, and a shotgun-based binning tool MetaBAT2 [138] only contain the binning analysis, whereas MetaTOR uses its own workflow and integrates the binning process with standard alignment and annotation softwares [86].

1.7.13.2 **HiCBin** is a Hi-C based binning pipeline to recover high-quality MAGs by employing the HiCzin normalization method [139] to removed experimental biases. HiCBin also detects and removes the spurious inter-species contacts for the first time among all Hi-C-based binning pipelines. Leiden algorithm [140], an advanced modularity-based community detection algorithm is used for binning. The principal behind HiCBin is to outperform experimental biases for raw metagenomic Hi-C contacts: the number of restriction sites, contig length, and contig coverage, and to demonstrated that normalization methods in publicly available Hi-C-based binning pipelines cannot correct all explicit biases. Hi-C metagenomics normalization procedures utilize restriction enzymes like Sau3AI and MluCI that experimentally could be unspecified under different sample complexity, therefore, bioinformatic pipelines using the information of the number of restriction sites to do normalization cannot be executed property, generating bias. Employing the HiCzin mode that merely normalized raw Hi-C contact maps by the length and coverage of contigs rendered the HiCBin pipeline more applicable. HiCzin normalization method and the Leiden algorithm combined with the Potts spin-glass model from the HiCBin provided the best results in genome binning [86]. HiCBin requires an

extra preprocessing step on running the TAXAassign software in order to annotate some contigs (https://github.com/umerijaz/TAXAassign). Comparing CheckM validation results, HiCBin retrieved 67 near-complete, 33 substantially complete, and 12 moderately complete MAGs. Near-complete MAGs ranged from 1,509,376 bp to 4,884,489 Mbp, while the substantially complete MAGs ranged from 1,549,630 to 2,662,928 bp, and moderately complete MAGs ranged from 1,336,581 to 2,881,511 bp. In comparison, MetaBAT2 generated 30 near-complete, 25 substantially complete, and 22 moderately complete MAGs. ProxiMeta created 247 bins with 50 near-complete, 14 substantially complete, and 11 moderately complete MAGs while bin3C constructed 138 bins with 60 near-complete, 24 substantially complete, and 11 moderately complete MAGs. Genomic bins recovered by MetaTOR suffered serious contaminations without the recursive Louvain clustering. The average contamination of bins with completeness larger than 50% was 80.94% [86].

1.7.13.3 **bin3C** is an unsupervised method that exploits the hierarchical nature of Hi-C proximity ligation interaction rates to resolve MAGs using a single time point. It is based on two stage normalization method to process the raw contact maps. It first divided the raw Hi-C counts by the product of the number of restriction sites and then used the Knight-Ruiz algorithm[141] to construct a general doubly stochastic matrix. Bin3C utilizes the Infomap software [142] to bin contigs. bin3C can retrieve MAGs from metagenomes by combining conventional metagenome shotgun and Hi-C sequencing data. Basic steps on bin3C are as follows: Read clean-up for both shotgun metagenomics and Hi-C read-sets using bbduk from the BBTools suite (https://www.sourceforge.net/projects/bbmap/), shotgun assembly grouping contigs into scaffolds done by SPAdes [99]. HiC reads are mapped to the corresponding scaffold

by BWAMEM[143]. The resulting BAM files were subsequently processed using samtools [144] to remove unmapped reads and supplementary and secondary alignments, then sorted by name and merged. bin3C constructs a contact map from the Hi-C read-pairs under constraints on minimum acceptable length (default, 1000 bp) and minimum acceptable raw signal (default, five non-self-observations) for contig inclusion. Based on the assumption that DNA-DNA interactions between loci within a same cell (intra-cellular proximity interactions) occur an order of magnitude more frequently than interactions between cells (inter-cellular), unsupervised cluster analysis by using Markov clustering (https://dl.acm.org/doi/book/10.5555/868986), the Louvain method [145], and infomap (https://github.com/mapequation/infomap) are use constructs genome bins while CheckM [101] for completeness and contamination.

A comparison between the CheckM validation of bin3C results to the CheckM validation of ProxiMeta a commercial metagenomic genome binning platform, and MaxBin [100] a well performing MAG retrieval tools [146], reports that ProxiMeta retrieved 35 nearly, 29 substantially, and 13 moderately complete MAGs, while MaxBin retrieved 20 nearly, 22 substantially, and 17 moderately complete MAGs, while bin3C retrieved 55 nearly, 29 substantially, and 12 moderately complete MAGs [45].

PART II

2. **Research Objective**

2.1 Evaluating the impact of depth of coverage on microbiome and resistome analysis

- To capturing most of DNA sequences carried by every microorganism on a microbiome.

2.2 Characterization of previously uncultured food animals' intestinal microbiota

- To better understand animal microbiome composition and their potential to become antimicrobial resistance reservoirs.

2.3 Attribution of antimicrobial resistance gene and plasmid to host bacterial cells.

- To help define potential antimicrobial resistance reservoirs present in gut microbial communities and their role in the spread of antimicrobial resistance.

The main goal of this graduate research is to map antimicrobial resistance genes and plasmids to the host bacterial cells for widespread distribution of the AMR genes by combining intra-cellular proximity signal captured Hi-C and shotgun metagenomics approaches. Thus, to measure changes in microbiome and resistomes composition on cecal microbiomes. This approach could facilitate identification of sources and hotspots of AMR along the food chain and measure the effect of changes in agricultural practices.

3. RESEARCH DATA SET

We utilized cecal metagenomics dataset(s) generated from a longitudinal study involving four-major food animal classes (cattle, swine, chickens, and turkeys), monitored by the National Antimicrobial Resistance Monitoring System (NARMS).

4. EVALUATIONG THE IMPACT OF DEPTH OF COVERAGE ON MICROBIOME AND RESISTOME ANALYSIS IN FOOD ANIMALS

4.1 Abstract

Metagenomic studies have dramatically increased our knowledge on microbial composition of different environments. Capturing all DNA sequences carried by every microorganism in the microbiome is still unlikely. In the context of bacteria taxonomic characterization and resistome profiling, sequencing depth of coverage is crucial in accurately characterizing the composition of microbial communities. Therefore, estimating a suitable sequencing depth to achieve microbial characterization is at most. This research study used shotgun metagenomics and different sequencing depth approaches utilizing Illumina sequencing strategies. Illumina NextSeq, Illumina HiSeq High Output mode, and Illumina HiSeq Rapid Run mode, to establish an optimal sequencing depth for gut metagenomics studies. A total of 72 cecal metagenomics samples from four major food animal production classes (cattle, swine, chicken, and turkey) monitored by the National Antimicrobial Resistance Monitoring System (NARMS) were utilized for this study, allowing a comprehensive profiling of the microbiome composition.

We obtain three different read average (i) 31 million reads on average, ii) 27.2 million reads on average, and iii) 7.1 million read on average.

Approximately 400 bacterial families with different relative abundance were characterized across all approaches among all seventy-two cecal metagenomic samples. i) 31 million reads/average showed a total of 153 bacteria families (39.13%), with an average per sample of

80 taxa-family rank with Bacteroidaceae, Lactobacillaceae, Enterobacteriaceae, Clostridiaceae, Enterococcaceae, and Campylobacteraceae, representing the most abundant families. On the other hand, ii) 27.2 million reads/average denoted a total of 106 families (27.11%), with an average per sample of 41 taxa-family ranks, were Bacteroidaceae, Lactobacillaceae , Enterobacteriaceae, Enterococcaceae, Clostridiaceae, and Campylobacteraceae, were the most abundant families, while iii) 7.1 million read/average outlined a total of 132 families (33.76%), with an average per sample of 64 taxa-family ranks, with Bacteroidaceae, Lactobacillaceae, Enterobacteriaceae, Enterococcaceae, Clostridiaceae, and Campylobacteraceae representing the most abundant families.

Over 350 acquired AMR genes across all approaches, representing 11 antimicrobial classes (aminoglycosides, betalactam, fosfomycin, macrolide, nitroimidazole, phenicols, quinolone, sulphonamide, rifampicin, tetracycline, trimethoprim, and vancomycin). At i) 31 million AMR RPKM profiled 104 unique AMR genes representing 11 AMR classes, ii) 27.2 million AMR-RPKM profiled 119 AMR genes corresponding to 11 AMR classes, and iii) 7.1 million AMR-RPKM profiled 94 AMR genes representing 11 AMR classes. Aminoglycoside, Macrolide, Beta-lactamase, and Tetracycline resistance genes were the most abundant.

We have used deep shotgun metagenomics sequencing approaches to estimate a suitable sequencing depth to achieve microbial and resistome characterization. The number of sequences required to characterize a sample depends primary on the goal of the study, the complexity of the microbiome, as well as the available resources. It is imperative to note that there are no abundance thresholds below which a microbe can be overlooked as unimportant; every member of a bacterial community plays critical role in the functional relationship to

maintain the good health of the microbiome. Overall, our data indicate that sequencing depth critically affects the inferred AMR gene content and taxonomic diversity of complex microbiomes. The more sequencing depth the better characterization of microbial composition. These results can help to better understand uncultured bacteria and measure potential new AMR gene reservoirs present in intestinal microbiomes and their latent hazards on animal and human health. This study provides insights into food animal intestinal microbiome and resistome composition. Furthermore, it establishes an optimal sequencing depth for animal microbiomes to provide more robust scientific findings that could avoid compromising the biological outcomes that can be achieved from an experimental design.

4.2 Introduction

Many efforts have been made to monitor the spread of antimicrobial resistance in humans and animals. Culture-based approaches provide limited information on microbial composition. Likewise, they provide incomplete characterization of antimicrobial resistance (AMR) genes within intestinal microbiomes, disregarding potential antimicrobial resistance reservoirs present in microorganisms that cannot be cultured on artificial media [147]. This dramatically impedes understanding of the role of resistance genes present in these organisms. However, advances in next generation sequencing technologies offer a culture independent approach, metagenomics, to study microbes in their natural environment to get a global snapshot of the microbial community and its antimicrobial resistance potential. Although AMR is a natural phenomenon, resistance develops more rapidly through the misuse and overuse of antimicrobial agents. AMR has increased at an alarming rate in clinical settings and in the environment, which poses a threat to human and animal health worldwide [16]. The alarming

resistance levels to antimicrobials and the great diversity of genes that potentially could be acquired and used by pathogens to counteract the effect of antibiotics have been reported in many countries, and treatment for common diseases has become increasingly challenging[147-149]. AMR threatens the effective prevention and treatment of an ever-increasing range of infections caused by a microorganism such as bacteria, fungi, viruses, and parasites [150, 151]. AMR can occur when microorganisms change because of exposure to antimicrobial drugs[149]. These AMR genes are in reservoirs held by other pathogens or commensal bacteria that combine to form an entire bacterial community [152]. The mere presence of these resistance genes in a bacterial community does not explain the spread of resistance. Instead, the mobilization of those genes to new hosts and their expression under different environmental conditions that allows the spread of resistance [15]. Mobile genetic elements (MGE) such as plasmids and integrative conjugative elements can transmit these AMR genes between bacterial communities, making this a public health concern [19, 152]. Metagenomics helps assess whether AMR genes are present in pathogens or commensals, track their spread within a community, and evaluate the potential hazards on animal and human health, particularly, the combination of metagenomics approaches, bioinformatic tools, and computational platforms have facilitated the characterization of microbial communities and has allowed access to genome data without the need for isolating and culturing microorganisms [153]. Metagenomics shows an unbiased estimation of the bacterial composition without isolating or culturing a specific organism. However, unaffordable sequencing costs often limit the number of sequences depth that can be experimentally obtained, consequently restricting the biological

importance achieved from an experimental design and thus, critically affecting microbial characterization and resistome composition.

Depending on the complexity of the metagenomic sample, it may require a sufficient sequencing depth to recover the total composition of a mixed bacteria population and it resistome potential [154]. Metagenomics could provide the solution to a cultureless future in clinical microbiology, food safety, and public health, as well as identify clinical causative agents or foodborne contamination detection of AMR genes and virulence factors, in addition provides high resolution subtyping data [26]. Whole Genome Sequencing (WGS) approaches that rely on combinations with Next Generation Sequencing (NGS) technology have become an essential tool for numerous applications in the field from the development of novel antimicrobials to clinical diagnostics, infection control, and antimicrobial resistant surveillance [60, 155]. Nevertheless, a non-optimal sequencing dept can critically affect the ability to fully interrogate complex microbiomes, to be able to identify minor genetic variants or to separate closed related bacteria [156, 157]. reduction of sequencing depth had major impact on the sensitivity of whole genome sequence for profilin, increasing the number of undetected species [158].

4.3 Sample collection.

Seventy-two cecal samples collected from food animals (7 from cattle, 18 from swine, 24 from chicken and 23 from turkey) were used in the study. Cecal samples were collected as part of the National Antimicrobial Resistance Monitoring System's (NARMS) routine sample collection from cattle, swine, chicken, and turkey at slaughter establishments across the United States. Samples were collected according to the United States Department of Agriculture (USDA) Food Sampling and Inspection Service (FSIS) FSIS Directive 10,100.1, "FSIS Sampling for

the National Antimicrobial Resistance Monitoring System (NARMS). https://www.fsis.usda.gov/policy/fsis-directives/10100.1." Cecum contents (cattle, swine, or turkey) and intact ceca (chicken or turkey) were shipped in 120mL sample containers in a chilled cooler. Upon arrival, samples were frozen at -80°C until DNA extraction.

4.4 Community DNA extraction, and quantification.

Metagenomic DNA was extracted from approximately 0.2g of cecal contents using PowerMicrobiome® Kit (QIAGEN, Inc., Germantown, MD, USA) following the manufacturer's recommendation. The concentration and quality of the extracted DNA were determined using Qubit™ 3.0 fluorometer (Thermo Fisher Scientific, Wilmington, DE, USA) and TapeStation system (Agilent Technologies, Inc., Santa Clara, CA, USA) and stored at −20 °C until sequencing.

4.5 Shotgun metagenomic sequencing.

Extracted DNA was diluted to a concentration of 0.2 ng/µl. Five microliter of the diluted DNA was used for subsequent library preparations. Sequencing libraries for each sample were constructed using the Illumina Nextera XT DNA library preparation kit (Illumina, Inc., San Diego, CA, USA) and sequenced on HisSeq 2500 and NextSeq 500 systems (Illumina, Inc., San Diego, CA, USA) according to the manufacturer's recommendations. Briefly, tagmentation and tagging of the DNA with unique adapter sequences were performed using Nextera transposome. Limited-cycle PCR was used to amplify the tagged DNA and add indexes. Libraries were pooled together and spitted to be sequenced on HiSeq 2500 High Output mode (2 x125 bp paired-end sequencing), HiSeq 2500 Rapid Run mode (2 x250 bp paired-end sequencing), and NextSeq 500 system (2 x150 bp paired-end sequencing). Base calls generated were converted to FASTQ files for further analysis. FASTQ sequences from the metagenomics sequencing were trimmed with

respect to adaptors and low quality sequences using fastp [159] with parameters ILLUMINACLIP:Illumina-Adapter.fa:2:30:10 LEADING:20 TRAILING:20 SLIDINGWINDOW:5:20 MINLEN:90. To remove host DNA sequences, the trimmed and filtered FASTQ sequences were then mapped to cattle, chicken, turkey, and swine genomes based on the source of the samples with Bowtie2 [160] using standard settings. The mapped and unmapped reads were separated using Samtools [144]. Reads that were trimmed, filtered and unmapped to host DNA were used as an input for all further downstream analysis.

4.6 Taxonomic and resistome analyses:

Taxonomic classification was performed using Kraken2 [95] using the latest NCBI database version built with the "kraken-build standard" method. Kraken2 with the default k-mer size and parameters were used for the analysis. A threshold of 0.2 was used for kraken-filter step. The microbiome composition and the taxa abundances were estimated by the Bracken. The presence and the abundance of AMR genes in metagenomic dataset were determined using the Short, Better Representative Extract Data set ([161]) . ShortBRED-Identify was used to generates unique peptide markers for AMR genes-protein sequences compiled from AMRFinderPlus version 3.10 database (downloaded 2021-09-30; Index of /pathogen/Antimicrobial_resistance/AMRFinderPlus/database (nih.gov)). ShortBRED-identify used an 85% amino acid identity threshold to cluster the AMRFinderPlus database into nonredundant representative sequences and to query representative sequences against one another and against the comprehensive protein reference datasets UNIREF100 (https://www.uniprot.org/uniref/) (data accessed 19 October 2018). ShortBRED-quantify then was used to map translated reads to protein markers at ≥95% amino acid identity across ≥95%

of the marker length and generate the abundance of the AMR genes, normalized to reads per gene kilobase per million reads (RPKM).

4.7 Statistical analysis:

Descriptive statistics was used to describe and understand the features (measures of central tendency, measures of variability, and frequency distribution) of the metagenomic data set by giving short summaries about the sample and measures of the data.

Negative binomial regression was used to model the probability that a certain number of metagenomic samples will experience changes on microbiome and resistome and profiling as a result changes on sequencing depth coverage.

Negative binomial regression shows significant differences between the means of three sequencing depth performance (i) 31 million reads on average, ii) 27.2 million reads on average, and iii) 7.1 million read on average, to profile animal microbiome and resistome composition.

One-way analysis of variance (ANOVA) was used to determine whether there are statistically significant differences between the means of three sequencing platform performance parameters to characterize animal microbiome and resistome composition on a complex microbiome. Data visualization was performed using ggplot2 in R .

4.8 Results

Seventy-two cecal samples from cattle, swine, chicken, and turkey, a major food production animal, monitored by the National Antimicrobial Resistance Monitoring System (NARMS), were utilized to evaluating the impact of sequencing depth on microbiome and resistome analysis from a complex microbiota. Community DNA was extracted as described in methods and paired ends (PE) sequenced using three different Illumina sequencing approaches with different sequence

read generation parameters: NextSeq500 sequencer (NS), HiSeq 2500 Rapid Run (RR), and HiSeq 2500 High Output (HO).

To interrogate microbiome and resistome composition of ceca's food animals, average sequencing depths by sequencing approach was examined. The total number of sequencing reads varies from sample to sample and across the sequencing approaches, HiSeq-HO ranges from 13,033,418 to 50,406,093 reads, with an average per sample of 31,015,644. NextSeq ranged from 1,418,862 to 76,463,359 reads, with an average read per sample of 27,184,285, and HiSeq-RR ranged from 1,295,004 to 10,269,362 reads, with an average read per sample of 7,057,574. (Figure 1). Table 1 shows a total sequencing depth of coverage (millions of reads) obtained by sequencing approach per cecal sample. Therefore, based on the sequencing performance approaches used, three different sequencing read averages were established: i) 31 million reads on average, ii) 27.2 million reads on average, and iii) 7.1 million read on average, to obtain an optimal sequencing depth read coverage for metagenomics studies specially, microbiome characterization and resistome composition.

Significant difference between the three sequencing approaches/modes was observed. Degrees of freedom (Df) = 2 , Sum of squares (Sum Sq) = 2.385e+16 , mean of the sum of squares (Mean Sq) = 1.193e+16, test statistic from the F test (F value) = 188.7, p-value of the F-statistic (Pr(>F))= 2e-16. Tukey's Honestly Significant Difference (Tukey's HSD) post-hoc test for pairwise comparisons helps to identify that there are statistically significant differences from one another: NS (27.2 million reads/average) / HiSeq-HO (31 million reads/average) 3831359, HiSeq-RR (7.1 million reads/average)/ HiSeq-HO (31 million reads/average) 23958070, and HiSeq-RR (7.1 million reads/average) / NS (27.2 million reads/average) 220126711.

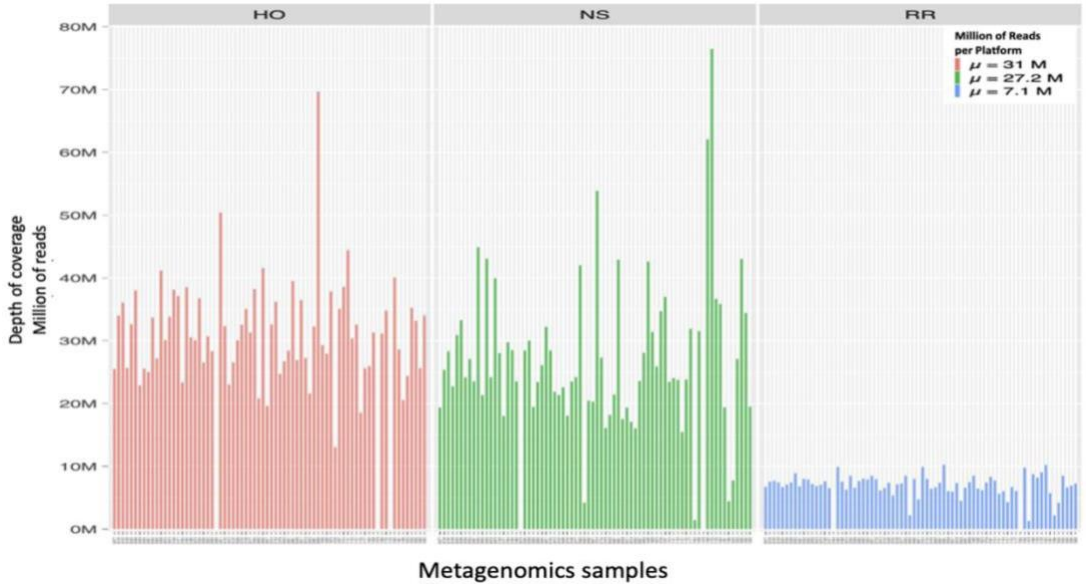


Figure A

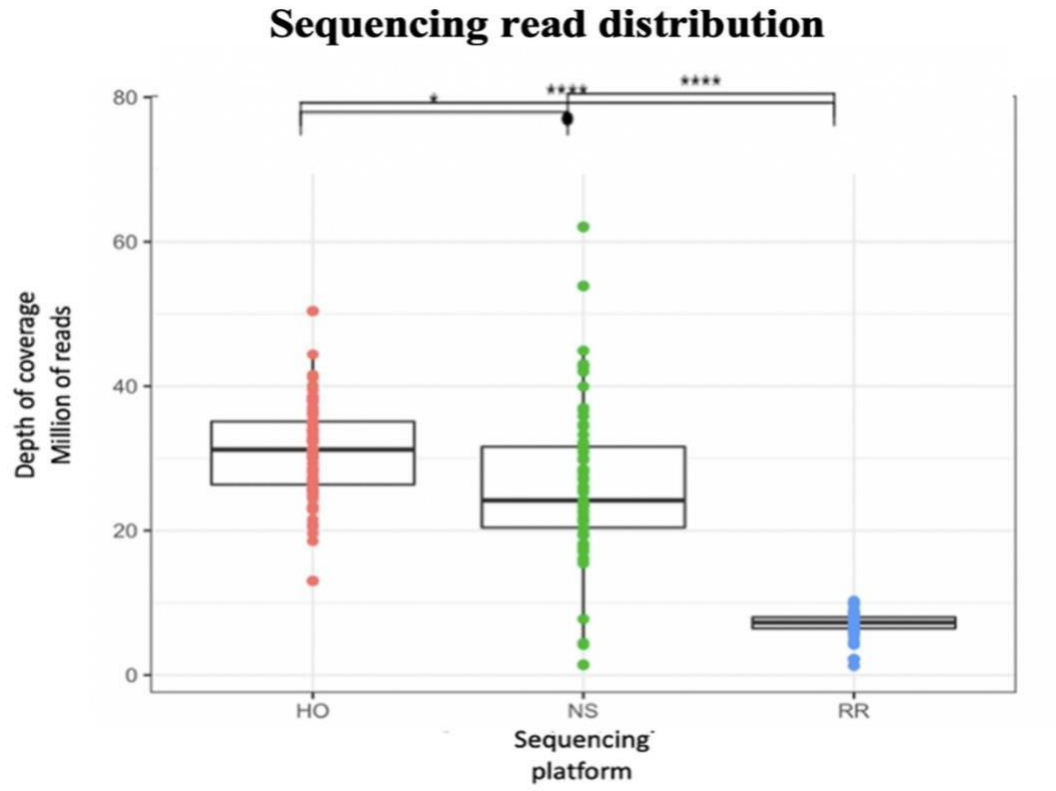


Figure B.

Figure 1. [A] Sequencing depth of coverage for metagenomics samples across multiple sequencing platforms. [B] Box plot showing sequencing read distribution. HO, i) 31 million reads/average, NS, ii) 27.2 million reads/average, RR, iii) 7.1 million read/average. Standard

deviation: i) 31 million reads/average: 6.5, ii) 27.2 million reads/average: 11.9, iii) 7.1 million read/average: 7.6.

Total Depth of Coverage by Averaged Sequenicing Depth on a Complex intestinal microbiomes					
Source ID	Depth Coverage/ Million Reads	Source ID	epth Coverage/ Million Rea	Source ID	Depth Coverage/ Million Reads
NS126	25365397	HO10496	29278916	RR10496	8497545
NS165	28291042	HO10589	27923088	RR10589	6451144
NS168	22754615	HO10674	37855386	RR10674	6192541
NS204	30874821	HO10756	13033418	RR10756	7403361
NS296	33258687	HO10787	35063350	RR10787	8305157
NS299	24161737	HO11014	38569447	RR11014	7728152
NS353	27062736	HO11131	44387620	RR11131	5629667
NS540	44905444	HO11145	30373290	RR11145	6031077
NS820	21326850	HO11151	32558527	RR11151	4288323
NS909	43032354	HO11152	18551524	RR11152	6705717
NS910	24163619	HO11177	25587717	RR11177	6080507
NS1349	28018278	HO11650	25960910	RR11656	9765285
NS1371	17999267	HO11656	31282623	RR11678	1295004
NS1844	29768291	HO11689	31130000	RR11689	8739843
NS1845	28475998	HO11720	34812869	RR11720	8172648
NS1883	23512519	HO11853	40075484	RR11725	9039683
NS2294	28428563	HO11991	28578371	RR11853	10229211
NS2599	29996621	HO11996	20537549	RR11991	5738230
NS2622	19471717	HO12023	24391077	RR11996	2191615
NS2650	23412370	HO12032	35238548	RR12023	4190448
NS3611	32214842	HO12062	33166841	RR12032	8485983
NS4152	28427544	HO12082	25636417	RR12062	6660193
NS5012	21847125	HO12083	33993467	RR12082	6925653
NS5124	21340904	HO126	33999990	RR12083	7258199
NS5399	22544963	HO1349	38136525	RR126	7545766
NS5702	18042100	HO1371	37115448	RR1349	7618102
NS5748	23502144	HO165	36060074	RR1371	6511311
NS5777	24197074	HO168	25648931	RR165	7701870
NS5902	42009893	HO1844	23343922	RR168	7435714
NS6655	4191256	HO1845	38528460	RR1845	9909936
NS6940	20452982	HO1883	30562437	RR1883	7544513
NS6971	20286202	HO2824	30098584	RR2010	6268555
NS7615	53870395	HO204	32640061	RR204	6702263
NS7744	27300358	HO2294	36763726	RR2294	8469097
NS7976	16104274	HO2599	26515074	RR2599	6605543
NS8035	18162808	HO2622	30697305	RR2622	7655953
NS8291	21402726	HO2650	28326498	RR2650	8041771
NS8991	42884749	HO296	38002240	RR296	7085714
NS9208	17503845	HO299	22878106	RR299	7439320
NS9211	19345536	HO353	25526155	RR353	8904789
NS9216	17083520	HO3611	50406093	RR3611	8492189
NS9223	16049408	HO4152	32312615	RR4152	7954835
NS9581	23580766	HO5012	23034471	RR5012	6150311
NS9858	28091973	HO5124	26558110	RR5124	6542420
NS10496	42630617	HO5399	30056949	RR5399	7375379
NS10589	31403548	HO540	33678077	RR540	8024221
NS10674	25899050	HO5702	32533229	RR5702	5365843
NS10756	34693446	HO5748	35049571	RR5748	7131832
NS10787	36981265	HO5777	31334707	RR5777	7278239
NS11014	23454302	HO5902	38239008	RR5902	8473546
NS11131	24043286	HO6655	20799972	RR6655	2208555
NS11145	23761002	HO6940	41567391	RR6940	7986081
NS11151	15458510	HO6971	19612761	RR6971	4747682
NS11152	23827415	HO7615	32600044	RR7615	9895073
NS11177	31888496	HO7744	36196064	RR7744	7994976
NS11650	1418862	HO7976	24735781	RR7976	6435826
NS11656	31519084	HO8035	26718056	RR8035	6687919
NS11689	62060908	HO820	27193350	RR820	7908664
NS11720	76463359	HO8291	28420478	RR8291	7377703
NS9858	36650572	HO8991	39493355	RR8991	10269362
NS11853	35850177	HO909	41164449	RR909	7175657
NS11991	19390852	HO910	30104158	RR910	6843822
NS11996	4426120	HO9208	26931414	RR9208	6044174
NS12023	7741434	HO9211	36455421	RR9211	5940937
NS12032	27106180	HO9216	27264175	RR9216	7351458
NS12062	43021389	HO9223	21560559	RR9223	4478137
NS12082	34415331	HO9581	32279509	RR9581	6603275
NS12083	19502059	HO9858	34836094	RR9858	7463227
NS2824	26105969	HO361	25033591	RR2824	7901956
NS361	23509112	HO77	25481389	RR361	6813105
NS77	19378235	HO974	33809502	RR77	6735144
NS974	39949669	HO9858	34836094	RR974	7048444

Table 1. Sequencing Depth Coverage

4.8.1 Impact of Sequencing Dept on Gut Microbiome Profiling

4.8.1.1 Taxonomic Profiling:

A total of 20 bacteria Phyla were identified. i) 31 million reads/average characterized 20 Phyla with Firmicutes (41%), Bacteroidetes (40.5%), and Proteobacteria (14.8%), representing the most abundant phyla. ii) 27.2 million reads/average outlined 20 with Firmicutes (40.4%), Bacteroidetes (40.2%), and Proteobacteria (15.7%) comprising the most abundant phyla. iii) 7.1 million read/average denoted 19 with Bacteroidetes (40.3%), Firmicutes (40%), and Proteobacteria (15.9%), representing the most abundant phyla.

Approximately 400 bacterial families with different relative abundance were characterized across all approaches among all seventy-two cecal metagenomic samples. i) 31 million reads/average depicted a total of 153 bacteria families (39.13%), with an average per sample of 80 taxa-family rank with Bacteroidaceae (29.6%), Lactobacillaceae (22.5%), Enterobacteriaceae (10.1%), Clostridiaceae (3%), Enterococcaceae (3.5%), and Campylobacteraceae (2.5%) representing the most abundant families. On the other hand, ii) 27.2 million reads/average denoted a total of 106 families (27.11%), with an average per sample of 41 taxa-family ranks, were Bacteroidaceae (30%), Lactobacillaceae (21.5%), Enterobacteriaceae (11.3%), Enterococcaceae (3.9%), Clostridiaceae (2.9%), and Campylobacteraceae (2.6%) were the most abundant families, while iii) 7.1 million read/average outlined a total of 132 families (33.76%), with an average per sample of 64 taxa-family ranks, with Bacteroidaceae (29%), Lactobacillaceae (22.3%), Enterobacteriaceae (11. %), Enterococcaceae (3.3%), Clostridiaceae (2.9%) and Campylobacteraceae (2.6%) representing the most abundant families. Figure 2

shows the number of bacteria families by sequencing depth of coverage samples and negative binomial analysis. Table 2 denotes family rank characterized by sequencing depth.

4.8.2 Descriptive Statistics:

Negative binomial regression shows no plateau or changing slope. It is linear and based on the regression, where depth of coverage does significantly predict Number of number of taxa profiled.

The one-way analysis of variance (ANOVA) was used to determine whether there are statistically significant differences between the means of three sequencing platform performance parameters to characterize animal microbiome at the family taxonomic level. There was a significant difference between the three averaged sequencing depths on the total microbiome at the family taxa level. Degrees of freedom (Df) = 2 , Sum of squares (Sum Sq) = 56977, mean of the sum of squares (Mean Sq) = 28488, test statistic from the F test (F value) = 249.6, p-value of the F-statistic (Pr(>F))= 2e-16. In pairwise comparisons, there are statistically significant differences from one another: ii) 27.2 million reads/average / i) 31 million reads/average 16.47222, iii) 7.1 million reads/average / i) 31 million reads/average 39.59722, iii) 7.1 million reads/average / ii) 27.2 million reads/average 23.12500.

Richness and diversity were calculated as average sequencing depths at family taxonomic rank and finding a significant difference between the three averaged sequencing depths approaches. Family taxonomic richness by i)31 million reads/average: Mean (μ) 27.8134, Standard deviation (SD) 27.15923. μ (Shannon) 2.023266, SD(Shannon) 1.237081. Family taxonomic richness by ii) 27.2 million reads/average: Mean (μ) 25.34973, Standard deviation (SD) 26.86572. μ (Shannon) 1.86202, SD (Shannon) 1.242713. Family taxonomic richness by iii) 7.1 million reads/average

sequencing: Mean (μ) 25.88696, Standard deviation (SD) 27.98112. μ (Shannon) 1.791294, SD(Shannon) 1.306915.

Two hundred and fifty-seven bacterial genera were characterized with i)31 million reads/average, 220 bacterial genera by ii) 27.2 million reads/average, and 164 by iii) 7.1 million reads/average respectably. Bacteroides 29.6%, Lactobacillus 22.5%, Escherichia 9.3%, Enterococcus 3.5%, Clostridium 2.9%, and Campylobacter 2.5 % were the most abundant characterized by i)31 million reads/average. Bacteroides 29.6%, Lactobacillus 22.3%, Escherichia 10%, Enterococcus 3.7%, Clostridium 2.9%, and Campylobacter 2.5 % were the most abundant genera characterized by ii) 27.2 million reads/average and Bacteroides 30%, Lactobacillus 21.5%, Escherichia 10.5%, Clostridium 2.9%, and Campylobacter 2.6% were the most abundant genus rank characterized by iii) 7.1 million reads/average.

We calculated a variety of diversity matrices per sequencing mode to establish the species richness characterized based on the averaged depth of coverage. i)31 million reads/average genera taxa level richness mean, and standard deviation were 21.0315 and 24.96111, respectably, while ii) 27.2 million reads/average, were mean was 20.25641 and standard deviation of 24.5660, compared to iii) 7.1 million reads/average mean was 20.06011 and standard deviation 24.96111 The Shannon-Wiener Diversity Index was calculated per sample per averaged depth of coverage to determine a significant difference between the three averaged depth of coverage. We found a significant difference between i)31 million reads/average, ii) 27.2 million reads/average, and iii) 7.1 million reads/average at genera taxonomic rank. i)31 million reads/average Shannon diversity mean 1.687921 and standard deviation 1.259115, compared to ii) 27.2 million reads/average were mean was 1.595077 and

standard deviation of 1.268 and iii) 7.1 million reads/average mean was 1.519076 and standard deviation of 1.311921.

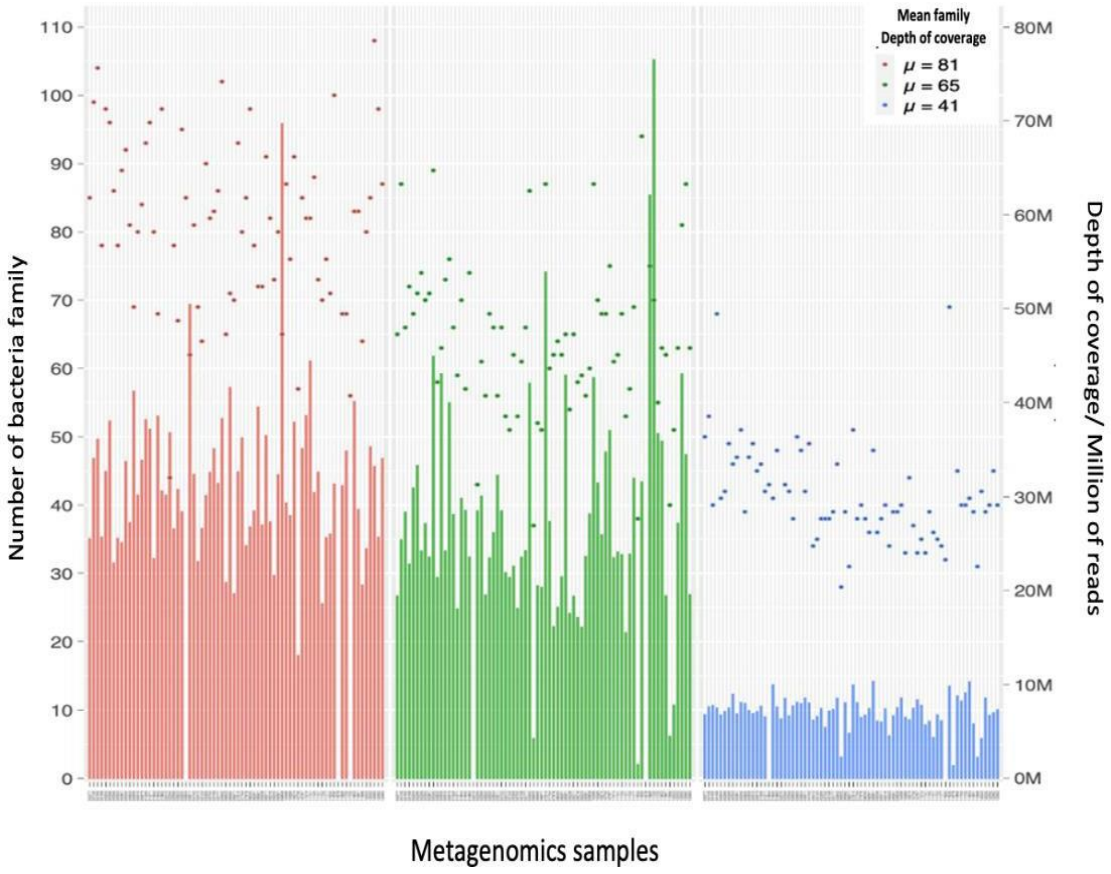


Figure 2A

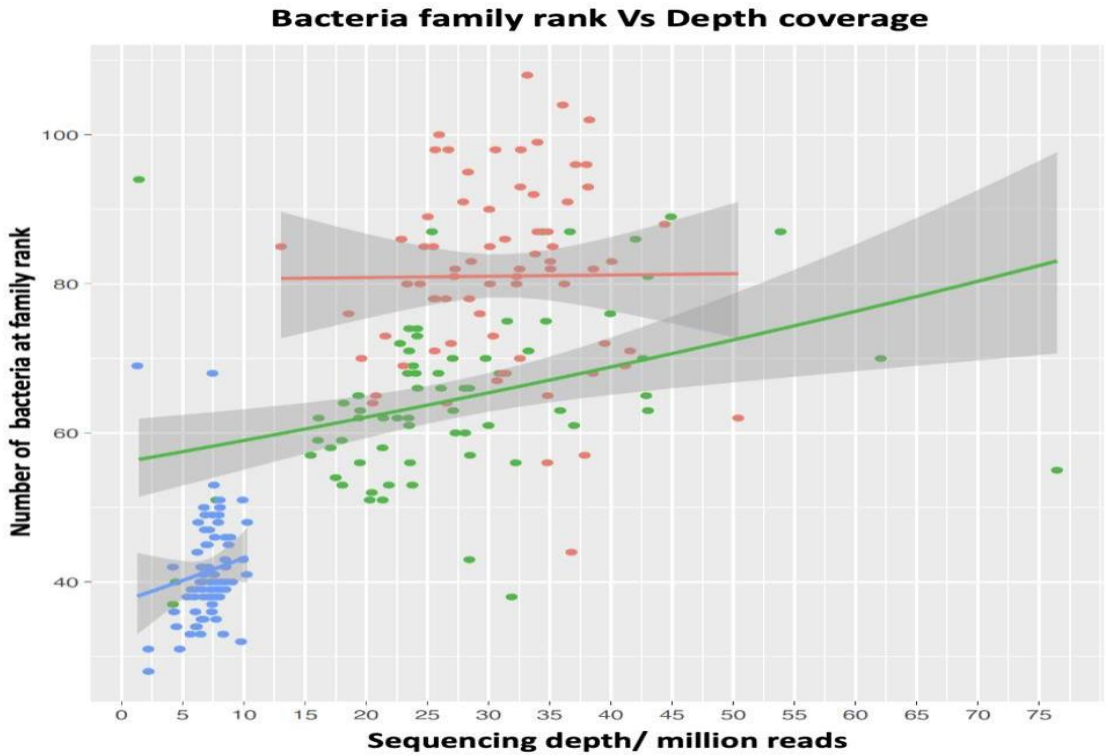


Figure 2A

Figure 2. [A] Bar chart showing the number of bacteria families by sequencing depth of coverage samples. [B] Negative binomial analysis showing number of bacteria family observed by sequencing depth coverage.

Total Families Identified in a Complex Cecal Microbiome				
		Taxa-family by Depth Coverage		
Source ID	Animal Source	31 million read averg	27.2 million read averg	7.1 million read averg
126	Swine	99	87	53
165	Turkey	104	66	40
168	Swine	78	72	68
204	Turkey	98	68	41
296	Turkey	96	71	42
299	Turkey	86	74	49
353	Turkey	78	70	46
540	Cattle	92	89	51
820	Turkey	81	58	39
909	Swine	69	63	47
910	Swine	80	73	49
1349	Turkey	93	66	46
1371	Turkey	96	59	42
1844	Swine	80	70	43
1845	Swine	68	57	41
1883	Turkey	98	74	48
2294	Swine	44	43	43
2599	Turkey	78	61	42
2622	Turkey	67	56	38
2650	Turkey	95	68	50
3611	Swine	62	56	42
4152	Swine	81	66	49
5012	Chicken	69	53	34
5124	Chicken	64	51	35
5399	Chicken	90	62	38
5702	Chicken	82	53	38
5748	Chicken	83	61	38
5777	Chicken	86	66	39
5902	Turkey	102	86	46
6655	Turkey	65	37	28
6940	Chicken	71	52	39
6971	Turkey	70	51	31
7615	Turkey	93	87	51
7744	Swine	80	60	38
7976	Turkey	85	62	40
8035	Chicken	98	64	38
8291	Chicken	78	62	36
8991	Swine	72	65	48
9208	Chicken	72	54	36
9211	Chicken	91	65	38
9216	Chicken	82	58	40
9223	Chicken	73	59	34
9581	Chicken	80	56	39
9858	Swine	65	60	39
10496	Chicken	87	87	40
10589	Chicken	76	70	33
10674	Swine	91	68	44
10756	Chicken	57	68	37
10787	Chicken	85	75	33
11014	Chicken	82	61	35
11131	Chicken	82	62	33
11145	Chicken	88	68	39
11151	Chicken	73	53	36
11152	Chicken	70	57	35
11177	Chicken	76	69	34
11650	Swine	71	38	32
11656	Swine	100	94	69
11689	Swine	68	75	45
11720	Swine	68	70	40
11725	Swine	56	55	40
11853	Cattle	83	63	41
11991	Turkey	83	62	39
11996	Turkey	64	40	31
12023	Turkey	80	51	42
12032	Cattle	85	63	39
12062	Turkey	108	81	40
12082	Turkey	98	87	45
12083	Turkey	87	63	40
2824	Cattle	85	66	48
361	Cattle	89	71	47
77	Cattle	85	65	50
974	Cattle	84	76	45

Table 2. Total number of families per metagenomic sample at different million reads.

4.8.3 Microbiome Analysis by Animal Source

4.8.3.1 Microbiome Characterization by Stablishing Average Sequencing Depth per Animal Source.

Further examination was made compare different animal gut complex microbiome. The seventy-two metagenomic samples were split by food animal source. 18 swine cecal metagenomics samples, 23 turkeys, 7 cattle, and 24 chicken samples were analyzed.

Swine metagenomics samples: Total millions of reads by HiSeq-HO was 618,675.739 with i) 34,370.874 reads average per sample. NextSeq was 59,4920.138 with ii)33,051.118 reads average HiSeq-RR was 140,348.933 with iii) 7,797.162 reads average per sample. ANOVA analysis shows a significant difference between the three-sequencing read average on swine microbiome. Df = 2, Sum Sq = 8.074e+15, Mean Sq = 4.037e+15, F value =40.29, Pr(>F) = 3.19e-11. Tukey's Honestly Significant Difference (Tukey's HSD) post-hoc test for pairwise comparisons helps to identify that there are statistically significant differences from one another: ii)33,051.118 reads average / i) 34,370.874 reads average: 1319756, iii) 7,797.162 reads average / i) 34,370.874 reads average: 26573711, iii) 7,797.162 reads average / ii)33,051.118 reads average: 25253956.

Turkey metagenomics samples: Total millions of reads by HiSeq-HO was 675,944.561 with i) 29,388.894 reads average per sample. NextSeq was 572.345.850 with ii) 24,884.602

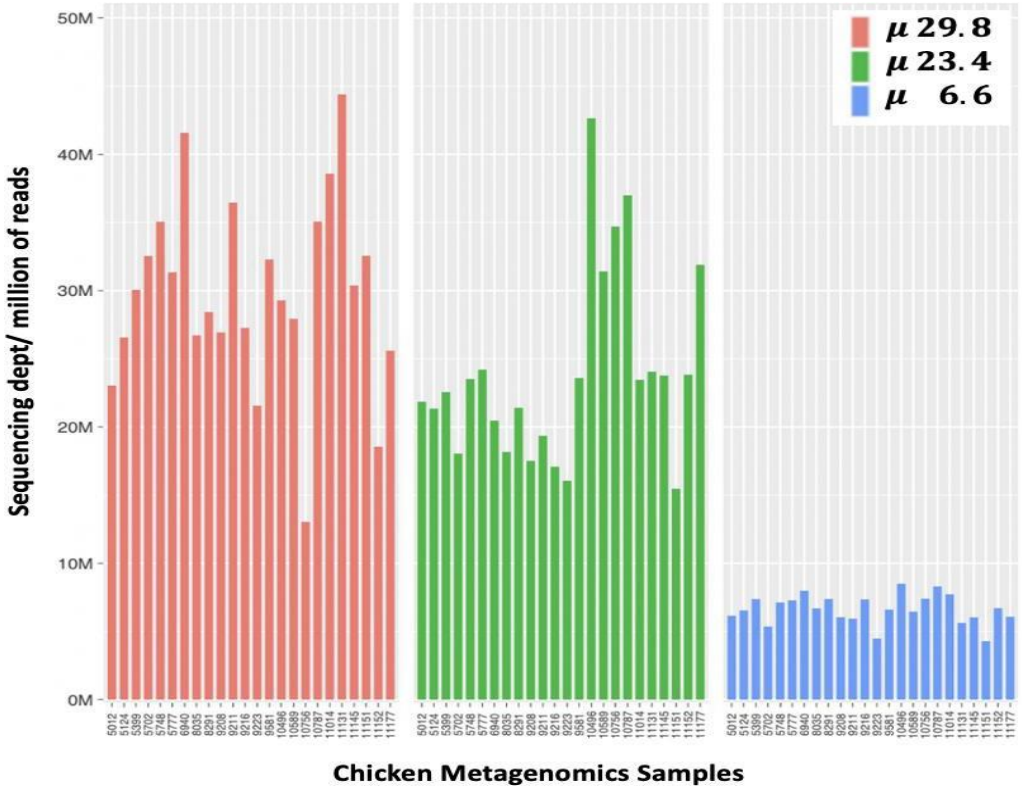
reads average per sample. HiSeq-RR was 154,444.833 million reads with iii) 6,714.992 reads average per sample.

Descriptive statistics shows a significant difference between the three-sequencing read average on turkey microbiome. Df = 2, Sum Sq = 6281e+15, Mean Sq = 3.3140e+15, F value = 55.312, Pr(>F) = 7.805e-15. Tukey's Honestly Significant Difference (Tukey's HSD) post-hoc test for pairwise comparisons helps to identify that there are statistically significant differences from one another: ii) 24,884.602 reads average / i) 29,388.894 reads average: 4504292, iii) 6,714.992 reads average / i) 29,388.894 reads average: 22673901, iii) 6,714.992 reads average / ii) 24,884.602 reads average: 18169609.

Cattle metagenomic samples: Total millions of reads by HiSeq-HO was 223,415.175 with i) 31,916.454 reads average per sample. NextSeq was 176,855.117 with ii) 25,265.016 reads average per sample. HiSeq-RR was 48,189.620 with iii) 6,884.231 reads average per sample. Descriptive statistics analysis shows a significant difference between the three-sequencing read average on cattle microbiome. Df = 2, Sum Sq =2.592e+15, Mean Sq =1.2960e+15, F value =32.731, Pr(>F) = 1.009e-06. Tukey's Honestly Significant Difference (Tukey's HSD) post-hoc test for pairwise comparisons helps to identify that there are statistically significant differences from one another: ii) 25,265.016 reads average / i) 31,916.454 reads average: 944341.3, iii) 6,884.231 reads average / i) 31,916.454 reads average: 24025301.6, iii) 6,884.231 reads average / ii) 25,265.016 reads average :23080960.3.

Chicken metagenomic samples: Total millions of reads by HiSeq-HO was 715,090.937 with i)29,795.456 reads average per sample. NextSeq was 573,197.788 with ii)23,883.241 reads average per sample. HiSeq-RR was 159,434.358 with iii)6,643.098 reads average per sample.

Descriptive statistics analysis shows a significant difference between the three-sequencing read average on chicken microbiome. Df = 2, Sum Sq =6.9457e+15, Mean Sq = 3.4728e+15, F value = 106.22 Pr(>F) = 2.2e-16. Tukey's Honestly Significant Difference (Tukey's HSD) post-hoc test for pairwise comparisons helps to identify that there are statistically significant differences from one another: ii)23,883.241 reads average / i)29,795.456 reads average: 5912215, iii)6,643.098 reads average / i)29,795.456 reads average: 23152357, iii)6,643.098 reads average / ii)23,883.241 reads average: 17240143. Figure 3 shows Sequencing Depth per Animal Source.



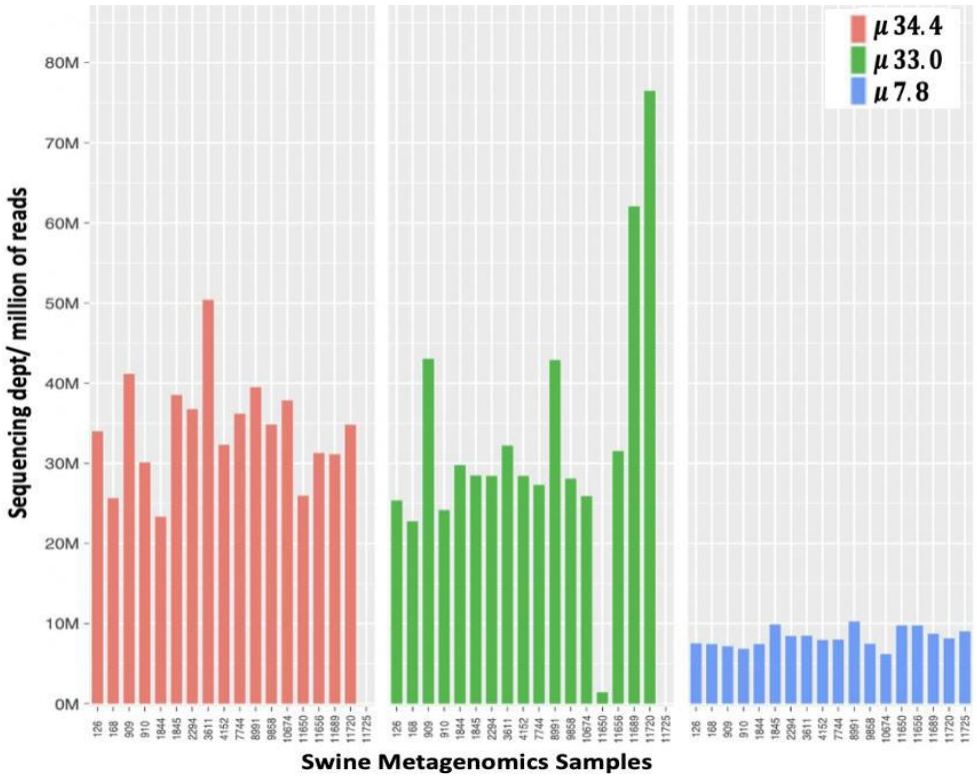


Figure 3 A

Figure 3 B.

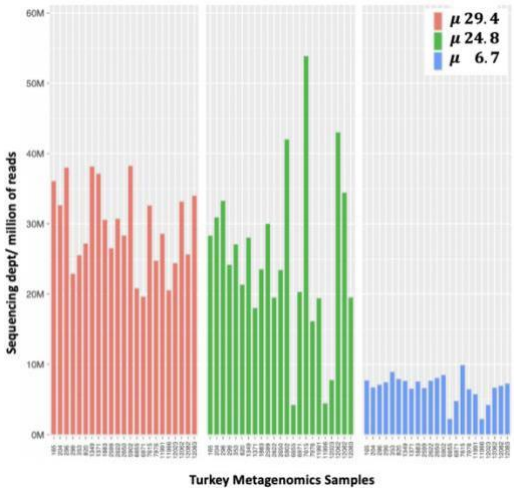


Figure 3 C

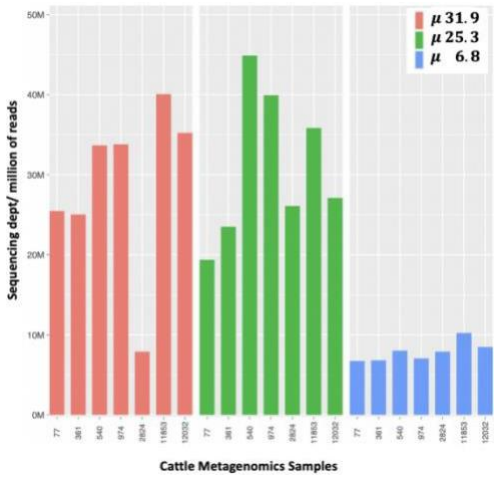


Figure 3 D

Figure 3. Bar chart Sequencing performance by animal animal gut complex microbiome. Chicken [A], swine [B], turkey [C], cattle [D].

4.8.4 Microbiota characterization

The number of sequencing reads varies from sample to sample by its complexity. We stablished three different sequencing read average to interrogate the taxonomic and resistome composition of the gut's animal dataset. i) 31,015,644 averaged read coverage, ii) 27,184,285 averaged read coverage, and iii) 7,057,574 averaged read coverage.

4.8.4.1 Taxonomic profiling at family rank:

Swine microbiome shows Lactobacillaceae, Prevotellaceae, Enterobacteriaceae, Desulfovibrionaceae, Desulfovibrionaceae, Peptostreptococcaceae, Bifidobacteriaceae, Ruminococcaceae, and Methanobacteriaceae as the most abundant families. At i) 31 million read averaged, 1772 bacteria families with 74 bacteria families on average per sample. 1332 bacteria families at ii) 27.2 million read averaged, with 61 bacteria families on average per

samples. 830 bacteria families at iii) 7.1 million read averaged with 46 bacteria families on average per sample.

Turkey microbiome shows higher abundance of Enterobacteriaceae, Bacteroidaceae, Lactobacillaceae, Clostridiales, Rikenellaceae, oriobacteriaceae, Ruminococcaceae, Veillonellaceae and Porphyromonadaceae. At i) 31 million read averaged, 2005 bacteria families with 87 bacteria families on average per sample. 498 bacteria families at ii) 27.2 million read averaged, with 65 bacteria families on average per samples. 956 bacteria families at iii) 7.1 million read averaged with 41 bacteria families on average per sample.

Cattle microbiomes have a large abundance of Enterobacteriaceae, Methanobacteriaceae, Bacteroidaceae, Prevotellaceae, Lachnospiraceae, Ruminococcaceae, Bifidobacteriaceae, Enterococcaceae, and Streptococcaceae. At i) 31 million read averaged, 603 bacteria families with 86 bacteria families on average per sample. 493 bacteria families at ii) 27.2 million read averaged, with 70 bacteria families on average per samples. 321 bacteria families at iii) 7.1 million read averaged with 46 bacteria families on average per sample.

Chicken microbiome reveals Lachnospiraceae, Lachnospiraceae, actobacillaceae, Porphyromonadaceae, Erysipelotrichaceae, Bacteroidaceae, Enterococcaceae, Enterobacteriaceae, Campylobacteraceae, and Clostridiaceae as the most abundant bacteria family. At i) 31 million read averaged, 1895 bacteria families with 79 bacteria families on average per sample. 1486 bacteria families at ii) 27.2 million read averaged, with 62 bacteria families on average per samples. 877 bacteria families at iii) 7.1 million read averaged with 37 bacteria families on average per sample.

Figure 4 represent the number of bacterial families characterized at different averaged read coverage.

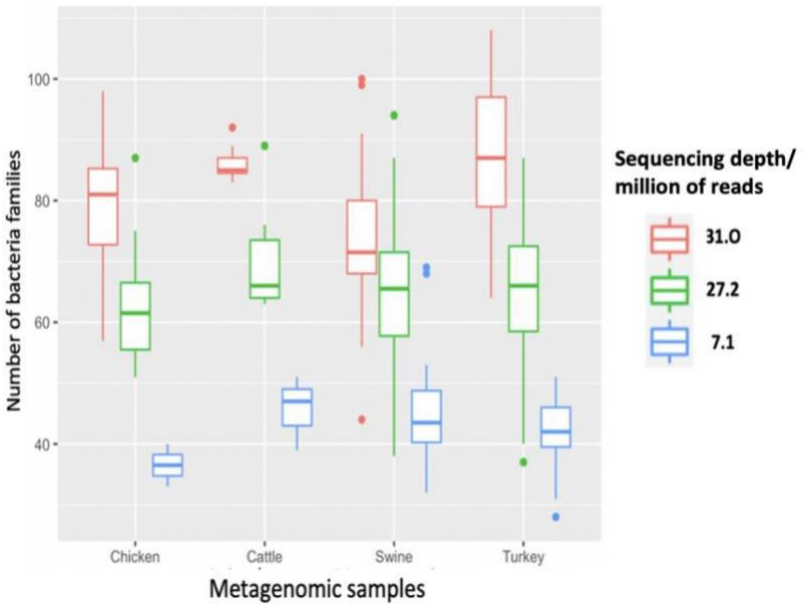


Figure 4. BoxPlot number of bacterial families characterized at different averaged read coverage.

Across all animal set, we were able to identify over 150 unique bacterial families, corresponding to four major bacterial phyla: Firmicutes, Proteobacteria, Bacteroidetes, and Actinobacteria. At an average sequencing depth of 31 million reads, 153 families were characterized; at an average sequencing depth of 27.2 million reads, 132 families were characterized; and at an average sequencing depth of 7.1 million reads, 106 families were characterized.

Overall, the data indicate that sequencing depth critically affects the inferred taxonomic diversity of complex microbiomes to provide more robust scientific findings that could avoid compromising the biological outcomes that can be achieved from an experimental design.

4.8.5 Impact of Sequencing Dept on Gut Microbiome Profiling

4.8.5.1 Resistome Profiling

Our finding reveals a significant difference in the number of AMR genes per cecal sample by the three different established averaged sequencing depth i) 31 million *reads per kilobase of transcript per million reads mapped (RPKM*) average, ii) 27.2 million RPKM average, and iii) 7.1 million RPKM average.

Negative binomial shows no plateau or changing slope. It is linear and based on the regression, where depth of coverage in AMR-RPKM does significantly predict Number of numbers of AMR genes profiled.

The total number of *reads per kilobase of transcript per million reads mapped (RPKM*) average reads varied across food animal sources within the data set. We characterized a full resistome composition of the data set by established averaged sequencing depth. At i) 31 million AMR-RPKM we characterized a total of 1651 AMR genes with an average of 23 AMR genes per cecal sample, ii) 27.2 million AMR-RPKM we identified a total of 1429 AMR genes with an average of 20 AMR genes per cecal sample, and iii) 7.1 million RPKM 1893 AMR-RPKM were identified with an average of 26 AMR genes per cecal sample. Table 3 denotes the number of genes found by source. Figure 5 show the number of antimicrobial resistance genes by sequencing depth of coverage. It is expected to have all these gene coming from different bacteria taxa. Therefore, counted more than one time as a global snapshot of total resistome from the sample set. Descriptive statistic was used to determine whether there are statistically significant differences between the means of three sequencing platform performance parameters to characterize resistome composition on a complex microbiome. We found a significant difference between i) 31 million AMR-RPKM, 27.2 million AMR-RPKM and iii) 7.1 million AMR-RPKM to characterize resistome composition.

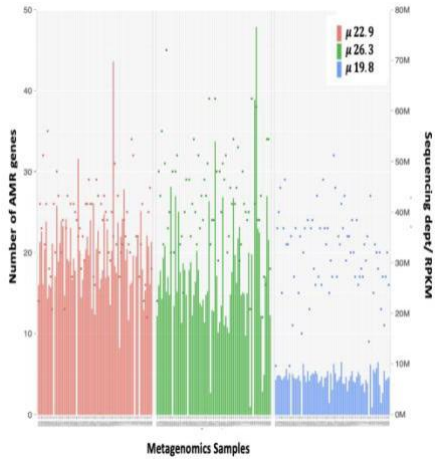


Figure 5A

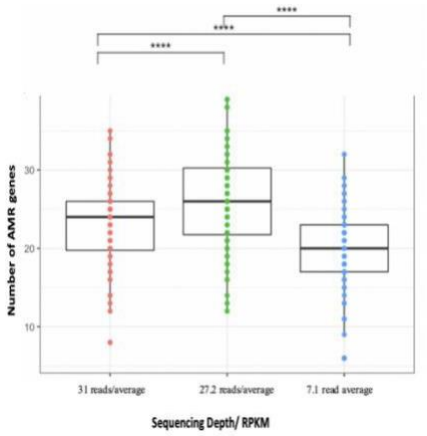


Figure 5B

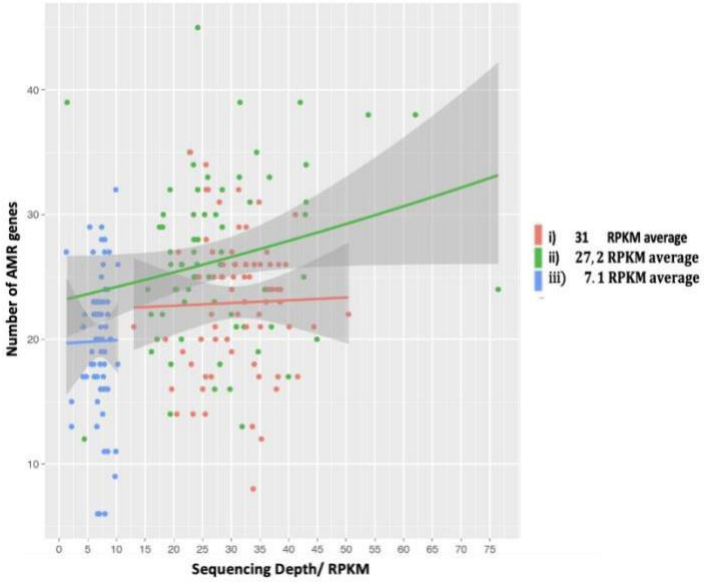


Figure 5C

Figure 5. [A]. Bar chart showing the number of Antimicrobial Resistance Genes by sequencing depth of coverage. [B]. Box Plot Resistome profiling showing total number of AMR genes distribution by sequencing depth of coverage. Antimicrobial Resistance Genes comparison by sequencing depth of coverage: 31 RPKM averaged: 23, 27,2 RPKM averaged: 26, 7.1 RPKM averaged: 19. Standard deviation: 31 RPKM averaged: 5.4, 27,2 RPKM averaged: 6.8, 26, 7.1 RPKM averaged: 5.5. [C]. Negative binomial model, Number of antimicrobial genes Vs Sequencing depth of coverage.

Resistome profiling on a complex cecal microbiome				
		Number of Genes by Depth Coverage		
Source ID	AnimalSource	31 million read averg	27.2 million read averg	7.1 million read averg
126	Swine	26	30	23
165	Turkey	23	27	26
168	Swine	32	35	28
204	Turkey	21	21	15
296	Turkey	24	31	23
299	Turkey	35	45	29
353	Turkey	17	23	20
540	Cattle	13	20	6
820	Turkey	21	20	11
909	Swine	30	31	18
910	Swine	27	32	23
1349	Turkey	17	18	14
1371	Turkey	24	22	22
1844	Swine	14	16	11
1845	Swine	23	25	20
1883	Turkey	26	28	23
2294	Swine	26	29	24
2599	Turkey	17	22	13
2622	Turkey	26	26	23
2650	Turkey	25	34	24
3611	Swine	22	21	16
4152	Swine	23	32	27
5012	Chicken	18	23	17
5124	Chicken	22	26	23
5399	Chicken	19	24	16
5702	Chicken	26	29	29
5748	Chicken	26	27	22
5777	Chicken	29	28	21
5902	Turkey	26	39	23
6655	Turkey	27	21	15
6940	Chicken	17	24	19
6971	Turkey	16	24	17
7615	Turkey	29	38	32
7744	Swine	27	30	28
7976	Turkey	20	19	17
8035	Chicken	27	30	22
8291	Chicken	25	26	22
8991	Swine	26	30	26
9208	Chicken	25	29	23
9211	Chicken	24	32	27
9216	Chicken	20	20	19
9223	Chicken	19	22	22
9581	Chicken	25	28	22
9858	Swine	22	26	20
10496	Chicken	31	33	22
10589	Chicken	20	25	20
10674	Swine	31	33	22
10756	Chicken	16	33	25
10787	Chicken	21	19	16
11014	Chicken	21	21	17
11131	Chicken	24	30	18
11145	Chicken	21	26	18
11151	Chicken	25	25	21
11152	Chicken	26	24	20
11177	Chicken	20	25	22
11650	Swine	34	13	9
11656	Swine	32	39	27
11689	Swine	32	39	27
11720	Swine	22	38	21
11725	Swine	17	24	20
11853	Cattle	21	24	18
11991	Turkey	26	27	19
11996	Turkey	14	12	13
12023	Turkey	25	16	17
12032	Cattle	12	16	11
12062	Turkey	25	34	17
12082	Turkey	28	35	20
12083	Turkey	18	18	16
2824	Cattle	24	25	18
361	Cattle	16	25	22
77	Cattle	14	14	6
974	Cattle	8	17	6

Table 3. Total number of AMR genes per metagenomics at different RPKM

Getting deeper into the analysis, to know the unique AMR genes characterized amount the total microbiome.

i) 31 million AMR RPKM profiled 104 unique AMR genes representing 11 AMR classes, ii) 27.2 million AMR-RPKM profiled 119 AMR genes corresponding to 11 AMR classes, and iii) 7.1 million AMR-RPKM profiled 94 AMR genes representing 11 AMR classes.

Our results show exciting findings on resistome composition where some AMR genes could be characterized by one or two of the established averaged sequencing depth i) 31 million *reads per kilobase of transcript per million reads mapped (RPKM*) average, ii) 27.2 million AMR-RPKM average, and iii) 7.1 million reads AMR-RPKM average.

At i) 31 million *reads* AMR-RPKM characterized: Aminoglycoside 23.1%, Macrolide 19.8%, Beta-lactamase 17.3%, and Tetracycline 14.8 %, represent the most abundant AMR classes. ii) 27.2 million reads AMR-RPKM: Aminoglycoside 16.7%, Macrolide 13.6%, Beta-lactamase 10.5%, and Tetracycline 12.4 % were found the most abundant AMR classes.

iii) 7.1 million reads AMR-RPKM: Aminoglycoside 23.2%, Macrolide 22%, Beta-lactamase 14.6%, and Tetracycline 17.2 %, illustrate the most abundant AMR classes.

Tet(Q) (11902.961160269 AMR-RPKM), *lnu(C)* (9446.753859174 AMR-RPKM), ap'(3')-III 3693.318764671 AMR-RPKM), *erm(B)* (2011.995263679 AMR-RPKM), *cfxA6* (522.508905158 AMR-RPKM) were the most abundant AMR genes characterized by i) 31 million *reads* AMR-RPKM. *tet(Q)* (11455.9826 AMR-RPKM), *lnu(C)* (9394.46147 AMR-RPKM), *ap'(3')-III* (3867.10641 AMR-RPKM), *erm(B)* (2007.26762 AMR-RPKM), and *cfxA6* (491.264347 AMR-RPKM) were identified by. ii) 27.2 million reads AMR-RPKM, while *tet(W)* (47140.50618 AMR-RPKM), *lnu(C)* (16723.24732 AMR-RPKM), *ap'(3')-III* (5355.36847 AMR-RPKM), *erm(B)* (1531.301673 AMR-RPKM), *cfxA6* (561.5914912 AMR-RPKM), *VanX-D* (116.8729619 AMR RPKM) were the most abundant AMR genes characterized by iii) 7.1 million reads AMR RPKM, .

In contrast, *QnrB41* (26.87730157 AMR-RPKM), *QnrA1* (1.801772390 AMR-RPKM), *QnrS1*(0.646768596 AMR-RPKM), *oqxB* (1.31356698 AMR-RPKM), *blaOXA-24* (0.177315866 AMR-RPKM), *blaOXA-51* (0.464433238 AMR-RPKM), and *nblaCTX-M-110* (0.28859773 AMR-RPKM) were some of the least abundant genes among the three established sequence depth of coverage.

4.8.6 Resistome characterization by animal source

Swine resistome characterization shows. i) 31 million *reads* AMR-RPKM, 466 AMR genes with an average of 26 AMR genes per cecal sample, ii) 27.2 million reads AMR-RPKM identifying 523 AMR genes with an estimated average of 29 AMR genes per cecal sample, iii) 7.1 million reads AMR-RPKM, characterized 390 AMR genes with an average of 22 AMR genes per cecal sample. Descriptive analysis shows a significant difference between the three sequencing dept on Swine cecal samples Df = 2, Sum Sq = 494.7, Mean Sq = 247.35, F value = 6.554, Pr(>F) = 0.00293. Tukey's Honestly Significant Difference (Tukey's HSD) post-hoc test determines that there are statistically significant differences from one another: ii) 27.2 million reads AMR-RPKM / 31 million *reads* AMR-RPKM:3.166667, iii) 7.1 million reads AMR-RPKM, i) 31 million *reads* AMR-RPKM:4.222222, iii) 7.1 million reads AMR-RPKM:7.388889.

Cattle resistome characterization reveal 108 AMR genes with an average of 15 AMR genes per cecal sample at 31 million *reads* AMR-RPKM, 141 AMR genes with an estimate average of 20 AMR genes per cattle sample at ii) 27.2 million reads AMR-RPKM, and 87 AMR genes with 12 AMR genes average per cecal sample at iii) 7.1 million reads AMR-RPKM. Descriptive statistical analysis shows a significant difference between the three approaches on Cattle cecal samples Df = 2, Sum Sq = 211.7, Mean Sq = 105.86, F value = 3.25, Pr(>F) = 0.0624.

Tukey's Honestly Significant Difference (Tukey's HSD) post-hoc test determines that there are statistically significant differences from one another: AMR-RPKM, 141 AMR genes with an estimate average of 20 AMR genes per cattle sample at ii) 27.2 million reads AMR-RPKM / 31 million *reads* AMR-RPKM:4.714286, iii) 7.1 million reads AMR-RPKM /31 million *reads* AMR RPKM :3.000000, iii) 7.1 million reads AMR RPKM 7.714286.

Turkey at 31 million *reads* AMR-RPKM shows 530 AMR genes with an average of 23 AMR genes per cecal sample ,while resistome total characterization of 600 AMR genes and an average of 26 AMR genes per ii) 27.2 million reads AMR-RPKM compare to iii) 7.1 million reads AMR-RPKM characterized 499 AMR genes with 19 AMR genes average per turkey descriptive analysis shows a significant difference between the three approaches on Turkey cecal samples Df = 2, Sum Sq = 496.6, Mean Sq = 248.28, F value = 6.072, Pr(>F) = 0.0038. Tukey's Honestly Significant Difference (Tukey's HSD) post-hoc test determines that there are statistically significant differences from one another: ii) 27.2 million reads AMR-RPKM / 31 million *reads* AMR-RPKM: 3.043478, iii) 7.1 million reads AMR RPKM c/31 million *reads* AMR RPKM :3.521739, iii) 7.1 million reads AMR RPKM :6.565217.

Chicken resistome analysis exhibit 31 million *reads* AMR-RPKM shows 547 AMR genes with an average of 22 AMR genes per cecal sample, ii) 27.2 million reads AMR-RPKM with total characterization of 629 AMR genes and an average of 26 AMR genes per cecal sample. iii) 7.1 million reads AMR RPKM characterized 503 AMR genes with 21 AMR genes average per chicken.

Descriptive statistical analysis showed a significant difference between the three approaches on Chicken cecal samples Df = 2, Sum Sq = 340.8, Mean Sq = 170.4, F value = 12.71, Pr(>F) = 2e-05.

Tukey's Honestly Significant Difference (Tukey's HSD) post-hoc test showed that they were significantly different from one another: ii) 27.2 million reads AMR-RPKM / 31 million *reads* AMR-RPKM 3.416667, iii) 7.1 million reads AMR-RPKM /31 million *reads* AMR-RPKM :1.833333, iii) 7.1 million reads AMR-RPKM:5.250000.

There were some significant differences between the 31 million *reads* AMR-RPKM, ii) 27.2 million reads AMR-RPKM,and iii) 7.1 million reads AMR-RPKM, on a one-to-one sequencing approach comparison by cecal sample. Below are a few examples of resistome characterization on food animal cecal samples.

Resitome profiling by individual samples at different sequencing dept show that while most of the genes were characterized by all sequencing depth coverages, we established, some genes were characterized by one or two sequencing deep only.

Here, we described a few AMR genes profiling at different sequencing dept of coverage.

Swine cecal sample identification (ID) 168 shows a total of 45 AMR genes, representing 11 AMR classes. At 31 million *reads* AMR-RPKM 32 AMR genes were characterized ii) 27.2 million reads AMR-RPKM 35 AMR genes, and at iii) 7.1 million reads AMR-RPKM 28 AMR genes. Some of the total 45 genes were characterized with all sequencing dept approaches but no all.

Aminoglycoside *ant(6)-Ia,* shows 3.980962-RPKM at ii) 27.2 million reads AMR-RPKM , at At 31 million *reads* AMR-RPKM (4.565389-RPKM) , and at iii) 7.1 million reads AMR-RPKM (7.807734-RPKM).

Beta-lactamase *blaCMY-18, at* ii) 27.2 million reads AMR-RPKM (1.894273-RPKM), 31 million *reads* AMR-RPKM (1.344413-RPKM), and at iii) 7.1 million reads AMR-RPKM, Beta-lactamase *blaCMY-18* was no delectated.

Macrolide *erm(B),* at ii) 27.2 million reads AMR-RPKM (3.8891-RPKM), 31 million *reads* AMR RPKM (4.140316-RPKM), and iii) 7.1 million reads AMR-RPKM (1.102345-RPKM).

Lincosamida *lnu(B),* neither at ii) 27.2 million reads AMR-RPKM or iii) 7.1 million reads AMR RPKM was detected while at 31 million *reads* AMR-RPKM (0.686153-RPKM).

Tetracycline *tet(W),* ii) 27.2 million reads AMR-RPKM (60.4273-RPKM), 31 million *reads* AMR-RPKM (0), and iii) 7.1 million reads AMR-RPKM (277.3775-RPKM).

Quinolone *QnrS1*, at 31 million *reads* AMR-RPKM (0.891911), and ii) 27.2 million reads AMR-RPKM (0.595071-RPKM) while detected at iii) 7.1 million reads AMR-RPKM was not

Quinolone *QnrB41*, ii) 27.2 million reads AMR-RPKM (2.746696-RPKM), while at 31 million *reads* AMR-RPKM, and iii) 7.1 million reads AMR-RPKM was not detected.

Swine cecal sample ID 3611: shows 35 AMR genes, representing 11 AMR classes. 31 million *reads* AMR-RPKM (24 AMR genes), ii) 27.2 million reads AMR-RPKM (31 AMR genes), and iii) 7.1 million reads AMR-RPKM (23 AMR genes).

Aminoglycoside *aa'(6')-Isa,* ii) 27.2 million reads AMR-RPKM (3.503891-RPKM), 31 million *reads* AMR RPKM (3.542377-RPKM) , and iii) 7.1 million reads AMR-RPKM (3.030931-RPKM).

Beta-lactamase *cfxA6,* ii) 27.2 million reads AMR-RPKM (11.44417-RPKM), 31 million *reads* AMR-RPKM (11.00307-RPKM), and iii) 7.1 million reads AMR-RPKM (13.85356-RPKM).

Macrolide *erm(O),* ii) 27.2 million reads AMR-RPKM (0.993331-RPKM), 31 million *reads* AMR-RPKM (0.158711-RPKM), and iii) 7.1 million reads AMR-RPKM (2.826126-RPKM)

Lincosamida *(C),* ii) 27.2 million reads AMR-RPKM (24.45504-RPKM), 31 million *reads* AMR-RPKM (22.69384-RPKM), iii) 7.1 million reads AMR-RPKM (36.48929-RPKM)

Turkey cecal sample ID 296: shows 28 AMR genes, representing 6 AMR classes. ii) 27.2 million reads AMR-RPKM (21 AMR genes), 31 million *reads* AMR RPKM (22 AMR genes), and iii) 7.1 million reads AMR-RPKM (16 AMR genes).

Aminoglycoside *aph(3')-III,* ii) 27.2 million reads AMR-RPKM (41.34182-RPKM), 31 million *reads* AMR-RPKM (38.79476-RPKM) , and iii) 7.1 million reads AMR-RPKM (53.5971-RPKM) respectively.

Beta-lactamase *blaOXA-347*, ii) 27.2 million reads AMR-RPKM (1.869301-RPKM), 31 million *reads* AMR-RPKM (0.469897-RPKM), and iii) 7.1 million reads AMR-RPKM (4.1239-RPKM).

Macrolide *mef(A),* ii) 27.2 million reads AMR-RPKM (0.893089-RPKM), 31 million *reads* AMR-RPKM (0.521074-RPKM), while iii) 7.1 million reads AMR-RPKM did not detected Macrolite *mef(A)*.

Lincosamida *lnu(C)),* ii) 27.2 million reads AMR-RPKM (142.504-RPKM), i)31 million *reads* AMR-RPKM (177.0959-RPKM), iii) 7.1 million reads AMR-RPKM (256.9082-RPKM).

Tetracycline *tet(W),* ii) 27.2 million reads AMR-RPKM (92.08121-RPKM), i) 31 million reads AMR-RPKM (0), and iii) 7.1 million reads AMR-RPKM (608.619-RPKM).

Quinolone *QnrA1*, ii) 27.2 million reads AMR-RPKM (2.886464-RPKM), i)31 (0.842056-RPKM), and iii) 7.1 million reads AMR-RPKM 4.516129-RPKM).

Fosfomycin *fosA,* ii) 27.2 million reads AMR-RPKM (2.645925-RPKM), i)31 million *reads* AMR-RPKM (2.631424-RPKM), and iii) 7.1 million reads AMR-RPKM (1.129032-RPKM) respectively.

Turkey cecal sample ID 299: shows 50 AMR genes, representing 11 AMR classes. ii) 27.2 million reads AMR-RPKM (45 AMR genes), i)31 million *reads* AMR-RPKM (35 AMR genes), and iii) 7.1 million reads AMR-RPKM (29 AMR genes).

Aminoglycoside *strB,* ii) 27.2 million reads AMR-RPKM (1.881261-RPKM), i)31 million *reads* AMR-RPKM (2.447755-RPKM), and iii) 7.1 million reads AMR-RPKM (4.301468-RPKM) respectively.

Beta-lactamase *blaOXA-51*, ii) 27.2 million reads AMR-RPKM (0), i)31 million *reads* AMR-RPKM (0.17484-RPKM), while at iii) 7.1 million reads AMR-RPKM was dot detected.

Macrolide *msr(D),* ii) 27.2 million reads AMR-RPKM (4.356606-RPKM), i)31 million *reads* AMR-RPKM was not detected, and at iii) 7.1 million reads AMR-RPKM (21.22435-RPKM).

Lincosamine *lnu(C),* ii) 27.2 million reads AMR-RPKM (93.99193-RPKM), i) 31 million *reads* AMR-RPKM (81.22146-RPKM), iii) 7.1 million reads AMR-RPKM (178.0512-RPKM.

Tetracycline *tet(B),* ii) 27.2 million reads AMR-RPKM (2.483265-RPKM), i) 31 million *reads* AMR-RPKM (3.496793-RPKM), and iii) 7.1 million reads AMR-RPKM (1.893252-RPKM).

Sulphonamide sul*2,* ii) 27.2 million reads AMR-RPKM (3.703457-RPKM), i)31 million *reads* AMR-RPKM (1.68115-RPKM), and iii) 7.1 million reads AMR-RPKM (6.093255-RPKM) AMR-RPKM.

Fosfomycin *fosA*, ii) 27.2 million reads AMR-RPKM (4.138775), i)31 million *reads* AMR-RPKM (4.196152-RPKM), and iii) 7.1 million reads AMR-RPKM (1.075367-RPKM) respectively.

Cattle cecal sample ID 11853: shows 27 AMR genes, representing 8 AMR classes. ii) 27.2 million reads AMR-RPKM (24 AMR genes), i) 31 million reads AMR-RPKM (21 AMR genes), and iii) 7.1 million reads AMR-RPKM (18 AMR genes).

Aminoglycoside *ant(6)-Ia,* ii) 27.2 million reads AMR-RPKM (15.41187-RPKM), i) 31 million *reads* AMR-RPKM (14.64695-RPKM) , and iii) 7.1 million reads AMR-RPKM (15.06772-RPKM).

Beta-lactamase *cfxA6,* ii) 27.2 million reads AMR-RPKM (28.4283-RPKM), i)31 million *reads* AMR-RPKM (31.66042-RPKM), and iii) 7.1 million reads AMR-RPKM (30.74077-RPKM).

Macrolide *erm(G),* N ii) 27.2 million reads AMR-RPKM (9.099918-RPKM), i) 31 million *reads* AMR-RPKM (5.509327-RPKM), and iii) 7.1 million reads AMR-RPKM (10.12491-RPKM) Lincosamide *lnu(C),* ii) 27.2 million reads AMR-RPKM (42.5147-RPKM), i) 31 million *reads* AMR-RPKM (40.02426-RPKM), iii) 7.1 million reads AMR-RPKM (85.64203-RPKM). Tetracycline *tet(B),* ii) 27.2 million reads AMR-RPKM (2.483265-RPKM), i) 31 million *reads* AMR-RPKM (3.496793-RPKM), and iii) 7.1 million reads AMR-RPKM (1.893252-RPKM) respectively. Vancomycin *VanX-D*, ii) 27.2 million reads AMR-RPKM (2.231509-RPKM), i) 31 million *reads* AMR-RPKM (1.996233-RPKM), and iii) 7.1 million reads AMR-RPKM was no detected. Fosfomycin *fosA,* ii) 27.2 million reads AMR-RPKM and iii) 7.1 million reads AMR-RPKM, was no detected, while at i) 31 million *reads* AMR-RPKM (0.399247-RPKM).

Cattle cecal sample ID 540: shows 25AMR genes, representing 8 AMR classes. ii) 27.2 million reads AMR-RPKM (20 AMR genes), i) 31 million *reads* AMR-RPKM (13 AMR genes), and iii) 7.1 million reads AMR-RPKM (6 AMR genes).

Aminoglycoside *strA*, ii) 27.2 million reads AMR-RPKM (0.095987-RPKM), i) 31 million *reads* AMR-RPKM (0.250156-RPKM), while at iii) 7.1 million reads AMR-RPKM it was not detected. Beta-lactamase *blaSHV-12*, ii) 27.2 million reads AMR-RPKM (0.089076-RPKM), i) 31 million *reads* AMR-RPKM, and iii) 7.1 million reads AMR-RPKM were not detected.

Macrolide *mph(B),* ii) 27.2 million reads AMR-RPKM (0.089076-RPKM), while i) 31 million *reads* AMR-RPKM, and iii) 7.1 million reads AMR-RPKM were not detected.

Lincosamida *lnu(C),* ii) 27.2 million reads AMR-RPKM (9.871349-RPKM), while i) 31 million *reads* AMR-RPKM (8.795588-RPKM), iii) 7.1 million reads AMR-RPKM (21.59952-RPKM).

Tetracycline *tet(W),* ii) 27.2 million reads AMR-RPKM (2.783627-RPKM), while at i) 31 million *reads* AMR-RPKM did not detected, and iii) 7.1 million reads AMR-RPKM (15.57784-RPKM). Vancomycin *VanX-D*, ii) 27.2 million reads AMR-RPKM (0.712608-RPKM), while i) 31 million *reads* AMR-RPKM (0.237543-RPKM), and iii) 7.1 million reads AMR-RPKM (1.993963-RPKM) respectively.

Chicken cecal sample ID 10496: denotates 39 AMR genes, representing 9 AMR classes. ii) 27.2 million reads AMR-RPKM (33 AMR genes), while i) 31 million *reads* AMR-RPKM (31 AMR genes), and iii) 7.1 million reads AMR-RPKM (22 AMR genes).

Aminoglycoside *aph(3')-III*, ii) 27.2 million reads AMR-RPKM (49.21463-RPKM), while at i) 31 million *reads* AMR-RPKM (51.06577-RPKM) , and iii) 7.1 million reads AMR-RPKM (103.4911-RPKM) respectively.

Beta-lactamase *blaFOX-1* ii) 27.2 million reads AMR-RPKM (0.941448-RPKM), while i) 31 million *reads* AMR-RPKM (0.136617-RPKM), and iii) 7.1 million reads AMR-RPKM (0.941448-RPKM).

Macrolide *erm(E),* at ii) 27.2 million reads AMR-RPKM was not detected, while at i) 31 million *reads* AMR-RPKM (0.273234-RPKM), and iii) 7.1 million reads AMR-RPKM did not detected.

Lincosamide *lnu(C),* ii) 27.2 million reads AMR-RPKM (565.1686-RPKM), at i) 31 million *reads* AMR-RPKM (0546.9706-RPKM), and iii) 7.1 million reads AMR-RPKM (976.3077-RPKM).

Tetracycline *tet(W),* ii) 27.2 million reads AMR-RPKM (60.10938-RPKM), while i) 31 million *reads* AMR-RPKM, did not detected, while iii) 7.1 million reads AMR-RPKM (485.4343-RPKM) respectively.

Chicken cecal sample ID 10787: shows 26 AMR genes, representing 9 AMR classes. ii) 27.2 million reads AMR-RPKM (19 AMR genes), while 31 million *reads* AMR-RPKM (21 AMR genes), and iii) 7.1 million reads AMR RPKM (16 AMR genes).

Aminoglycoside *aph(2")-Id*, ii) 27.2 million reads AMR-RPKM (1.423196-RPKM), while i) 31 million *reads* AMR-RPKM (0.718987-RPKM) , and iii) 7.1 million reads AMR-RPKM (1.09461-RPKM) respectively.

Beta-lactamase *blaFOX-1*, ii) 27.2 million reads AMR-RPKM (0.540814-RPKM), while i) 31 million *reads* AMR-RPKM and iii) 7.1 million reads AMR-RPKM, were not able to detect.

Macrolide *erm(G),* ii) 27.2 million reads AMR-RPKM (5.840503-RPKM), while at i) 31 million *reads* AMR-RPKM (4.30984-RPKM), and iii) 7.1 million reads AMR-RPKM (13.44579-RPKM).

Lincosamida *lnu(C),* ii) 27.2 million reads AMR-RPKM (196.4145-RPKM), while i) 31 million *reads* AMR-RPKM (194.2463-RPKM), and iii) 7.1 million reads AMR-RPKM (369.4424-RPKM) respectably.

Tetracycline tet(32), at ii) 27.2 million reads AMR-RPKM and i) 7.1 million reads AMR-RPKM, was not detected while at i) 31 million *reads* AMR-RPKM was (0.582037-RPKM).

Figure 6 shows Number of AMR genes per animal source and Figure 7. Bar chart. Number of antimicrobial resistance per source (chicken [A], swine [B], turkey [C], cattle [D]) per depth of coverage.

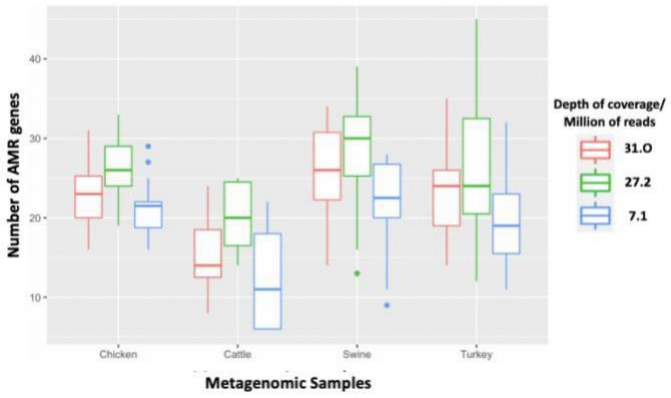


Figure 6. Number of AMR genes per animal source (chicken, swine, turkey, cattle) per sequencing depth of coverage.

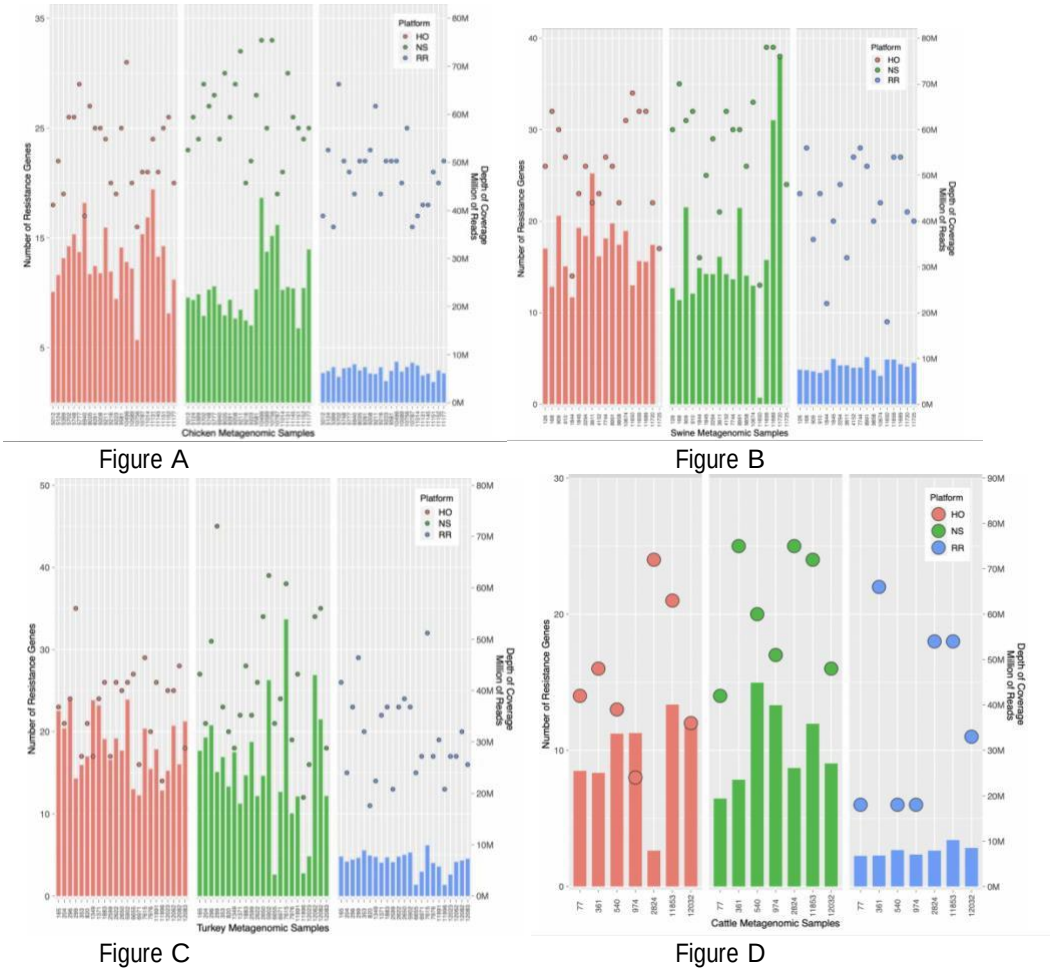


Figure A

Figure B

Figure C

Figure D

Figure 7. Bar chart. Number of antimicrobial resistance per source (chicken [A], swine [B], turkey [C], cattle [D]) per depth of coverage

4.9 Discussion:

Antimicrobial resistance is a natural phenomenon that can be accelerated by the use and or misused of antimicrobials in both humans and animals. The emergence of resistant pathogenic bacteria represents a serious public health challenge worldwide that causes approximately 700,000 deaths annually. AMR is estimated to account for 10 million deaths and US $100 trillion economic loss every year by 2050 if no urgent actions are needed taken [11]. Even though antimicrobial resistance represents a public health threat, there are still many unsolved questions regarding bacterial composition, resistome characterization, and new AMR reservoirs in complex and poorly understood microbiomes, especially for those bacteria unable to grow using current laboratory culturing techniques [162] . Microbial communities present in the gastrointestinal tract of food animals (gut microbiome) are known as the source of novel genes for biotechnology applications [163], as well as a prime reservoir for AMR genes [156]. Metagenomic shotgun sequencing approaches provide a comprehensive microbial community profile, with datasets yielding millions of short reads to help understand the gut microbiome, including the bacterial diversity, identifying of novel genes, and potential new AMR gene reservoirs. However, sequencing approaches and costs often limit the number of sequences reads that can be experimentally obtained, consequently limiting the biological findings that can be achieved from an experimental design. We employed depth shotgun metagenomics using different Illumina sequencers, HiSeq High Output chemistry, HiSeq Rapid Run chemistry, and NextSeq, establishing three different averaged sequencing depth coverage: i) 31 million *reads*, ii) 27.2 million reads, and iii) 7.1 million reads to determine an optimal sequencing depth of coverage for gut metagenomics studies by interrogating intestinal microbiome and resistome

composition from complex food animal gut microbiomes. To our knowledge, this gut animal microbiome study is the first to utilize three different sequencing depth parameters to interrogate complex food animal cecal microbiome and resistome composition.

Shotgun metagenomic studies have exceptionally expanded our knowledge of microbial communities. Furthermore, the quantity of sample for sequencing has dramatically increased with the advances of high-throughput sequencing technologies. Nevertheless, a complete characterization of microbiomes is still a challenge due to the high complexity of microbial communities and the relative bacterial abundance in the microbiome. A total of seventy-two cecal samples from four major food animal production classes (cattle, swine, chicken, and turkey) monitored by the National Antimicrobial Resistance Monitoring System (NARMS) were utilized for this study. Shotgun metagenomic approaches with a different sequencing depth, allowed a comprehensive microbial profiling and resistome characterization of these complex environments.

4.9.1 Microbiome

The bacteria that can be grown in the laboratory are only a small fraction of the total diversity in the microbiome, our ongoing work on HiC and Shotgun metagenomics estimates that over 90% of the whole animal cecal microbe diversity is uncharacterized (data not yet published). Thus, there is a need for better curated databases for reliable annotation of metagenomics taxonomic assignment. In addition, many microbial community characterization studies are based on time represented in the metagenomic data set, termed coverage [164]. However, small data sets from complex microbiomes could mislead a microbial characterization where lower abundance microorganisms could be left behind. Our study expands upon the

findings of previous studies by utilizing three different established read performance parameters to provide a full identification of previously uncharacterized microorganism, unculturable bacterium, including anaerobic and those with low relative abundance within the food animal cecal microbial community.

Our findings revealed four major bacterial phyla, Bacteroidetes, Firmicutes, Proteobacteria, and Actinobacteria comprising over 95% of the cecal microbiota in the food animal cecal microbiome. These findings were very comparable by the three sequencing depth approaches with some differences in the percentage composition. Bacteroidetes and Firmicutes were over 40%, while Proteobacteria and Actinobacteria were around 15% and 1.5%, respectively. At higher taxonomic rank - Phylum we characterized 20 different phyla at i) 31 million reads coverage and ii) 27.2 million read coverage while 19 at iii) 7 million reads coverage. Meanwhile at family taxonomic rank, at i) 31 million reads coverage 153 different families, representing 39.13% of total characterization, ii) 27.2 million read coverage 132 families representing 33.76%, while 19 at iii) 7 million reads coverage we profiled 106 bacteria family, representing 27.11%.

Another study on wild and domestic animals including pig, chicken, and turkey, were sequencing depth ranged between 111 to ~196 million reads per sample, established a phyla composition of Bacteroidetes, Firmicutes, Proteobacteria, and Actinobacteria as the most abundant with a higher level of Bacteroidetes, over 50% on farm animals [165]. Regarding the taxonomic assignments in a metagenomic dataset, we consider that it depends not only on the complexity of the microbiome but also on sequencing depth and the availability of the computational tools and curated databases to extract valuable information about the

microbiome. However, errors in public databases can lead to misclassification from biological data [166, 167]. The descriptive statistical analysis, one-way analysis of variance showed significant differences between the means of three the three sequencing platforms approaches: i) 31 million reads coverage (means 81.0416667), ii) 27.2 million reads coverage (means 64.5694444) and iii) 7.1 million reads coverage (means 41.4444444). Similar results were obtained in a study about the impact of sequencing depth on the characterization of microbiome and resistome, where new bacteria taxa were increasingly identified by increasing sequence depth [154].

The taxonomic affiliations from this shotgun metagenomics comparative sequencing depth between three different performance approach, allowed to characterize to a lower taxonomic category. We characterized 257 bacteria genera. At i) 31 million reads coverage we identified 257 genera, 220 at ii) 27.2 million reads coverage and 164 genera at iii) 7.1 million reads coverage. Study on taxonomic composition of the fish gut microbiome studies, using 23.5 million reads for taxonomic assignment, was able to identify the presence of 36 phyla, 326 families and 985 genera [167], this could be a confirmation of the complexity of the food animal gut microbiome. Rodriguez-R and Konstantinidis [162] postulate that coverage is not simply a function of dataset size. Instead, the relationship heavily depends on the complexity of the microbial communities sampled.

The most abundant genera characterized utilizing three different read performance approaches, were Bacteroides, Lactobacillus, Escherichia, Odoribacter, Enterococcus, Streptococcus, Clostridium, and Campylobacter at different millions of reads coverage. On the other hand, we were able to identify Bifidobacterium, Prevotella, and Eubacterium on lower

abundance-million reads coverage. It is well studied that both harmless and beneficial bacteria, apart from contributing to a heathy microbiome, act as reservoirs of AMR genes[168] . The Shannon-Wiener Diversity Biodiversity Index calculated based on relative abundance-million reads per sample per sequencing depth performance parameters demonstrated a significant difference within the three sequencing performers. Thus, we report a significant in biodiversity denotated when different sequence dept coverage is utilized.

Studies on the impact of diet on bovine rumen function and outlines what is known about the rumen microbiome [169] reported the presence of Lactobacillus or Bacteroides as the most predominant genera, while in similar studies on profile the taxonomic and functional composition of the small intestinal luminal microbiome of neonatal calves using whole-genome sequencing of the metagenome [170], reported Lactobacillus as the most abundant in the microbiome while Clostridium and Eubacterium, Prevotella, Bifidobacterium and Bacteroides were observed in lower abundance. Other studies on the dynamics of the fecal microbiome and antimicrobial resistome in commercial piglets during the weaning period [171], reported Bacteroidetes and Firmicutes constituted approximately 80% of all bacteria phyla while Acteroidetes, Firmicutes, Proteobacteria, Synergistetes, Spirochaetes, Chlamydiae, Lentisphaerae , Actinobacteria, Alistipes, Bacteroides, Prevotella, Clostridium, Parabacteroides, Lactobacillus, Eubacterium, Pyramidobacter, Butyricimonas, Phascolarctobacterium, Treponema, Fusobacterium, Oscillibacter, Desulfovibrio, Faecalibacterium, Butyricimonas, Phascolarctobacterium, Treponema, Fusobacterium, Oscillibacter were the most abundant genera. Those results are very comparable with our finding at different sequencing depth coverage. At i) 31 million redas, Veillonella 0.00261972 %, Clostridium 2.97150867 %,

Bifidobacterium 0.61426761 %, Ruminococcus 0.61426761 %, Roseburia 2.91313771 %, Eubacterium 0.52570423 %, Prevotella 1.00869094 %. At ii) 27.2 Veillonella 0.00139437 %, Clostridium 2.87053674 %, Bifidobacterium 0.71390141 %, Ruminococcus 0.13383099%, Roseburia 2.96473453 % while at Iii) 7.1 million redas, Veillonella 0.00169014 %, Clostridium 2.89207095 %, Bifidobacterium 0.76507042 %, Ruminococcus 0.10608451 %, Roseburia 3.22531065 %, Eubacterium 0.47097183%, respectably. This alone indicates the high complexity of the cecal microbiome and the potential cross interaction between pathogenic and nonpathogenic bacteria that could increase the horizontal gene transfer on AMR genes from one to another.

4.9.2 Resitome

Despite the increasing characterization of antimicrobial resistant bacteria and antimicrobial resistance genes in different microbiomes influenced by human activity [172], there is still a lack of knowledge on the diversity and resistome abundance of intestinal microbiomes in food animal, particularly the total composition of microbiota of those unculturable bacteria [39]. In addition, tracking AMR genes drivers is challenging due to the array of microbiome complexity and sometimes contradictory AMR literature [165]. Increasing a sequencing depth could help solve in part the lack of complete characterization of complex microbiomes.

Our study reveals a significant difference in AMR genes per cecal sample by the three different sequencing depth approaches. At i) 31 million reads AMR-RPKM we profiled 104 AMR genes representing 13 AMR classes with 23 AMR genes per cecal sample on average, while 119 AMR genes, representing 13 AMR classes with an average of 26 genes per cecal sample were profiled

at ii)27.2 million reads ARM-RPKM, in contrast, 94 AMR genes, representing 13 AMR classes with an average of 20 genes per cecal sample were characterized at iii) 7.1 million reads ARM-RPKM.

Among the most abundance AMR classes characterized withing the three sequencing dept approaches are aminoglycoside, betalactamase, fosfomycin, macrolide, lincosamide, nitroimidazole, phenicol, quinolone, rifampicin, sulphonamide, tetracycline, trimethoprim, and vancomycin at different read-RPKM. However, with more sequencing depth, the abundance of AMR determinants increases. Same AMR patterns were reported [171] on a study regarding the dynamics of the fecal microbiome and antimicrobial resistome in commercial piglets during the weaning period.

The intestinal microbiome is considered one of the primary reservoirs of AMR genes [163]. AMR genes can be transferred within the microbial community via horizontal gene transfer (HGT)[173]. Similarly, AMR genes in food animals can be transmitted to humans through food animal goods [174].

It is now well established that antimicrobial in a microbial community leads to the rapid emergence and propagation of AMR genes. *tet(Q), tet(W), lnu(C), aph(3')-III, erm(B), cfxA6*, were the most abundant AMR genes characterized by at i)31 million reads AMR-RPKM. while, *tet(Q), lnu(C), aph(3')-III, erm(B), and cfxA6* were the most abundance genes identified at 37.2 million reads AMR-RPKM In contrast to *tet(W), lnu(C), aph(3')-III, erm(B), cfxA6,VanX-D*, were the most abundant AMR genes characterized at iii)7.1 million reads AMR-RPKM. *tetW, tetQ, tet44, tet37, tet40, mefA, aadE, ant(9)-1, ermB and cfxA2*. has been previously reported at most

abundant AMR genes on pig faeces [70], *tetB, tetQ, ermA, ermB* and dfrA1 AMR gene dynamics on a commercial pig farm with high antimicrobial usage [175].

Preclinical studies demonstrated the transfer of resistance genes from probiotics to pathogenic bacteria within the gut microbial communities via HGT, including *erm(B) and tet(M),* genes which encode for macrolide and tetracycline resistance respectively, and are shown to transfer from Lactobacillus to potential pathogens such Enterococcus faecalis, and Listeria monocytogenes, spreading resistant traits into these pathogenic bacteria [176] . On the other hand, *cfxA6* genes were present in three Prevotella spp. and one Porphyromonas spp. strains containing *cfxA* genes (56%) that were resistant to the β-lactam antimicrobial drug class [177, 178] In contrast, *qnrB41, qnrA1, qnrS1, oqxB, blaOXA-24, blaOXA-51, blaCTX-M*, and *blaROB-1* were some of the less abundant genes among the three sequencing depth approaches we used. *Qnr* genes are known to confer low-level resistance to fluoroquinolone in Enterobacteriacea, *qnrA, qnrB and qnrS* gene among extended-spectrum β-lactamas (ESBL)-producing E. coli and Klebsiella spp from Lome, Togo [179]. These qnr gene positive strains were highly resistant to fosfomycin, ceftazidim, ceftriaxon and gentamycin. However, they remain susceptible to fosfomycin amikacin and Fosfomycin [179-181] . *oqxB* a well-known efflux pump gene that has emerged as a factor contributing to the antibiotic resistance in Klebsiella pneumoniae, *OqxB* underwent horizontal gene transfer and is now seen in other bacterial pathogens including Escherichia coli, Enterobacter cloacae and Salmonella spp [182, 183]. *blaOXA-24, blaOXA-51* multi drug resistance in Acinetobacter baumannii [184]. *blaCTX-M*, extended-spectrum β-lactamas (ESBL), multidrug resistance was reported in low abundance in food animals [185, 186].

After we confirmed significant difference on samples resistome and microbiome between the three different sequencing depth coverage, using Negative binomial regression and ANOVA. Figure 8 show best sequencing depth on microbiome and resistome profiling.*repeated*

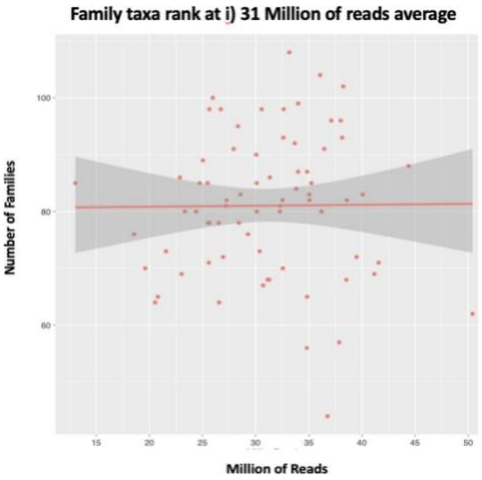


Figure 8 A

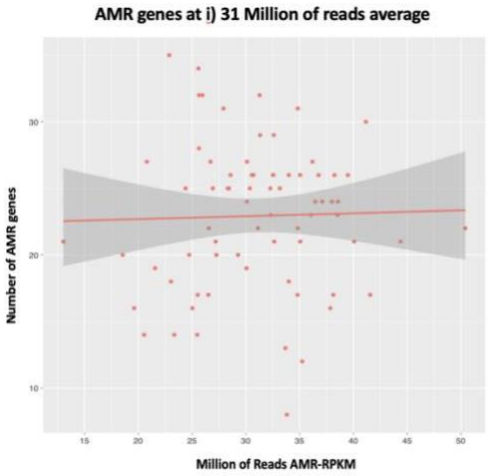


Figure 8 B

Figure 8. sequencing depth on microbiome and resistome profiling. A. Microbiome at 31 million reads depth coverage/ family rank. B. Resistome at 31 million reads depth coverage.

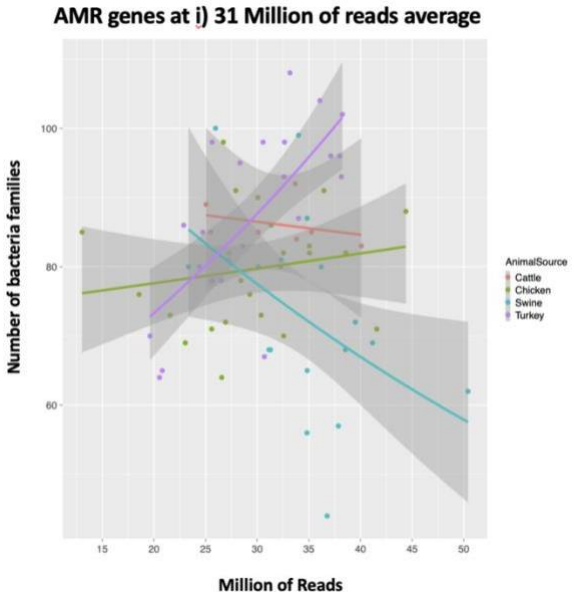


Figure 9 A

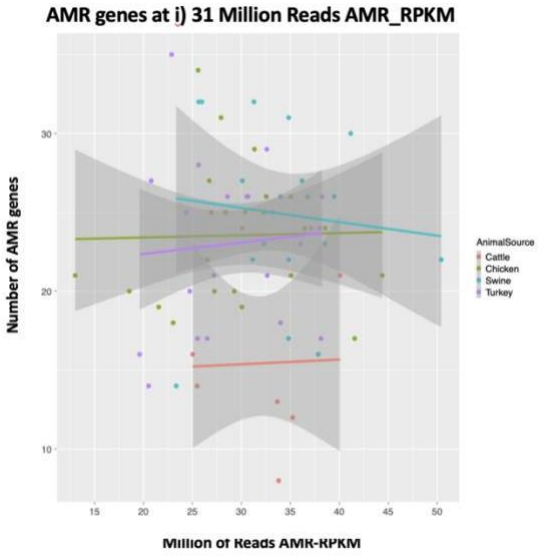


Figure 9 B

Figure 9. sequencing depth on microbiome and resistome profiling by Animal source. A. Microbiome at 31 million reads depth coverage/number of family by animal source. B. Resistome at 31 million reads depth coverage by animal source.

4.9 **Conclusions**

The study reveals significant differences between the three sequencing depth approach. Descriptive statistics show significant differences between the means of sequencing dept performance. i) 31 million reads mean 31.015644.6, and 27.2 million reads mean 27.184285.6, and iii) 7.1 million reads mean 7.057574.93. Descriptive statistics of the microbiome showed significant differences between the means of sequencing depth performance. We were able to identify a total of 20 bacteria Phyla. i) 31 million reads/average Firmicutes, Bacteroidetes, and Proteobacteria, representing the most abundant phyla. ii) 27.2 million reads/average outlined Firmicutes, Bacteroidetes, and Proteobacteria, comprising the most abundant phyla and at iii) 7.1 million read/average we denoted 19 were Bacteroidetes Firmicutes, and Proteobacteria, representing the most abundant phyla.

At family taxonomic rank, approximately 400 bacterial families with different relative abundance were characterized across all sequencing depth approaches among all seventy-two cecal metagenomic samples. i) 31 million reads/average depicted a total of 153 bacteria families, with an average per sample of 80 taxa-family rank with Bacteroidaceae, Lactobacillaceae, Enterobacteriaceae, Clostridiaceae, Enterococcaceae , and Campylobacteraceae represented the most abundant families, meanwhile at ii) 27.2 million reads/average denoted a total of 106 families, with an average per sample of 41 taxa-family ranks, were Bacteroidaceae, Lactobacillaceae, Enterobacteriaceae, Enterococcaceae, Clostridiaceae, and Campylobacteraceae, were the most abundant families, compare to iii) 7.1 million read/average outlining a total of 132 families, with an average per sample of 64 taxa-

family ranks, were Bacteroidaceae, Lactobacillaceae, Enterobacteriaceae, Enterococcaceae, Clostridiaceae, and Campylobacteraceae represented the most abundant families.

Resistome analysis reveals a significant difference in the number of AMR genes per cecal sample by the three different sequencing depth i) 31 million *AMR-RPKM*, ii) 27.2 million AMR-RPKM and iii) 7.1 million AMR_RPKM. We were able to characterize a full resistome composition of the data set. At i) 31 million AMR-RPKM we characterized a total of 1651 AMR genes with an average of 23 AMR genes per cecal sample, at ii) 27.2 million AMR-RPKM we identified a total of 1429 AMR genes with an average of 20 AMR genes per cecal sample, and at iii) 7.1 million AMR-RPKM 1893 AMR were identified with an average of 26 AMR genes per cecal sample. By deeper analysis on the resistome composition at i) 31 million AMR-RPKM profiled 104 unique AMR genes representing 11 AMR classes, ii) 27.2 million AMR-RPKM profiled 119 AMR genes corresponding to 11 AMR classes, and iii) 7.1 million RPKM profiled 94 AMR genes representing 11 AMR classes.

It is important to clarify that due to limitation on Illumina sequencer availability in our lab, all results from ii) 27.2 million reads averaged depth coverages were collected from three different runs as we used NextSeq 500. Based on our high sequencing depth coverage goal, NextSeq 500 total performance parameters allows to run no more than 25 samples at the time. Therefore, we run our gut metagenomic set of 72 food animal sample witting three different sequencing runs.

We have used deep shotgun metagenomics sequencing approaches to estimate a suitable sequencing depth to achieve microbial and resistome characterization. The number of sequences required to characterize a sample depends primarily on the goal of the study, the

complexity of the microbiome, as well as the available resources. It is imperative to note that there are no abundance thresholds below which a microbe can be overlooked as unimportant; every member of a bacterial community plays a critical role in the functional relationship to maintain the good health of the microbiome. Overall, our data indicate that sequencing depth critically affects the inferred AMR gene content and taxonomic diversity of complex microbiomes. the more sequencing depth the better characterization of microbial composition. It is noteworthy that sequencing depth alone is not sufficient for accurate taxonomic or resistome profiling. Proper bioinformatics analysis, including read quality control, data normalization, and appropriate curated reference databases, is essential for obtaining reliable results. Increasing the sequencing depth of coverage in bacterial taxonomic and resistome profiling studies enhances the accuracy, resolution, and sensitivity of the analyses. However, the optimal sequencing depth depends on the specific research goals, the complexity of the microbial community, and the expected abundance of rare taxa or resistance genes. These results not only help to better understand uncultured bacteria and measure potential new AMR gene reservoirs present in intestinal microbiomes and their latent hazards on animal and human health but also, provides insights into food animal intestinal microbiome and resistome characterization.

Acknowledgments

The views expressed in this article are those of the authors and do not necessarily reflect the official policy of the Department of Health and Human Services, the U.S. Food and Drug Administration, or the U.S. government. Reference to any commercial materials, equipment, or

process does not in any way constitute approval, endorsement, or recommendation by the Food and Drug Administration.

PART II

5. LINKING ANTIMICROBIAL RESISTANCE GENES AND PLASMIDS TO HOST BACTERIA

5.1 Abstract

Antimicrobial resistance (AMR) is a natural phenomenon that occurs over time through genetic changes. Antimicrobial resistance is a global public health threat were selective pressure exerted by antimicrobial use has been the major driving force. However, the true magnitude of antimicrobial resistance on human and animal health is not known. We employed deep shotgun metagenomic sequencing and proximity ligation (Hi-C) library sequencing to characterize the resistome, assemble genomes from metagenomic samples, and accurately attribute antibiotic resistance genes and plasmids to host cells. We randomly selected 21 cecal samples from food animal sources (cattle n= 6, swine n= 6, chicken n= 3, and turkey n= 6). More than 75 million reads/sample for Hi-C and more than 100 million reads/sample for shotgun metagenomic sequence reads were generated.

Bioinformatic analysis revealed over 200 bins containing metagenome assembled genomes (MAGs) with different levels of completeness, novelty score, and contamination based on CheckM. A total of 245 previously uncharacterized genomes were reconstructed using MAGs with high level of confidence (>90% Completeness, >90% Novelty, < 5% contamination). On the other hand, 78 previously characterized MAGs were identified with high level of confidence (>90% Completeness, <90% Novelty, < 5% contamination). Of the 245 newly reconstructed MAGs, 24 were at Bacteria taxonomic rank level, 5 at Phyla (Actinobacteria [11 genomes from

UID2112], Firmicutes [3 genomes from UID1022], Bacteroidetes [17 genomes from UID2605], Euryarchaeota [2 genomes from UID3]), 4 at Class (Bacilli [1 genome from UID259], Clostridia [9 genomes from UID1118]), Deltaproteobacteria [3 genomes from UID3218], Gammaproteobacteria [1 genome from UID4201]), 5 at Order level (Actinomycetales[1 genome from UID1663, and one genome from UID1812], Bacteroidales [9 genomes from UID2617, 3 genomes from UID2621, 12 genomes from UID2657, 1 genome from UID2716], Clostridiales [30 genomes from UID1120, 67 genomes from UID1212, 17 genomes from UID1226]), Lactobacillales [1 genome from UID544, 1 genome from UID355], and Selenomonadales [2 genomes from UID1024]), and 3 at family level (Lachnospiraceae [5 genomes from UID1255, 13 genomes from UID1256, 5 genomes from UID1286, 1 genome from UID4932], Spirochaetaceae [1 genome from UID2535], and Spirochaetaceae [2 genomes from UID2535]).

Plasmid characterization using Hi-C and shotgun metagenomics allowed identification of 146 plasmids (>= 85% completeness, >= 90% reference sequence similarity), and over 13000 plasmid-contigs (<85% completeness, < 90% reference sequence similarity).

We identified over 400 antimicrobial resistance genes representing 22 antimicrobial classes including across the sample set: aminoglycoside (40 gene variants), beta-lactam(37 gene variants), bleomycin (2 gene variants), colistin (3 gene variants), efflux (6 gene variants), fosfomycin (4 gene variants), glycopeptide (6 gene variants), lincosamide (9 gene variants), lincosamide/streptogramin (2 gene variants), macrolide (16 gene variants), macrolide/lincosamide/streptogramin (4 gene variants), nitroimidazole (1 gene variant), phenicol (9 gene variants), phenicol/oxazolidinone (1 gene variant), phenicol/quinolone (2

gene variants), pleuromutilin (1 gene variant), quinolone (5 gene variants), streptogramin (1 gene variant), streptothricin (3 gene variants), sulfonamide (3 gene variants), tetracycline (25 gene variants), and trimethoprim (8 gene variants).

Hi-C detected both previously known and novel associations between AMR genes, plasmids, and host genomes. Thus, Hi-C proximity ligation and deep shotgun metagenomics sequencing were able to link antimicrobial resistance genes and plasmids to host bacteria to help explain the reservoirs of antibiotic resistance genes accessible to pathogens and their potential for widespread dissemination along the 'farm to fork' continuum.

5.2 Introduction

The clinical application of antimicrobials has greatly impacted the morbidity and mortality of bacterial infectious diseases in humans and animals [187], but the increasing medical and veterinary problems arising from the emergence of antimicrobial resistance genes is one of the greatest threats in human history [188, 189]. It is the leading cause of death worldwide, and it has been estimated that AMR infections cause at least 700,000 deaths annually, is astonishingly projected to increase to 10 million deaths per year by 2050 across the globe [36, 190], and affects low, middle, and high income countries [191]. AMR infections are rising exponentially, have reached an alarming level of concern due to the high manifestation of multi-drug resistance (MDR) microbes [192], making them harder to treat [193]. Antimicrobials are used in livestock (cattle, swine, chicken, turkey), aquaculture, pets, crops, and humans which makes a perfect storm for pathogens and AMR bacteria to create new genetic AMR reservoirs which can be transferred by horizontal gene transfer(HGT)[168, 194], which is mainly

mediated by mobile genetic elements (MGEs) like plasmids [195, 196]. Therefore, it is necessary to understand the composition and distribution of antimicrobial resistance genes in the animal gut. The swine gut microbiota is an important reservoir of AMR genes, which have a great risk of transferring to humans and the environment[195]. Deep metagenomic sequencing on swine feces identified a total 349 unique AMR gene clusters representing 69 antimicrobial resistance classes, where swine raised on commercial farms had a significantly higher AMR level than swine under semi-free ranging conditions [197]. Shotgun metagenomics analysis on broiler chicken ceca showed over 500 AMR genes, representing 20 antimicrobial resistance class [198]. Similar studies to interrogate changes of intestinal microflora and antimicrobial resistance genes in swine and chickens show *tetW* as the resistance gene with the highest relative abundance, followed by *tetW/N/W,* then *lnuA*; andothers from high to low were *mdtB, lnuC, ErmB, mdtC, ErmQ, tetBP, vatE, evgS, acrB, cpxA, mefA mdtB* [199]. Gastrointestinal resistome analysis in fecal and endoscopic samples, showed that the fecal samples underestimate the real resistome composition on the of the gastrointestinal track[200]. Therefore, it is necessary to understand the composition and distribution of antimicrobial resistance genes in the animal gut to counteract further AMR transmission. Thus, understanding who, what, and how these genes spread within different microbial communities, could provide ways to control and predict the emergence of antimicrobial resistance and potentially decrease the rates of infectious diseases seen in humans and animals.

5.2.1 Shotgun metagenomics:

Next-generation sequencing (NGS) technologies have revolutionized multiple biomedical research areas, from clinical diagnosis, pathogen discovery to characterization of AMR genes[201] . Shotgun metagenomic sequencing, a NGS approaches allows to comprehensively interrogate the total microbiota on a complex environmental sample [202]. The method enables assessment of bacterial taxonomic diversity, relative abundance of microbes, as well as full characterization of resistome composition in specific environments. Shotgun metagenomics also provides a means to discover and characterize unculturable microorganisms that are otherwise difficult or impossible to evaluate with traditional culture-based approaches. Advances in sequencing technology, particularly the development of high-throughput DNA sequencing has made taxonomic profiling and resistome characterization more accessible and practical [203]. In particular, the combination of deep shotgun and Hi-C metagenomics, along with powerful bioinformatics tools and computational platforms have facilitated not only the identification of microbial communities [30, 32, 126] and retrieval of complete genomes[204], but also to link antimicrobial resistance genes to the AMR plasmid and the host bacteria [35, 205] by bypassing the need for culture-based approaches.

5.2.2 Hi-C metagenomics:

Metagenomics a culture-independent genomic approach, eludes the limitation of bacterial cultivation and isolation and provides a snapshot of the community structure and the functional capabilities present in complex microbiomes [30, 32, 126]. Thus, metagenomics offers insights into the resistomes, metabolic, and virulome potential of microbial communities. However, gene attribution to specific bacterial strains present in the microbiome is a challenge

of the metagenomics approach. Thus, when traditional metagenomics is used to detect specific genes, it is difficult to predict if they belong to a particular bacterium or to the background microbiota. High-throughput chromosome conformation capture (Hi-C), a DNA proximity ligation technique has the power to retrieve metagenomic assembly genomes (MAGs) accurately from a single sample, could generate millions of paired-end reads based of sequence depth coverage, and link DNA fragments in close proximity in 3D space within cells [21, 44]. When the Hi-C technique is combined with traditional shotgun sequencing, it shows great capabilities for genome binning and retrieval of high-quality MAGs from complex microbiome samples[26, 30, 126, 206, 207].

5.3 Material and Methods

We randomly selected of 21cecal samples from four major food animal production classes (cattle n= 6, swine n= 6, chicken n= 3 and turkey n= 6). Cecal samples were collected as part of the National Antimicrobial Resistance Monitoring System's (NARMS) routine sample collection from slaughter establishments across the United States. Samples were collected according to the United States Department of Agriculture (USDA) Food Sampling and Inspection Service (FSIS) FSIS Directive 10,100.1, "FSIS Sampling for the National Antimicrobial Resistance Monitoring System (NARMS). https://www.fsis.usda.gov/policy/fsis-directives/10100.1." Cecum contents (cattle, swine, or turkey) and intact ceca (chicken or turkey) were shipped in 120mL sample containers in a chilled cooler. Upon arrival, samples were frozen at -80°C until DNA extraction.

5.3.1 Shotgun metagenomics community DNA extraction, and quantification.

Metagenomic DNA was extracted from approximately 0.2g of cecal contents using PowerMicrobiome® Kit (QIAGEN, Inc., Germantown, MD, USA) following the manufacturer's recommendation. The concentration of the extracted DNA was determined using Qubit™ 3.0 fluorometer (Thermo Fisher Scientific, Wilmington, DE, USA), and stored at −20 °C until sequencing.

5.3.2 Hi-C metagenomics.

We utilized the ProxiMeta™ Metagenome Deconvolution Platform (Phase Genomics, Seattle, WA, USA). Breifly, the proximoMeta Hi-C deconvolution workflow is as follows: formaldehyde induces covalent bonds between DNAs that are close by in three-dimensional space (bonds are depicted by black arcs), and therefore within the same cell. Hi-C library preparation via restriction enzyme digestion (top left), ligation and junction enrichment (bottom left), preparation of the fragment for paired-end sequencing (top right), reads from fragments generated by Hi-C provide linkage information of non-contiguous DNA originated from the same cell (bottom right). Hi-C links can be used in a graph clustering context to deconvolute contigs into their original cellular groupings, including both chromosomes and plasmids. Figure 1. Shotgun metagenomics and HiC metagenomics library preparation illustration.

5.3.3. Deep shotgun metagenomic sequencing.

Community DNA extracted was diluted to a concentration of 0.2 ng/µl. Five microliters of diluted DNA was used for subsequent library preparations. Sequencing libraries for each sample were constructed using the Illumina Nextera XT DNA library preparation kit (Illumina,

Inc., San Diego, CA, USA) and sequenced on the HiSeq 2500 system (Illumina, Inc., San Diego, CA, USA) according to manufacturer's recommendations. Briefly, tagmentation and tagging of the DNA with unique adapter sequences were performed using Nextera transposase. Limited-cycle PCR was used to amplify the tagged DNA and add indexes. Libraries were pooled together to be sequenced on HiSeq 2500 High Output mode (2 x125 bp paired-end sequencing), paired-end sequencing). Library quantity and quality was measured by Qubit™ 3.0 fluorometer (Thermo Fisher Scientific, Wilmington, DE, USA) and TapeStation system (Agilent Technologies, Inc., Santa Clara, CA, USA) Figure 10. Shotgun metagenomics and HiC metagenomics library preparation illustration.

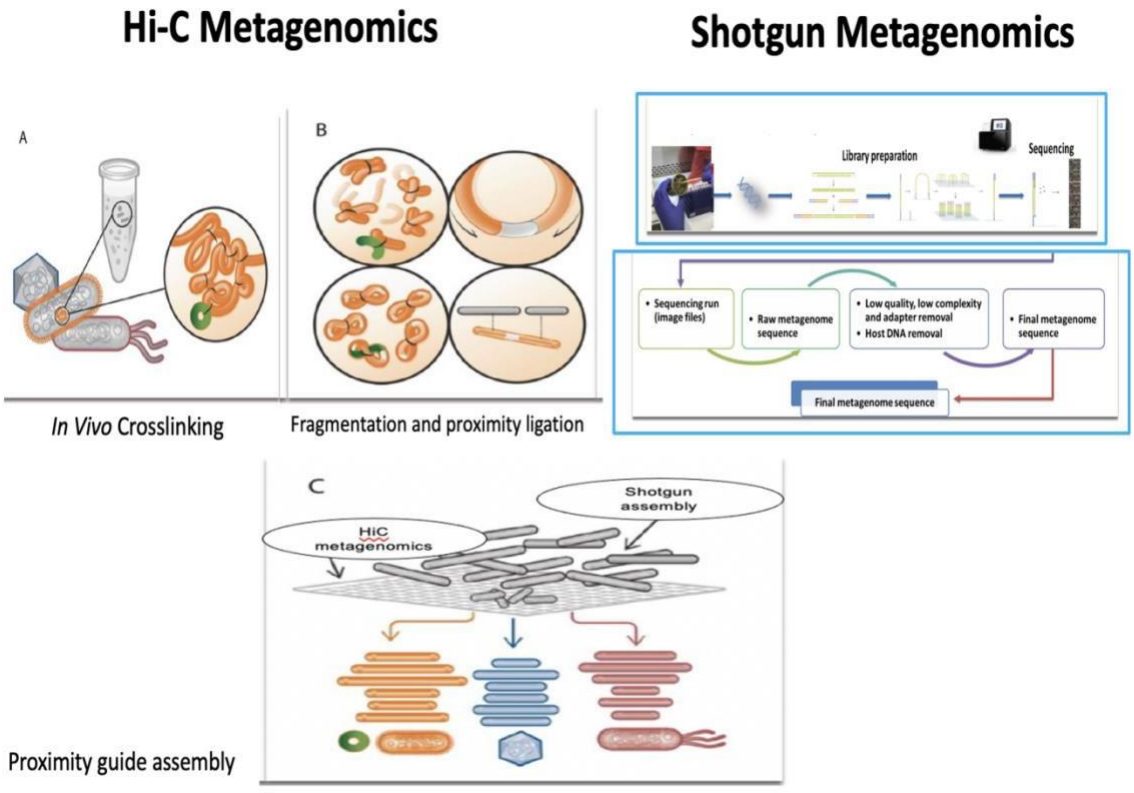


Figure 10. Shotgun metagenomics and HiC metagenomics library preparation illustration.

5.4 Data processing:

The Illumina bcl2fastq2 Conversion Software v2.20 demultiplexes sequencing data and converts base call (BCL) files into FASTQ files for further analysis. A standard cleaning procedure is applied to all shotgun and Hi-C FASTQ sequences with respect to adaptors and low quality sequences using fastp [159] with parameters ILLUMINACLIP:Illumina-Adapter.fa:2:30:10 LEADING:20 TRAILING:20 SLIDINGWINDOW:5:20 MINLEN:90. To remove host DNA sequences, the trimmed and filtered FASTQ sequences were then mapped to cattle, chicken, turkey, and swine genomes based on the source of the samples with Bowtie2 [160] using standard settings. The mapped and unmapped reads were separated using Samtools [144]. Reads that were trimmed, filtered and unmapped to host DNA were used as an input for all further downstream analysis.

5.4.1 Genome Sequence Assembly

Metagenomic assembly is a computational process aimed at reconstructing genes and genomesfrom metagenomic mixtures. MetaSPAdes (v3.11.1) [99] was used with default parameters.

5.4.2 Processing Hi-C reads

Hi-C read sets were mapped to the metagenomic assembly by Burrows-Wheeler Alignment with the option -5 (v0.7.17) [208]. SAM files were sorted by name using samtools (v1.10) [144].

5.4.3 Deconvolution of Hi-C data

The binning process is done by bin3C (v0.1.1)[45]. Briefly, bin3C is an unsupervised method that exploits the hierarchical nature of Hi-C proximity ligation interaction rates to resolve MAGs using a single time point. It is based on two stage normalization method to process the raw contact maps. It first divides the raw Hi-C counts by the product of the number of restriction sites and then uses the Knight-Ruiz algorithm[141] to construct a general doubly stochastic matrix. Bin3C utilizes Infomap software [142] to bin contigs. bin3C can retrieve MAGs from metagenomes by combining conventional metagenome shotgun and Hi-C sequencing data.

Basic steps on bin3C are as follows: Read clean-up for both shotgun metagenomics and Hi-C read-sets using Fastp [159] a tool designed to provide fast all-in-one preprocessing for FastQ files, shotgun assembly grouping contigs into scaffolds by MetaSPAdes [99]. Hi-C reads are mapped to the corresponding scaffold by BWA-MEM[143]. The resulting BAM files were subsequently processed using samtools [144] to remove unmapped reads and supplementary and secondary alignments, then sorted by name and merged. bin3C constructs a contact map from the Hi-C read-pairs under constraints on minimum acceptable length (default, 1000 bp) and minimum acceptable raw signal (default, five non-self-observations) for contig inclusion. Based on the assumption that DNA-DNA interactions between loci within the same cell (intra-

cellular proximity interactions) occur an order of magnitude more frequently than interactions between cells (inter-cellular), unsupervised cluster analysis by using Markov clustering (https://dl.acm.org/doi/book/10.5555/868986), the Louvain method [145], and infomap (https://github.com/mapequation/infomap) are used to construct genome bins.

Genome clusters were BLASTed to RefSeq genomes using Mash [209] to identify close matches for genome clustering. Genome clusters were further examined using CheckM [101] lineage_wf workflow to assess genome quality and estimate high-level phylogenetic placements for each cluster based on single-copy marker gene analysis (completeness and contamination).

All Hi-C linkages between contigs of interest and any other contig were considered, except for those contig-cluster linkages represented by less than three reads. We next used the same criteria to infer all linkages between AMR and plasmid contigs.

Phylogenetic analysis of the genomes linked to contigs of interest was performed using the CheckM tree workflow. CheckM provides a set of tools for assessing the quality of genomes recovered from isolates, single cells, or metagenomes. It provides robust estimates of genome completeness and contamination by using collocated sets of genes that are ubiquitous and single copy within a phylogenetic lineage. Figure 11. Hi-C and Shotgun metagenomics data processing.

5.3.4 Taxonomic, resistome and palsmidome analyses:

Taxonomic classification was performed using Kraken2 [95] using the latest NCBI database version built (need version) with the “kraken-build standard” method. Kraken2 with the default k-mer size and parameters were used for the analysis. We used threshold of 0.2 for

kraken-filter step. The microbiome composition and the taxa abundances were estimated by the Bracken.

The presence and the abundance of AMR genes in metagenomic datasets were determined using the Short, Better Representative Extract Data set (ShortBRED)[161]. ShortBRED-Identify was used to generates unique peptide markers for AMR genes-protein sequences compiled from AMRFinderPlus version 3.10 database (downloaded 2021-09-30; Index of /pathogen/Antimicrobial_resistance/AMRFinderPlus/database (nih.gov)). ShortBRED-identify used an 85% amino acid identity threshold to cluster the AMRFinderPlus database into nonredundant representative sequences and to query representative sequences against one another and against the comprehensive protein reference datasets UNIREF100 (https://www.uniprot.org/uniref/) (data accessed 13 January 2022). ShortBRED quantify then was used to map translated reads to protein markers at ≥95% amino acid identity across ≥95% of the marker length and generate the abundance of the AMR genes, normalized to reads per gene kilobase per million reads (RPKM). PLDS, plasmid database [210] was accessed 13 January 2022.

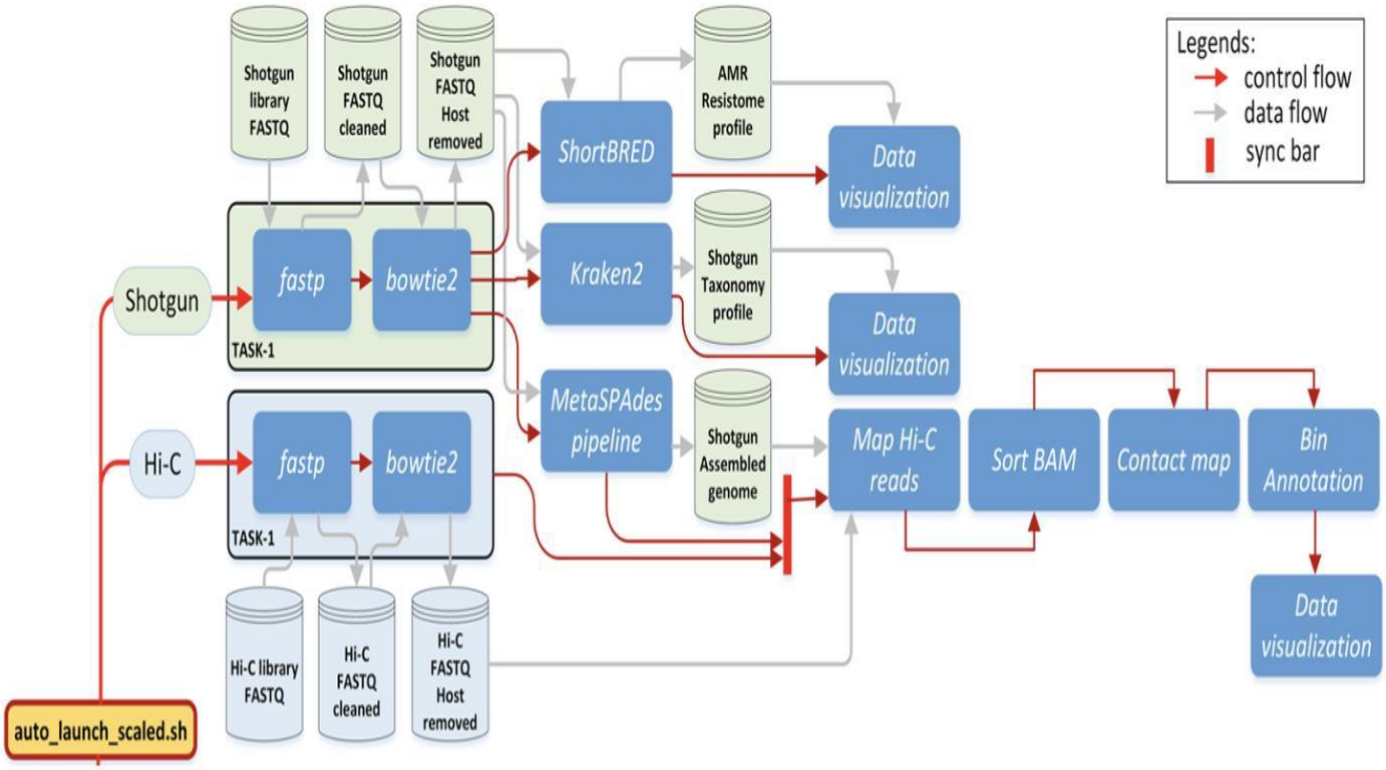


Figure 11. Hi-C and Shotgun metagenomics data processing.

5.5 Statistical analysis:

We utilized descriptive statistics to help describe and understand the features (measures of central tendency, measures of variability, and frequency distribution) of our metagenomic data set by giving short summaries about the sample and measures of the data. Data visualization was performed using R studio and R (https://www.r-project.org/foundation/).

5.6 Results. And Discussion

Hi-C is mostly applied in cancer studies and in the biomedical field [211]. However, its applicability on the gut microbiome is still limited but showing increasing interest for the impact

on human health research [44]. Many studies involving Hi-C and shotgun metagenomics are originated based on construction of a synthetic microbial community and simulated paired-end sequencing of the synthetic microbial community to obtain metagenomic assemblies that could subsequently attempt to bin using experimentally derived Hi-C reads. On the contrary, our research work includes two components, wet lab where we process gut metagenomic samples and dry lab where we attempt to bin bacterial communities by proximity guide assembly. We originally randomly selected of 24 cecal samples from four major food animal production classes (cattle n= 6, swine n= 6, chicken n= 6, and turkey n= 6) monitored by the NARMS. We generated more than 75 million reads/sample for Hi-C and more than 100 million reads/sample for shotgun metagenomic sequence reads to link antimicrobial resistance genes to plasmids and bacterial hosts by using proximity guided assembly. Other studies utilized simulated Illumina paired-end sequencing of the synthetic microbial community to obtain metagenomic assemblies that we could subsequently attempt to bin using experimentally derived Hi-C reads [30]. The gut microbiome is a complex and diverse bacterial community of millions of microorganisms that colonize the host's gastrointestinal track [212, 213]. Thus, natural complexity of the samples plus cellular and other type of environmental debris discovered within the animal guts, made some of the samples very difficult for processing Hi-C libraries. Three chicken samples were particularly challenging, with low DNA quality that carried over into lower sequencing dept results. These three samples underperformed after several trials of Hi-C library preparation and deep sequencing approaches with less than 75 million reads targeted depth of coverage. Therefore, were exclude from the final sample set. Thus, a 21 total cecal samples were used for this study.

5.6.1 Genome discovery from complex animal gut microbiome

Little is known about the true composition of food animal intestinal microbiomes due to their environmental complexity. Taxonomic classification was performed using Kraken2 [95] using the latest NCBI database version built with the “kraken-build standard” method. The microbiome composition and the taxa abundances were estimated by the Bracken. Binning methods help assess issues with assembly fragmentation and to organize contigs into candidate MAGs based on assumptions of shared sequence composition [108] or orthologous linkage data [126]. The presence of single copy genes (SCGs) expected to be in all bacterial lineages has been recommended as a measure of the completeness and redundancy within these bins[214]. We generated over 200 bins containing 1780 MAGs with different levels of completeness and contamination based on CheckM. 88 MAGs (4.9 %) were previously characterized with a high level of confidence (>90% Completeness, < 15% contamination) called” known complete genomes," while 1692 MAGs (95.1 %) were not previously characterized, “Unknown Genomes”. We were able to reconstruct 338 new complete genomes (19%) with high confidence (>90% Completeness, < 15% contamination) at different taxonomic rank. Among the new completed genome, 86 bacteria were at species level including: *Akkermansia muciniphila*, a commensal bacterium commonly found in a healthy gut microbiota, and harbors the gene *ade*F and *Tet*W, potentially involved in ciprofloxacin and tetracycline resistance, respectedly [215]. *Enterococcus cecorum*, which was initially identified as a harmless commensal of the gastrointestinal tract of chickens. However, over the past 15 years, pathogenic strains of *E. cecorum* have become a significant cause of morbidity and mortality in

broiler breeders, and repeated outbreaks occur, with high resistance to macrolides[216]. Eubacterium sp, organisms generally susceptible to penicillins, carbapenems, and clindamycin were found. However, Lactobacillus, some eubacteria, and other facultative anaerobes are resistant to nitroimidazoles. Cephalosporins, erythromycin, and tetracycline have variable activity [217]. *Bacteroides dorei,* a species of bacteria within the genus Bacteroides, is found in the intestinal systems of humans and animals with high numbers of antibiotic resistance genes; it is especially resistance to amoxicillin [218]. Clostridium sp were characterized andare ubiquitous and associated with various diseases in animals and humans. *Clostridium difficile* (*C. difficile*), *Clostridium perfringens* (*C. perfringens*), *Clostridium tetani* (*C. tetani*) and Clostridium botulinum (*C. botulinum*) are the most common pathogens associated with human and animal disease [219]. Bacteroidetes (51.5%), Actinobacteria (32.5%), Firmicutes (13.5%), and Proteobacteria (2.7%) were the most abundant phyla. Clostridia (69%), Deltaproteobacteria (23%), Bacilli (8%) were the most representative families. Clostridiales (76%) and Bacteroidales (18%) represented the most abundant order taxonomic rank. Table 4. Snapshot of previously uncharacterized genome. Table 5. Snapshot on new characterized genomes.

SampleSource	bin_id	Taxa-rank	Completeness (%)	Contamination (%)	Abundance	n50_contigs	Genome Size	Num Contigs	GC (%)
cattle_4959	bin_5	k_Bacteria (UID203)	98.25	8.77	0.09	45948	3424769	99	38.17
cattle_4959	bin_3	c_Bacilli (UID259)	96.03	3.86	0.13	4477	4446908	477	36.2
cattle_4959	bin_8	k_Bacteria (UID2982)	91.42	1.7	0.22	37836	2276740	131	55.67
cattle_6598	bin_15	o_Clostridiales (UID1226)	98.73	4.01	0.36	6966	2899128	310	56.44
cattle_6598	bin_21	o_Clostridiales (UID1226)	98.73	2.32	0.22	35238	2526364	194	42.36
chicken_11905	bin_34	*o_Lactobacillales (UID544)*	97.57	0.75	0.07	5,495	1,484,466	99	40.9
chicken_11905	bin_17	*o_Clostridiales (UID1212)*	96.55	8.38	2.02	24,612	2,778,638	326	53.66
chicken_11905	bin_27	*o_Clostridiales (UID1212)*	89.73	0.67	0.9	5,006	2,175,369	358	54.34
chicken_11905	bin_33	*o_Clostridiales (UID1120)*	89.16	0.93	0.02	38,905	1,631,064	37	43.54
chicken_11905	bin_43	*k_Bacteria (UID2495)*	89.01	3.3	0.2	419,446	1,049,499	8	44.87
chicken_17552	bin_6	*p_Bacteroidetes (UID2605)*	99.52	0.24	1.17	77,822	2,235,576	38	57.33
chicken_17552	bin_12	*k_Bacteria (UID2566)*	98.53	0	0.57	154,271	1,948,592	43	54.47
turkey_2145	bin_4	*o_Bacteroidales (UID2657)*	92.39	2.16	0.09	51,744	2,837,316	108	50.32
turkey_2145	bin_49	*k_Bacteria (UID2495)*	89.01	0.37	0.21	176,822	974,222	20	44.63
turkey_4307	bin_13	*o_Bacteroidales (UID2617)*	99.43	1.51	0.93	300,667	2,814,598	47	50.86
turkey_4307	bin_21	*o_Clostridiales (UID1212)*	98.99	2.01	0.09	182,054	2,434,236	99	41.33
turkey_4307	bin_44	*p_Bacteroidetes (UID2605)*	98.99	0.72	0.08	125,116	1,895,650	31	59.65
turkey_4307	bin_24	*p_Firmicutes (UID1022)*	98.31	3.16	0.4	21,104	2,350,212	130	43.17
turkey_4307	bin_17	*p_Bacteroidetes (UID2605)*	97.6	1.2	0.08	47,917	2,535,232	38	57.06
turkey_4307	bin_15	*o_Clostridiales (UID1212)*	97.46	3.27	3.12	18,470	2,723,666	182	62.34
turkey_4307	bin_19	*f_Lachnospiraceae (UID1286)*	94.71	3.43	0.17	60,701	2,467,512	186	53.69
turkey_4307	bin_38	*o_Selenomonadales (UID1024)*	94.36	3.59	0.43	3,704	2,063,685	192	54.05
turkey_4307	bin_4	*o_Bacteroidales (UID2657)*	94.12	6.44	1.03	44,109	3,383,043	157	50.04
swine_1950	bin_61	*o_Clostridiales (UID1212)*	96.64	0.34	0.12	59,660	1,684,678	52	42.92
swine_1950	bin_53	*k_Bacteria (UID2372)*	96.19	0.95	0.11	9,654	1,808,054	175	37.65
swine_1950	bin_12	*o_Clostridiales (UID1212)*	95.75	3.91	0.7	84,331	2,882,360	97	35.08
swine_1950	bin_21	*f_Lachnospiraceae (UID1256)*	95.61	0.78	0.54	107,627	2,618,979	59	44.14
swine_1950	bin_36	*f_Pasteurellaceae (UID4932)*	95.45	0.11	0.21	169,925	2,122,410	34	40.89
swine_1950	bin_17	*o_Clostridiales (UID1212)*	94.97	3.67	0.89	24,608	2,733,659	101	62.13
swine_1950	bin_69	*o_Clostridiales (UID1120)*	94.33	0	0.23	52,219	1,560,133	96	48.04
swine_1950	bin_18	*o_Clostridiales (UID1226)*	94.3	1.96	0.3	127,397	2,714,435	134	46.4
swine_1950	bin_31	*k_Bacteria (UID203)*	93.1	8.62	1.31	49,844	2,276,333	242	34.35
swine_1950	bin_47	*p_Bacteroidetes (UID2605)*	93.1	3.51	0.73	51,343	1,895,769	80	54.36
swine_1950	bin_28	*o_Bacteroidales (UID2716)*	92.58	4.78	0.07	25,110	2,384,437	107	49.43
swine_1950	bin_55	*o_Clostridiales (UID1212)*	92.56	2.18	1.77	51,824	1,802,315	184	49.97
swine_1950	bin_45	*o_Clostridiales (UID1212)*	92.48	2.01	0.33	147,386	1,915,588	30	51.58
swine_1950	bin_39	*o_Clostridiales (UID1120)*	92.13	1.28	0.16	18,803	2,083,104	186	43.94
swine_1950	bin_62	*f_Lachnospiraceae (UID1286)*	91.96	0.48	0.07	21,378	1,683,729	125	44.17

Table 4. Snapshot of previously uncharacterized genome

SampleSource	bin_id	Taxa-rank	Completeness (%)	Contamination (%)	Abundance	n50_contigs	Genome Size	Num Contigs	GC (%)
Cattle_4959	bin_4	SaccharomonosporaglaucaK62	98.03	1.15	0.24	13928	4306786	196	69.32
Cattle_4959	bin_6	BacilluspsychrosaccharolyticusATCC23296	91.29	4	0.08	60754	3377297	427	39.33
Cattle_6598	bin_36	Bifidobacteriumpseudolongumsubsp.globosum	98.89	5.2	0.1	63779	1870304	104	63.74
Cattle_6598	bin_11	ThermobifidafuscaYX	98.36	0.00	0.08	53,459	3,494,955	164	67.58
Cattle_6598	bin_29	Eubacteriumsp.AB3007	95.57	4.05	0.42	14,795	2,075,380	394	53.76
Cattle_8907	bin_10	DesulfovibriopigerATCC29098	99.29	4.73	2.23	3,470	2,754,115	71	63.33
Cattle_12947	bin_21	Eubacteriumsp.AB3007	89.23	1.44	0.88	3,445	1,734,734	216	53.88
Cattle_18815	bin_26	Bifidobacteriumpseudolongumsubsp.globosumDSM20092	97.71	1.11	0.23	4085	1694427	70	64
Chicken_11905	bin_10	BarnesiellaintestinihominisYIT11860	99.62	0.00	0.13	340,759	3,165,978	33	43.37
Chicken_11905	bin_30	Phascolarctobacteriumsp.CAG266	99.4	2.32	0.9	63259	2000988	20	45.53
Swine_6347	bin_11	AkkermansiamuciniphilaCAG154	97.96	0.00	62.82	190,018	2,970,703	77	54.93
Chicken_11905	bin_32	LactobacillusjohnsoniiFI9785	99.22	1.36	0.97	7727	1647554	200	34.48
Chicken_11905	bin_2	BacteroidesuniformisCL03T00C23	98.82	3.04	2.32	69,099	4,946,860	142	46.24
Chicken_11905	bin_1	Bacteroidesdorei	97.37	0.97	2.36	42,019	5,727,531	311	41.61
Chicken_11905	bin_5	BacteroidesclarusCAG160	97.31	0.9	0.29	31891	3812621	83	45.18
Chicken_11905	bin_26	LactobacilluscrispatusST1	97.27	4.46	1.56	29680	2194983	251	36.69
Chicken_11905	bin_28	ClostridiumspiroformeDSM1552	97.17	0.94	0.54	25,623	2,171,633	266	28.89
Chicken_11905	bin_3	ParabacteroidesjohnsoniiCAG246	93.85	0	2.99	178043	4482569	74	44.85
Chicken_11905	bin_12	Alistipessp.AL1	91.70	2.36	0.37	109,023	3,089,162	310	58.90
Chicken_16335	bin_1	BacteroidesuniformisATCC8492	97.41	4.17	11.08	72623	4359076	167	46.33
Chicken_17552	bin_1	Alistipessp.AL1	93.28	3.77	2.4	24101	3066870	324	59.62
Turkey_2145	bin_22	LactobacillusvaginalisDSM5837ATCC49540	97.28	0.54	2.14	53398	1821579	128	40.29
Turkey_2145	bin_21	LactobacilluscrispatusST1	95.78	2.03	2.46	4,460	1,826,917	350	37.16
Turkey_2145	bin_9	Prevotellasp.CAG1320	95.44	2.2	0.35	41764	2474073	65	50.43
Turkey_2145	bin_11	Alistipessp.CAG831	93.59	1.9	0.11	36482	2349217	50	55.03
Turkey_2145	bin_29	FirmicutesbacteriumCAG308	89.76	2.92	0.31	17118	1530472	208	35.29
Turkey_4307	bin_20	Prevotellasp.CAG1320	98.07	2.70	0.15	57,484	2,460,410	61	51.42
Turkey_4307	bin_54	FirmicutesbacteriumCAG308	94.88	1.08	0.49	11852	1540726	203	35.34
Turkey_4307	bin_2	BacteroidesplebeiusCAG211	94.28	6.63	0.1	47506	3561263	155	44.78
Turkey_5824	bin_19	Prevotellasp.CAG1320	100	7.24	0.51	46139	2442903	125	51.22
Turkey_5824	bin_1	Prevotellasp.CAG1124	98.75	9.17	2.80	92,522	4,258,228	151	48.07
Turkey_5824	bin_14	DesulfovibriopigerATCC29098	98.11	2.96	0.41	35962	2618703	103	63.57
Turkey_5824	bin_20	Clostridiumsp.CAG678	97.32	0.00	0.52	575,442	2,291,030	43	45.13
Turkey_5824	bin_7	Prevotellasp.CAG1058	97.22	4.94	0.75	44196	3250770	133	48.36
Turkey_5824	bin_4	Bacteroidessp.CAG598	96.33	5.84	0.54	54517	3626107	93	49.17
Turkey_5824	bin_3	BacteroidescoprocolaCAG162	96.17	5.34	1.13	54095	3674258	145	41.1
Turkey_5824	bin_13	Clostridium sp.CAG149	94.48	0.32	0.06	53959	2664873	154	51
Turkey_5824	bin_22	Bifidobacteriumpullorum	92.82	2.58	5.38	17062	2112670	260	64.04
Turkey_5824	bin_27	Collinsellasp.CAG398	90.08	0.54	2.21	16,865	1,994,461	249	62.12
Turkey_8374	bin_24	Alistipessp.N15.MGS157	99.52	0	0.29	83569	2263932	34	56.49
Turkey_8374	bin_49	Corallococcussp.CAG1435	93.15	1.61	0.12	79684	1617032	37	46.31
Swine_1950	bin_65	MethanobrevibactersmithiiATCC35061	100	0	6.9	119650	1601182	48	31.52
Swine_1950	bin_4	Coprobacillussp.3356FAA	97.17	0.94	1.57	186,090	3,786,325	127	31.50
Swine_1950	bin_3	Bifidobacteriumpseudolongumsubsp.pseudolongumDSM20099	96.89	0.15	0.23	96386	1744660	35	63.43

Table 5. Snapshot of new characterized genome

5.6.2 Characterization analysis by animal source

Swine genome discovery shows 660 total MAGs of which 155 MAGs (23.5%) are considered new complete genomes with high confidence (>90% Completeness, < 15% contamination). Actinobacteria (37.5%), Firmicutes (37.5%) and Bacteroidetes (25%) constitute the phyla, while Clostridiales (86.4%) was the most abundant order-taxa. The fecal bacterial diversity in the current study supports what has been shown previously, the swine gut is dominated by Firmicutes (~30%), Bacteroidetes (~50%), and Proteobacteria (~10%) [220]. Turkey shows 551 total MAGs of which 92 (16.7%) MAGs, considered new complete genomes with high confidence (>90% Completeness, < 15% contamination). Bacteroidetes (56.3%) and Actinobacteria (25%) represent the most abundant phyla, while Bacteroidales (50%) and Clostridiales (45.5%) were the most representative order-taxa. Sequencing data characterizing microbiome along the gastrointestinal track in turkey on a genus level, showed an abundance of Lactobacillus 38.2%, Streptococcus 28.1%, and Clostridium 13.0% respectably [221]. Cattle shows 413 total MAGs of which 58 MGAs (14%) are considered new complete genomes with high confidence (>90% Completeness, < 15% contamination). Actinobacteria (44.4%) and Bacteroidetes (44.4%) were the most abundant phyla-taxa while Clostridiales (93%) was the most abundance taxa-order. Study of the gastrointestinal microbiome in dairy cattle observed. Firmicutes (46.5%), Bacteroidetes (33.6%), Proteobacteria (7.5%), and Fibrobacteres (3.0%) as the most abundant [222], In the analysis of the abundance of bacteria, at the phyla level, the dominant bacteria were Firmicutes (47.40 %), Bacteroidetas (16.82 %), and Proteobacteria(1.79 %) [223].

Chicken reveals 156 total MAGs of which, 33 (21.2%) MAGs are considered new complete genomes with high confidence (>90% Completeness, < 15% contamination). Bacteroidetes phyle was discovered (100%) and Clostridiales (90.9%) and Lactobacillales (9.1%) represent the taxa-order. Figure 12 represent the known genome of metagenomic sample and Metagenomic 11905 and assembly genomes discovered. figure 13. Represents total of genome discovery and genome quality. The dominant ceca microbiota in adult laying hens belonged to the phyla Firmicutes and Bacteriodetes [224], similar results were observed Firmicutes, Bacteroides, and Proteobacteria are the most common phyla in the chicken ceca[225]. Interestedly, none of the studies cited addresses the level of genome completeness for taxonomic assignments.

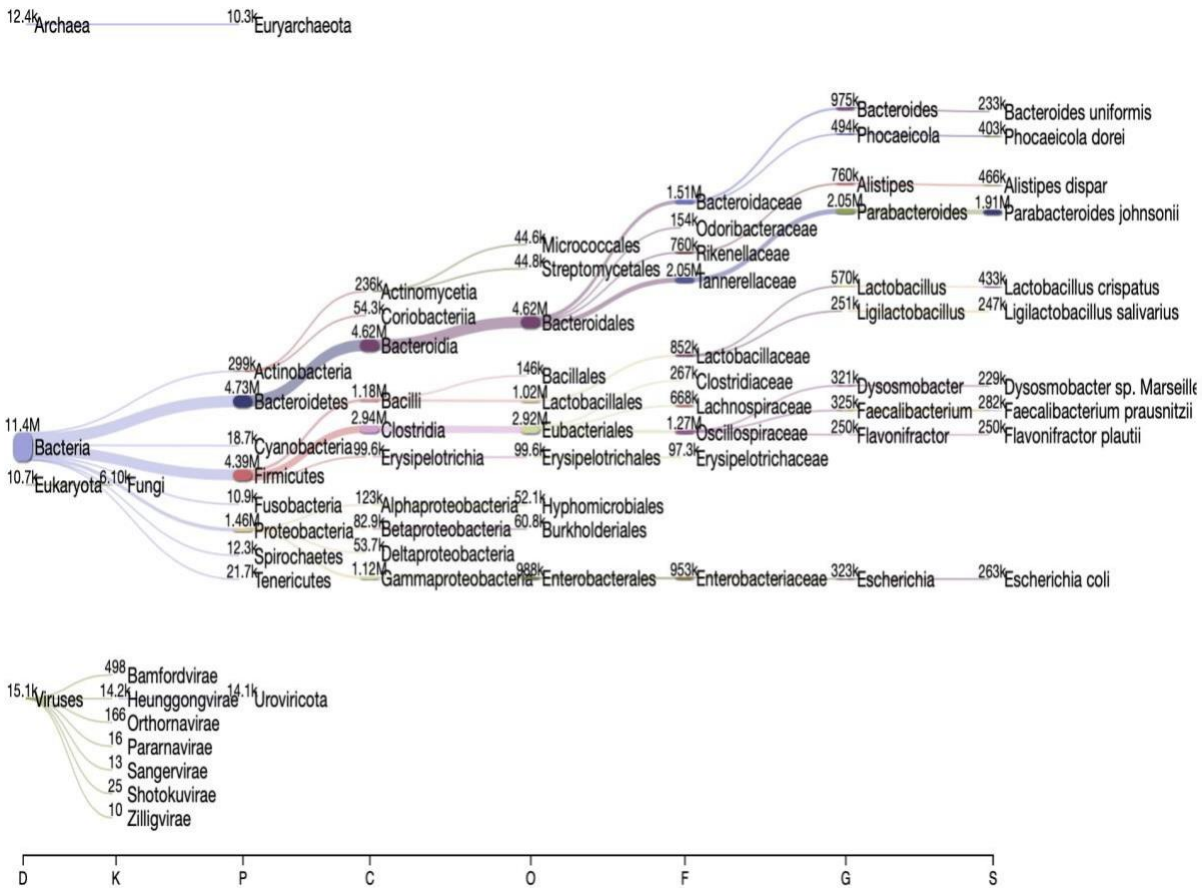


Figure 12 A

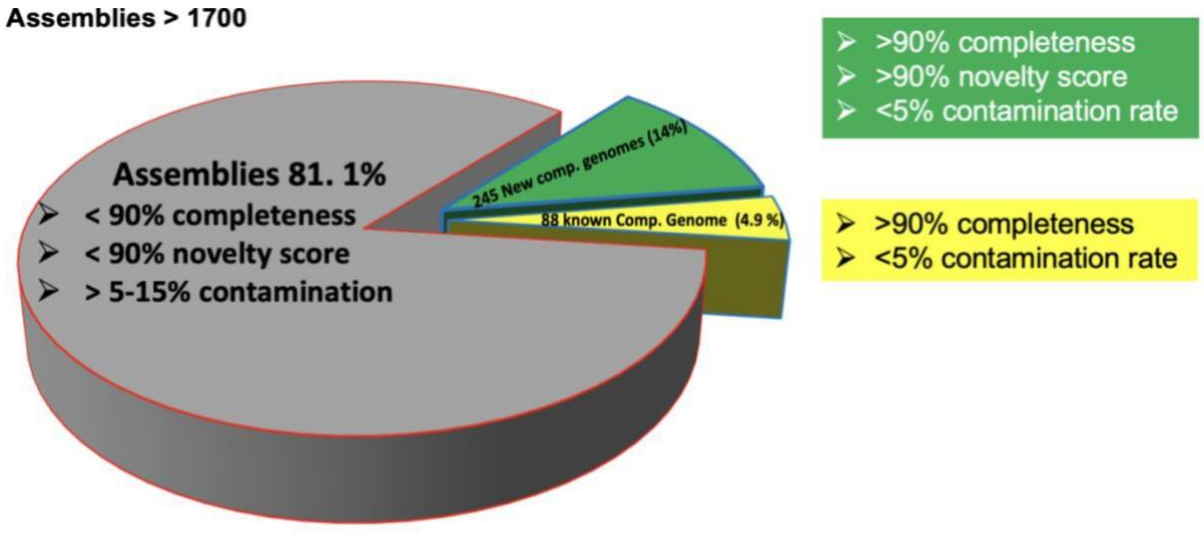


Figure 12 B

Figure 12 [A] represent the known genome of metagenomic sample and Metagenomic 11905 and [B] represent assembly genomes discovered.

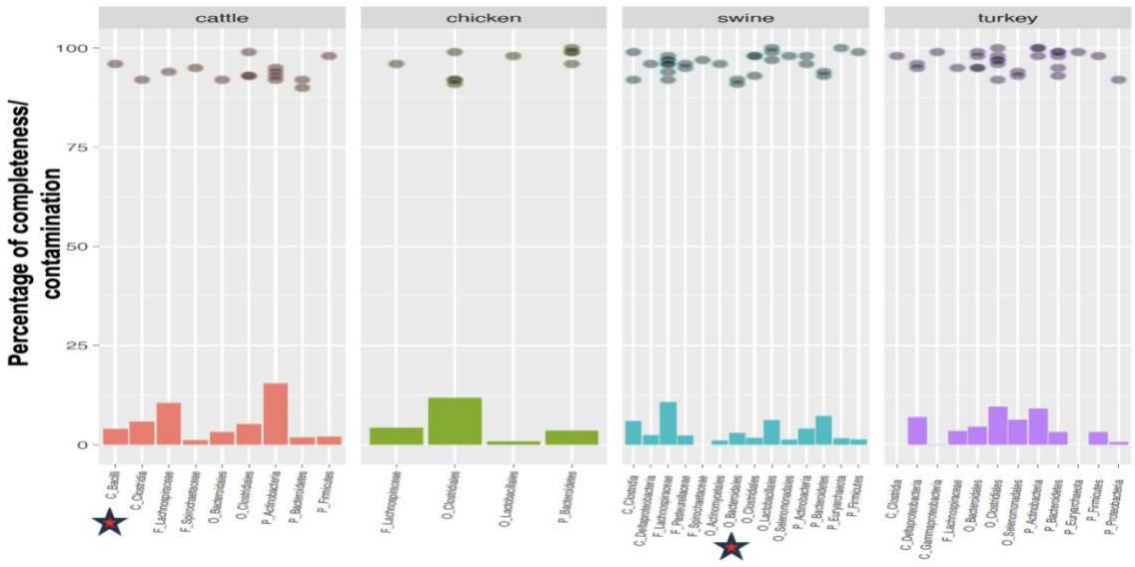


Figure 13 A

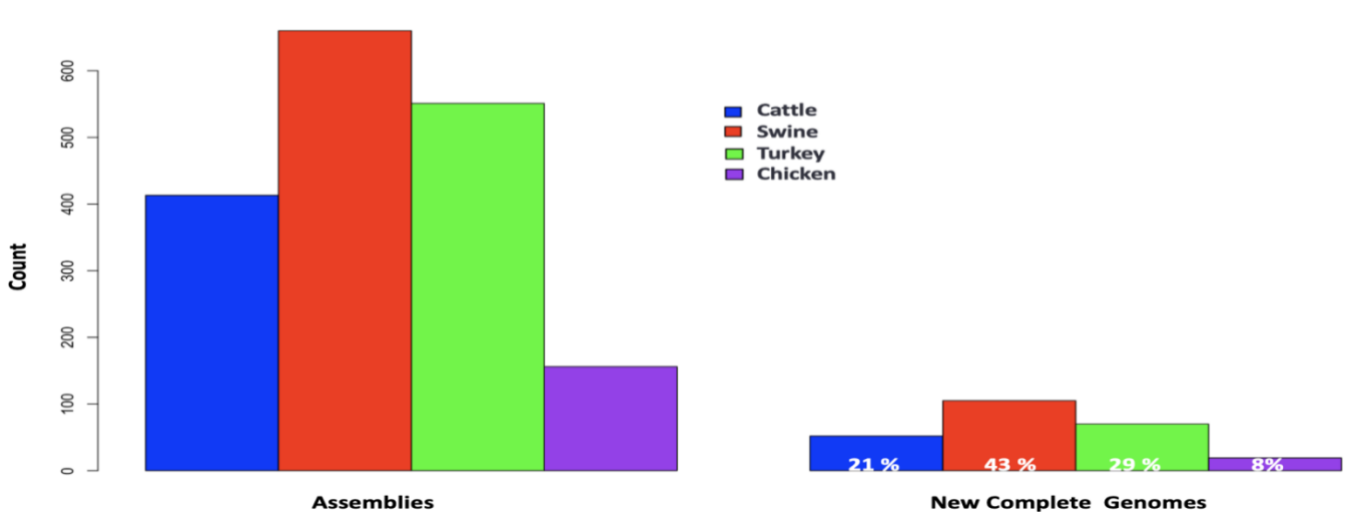


Figure 13 B

Figure 13 [A] represent the checkM completeness and contamination on genome discovery [B] represent total number of assemblies and new metagenomic assembled genomes.

5.6.3 Resistome profiling utilizing Hi-C and shotgun metagenomics by proximity guide assembly.

The presence and the abundance of AMR genes in the metagenomic datasets were determined using the ShortBRED [161]. ShortBRED-Identify was used to generate unique peptide markers for AMR genes-protein sequences compiled from AMRFinderPlus version 3.10 database.

We characterized over 1100 antimicrobial resistance genes with >70% reference sequence coverage from the sample set. 186 unique AMR gene variants, representing 22 AMR classes were found. Aminoglycoside (23%), Tetracycline (19%), Beta-lactam (15%), and Macrolide (10%), were the most abundant AMR classes. While *aph(6)-Id , aph(3')-IIIa , aad9, aadE , aadA1,*

ant(6)-Ia , aadE , aph(2'')-IIa, blaEC , blaTEM-1 , cfxA , blaACI-1, cfxA , blaTEM-1 , blaOXA , blaEC , blaFOX, mef(A) , erm(X) , mef(A) , msr(D) , erm(F) , mef(En2), erm(B) ,erm(Q), tet(W) , tet(Q), tet(32), tet(O) , tet(40), tet(44) , and *tet(L)* were the most abundant AMR genes across the sample set. Table 6.

Cattle						
Total AMR Class	**Total AMRgene**	**Unique AMR gene**	**AMR Classes**	**Gene varients**	**percentage**	**More abundan genes**
18	299	105	AMINOGLYCOSIDE	20	19	aph(6)-Id , aph(3')-IIIa , aad9, aadE , aadA1
			MACROLIDE	12	11	mef(A) , erm(B) , mef(En2) , msr(D)
			BETA-LACTAM	15	14	cfxA , blaTEM-1 , blaOXA , blaEC , blaFOX
			COLISTIN	2	2	mcr-3 , mcr-3.17
			EFFLUX	3	3	mdtM , emrD , acrF
			FOSFOMYCIN	2	2	fosA2 , fosA7.5
			GLYCOPEPTIDE	3	3	vanS-C , vanR-C , vanR-D
			LINCOSAMIDE	6	6	lnu(C) , lnu(AN2) , lnu(G) , lnu(P)
			LINCOSAMIDE/STREPTOGRAMIN	1	1	lsa(E)
			MACROLIDE/LINCOSAMIDE/STREPTOGRAMIN	3	3	cfr(E) , cfr(C) , cfr(B)
			PHENICOL	6	6	floR , cmx
			PHENICOL/QUINOLONE	1	1	oqxA
			QUINOLONE	3	3	qnrB10 , qnrD , qnrB19
			PLEUROMUTILIN	1	1	eat(A)
			STREPTOTHRICIN	2	2	sat2 , sat4
			SULFONAMIDE	3	3	sul1 , sul2 , sul3
			TRIMETHOPRIM	4	4	dfrA1 , dfrA12 ,dfrA14 , dfrA17
			TETRACYCLINE	18	17	tet(W) , tet(Q) , tet(32) , tet(40) , tet(L), tet(O)
Swine						
Total AMR Class	**Total AMRgene**	**Unique AMR gene**	**AMR Classes**	**Gene varients**	**percentage**	**More abundan genes**
20	523	152	AMINOGLYCOSIDE	28	18	ant(6)-Ia , aac(6')-Im , aph(3')-IIIa , aad9 , aph(3'')-Ib
			BETA-LACTAM	29	19	blaEC , blaTEM-1 , cfxA , blaACI-1
			BLEOMYCIN	2	1	bleO , ble
			COLISTIN	2	1	mcr-9.1 , mcr-3
			EFFLUX	6	4	emrD , mdtM , acrF
			FOSFOMYCIN	2	1	fosA7, fosA
			LINCOSAMIDE	9	6	lnu(C) , lnu(AN2) , lnu(A) ,lnu(B) , lnu(G)
			LINCOSAMIDE/STREPTOGRAMIN	2	1	lsa(E) , lsa(A)
			MACROLIDE	13	9	erm(X) , mef(A) , msr(D) , erm(F) , mef(En2), erm(B) ,erm(Q)
			MACROLIDE/LINCOSAMIDE/STREPTOGRAMIN	4	3	cfr(E) , cfr(C) , cfr(B) , cfr
			NITROIMIDAZOLE	1	1	nimA
			PHENICOL	7	5	cmlA1 , floR , cmx ,
			PHENICOL/OXAZOLIDINONE	1	1	optrA
			PHENICOL/QUINOLONE	3	2	oqxA , oqxB
			QUINOLONE	3	2	qnrB19 , qnrS1 , qnrD1
			STREPTOGRAMIN	1	1	vat(E)
			STREPTOTHRICIN	3	2	sat4 , sat2 , sat3
			SULFONAMIDE	3	2	sul3 , sul2 , sul1
			TRIMETHOPRIM	6	4	dfrA12
			TETRACYCLINE	27	18	tet(Q) , tet(W) , tet(M) , tetA(P) , tet(40) , tet(X2), tet(A) , tet(B)
Chicken						
Total AMR Class	**Total AMRgene**	**Unique AMR gene**	**AMR Classes**	**Gene varients**	**percentage**	**More abundan genes**
16	135	66	AMINOGLYCOSIDE	18	27	aph(2'')-Ia , aac(6')-Im , ant(6)-Ia , aph(3')-IIIa
			BETA-LACTAM	6	9	cblA , blaEC , cepA
			BLEOMYCIN	1	2	ble
			EFFLUX	4	6	mdtM , emrD , bexA
			GLYCOPEPTIDE	5	8	vanS-D , vanR-D, vanH-D , vanD , vanX-D
			LINCOSAMIDE	3	5	lnu(C) , lnu(AN2) , lnu(A)
			MACROLIDE	6	9	mef(En2) , erm(G) , mef(A)
			LINCOSAMIDE/STREPTOGRAMIN	1	2	lsa(E)
			PHENICOL	3	5	catA , floR , catA
			QUINOLONE	1	2	qnrB19
			MACROLIDE/LINCOSAMIDE/STREPTOGRAMIN	1	2	cfr(C)
			SULFONAMIDE	2	3	sul1 , sul2
			TRIMETHOPRIM	1	2	dfrA1
			STREPTOGRAMIN	2	3	sat4 , vat€
			TETRACYCLINE	12	18	tet(Q) , tet(W), tet(L) , tet(32) , tet(40)
Turkey						
Total AMR Class	**Total AMRgene**	**Unique AMR gene**	**AMR Classes**	**Gene varients**	**percentage**	**More abundan genes**
15	290	88	AMINOGLYCOSIDE	28	32	aac(6')-Im , aph(3')-Iia , ant(6)-Ia , aadE , aph(2'')-Iia, aad9 , aph(3')-VIIa
			BETA-LACTAM	10	11	blaOXA-347 , cblA , blaEC
			BLEOMYCIN	1	1	ble
			EFFLUX	3	3	mdtM , emrD , acrF
			LINCOSAMIDE	6	7	lnu(AN2) , lnu(C) , lnu(A)
			LINCOSAMIDE/STREPTOGRAMIN	1	1	lsa(E)
			MACROLIDE	9	10	erm(B) , mef(A) , mef(En2) , erm(G) , erm(F)
			MACROLIDE/LINCOSAMIDE/STREPTOGRAMIN	3	3	cfr(E) , cfr(C) , cfr(B)
			NITROIMIDAZOLE	1	1	nimA
			PHENICOL	1	1	catA
			PLEUROMUTILIN	1	1	eat(A)
			STREPTOGRAMIN	1	1	vat(E)
			STREPTOTHRICIN	1	1	sat4
			SULFONAMIDE	2	2	sul1 , sul2
			TETRACYCLINE	20	23	tet(W) , tet(Q) , tet(32) , tet(O) , tet(40) , tet(44) , tet(L)

Table 6. Snapshot of AMR gene profiling

Characterization of AMR genes by animal source shows: Swine resistome profiling showed 152 AMR genes representing 20 AMR classes where Beta-lactam (19%), AMR genes: *blaEC , blaTEM-1 , cfxA , blaACI-1* . Aminoglycoside (18%), *AMR* genes: *ant (6)-Ia , aac(6')-Im , aph(3')-IIIa , aad9 , aph(3")-Ib,* and Tetracycline (18%), AMR genes: *tet(Q) , tet(W) , tet(M) , tetA(P) , tet(40) , tet(X2), tet(A) , tet(B)* were the most abundant. Studies on the effects of antimicrobial and heavy metal usage on the antimicrobial resistance profiles of pigs reported that genes encoding tetracycline and macrolide resistance were the most abundant resistance determinants [70, 177, 220].

Turkey AMR profiling shows Aminoglycoside (32%) AMR genes: *aac(6')-Im , aph(3')-Iia , ant(6)-Ia, aadE , aph(2")-Iia, aad9 , aph(3')-VIIa*. Tetracycline (23%), AMR genes: *tet(W) ,tet(Q) , tet(32) , tet(O) ,tet(40) ,tet(44) , tet(L).* Beta-lactam (11%) and AMR genes: *blaOXA-347, cblA, blaEC*. *Microlide* (10%), AMR genes: *erm(B), mef(A), mef(En2),erm(G), erm(F)* were the most representative. Investigation to assess the epidemiology and genetic heterogeneity of *C. jejuni* recovered from commercial turkey shows the genes for resistance to ampicillin (*bla*OXA), tetracycline [*tet*(O)], neomycin [*aph*(3')-IIIa], streptomycin (*aadE*) and streptothricin (*sat4*) as most prevalent, as well as a gene cluster comprising the genes *sat4, aph*(3')-IIIa and *aadE* was present[226].

The Cattle resistome profile was dominated by Aminoglycoside (19%), AMR genes: *aph(6)-Id , aph(3')-IIIa, aad9, aadE , aadA1*. Tetracycline (17%), AMR genes: *tet(W), tet(Q), tet(32) , tet(40), tet(L), tet(O).* Beta-lactam (14%), AMR genes: *cfxA, blaTEM-1, blaOXA, blaEC , blaFOX*. Macrolide (11%), AMR genes: *mef (A, erm(B), mef(En2) , msr(D)*.

Deep sequencing of the gut microbiomes of swine and cattle shows aminoglycoside, β-lactam, lincosamide, streptogramin, and tetracycline were the prevalent resistance determinants in both swine and cattle. *tetQ, tetW, tetO, tet32*, and t*et44* were the 5 most abundant and prevalent tetracycline resistance genes, while *ANT(6) and* aph(3") were the dominant aminoglycoside AMR genes in swine [227].

The Chicken resitome profile ‘s most representative AMR classes and genes were: Aminoglycoside (27%): *aph(2")-Iia , aac(6')-Im , ant(6)-Ia , aph(3')-IIIa.* Tetracycline (18%): *tet(Q), tet(W), tet(L) , tet(32) , tet(40).* Lincosamide/Streptogramin (9%): *mef(En2) , erm(G) , mef(A),* and beta-lactam (9%) genes: *cblA , blaEC , cepA.* Even though the use of antimicrobials as growth promoters in veterinary medicine has been regulated in developed countries, undeveloped countries continue wide administration for therapeutic purposes and growth promotion in poultry [228, 229]. Gut resistome profiling of broiler chickens infected withmultidrug-resistant Escherichia coli, shows 26.1% aminoglycoside, 15.9% tetracycline, and 9.6% macrolide-lincosamide-streptogramin the most prevalent AMR classes[230].

We characterized within the animal resistome 58 genes-stress factors, metal, heat, biocide and acid across the sample set. Antimicrobial stress agents may amplify bacterial resistance to a wide variety of antibiotics [231]. Antimicrobial stresses, including soil salinity and water pollutants, can affect AMR in soils, which in turn reduces the yield and quality of agricultural products [232]. The US Centers for Disease Control and Prevention (CDC) (https://www.cdc.gov/drugresistance/biggest-threats.html) reported that antibiotic-resistant diseases affect approximately 8.2 million Americans every year. While WHO Antimicrobial

Resistance (https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance) considers that one of the top ten worldwide healthcare issues is AMR.

The overuse of antimicrobials in the agriculture has raised concerns regarding the effects of antimicrobial and the threat of AMR dissemination within different environments [233]. The enhanced selective pressure, helps accelerated the evolution of resistant bacteria[234]. Adding penicillin (100 mg kg−1 soil) to the soil decreased the probiotic Lactobacillus and increased enterobacteria and bacteroid, which are resistant to penicillin in the ileum of mice [235]. Figure 14 represents resistome and Biocide factors. Figure 15 Number of AMR gene variant by AMR class.

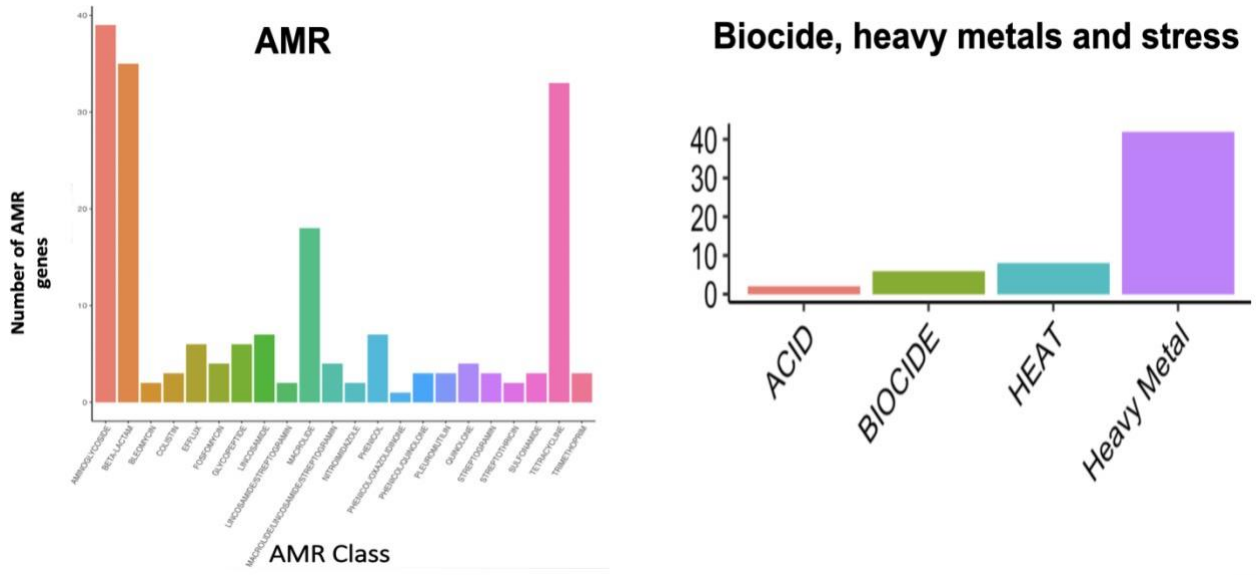


Figure 14 represents resistome and Biocide factors.

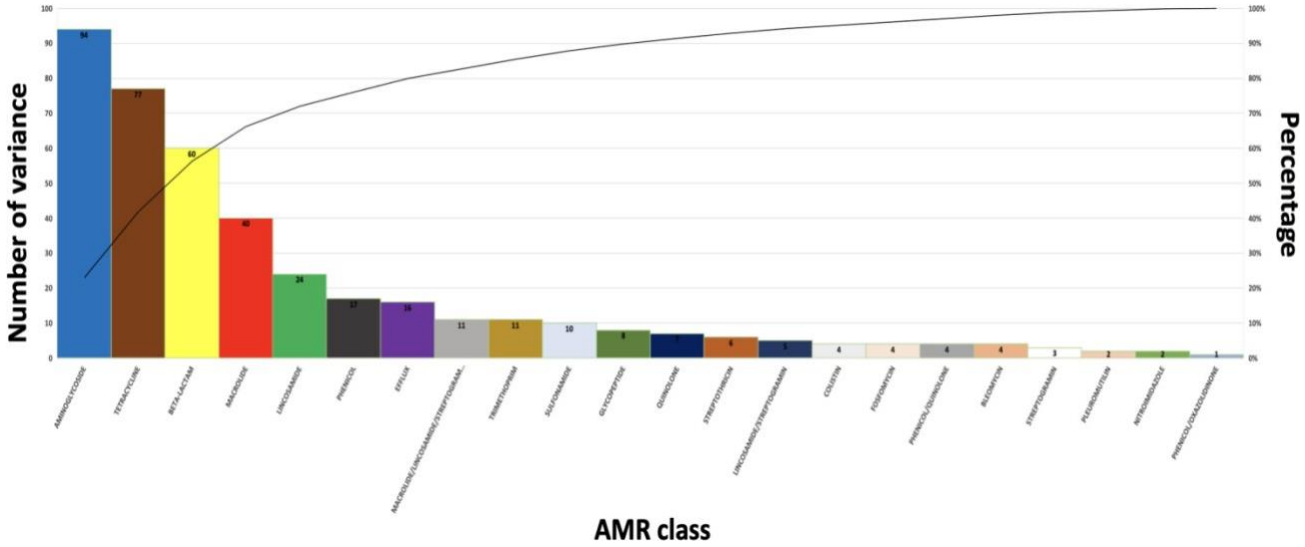


Figure 15 Number of AMR gene variant by AMR class.

5.6.4 Plasmidome profiling utilizing Hi-C and shotgun metagenomics by proximity guide assembly.

We observed over 14000 contig-plasmids with different levels of confidence based on CheckM. PLDS plasmid data base [210] was used. Plasmid characterization using Hi-C and shotgun metagenomics allowed identification of 146 plasmids (>= 85% completeness, >= 90% reference sequence similarity), and over 13000 plasmid-contigs (<85% completeness, < 90% reference sequence similarity). Table 7. Snapshot plasmidome characterization.

Contig_id	MetaSample_id	Animal Source	Length	Aligned_length	aligned %	similarity %	Plasmid_id	Reference_length	Completeness %	Contamination %	Unaligned %
k141_2182	4959	Cattle	5167	5165	99.96	99.9	NC_010375.1	5031	99.98	0	0.04
k141_1006369	4959	Cattle	1829	1781	97.38	98.18	NZ_CP056765.1	1697	98	0	2.62
k141_248689	4959	Cattle	4406	3734	84.75	97.97	NZ_LT985290.1	4074	91.68	0	15.25
k141_184409	4959	Cattle	6899	6897	99.97	99.32	NC_023328.1	6822	99.99	0	0.03
k141_953344	4959	Cattle	6042	6041	99.98	99.5	NZ_CP027259.1	6812	88.7	0	0.02
k141_349370	4959	Cattle	1218	1128	92.61	98.86	NZ_LT968700.1	1296	87.04	0	7.39
k141_58672	6598	Cattle	4543	4541	99.96	98.31	NZ_CP026411.1	4613	98.37	0	0.04
k141_1008299	6598	Cattle	40159	32769	81.6	96.47	NZ_AP022265.1	37223	87.58	0	18.4
k141_1008306	6598	Cattle	4508	4394	97.47	99.89	NZ_CP017027.1	4454	95.49	0	2.53
k141_282285	8907	Cattle	6681	6133	91.8	99.95	NZ_MH229772.1	6961	88.11	0	8.2
k141_170741	8907	Cattle	6856	6739	98.29	98.22	NC_019078.1	6952	94.92	0	1.71
k141_1565840	12947	Cattle	1404	1402	99.86	99.86	NC_016841.1	1308	99.92	0	0.14
k141_637627	12947	Cattle	2187	2185	99.91	97.31	NZ_AP022132.1	2049	99.95	0	0.09
k141_493893	12947	Cattle	1948	1946	99.9	99.59	NZ_CP054386.1	1822	99.95	0	0.1
k141_339967	12947	Cattle	1544	1512	97.93	98.68	NZ_CP051724.1	1538	97.79	0	2.07
k141_656396	12947	Cattle	2986	2736	91.63	98.98	NZ_CP041418.1	3214	85.13	0	8.37
k141_2065499	12947	Cattle	2874	2872	99.93	98.3	NZ_AP019859.1	2732	99.96	0	0.07
k141_2068215	12947	Cattle	4774	4695	98.35	99.94	NZ_MG649064.1	4773	98.37	0	1.65
k141_87608	12947	Cattle	3866	3101	80.21	99.84	NZ_CP035348.1	3103	99.97	0	19.79
k141_1338161	12947	Cattle	2583	2441	94.5	99.51	NZ_CP036185.1	2444	99.96	0	5.5
k141_1051314	12947	Cattle	1512	1510	99.87	99.87	NZ_CP035726.1	1506	99.93	0	0.13
k141_1055571	12947	Cattle	4331	4330	99.98	99.95	NZ_CP043668.1	4593	94.27	0	0.02
k141_2491131	12947	Cattle	3478	2867	82.43	97.52	NZ_KY978588.1	3337	85.92	0	17.57
k141_2491323	12947	Cattle	4379	4368	99.75	99.45	NZ_CP008717.1	4237	99.98	0	0.25
k141_2491380	12947	Cattle	1381	1379	99.86	95.22	NZ_CP058145.1	1240	99.92	0	0.14
k141_2491388	12947	Cattle	2432	2103	86.47	93.92	NZ_CP041446.1	2137	99.91	0	13.53
k141_801668	2145	turkey	45147	551	1.22	100	NZ_AJ223173.1	552	99.82	0	98.78
k141_748597	2145	turkey	3517	2565	72.93	97.7	NZ_FN597255.1	2797	91.53	0	27.07
k141_298266	2145	turkey	30206	4011	13.28	99.1	NC_019049.1	4012	99.98	0	86.72
k141_197805	2145	turkey	2758	1906	69.11	97.27	NZ_CP051612.1	1907	99.95	0	30.89
k141_167840	2145	turkey	4864	1078	22.16	99.72	NZ_CP059912.1	1079	99.91	0	77.84
k141_610719	5824	turkey	3199	2243	70.12	98.4	NZ_CP010877.1	1506	99.93	0	29.88
k141_608498	8374	turkey	3519	2417	68.68	97.58	NZ_CP035326.1	3019	99.8	0	31.32
k141_784427	8374	turkey	16853	4011	23.8	99.23	NC_019049.1	4012	99.98	0	76.2
k141_337586	13145	turkey	3075	1975	64.23	95.65	NZ_CP040583.1	1925	96.21	0	35.77
k141_332399	16335	Chicken	2702	2696	99.78	99.74	NZ_CP046339.1	2579	99.96	0	0.22
k141_135743	16335	Chicken	42614	41557	97.52	97.76	NZ_CP059935.1	42400	97.99	0	2.48
k141_261262	16335	Chicken	3525	3523	99.94	99.15	NZ_CP057103.1	3555	98.93	0	0.06
k141_395889	16335	Chicken	2925	2924	99.97	98.06	NC_019075.1	2779	99.96	0	0.03
k141_395905	16335	Chicken	4961	4960	99.98	98.71	NZ_CP034822.1	9606	99.99	0	0.02
k141_284873	11905	Chicken	1647	1508	91.56	100	NZ_CP023506.1	1509	99.93	0	8.44
k141_284903	11905	Chicken	1698	1696	99.88	99.82	NZ_CP050377.1	1565	99.94	0	0.12
k141_254827	11905	Chicken	5931	5929	99.97	99.41	NC_017655.1	5954	99.75	0	0.03
k141_567624	11905	Chicken	2230	2228	99.91	99.73	NZ_CP048348.1	2088	99.95	0	0.09
k141_252475	17552	Chicken	1871	1505	80.44	99.87	NZ_CP035726.1	1506	99.93	0	19.56
k141_472061	17552	Chicken	2230	2216	99.37	99.33	NZ_CP058665.1	2101	99.95	0	0.63
k141_472134	17552	Chicken	2237	2235	99.91	100	NZ_CP059921.1	2096	99.95	0	0.09
k141_472187	17552	Chicken	1889	1887	99.89	99.74	NZ_CP047465.1	1749	99.94	0	0.11
k141_100703	1950	Swine	7102	7100	99.97	99.96	NZ_MH229772.1	6961	99.99	0	0.03
k141_492516	1950	Swine	4361	4302	98.65	99.26	NZ_CP041398.1	4148	99.98	0	1.35
k141_639841	1950	Swine	2109	1774	84.12	95.89	NZ_CP055813.1	1902	92.95	0	15.88
k141_250882	1950	Swine	42502	39757	93.54	99.2	NZ_CP025711.1	41544	95.34	0	6.46
k141_350447	1950	Swine	1662	1660	99.88	99.28	NZ_CP044966.1	1541	99.94	0	0.12
k141_550839	1950	Swine	4532	4530	99.96	99.25	NZ_CP053241.1	5237	86.54	0	0.04
k141_306396	1950	Swine	46811	45263	96.69	99.9	NC_010937.1	47263	95.77	0	3.31
k141_701381	1950	Swine	31374	31311	99.8	99.19	NZ_CP040529.1	33036	94.91	0	0.2
k141_268296	1950	Swine	2210	2089	94.52	99.81	NZ_CP050362.1	2101	99.43	0	5.48
k141_274746	1950	Swine	5698	5696	99.96	99.77	NZ_CP040632.1	5595	99.98	0	0.04
k141_423200	1950	Swine	5630	5452	96.84	94.97	NZ_CP050056.1	4786	99.98	0	3.16
k141_277761	1950	Swine	3559	3557	99.94	99.44	NZ_CP057103.1	3555	99.97	0	0.06
k141_782695	1950	Swine	4005	4003	99.95	99.75	NZ_CP053243.1	3863	99.97	0	0.05
k141_782703	1950	Swine	3492	3490	99.94	99.91	NZ_CP044964.1	3353	99.97	0	0.06
k141_782715	1950	Swine	7142	7140	99.97	99.89	NZ_KJ830769.1	6994	99.99	0	0.03
k141_782784	1950	Swine	1516	1448	95.51	96.91	NC_004345.1	1338	97.91	0	4.49
k141_782833	1950	Swine	2893	2605	90.04	99.58	NZ_CP044099.1	2604	99.96	0	9.96

Table 7. Snapshot plasmidome characterization.

5.6.5 Linking AMR genes to host bacteria by proximity guide assembly.

For every AMR with >70% reference sequence coverage from the sample set, we considered all Hi-C linkage with other contigs corresponding to bacteria taxa rank present in that cluster. Hi-C contacts were normalized based on their abundance within the genome cluster. Phylogenetic analysis of the genomes linked to AMR shows linkage between different AMR genes and bacterial genomes at different taxonomic ranks. Over 400 antimicrobial resistance genes representing 22 antimicrobial classes were identified and include:

aminoglycoside (40 gene variants), beta-lactam (37 gene variants), bleomycin (2 gene variants), colistin (3 gene variants), efflux (6 gene variants), fosfomycin (4 gene variants), glycopeptide (6 gene variants), lincosamide (9 gene variants), lincosamide/streptogramin (2 gene variants), macrolide (16 gene variants), macrolide/lincosamide/streptogramin (4 gene variants), nitroimidazole (1 gene variant), phenicol (9 gene variants), phenicol/oxazolidinone (1 gene variant), phenicol/quinolone (2 gene variants), pleuromutilin (1 gene variant), quinolone (5 gene variants), streptogramin (1 gene variant), streptothricin (3 gene variants), sulfonamide (3 gene variants), tetracycline (25 gene variants), and trimethoprim (8 gene variants). Figure 16. represent the linking of AMTR gens at different bacteria Order. Phylogenetic analysis shows linkage between different AMR genes and genomes at different bacteria taxonomic ranks. Clostridiales, Bacteridales and Lactobaciliales order shows to be prone to host 13 AMR classes including: macrolide, aminoglycoside, lincosamide, beta-lactam, streptogramin, tetracycline, trimethoprim, metronidazole, lincosamide/streptogramin, and macrolide/lincosamide/streptogramin. Enterobacteriaceae family that includes a number of pathogens such as Klebsiella, Enterobacter, Citrobacter, Salmonella was associate with different AMR genes including, beta-lactamase *blaEC*, aminoglycoside *APH(3'')-Ib* an, multidrug transporter from the Major Facilitator Superfamily (MFS) EmrD. AMR genes*: lnu(C, aph(2'')-Ig, erm(B), mef(A), tet(W, tet(32, aph(2'')-IIa, aph(3')-IIIa, ant(6)-Ia, Sat 4, blaEC* were the most abundance.

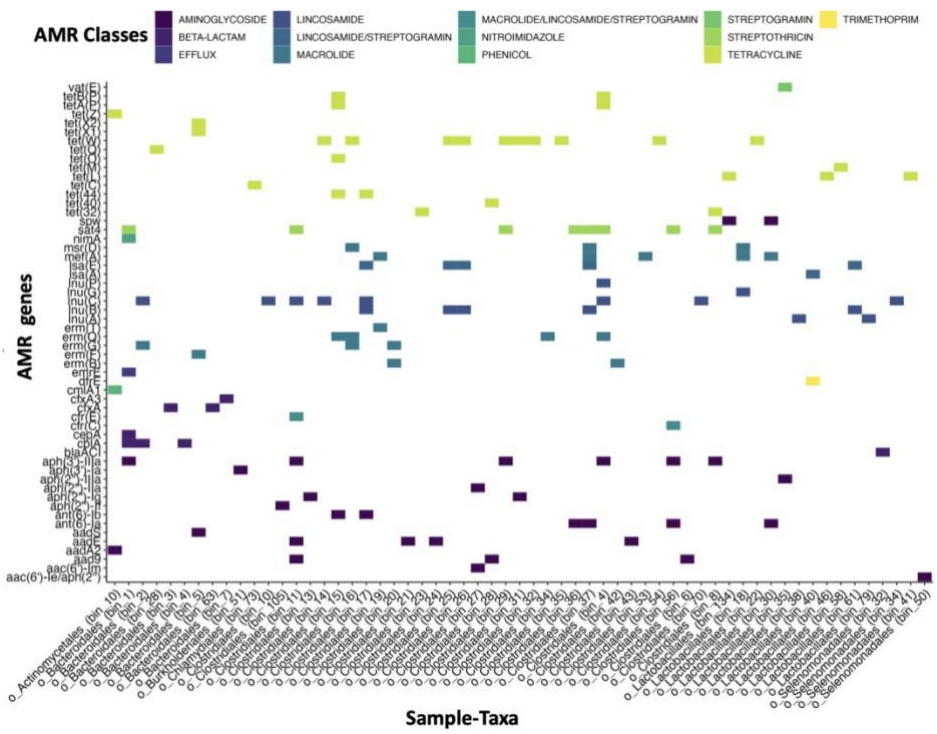


Figure 16. represent the linking of AMTR gens at different bacteria Order

5.6.6 Linking AMR genes to plasmids by proximity guide assembly.

We followed the same criteria used to link AMR genes to host bacteria thus, for every AMR with >70% reference sequence coverage from the sample set, were considered all Hi-C linkages (> 3 contacts), with other contigs corresponding to plasmid-contigs present in that cluster. We identified that over 60% of total AMR were originated from plasmids.

149 plasmids had a higher level of confidence (>= 85% completeness, >= 90% reference sequence similarity). 67 plasmid-contigs were identified with >= 70 < 85% completeness, >= 90% reference sequence similarity. Overall, 13000 + plasmid-contigs were identified with <=69% completeness, < 90% reference sequence similarity. Some higher level of confidence plasmid-AMR contig linkages found are as follows:

Plasmid NZ_CP040912.1: Acinetobacter baumannii strain 29FS20 plasmid p29FS20-1 complete sequence containing aminoglycoside AMR-contig gene *aac(3)-IId* and beta-lactamase AMR-contig gene *blaTEM-1.*

Plasmid NZ_CP018901.1: Campylobacter coli strain YH502 plasmid pCOS502 complete sequence containing aminoglycoside AMR-contig gene *aph(3')VIIa* and lincosamide *lnu(C).*

Plasmid NZ_MG591698.1: Escherichia coli strain PN43 plasmid unnamed, complete sequence containing lincosamide *lnu(B)* and *aac(6')-Isa.*

Plasmid NZ_CP047617.: Lactococcus raffinolactis strain Lr_19_5 plasmid pLraf_19_5_1, complete sequence containing macrolide *erm(G), met (A),* and *msr(D).*

Plasmid NZ_LR135483.1. Enterococcus faecium isolate E4456 plasmid 2 containing aminoglycoside *aph(3')IIIA* ,and streptothricin acetyltransferase *SAT-4*

Plasmid NZ_CP029755.1. Lactobacillus amylovorus strain PMRA3 plasmid pPMRA301, sequence containing aminoglycoside, complete aadD1, aac(6) and lincosamide lnu(C).

Plasmid NZ_KY991369.1. Salmonella enterica subsp. enterica serovar Anatum strain CVM N57952F plasmid N57952F, complete sequence containing quinolone qnrB10. Linking AMR genes to Plasmid (>85% completeness, > 90% reference sequence similarity. Figure 17. Linking AMR genes to Plasmid (>85% completeness, > 90% reference sequence similarity

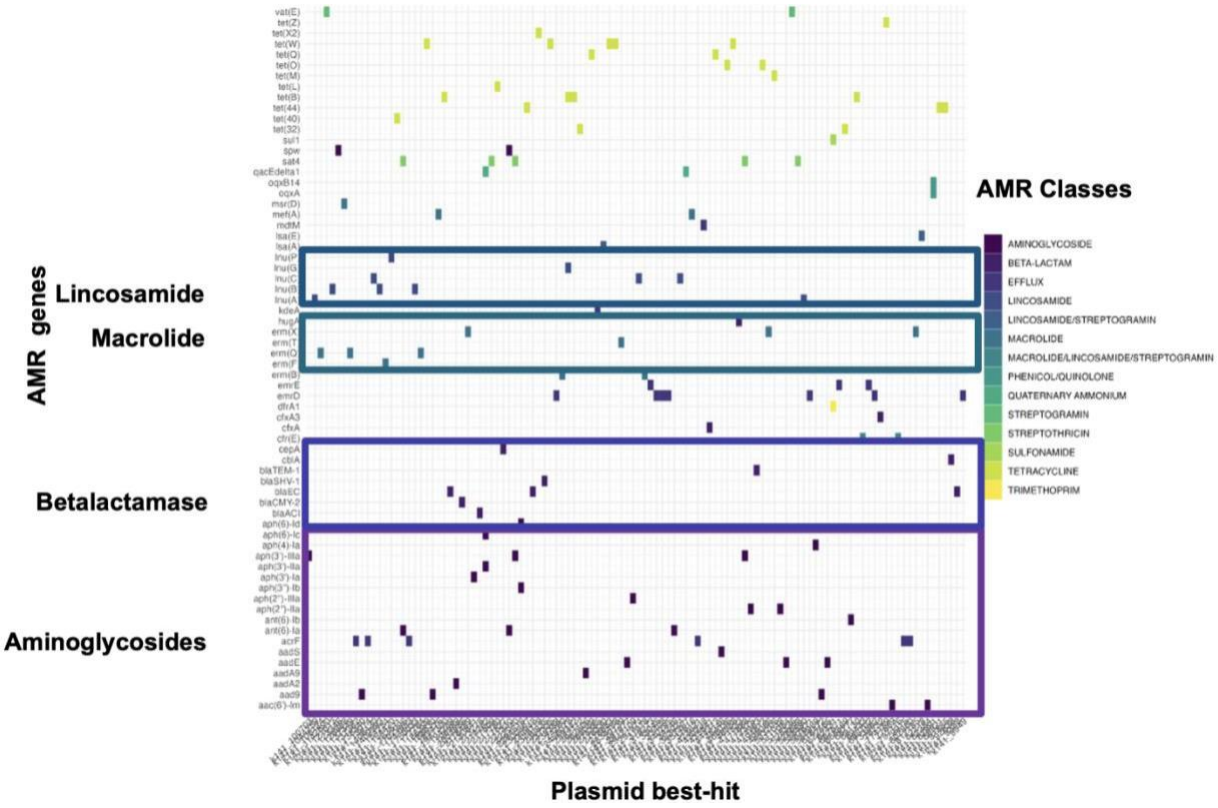


Figure 17. Linking AMR genes to Plasmid (>85% completeness, > 90% reference sequence similarity**)**

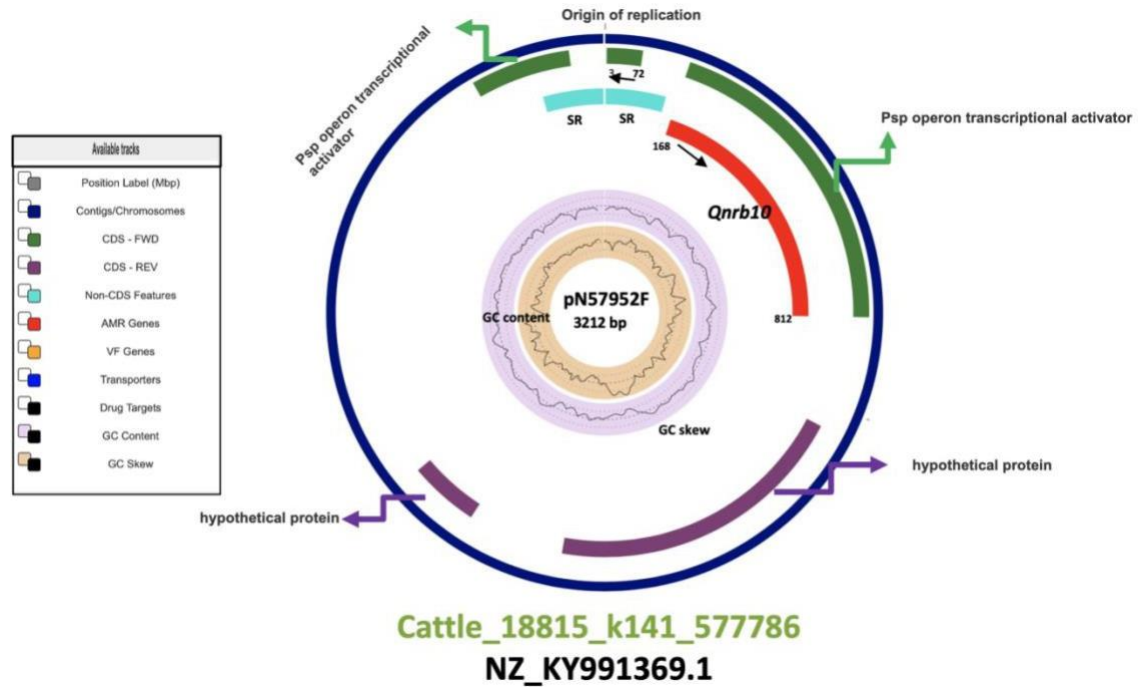


Figure 18. AMR gene *Qnrb10 linked to* Circular plasmid pN57952F (>85% completeness, > 90% reference sequence similarity**)**

5.6.7 Linking AMR genes and plasmid to host bacteria

Resistance plasmids were linked to AMR genes and host bacteria. We generated assemblies from high sequence depth coverage (>100 million reads per sample) shotgun metagenomics, bin3C- modified to cluster metagenomic contigs into putative metagenomic assembly genomes MAG and R for visualization. All clustered contigs were blasted against the resistome and taxonomic profile generated from shotgun metagenomics data and to the plasmid database to retrieve AMR gene frequency, plasmidome contents and bacterial taxonomic ranks. The sum of interactions was normalized by the number of contigs that all clusters have. Only clusters with interactions between AMR genes, plasmids and bacterial taxa were considered for final linking, excluding interaction with less than two contacts. We generated over 1700 assamblies with thousands of resistance plasmid-contigs, AMR-contigs and bacteria taxonomic-contigs with different levels of confidence base on CheckM, to explain the host bacteria proficiency to uptake and exchange AMR genes within the microbiome. We linked over 600 AMR genes to plasmid and host bacteria at different confident levels (completeness and contamination). 80 links were stablished by Hi-C with high levels of confidence (AMR > 70% coverage and Plasmid >85% completeness and <15% contamination). Similar study reconstructing a known plasmid-host association from a wastewater sample, shows the ability of Hi-C to assemble the genome of a completely sequenced plasmid-bearing bacterium from a wastewater metagenome[44], leveraged Hi-C data generated from a synthetic metagenome sample to accurately cluster MAGs into groups that contain nearly complete genomes and show the plasmid association with the chromosomes of their host[30]. Figure 19 . represents the Hi-C contact map to create linkage clusters.

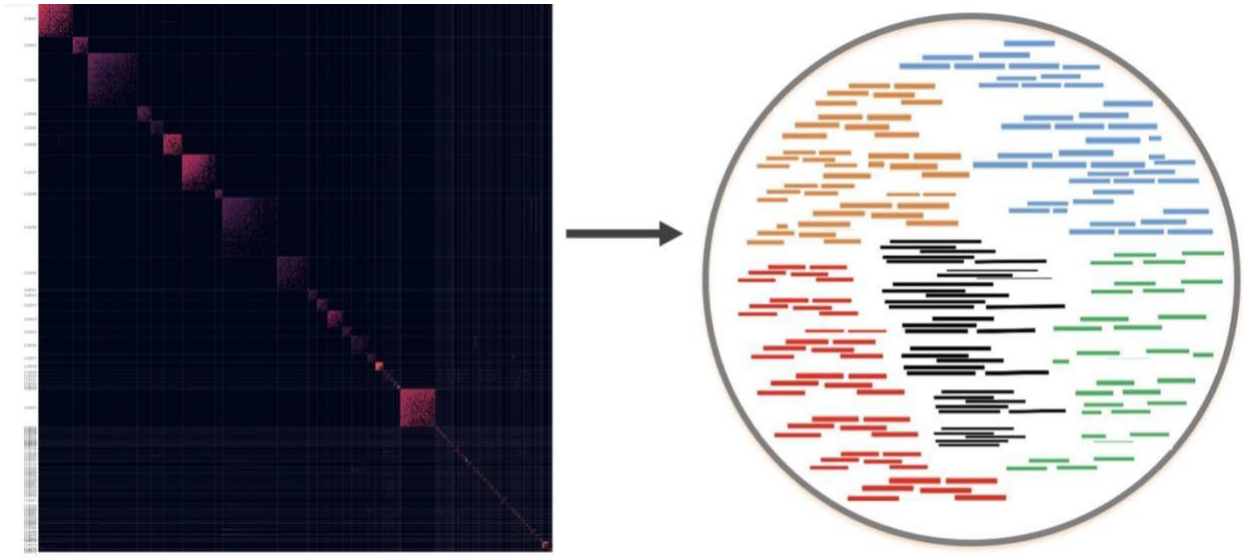


Figure 19 . Hi-C contact map to create linkage clusters metagenomic sample 1409

Other studies described the use of Hi-C to perform genome scaffolding in eukaryotic genomes, by inferring chromosomal groupings of contigs and then ordering sequences along the chromosome [132] and then use of Hi-C data to phase haplotypes in a diploid organism [236]. Plasmid NZ_CP013691.1: Myroides odoratimimus strain PR63039 plasmid p63039, complete sequence. Myroides are broadly resistant to β-lactams, including carbapenems, with variable susceptibility to aminoglycosides, quinolones, and sulfamethoxazole. We found AMR gene *aadS* (*ANT (6)*), a novel Tn4551 streptomycin-resistance gene that was phenotypically silent in wild-type Bacteroides [237]. *aadS* is an Aminoglycoside microbial resistance gene with a perfect resistome match to Bacteroides fragilis, Bacteroides ovatus, Bacteroides thetaiotaomicron, Elizabethkingia anophelis, Myroides odoratimimusg+p, Myroides phaeus, Parabacteroides distasonis, Riemerella anatipestifer, Sphingobacterium hotanense, linked Bacteroides taxonomic rank.

Plasmid NZ_AP022193.1: Klebsiella sp. WP7-S18-CRE-02 plasmid pWP7-S18-CRE-02_4, complete sequence. Klebsiella is β-lactam resistance that was linked to spew(ANT(9)-Ia) an Aminoglycoside resistance gene with perfect resistome match to Bacillus subtilis, Bacillus velezensis, Enterococcus avium, Enterococcus faecalis, Enterococcus faecium, Enterococcus hirae, Staphylococcus aureus, Staphylococcus capitis, Staphylococcus epidermidis, Staphylococcus haemolyticus, Staphylococcus lugdunensis, Staphylococcus simulans [238-240] linked to Chlamydiales taxonomic rank.

Plasmid NZ_LR135334.1: Enterococcus faecium isolate E7471 plasmid 4 Enterococcus Resistance to penicillin and ampicillin, was liked to Aminoglycoside resistance gene APH(2'')-IIa, resistome with sequence variants Campylobacter coli, Citrobacter freundii, Enterococcus faecium, Klebsiella pneumoniae and mapped to Clostridiales taxa-rank.

Plasmid NZ_MG957432.1: Enterococcus faecium strain 37BA plasmid pEf37BA was linked to Lincosamide resistance gene lnuC, resistome with sequence variants Brachyspira pilosicoli, Campylobacter jejuni, Enterococcus faecalis, Eubacterium maltosivorans, Fusobacterium nucleatum, Glaesserella parasuis, Lactobacillus crispatus, Lactobacillus johnsonii, Ligilactobacillus animalis, Megasphaera stantonii, Streptococcus agalactiaeg, Streptococcus lutetiensis, Streptococcus mitis, Streptococcus pasteurianus, Streptococcus suis , linked to Lactobacilliales taxa rank.

Plasmid NZ_CP010147.1: Escherichia coli strain D5 plasmid B, complete sequence, was linked to Crf multidrug resistance gene (oxazolidinone antibiotic, streptogramin antibiotic, phenicol antibiotic, lincosamide antibiotic) [241-246] and also to APH(3')-III Aminoglycoside antimicrobial class [247, 248], linked to Clostridiales taxa rank. Linking resistome and

plasmidone to host bacteria with higher levels of confidence - AMR-genes > 70% completeness, Plasmid >85% completeness and <15% contamination. Figure 20 represents a snapshot of linking AM gene and plasmid to host bacterial cell at Family level.

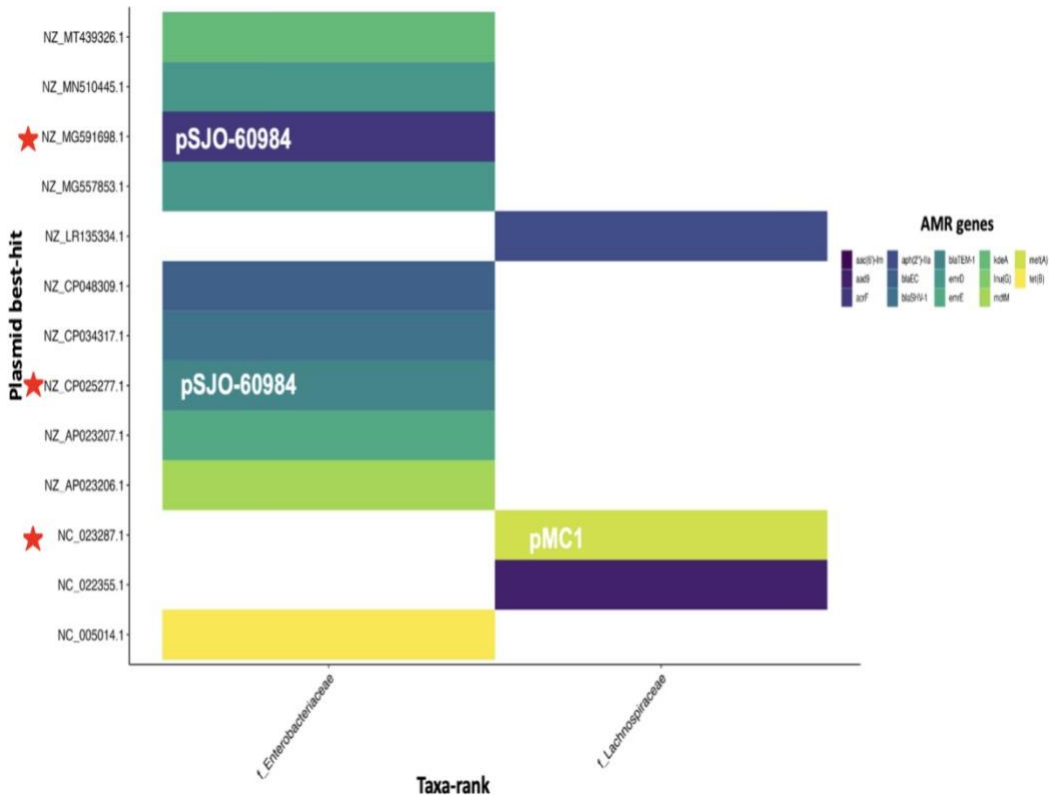


Figure 20. represents a snapshot of linking AMR gene and plasmid to host bacterial cell at Family level.

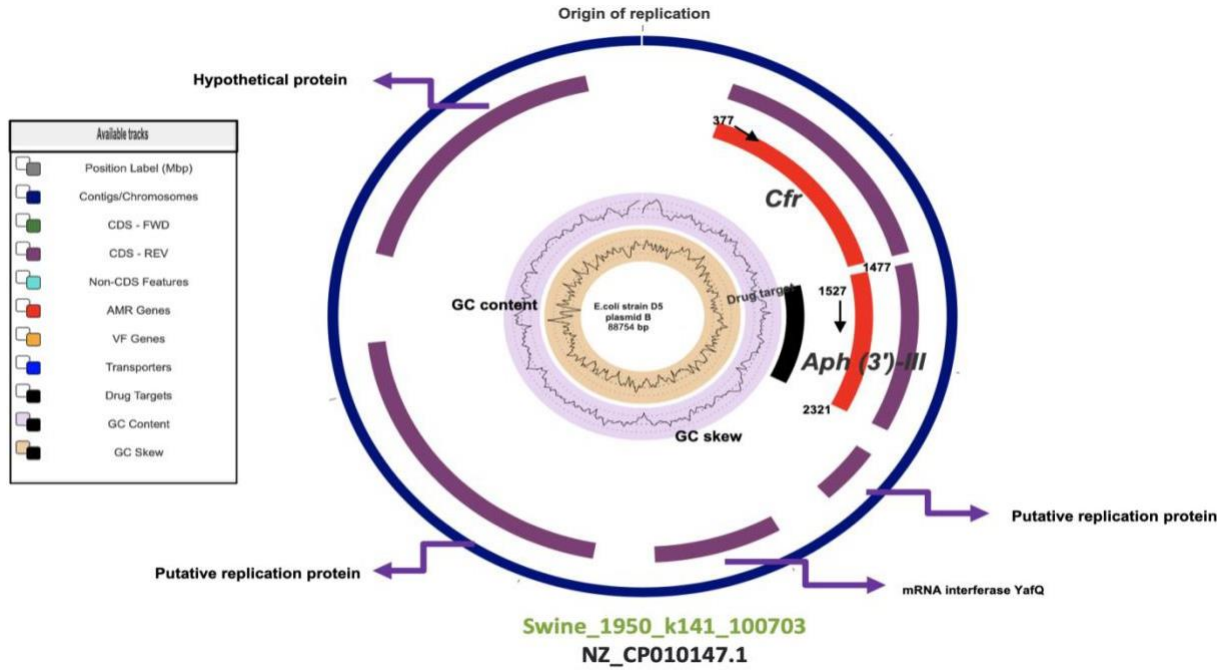


Figure 21. Linked AMR genes (Cfr and Aph(3')-III) and plasmid B to host bacterial Clostridiales_(bin_56)

Figure 22 show resistome analysis at food animal source and Figure 23 show linking AMR genes and plasmid at Order level by food animal source.

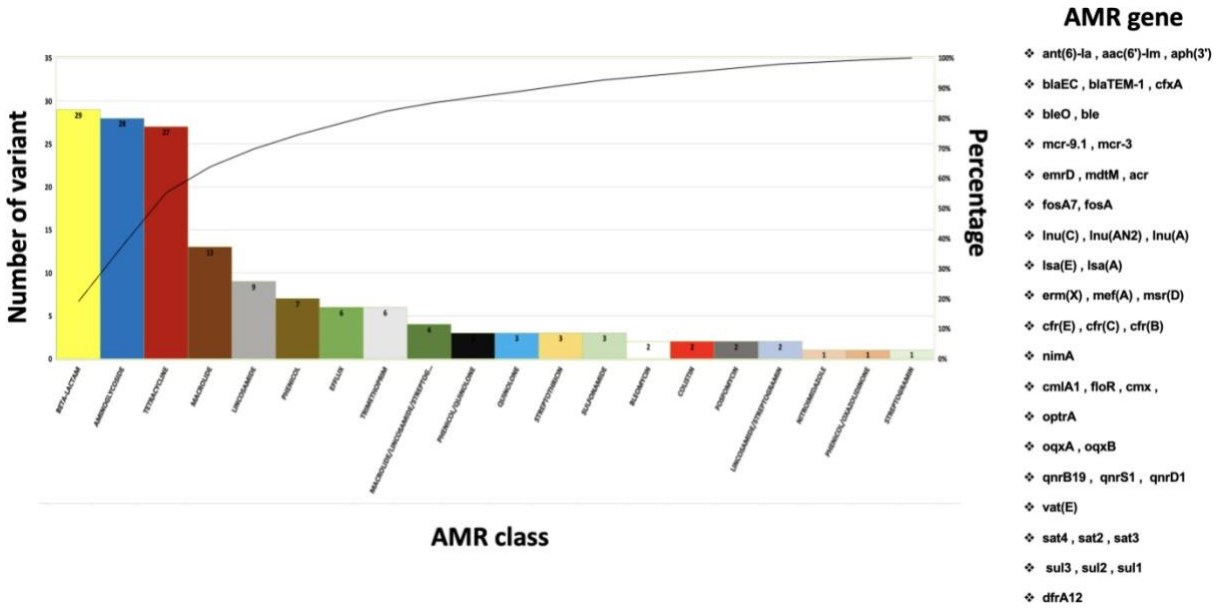


Figure 22 [A] Swine

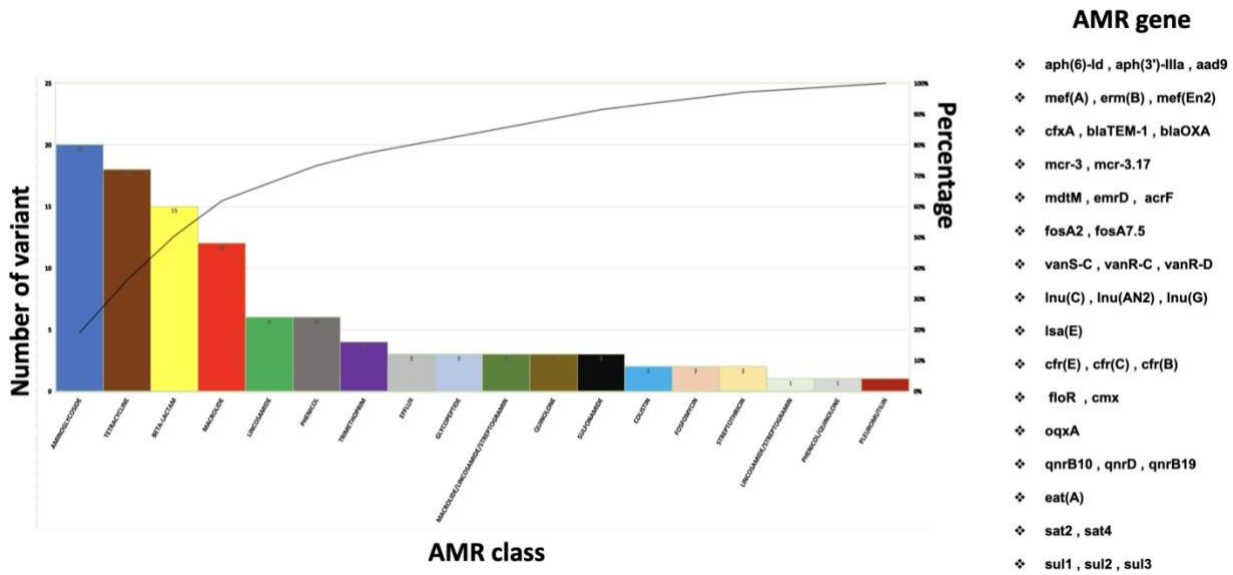


Figure 22 [B] Turkey

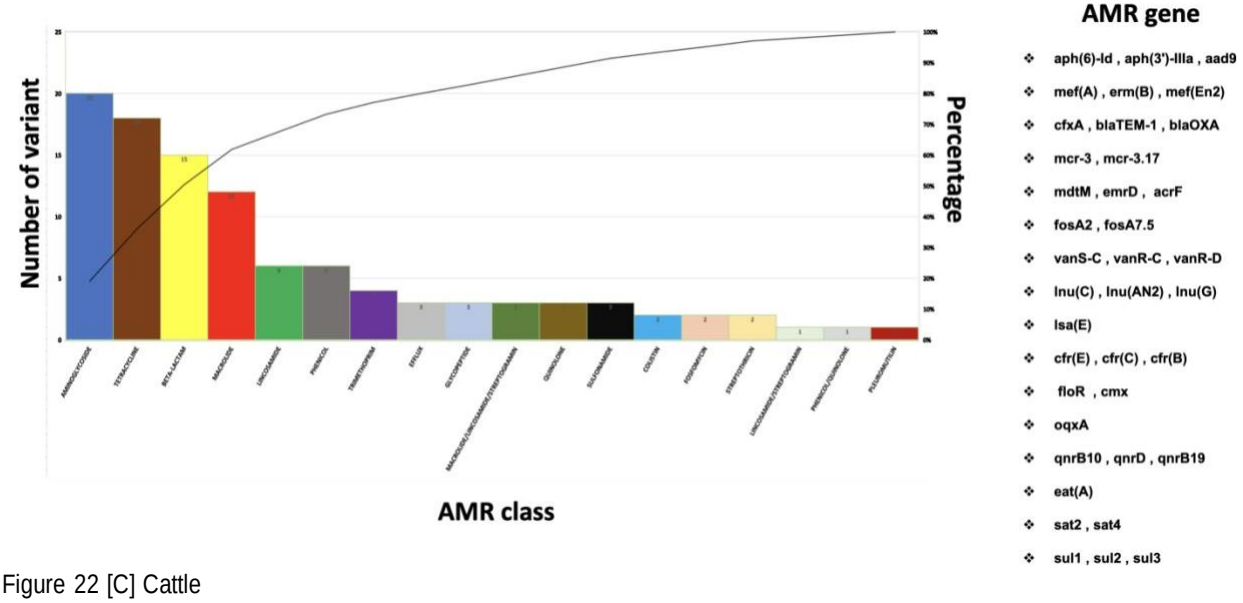


Figure 22 [C] Cattle

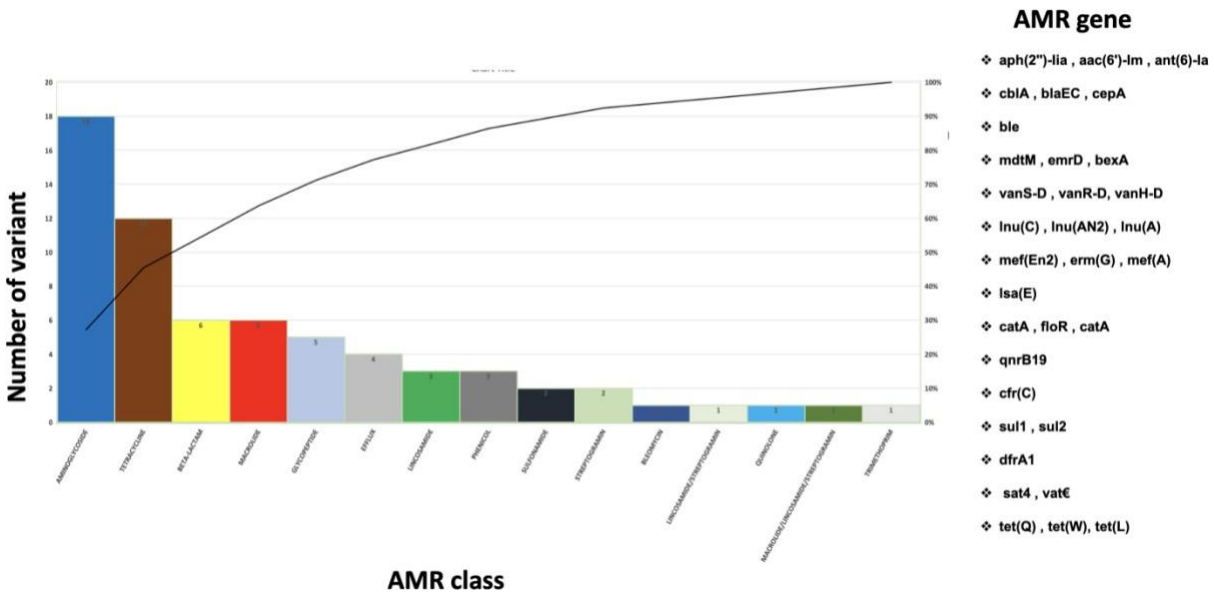


Figure 22 [D] Chicken

Figure 22. resistome analysis at food animal source. Number AMR gene variants by AMR class and more abundant AMR gene by animal source. [A] Swine, [B] Turkey. [C] Cattle, [D] Chicken

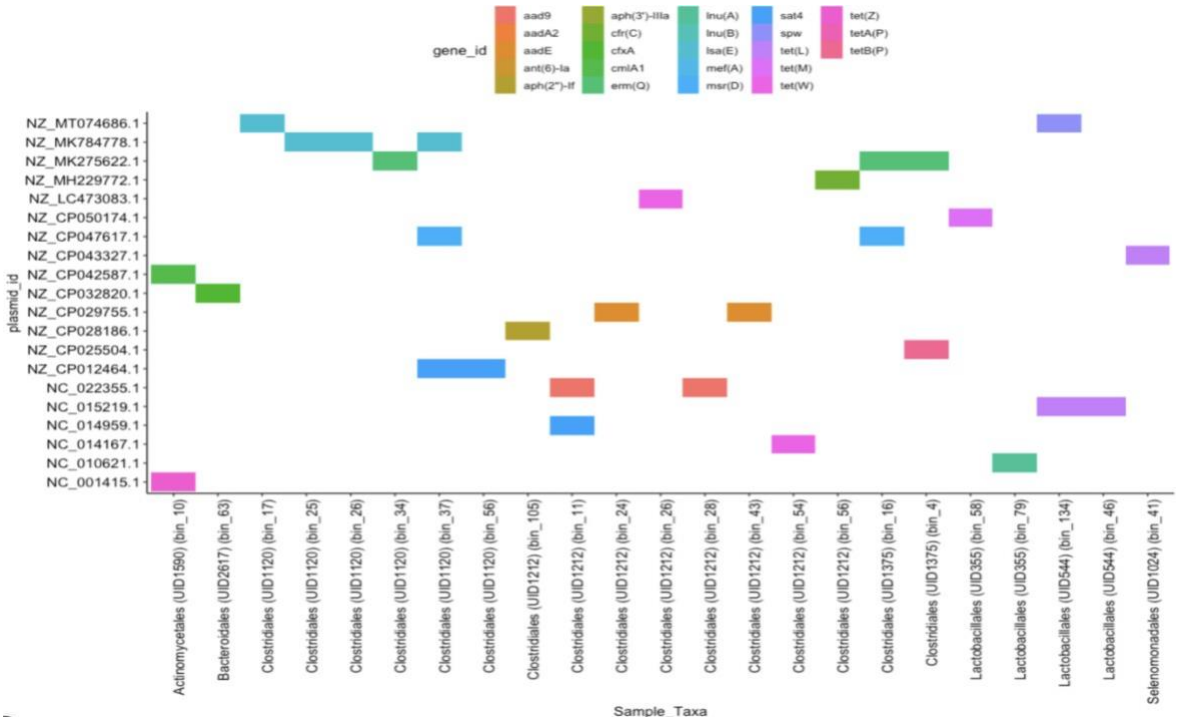


Figure 23 [A] Swine

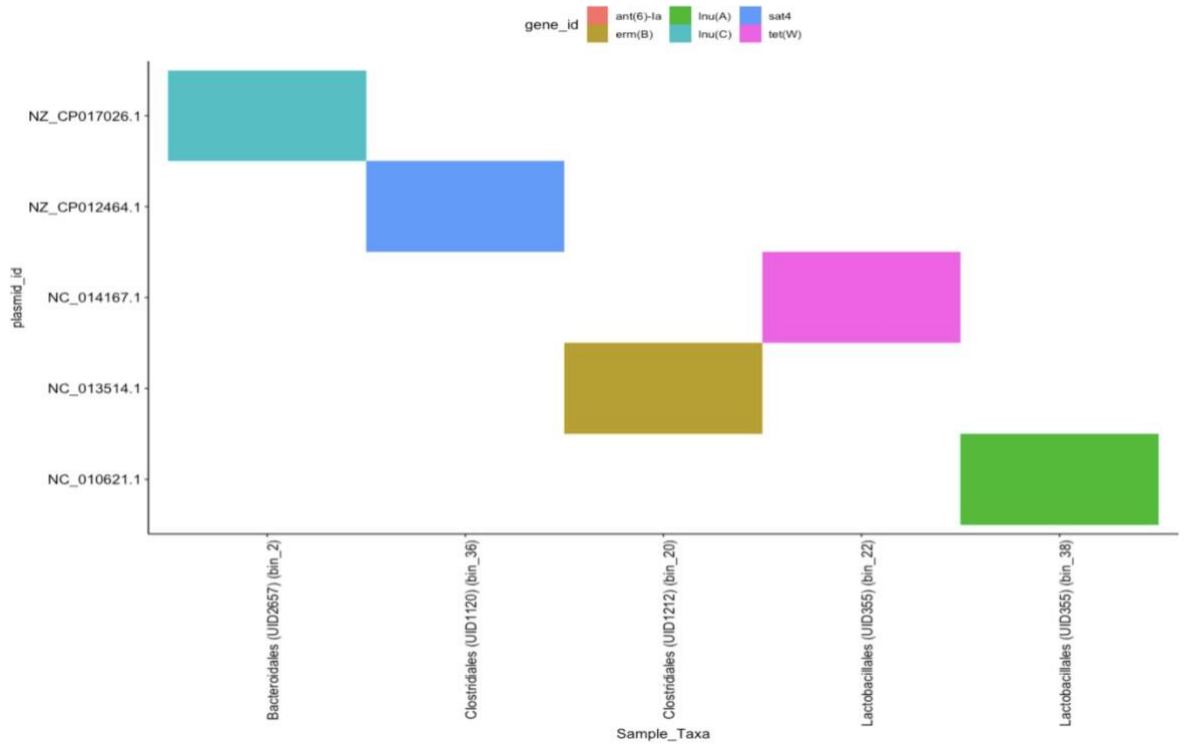


Figure 23 [B] Turkey

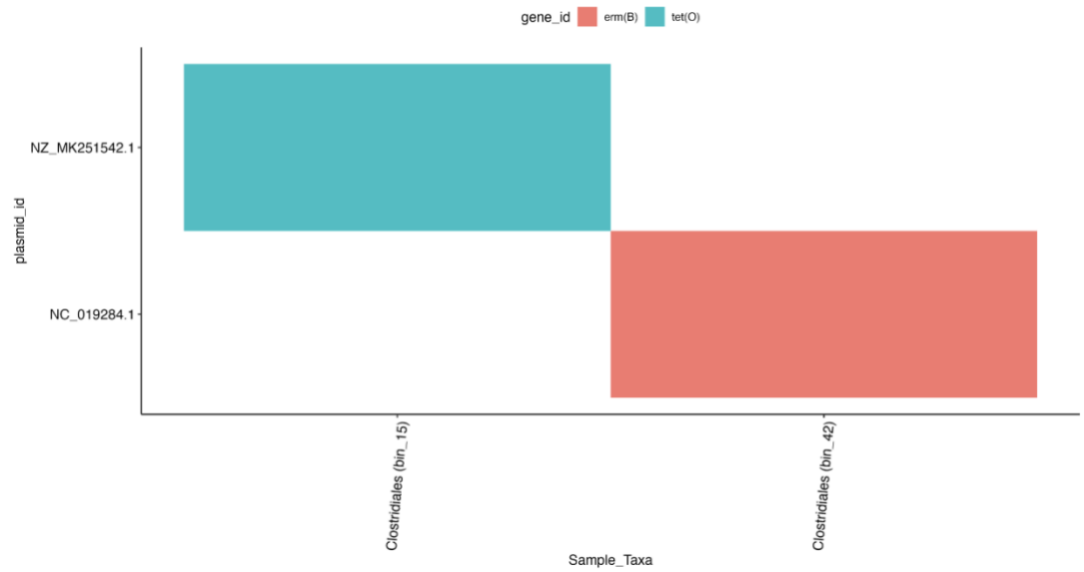


Figure 23 [C] Cattle

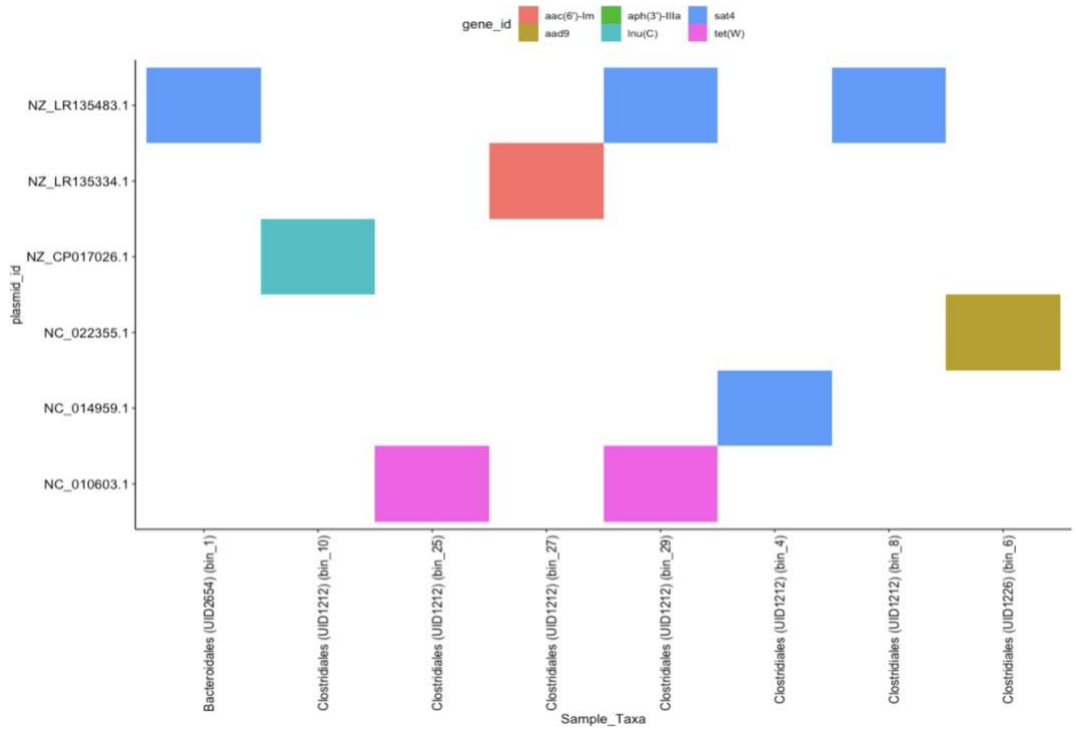


Figure 23 [D] Chicken

Figure 23. Linking AMR genes and plasmid to host bacteria source at food animal source [A] Swine, [B] Turkey. [C] Cattle, [D] Chicken

5.7 Conclusion:

The rapid increase of antimicrobial resistance among bacterial pathogens has become a serious human health threat. We lack understanding of the origins of these antimicrobial resistance genes and their spread within the microbiome. This is somewhat due to the inability to identify the bacterial hosts of AMR genes and the mobile genetic elements that mediate the spread of resistance on a complex microbial community. Metagenomics is a valuable tool for the study of microbial communities but has limitation by the inability to cluster sequencing outputs into specific bins corresponding to the individual taxa rank within the microbiome. Moreover, it can't track the flow of mobile genetic elements such as plasmids responsible for carrying and transfering AMR genes through bacterial communities or to determine from which plasmid a particular AMR genes is coming from. We employed deep shotgun metagenomic sequencing and proximity ligation (Hi-C) technology to overcome this limitation to characterize the resistome, assemble genomes from metagenomic samples, and accurately attribute antibiotic resistance genes and plasmids to host cells.

From a random selection of 21 cecal food animal sources (cattle, swine, chicken, and turkey) samples, we generated more than 75 million reads/sample for Hi-C and more than 100 million reads/sample for shotgun metagenomic sequence reads. Accurately, we clustered metagenome assembly genomes into MAGs that contain nearly complete genomes with high levels of confidence. A total of 245 previously uncharacterized genomes and 78 previously characterized genomes were reconstructed using MAGs with high level of confidence (>90% Completeness, >90% Novelty, < 5% contamination).

Plasmid characterization using Hi-C and shotgun metagenomics allowed identification of 149 plasmids (>= 85% completeness, >= 90% reference sequence similarity), and over 13000 plasmid-contigs (<85% completeness, < 90% reference sequence similarity).

We identified over 400 antimicrobial resistance genes representing 22 antimicrobial classes. Hi-C detected both previously known and novel associations between AMR genes, plasmids, and host genomes. Thus, Hi-C proximity ligation and deep shotgun metagenomics sequencing successfully linked antimicrobial resistance genes and plasmids to host bacteria to help explain the reservoirs of antibiotic resistance genes accessible to pathogens.

Acknowledgements

The views expressed in this article are those of the authors and do not necessarily reflect the official policy of the Department of Health and Human Services, the U.S. Food and Drug Administration, or the U.S. government. Reference to any commercial materials, equipment, or process does not in any way constitute approval, endorsement, or recommendation by the Food and Drug Administration.

PART III:

6. DISCUSSION

In the present study, we tackled a big gap on the sequencing research field to evaluate the impact of depth of coverage on microbiome and resistome analysis in food animal. Depending on the complexity of the metagenomic sample, it may require a sufficient sequencing depth to recover the total composition of a mixed bacteria population and it resistome potential [154]. Metagenomics could provide the solution to a cultureless future in clinical microbiology, food safety, and public health, as well as identify clinical causative agents

or foodborne contamination detection of AMR genes and virulence factors, in addition provides high resolution subtyping data [26]. Nevertheless, a non-optimal sequencing dept can critically affect the ability to fully interrogate complex microbiomes, to be able to identify minor genetic variants or to separate closed related bacteria [156, 157]. reduction of sequencing depth had major impact on the sensitivity of whole genome sequence for profilin, increasing the number of undetected species [158].

We also characterized previously uncultured food animals' intestinal microbiota, utilizing shotgun metagenomics approaches and Hi-C proximity ligation to better understand animal microbiome composition and their potential to become antimicrobial resistance reservoirs. Advances in sequencing technologies and the decreasing costs for Next Generation Sequencing technology has made it easier to use metagenomic approaches to sequence bacterial community DNA from a single sample *en masse*. This provides a comprehensive understanding of an entire microbial population, including those that cannot be cultured on artificial media and anaerobic [59]. Most research on microbiome and resistome composition relevant to human and animal health are based on the ability to culture and isolate a bacteria colony in a laboratory setting. Because of this, bacteria that are incapable of growing on artificial media are unculturable and disregarded in the majority of antimicrobial resistance studies. These culture-based techniques lack the ability to analyze whole microbial populations and could only provide information that is limited to few marker organisms, representing a potential towards culturable bacteria. On the contrary, culture free metagenomics studies will reveal all of the bacteria present in a microbiome that do not grow in artificial medium [44, 63, 71]. Examining unculturable bacteria can provide invaluable biological information about

microbial communities and their potential as new antimicrobial resistance reservoirs [70]. Thus, metagenomics will provide a complete picture of bacterial taxonomic and resistome composition in a complex gut microbiome system.

We randomly selected 21 cecal samples from food animal sources, more than 75 million reads/sample for Hi-C and more than 100 million reads/sample for shotgun metagenomic sequence reads were generated.

Analysis revealed over 200 bins containing metagenome assembled genomes (MAGs) with different levels of completeness, novelty score, and contamination based on CheckM. A total of 245 previously uncharacterized genomes were reconstructed using MAGs with high level of confidence (>90% Completeness, >90% Novelty, < 5% contamination). On the other hand, 78 previously characterized MAGs were identified with high level of confidence (>90% Completeness, <90% Novelty, < 5% contamination). Of the 245 newly reconstructed MAGs, 24 were at Bacteria taxonomic rank level, 5 at Phyla (Actinobacteria [11 genomes from UID2112], Firmicutes [3 genomes from UID1022], Bacteroidetes [17 genomes from UID2605], Euryarchaeota [2 genomes from UID3]), 4 at Class (Bacilli [1 genome from UID259], Clostridia [9 genomes from UID1118]), Deltaproteobacteria [3 genomes from UID3218], Gammaproteobacteria [1 genome from UID4201]), 5 at Order level (Actinomycetales[1 genome from UID1663, and one genome from UID1812], Bacteroidales [9 genomes from UID2617, 3 genomes from UID2621, 12 genomes from UID2657, 1 genome from UID2716], Clostridiales [30 genomes from UID1120, 67 genomes from UID1212, 17 genomes from UID1226]), Lactobacillales [1 genome from UID544, 1 genome from UID355], and Selenomonadales [2 genomes from UID1024]), and 3 at family level (Lachnospiraceae [5

genomes from UID1255, 13 genomes from UID1256, 5 genomes from UID1286, 1 genome from UID4932], Spirochaetaceae [1 genome from UID2535], and Spirochaetaceae [2 genomes from UID2535]).

Plasmid characterization using Hi-C and shotgun metagenomics allowed identification of 146 plasmids (>= 85% completeness, >= 90% reference sequence similarity), and over 13000 plasmid-contigs (<85% completeness, < 90% reference sequence similarity).

We identified over 400 antimicrobial resistance genes representing 22 antimicrobial classes including across the sample set: aminoglycoside (40 gene variants), beta-lactam(37 gene variants), bleomycin (2 gene variants), colistin (3 gene variants), efflux (6 gene variants), fosfomycin (4 gene variants), glycopeptide (6 gene variants), lincosamide (9 gene variants), lincosamide/streptogramin (2 gene variants), macrolide (16 gene variants), macrolide/lincosamide/streptogramin (4 gene variants), nitroimidazole (1 gene variant), phenicol (9 gene variants), phenicol/oxazolidinone (1 gene variant), phenicol/quinolone (2 gene variants), pleuromutilin (1 gene variant), quinolone (5 gene variants), streptogramin (1 gene variant), streptothricin (3 gene variants), sulfonamide (3 gene variants), tetracycline (25 gene variants), and trimethoprim (8 gene variants). Thus, this is an enormous contribution to the scientific community.

One if the biggest challenge on the war against antimicrobial resistance gene and their spread across different environments is the capability to attribute AMR genes to the plasmid and host bacterial cell, so that we could better understand the potential of antimicrobial resistance reservoirs present in gut microbial communities and their role in the spread of antimicrobial resistance.

Antimicrobial resistance is considered a global public health threat, were more than 2.8 million AMR infections occur yearly in the United States and each year more than 35,000 people die as a result of AMR infections **(**https://www.cdc.gov/drugresistance/national-estimates.html#:~:text=In%20the%20U.S.%2C%20more%20than,Resistance%20(AR)%20Threats%20Report**).** The spread of AMR genes that are present food animals [188, 197, 249-252] and livestock manures, which are discharged into the environment, are severe threat to human and animal health [251, 253]. Food animals' gut microbiome is an important source of AMR, swine farm has shown tetracycline and sulfonamide as concerning resistance genes [254, 255], aminoglycoside and tetracycline [197], beta-lactamase and tetracycline in turkey and chickens [256]. Lastly, to overcome such challenges of gene attribution, we utilized a combination of deep metagenomics sequencing and Hi-C (high- throughput chromosome conformation capture -3C) proximity ligation. Hi-C can identify long-range interactions in an unbiased, genome-wide fashion [45, 65-67]. It can produces high-quality metagenome-assembled genomes (MAGs) and has been applied in microbial genomics studies to help answer the question of attributing genomic material to individual isolates of bacterial species [44, 67, 68]. It could be used as a direct indicator of which sequences originated in the same cell, enabling of mixed genomes without any reliance on priori information [65]. Thus Hi-C links can be used in a map clustering context to deconvolute contigs into their original cellular groupings and plasmids [44, 69].

PART IV:

7. IMPACT OF RESEARCH, FUTURE WORK, AND CONCLUDING REMARKS

IMPACT OF RESEARCH:

The use of shotgun metagenomics and Hi-C proximity ligation offers a cost-effective and comprehensive means to characterized previously unknown, unculturable genomes and decipher the effects of antimicrobials on microbial communities and the public health risk posed by food animal resistomes thereby provide a better scientific basis for regulatory decision making on the safety of food animal antimicrobials. In addition, attribution of AMR genes to specific host bacteria greatly augments and impact antimicrobial resistance surveillance. This knowledge can help guide decision makers during hazard characterization, risk assessment, and risk mitigation for proposed uses of antimicrobial new animal drugs, or supplemental uses of previously approved antimicrobial new animal drugs.

FUTURE WORK:

I believe our research have room for improvements and continue the characterization on animal gut microbiota. We were able to characterize 245 new complete genomes from food animal gut metagenomics and reconstructed 88 genomes previously characterized. That comprise approximately 20% of total microbiome form our sample set. Therefore, there is 80% of gut microbiome still need to be characterized. We also identified 149 plasmids with higher level of confidence (> 85% completeness, > 90% reference sequence similarity), and over 13000 plasmid contigs with lower level of confidence (< 85% completeness, < 90% reference sequence similarity) that still need to be characterized. This future

characterization can be compared with single genome sequencing data from food animals and culture bacteria for more comprehensive snapshot of total microbiome, resistome and plasmidome composition of animal's gut.

CONCLUDING REMARKS:

Overall, our data indicate that sequencing depth critically affects the inferred taxonomic diversity of complex microbiomes and its resistome content, but sequencing depth alone is not sufficient for accurate taxonomic or resistome profiling. Proper bioinformatics analysis, including read quality control, data normalization, and appropriate reference databases, are essential for obtaining reliable results.

Our results can help to better understand uncultured bacteria and measure potential new antimicrobial resistance genes reservoirs present in intestinal microbiomes and their latent hazards on animal and human health.

Our data provide valuable insights into the diversity and identity of bacteria, plasmid, and resistance gene present in this complex gut microbiome.

Lastly, Hi-C detected both previously known and novel associations between AMR genes, plasmids, and host genomes. Thus, Hi-C proximity ligation and deep shotgun metagenomics sequencing are able to link antimicrobial resistance genes and plasmids to host bacteria to help explain the reservoirs of antibiotic resistance genes accessible to pathogens and their potential for widespread dissemination along the 'farm to fork' continuum.

PART V:

8. REFERENCES

1. Kellenberger E: **Exploring the unknown. The silent revolution of microbiology**. *EMBO Rep* 2001, **2**(1):5-7.
2. Burnham CD, Leeds J, Nordmann P, O'Grady J, Patel J: **Diagnosing antimicrobial resistance**. *Nat Rev Microbiol* 2017, **15**(11):697-703.
3. Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM: **Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products**. *Chem Biol* 1998, **5**(10):R245-249.
4. Watson EJ, Giles J, Scherer BL, Blatchford P: **Human faecal collection methods demonstrate a bias in microbiome composition by cell wall structure**. *Sci Rep* 2019, **9**(1):16831.
5. Salonen A, Nikkila J, Jalanka-Tuovinen J, Immonen O, Rajilic-Stojanovic M, Kekkonen RA, Palva A, de Vos WM: **Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: effective recovery of bacterial and archaeal DNA using mechanical cell lysis**. *J Microbiol Methods* 2010, **81**(2):127-134.
6. Hsieh YH, Peterson CM, Raggio A, Keenan MJ, Martin RJ, Ravussin E, Marco ML: **Impact of Different Fecal Processing Methods on Assessments of Bacterial Diversity in the Human Intestine**. *Front Microbiol* 2016, **7**:1643.
7. Sui HY, Weil AA, Nuwagira E, Qadri F, Ryan ET, Mezzari MP, Phipatanakul W, Lai PS: **Impact of DNA Extraction Method on Variation in Human and Built Environment Microbial Community and Functional Profiles Assessed by Shotgun Metagenomics Sequencing**. *Front Microbiol* 2020, **11**:953.
8. Yang F, Sun J, Luo H, Ren H, Zhou H, Lin Y, Han M, Chen B, Liao H, Brix S *et al*: **Assessment of fecal DNA extraction protocols for metagenomic studies**. *Gigascience* 2020, **9**(7).
9. Head SR, Komori HK, LaMere SA, Whisenant T, Van Nieuwerburgh F, Salomon DR, Ordoukhanian P: **Library construction for next-generation sequencing: overviews and challenges**. *Biotechniques* 2014, **56**(2):61-64, 66, 68, passim.
10. Chaudhari HG, Prajapati S, Wardah ZH, Raol G, Prajapati V, Patel R, Shati AA, Alfaifi MY, Elbehairi SEI, Sayyed RZ: **Decoding the microbial universe with metagenomics: a brief insight**. *Front Genet* 2023, **14**:1119740.
11. Navgire GS, Goel N, Sawhney G, Sharma M, Kaushik P, Mohanta YK, Mohanta TK, Al-Harrasi A: **Analysis and Interpretation of metagenomics data: an approach**. *Biol Proced Online* 2022, **24**(1):18.
12. Cho I, Blaser MJ: **The human microbiome: at the interface of health and disease**. *Nat Rev Genet* 2012, **13**(4):260-270.
13. Port JA, Cullen AC, Wallace JC, Smith MN, Faustman EM: **Metagenomic frameworks for monitoring antibiotic resistance in aquatic environments**. *Environ Health Perspect* 2014, **122**(3):222-228.

14. Sukhum KV, Diorio-Toth L, Dantas G: **Genomic and Metagenomic Approaches for Predictive Surveillance of Emerging Pathogens and Antibiotic Resistance**. *Clin Pharmacol Ther* 2019, **106**(3):512-524.
15. Davies J, Davies D: **Origins and evolution of antibiotic resistance**. *Microbiol Mol Biol Rev* 2010, **74**(3):417-433.
16. Prestinaci F, Pezzotti P, Pantosti A: **Antimicrobial resistance: a global multifaceted phenomenon**. *Pathog Glob Health* 2015, **109**(7):309-318.
17. Zheng D, Liwinski T, Elinav E: **Interaction between microbiota and immunity in health and disease**. *Cell Res* 2020, **30**(6):492-506.
18. Reygaert WC: **An overview of the antimicrobial resistance mechanisms of bacteria**. *AIMS Microbiol* 2018, **4**(3):482-501.
19. Peterson E, Kaur P: **Antibiotic Resistance Mechanisms in Bacteria: Relationships Between Resistance Determinants of Antibiotic Producers, Environmental Bacteria, and Clinical Pathogens**. *Front Microbiol* 2018, **9**:2928.
20. Piewngam P, Quinones M, Thirakittiwatthana W, Yungyuen T, Otto M, Kiratisin P: **Composition of the intestinal microbiota in extended-spectrum beta-lactamase-producing Enterobacteriaceae carriers and non-carriers in Thailand**. *Int J Antimicrob Agents* 2019, **53**(4):435-441.
21. Donkor ES, Muhsen K, Johnson SAM, Kotey FCN, Dayie N, Tetteh-Quarcoo PB, Tette EMA, Osei MM, Egyir B, Nii-Trebi NI *et al*: **Multicenter Surveillance of Antimicrobial Resistance among Gram-Negative Bacteria Isolated from Bloodstream Infections in Ghana**. *Antibiotics (Basel)* 2023, **12**(2).
22. Ahmed SA, Kumar A, Sethi P, Kapil A, Pandey RM, Wig N: **Effectiveness of education and antibiotic control programme at All India Institute of Medical Sciences, New Delhi**. *Natl Med J India* 2018, **31**(5):262-267.
23. Bhatia R: **Antimicrobial Resistance: Threat, consequences and options**. *Natl Med J India* 2018, **31**(3):133-135.
24. Sutherland ME: **Antibiotic use across the globe**. *Nat Hum Behav* 2018, **2**(6):373.
25. Host institution: Department of Food H, Technology IoFS, Technology UdLLS, Likotrafiti E, Oniciuc EA, Prieto M, Santos JA, Lopez S, Alvarez-Ordonez A: **Risk assessment of antimicrobial resistance along the food chain through culture-independent methodologies**. *EFSA J* 2018, **16**(Suppl 1):e160811.
26. Forbes JD, Knox NC, Ronholm J, Pagotto F, Reimer A: **Metagenomics: The Next Culture-Independent Game Changer**. *Front Microbiol* 2017, **8**:1069.
27. Dos Santos DF, Istvan P, Quirino BF, Kruger RH: **Functional Metagenomics as a Tool for Identification of New Antibiotic Resistance Genes from Natural Environments**. *Microb Ecol* 2017, **73**(2):479-491.
28. Boolchandani M, Patel S, Dantas G: **Functional Metagenomics to Study Antibiotic Resistance**. *Methods Mol Biol* 2017, **1520**:307-329.
29. Bengtsson-Palme J, Larsson DGJ, Kristiansson E: **Using metagenomics to investigate human and environmental resistomes**. *J Antimicrob Chemother* 2017, **72**(10):2690-2703.

30. Beitel CW, Froenicke L, Lang JM, Korf IF, Michelmore RW, Eisen JA, Darling AE: **Strain- and plasmid-level deconvolution of a synthetic metagenome by sequencing proximity ligation products**. *PeerJ* 2014, **2**:e415.
31. Kalmar L, Gupta S, Kean IRL, Ba X, Hadjirin N, Lay EM, de Vries SPW, Bateman M, Bartlet H, Hernandez-Garcia J *et al*: **HAM-ART: An optimised culture-free Hi-C metagenomics pipeline for tracking antimicrobial resistance genes in complex microbial communities**. *PLoS Genet* 2022, **18**(3):e1009776.
32. Bickhart DM, Kolmogorov M, Tseng E, Portik DM, Korobeynikov A, Tolstoganov I, Uritskiy G, Liachko I, Sullivan ST, Shin SB *et al*: **Generating lineage-resolved, complete metagenome-assembled genomes from complex microbial communities**. *Nat Biotechnol* 2022, **40**(5):711-719.
33. Ivanova V, Chernevskaya E, Vasiluev P, Ivanov A, Tolstoganov I, Shafranskaya D, Ulyantsev V, Korobeynikov A, Razin SV, Beloborodova N *et al*: **Hi-C Metagenomics in the ICU: Exploring Clinically Relevant Features of Gut Microbiome in Chronically Critically Ill Patients**. *Front Microbiol* 2021, **12**:770323.
34. Beyi AF, Hassall A, Phillips GJ, Plummer PJ: **Tracking Reservoirs of Antimicrobial Resistance Genes in a Complex Microbial Community Using Metagenomic Hi-C: The Case of Bovine Digital Dermatitis**. *Antibiotics (Basel)* 2021, **10**(2).
35. Bickhart DM, Watson M, Koren S, Panke-Buisse K, Cersosimo LM, Press MO, Van Tassell CP, Van Kessel JAS, Haley BJ, Kim SW *et al*: **Assignment of virus and antimicrobial resistance genes to microbial hosts in a complex microbial community by combined long-read assembly and proximity ligation**. *Genome Biol* 2019, **20**(1):153.
36. de Kraker ME, Stewardson AJ, Harbarth S: **Will 10 Million People Die a Year due to Antimicrobial Resistance by 2050?** *PLoS Med* 2016, **13**(11):e1002184.
37. McEwen SA, Fedorka-Cray PJ: **Antimicrobial use and resistance in animals**. *Clin Infect Dis* 2002, **34 Suppl 3**:S93-S106.
38. Manyi-Loh C, Mamphweli S, Meyer E, Okoh A: **Antibiotic Use in Agriculture and Its Consequential Resistance in Environmental Sources: Potential Public Health Implications**. *Molecules* 2018, **23**(4).
39. Thanner S, Drissner D, Walsh F: **Antimicrobial Resistance in Agriculture**. *mBio* 2016, **7**(2):e02227-02215.
40. Scott HM, Acuff G, Bergeron G, Bourassa MW, Gill J, Graham DW, Kahn LH, Morley PS, Salois MJ, Simjee S *et al*: **Critically important antibiotics: criteria and approaches for measuring and reducing their use in food animal agriculture**. *Ann N Y Acad Sci* 2019, **1441**(1):8-16.
41. Luiken REC, Van Gompel L, Munk P, Sarrazin S, Joosten P, Dorado-Garcia A, Borup Hansen R, Knudsen BE, Bossers A, Wagenaar JA *et al*: **Associations between antimicrobial use and the faecal resistome on broiler farms from nine European countries**. *J Antimicrob Chemother* 2019, **74**(9):2596-2604.
42. Tyagi A, Singh B, Billekallu Thammegowda NK, Singh NK: **Shotgun metagenomics offers novel insights into taxonomic compositions, metabolic pathways and antibiotic resistance genes in fish gut microbiome**. *Arch Microbiol* 2019, **201**(3):295-303.

43. Zeng J, Pan Y, Yang J, Hou M, Zeng Z, Xiong W: **Metagenomic insights into the distribution of antibiotic resistome between the gut-associated environments and the pristine environments**. *Environ Int* 2019, **126**:346-354.
44. Stalder T, Press MO, Sullivan S, Liachko I, Top EM: **Linking the resistome and plasmidome to the microbiome**. *ISME J* 2019, **13**(10):2437-2446.
45. DeMaere MZ, Darling AE: **bin3C: exploiting Hi-C sequencing data to accurately resolve metagenome-assembled genomes**. *Genome Biol* 2019, **20**(1):46.
46. Liao H, Friman VP, Geisen S, Zhao Q, Cui P, Lu X, Chen Z, Yu Z, Zhou S: **Horizontal gene transfer and shifts in linked bacterial community composition are associated with maintenance of antibiotic resistance genes during food waste composting**. *Sci Total Environ* 2019, **660**:841-850.
47. Shan X, Li XS, Schwarz S, Chen Y, Xu C, Du XD: **Plasmid-Assisted Horizontal Transfer of a Large lsa(E)-Carrying Genomic Island in Enterococcus faecalis**. *Microbiol Spectr* 2022, **10**(4):e0015422.
48. Smorawinska M, Szuplewska M, Zaleski P, Wawrzyniak P, Maj A, Plucienniczak A, Bartosik D: **Mobilizable narrow host range plasmids as natural suicide vectors enabling horizontal gene transfer among distantly related bacterial species**. *FEMS Microbiol Lett* 2012, **326**(1):76-82.
49. Brooks LE, Kaze M, Sistrom M: **Where the plasmids roam: large-scale sequence analysis reveals plasmids with large host ranges**. *Microb Genom* 2019, **5**(1).
50. Higuera-Llanten S, Vasquez-Ponce F, Barrientos-Espinoza B, Mardones FO, Marshall SH, Olivares-Pacheco J: **Extended antibiotic treatment in salmon farms select multiresistant gut bacteria with a high prevalence of antibiotic resistance genes**. *PLoS One* 2018, **13**(9):e0203641.
51. Crits-Christoph A, Hallowell HA, Koutouvalis K, Suez J: **Good microbes, bad genes? The dissemination of antimicrobial resistance in the human microbiome**. *Gut Microbes* 2022, **14**(1):2055944.
52. Leekitcharoenphon P, Johansson MHK, Munk P, Malorny B, Skarzynska M, Wadepohl K, Moyano G, Hesp A, Veldman KT, Bossers A *et al*: **Genomic evolution of antimicrobial resistance in Escherichia coli**. *Sci Rep* 2021, **11**(1):15108.
53. Aguiar-Pulido V, Huang W, Suarez-Ulloa V, Cickovski T, Mathee K, Narasimhan G: **Metagenomics, Metatranscriptomics, and Metabolomics Approaches for Microbiome Analysis**. *Evol Bioinform Online* 2016, **12**(Suppl 1):5-16.
54. Guardabassi L, Dalsgaard A, Olsen JE: **Phenotypic characterization and antibiotic resistance of Acinetobacter spp. isolated from aquatic sources**. *J Appl Microbiol* 1999, **87**(5):659-667.
55. Seifert H, Dijkshoorn L, Gerner-Smidt P, Pelzer N, Tjernberg I, Vaneechoutte M: **Distribution of Acinetobacter species on human skin: comparison of phenotypic and genotypic identification methods**. *J Clin Microbiol* 1997, **35**(11):2819-2825.
56. Coculescu BI: **Antimicrobial resistance induced by genetic changes**. *J Med Life* 2009, **2**(2):114-123.
57. Dimitriu T: **Evolution of horizontal transmission in antimicrobial resistance plasmids**. *Microbiology (Reading)* 2022, **168**(7).

58. Conlan S, Lau AF, Deming C, Spalding CD, Lee-Lin S, Thomas PJ, Park M, Dekker JP, Frank KM, Palmore TN *et al*: **Plasmid Dissemination and Selection of a Multidrug-Resistant Klebsiella pneumoniae Strain during Transplant-Associated Antibiotic Therapy**. *mBio* 2019, **10**(5).
59. Pitta DW, Dou Z, Kumar S, Indugu N, Toth JD, Vecchiarelli B, Bhukya B: **Metagenomic Evidence of the Prevalence and Distribution Patterns of Antimicrobial Resistance Genes in Dairy Agroecosystems**. *Foodborne Pathog Dis* 2016, **13**(6):296-302.
60. Koser CU, Ellington MJ, Peacock SJ: **Whole-genome sequencing to control antimicrobial resistance**. *Trends Genet* 2014, **30**(9):401-407.
61. Noyes NR, Yang X, Linke LM, Magnuson RJ, Cook SR, Zaheer R, Yang H, Woerner DR, Geornaras I, McArt JA *et al*: **Characterization of the resistome in manure, soil and wastewater from dairy and beef production systems**. *Sci Rep* 2016, **6**:24645.
62. Thomas M, Webb M, Ghimire S, Blair A, Olson K, Fenske GJ, Fonder AT, Christopher-Hennings J, Brake D, Scaria J: **Metagenomic characterization of the effect of feed additives on the gut microbiome and antibiotic resistome of feedlot cattle**. *Sci Rep* 2017, **7**(1):12257.
63. Harvey R, Funk J, Wittum TE, Hoet AE: **A metagenomic approach for determining prevalence of tetracycline resistance genes in the fecal flora of conventionally raised feedlot steers and feedlot steers raised without antimicrobials**. *Am J Vet Res* 2009, **70**(2):198-202.
64. Munk P, Knudsen BE, Lukjancenko O, Duarte ASR, Van Gompel L, Luiken REC, Smit LAM, Schmitt H, Garcia AD, Hansen RB *et al*: **Abundance and diversity of the faecal resistome in slaughter pigs and broilers in nine European countries**. *Nat Microbiol* 2018, **3**(8):898-908.
65. van Berkum NL, Lieberman-Aiden E, Williams L, Imakaev M, Gnirke A, Mirny LA, Dekker J, Lander ES: **Hi-C: a method to study the three-dimensional architecture of genomes**. *J Vis Exp* 2010(39).
66. Li Y, Ren Y, Zhang D, Jiang H, Wang Z, Li X, Rao D: **Chromosome-level assembly of the mustache toad genome using third-generation DNA sequencing and Hi-C analysis**. *Gigascience* 2019, **8**(9).
67. Pal K, Forcato M, Ferrari F: **Hi-C analysis: from data generation to integration**. *Biophys Rev* 2019, **11**(1):67-78.
68. Cremazy FG, Rashid FM, Haycocks JR, Lamberte LE, Grainger DC, Dame RT: **Determination of the 3D Genome Organization of Bacteria Using Hi-C**. *Methods Mol Biol* 2018, **1837**:3-18.
69. Brito IL, Yilmaz S, Huang K, Xu L, Jupiter SD, Jenkins AP, Naisilisili W, Tamminen M, Smillie CS, Wortman JR *et al*: **Mobile genes in the human microbiome are structured from global to individual scales**. *Nature* 2016, **535**(7612):435-439.
70. Joyce A, McCarthy CGP, Murphy S, Walsh F: **Antibiotic resistomes of healthy pig faecal metagenomes**. *Microb Genom* 2019, **5**(5).
71. Cantas L, Fraser TW, Fjelldal PG, Mayer I, Sorum H: **The culturable intestinal microbiota of triploid and diploid juvenile Atlantic salmon (Salmo salar) - a comparison of composition and drug resistance**. *BMC Vet Res* 2011, **7**:71.

72. Anjum MF, Zankari E, Hasman H: **Molecular Methods for Detection of Antimicrobial Resistance**. *Microbiol Spectr* 2017, **5**(6).
73. Amos GC, Zhang L, Hawkey PM, Gaze WH, Wellington EM: **Functional metagenomic analysis reveals rivers are a reservoir for diverse antibiotic resistance genes**. *Vet Microbiol* 2014, **171**(3-4):441-447.
74. Yu FB, Blainey PC, Schulz F, Woyke T, Horowitz MA, Quake SR: **Microfluidic-based mini-metagenomics enables discovery of novel microbial lineages from complex environmental samples**. *Elife* 2017, **6**.
75. Suzuki H, Yano H, Brown CJ, Top EM: **Predicting plasmid promiscuity based on genomic signature**. *J Bacteriol* 2010, **192**(22):6045-6055.
76. Krakau S, Straub D, Gourle H, Gabernet G, Nahnsen S: **nf-core/mag: a best-practice pipeline for metagenome hybrid assembly and binning**. *NAR Genom Bioinform* 2022, **4**(1):lqac007.
77. Van der Jeugt F, Maertens R, Steyaert A, Verschaffelt P, De Tender C, Dawyndt P, Mesuere B: **UMGAP: the Unipept MetaGenomics Analysis Pipeline**. *BMC Genomics* 2022, **23**(1):433.
78. Tamames J, Puente-Sanchez F: **SqueezeMeta, A Highly Portable, Fully Automatic Metagenomic Analysis Pipeline**. *Front Microbiol* 2018, **9**:3349.
79. Garfias-Gallegos D, Zirion-Martinez C, Bustos-Diaz ED, Arellano-Fernandez TV, Lovaco-Flores JA, Espinosa-Jaime A, Avelar-Rivas JA, Selem-Mojica N: **Metagenomics Bioinformatic Pipeline**. *Methods Mol Biol* 2022, **2512**:153-179.
80. Ferravante C, Memoli D, Palumbo D, Ciaramella P, Di Loria A, D'Agostino Y, Nassa G, Rizzo F, Tarallo R, Weisz A *et al*: **HOME-BIO (sHOtgun MEtagenomic analysis of BIOlogical entities): a specific and comprehensive pipeline for metagenomic shotgun sequencing data analysis**. *BMC Bioinformatics* 2021, **22**(Suppl 7):106.
81. Kuczynski J, Stombaugh J, Walters WA, Gonzalez A, Caporaso JG, Knight R: **Using QIIME to analyze 16S rRNA gene sequences from microbial communities**. *Curr Protoc Bioinformatics* 2011, **Chapter 10**:10 17 11-10 17 20.
82. Edgar RC: **UPARSE: highly accurate OTU sequences from microbial amplicon reads**. *Nat Methods* 2013, **10**(10):996-998.
83. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ *et al*: **Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities**. *Appl Environ Microbiol* 2009, **75**(23):7537-7541.
84. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP: **DADA2: High-resolution sample inference from Illumina amplicon data**. *Nat Methods* 2016, **13**(7):581-583.
85. Eren AM, Morrison HG, Lescault PJ, Reveillaud J, Vineis JH, Sogin ML: **Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences**. *ISME J* 2015, **9**(4):968-979.
86. Du Y, Sun F: **HiCBin: binning metagenomic contigs and recovering metagenome-assembled genomes using Hi-C contact maps**. *Genome Biol* 2022, **23**(1):63.
87. DeMaere MZ, Darling AE: **qc3C: Reference-free quality control for Hi-C sequencing data**. *PLoS Comput Biol* 2021, **17**(10):e1008839.

88. Foutel-Rodier T, Thierry A, Koszul R, Marbouty M: **Generation of a Metagenomics Proximity Ligation 3C Library of a Mammalian Gut Microbiota**. *Methods Enzymol* 2018, **612**:183-195.
89. Baudry L, Foutel-Rodier T, Thierry A, Koszul R, Marbouty M: **MetaTOR: A Computational Pipeline to Recover High-Quality Metagenomic Bins From Mammalian Gut Proximity-Ligation (meta3C) Libraries**. *Front Genet* 2019, **10**:753.
90. Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, Wu D, Paulsen I, Nelson KE, Nelson W *et al*: **Environmental genome shotgun sequencing of the Sargasso Sea**. *Science* 2004, **304**(5667):66-74.
91. Morais DAA, Cavalcante JVF, Monteiro SS, Pasquali MAB, Dalmolin RJS: **MEDUSA: A Pipeline for Sensitive Taxonomic Classification and Flexible Functional Annotation of Metagenomic Shotgun Sequences**. *Front Genet* 2022, **13**:814437.
92. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD *et al*: **SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing**. *J Comput Biol* 2012, **19**(5):455-477.
93. Antipov D, Korobeynikov A, McLean JS, Pevzner PA: **hybridSPAdes: an algorithm for hybrid assembly of short and long reads**. *Bioinformatics* 2016, **32**(7):1009-1015.
94. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH: **GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database**. *Bioinformatics* 2019, **36**(6):1925-1927.
95. Wood DE, Lu J, Langmead B: **Improved metagenomic analysis with Kraken 2**. *Genome Biol* 2019, **20**(1):257.
96. Kim D, Song L, Breitwieser FP, Salzberg SL: **Centrifuge: rapid and sensitive classification of metagenomic sequences**. *Genome Res* 2016, **26**(12):1721-1729.
97. Ondov BD, Bergman NH, Phillippy AM: **Interactive metagenomic visualization in a Web browser**. *BMC Bioinformatics* 2011, **12**:385.
98. Rho M, Tang H, Ye Y: **FragGeneScan: predicting genes in short and error-prone reads**. *Nucleic Acids Res* 2010, **38**(20):e191.
99. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA: **metaSPAdes: a new versatile metagenomic assembler**. *Genome Res* 2017, **27**(5):824-834.
100. Wu YW, Tang YH, Tringe SG, Simmons BA, Singer SW: **MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm**. *Microbiome* 2014, **2**:26.
101. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW: **CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes**. *Genome Res* 2015, **25**(7):1043-1055.
102. Wingett SW, Andrews S: **FastQ Screen: A tool for multi-genome mapping and quality control**. *F1000Res* 2018, **7**:1338.
103. Ewels P, Magnusson M, Lundin S, Kaller M: **MultiQC: summarize analysis results for multiple tools and samples in a single report**. *Bioinformatics* 2016, **32**(19):3047-3048.
104. Fu L, Niu B, Zhu Z, Wu S, Li W: **CD-HIT: accelerated for clustering the next-generation sequencing data**. *Bioinformatics* 2012, **28**(23):3150-3152.
105. Treangen TJ, Sommer DD, Angly FE, Koren S, Pop M: **Next generation sequence assembly with AMOS**. *Curr Protoc Bioinformatics* 2011, **Chapter 11**:Unit 11 18.

106. Li D, Liu CM, Luo R, Sadakane K, Lam TW: **MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph**. *Bioinformatics* 2015, **31**(10):1674-1676.
107. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM: **Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation**. *Genome Res* 2017, **27**(5):722-736.
108. Kang DD, Froula J, Egan R, Wang Z: **MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities**. *PeerJ* 2015, **3**:e1165.
109. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF: **Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy**. *Nat Microbiol* 2018, **3**(7):836-843.
110. Boers SA, Jansen R, Hays JP: **Understanding and overcoming the pitfalls and biases of next-generation sequencing (NGS) methods for use in the routine clinical microbiological diagnostic laboratory**. *Eur J Clin Microbiol Infect Dis* 2019, **38**(6):1059-1070.
111. Chakravorty S, Helb D, Burday M, Connell N, Alland D: **A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria**. *J Microbiol Methods* 2007, **69**(2):330-339.
112. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P: **An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea**. *ISME J* 2012, **6**(3):610-618.
113. Federhen S: **The NCBI Taxonomy database**. *Nucleic Acids Res* 2012, **40**(Database issue):D136-143.
114. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glockner FO: **SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB**. *Nucleic Acids Res* 2007, **35**(21):7188-7196.
115. Cole JR, Chai B, Farris RJ, Wang Q, Kulam SA, McGarrell DM, Garrity GM, Tiedje JM: **The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis**. *Nucleic Acids Res* 2005, **33**(Database issue):D294-296.
116. Durazzi F, Sala C, Castellani G, Manfreda G, Remondini D, De Cesare A: **Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota**. *Sci Rep* 2021, **11**(1):3030.
117. Schloss PD, Handelsman J: **Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness**. *Appl Environ Microbiol* 2005, **71**(3):1501-1506.
118. Schloss PD, Larget BR, Handelsman J: **Integration of microbial ecology and statistics: a test to compare gene libraries**. *Appl Environ Microbiol* 2004, **70**(9):5485-5492.
119. Schloss PD, Handelsman J: **Introducing SONS, a tool for operational taxonomic unit-based comparisons of microbial community memberships and structures**. *Appl Environ Microbiol* 2006, **72**(10):6773-6779.
120. Schloss PD, Handelsman J: **Introducing TreeClimber, a test to compare microbial community structures**. *Appl Environ Microbiol* 2006, **72**(4):2379-2384.

121. Lozupone C, Hamady M, Knight R: **UniFrac--an online tool for comparing microbial community diversity in a phylogenetic context**. *BMC Bioinformatics* 2006, **7**:371.
122. Huse SM, Dethlefsen L, Huber JA, Mark Welch D, Relman DA, Sogin ML: **Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing**. *PLoS Genet* 2008, **4**(11):e1000255.
123. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI *et al*: **QIIME allows analysis of high-throughput community sequencing data**. *Nat Methods* 2010, **7**(5):335-336.
124. Chen YA, Park J, Natsume-Kitatani Y, Kawashima H, Mohsen A, Hosomi K, Tanisawa K, Ohno H, Konishi K, Murakami H *et al*: **MANTA, an integrative database and analysis platform that relates microbiome and phenotypic data**. *PLoS One* 2020, **15**(12):e0243609.
125. Sanderson LA, Caron CT, Tan RL, Bett KE: **A PostgreSQL Tripal solution for large-scale genotypic and phenotypic data**. *Database (Oxford)* 2021, **2021**.
126. Burton JN, Liachko I, Dunham MJ, Shendure J: **Species-level deconvolution of metagenome assemblies with Hi-C-based contact probability maps**. *G3 (Bethesda)* 2014, **4**(7):1339-1346.
127. Calderon-Franco D, van Loosdrecht MCM, Abeel T, Weissbrodt DG: **Catch me if you can: capturing microbial community transformation by extracellular DNA using Hi-C sequencing**. *Antonie Van Leeuwenhoek* 2023, **116**(7):667-685.
128. Beyi AF, Brito-Goulart D, Hawbecker T, Ruddell B, Hassall A, Dewell R, Dewell G, Sahin O, Zhang Q, Plummer PJ: **Enrofloxacin Alters Fecal Microbiota and Resistome Irrespective of Its Dose in Calves**. *Microorganisms* 2021, **9**(10).
129. Marbouty M, Baudry L, Cournac A, Koszul R: **Scaffolding bacterial genomes and probing host-virus interactions in gut microbiome by proximity ligation (chromosome capture) assay**. *Sci Adv* 2017, **3**(2):e1602105.
130. Marbouty M, Koszul R: **Metagenomes Binning Using Proximity-Ligation Data**. *Methods Mol Biol* 2022, **2301**:163-181.
131. Marie-Nelly H, Marbouty M, Cournac A, Flot JF, Liti G, Parodi DP, Syan S, Guillen N, Margeot A, Zimmer C *et al*: **High-quality genome (re)assembly using chromosomal contact data**. *Nat Commun* 2014, **5**:5695.
132. Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J: **Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions**. *Nat Biotechnol* 2013, **31**(12):1119-1125.
133. Chen MS, Niu L, Zhao ML, Xu C, Pan BZ, Fu Q, Tao YB, He H, Hou C, Xu ZF: **De novo genome assembly and Hi-C analysis reveal an association between chromatin architecture alterations and sex differentiation in the woody plant Jatropha curcas**. *Gigascience* 2020, **9**(2).
134. Varoquaux N, Liachko I, Ay F, Burton JN, Shendure J, Dunham MJ, Vert JP, Noble WS: **Accurate identification of centromere locations in yeast genomes using Hi-C**. *Nucleic Acids Res* 2015, **43**(11):5331-5339.
135. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A *et al*: **The metagenomics RAST server - a public resource for the**

automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 2008, **9**:386.
136. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ: **Prodigal: prokaryotic gene recognition and translation initiation site identification**. *BMC Bioinformatics* 2010, **11**:119.
137. Box M, Jones MW, Whiteley N: **A hidden Markov model for decoding and the analysis of replay in spike trains**. *J Comput Neurosci* 2016, **41**(3):339-366.
138. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z: **MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies**. *PeerJ* 2019, **7**:e7359.
139. Du Y, Laperriere SM, Fuhrman J, Sun F: **Normalizing Metagenomic Hi-C Data and Detecting Spurious Contacts Using Zero-Inflated Negative Binomial Regression**. *J Comput Biol* 2022, **29**(2):106-120.
140. Traag VA, Waltman L, van Eck NJ: **From Louvain to Leiden: guaranteeing well-connected communities**. *Sci Rep* 2019, **9**(1):5233.
141. Marbouty M, Cournac A, Flot JF, Marie-Nelly H, Mozziconacci J, Koszul R: **Metagenomic chromosome conformation capture (meta3C) unveils the diversity of chromosome organization in microorganisms**. *Elife* 2014, **3**:e03318.
142. Rosvall M, Bergstrom CT: **Maps of random walks on complex networks reveal community structure**. *Proc Natl Acad Sci U S A* 2008, **105**(4):1118-1123.
143. Jung Y, Han D: **BWA-MEME: BWA-MEM emulated with a machine learning approach**. *Bioinformatics* 2022, **38**(9):2404-2413.
144. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S: **The Sequence Alignment/Map format and SAMtools**. *Bioinformatics* 2009, **25**(16):2078-2079.
145. Marbouty M, Koszul R: **Metagenome Analysis Exploiting High-Throughput Chromosome Conformation Capture (3C) Data**. *Trends Genet* 2015, **31**(12):673-682.
146. Meyer F, Fritz A, Deng ZL, Koslicki D, Lesker TR, Gurevich A, Robertson G, Alser M, Antipov D, Beghini F *et al*: **Critical Assessment of Metagenome Interpretation: the second round of challenges**. *Nat Methods* 2022, **19**(4):429-440.
147. Larsson DGJ, Flach CF: **Antibiotic resistance in the environment**. *Nat Rev Microbiol* 2022, **20**(5):257-269.
148. Aslam B, Wang W, Arshad MI, Khurshid M, Muzammil S, Rasool MH, Nisar MA, Alvi RF, Aslam MA, Qamar MU *et al*: **Antibiotic resistance: a rundown of a global crisis**. *Infect Drug Resist* 2018, **11**:1645-1658.
149. EclinicalMedicine: **Antimicrobial resistance: a top ten global public health threat**. *EClinicalMedicine* 2021, **41**:101221.
150. Ayukekbong JA, Ntemgwa M, Atabe AN: **The threat of antimicrobial resistance in developing countries: causes and control strategies**. *Antimicrob Resist Infect Control* 2017, **6**:47.
151. Salam MA, Al-Amin MY, Salam MT, Pawar JS, Akhter N, Rabaan AA, Alqumber MAA: **Antimicrobial Resistance: A Growing Serious Threat for Global Public Health**. *Healthcare (Basel)* 2023, **11**(13).

152. von Wintersdorff CJ, Penders J, van Niekerk JM, Mills ND, Majumder S, van Alphen LB, Savelkoul PH, Wolffs PF: **Dissemination of Antimicrobial Resistance in Microbial Ecosystems through Horizontal Gene Transfer**. *Front Microbiol* 2016, **7**:173.
153. Pillay S, Calderon-Franco D, Urhan A, Abeel T: **Metagenomic-based surveillance systems for antibiotic resistance in non-clinical settings**. *Front Microbiol* 2022, **13**:1066995.
154. Zaheer R, Noyes N, Ortega Polo R, Cook SR, Marinier E, Van Domselaar G, Belk KE, Morley PS, McAllister TA: **Impact of sequencing depth on the characterization of the microbiome and resistome**. *Sci Rep* 2018, **8**(1):5890.
155. Dookie N, Khan A, Padayatchi N, Naidoo K: **Application of Next Generation Sequencing for Diagnosis and Clinical Management of Drug-Resistant Tuberculosis: Updates on Recent Developments in the Field**. *Front Microbiol* 2022, **13**:775030.
156. Gweon HS, Shaw LP, Swann J, De Maio N, AbuOun M, Niehus R, Hubbard ATM, Bowes MJ, Bailey MJ, Peto TEA *et al*: **The impact of sequencing depth on the inferred taxonomic composition and AMR gene content of metagenomic samples**. *Environ Microbiome* 2019, **14**(1):7.
157. Yan Y, Nguyen LH, Franzosa EA, Huttenhower C: **Strain-level epidemiology of microbial communities and the human microbiome**. *Genome Med* 2020, **12**(1):71.
158. Pereira-Marques J, Hout A, Ferreira RM, Weber M, Pinto-Ribeiro I, van Doorn LJ, Knetsch CW, Figueiredo C: **Impact of Host DNA and Sequencing Depth on the Taxonomic Resolution of Whole Metagenome Sequencing for Microbiome Analysis**. *Front Microbiol* 2019, **10**:1277.
159. Chen S, Zhou Y, Chen Y, Gu J: **fastp: an ultra-fast all-in-one FASTQ preprocessor**. *Bioinformatics* 2018, **34**(17):i884-i890.
160. Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2**. *Nat Methods* 2012, **9**(4):357-359.
161. Kaminski J, Gibson MK, Franzosa EA, Segata N, Dantas G, Huttenhower C: **High-Specificity Targeted Functional Profiling in Microbial Communities with ShortBRED**. *PLoS Comput Biol* 2015, **11**(12):e1004557.
162. Schmieder R, Edwards R: **Insights into antibiotic resistance through metagenomic approaches**. *Future Microbiol* 2012, **7**(1):73-89.
163. Anthony WE, Burnham CD, Dantas G, Kwon JH: **The Gut Microbiome as a Reservoir for Antimicrobial Resistance**. *J Infect Dis* 2021, **223**(12 Suppl 2):S209-S213.
164. Wang WL, Xu SY, Ren ZG, Tao L, Jiang JW, Zheng SS: **Application of metagenomics in the human gut microbiome**. *World J Gastroenterol* 2015, **21**(3):803-814.
165. Collignon P, Beggs JJ, Walsh TR, Gandra S, Laxminarayan R: **Anthropological and socioeconomic factors contributing to global antimicrobial resistance: a univariate and multivariable analysis**. *Lancet Planet Health* 2018, **2**(9):e398-e405.
166. Deiner K, Bik HM, Machler E, Seymour M, Lacoursiere-Roussel A, Altermatt F, Creer S, Bista I, Lodge DM, de Vere N *et al*: **Environmental DNA metabarcoding: Transforming how we survey animal and plant communities**. *Mol Ecol* 2017, **26**(21):5872-5895.
167. Sierra MA, Li Q, Pushalkar S, Paul B, Sandoval TA, Kamer AR, Corby P, Guo Y, Ruff RR, Alekseyenko AV *et al*: **The Influences of Bioinformatics Tools and Reference Databases in Analyzing the Human Oral Microbial Community**. *Genes (Basel)* 2020, **11**(8).

168. Martinez JL, Coque TM, Baquero F: **What is a resistance gene? Ranking risk in resistomes**. *Nat Rev Microbiol* 2015, **13**(2):116-123.
169. Matthews C, Crispie F, Lewis E, Reid M, O'Toole PW, Cotter PD: **The rumen microbiome: a crucial consideration when optimising milk and meat production and nitrogen utilisation efficiency**. *Gut Microbes* 2019, **10**(2):115-132.
170. Malmuthuge N, Liang G, Griebel PJ, Guan LL: **Taxonomic and Functional Compositions of the Small Intestinal Microbiome in Neonatal Calves Provide a Framework for Understanding Early Life Gut Health**. *Appl Environ Microbiol* 2019, **85**(6).
171. Suriyaphol P, Chiu JKH, Yimpring N, Tunsagool P, Mhuantong W, Chuanchuen R, Bessarab I, Williams RBH, Ong RT, Suriyaphol G: **Dynamics of the fecal microbiome and antimicrobial resistome in commercial piglets during the weaning period**. *Sci Rep* 2021, **11**(1):18091.
172. Huijbers PM, Blaak H, de Jong MC, Graat EA, Vandenbroucke-Grauls CM, de Roda Husman AM: **Role of the Environment in the Transmission of Antimicrobial Resistance to Humans: A Review**. *Environ Sci Technol* 2015, **49**(20):11993-12004.
173. Boolchandani M, D'Souza AW, Dantas G: **Sequencing-based methods and resources to study antimicrobial resistance**. *Nat Rev Genet* 2019, **20**(6):356-370.
174. Zaheer R, Lakin SM, Polo RO, Cook SR, Larney FJ, Morley PS, Booker CW, Hannon SJ, Van Domselaar G, Read RR *et al*: **Comparative diversity of microbiomes and Resistomes in beef feedlots, downstream environments and urban sewage influent**. *BMC Microbiol* 2019, **19**(1):197.
175. Pollock J, Muwonge A, Hutchings MR, Mainda G, Bronsvoort BM, Gally DL, Corbishley A: **Resistance to change: AMR gene dynamics on a commercial pig farm with high antimicrobial usage**. *Sci Rep* 2020, **10**(1):1708.
176. Jacobsen L, Wilcks A, Hammer K, Huys G, Gevers D, Andersen SR: **Horizontal transfer of tet(M) and erm(B) resistance plasmids from food strains of Lactobacillus plantarum to Enterococcus faecalis JH2-2 in the gastrointestinal tract of gnotobiotic rats**. *FEMS Microbiol Ecol* 2007, **59**(1):158-166.
177. Holman DB, Gzyl KE, Mou KT, Allen HK: **Weaning Age and Its Effect on the Development of the Swine Gut Microbiome and Resistome**. *mSystems* 2021, **6**(6):e0068221.
178. Binta B, Patel M: **Detection of cfxA2, cfxA3, and cfxA6 genes in beta-lactamase producing oral anaerobes**. *J Appl Oral Sci* 2016, **24**(2):142-147.
179. Salah FD, Soubeiga ST, Ouattara AK, Sadji AY, Metuor-Dabire A, Obiri-Yeboah D, Banla-Kere A, Karou S, Simpore J: **Distribution of quinolone resistance gene (qnr) in ESBL-producing Escherichia coli and Klebsiella spp. in Lome, Togo**. *Antimicrob Resist Infect Control* 2019, **8**:104.
180. Pallecchi L, Riccobono E, Mantella A, Bartalesi F, Sennati S, Gamboa H, Gotuzzo E, Bartoloni A, Rossolini GM: **High prevalence of qnr genes in commensal enterobacteria from healthy children in Peru and Bolivia**. *Antimicrob Agents Chemother* 2009, **53**(6):2632-2635.
181. Martinez-Martinez L, Eliecer Cano M, Manuel Rodriguez-Martinez J, Calvo J, Pascual A: **Plasmid-mediated quinolone resistance**. *Expert Rev Anti Infect Ther* 2008, **6**(5):685-711.

182. Bharatham N, Bhowmik P, Aoki M, Okada U, Sharma S, Yamashita E, Shanbhag AP, Rajagopal S, Thomas T, Sarma M *et al*: **Structure and function relationship of OqxB efflux pump from Klebsiella pneumoniae**. *Nat Commun* 2021, **12**(1):5400.
183. Hansen LH, Jensen LB, Sorensen HI, Sorensen SJ: **Substrate specificity of the OqxAB multidrug resistance pump in Escherichia coli and selected enteric bacteria**. *J Antimicrob Chemother* 2007, **60**(1):145-147.
184. Hou C, Yang F: **Drug-resistant gene of blaOXA-23, blaOXA-24, blaOXA-51 and blaOXA-58 in Acinetobacter baumannii**. *Int J Clin Exp Med* 2015, **8**(8):13859-13863.
185. Tadesse DA, Li C, Mukherjee S, Hsu CH, Bodeis Jones S, Gaines SA, Kabera C, Loneragan GH, Torrence M, Harhay DM *et al*: **Whole-Genome Sequence Analysis of CTX-M Containing Escherichia coli Isolates from Retail Meats and Cattle in the United States**. *Microb Drug Resist* 2018, **24**(7):939-948.
186. Maheshwari M, Ahmad I, Althubiani AS: **Multidrug resistance and transferability of blaCTX-M among extended-spectrum beta-lactamase-producing enteric bacteria in biofilm**. *J Glob Antimicrob Resist* 2016, **6**:142-149.
187. Sandelin A, Halli O, Hartel H, Herva T, Kaartinen L, Tuunainen E, Rautala H, Soveri T, Simojoki H: **Effect of Farm Management Practices on Morbidity and Antibiotic Usage on Calf Rearing Farms**. *Antibiotics (Basel)* 2022, **11**(2).
188. Auffret MD, Dewhurst RJ, Duthie CA, Rooke JA, John Wallace R, Freeman TC, Stewart R, Watson M, Roehe R: **The rumen microbiome as a reservoir of antimicrobial resistance and pathogenicity genes is directly affected by diet in beef cattle**. *Microbiome* 2017, **5**(1):159.
189. Cameron A, McAllister TA: **Antimicrobial usage and resistance in beef production**. *J Anim Sci Biotechnol* 2016, **7**:68.
190. Van Boeckel TP, Brower C, Gilbert M, Grenfell BT, Levin SA, Robinson TP, Teillant A, Laxminarayan R: **Global trends in antimicrobial use in food animals**. *Proc Natl Acad Sci U S A* 2015, **112**(18):5649-5654.
191. Pokharel S, Raut S, Adhikari B: **Tackling antimicrobial resistance in low-income and middle-income countries**. *BMJ Glob Health* 2019, **4**(6):e002104.
192. Wallace MJ, Fishbein SRS, Dantas G: **Antimicrobial resistance in enteric bacteria: current state and next-generation solutions**. *Gut Microbes* 2020, **12**(1):1799654.
193. Jasovsky D, Littmann J, Zorzet A, Cars O: **Antimicrobial resistance-a threat to the world's sustainable development**. *Ups J Med Sci* 2016, **121**(3):159-164.
194. D'Costa VM, McGrann KM, Hughes DW, Wright GD: **Sampling the antibiotic resistome**. *Science* 2006, **311**(5759):374-377.
195. Xu J, Wang H, Xu R, Li Q, Su Y, Liu J, Zhu W: **The diurnal fluctuation of colonic antibiotic resistome is correlated with nutrient substrates in a pig model**. *Sci Total Environ* 2023, **891**:164692.
196. Partridge SR, Kwong SM, Firth N, Jensen SO: **Mobile Genetic Elements Associated with Antimicrobial Resistance**. *Clin Microbiol Rev* 2018, **31**(4).
197. Zhou Y, Fu H, Yang H, Wu J, Chen Z, Jiang H, Liu M, Liu Q, Huang L, Gao J *et al*: **Extensive metagenomic analysis of the porcine gut resistome to identify indicators reflecting antimicrobial resistance**. *Microbiome* 2022, **10**(1):39.

198. Yang C, Das Q, Rehman MA, Yin X, Shay J, Gauthier M, Lau CH, Ross K, Diarra MS: **Microbiome of Ceca from Broiler Chicken Vaccinated or Not against Coccidiosis and Fed Berry Pomaces**. *Microorganisms* 2023, **11**(5).
199. Rui Y, Qiu G: **Analysis of Gut Microbial Communities and Resistance Genes in Pigs and Chickens in Central China**. *Animals (Basel)* 2022, **12**(23).
200. Montassier E, Valdes-Mas R, Batard E, Zmora N, Dori-Bachash M, Suez J, Elinav E: **Probiotics impact the antibiotic resistance gene reservoir along the human GI tract in a person-specific and antibiotic-dependent manner**. *Nat Microbiol* 2021, **6**(8):1043-1054.
201. Dulanto Chiang A, Dekker JP: **From the Pipeline to the Bedside: Advances and Challenges in Clinical Metagenomics**. *J Infect Dis* 2020, **221**(Suppl 3):S331-S340.
202. Duarte ASR, Roder T, Van Gompel L, Petersen TN, Hansen RB, Hansen IM, Bossers A, Aarestrup FM, Wagenaar JA, Hald T: **Metagenomics-Based Approach to Source-Attribution of Antimicrobial Resistance Determinants - Identification of Reservoir Resistome Signatures**. *Front Microbiol* 2020, **11**:601407.
203. Cao J, Hu Y, Liu F, Wang Y, Bi Y, Lv N, Li J, Zhu B, Gao GF: **Metagenomic analysis reveals the microbiome and resistome in migratory birds**. *Microbiome* 2020, **8**(1):26.
204. Du Y, Fuhrman JA, Sun F: **ViralCC retrieves complete viral genomes and virus-host pairs from metagenomic Hi-C data**. *Nat Commun* 2023, **14**(1):502.
205. Cusco A, Perez D, Vines J, Fabregas N, Francino O: **Novel canine high-quality metagenome-assembled genomes, prophages and host-associated plasmids provided by long-read metagenomics together with Hi-C proximity ligation**. *Microb Genom* 2022, **8**(3).
206. Adkins PRF, Ericsson AC, Middleton JR, Witzke MC: **The effect of intramammary pirlimycin hydrochloride on the fecal microbiome of early-lactation heifers**. *J Dairy Sci* 2020, **103**(4):3459-3469.
207. Sanchez-Reyes A, Breton-Deval L, Mangelson H, Salinas-Peralta I, Sanchez-Flores A: **Hi-C deconvolution of a textile dye-related microbiome reveals novel taxonomic landscapes and links phenotypic potential to individual genomes**. *Int Microbiol* 2022, **25**(1):99-110.
208. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform**. *Bioinformatics* 2009, **25**(14):1754-1760.
209. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM: **Mash: fast genome and metagenome distance estimation using MinHash**. *Genome Biol* 2016, **17**(1):132.
210. Galata V, Fehlmann T, Backes C, Keller A: **PLSDB: a resource of complete bacterial plasmids**. *Nucleic Acids Res* 2019, **47**(D1):D195-D202.
211. Orlando G, Law PJ, Cornish AJ, Dobbins SE, Chubb D, Broderick P, Litchfield K, Hariri F, Pastinen T, Osborne CS *et al*: **Promoter capture Hi-C-based identification of recurrent noncoding mutations in colorectal cancer**. *Nat Genet* 2018, **50**(10):1375-1380.
212. Jarett JK, Kingsbury DD, Dahlhausen KE, Ganz HH: **Best Practices for Microbiome Study Design in Companion Animal Research**. *Front Vet Sci* 2021, **8**:644836.
213. Berg G, Rybakova D, Fischer D, Cernava T, Verges MC, Charles T, Chen X, Cocolin L, Eversole K, Corral GH *et al*: **Microbiome definition re-visited: old concepts and new challenges**. *Microbiome* 2020, **8**(1):103.

214. Parks DH, Rinke C, Chuvochina M, Chaumeil PA, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW: **Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life**. *Nat Microbiol* 2017, **2**(11):1533-1542.
215. Filardi R, Gargari G, Mora D, Arioli S: **Characterization of antibiotic-resistance traits in Akkermansia muciniphila strains of human origin**. *Sci Rep* 2022, **12**(1):19426.
216. Jung A, Chen LR, Suyemoto MM, Barnes HJ, Borst LB: **A Review of Enterococcus cecorum Infection in Poultry**. *Avian Dis* 2018, **62**(3):261-271.
217. Paju S, Carlson P, Jousimies-Somer H, Asikainen S: **Heterogeneity of Actinobacillus actinomycetemcomitans strains in various human infections and relationships between serotype, genotype, and antimicrobial susceptibility**. *J Clin Microbiol* 2000, **38**(1):79-84.
218. Davis-Richardson AG, Ardissone AN, Dias R, Simell V, Leonard MT, Kemppainen KM, Drew JC, Schatz D, Atkinson MA, Kolaczkowski B *et al*: **Bacteroides dorei dominates gut microbiome prior to autoimmunity in Finnish children at high risk for type 1 diabetes**. *Front Microbiol* 2014, **5**:678.
219. Khademi F, Sahebkar A: **The prevalence of antibiotic-resistant Clostridium species in Iran: a meta-analysis**. *Pathog Glob Health* 2019, **113**(2):58-66.
220. Allen HK, Looft T, Bayles DO, Humphrey S, Levine UY, Alt D, Stanton TB: **Antibiotics in feed induce prophages in swine fecal microbiomes**. *mBio* 2011, **2**(6).
221. Wilkinson TJ, Cowan AA, Vallin HE, Onime LA, Oyama LB, Cameron SJ, Gonot C, Moorby JM, Waddams K, Theobald VJ *et al*: **Characterization of the Microbiome along the Gastrointestinal Tract of Growing Turkeys**. *Front Microbiol* 2017, **8**:1089.
222. Lin L, Lai Z, Zhang J, Zhu W, Mao S: **The gastrointestinal microbiome in dairy cattle is constrained by the deterministic driver of the region and the modified effect of diet**. *Microbiome* 2023, **11**(1):10.
223. Liu L, Wu P, Chen F, Zhou J, Guo A, Shi K, Zhang Q: **Multi-omics analyses reveal that the gut microbiome and its metabolites promote milk fat synthesis in Zhongdian yak cows**. *PeerJ* 2022, **10**:e14444.
224. Birkl P, Bharwani A, Kjaer JB, Kunze W, McBride P, Forsythe P, Harlander-Matauschek A: **Differences in cecal microbiome of selected high and low feather-pecking laying hens**. *Poult Sci* 2018, **97**(9):3009-3014.
225. Oakley BB, Lillehoj HS, Kogut MH, Kim WK, Maurer JJ, Pedroso A, Lee MD, Collett SR, Johnson TJ, Cox NA: **The chicken gastrointestinal microbiome**. *FEMS Microbiol Lett* 2014, **360**(2):100-112.
226. El-Adawy H, Hotzel H, Garcia-Soto S, Tomaso H, Hafez HM, Schwarz S, Neubauer H, Linde J: **Genomic insight into Campylobacter jejuni isolated from commercial turkey flocks in Germany using whole-genome sequencing analysis**. *Front Vet Sci* 2023, **10**:1092179.
227. Lim SK, Kim D, Moon DC, Cho Y, Rho M: **Antibiotic resistomes discovered in the gut microbiomes of Korean swine and cattle**. *Gigascience* 2020, **9**(5).
228. Zhou Y, Li Y, Zhang L, Wu Z, Huang Y, Yan H, Zhong J, Wang LJ, Abdullah HM, Wang HH: **Antibiotic Administration Routes and Oral Exposure to Antibiotic Resistant Bacteria as Key Drivers for Gut Microbiota Disruption and Resistome in Poultry**. *Front Microbiol* 2020, **11**:1319.

229. Dsani E, Afari EA, Danso-Appiah A, Kenu E, Kaburi BB, Egyir B: **Antimicrobial resistance and molecular detection of extended spectrum beta-lactamase producing Escherichia coli isolates from raw meat in Greater Accra region, Ghana**. *BMC Microbiol* 2020, **20**(1):253.
230. Afridi OK, Ali J, Chang JH: **Next-Generation Sequencing Based Gut Resistome Profiling of Broiler Chickens Infected with Multidrug-Resistant Escherichia coli**. *Animals (Basel)* 2020, **10**(12).
231. Zhang YJ, Hu HW, Yan H, Wang JT, Lam SK, Chen QL, Chen D, He JZ: **Salinity as a predominant factor modulating the distribution patterns of antibiotic resistance genes in ocean and river beach soils**. *Sci Total Environ* 2019, **668**:193-203.
232. Kaviani Rad A, Astaykina A, Streletskii R, Afsharyzad Y, Etesami H, Zarei M, Balasundram SK: **An Overview of Antibiotic Resistance and Abiotic Stresses Affecting Antimicrobial Resistance in Agricultural Soils**. *Int J Environ Res Public Health* 2022, **19**(8).
233. Laconi A, Mughini-Gras L, Tolosi R, Grilli G, Trocino A, Carraro L, Di Cesare F, Cagnardi P, Piccirillo A: **Microbial community composition and antimicrobial resistance in agricultural soils fertilized with livestock manure from conventional farming in Northern Italy**. *Sci Total Environ* 2021, **760**:143404.
234. Beceiro A, Tomas M, Bou G: **Antimicrobial resistance and virulence: a successful or deleterious association in the bacterial world?** *Clin Microbiol Rev* 2013, **26**(2):185-230.
235. Li MY, Chen XQ, Wang JY, Wang HT, Xue XM, Ding J, Juhasz AL, Zhu YG, Li HB, Ma LQ: **Antibiotic exposure decreases soil arsenic oral bioavailability in mice by disrupting ileal microbiota and metabolic profile**. *Environ Int* 2021, **151**:106444.
236. Selvaraj S, J RD, Bansal V, Ren B: **Whole-genome haplotype reconstruction using proximity-ligation and shotgun sequencing**. *Nat Biotechnol* 2013, **31**(12):1111-1118.
237. Smith CJ, Owen C, Kirby L: **Activation of a cryptic streptomycin-resistance gene in the Bacteroides erm transposon, Tn4551**. *Mol Microbiol* 1992, **6**(16):2287-2297.
238. Murphy E, Huwyler L, de Freire Bastos Mdo C: **Transposon Tn554: complete nucleotide sequence and isolation of transposition-defective and antibiotic-sensitive mutants**. *EMBO J* 1985, **4**(12):3357-3365.
239. Murphy E: **Nucleotide sequence of a spectinomycin adenyltransferase AAD(9) determinant from Staphylococcus aureus and its relationship to AAD(3") (9)**. *Mol Gen Genet* 1985, **200**(1):33-39.
240. Wendlandt S, Li B, Lozano C, Ma Z, Torres C, Schwarz S: **Identification of the novel spectinomycin resistance gene spw in methicillin-resistant and methicillin-susceptible Staphylococcus aureus of human and animal origin**. *J Antimicrob Chemother* 2013, **68**(7):1679-1680.
241. Morales G, Picazo JJ, Baos E, Candel FJ, Arribi A, Pelaez B, Andrade R, de la Torre MA, Fereres J, Sanchez-Garcia M: **Resistance to linezolid is mediated by the cfr gene in the first report of an outbreak of linezolid-resistant Staphylococcus aureus**. *Clin Infect Dis* 2010, **50**(6):821-825.
242. Tsiodras S, Gold HS, Sakoulas G, Eliopoulos GM, Wennersten C, Venkataraman L, Moellering RC, Ferraro MJ: **Linezolid resistance in a clinical isolate of Staphylococcus aureus**. *Lancet* 2001, **358**(9277):207-208.

243. Nannini E, Murray BE, Arias CA: **Resistance or decreased susceptibility to glycopeptides, daptomycin, and linezolid in methicillin-resistant Staphylococcus aureus**. *Curr Opin Pharmacol* 2010, **10**(5):516-521.
244. Toh SM, Xiong L, Arias CA, Villegas MV, Lolans K, Quinn J, Mankin AS: **Acquisition of a natural resistance gene renders a clinical strain of methicillin-resistant Staphylococcus aureus resistant to the synthetic antibiotic linezolid**. *Mol Microbiol* 2007, **64**(6):1506-1514.
245. Atkinson GC, Hansen LH, Tenson T, Rasmussen A, Kirpekar F, Vester B: **Distinction between the Cfr methyltransferase conferring antibiotic resistance and the housekeeping RlmN methyltransferase**. *Antimicrob Agents Chemother* 2013, **57**(8):4019-4026.
246. Hansen LH, Planellas MH, Long KS, Vester B: **The order Bacillales hosts functional homologs of the worrisome cfr antibiotic resistance gene**. *Antimicrob Agents Chemother* 2012, **56**(7):3563-3567.
247. Kotra LP, Haddad J, Mobashery S: **Aminoglycosides: perspectives on mechanisms of action and resistance and strategies to counter resistance**. *Antimicrob Agents Chemother* 2000, **44**(12):3249-3256.
248. Mingeot-Leclercq MP, Glupczynski Y, Tulkens PM: **Aminoglycosides: activity and resistance**. *Antimicrob Agents Chemother* 1999, **43**(4):727-737.
249. Alvarado AC, Chekabab SM, Predicala BZ, Korber DR: **Impact of Raised without Antibiotics Measures on Antimicrobial Resistance and Prevalence of Pathogens in Sow Barns**. *Antibiotics (Basel)* 2022, **11**(9).
250. Dias D, Fonseca C, Mendo S, Caetano T: **A closer look on the variety and abundance of the faecal resistome of wild boar**. *Environ Pollut* 2022, **292**(Pt B):118406.
251. Do TT, Nolan S, Hayes N, O'Flaherty V, Burgess C, Brennan F, Walsh F: **Metagenomic and HT-qPCR analysis reveal the microbiome and resistome in pig slurry under storage, composting, and anaerobic digestion**. *Environ Pollut* 2022, **305**:119271.
252. Ma T, McAllister TA, Guan LL: **A review of the resistome within the digestive tract of livestock**. *J Anim Sci Biotechnol* 2021, **12**(1):121.
253. Begmatov S, Beletsky AV, Gruzdev EV, Mardanov AV, Glukhova LB, Karnachuk OV, Ravin NV: **Distribution Patterns of Antibiotic Resistance Genes and Their Bacterial Hosts in a Manure Lagoon of a Large-Scale Swine Finishing Facility**. *Microorganisms* 2022, **10**(11).
254. Holzel CS, Harms KS, Kuchenhoff H, Kunz A, Muller C, Meyer K, Schwaiger K, Bauer J: **Phenotypic and genotypic bacterial antimicrobial resistance in liquid pig manure is variously associated with contents of tetracyclines and sulfonamides**. *J Appl Microbiol* 2010, **108**(5):1642-1656.
255. Mencia-Ares O, Cabrera-Rubio R, Cobo-Diaz JF, Alvarez-Ordonez A, Gomez-Garcia M, Puente H, Cotter PD, Crispie F, Carvajal A, Rubio P *et al*: **Antimicrobial use and production system shape the fecal, environmental, and slurry resistomes of pig farms**. *Microbiome* 2020, **8**(1):164.
256. Skarzynska M, Leekitcharoenphon P, Hendriksen RS, Aarestrup FM, Wasyl D: **A metagenomic glimpse into the gut of wild and domestic animals: Quantification of antimicrobial resistance and more**. *PLoS One* 2020, **15**(12):e0242987.