Norwegian University
of Life Sciences

Master's Thesis 2022 30 ECTS
Faculty of Veterinary Medicine

When Methods Matter; Disagreement Between Multiplex qPCR and Metagenomic Shotgun Sequencing in Defining the Resistome of Canine Fecal Samples

Oda Marie Bjørgum Karlsen
Chemistry and Biotechnology

Acknowledgements

This master thesis was carried out at the Norwegian University of Life Sciences at the Faculty of Veterinary Medicine in collaboration with the Norwegian Veterinary Institute, Ås, Norway. The thesis was written during the spring of 2022 with a scope of 30 credits.

I would like to thank my external supervisors Øivind Øines, senior researcher, and Håkon Kaspersen, postdoctoral researcher, at the Norwegian Veterinary Institute for the opportunity to write my thesis here and for introducing me to such an exciting field! They have been engaged in my work from beginning to end and have been of great help throughout the whole process! I have enjoyed working with them at my time here. I also want to thank my supervisor Ann-Katrin Llarena, Associate professor, Food Safety Unit, Department of Paraclinical Sciences, for all the help and guidance she has given me. Her feedback throughout the process has been very valuable! I want to thank Mari Røken for letting me use her findings in my method comparison. I would also like to thank senior researcher Eric de Muinck, University of Oslo, for agreeing to oppose my thesis.

A thank you to the HUNT One Health project for the opportunity to use the data that my thesis is based on. The computations were performed on resources provided by UNINETT Sigma2 – the National Infrastructure for High Performance Computing and Data Storage in Norway.

Lastly, I want to thank Andrea Obstsfelder for being a good friend and study partner throughout our whole bachelor and master's degree. I would like to thank my family and friends for the kind words and support. And a special thank you to my grandparents Ellen, Kari and Kjell!

Abstract

Antimicrobial resistance (AMR) is a global health threat to humans and animals. It is of great importance to study and monitor the spread of AMR. This thesis intends to characterize the resistome in 35 canine fecal samples with bioinformatic tools using acquired shotgun metagenomic sequence data, and to compare the resistome results to those obtained using an extended multiplex qPCR method used in an earlier study. Fecal samples collected by dog owners for the HUNT-One Health project in addition to blank samples and mocks was analyzed. FastQC, MultiQC and Trim Galore were used for quality control and trimming of sequence data. AMRPlusPlus with MEGARes was used for AMR gene detection and resistome analysis. Only 8,2% of the AMR genes detected by qPCR were also detected in the metagenomic shotgun sequencing data. There were also AMR detections in the resistome that were not detected using qPCR as the corresponding detections accounted for 22,8% of the total detections made with the metagenomic shotgun sequencing data. Various factors such as extraction and detection method, sequencing depth, and different starting material could explain some discrepancy observed between the qPCR and resistome analysis. This master study does however tell a cautionary tale that resistome results vary depending on analysis method chosen, and that care should be taken when interpreting such results, especially if just one method is used. It would be valuable to have more starting material, increased sequence depth and use multiple resistome pipelines/databases in an effort to describe the “true” resistome better, as well as investigating taxonomic classification to study the bigger picture in the canine fecal microbiota.

List of Abbreviations

AMR	Antimicrobial Resistance
AST	Antimicrobial Susceptibility Tests
bp	base pair
BWA	Burrows-Wheeler Aligner
CARD	Comprehensive Antibiotic Resistance Database
ds	double stranded
GI	Gastrointestinal
HGT	Horizontal Gene Transfer
HUNT	Health Examination in Trøndelag
MAG	Metagenome-Assembled Genomes
MDR	Multi Drug Resistance
MDS	(Non-metric) Multi-Dimensional Scaling
MLS	Macrolide, Lincosamide and Streptogramin
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
NMBU	Norwegian University of Life Sciences
NORM/NORM-VET	Norwegian Surveillance of Resistance in Microbes
NVI	The Norwegian Veterinary Institute
PBP	Penicillin-Binding Protein
PCA	Principal Component Analysis
RGI	Resistance Gene Identifier
SAM	Sequence Alignment/Map
SBS	Sequencing By Synthesis
ss	single stranded
TSS	Total Sum Scaling
WGS	Whole Genome Sequencing

Table of Contents

1 Background

1.1 Microbiota

All individuals, both humans and animals, have a microbiota unlike anybody else which can be compared to a fingerprint in terms of uniqueness. The microbiota is the community of all microorganisms present in a certain environment. Bacteria, viruses, protozoa, and fungi are amongst the microorganisms that inhabit the gastrointestinal (GI) tract and collectively compose the gut microbiota. The composition of the microbiota plays a central role in functions related to metabolism, development of the immune system and to protect against pathogens (Rinninella et al., 2019). When the microbiota is in its healthy state, it is in a so-called normbiosis. However, there can be deviations from normbiosis to dysbiosis, a state associated with several diseases including diabetes, asthma, and irritable bowel syndrome (Casen et al., 2015). A factor that can contribute to dysbiosis is the use of antibiotics (Zhang & Chen, 2019).

Studies shows that humans and dogs have similar microbiota (Deng & Swanson, 2015). The richness and abundance of species varies throughout the GI tract for both humans and dogs. However, the GI tract is shorter in dogs compared to in humans and it has been shown that fecal samples from dogs are reliable in terms of presenting the most relevant taxa. The bacterial phyla *Firmicutes*, *Bacteroidetes*, *Fusobacteria, Proteobacteria* and *Actinobacteria* accounts for the majority of the bacterial sequences that are identified in the GI tract in dogs (Pilla & Suchodolski, 2020). In the human gut microbiota the major bacterial phyla present are similar to the ones in dogs with the exception of *Fusobacteria* whereas *Verrucomicrobia* is the more common phyla in humans (Zhang & Chen, 2019).

1.2 One Health

One Health is a concept used to describe the recognition that human, animal, and environmental health are interdependent, and that a holistic One Health approach involving various sectors is needed to improve and secure health for all. This involves sectors such as research, politics and legislation (World Health Organization, 2017). Amongst the major relevant areas for the One Health approach is the fight against AMR, securing food safety and controlling zoonoses. One Health is especially relevant in the work against AMR, as AMR bacteria knows no species nor geographical boundaries. The World Health Organization

(WHO) acknowledges that the One Health approach is necessary to combat public health threats (World Health Organization, 2017). When it comes to combatting antibiotic resistance, it is crucial to look at humans, animals, and the environment, as resistance bacteria can transmit between these various niches (Figure 1).

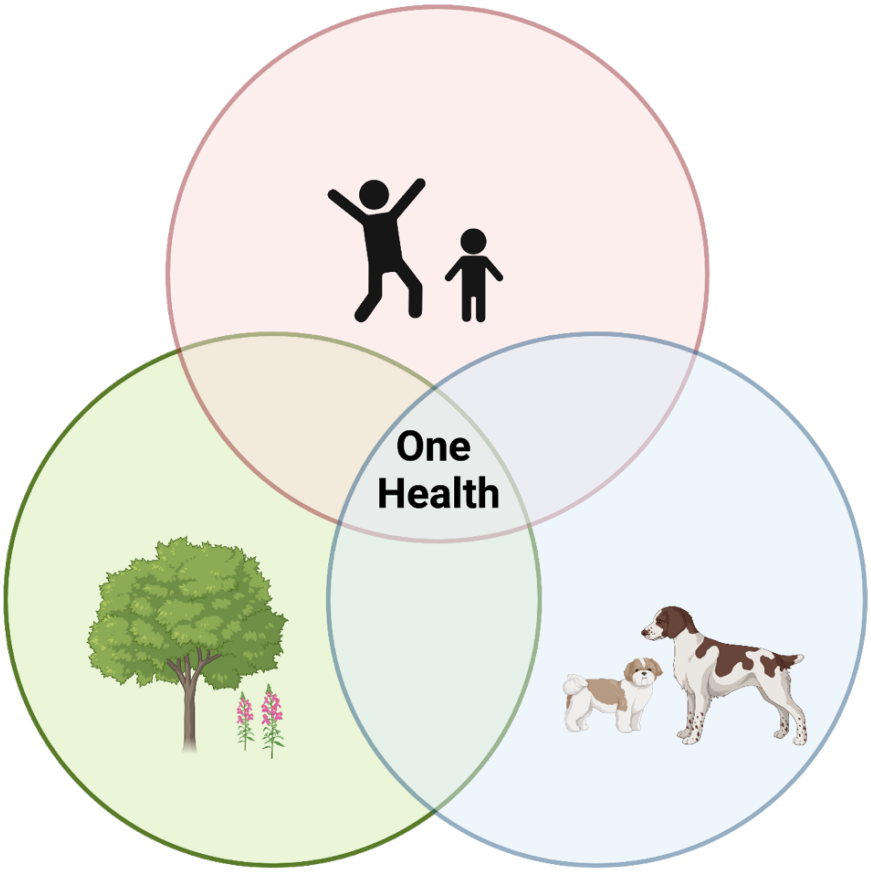

Figure 1: One health: human, environment and animal health are dependent on each other. Created with BioRender.com

1.2.1 **Dogs**

Dogs are known to be man´s best friend, making it one of the animals that have the closest contact with humans. Dogs live under the same roof, eat some of the same food, and even sleep in the same beds as humans. In other words, humans and dogs live in the same environment and are exposed to many of the same microorganisms daily. Since bacteria can transfer between humans and animals there is reason to believe that dogs and humans share many of the same bacteria. This makes the dog an interesting surveillance animal in a One Health paradigm and could mirror the occurrence of AMR bacteria in humans (Thomson et al., 2022).

1.3 Antibiotics & antibiotic resistance

Ever since Alexander Fleming discovered penicillin in 1928, antibiotics have saved countless lives. Since then, several different antibiotics have been discovered or synthesized. Antibiotics can kill or prevent growth of bacteria and are therefore used to treat human and animal infections. In many parts of the world, antibiotics is also used prophylactically, and as a growth promotor in animal husbandry (McEwen & Collignon, 2018).

Antimicrobial resistant (AMR) bacteria are recognized as one of the major threats to global health, development, and food security (World Health Organization, 2021). Balaban *et al.* defined AMR the following way: "the ability of microorganisms to survive and grow in the presence of antimicrobials" (Balaban et al., 2019). Amongst numerous consequences is the fact that infections and diseases will be harder, or near impossible, and more expensive to treat (Bengtsson & Greko, 2014). In ultimate circumstances, simple infections can turn deadly.

The main drivers of AMR includes the misuse and overuse of antibiotics (Aslam et al., 2018). Using antibiotics inappropriately, for instance administrating a too low dose, or prescribing a too-short treatment duration, can result in a sublethal concentration of the antibiotic in an infected tissue. This will provide the bacteria with time to adapt and select resistant sub-clones, resulting in the emergence of AMR bacteria (Milken Institute School of Public Health, 2017). As a result, AMR against all existing antibiotics have been reported (Arukovic et al., 2019). There are numerous mechanisms that ensures AMR in bacteria, some of which arise from mutations in existing gene pools or those that are acquired through horizontal gene transfer (HGT).

However, bacterial resistance towards antibiotics existed long before antibiotics were discovered and used by humans as a way to treat infectious diseases (Leisner et al., 2016). Certain bacteria have natural resistance, such as microorganisms that produce and excrete antibiotics as a defensive mechanism. (Perry et al., 2016). *Pseudomonas aeruginosa* is an example of a bacterium with several intrinsic resistance mechanisms conferring resistance to antibiotics such as macrolides and tetracyclines (Botelho et al., 2019).

A bacterium is categorized as multidrug resistant (MDR) if it is resistant to a minimum of three different types of antibiotic classes (Magiorakos et al., 2012). An example of a bacterium that can be MDR is the methicillin resistant *Staphylococcus aureus* (MRSA), a *S. aureus* that has acquired resistance towards all β-lactam antibiotics, which includes methicillins and penicillins (Shebl et al., 2020) (Lade & Kim, 2021).

1.4 The Antibiotic Resistome

The antibiotic resistome includes all genes that either directly or indirectly contribute towards AMR in an environment (Wright, 2010) (Crofts et al., 2017). There are various ways of defining the antibiotic resistome, depending on the nature of the research. In this thesis, the resistome is defined as the detected acquired resistance genes present in the canine gut microbiota obtained from fecal samples.

Studies have shown that antibiotic presence in a microbial community causes a significant decrease in the richness, structural changes and a selective pressure on AMR genes (Zhao et al., 2019). Depending on the type of antibiotic spectrum, different bacteria present in the gut microbiota are affected. A study on antibiotic selection pressure determination with the use of ciprofloxacin (a fluoroquinolone antibiotic) over the span of six days was performed (Bortolaia et al., 2020). The study revealed a trend for positive selection in class D β-lactamases, and a negative one in class A β -lactamases. After four weeks the AMR gene composition in the microbiota was closer to initial state however remained changed to some degree (Bortolaia et al., 2020). This shows that the antibiotic resistome is dynamic and dependent on factors such as antibiotic use.

1.5 Antibiotic targets

Antibiotics are either bactericidal or bacteriostatic depending on whether they kill the bacteria or slow down their growth. Bactericidal antibiotics, such as penicillins, kills the bacteria by for instance destroying the cell wall and bacteriostatic antibiotics stops the bacteria from multiplying by targeting different processes such as replication, metabolism, and protein production (Norwegian Medicines Agency, 2014). The bacterial ribosome is a common antibiotic target for many clinically relevant antibiotics, where the antibiotic inhibits the ribosome by binding to different sites (Polikanov et al., 2018). An example of this is the antibiotic class macrolides of which binds to the exit of the ribosome tunnel and thereby

inhibits protein synthesis (Nguyen et al., 2019). Another important antibiotic targets includes the cell wall synthesis. β-lactam antibiotics primarily target penicillin-binding proteins (PBPs) to prevent the final step of cell wall synthesis in bacteria (Lade & Kim, 2021).

Different antibiotics target different spectra of bacteria. Some antibiotics act narrowly, whereas others act broadly: the narrow-spectrum antibiotics only target specific bacterial species or genera, while broad-spectrum antibiotics work on several genera or all. Using narrow-spectrum antibiotics are preferred, as it will only affect the bacteria presumed to cause the disease. This will limit the antibiotic exposure to other bacterial species dwelling within the same individual, thereby also limiting the development of AMR. Correct diagnosis and confirmation of causative agents or knowledge-based antimicrobial treatment guidance are of paramount importance. However, when a diagnosis is uncertain and condition critical, the broad-spectrum antibiotics are useful (Melander et al., 2018).

1.6 Antibiotic use in dogs and their owners

Several studies have been conducted to compare the use of antibiotics in humans in different countries. The use of antibiotics in humans in Norway is considerably lower compared to most other countries. A study conducted on paediatric antibiotic use in Norway, Hungary and Portugal revealed that the use of narrow spectrum antibiotics is more common in Norway. Tetracyclines were more frequently prescribed in Norway whereas newer and broad-spectrum antibiotics were commonly prescribed in Hungary and Portugal (Benko et al., 2019). As of 2020, the most prescribed class of antibiotics in Norway to humans was penicillins (NORM/NORM-VET).

In Norway, the development of antibiotic resistance is monitored in humans and animals, by the Norwegian Surveillance of Resistance in Microbes (NORM/NORM-VET), as an action plan against antibiotic resistance (The Norwegian Vetarinary Institute, 2020). The NORM-VET report from 2020 states that the total amount of antibiotics sold to dogs (and cats) has decreased in the recent years and that the most sold antibiotic for dogs is the penicillin amoxicillin in a combination with clavulanic acid (NORM/NORM-VET, 2021). Amoxicillin is a broad-spectrum aminopenicillin and belong to the β-lactams class (European Medicines Agency, 2019). The β -lactamases act in such a way that it binds to the cell wall synthesis enzymes. These enzymes are also called penicillin-binding protein, and by binding to these

proteins, the cell wall is weakened and further preventing cell growth and promoting cell death (Krause et al., 2016). When it comes to the use of antibiotics in veterinary medicine, the European Medicines Agency has certain guidelines which categorizes antibiotics in a scale from A to D (A: Avoid, B: Restrict, C: Caution and D: Prudence), where category D is the first line treatment. The amoxicillin-clavulanate combination is categorized as «C» (European Medicines Agency, 2019).

1.7 Resistance mechanisms

There are numerous ways a bacteria can protect itself and resist antibiotics. However, all these different ways can be divided into three general resistance mechanisms. The first mechanism is pumping of antibiotics out of the cell (efflux). Secondly, the bacteria can prevent interaction between the target and the drug. Finally, the bacteria can modify or destroy the antibiotic compound (Wright, 2005). In this section some of the most common resistance mechanisms are introduced.

Active efflux

Efflux pumps are used by bacteria to actively export/pump unwanted substances out from the cells and thereby control their internal environment. Amongst the substrates that can be exported out of the cells are antimicrobial agents. The specificity of the efflux pumps can vary from being active for just a single substrate or to several types of antibiotics. Hence, the efflux pump can work as a MDR mechanism (Piddock, 2006). Efflux pumps are used by bacteria such as *Streptococcus pneumoniae* and *P. aeruginosa* to resist many classes of antibiotics such as macrolides and tetracyclines (Sikri et al., 2018). All bacteria encode for efflux pumps, regardless of antibiotic resistance. This suggests that the efflux pumps evolved for other reasons than just combatting antibiotics (Cox & Wright, 2013).

Modification of antibiotic targets

Secondly, a common resistance mechanism is the modification the antibiotic target. Antibiotic target site resistance is often caused by mutations in a gene located on the chromosome as well as selection for these mutations in the presence of the antibiotic. Quinolone resistance (in for example *S. pneumoniae*) frequently occurs as a result of alteration of the antibiotic target enzymes (Jacoby, 2005). In addition, this resistance can be acquired by HGT from other bacteria of which acquire this mutation. The result of this

resistance mechanism is that the antibiotic binds less effective or not at all to the target and thereby the bacteria survives (Lambert, 2005).

Enzymatic modification or degradation of antibiotic

Another common resistance mechanism amongst bacteria is that they can induce alteration or even degradation of the antibiotic molecule. The alteration or degradation of the microbial leads to a reduction or elimination of the antimicrobial activity (Harbottle et al., 2006). An example is the modification or cleaving of the β -lactam ring of penicillins executed by the β-lactamases leading to antibiotic inactivation (Wright, 2005).

Changing the permeability of the cell membrane

A final common resistance mechanism is modification of permeability of the cell membrane and with this restricting the antimicrobial agents from reaching the target sites. The cytoplasmic membrane works as a barrier and separator between the external and internal environment in bacteria. Decreasing the permeability of the cell membrane leads to a decreased fluidity, which has negative impacts on the membrane proteins when it comes to activity and structural factors. Therefore, certain bacteria have created a permeability barrier on the outside of the membrane. This permeability barrier is a resistance mechanism as it prevents the antimicrobial substances to getting to the target sites (Cox & Wright, 2013).

1.8 Ways bacteria acquire antibiotic resistance

1.8.1 Mutations

The National Human Genome Institute defines a mutation as the following: “A mutation is a change in the DNA sequence of an organism (National Human Genome Institute, 2022b). Mutations cause changes in the genotype which can lead to changes in phenotype such as the acquisition of antibiotic resistance. Two major types of mutations include base substitution/ point mutations and frame shift mutations (insertion/deletion). A single nucleotide polymorphism is a mutation where there is one nucleotide difference between two strands from the same region of DNA in two different individuals.

Mutations has been one of the main drivers of evolution (Arenas et al., 2018). Mutations as an evolutionary force has been crucial to create genetic variation, ensuring not only survival of the fittest clone, but ultimately creation of novel species. It is only natural that mutations can be a result of bacterial adaption to antibiotics by creating AMR clones.

Mutations that contribute towards AMR usually occurs in genes which encode for antibiotic targets, transporters, or regulators. An example of a mutation that can happen is in the MDR gene *MarA,* in which mutations can contribute to a higher expression of efflux pump genes (Barbosa & Levy, 2000). However, even though the mutation conveys resistance does not make it a resistance gene. This is because the gene was already present in the bacteria with the function that it had before the mutation. Although mutations have contributed to the phenotype AMR, the gene with the mutation should not be classified as an AMR gene. This could lead to inaccurate conclusions when it for instance comes to metagenomic assessments based on sequence (Martinez, 2014). Therefore, as well as for the simplicity of this thesis, only the acquired AMR genes that have been horizontally transferred will be of interest.

1.8.2 **Horizontal gene transfer**

Bacteria have several natural mechanisms to transfer, exchange and take up genetic material. The occurrence of HGT can be traced back to the origin of bacteria and has since then allowed bacteria to evolve, adapt and survive in different environments (Villa et al., 2019). HGT includes the three mechanisms transformation, transduction and conjugation (McInnes et al., 2020) (Burmeister, 2015) and the mechanisms are demonstrated in Figure 2. As HGT has been crucial for the evolution of the bacteria that has survived until this day, it also causes threats. Bacteria can share and acquire AMR genes through HGT, and they can take up genetic material either from other bacteria or from the environment. If the genetic material taken up by the bacteria contains AMR genes, the bacteria has acquired antibiotic resistance genes and can either evolve with these or lose them. Bacteria might acquire multiple resistance genes through HGT. The AMR genes are not necessarily expressed though they are acquired.

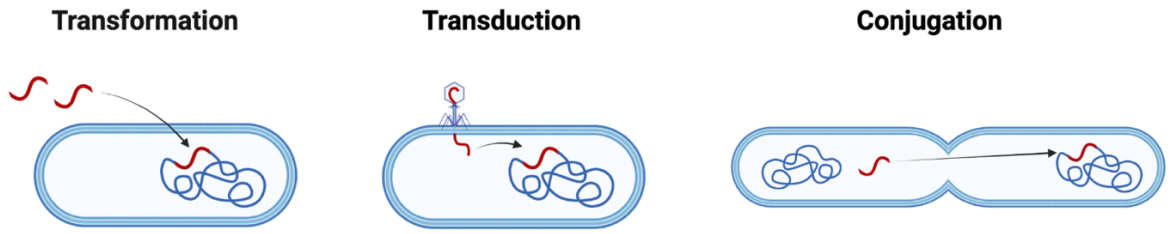

Figure 2: The 3 types of horizontal gene transfer (HGT): Transformation, Transduction and Conjugation. Transformation is the process of which a bacterium takes up genes from the environment. Transduction is when bacteriophages are used to move genes from one bacterium to another. Conjugation is when there is a direct gene transfer between the bacteria (Burmeister, 2015). Created with BioRender.com.

1.9 Detection of AMR

There are multiple methods for detection of AMR and these can be separated into phenotypic and genotypic methods. Phenotypic methods include the gold standard antimicrobial susceptibility tests (ASTs) and genotypic methods here are molecular tools for AMR detection.

1.9.1 Gold standard antimicrobial susceptibility tests

ASTs are different methods that can be performed to investigate what specific antibiotic a bacterium is phenotypically susceptible to and thereby evaluate the antibiotic resistance. These tests are for instance used in surveillance and in clinical laboratories to ensure an effective treatment for patients by detecting possible presence of resistance prior to the treatment. All tests provide qualitative results, and some also provide qualitative results such as minimum inhibitory concentration (MIC) or epidemiological cut-off value (ECOFF) (surveillance of animals or environment). Amongst the most frequently used ASTs are disk diffusion and agar/broth dilution (Mercer et al., 2020). Common for the different ASTs is that they in general provide an accurate detection of resistance for isolates. However, as the antibiotic resistance is increasing in occurrence, it is crucial to pay close attention to the selection of test methods to ensure a still correct AMR detection (Balouiri et al., 2016; Jorgensen & Ferraro, 2009).

1.9.2 Molecular tools for AMR detection

In this section a brief explanation of some relevant molecular tools for AMR gene detection; PCR and the sequenced based methods Sanger and next-generation sequencing (NGS).

1.9.2.1 PCR based AMR gene detection

PCR is a frequently used method for detection of AMR genes due to its high sensitivity and specificity (Aarts et al., 2001). The method is based on using polymerase enzymes to copy specific regions of AMR genes via repeating temperature cycles (denaturation, annealing and elongation) to make an exponentially increasing number of copies of the targeted DNA, which are thereafter detected using fluorometric methods or separation and staining of the amplified bands on a gel, or other downstream analyses.

Multiplex PCRs can detect multiple AMR genes simultaneously in a single assay. Studies show that multiplex PCR can be an accurate approach for simultaneous detection of different AMR genes, making it an optional tool for important matters such as identification and surveillance of antibiotic resistance genes (Strommenger et al., 2003).

1.9.2.2 Sanger sequencing

The human genome sequencing race ended in 2003 when the first ever human genome was completely sequenced. During the Human Genome Project, Sanger sequencing was used and the cost of a single sequenced human genome in 2001 was around 100 million dollars (National Human Genome Institute, 2021). Sanger sequencing is based on DNA polymerase adding fluorescent nucleotides on to a DNA template stands, which grows by one nucleotide at the time. The identification of the nucleotide happens with the use of fluorescence tags on the nucleotides. Sanger sequencing obtains high quality results for relatively long stretches of reads (up to 900bp) (Illumina, 2022), but it is challenging to have multiple targets and can be laborious.

1.9.2.3 Next generation sequencing

Nowadays, more efficient, and cheaper sequencing methods allowing larger genomic templates such as whole genomes and metagenomes to be generated/characterized/sequenced. The collective of these new sequencing technologies is referred to as NGS.

While it earlier took a decade to sequence the human genome using Sanger sequencing, NGS has shorten this time down to a day. Along with the turnaround time, the sequencing cost dropped to below ~1500$ in 2015 (National Human Genome Institute, 2021). NGS technologies are distinguished from Sanger sequencing as instead of sequencing one

fragment at a time, multiple DNA fragments are sequenced in parallel (Illumina, 2022). There are multiple technologies that are a part of the NGS technologies, which take into use different platforms (Hu et al., 2021).

1.9.2.3.1 Short read sequencing technologies

Short read sequencing technologies have in common that millions of shorter reads (smaller DNA fragments of varying lengths) are sequenced in parallel. These technologies allow for multiple individual targets to be sequenced in parallel which require less time, are less expensive, require less space (microchip) and each read is generally shorter than the typical read length from Sanger sequencing. (Behjati & Tarpey, 2013). A frequently used NGS technology platform is Illumina, which is based on sequencing by synthesis (SBS) technology. The SBS method has four major steps, which are sample preparation, cluster generation, sequencing and data analysis (Illumina, 2017).

Sequencing by synthesis

The DNA is sequenced while attached to a flow cell, a unit that physically is placed within the sequencer. Flow cells consist of channel lanes embedded on a glass slide. The coat is made up of two different oligonucleotides, namely short synthetic ssDNA and ssRNA. The DNA subjected to sequencing is randomly fragmented and adapters are ligated to both the 3`end and 5`end. With the help of the adapters and the ssDNA fragments on the lanes, the fragmented template DNA is bound to the flow cell. Bridge amplification occurs as nucleotides and enzymes are added and the ssDNA becomes dsDNA on the surface of the flow cell. Denaturation occurs, leaving ssDNA templates attached to the flow cell followed by complete amplification where millions of dsDNA clusters are made. The clusters are generated in the flow cell channels. Sequencing occurs in multiple cycles to determine every base in the fragment, one by one. Each base is determined by adding labeled reversible terminators, DNA polymerase and primers. By then using laser excitation, fluorescence is emitted from each cluster, and this is captured making it possible to determine the base. Steps after sequencing includes alignment and data analysis (Illumina, 2010). The NovaSeq 6000 is amongst the newer models and reduces time and data processing steps by utilizing a two channel SBS technology (Hu et al., 2021).

1.9.2.3.2 Long read sequencing technologies

Long read sequencing allow longer fragments to be sequenced, which is an advantage when it for instance comes to sequencing genomes with highly repetitive regions. Long read sequencing has significant potential to improve several aspects of the sequencing performed. Amongst important aspects that distinguishes this technology from other technologies is the ability to sequence considerably longer molecules, potentially up to millions of bp's long (Nanopore). Examples of long read sequencing technology platforms are Oxford Nanopore Technologies sequencing platform (Oxford Nanopore Technologies, 2022) and PacBio Single Molecule Real Time sequencing (PACBIO, 2022).

1.9.2.1 Sequencing of isolates and metagenomics

Sequencing all the genomic content of a bacterial isolate can provide valuable information on potential phenotype of the isolate and its heritage. However, a severe limitation is the fact that the bacteria must be cultivatable to acquire an isolate for sequencing. Metagenomics is the study of genomic material from multiple organisms at the same time (National Human Genome Institute, 2022a). The study of metagenomics can provide taxonomic and functional, microbial and eucaryotic composition of a matrix. The technique is culture independent, as the whole collection of DNA can be extracted directly from the studied matrix. This makes metagenomics a suitable option when it comes to studying and characterizing the collection of AMR genes in a matrix, the so-called resistome, in complex bacterial communities such as in fecal samples. A typical metagenomic shotgun sequencing workflow is demonstrated in Figure 3.

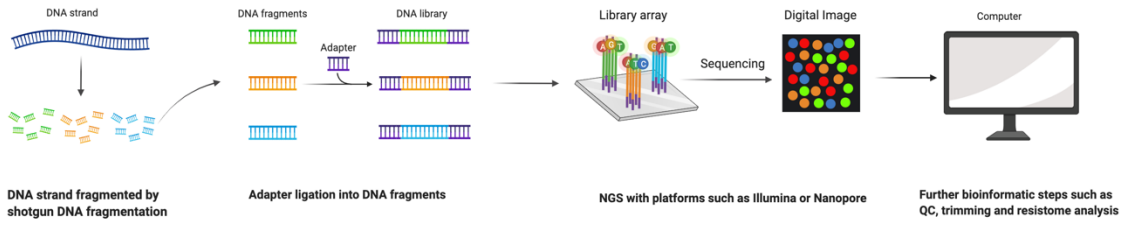

Figure 3: Metagenomic shotgun sequencing workflow. Created with BioRender.com

Several techniques are available for defining the metagenome. Shotgun sequencing is a technique that is used by Illumina and normally 150-300bp can be sequenced on this platform. The DNA is randomly fragmented into smaller pieces and thereby sequenced individually. This is done for both metagenomes and isolate sequencing. For the former,

microbes can be assembled back to the original genome generating metagenomic assembled genomes (MAGs) or by advanced algorithms based on overlap or k-mer based de bruijn graphs (National Human Genome Institute, 2022c).

The field of metagenomics has recently been revolutionized. This is primarily due to improvements regarding sequencing technologies and the field of bioinformatics. It is of great importance to choose both databases and pipelines that are suitable for the study. Available curated AMR gene databases includes MEGARes (Doster et al., 2019), Comprehensive Antibiotic Resistance Database (CARD) (Alcock et al., 2020) and ResFinder (Bortolaia et al., 2020), among others. Tools for resistome analysis includes AMRPlusPlus (Doster et al., 2019), Resistance Gene Identifier (RGI) (Alcock et al., 2020) and GROOT (Rowe & Winn, 2018).

1.10 HUNT and HUNT – One Health

HUNT stands for the Health Examination in Trøndelag, Norway, and is the longest and largest health cohort survey, including more than 240 000 participants since the project started in 1984. HUNT is Norway´s greatest collection of various health data and biological material from one single population. The project HUNT4 is the most recent survey and was completed in 2019 (Norwegian Veterinary Institute, 2021).

HUNT- One Health is a project inspired by HUNT4. The project is a collaboration between the Norwegian Veterinary Institute, the Norwegian University of Life Sciences and the Norwegian University of Science and Technology. The main goal is to facilitate research investigating the relationship between human and animal health and how these impact each other. The project has a collection of animals feces from dog, swine, cow, pig and sheep with corresponding metadata collected between 2017 and 2019 by owners participating in HUNT4. The fecal material has been subjected to DNA extraction and shotgun metagenomic sequencing. In total, sequence data from approximately 3000 animal fecal samples are currently available to the research community (Norwegian University of Life Sciences, 2021). The metadata available is especially rich for the dogs, where information on health status, diet, use, medicine-use, and surroundings is collected.

1.11 Purpose of study

To prevent spread and development of AMR, efficient and correct surveillance is of paramount importance. When new methodology rise, such as metagenomics, comparisons of performances are needed. Therefore, this thesis aims to test the hypothesis of agreement in resistomes in canine fecal samples determined by two methods; a multiplex qPCR and metagenomic shotgun sequencing. To test this hypothesis, the resistome of 35 canine fecal samples were characterized using bioinformatic tools on metagenomic datasets acquired from HUNT One Health. The findings were compared to the multiplex qPCR results, obtained by Røken and colleges (Røken et al., 2022), in terms of presence and absence of 34 different AMR genes.

Objectives:

- Characterize the resistome in the feces of 35 dogs using shotgun metagenomic datasets and bioinformatic tools
- Compare the resistome achieved by analysis of shotgun metagenomic datasets with those of an extended multiplex qPCR for the same 35 canine fecal samples.

2 Materials & Methods

The sequence data used in this thesis was obtained after sampling, extraction and sequencing had been performed. For the readers information, these steps are briefly explained in the material section. In addition, a short description on how the qPCR presence/absence data was obtained is provided in the material section as this data is unpublished. Both methods are based on the same 35 canine fecal samples and sampling methods, but the sample preparation and extraction differed. The bioinformatic steps performed as a part of this thesis are provided in the method section. The overall workflow of the project from sampling to bioinformatic analysis is shown in Figure 4.

Figure 4: Workflow of the project from collecting dog fecal samples to resistome analysis. Created with BioRender.com

2.1 Materials

2.1.1 Sampling, sample selection and the multiplex qPCR results

The fecal samples were collected through the HUNT One Health study. In short, owners in Nord-Trøndelag collected fecal samples from dogs on dry fecal cards, totalling a collection of 1800 samples. Fecal material dried for at least 2 hours on the paper before being shipped to the lab, where they were stored at -20°C. In a separate work performed by Mari Røken and colleagues, the presence of 34 AMR genes in 35 of these fecal samples was tested by a multiplex qPCR method (Røken et al., 2022). In her study, the fecal samples selected were chosen based on the following; only family dogs of good or very good health at the time the sampling was included, and no dogs received antibiotics at the time of sampling.

Data on the presence of AMR genes as generated by the multiplex qPCR in these 35 samples were kindly provided to us by Mari Røken. In short, the qPCR results were generated after bead-beating of the samples followed by DNA extraction using QIAamp PowerFecal Pro DNA kit (Qiagen, GmbH, Hilden, Germany) and a high-throughput qPCR (Røken et al., 2022) performed at NIBIO Svanhovd, Finnmark, Norway. Because of a low sample volume, 34 AMR genes were pre-amplified from the samples using PCR with a primer pool of

forward and reverse primers prior to the use of a qPCR-chip containing 46 assays. In addition to the 35 fecal samples, positive and negative control samples were included in the qPCR analysis. Here the relevant results from Røkens unpublished work (Røken et al., 2022) was extracted and compared to those acquired through a metagenomic approach.

2.1.2 Sequence data

2.1.2.1 Fecal samples

DNA extraction

Extraction of DNA from fecal material streaked on dry paper was conducted in the HUNT One Health project, with the aim of performing shot-gun sequencing. For the readers understanding, the method is briefly described here: Four circles of 8mm in diameter of each fecal card was first homogenized on the FastPrep-24™ Classic instrument, before being subjected a lysis step by proteinase K. Thereafter, inhibitors were removed and gDNA purified using the high-throughput automated QIAsymphony instrument with the QIAamp PowerFecal Pro kit. Every batch of samples were followed by extraction of blanks (paper cards only) and mock communities (cards and ZymoBIOMICS^{TM} Microbial Community Standard II with log distribution, Zymobiomics, USA).

Library preparation, WGS shotgun sequencing and preparing the reads

Sequencing of the gDNA from the fecal material was conducted in the HUNT One Health project. For the readers understanding, the method is briefly described here: Both the library preparation and sequencing was performed by BGI Tech Solutions (Hong Kong, China). Libraries were prepared using the ThruPLEX® DNA-seq by Takara. Libraries were thereafter sequenced on the Illumina NovaSeq 6000 using the 150 bp paired end (PE) sequencing strategy. A minimum of 5GB data per sample, with targeted 20M PE reads per sample, was delivered by BGI. BGI processed the raw data by removing adapters, low quality reads and contamination and the criteria for removal are provided in Table 1. Reads that did not fulfill the criteria was not a part of the dataset used in this thesis.

Table 1: The criteria set by BGI to remove reads in terms of adapter sequence, quality score and N bases.

Reason for removal	Threshold
Adapter sequence	Above 25%
Quality score below 20	Above 50%
N bases	Above 3%

2.1.2.2 Mock and blank samples

Microbial community samples (ZymoBIOMICS™ Microbial Community Standard II with log distribution, Zymobiomics, USA) mixed with clean fecal cards were subjected to DNA extraction, library preparation and Illumina sequencing in each batch of fecal samples. The mock samples contained *Listeria monocytogenes*, *Pseudomonas aeruginosa, Bacillus subtilis*, *Saccharomyces cerevisiae*, *Escherichia coli*, *Salmonella enterica*, *Lactobacillus fermentum, Enterococcus faecalis, Cryptococcus neoformans* and *Staphylococcus aureus* (ZYMO RESEARCH). Clean paper cards (4*8mm in diameter punches) served as the blank control. Blank samples were included in each batch of DNA extraction through to sequencing. Both blank samples and mocks were included in library preparation, sequencing and the downstreams analysis.

In this thesis, 30 blank samples and 31 mock samples were included. The reason behind the high number of blanks was that the fecal samples were allocated on multiple trays and blanks were needed from all trays to identify potential contamination. Since there is no taxonomic classification included in this thesis, the mock positive control samples mainly provide discussion material on the method.

2.2 Methods

The workflow of the bioinformatic tools that were used in this master project is presented in Figure 5. Quality control and trimming was performed before resistome analysis. The resistome analysis in this thesis refers to the resistome defined by the metagenomic shotgun sequence data. Visualization of AMR data from metagenomic sequencing was performed in ResistoXplorer version 2021.11.04 (Dhariwal et al., 2021) and R-studio version 2022.02.1 (R version 4.1.3). The resistome analysis was compared to qPCR in terms of absence and presence of 34 specific AMR genes. The resistome analysis was performed on AMR gene level and are visualized in terms of AMR gene or AMR class. All computations were

performed on SAGA, provided by UNINETT Sigma2 (the National Infrastructure for High Performance Computing and Data Storage in Norway).

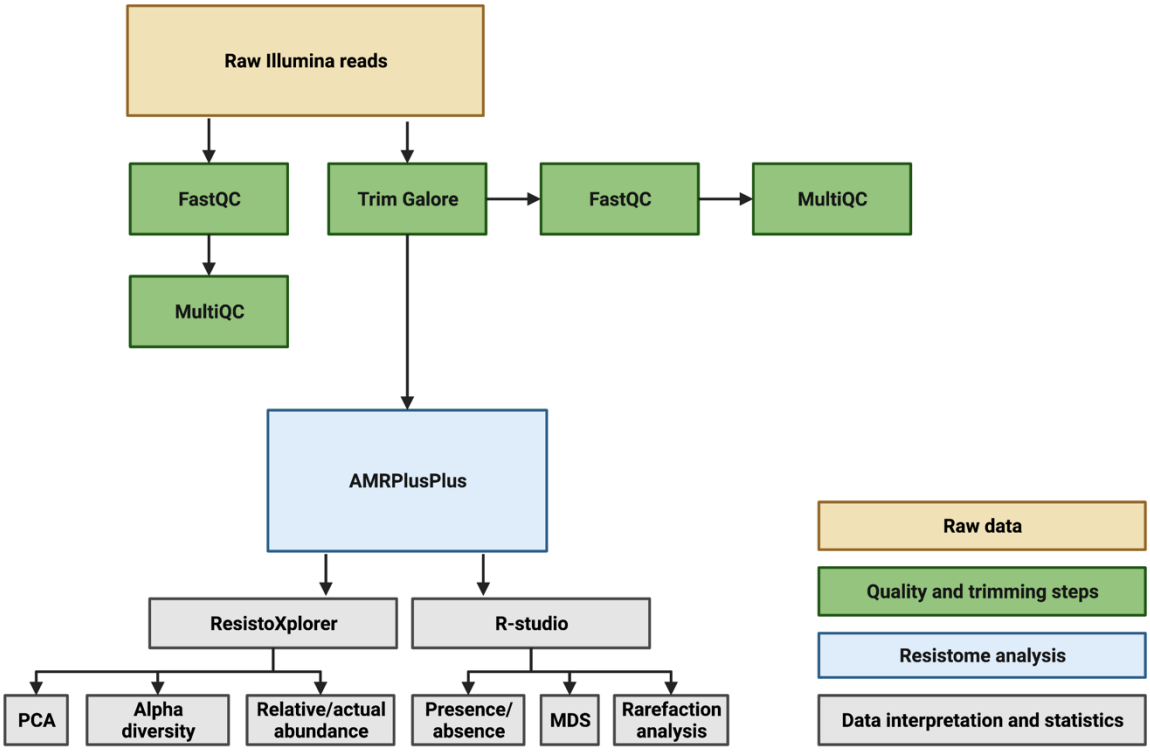

Figure 5: Pipeline depicting the steps of quality control of the Illumina reads data. The green boxes are the steps of quality control and preprocessing of data, the blue box represents the resistome analysis and the grey boxes are the data processing and statistics steps performed with ResistoXplorer or R. Created with BioRender.com.

2.2.1 Quality control

FastQC version 0.11.9 (https://github.com/s-andrews/FastQC), MultiQC version 1.9 (Ewels et al., 2016) and Trim Galore version 0.6.4 (https://github.com/FelixKrueger/TrimGalore), were used for quality control and trimming. FastQC and MultiQC were used to present the quality of the sequence data from BGI before and after trimming with Trim Galore.

Quality control with FastQC and MultiQC

FastQC and MultiQC were used to control the quality of high throughput sequence data. FastQC detects adapters and primers in addition to flagging low quality sequences, duplicates etc. FastQC takes SAM, BAM or fastq files as input, and produce html reports that includes basic read statistics, “per base sequence quality”, “per sequence quality scores” etc.. This gives the user an overview of different modules. The software also assigns a flag to the

different modules in the form of a “pass”, “fail” or “warning”. These modules and flags give an insight to potential obstacles and areas that might be problematic at an early stage. MultiQC is a tool written in Python, which generates one single report summarizing all the FastQC data, giving a nice overview of the sequence statistics and quality in the dataset.

In total, 96 datasets consisting of forward and reverse fastq files from the HUNT One Health project were received. These were generated from 35 canine fecal samples, 31 mock samples and 30 blank samples. FastQC was run on all 192 of the raw compressed fastq files and the script is found in Appendix 1. FastQC gave an output of one html and one zip file per input file. The zip files from the raw data were further used as input in MultiQC and the script can be found in Appendix 3. The MultiQC software was provided with the path to the folder, and it automatically chose the information that it recognized. The output of MultiQC was one html file, which was a summary of all the FastQC html files and a folder containing plots and text files. These steps were repeated for the trimmed data as FastQC was also run on the trimmed reads to confirm adapter removal after trimming. This script can be found in Appendix 2. The MultiQC script on trimmed reads is provided in Appendix 4. Bad quality reads were removed from all samples on basis of criteria in Table 1 by the provider, BGI. Based on this, no data was excluded due to low quality. The range number of reads per sample as well as average number of reads in the different sample types are presented in Table 2.

Trimming with Trim Galore

Whereas FastQC and MultiQC gave an overview of multiple quality measures, Trim Galore is a preprocessing tool based on FastQC. Trim Galore cuts technical and poor-quality sequences. Technical sequences such as adapters and bad quality sequences are removed (Bolger et al., 2014).

Here, Trim Galore was used for adapter removal trimming of all 96 raw datasets of fecal, mock, and blank samples. The settings used were default. The input files were compressed fastq files and the output was one text file and one fastq file for each input file. The text file included a summary of parameters, information about the trimming The output fastq files were further used in the downstream resistome analysis. The Trim Galore script can be found in Appendix 5.

2.2.2 Resistome analysis with AMRPlusPlus

The pipeline AMRPlusPlus version 2.0 with the MEGARes database version 2.0 (Doster et al., 2019) was used to analyze for the presence of AMR genes in the acquired datasets. The inputs in AMRPlusPlus were the MEGARes database (fasta file), and the forward and reverse trimmed reads (fastq format). The pipeline is put together by several software and the ones utilized in this thesis are Burrows-Wheeler Aligner (BWA), ResistomeAnalyzer and RarefactionAnalyzer (Microbial Ecology Group, 2019). The Nextflow version 20.07.1 (Di Tommaso et al., 2017) script is to be found in Appendix 6. The softwares used are briefly explained below.

BWA (Burrows-Wheeler Aligner)

BWA is an alignment tool and is used to align the reads to the MEGARes database (Li & Durbin, 2010) in the AMRPlusPlus pipeline. The output from this process is a SAM file.

ResistomeAnalyzer

ResistomeAnalyzer takes a SAM file as input along with a reference database in fasta format (MEGARes) and an annotation database in CSV format (included in the program). All parameters were set to default. All alignments of the target genes from the SAM formatted alignment file were counted. The output of this program was four text files in TSV format. These four files all represented a different level: class, mechanism, gene and group. The TSV file on gene level was the one used for the resistome analysis in this thesis. The gene level file had four columns containing information on the sample that was analyzed, the level of identification, the number of reads aligned to the targeted level, and gene fraction (nucleotides in reference fasta file aligned to by a minimum of one sequence read).

RarefactionAnalyzer

The RarefactionAnalyzer is a tool performing rarefaction analysis by counting the number of new AMR hits as a function of reads. This is useful to ascertain whether the sequencing depth was adequate for detection of AMR genes. The inputs in this tool were the same as for ResistomeAnalyzer, which included an alignment file (SAM), a reference database (FASTA), and an annotation database (CSV). The outputs of the program were four text files (TSV) with two columns providing information on proportion of reads sampled and the number of unique genes detected. The rarefaction data was obtained from this part of the AMRPlusPlus pipeline and was exported to R- studio for visualization at gene level.

2.2.3 **Sorting of data in R-studio**

The AMR resistome data obtained from AMRPlusPlus was a matrix with data that needed to be sorted for further use to fit the format of ResistoXplorer and to simplify the visualizations. The MEGARes database gave several hits for the same AMR gene in different bacterial species, as seen for the *tetM* gene; MEGARes provided altogether 10 different hits for *tetM*, where both the sequence and length differed between the AMR gene in different bacterial species. The *tetM* with genebank ID X92947 is from *Enterococcus faecalis* while *tetM* with genebank ID FR671418 is *S. pneumoniae*. In this thesis the taxonomic composition of the ARGs is not of importance because the interest lays in studying the presence of the AMR genes regardless of origin bacteria. Therefore, different *tetM* hits were merged and the same is true for all other AMR genes referring to the same gene. An overview of the total amount of unique AMR genes for each sample group is provided in Table 3.

2.2.4 **ResistoXplorer**

ResistoXplorer (Dhariwal et al., 2021) was used to visualize the resistome from data obtained from AMRPlusPlus in terms of alpha diversity, PCA, and relative and actual abundance. The settings used were default which included normalization using total sum scaling (TSS). TSS means that every AMR gene count was divided with the total number of counts for that specific AMR gene. ResistoXplorer did not account for the total number of reads as only the number of hits were imported. Other settings included a low count filter where the default was set to 2 reads. The low count filter was chosen as features with low abundance and prevalence are difficult to distinguish from errors in the sequencing or contaminations. The filtered data was used in visualization of PCA and relative/actual abundance (not for alpha diversity). Figures obtained from ResistoXplorer are Figure 8, Figure 9, Figure 10, and Figure 11.

PCA was performed to reduce the dimensions of the data to show the variety between the samples in two dimensions. The samples are clustered based on similarity and the dimensionality is reduced using permutational MANOVA. Alpha diversity can the richness and/or evenness of the AMR genes within a matrix and can be calculated in several ways (Willis, 2019). The richness indicates the amount of unique AMR genes present in the sample and the evenness accounts the abundance of the AMR genes. Here, the alpha diversity

analysis was conducted based on normalized counts of AMR genes as produced by TSS using the t-test/ANNOVA. The alpha diversity was calculated using three different methods: Shannon, Chao1 and Pielou`s evenness (Jost, 2007).

2.2.5 **R-studio**

All visualizations performed in R-studio was on a gene level. R-Studio was used to visualize rarefaction and resistome data from AMRPlusPlus in terms of rarefaction, beta diversity, and presence/absence. A rarefaction diagram was created based on the data obtained from the RarefactionAnalyzer part of the AMRPlusPlus pipeline on gene level. The comparison of qPCR and metagenomic shotgun sequencing was done by looking at absence and presence of all 35 AMR genes tested for in the qPCR assay. The two samples that were outliers were removed from the qPCR results so a fair comparison could be made as these samples were not sequenced well enough (0M and 0,4M reads, median ≈ 28M reads). AMR genes absent in all samples were removed from the visualization. The criteria for the presence of an AMR gene for the resistome data was set to be a minimum of 10 reads. The reason for this was to remove AMR with low abundance as this could be from for instance contaminations. Beta diversity was calculated using the Jaccard distance. This distance was based on presence/absence data in terms of zeros and ones. The distance was calculated in R and the script is in Appendix 7.

3 Results

The result section includes results on quality, resistome analysis and method comparison between qPCR and metagenomic shotgun sequencing. All analyses were performed on gene level and the visualization in R and ResistoXplorer is based on gene level data.

3.1 Quality

A summary of factors including quality, adapter content, and number of sequences was provided by FastQC and MultiQC. The reads were established to be of overall good quality both before and after trimming (phred score above 28), and the adapter content was removed from the sequences with Trim Galore. This resulted in an adapter content of less than 0,1%.

3.2 Sequencing depth

In this thesis, the sequencing depth refers to the number of reads per sample. There were variations in the sequencing depth across the different sample types as well as within the same sample groups. Figure 6 presents an overview of all samples in terms of number of sequences (M) and AMR gene hits per M sequences. The color of the points represents the sample type where blank samples are red, fecal samples are green and mock samples are blue.

From Figure 6 one can see that the blank samples generally had a lower sequencing depths and fewer AMR gene hits per M sequences compared to the two other groups. Most of the fecal and mock samples contained similar number of reads. However, the number of AMR gene hits per M sequences separated the two clusters as the mock had higher abundances of AMR genes.

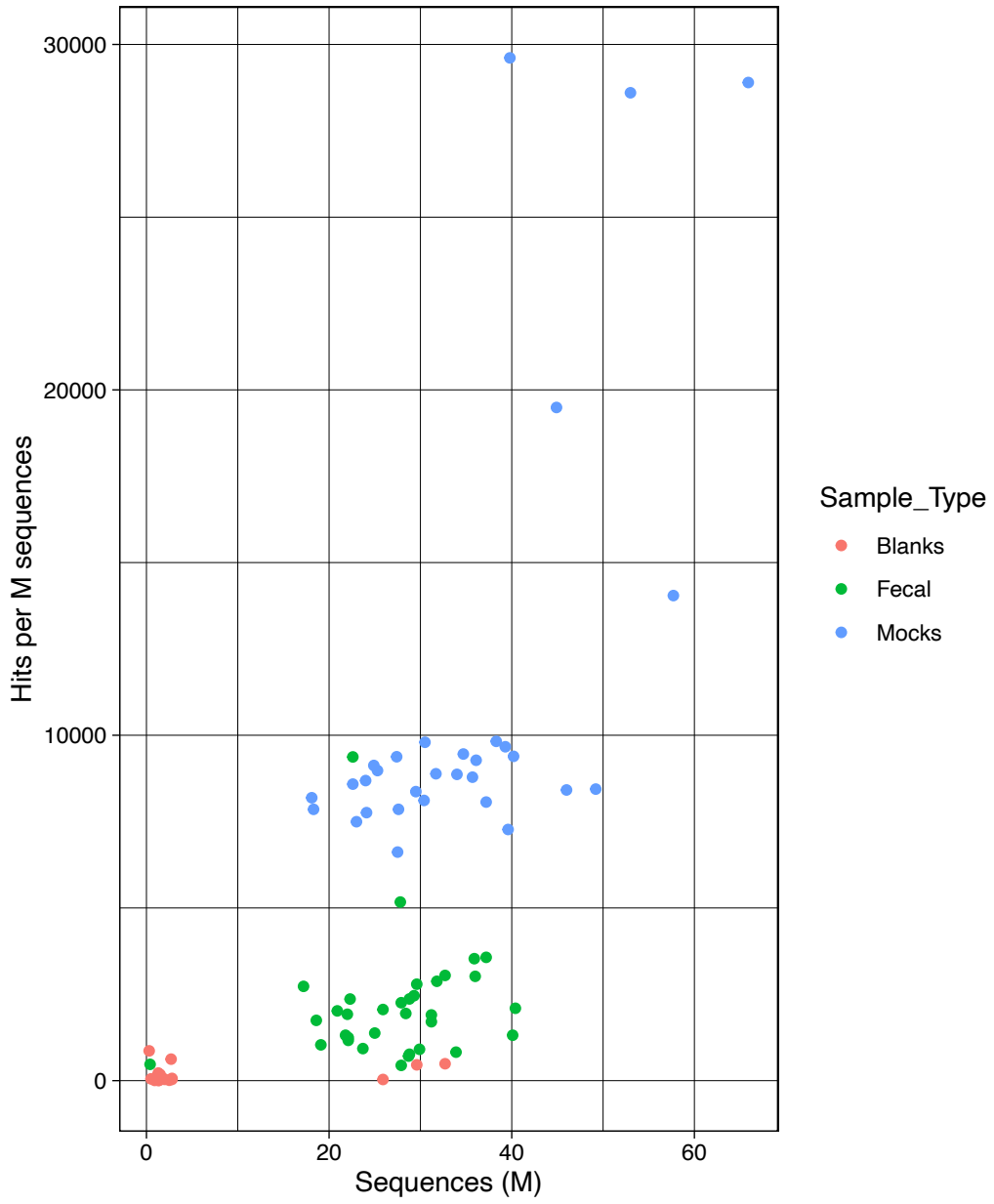

Figure 6: The amount of hits per million sequences vs amounts of sequences in millions. The x-axis provides the number of sequences (M) and the y-axis is the number of AMR gene hits per M sequences.

Amongst the fecal samples, two outliers were identified and excluded from further analysis due to low sequencing depth. The cutoff read depth was determined to be 17.2M reads (Figure 7). The outliers removed had 0.0M and 0.4M sequences per sample and got 0 and 189 hits respectively. After removing outliers, 33 fecal were further used in resistome analysis and method comparison. The sequencing depth (M) of the fecal samples are presented in the boxplot in Figure 7. The arrows show the two outliers.

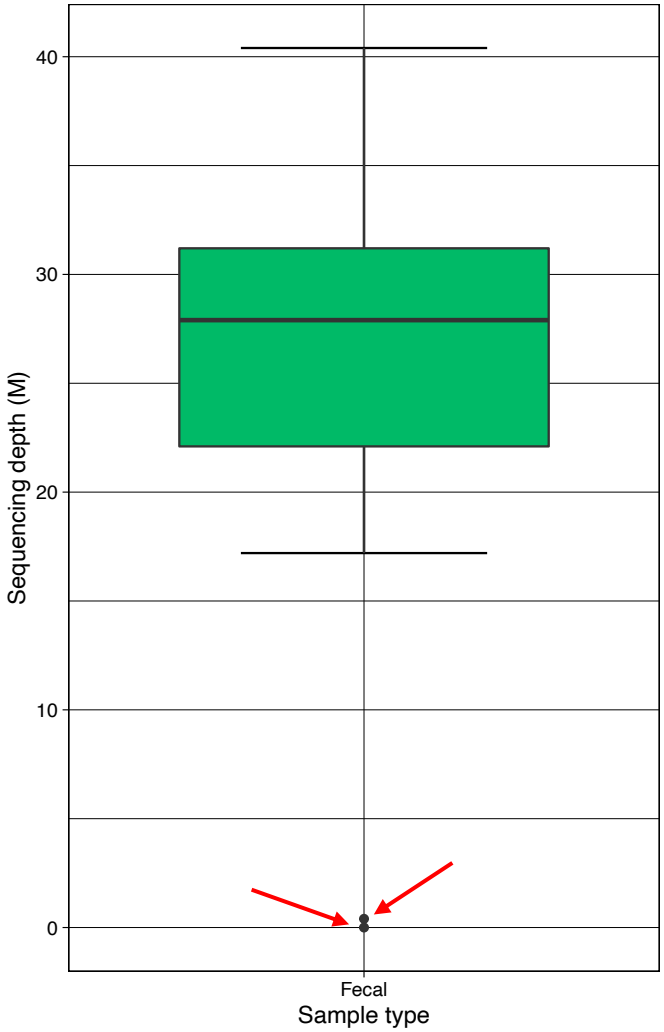

Figure 7: Boxplot of sequencing depth from the fecal samples. Based on too low sequencing depth, two outliers were removed and excluded from further analysis.

The range and average number of reads and AMR gene read hits for all sample types are presented in Table 2, and Table 3 shows the number of samples with AMR gene hits and the number of unique genes in the present altogether in one sample type. The 33 fecal samples on average had 27.1 M reads and had 60 129 hits (reads aligned to AMR genes in the MEGARes database). The sequencing depth per sample ranged between 17.7M to 40.4M reads.

AMR genes were detected in 20 out of 30 blank samples. These samples on average had a sequencing depth of 5.7M reads and from this an average of 1 697 hits. In the 31 mock samples, a total of 128 unique AMR genes were detected. The mock samples had the highest average in terms of sequencing depth and hits per sample with an average sequencing depth of 34.7M reads and 429 824 hits.

Table 2: Table provides data on the range of numbers of reads per sample type (M), the average read per sample type (M), the range of AMR gene hits per sample type and the average hits on AMR genes in the sample types.

Sample type	Range of reads per sample (M)	Average reads per sample (M)	Range of hits per sample	Average hits per sample
Blank	0,3 – 32,7	5,7	5 – 16 022	1 697
Fecal	17,2 – 40,4	27,1	12 336 – 211 813	60 129
Mock	18,1 – 65,9	34,7	143 751 – 1 904 382	429 824

3.3 Resistome analysis

Results from AMRPlusPlus included a matrix containing various information of the different genes from the MEGARes database, such as annotation, class, and mechanism. The amount of hits were numbered for all samples for all AMR genes. The result matrix excluded samples without hits on AMR genes. In this section, results obtained from the AMRPlusPlus pipeline are presented with the use of ResistoXplorer to get an overview of AMR genes present in the canine fecal samples, blanks, and mocks.

3.3.1 Data filtering in ResistoXplorer

The default data filtering in ResistoXplorer was used (Section 2.2.3). In total for the fecal, blank and mock samples there were in total unique 167 AMR genes detected. After the data filtration, 54 unique AMR genes were removed leaving 103 unique AMR genes for the resistome characterization.

3.3.2 Relative abundance per sample

The relative abundance (y-axis) of the different AMR classes is presented in Figure 8. The sample type is indicated with color bars under the x-axis; red sample bar is blanks, green is fecal samples and bule is mock.

The major AMR gene classes present in all 33 fecal samples included tetracycline resistance, macrolide-lincosamide-streptogramin (MLS) resistance, and elfamycine resistance. In two fecal samples, 50% of the AMR genes belonged to the β-lactamase resistance class. MDR mechanisms, fluoroquinolone resistance and elfamycine resistance genes accounted for most of the resistance classes in the mock samples (more than 90% of the AMR genes detected).

The relative abundance of the MDR mechanisms was around 50% in all mock samples and represents the biggest portion of AMR genes in this sample type. In blank samples MLS resistance had a relative abundance above 75% in six samples and in two samples the relative abundance of elfamycines was 1. Two mock samples also had a relative abundance of above 50% for MDR mechanisms. The results are visualized in Figure 8.

Table 3: The total number of samples, the number of samples with hits and the amount of unique AMR genes summed up for each sample type.

Sample Type	Total number of samples	Number of samples with AMR hits	Number of unique AMR genes
Blank	30	20	40
Fecal	33	33	124
Mock	31	31	128

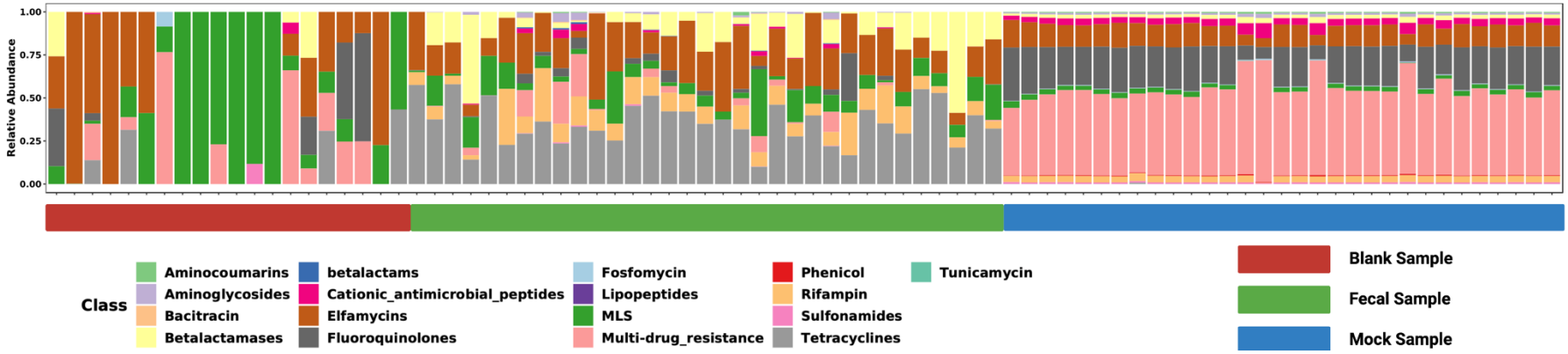

Figure 8: Relative abundance of AMR gene classes in blank (red), fecal (green) and mock samples (blue). The color of the bars represents the AMR class. The color of the x-axis shows the sample type, and the y-axis shows the relative

3.3.3 Relative and actual abundance per sample group

The relative abundance of AMR classes within the sample groups are shown in Figure 9. One can see that in general, the same classes of AMR resistance are present in all types of samples with varying abundances. MDR mechanisms dominates within the mock samples, tetracycline resistance has the highest abundance within the fecal samples and elfamycine resistance dominates within the blank samples.

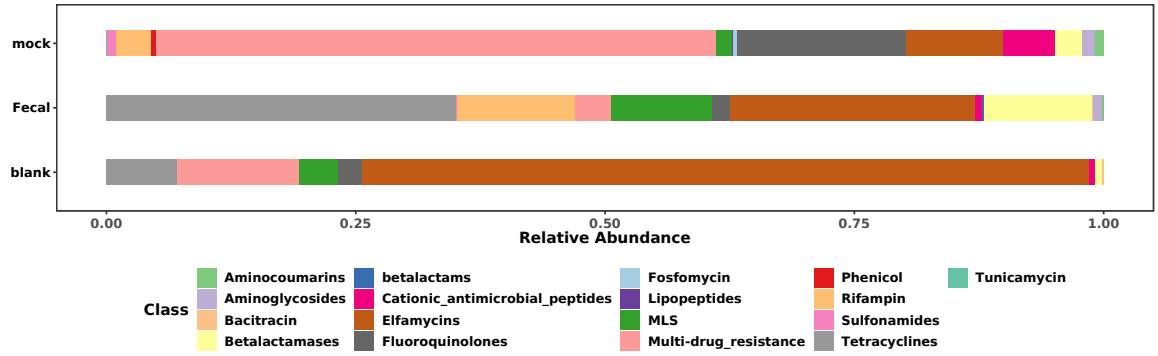

Figure 9: The relative abundance of different AMR classes presented for each of the three sample groups: blanks, mocks, and fecal samples. The x-axis presents the relative abundance levels, and the y-axis presents the sample groups.

The composition of classes varies across the sample types. Another factor creating visible differences between the sample types is the actual abundance (Figure 10). There is a big difference in the actual abundance. In the mock samples, there are generally higher levels of actual abundance compared to the fecal and blank samples. Conclusively, low amounts of AMR genes were found in fecal samples and close to neglectable amounts were detected in blank controls.

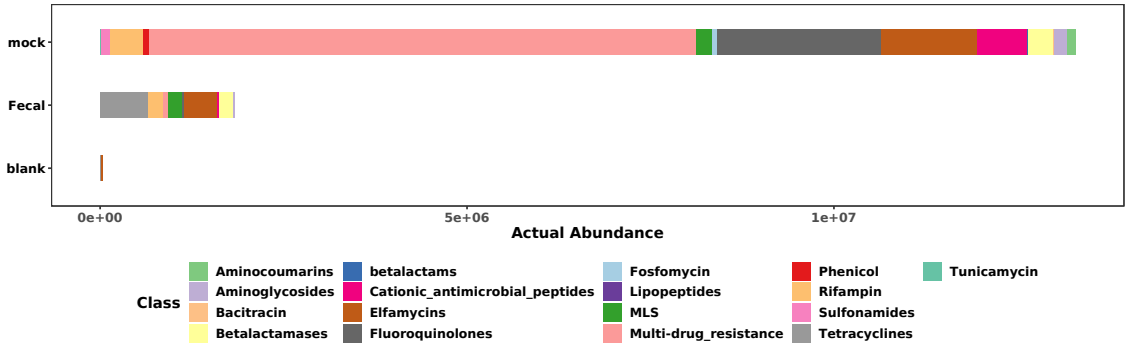

Figure 10: Actual abundance of the different antibiotic classes the AMR genes are resistant towards for all three sample types. The x-axis represents the actual abundance of the classes (colors) and the y-axis provides the sample type (mock, fecal and blank samples).

3.4 Alpha diversity

Here, the alpha diversity reflects the richness and/or evenness of AMR genes within the different sample types. The richness in a sample indicates the amount of unique AMR genes present, and evenness indicates the abundance of AMR genes present in the sample. The diversity of the resistome in the samples are demonstrated with Figure 11.

Shannon takes both the richness and evenness into account (Figure 11A). The medians of the datasets are clearly distinguished for blank, fecal and mock samples with medians of around 0.7, 2.2 and 3.3 respectively in terms of the Shannon index. Chao1 accounts for richness only (Figure 11B). Also here are the medians and majority of data are well separated in terms of the Chao index. The blank and mock samples are the most different. Pielou's evenness index only considers evenness (Figure 11C). The mock sample data completely overlaps with the blank and fecal samples in terms this index, meaning that approximately 50% of the mock datasets cannot be distinguished from the other two in terms of evenness with Pielou´s evenness index.

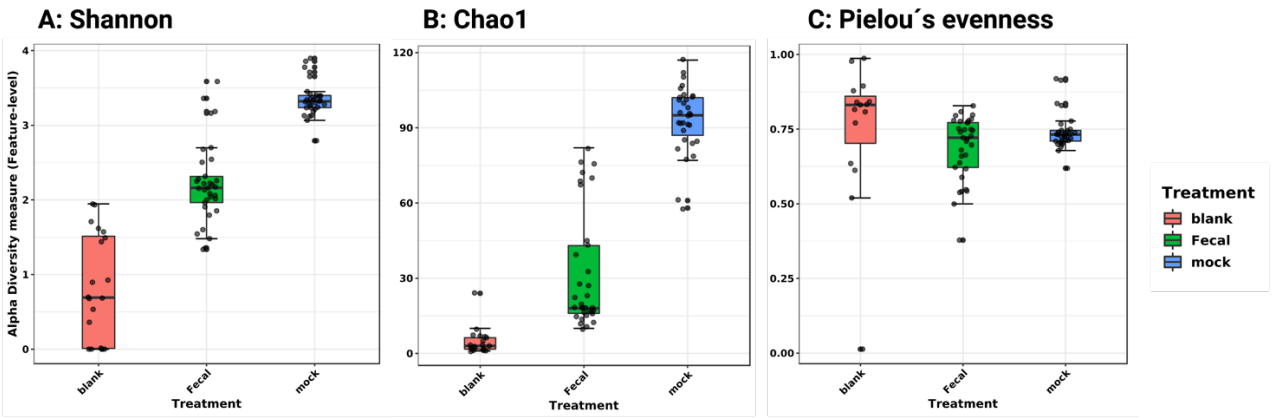

Figure 11: Alpha diversity calculated using three different diversities measurements: Shannon (A), Chao1 (B), and Pielou`s evenness (C). The x-axis presents the sample type and the y axis is the alpha diversity measure with different indices depending on the diversity measurement used.

3.5 PCA

Strength and statistical significance of the sampling groups are shown in the PCA plot in Figure 12. This shows beta diversity based on the AMR genes to show the difference between the different groups. The high dimensional data is reduced into PCA1 and PCA2. PCA1 (x-axis) explains 45.8% of the variation and PCA2 (y-axis) explains 21% of the variation. The PCA method is arbitrary, and the distances plotted are based on the input data.

Figure 12 shows three clusters which are somewhat separated. The mock samples are clearly separated from the blank and fecal samples over the PC1. The mock samples have a lot of variation explained by PC1, whereas the fecal samples have a higher variety explained by PC2. There is slight overlap between fecal and blank samples.

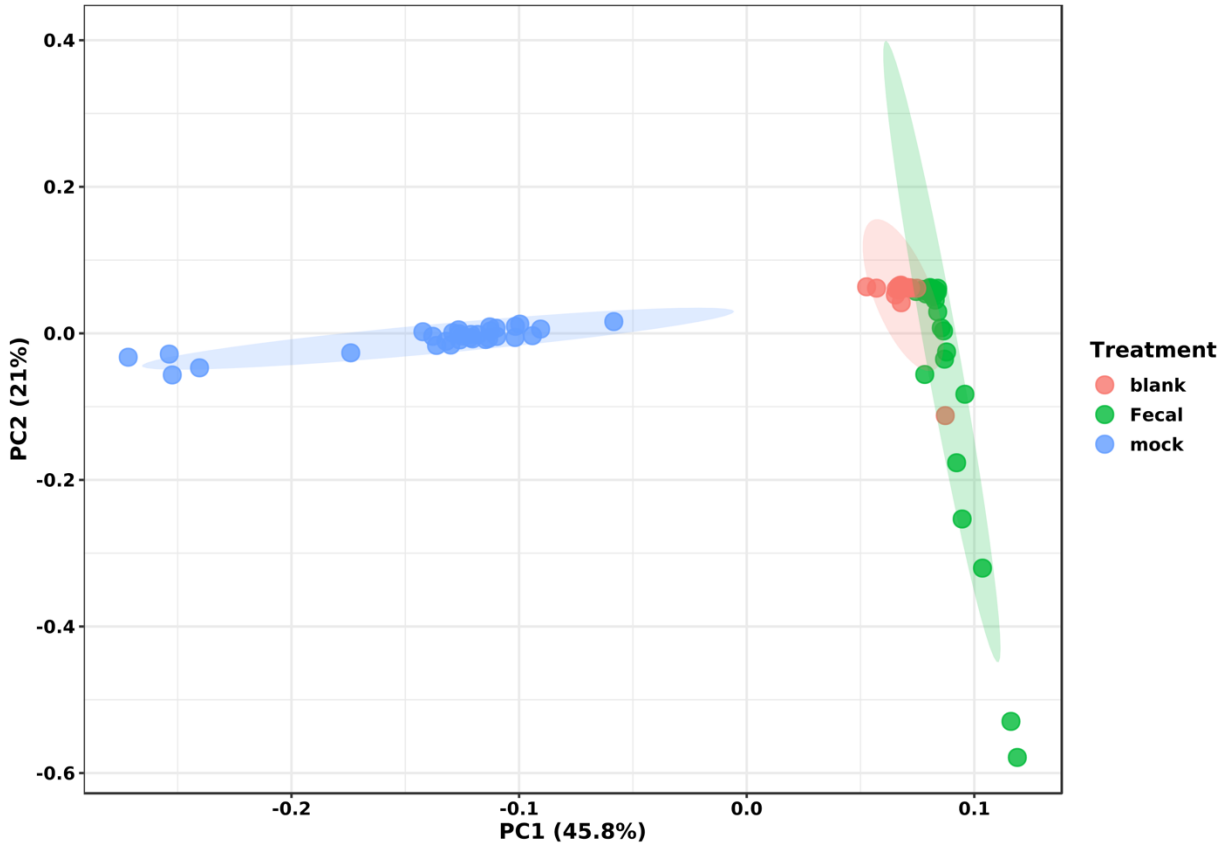

Figure 12: PCA plot for blank (red), fecal (green) and mock samples (blue). The x-axis, PC1, explains 45.8% of the variation and the y-axis, PC2, explains 21% of the variation.

3.6 Rarefaction analysis

A rarefaction analysis was conducted to investigate if the samples had been sequenced deep enough to capture the diversity of the AMR genes present. The rarefaction analysis was performed on gene level. Figure 13 shows the rarefaction plots for all sample types and Figure 14 shows the rarefaction plot for the fecal samples only. The analysis subsamples of the total reads, and identifies how many unique AMR genes are identified within those reads. According to Figure 13 some samples in each group (mock, negative, fecal) have a flattening rarefaction curve, indicating that the sequencing effort was sufficient for detection of the AMR genes present. However, there are also some of the fecal and mock samples that are

still left with rising rarefaction curves, meaning that more reads per sample might be necessary for better characterization of the ARGs in these samples. There is one blank sample that has a higher number of unique AMR genes compared to all other blank samples.

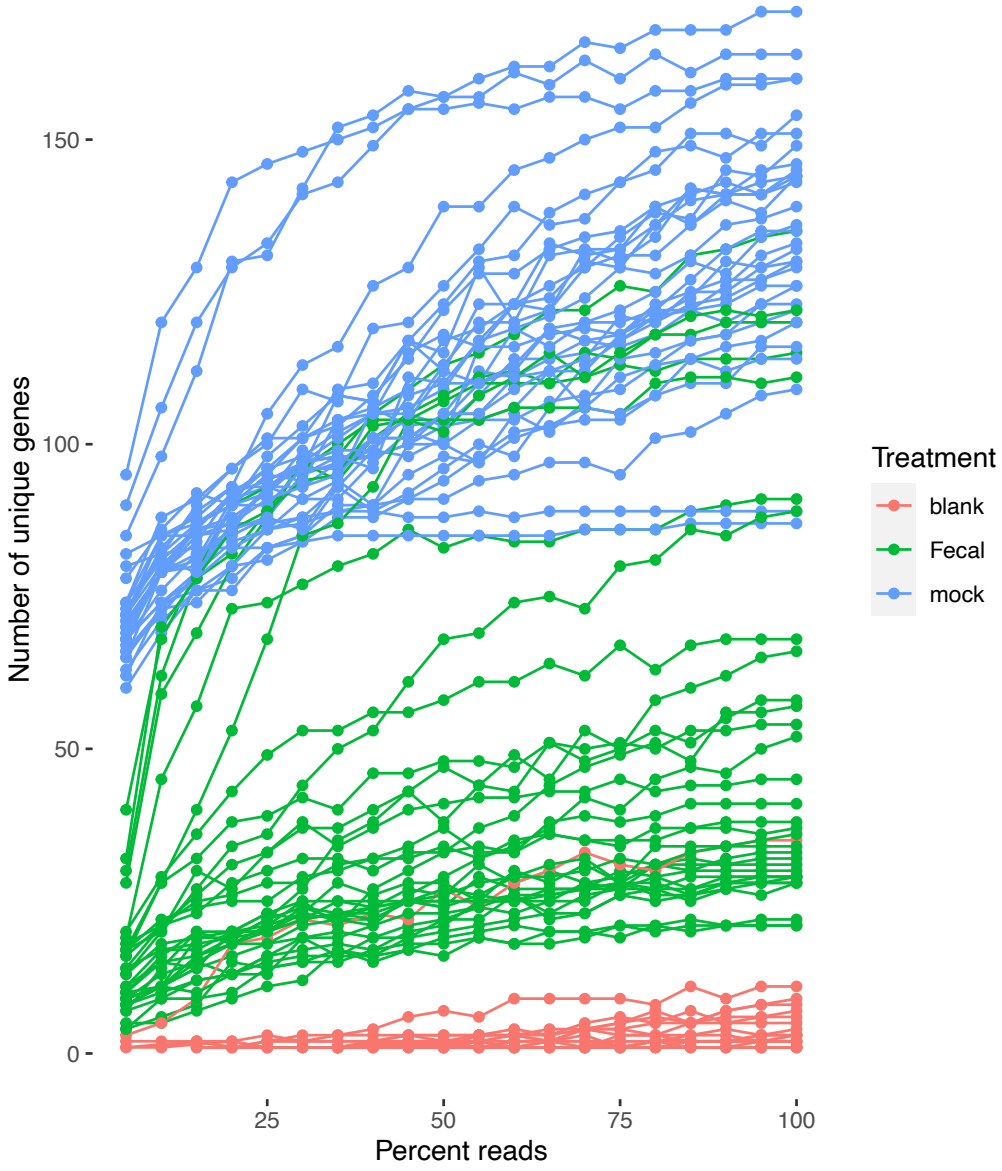

Figure 13: Rarefaction analysis for blank (red), fecal (green) and mock samples (blue). The rarefaction plot shows the number of unique genes (y-axis) found in an increasing percentage of total reads.

Figure 14 presents the rarefaction plot for the fecal samples alone. There was an increase in number of unique genes as the 100% read mark was approached for certain samples. This was generally the case for samples with general higher numbers of unique AMR gene hits.

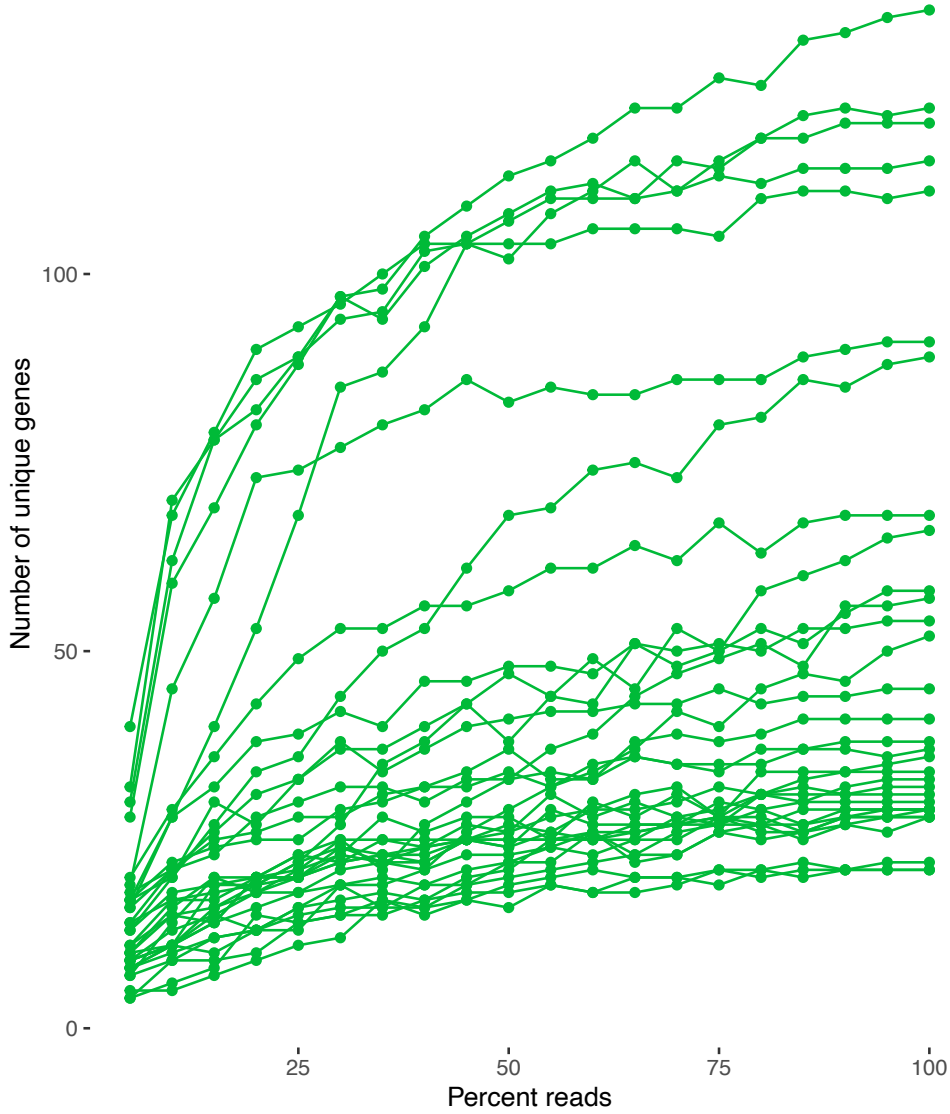

Figure 14: Rarefaction plot showing only fecal samples. The rarefaction plot shows the number of unique genes (y-axis) found in the different percentage of reads. Fecal samples with general higher number of unique AMR genes (above 50) tend to rise more, indicating a too low sequencing depth.

3.7 Comparison of resistome defined by qPCR and metagenomic sequencing

The multiplex qPCR assay could detect 34 specific AMR genes, while the metagenomic sequencing and resistome analysis could detect approximately 8000 curated AMR genes (Doster et al., 2019). The results achieved by qPCR and metagenomic sequencing were compared to investigate if the results correspond regarding the 34 AMR genes. Only 33 fecal samples were subjected to the comparison as these were the only samples that could be compared to the qPCR method while simultaneously having sufficient sequence depth.

A presence/absence of shared detectable ARGs was constructed and presented in Figure 15. Here, the 33 samples are numbered on the y-axis and the name of the AMR genes are presented on the x-axis for both the qPCR and resistome method. The number of reads in all fecal samples are presented to the right. AMR genes absent in all samples are not included to simplify the figure. However, it is important to keep in mind that there were other AMR genes with that were not detected using either method.

A total of 23 out of the 34 AMR genes (67.7%) were detected altogether in the 33 fecal samples with the use of qPCR. With the resistome method, 11 out of the 34 AMR genes (32.4%) were detected. Table 4 provides the number of samples with the different AMR genes detected with both methods, along with the number of samples where the AMR gene detections corresponded. The percentages of corresponding samples as a part of qPCR detection and resistome detection are presented for each AMR gene.

In the qPCR results 219 detections of the 34 AMR genes in the 33 canine fecal samples were made. By comparing the AMR detections to the ones in the resistome analysis, a total of 8.2% AMR hits corresponded. Out of the 79 AMR gene detections that were made in the resistome analysis, 22.8% of these were also made by qPCR.

In total there were 18 instances of correspondence between the two different methods. The AMR genes *tetA* and *tetB* were detected in 5 and 2 out of 33 samples with the use of qPCR detection, respectively, while resistome analysis detected *tetA* and *tetB* in 29 and 30 samples, respectively. The resistome method detected 100% of the samples that were detected in qPCR and the corresponding detections accounted for 17.2% and 6.67% of the total detections made in resistome. *ErmB* had a high prevalence with qPCR being detected in 29 samples. The same AMR gene was detected in two samples in the resistome analysis, which corresponded 100% with the qPCR despite that they were only detected in 6.90% of the samples compared to qPCR. *TetM* with a prevalence of 31 samples with qPCR was detected in 5 samples (16.1%) in resistome.

Table 4: The table shows the number of samples of which the different AMR genes were detected with qPCR, resistome analysis and the number of samples in which both methods had corresponding findings in terms of presence/absence. The corresponding presence in both methods is divided by the qPCR presence to see the percentage of samples where the AMR was also detected using resistome analysis. In addition the corresponding presence is also divided by the resistome presence to see the percentage of samples in which the AMR genes were detected only with the resistome method.

AMR gene	qPCR presence	Resistome presence	Corresponding presence	Corresponding presence / qPCR presence	Corresponding samples / resistome presence
aac6	4	1	0	-	-
ant3	19	0	-	-	-
aph3	18	1	1	5.56 %	100 %
blaACT	1	0	-	-	-
blaCTX	0	5	-	-	-
blaDHA	1	1	1	100 %	100 %
blaSHV	1	0	-	-	-
blaTEM	13	0	-	-	-
DfrA	3	0	-	-	-
ermB	29	2	2	6.90 %	100 %
ermF	15	3	2	13.3 %	66.7 %
floR	2	0	-	-	-
intl	7	0	-	-	-
mecA	7	0	-	-	-
oqxA	1	1	0	-	-
oqxB	1	1	0	-	-
strA	16	0	-	-	-
strB	13	0	-	-	-
sul1	19	0	-	-	-
sul2	11	0	-	-	-
tetA	5	29	5	100 %	17.2 %
tetB	2	30	2	100 %	6.67 %
tetM	31	5	5	16.1 %	100 %
vanA	1	0	-	-	-
Total	219	79	18	8.22%	22.8%

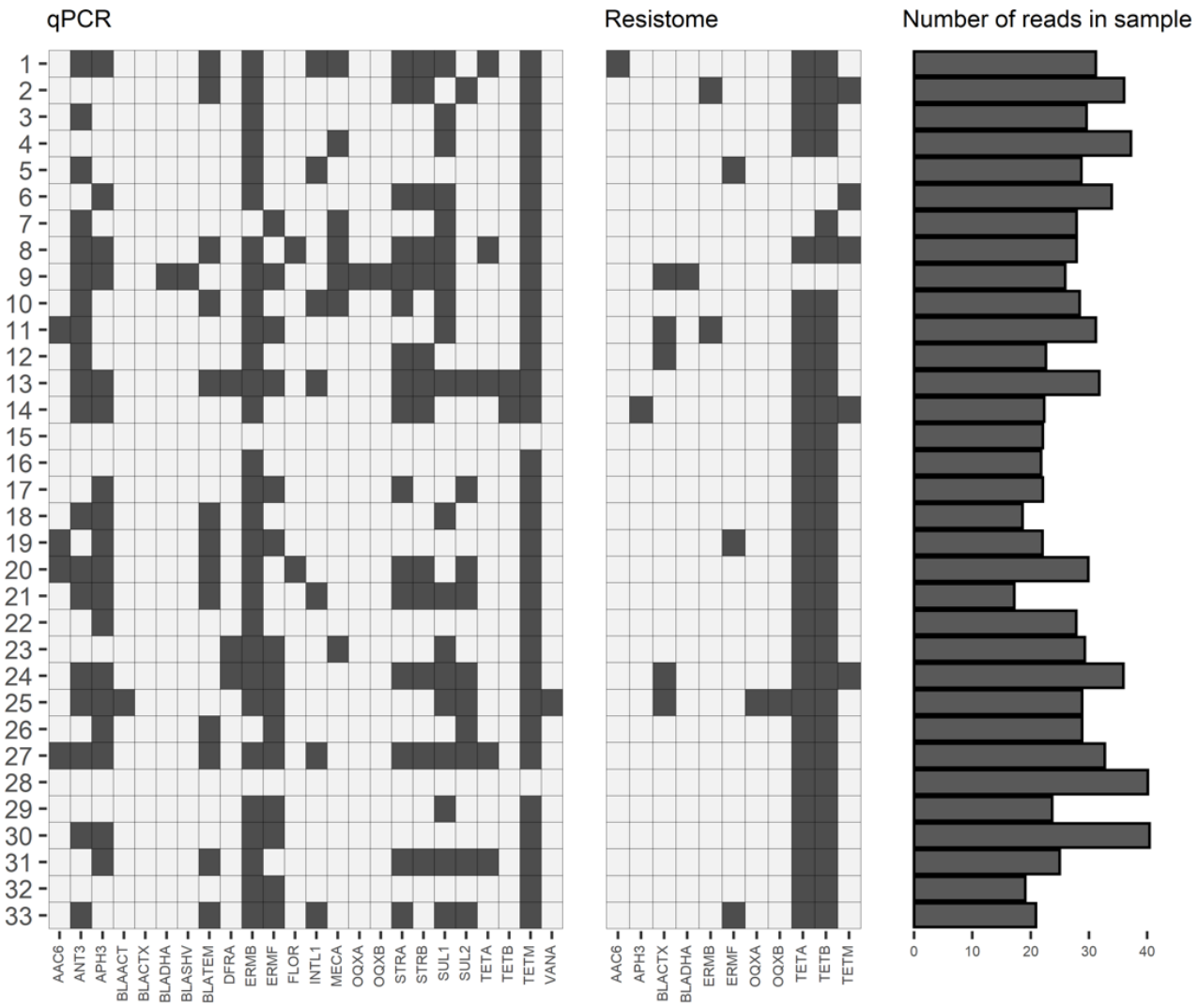

Figure 15: Presence (dark grey)/absence (white) of AMR genes from qPCR and resistome results for all fecal samples. The number of reads for each sample is also included.

The beta diversity measured the variation of AMR genes in terms of absence/presence (Jaccard distance) between detections in the two different methods: qPCR and resistome. The closer two points are to each other in the plot, the more similar they are with regards to which genes were identified with either method. The beta diversity is plotted in Figure 16. The figure provides a visualization of the results from the non-metric multidimensional scaling analysis and shows a clear distinction between the qPCR and resistome results. This highlights the discrepancies between the two methods.

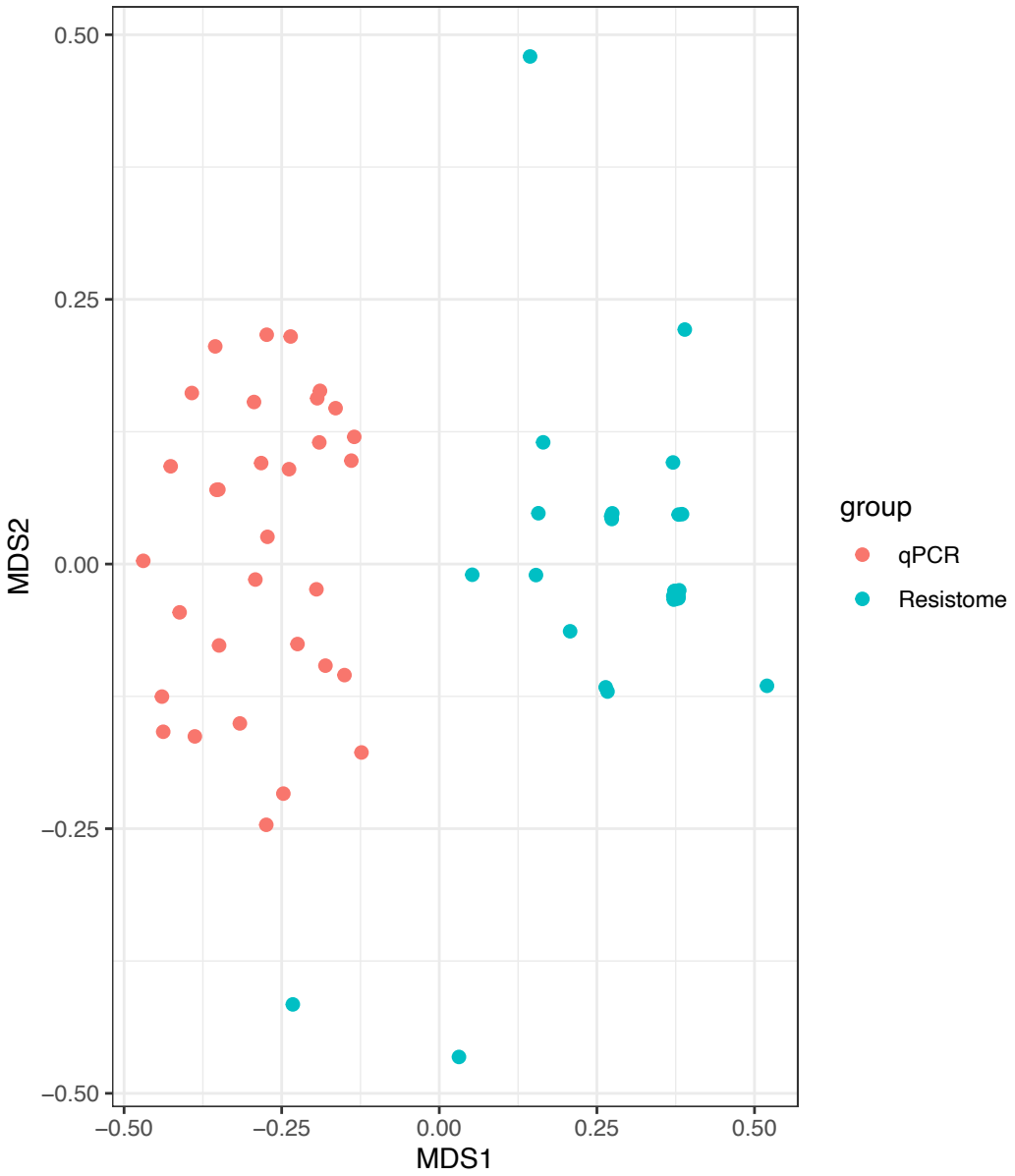

Figure 16: Beta diversity (Jaccard distance) plot presenting the dissimilarities between the two different methods used for AMR gene detection: qPCR and shotgun metagenomic sequencing (resistome). The qPCR method is shown with red and the resistome method is shown with blue.

4 Discussion

The discussion section is divided into two main parts: methodology and findings.

4.1 Methodology

This part of the discussion is focused on the methods performed, in advance, to obtain the data and the bioinformatic methods performed in this thesis. All method descriptions are found in Section 2.

4.1.1 Dataset

The dataset used in this thesis was obtained from 35 dogs. This is a relatively low number of dogs when it comes to concluding something about the resistome in the feces of Norwegian dogs. In other words, assumptions about trends in the Norwegian dog resistome would be limited as this would require a larger, more representative dataset. The findings are also restricted geographically to Trøndelag and one might find varying results if dogs from for instance Oslo or Tromsø were included. However, the dataset is suitable for the comparison of two different methods although more samples would have improved the assurance in our results.

4.1.2 Sampling

The sampling was performed by the dog owners themselves with the use of fecal cards. The procedure is relatively simple, and the owners were instructed on how to do it. However, there are several uncertainties associated with this procedure. Primarily, there might be contaminations in the samples from surroundings, air, skin, and other things that has been in touch with the sample or fecal paper. The fact that different dog owners performed the sampling is an uncertainty in itself whereas they all might have a unique way of performing the sampling, which can influence the results in different directions.

Fecal cards provide a relatively easy approach for sampling and logistics. However, the use of fecal cards might decrease the quality of the sample. The samples were dried in room temperature for hours before being sent in the mail. The environmental temperatures of where fecal samples are stored has an impact on the quality of the fecal DNA as it may degrade or some microbes might grow while being transported. A study showed the degradation rate in fecal samples was significantly higher in temperature of 28°C compared to 15°C. At the

higher temperature, enzyme activity that is related to fecal degradation is higher and affect the samples more as well as the growth and will influence the reproduction of fecal microorganisms (Zhang et al., 2019). However, these samples were dried on paper, which has shown to be stable for prolonged amount of time (Koster et al., 2021).

A challenge with this collection method is the risk of too little material for the DNA extraction. It is therefore a risk that AMR genes, which are present in the microbiota, are not represented in the samples. The fecal material might also be distributed unevenly on the surface of the fecal cards. Punches from different areas on the fecal cards might therefore contribute to a difference in the amount of material. More material does generally lead to more DNA that can be extracted and sequenced, which again increases the likelihood of detecting more AMR genes. Increasing the abundance of specific AMR genes might increase the chances of being detected and is especially important using less sensitive methods. The abundance levels of the AMR genes are not emphasized in this thesis and the main focus is the presence/absence of AMR genes.

The challenges/risks regarding the sampling methods that are mentioned above are the same for both the qPCR and metagenomics results because the same samples were used. Uncertainties regarding the sampling method will therefore only influence the characterization of the resistome and not the method comparison. Sampling errors are difficult to discover as it is expected to have some variations between the fecal samples due to factors such as different environments.

4.1.3 **Sample preparation and extraction**

One major source of uncertainty regarding the comparison between qPCR and metagenomic shotgun sequencing occurs in the sample preparation and extraction steps. The samples were prepared and extracted using two similar, but different methods. In addition, different amounts of starting material were used as one or two punches of fecal paper were used in the qPCR procedure and four punches were used in the shotgun metagenomic sequencing procedure. Therefore, there is more starting material, including more fecal paper, in the metagenome sequencing method. The consequence of this can be a higher actual abundance of AMR genes in this method, which can increase the possibility of detecting the genes for a

less sensitive method. However, as mentioned above, this also depends on the sequence depth of the sample being sufficient.

4.1.4 Sequencing with Illumina

Illumina sequencing was utilized to obtain the sequence data. The number of reads sequenced per sample (sequencing depth) varied between samples and also within the sample groups (Figure 6). The method gives relatively short reads (here 150pb PE reads) with good quality. The fecal material and dry cards provided a low DNA concentration/yield. Illumina sequencing was therefore a suitable approach for the detection of AMR genes from the samples used.

4.1.5 Read based vs *de novo*-based approach

In this thesis a read based profiling of AMR genes was performed. This method was chosen as the objective was to characterize the resistome (find AMR genes) of fecal samples and not taxonomic profiling. Read mapping does not require assembly resulting in less computational power needed as well as less time, which was important in this work. Another important factor was that since the fecal samples are from Norwegian dogs, one does not expect to find high levels of AMR genes (NORM/NORM-VET, 2021). Since the abundance and richness of AMR genes were most likely to be low, it would be difficult to assemble metagenome-assembled genomes (MAGs). Placing the AMR genes in MAGs can be extremely challenging and especially with short read data. Therefore, a read based profiling was a suitable approach in this situation.

On the other hand, if there were dogs from other countries or other samples of where a higher level of AMR genes were expected, it would be interesting to perform a *de novo* approach and compare the findings with the read based approach. Some of the main challenges with a *de novo* approach is that information (often low abundance reads) can be lost when assembling the reads to contigs. It would therefore be desirable to find out if the method would detect less AMR genes due to this or if there would have been other AMR genes detected. Comparing the findings would contribute to a better understanding on how the bioinformatic approach could impact the results for such analyses.

However, taxonomical data can also be obtained form a read based approach, which can be of great importance. To assess AMR threats this is valuable knowledge as pathogens with AMR genes would be important to monitor. The origin of the AMR gene is of valuable knowledge as some bacteria are more dangerous than others. In addition, the mock samples would have a clear task and be used to validate the presence of the species that are known to be there and could also be used to clarify whether the sequencing depth was enough. The mock would also help to identify bias of the method.

4.1.6 Choice of resistome pipeline and database

The resistome pipeline used in this thesis was AMRPlusPlus. AMRPlusPlus is relatively user friendly and is used by many researchers. This can make the results more accessible for comparison. MEGARes is a relatively large, curated database and was the database that was compatible with AMRPlusPlus.

The MEGARes database is a curated meta-database, as it consists of various databases: RESFinder, ARG-ANNOT (Gupta et al., 2014), CARD, and the National Center for Biotechnology Information (NCBI) (Sayers et al., 2020). The positive aspect of combining several curated databases together is that it increases the size of the database, and thus makes it possible to detect several genes. This is a positive aspect in the way that it is less likely for AMR genes to be missed and in for instance surveillance as it is important to not miss potential dangers. On the other hand, there are also challenges with a large database. A challenge with the large database is that different databases might use different naming traditions for the hits and the results might end up being confusing and hard to interpret.

An alternative that could have been interesting to try would be the CARD database with the compatible program RGI. RGI is a program that can be used to predict the resistome in various types of sequences such as genomic and protein. Currently there is also an option in beta-testing that can predict the resistome from metagenomic reads (Alcock et al., 2020). The choices of pipeline and database are important, as it can influence the results. Ideally, in this thesis, different pipelines, and databases should have been used to compare the results. One could then see if the results correspond, and to what degree the choice of pipeline and database influence the findings of AMR genes. However, due to time constraints this was not possible, but highlights the importance of more research and to improve our understanding within this field.

4.2 Findings

This part of the discussion is focused on the findings in the results in Section 3.

4.2.1 Sequencing depth

Rarefaction analysis was conducted to investigate if the sequence depth was sufficient to see the full diversity of ARGs present. Figure 13 and 14 shows that for some samples, the rarefaction curve flattens out, while others continue to rise. The curves which flatten out indicates that most unique hits in the samples were found with the chosen read depth and that the sequencing depth was sufficient to capture the diversity of ARGs present in those samples. There are also curves that keep rising as the amount of reads approach 100%, which indicates that there are still more unique AMR genes to be found in the samples and that the sequencing depth might be insufficient. The sequencing depth was not a choice in this thesis as it was decided by the HUNT One Health project.

The DNA concentrations of the samples sequenced were low (Øivind Øines, HUNT One Health, personal communication), which can indicate that there might in some samples not be more DNA to sequence. This could be the reason to why some fecal sample curves flatten out and does seem to have enough reads to uncover all unique AMR genes. However, this may not be case for all fecal samples, as some curves kept rising indicating that a deeper sequencing is necessary to detect more unique AMR genes. Another explanation as to why some sample curves flatten out might be biological; that there are no more AMR genes to detect as there were no more AMR genes to find in the gut microbiota of that specific dog. Some variation in terms of AMR gene content between the dogs based on environment, lifestyle and earlier antibiotic treatments is to be expected. Two outliers were removed due to a low sequencing depth (Figure 7) as these were clear outliers and far below the cutoff of 17.2 M reads.

There is one negative control sample that stands out with having a higher amount of unique AMR genes compared to the other negative controls (Figure 13). An explanation to this might be contamination, mislabeling, or operator error. The rarefaction analysis does not account for the abundances of the AMR genes and contamination might lead to a higher number of unique genes without the abundance of these necessarily being high. The samples from the same tray should have been omitted from further analysis as this could give false

resistome results. However, the main objective in this thesis was to compare the two methods and since there are not that many samples to compare it was important to compare as many samples as possible with acceptable sequencing depth. It is however important to keep in mind that there can be contaminations. If the sample on the same tray showed results with variation in AMR genes detected, contamination would be a likely reason.

A study conducted on the impact of sequencing depth on the characterization of the microbiome and resistome presented indications that when it comes to abundance of AMR classes, the relative portion of classes remained similar despite sequencing depth (Zaheer et al., 2018). The numbers of reads that was aligned to an AMR gene, on the other hand, increased significantly as the number of reads sequenced per sample increased. It was found that a sequencing depth of 59M was suitable for resistome and microbiome characterization in cattle fecal samples. The average sequencing depth of the fecal samples in this thesis was 27,1M reads (Table 2), which is 46% of the sequencing depth that was established to be suitable for the resistome characterization of cattle fecal samples. However, this does not necessarily mean that 27,1M is too low as there are many factors that plays a role in this, as mentioned above. This also indicates that the relative abundance found in this thesis is well represented whether the sequencing was deep enough or not. Another study on phylogenetic microbiota profiling also showed that the profiling was dependent on the sequencing depth (Rajan et al., 2019). The sequencing depth of the samples had a range of 5-200M and in general deeper sequencing resulted in higher richness and evenness, however, it is also suggested that the classification did not improve above 60 M. In the future, it might therefore be suitable to perform a sequencing depth closer to 59M reads per sample, if this is acceptable for the samples used. Due to the low starting material, this might not have been possible in the samples used here.

4.2.2 **Resistome analysis**

Figure 8 and 9 shows the relative abundance of AMR classes present on sample and treatment (sample type) level. The overall presence of AMR classes per sample group is shown in Figure 10. A higher relative abundance indicates that the antibiotic would be less efficient compared to antibiotics from other classes. It is important to monitor this data and pay close attention to the AMR classes present in the microbiota of dogs (companion animals) as well as humans and other animals. If the presence of certain AMR classes

increases, it could be an indication that the antibiotics will be less and less effective and action plans need to be made.

As one can see in Figure 10 the actual abundances of AMR genes in the blanks are a lot lower than in the fecal and especially the mock samples. Blanks samples had an average of 1697 AMR gene hits per sample as to 60129 and 429824 hits per sample for fecal and mock samples, respectively. There are several explanations as to why the actual abundance of AMR genes are higher in mock samples than of fecal samples. The actual abundance might be influenced by the amount of DNA available to sequence. Mock samples have been sequenced deeper than the fecal samples as the average number of reads per mock sample is 34.7M compared to fecal samples with an average of 27.1M reads per sample (Table 2). This is a factor that might have influenced the actual abundance and it is therefore more important to study the resistome in terms of relative abundance with this sequence data, even though there were also more hits per read in the mock samples.

The detection of AMR genes in the blank samples might be due to presence of genes from microorganisms that that are not from the sample. The population of microorganisms that are present in reagents or plasticware, can be referred to as *kitome*. These may be microorganisms that have contaminated reagents over time, or other organisms forming biofilms on plastics. In the blank samples these includes microorganisms that may be present in the buffers and the fecal paper, the plasticware, as well as from the equipment used in the lab.

The alpha diversity was calculated using three different indices and it can from this be established that the diversity within the different sample types comes from richness of AMR genes (Figure 11). The metagenomic shotgun sequencing method could distinguish the sample types in terms of richness using the Shannon and Chao1 indices. The diversity of the blank samples are nested within the diversity of fecal samples, which can indicate that the background “noise” of *kitome* etc. can also to some small degree present in fecal samples. When it comes to the mock samples, the species present are known, and certain predictions can be made on basis of the species. Amongst the species in the mock community is *P. aeruginosa*, which possesses multiple types of intrinsic antibiotic resistance. The resistance mechanisms *P. aeruginosa* includes efflux pumps, restricted outer membrane permeability and β-lactamase production and thus shows resistance towards many types of antibiotics

(Pang et al., 2019). The group with highest relative abundance in the mock samples are MDR mechanisms (Figure 9), and the intrinsic resistance in *P. aeruginosa* might contribute to this . The production of β-lactamases contribute towards resistance of the β-lactams antibiotic, which is also illustrated in the same figure.

The 35 dogs of which the fecal samples were from were of good health and did not undergo antibiotic treatment at the time of sampling. The occurrence of AMR genes in these samples are therefore expected to be low. It is also not expected to find alarming AMR genes such as *mcr,* a colistin (last resort treatment) AMR gene that is often present in multi resistant bacteria. This AMR gene was not detected in any samples using either method (Li et al., 2020).

The penicillin amoxicillin in combination with clavulanic acid was the most sold antibiotic for companion animals (dogs and cats) in 2020 (NORM/NORM-VET, 2021). Amoxicillins are antibiotics in the β-lactams class and the relative/actual abundance of β-lactams resistance genes can be seen in Figure 8/Figure 9. This is however not the class that makes up the biggest portion of the detected AMR. This is a good indication that despite the high use, the antibiotic class can still be efficient to treat infections in dogs. Just because it is the most used antibiotic prescribed for dogs does not mean that the usage is high. In Norway, a total of 360kg antibiotics was purchased for companion animals which accounts for 7.2% of the total amount of antibiotics purchased as antibacterial veterinary products for terrestrial animals (NORM/NORM-VET, 2021). The use of antibiotics in dogs (and cats) therefore make up a relatively small portion of the antibiotics used for animals.

Figure 12 shows the PCA plot for the different sample types. The datasets were visualized using ResistoXplorer in which the high dimensional data was projected into two variables (PCA1 and PCA2) that explains the principal variation in two dimensions. The PCA plot was generated with AMR data on a gene level. There is a clear distinction between the mock samples compared to the other two sample types. Factors that could have caused this distinction includes the relative and actual abundance (Figure 9 and 10), as distinctions are observable here too. The challenge with the representation is that the origin of variation is not given. The variations or similarities in the samples can for instance be explained by differences in presence/absence of AMR genes, the abundance of AMR genes and the

resistance mechanisms of the AMR genes. Here it is also important to keep in mind that figures obtained from ResistoXplorer does not include information on the number of reads per sample and only the number of AMR gene hits. ResistoXplorer was used to visualize multiple figures from the dataset on AMR resistance genes and abundances, and it was not possible to include this type of metadata. Therefore, the PCA plot does not account for the fact that the blanks had a lower sequencing depth compared to the other sample groups. A PCA plot where this type of sequence data is included would perhaps show different distances between the samples and sample groups.

4.2.3 Comparison of PCR and shotgun metagenomics

Figure 15 shows the AMR gene hits for the two different detection methods: PCR and metagenomic shotgun sequencing and Table 4 gives a numerical overview of what is depicted in the figure. The beta diversity analysis (Figure 16) shows the differences in presence/absence findings from the two different methods suggesting a low agreement.

It is particularly difficult to compare the two methods because of the differences in sample preparation and extraction as differences might already emerge in the lab. Another major challenge with comparing the two methods has been to interpret the AMR gene nomenclature to properly group the genes from qPCR and metagenomic resistome analysis (Hall & Schwarz, 2016).

In this thesis, the PCR method is considered as a «gold standard» due to amongst other factors, the sensitivity and specificity of the method. The metagenomic sequencing method is not as sensitive (Waseem et al., 2019), which is also indicated in Table 4. As a result, genes detected with qPCR were not necessarily detected with metagenomic resistome analysis. Out of the 34 AMR genes that were in the qPCR panel, 23 were detected using qPCR and 11 were detected using metagenome shotgun sequencing. The AMR gene with the highest prevalence (32 samples) detected with the qPCR was *tetM. tetM* was discovered in only five of these samples using metagenomic shotgun sequencing which is equivalent to 16% of the qPCR detection. The metagenomic shotgun sequencing method did detect the presence in some, but not all samples suggesting that it is not as sensitive as the qPCR method. This was to be expected. However, there are also cases of metagenomic shotgun sequencing detecting AMR genes in samples that are not detected in qPCR. *BlaCTX* was detected in five fecal samples

using metagenomic shotgun sequencing and was not detected at all in qPCR. *tetA* and *tetB* were detected in more samples using metagenomic shotgun sequencing. For these specific AMR genes, it can seem like the metagenomic shotgun sequencing method is more sensitive. However, there are also other reasons as to why this is the case such as lack of specificity.

Table 4 shows that the qPCR method detected more AMR genes in more samples than what the metagenomic shotgun sequencing method did. However, despite qPCR being the more sensitive method, there were also instances where AMR genes were detected in the resistome and not in qPCR, as for instance *tetA* and *tetB*. PCR-inhibitors can be a possible explanation to this. PCR-inhibitors are chemical substances that are found in several different biological materials such as fecal materials. The consequences of PCR-inhibitors false negatives and a decrease in sensitivity (Schrader et al., 2012). Since the inhibitors do not impact the metagenomic shotgun sequencing this might be the reason as to why AMR genes were detected only in the resistome.

A challenge in the method comparison was the different names used in qPCR compared to those in the MEGARes database. AMR gene names can be complicated as it might occur that multiple databases use different naming conventions. Another experience was that the AMR databases used slightly different names compared to the naming in qPCR. There are also chances that reads align to an AMR gene that is very similar but with a different name. Some of the AMR genes detected with metagenomic shotgun sequencing were also separated in terms of which species the AMR gene was detected in. Due to the different naming conventions, the same gene may have different names depending on which species it originated from. Extensive knowledge about each gene is necessary to make sure that such grouping is correct.

In the metagenome shotgun dataset, there were AMR genes aligned to ten different *tetM* AMR genes. The different tetM genes were of varying lengths and were found in different bacteria. All *tetM* genes with different origins and lengths were merged into "*tetM*" to be comparable to the one *tetM* in the qPCR dataset. The AMR genes found in different bacteria are different enough to be separated in the database. However, in qPCR there was only one set of primers for the *tetM* AMR gene. This might explain the discrepancies observed in the occurrence of this gene with qPCR and resistome analysis.

Both methods have positive and negative aspects to them. The qPCR method shows an overall higher sensitivity as it in general detected a higher number of the AMR genes in samples. However, with qPCR one must know what to look for and use primers accordingly. This might indicate that this method is an advantage when there are certain AMR genes that are checked for. On the other hand, a positive aspect of metagenomic shotgun sequencing is that the reads in the samples are aligned to a whole database and thereby detecting any AMR gene that is in the database. By constantly updating and curating the database, all known AMR genes can in theory be detected. This is an advantage when it comes to surveillance as unexpected AMR genes might occur. The risk with using this method would be that it is not sensitive enough. Therefore, it would be favorable to investigate the two methods by also studying the relationship between qPCR abundance and shotgun metagenomic detection. Finding a relationship between the two factors would make it possible to determine a threshold AMR gene detection using metagenomic shotgun sequencing in this dataset. When it comes to surveillance, the AMR genes of higher abundances might in some cases be the ones that are most important to detect. The higher abundance AMR genes that are present in more samples can be looked at as more of threat. However, it is also important to find the low abundance AMR genes found in fewer samples that are potential threats that might spread. This depends on the aim of the research/surveillance.

Metagenomics is a relatively new field and is under rapid development. There is room for improvement when it comes to the methods and the bioinformatic tools and therefore great potential for the method to improve in terms of for instance sensitivity. The low sensitivity of the metagenomic shotgun sequencing method is a reoccurring challenge when it comes to the detection of low abundance populations that are present below the threshold of detection (Lanza et al., 2018). Targeted enrichment of resistance genes is a method with a higher sensitivity and accuracy compared to the metagenomic sequencing. To execute this method 38 thousand probes were designed to target the sequences that are found in CARD. This method was able to identify AMR genes present in the human gut microbiota that were not detected using shotgun sequencing (Guitor et al., 2020). Further, it would be interesting comparing metagenomic shotgun sequencing to the targeted enrichment method as it would here be a possibility to use the same database (CARD) and detect for the same sequences using the same naming traditions.

5 Conclusion

In this thesis the resistome of 35 Norwegian canine fecal samples were characterized, with data acquired from metagenomic shotgun sequencing. The metagenomic shotgun sequencing resistome analysis detected a total of 124 unique AMR genes. The results were compared to the results from a qPCR multiplex assay based on 34 AMR genes. The qPCR method detected approximately 2/3 of the AMR genes in the panel, compared to approximately 1/3 for the metagenomic resistome analysis. However, only 8.2% of the findings in the metagenomic resistome analysis corresponded to the qPCR findings and this was 22.8% of the total findings in the resistome for the 34 AMR genes. This shows that there was a low agreement between the two methods. Several factors that could have contributed to this are discussed including differences in extraction, low sequencing depth, and choice of bioinformatic tools. In future studies it would be interesting to investigate the comparison further with the use of different programs and databases for detection, as well as taxonomic classification in addition to the resistome characterization to get a better picture of the microbiota.

References

Aarts, H. J. M., Boumedine, K. S., Nesme, X., & Cloeckaert, A. (2001). Molecular tools for the characterisation of antibiotic-resistant bacteria. *Veterinary Research*, *32*(3-4), 363-380.://WOS:000169544800013

Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A. L. V., Cheng, A. A., Liu, S. H., Min, S. Y., Miroshnichenko, A., Tran, H. K., Werfalli, R. E., Nasir, J. A., Oloni, M., Speicher, D. J., Florescu, A., Singh, B., . . . McArthur, A. G. (2020). CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Research*, *48*(D1), D517-D525. https://doi.org/10.1093/nar/gkz935

Arenas, M., Araujo, N. M., Branco, C., Castelhano, N., Castro-Nallar, E., & Perez-Losada, M. (2018). Mutation and recombination in pathogen evolution: Relevance, methods and controversies. *Infection Genetics and Evolution*, *63*, 295-306. https://doi.org/10.1016/j.meegid.2017.09.029

Arukovic, E., Fetahovic, D., & Pehlivanovic, B. (2019, May 16-18). Impact of Antibiotic Misuse on Genetics Alterations of Bacteria.*IFMBE Proceedings* [Proceedings of the international conference on medical and biological engineering, cmbebih 2019]. International Conference on Medical and Biological Engineering in Bosnia and Herzegovina (CMBEBIH), Banja Luka, BOSNIA & HERCEG.

Aslam, B., Wang, W., Arshad, M. I., Khurshid, M., Muzammil, S., Rasool, M. H., Nisar, M. A., Alvi, R. F., Aslam, M. A., Qamar, M. U., Salamat, M. K. F., & Baloch, Z. (2018). Antibiotic resistance: a rundown of a global crisis. *Infection and Drug Resistance*, *11*, 1645-1658. https://doi.org/10.2147/idr.S173867

Balaban, N. Q., Helaine, S., Lewis, K., Ackermann, M., Aldridge, B., Andersson, D. I., Brynildsen, M. P., Bumann, D., Camilli, A., Collins, J. J., Dehio, C., Fortune, S., Ghigo, J. M., Hardt, W. D., Harms, A., Heinemann, M., Hung, D. T., Jenal, U., Levin, B. R., . . . Zinkernagel, A. (2019). Definitions and guidelines for research on antibiotic persistence. *Nature Reviews Microbiology*, *17*(7), 441-448. https://doi.org/10.1038/s41579-019-0196-3

Balouiri, M., Sadiki, M., & Ibnsouda, S. K. (2016). Methods for in vitro evaluating antimicrobial activity: A review. *Journal of Pharmaceutical Analysis*, *6*(2), 71-79. https://doi.org/10.1016/j.jpha.2015.11.005

Barbosa, T. M., & Levy, S. B. (2000). Differential expression of over 60 chromosomal genes in Escherichia coli by constitutive expression of MarA. *Journal of Bacteriology*, *182*(12), 3467-3474. https://doi.org/10.1128/jb.182.12.3467-3474.2000

Behjati, S., & Tarpey, P. S. (2013). What is next generation sequencing? *Archives of Disease in Childhood-Education and Practice Edition*, *98*(6), 236-238. https://doi.org/10.1136/archdischild-2013-304340

Bengtsson, B., & Greko, C. (2014). Antibiotic resistance-consequences for animal health, welfare, and food production. *Upsala Journal of Medical Sciences*, *119*(2), 96-102. https://doi.org/10.3109/03009734.2014.901445

Benko, R., Matuz, M., Silva, A., Ferreira, J., Machado, M. C., Furtado, C., Galistiani, G. F., Bordas, R., & Blix, H. S. (2019). Cross-national comparison of paediatric antibiotic use in Norway, Portugal and Hungary. *Basic & Clinical Pharmacology & Toxicology*, *124*(6), 722-729. https://doi.org/10.1111/bcpt.13198

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120. https://doi.org/10.1093/bioinformatics/btu170

Bortolaia, V., Kaas, R. S., Ruppe, E., Roberts, M. C., Schwarz, S., Cattoir, V., Philippon, A., Allesoe, R. L., Rebelo, A. R., Florensa, A. F., Fagelhauer, L., Chakraborty, T., Neumann, B., Werner, G., Bender, J. K., Stingl, K., Nguyen, M., Coppens, J., Xavier, B. B., . . . Aarestrup, F. M. (2020). ResFinder 4.0 for predictions of phenotypes from genotypes. *Journal of Antimicrobial Chemotherapy*, *75*(12), 3491-3500. https://doi.org/10.1093/jac/dkaa345

Botelho, J., Grosso, F., & Peixe, L. (2019). Antibiotic resistance in Pseudomonas aeruginosa - Mechanisms, epidemiology and evolution. *Drug Resistance Updates*, *44*, 26-47. https://doi.org/10.1016/j.drup.2019.07.002

Burmeister, A. R. (2015). Horizontal Gene Transfer. 193–194.

Casen, C., Vebo, H. C., Sekelja, M., Hegge, F. T., Karlsson, M. K., Ciemniejewska, E., Dzankovic, S., Froyland, C., Nestestog, R., Engstrand, L., Munkholm, P., Nielsen, O. H., Rogler, G., Simren, M., Ohman, L., Vatn, M. H., & Rudi, K. (2015). Deviations in human gut microbiota: a novel diagnostic test for determining dysbiosis in patients with IBS or IBD. *Alimentary Pharmacology & Therapeutics*, *42*(1), 71-83. https://doi.org/10.1111/apt.13236

Cox, G., & Wright, G. D. (2013). Intrinsic antibiotic resistance: Mechanisms, origins, challenges and solutions. *International Journal of Medical Microbiology*, *303*(6-7), 287-292. https://doi.org/10.1016/j.ijmm.2013.02.009

Crofts, T. S., Gasparrini, A. J., & Dantas, G. (2017). Next-generation approaches to understand and combat the antibiotic resistome. *15*. https://doi.org/10.1038/nrmicro.2017.28

Deng, P., & Swanson, K. S. (2015). Gut microbiota of humans, dogs and cats: current knowledge and future opportunities and challenges. *British Journal of Nutrition*, *113*, S6-S17. https://doi.org/10.1017/s0007114514002943

Dhariwal, A., Junges, R., Chen, T., & Petersen, F. C. (2021). ResistoXplorer: a web-based tool for visual, statistical and exploratory data analysis of resistome data. *Nar Genomics and Bioinformatics*, *3*(1). https://doi.org/10.1093/nargab/lqab018

Di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., & Notredame, C. (2017). Nextflow enables reproducible computational workflows. *Nature Biotechnology*, *35*(4), 316-319. https://doi.org/10.1038/nbt.3820

Doster, E., Lakin, S. M., Dean, C. J., Wolfe, C., Young, J. G., Boucher, C., Belk, K. E., Noyes, N. R., & Morley, P. S. (2019). MEGARes 2.0: a database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. *Nucleic Acids Research*, *48*(D1), D561-D569. https://doi.org/10.1093/nar/gkz1010

European Medicines Agency. (2019). Categorisation of antibiotics in the European Union. Retrieved 03.05.2022, from https://www.ema.europa.eu/en/documents/report/categorisation-antibiotics-european-union-answer-request-european-commission-updating-scientific_en.pdf

Ewels, P., Magnusson, M., Lundin, S., & Kaller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047-3048. https://doi.org/10.1093/bioinformatics/btw354

Guitor, A. K., Raphenya, A. R., Klunk, J., Kuch, M., Alcock, B., Surette, M. G., McArthur, A. G., Poinar, H. N., & Wright, G. D. (2020). Capturing the Resistome: a Targeted Capture Method To Reveal Antibiotic Resistance Determinants in Metagenomes. *Antimicrobial Agents and Chemotherapy*, *64*(1), Article e01324-19. https://doi.org/10.1128/aac.01324-19

Gupta, S. K., Padmanabhan, B. R., Diene, S. M., Lopez-Rojas, R., Kempf, M., Landraud, L., & Rolain, J.-M. (2014). ARG-ANNOT, a New Bioinformatic Tool To Discover

Antibiotic Resistance Genes in Bacterial Genomes. *Antimicrobial Agents and Chemotherapy*, *58*(1), 212-220. https://doi.org/doi:10.1128/AAC.01310-13
Hall, R. M., & Schwarz, S. (2016). Resistance gene naming and numbering: is it a new gene or not? *Journal of Antimicrobial Chemotherapy*, *71*(3), 569-571. https://doi.org/10.1093/jac/dkv351
Harbottle, H., Thakur, S., Zhao, S., & White, D. G. (2006). Genetics of antimicrobial resistance. *Animal Biotechnology*, *17*(2), 111-124. https://doi.org/10.1080/10495390600957092
Hu, T. S., Chitnis, N., Monos, D., & Dinh, A. (2021). Next-generation sequencing technologies: An overview. *Human Immunology*, *82*(11), 801-811. https://doi.org/10.1016/j.humimm.2021.02.012
Illumina. (2010). Illumina Sequencing Technology. In: Illumina.
Illumina. (2017). An introduction to Next-Generation Sequencing Technology. In: Illumina.
Illumina. (2022). *Key differences between next-generation sequencing and Sanger sequencing*. Retrieved 01.05 from https://www.illumina.com/science/technology/next-generation-sequencing/ngs-vs-sanger-sequencing.html
Jacoby, G. A. (2005). Mechanisms of Resistance to Quinolones. *Clinical Infectious Diseases*, *41*(Supplement_2), S120-S126. https://doi.org/10.1086/428052
Jorgensen, J. H., & Ferraro, M. J. (2009). Antimicrobial Susceptibility Testing: A Review of General Principles and Contemporary Practices. *Clinical Infectious Diseases*, *49*(11), 1749-1755. https://doi.org/10.1086/647952
Jost, L. (2007). Partitioning diversity into independent alpha and beta components. *Ecology*, *88*(10), 2427-2439. https://doi.org/10.1890/06-1736.1
Koster, P. C., Bailo, B., Dashti, A., Hernandez-Castro, C., Calero-Bernal, R., Ponce-Gordo, F., Gonzalez-Barrio, D., & Carmena, D. (2021). Long-Term Preservation and Storage of Faecal Samples in Whatman(R) Cards for PCR Detection and Genotyping of Giardia duodenalis and Cryptosporidium hominis. *Animals*, *11*(5), Article 1369. https://doi.org/10.3390/ani11051369
Krause, K. M., Serio, A. W., Kane, T. R., & Connolly, L. E. (2016). Aminoglycosides: An Overview. *Cold Spring Harbor Perspectives in Medicine*, *6*(6), Article a027029. https://doi.org/10.1101/cshperspect.a027029
Lade, H., & Kim, J. S. (2021). Bacterial Targets of Antibiotics in Methicillin-Resistant Staphylococcus aureus. *Antibiotics-Basel*, *10*(4), Article 398. https://doi.org/10.3390/antibiotics10040398
Lambert, P. A. (2005). Bacterial resistance to antibiotics: Modified target sites. *Advanced Drug Delivery Reviews*, *57*(10), 1471-1485. https://doi.org/10.1016/j.addr.2005.04.003
Lanza, V. F., Baquero, F., Martinez, J. L., Ramos-Ruiz, R., Gonzalez-Zorn, B., Andremont, A., Sanchez-Valenzuela, A., Ehrlich, S. D., Kennedy, S., Ruppe, E., van Schaik, W., Willems, R. J., de la Cruz, F., & Coque, T. M. (2018). In-depth resistome analysis by targeted metagenomics. *Microbiome*, *6*, Article 11. https://doi.org/10.1186/s40168-017-0387-y
Leisner, J. J., Jorgensen, N. O. G., & Middelboe, M. (2016). Predation and selection for antibiotic resistance in natural environments. *Evolutionary Applications*, *9*(3), 427-434. https://doi.org/10.1111/eva.12353
Li, B., Yin, F., Zhao, X., Guo, Y., Wang, W., Wang, P., Zhu, H., Yin, Y., & Wang, X. (2020). Colistin Resistance Gene mcr-1 Mediates Cell Permeability and Resistance to Hydrophobic Antibiotics [Original Research]. *Frontiers in Microbiology*, *10*. https://doi.org/10.3389/fmicb.2019.03015

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, *26*(5), 589-595. https://doi.org/10.1093/bioinformatics/btp698
Martinez, J. L. (2014). General principles of antibiotic resistance in bacteria. https://doi.org/10.1016/j.ddtec.2014.02.001
McEwen, S. A., & Collignon, P. J. (2018). Antimicrobial Resistance: a One Health Perspective. *Microbiology Spectrum*, *6*(2), Article Arba-0009-2017. https://doi.org/10.1128/microbiolspec.ARBA-0009-2017
McInnes, R. S., McCallum, G. E., Lamberte, L. E., & van Schaik, W. (2020). Horizontal transfer of antibiotic resistance genes in the human gut microbiome. *Current Opinion in Microbiology*, *53*, 35-43. https://doi.org/10.1016/j.mib.2020.02.002
Melander, R. J., Zurawski, D. V., & Melander, C. (2018). Narrow-spectrum antibacterial agents. *Medchemcomm*, *9*(1), 12-21. https://doi.org/10.1039/c7md00528h
Mercer, D. K., Torres, M. D. T., Duay, S. S., Lovie, E., Simpson, L., von Kockritz-Blickwede, M., de la Fuente-Nunez, C., O'Neil, D. A., & Angeles-Boza, A. M. (2020). Antimicrobial Susceptibility Testing of Antimicrobial Peptides to Better Predict Efficacy. *Frontiers in Cellular and Infection Microbiology*, *10*, Article 326. https://doi.org/10.3389/fcimb.2020.00326
Microbial Ecology Group. (2019). *Pipeline Description*. Retrieved 20.04 from https://megares.meglab.org/amrplusplus/latest/html/pipeline.html?fbclid=IwAR2ggywkFVHGpKrysAE8yxEXP2CnhcsiBZcwJybagM9QFGNcQd0KBnkYT88
Milken Institute School of Public Health. (2017). *How Bacteria Build Resistance at the Cellular Level*. Retrieved 15.03 from https://onlinepublichealth.gwu.edu/resources/antibiotic-resistance-at-cellular-level/
National Human Genome Institute. (2021). *The Cost of Sequencing a Human Genome*. Retrieved 18.03 from https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost
National Human Genome Institute. (2022a). *METAGENOMICS* Retrieved 20.04 from https://www.genome.gov/genetics-glossary/Metagenomics
National Human Genome Institute. (2022b). *Mutation*. Retrieved 02.05 from https://www.genome.gov/genetics-glossary/Mutation
National Human Genome Institute. (2022c). *SHOTGUN SEQUENCING*. Retrieved 29.04 from https://www.genome.gov/genetics-glossary/Shotgun-Sequencing
Nguyen, H. L., An, P. H., Thai, N. Q., Linh, H. Q., & Li, M. S. (2019). Erythromycin, Cethromycin and Solithromycin display similar binding affinities to the E. coli's ribosome: A molecular simulation study. *Journal of Molecular Graphics & Modelling*, *91*, 80-90. https://doi.org/10.1016/j.jmgm.2019.06.002
NORM/NORM-VET. (2021). Usage of Antimicrobial Agents and Occurrence of Antimicrobial Resistance in Norway.
Norwegian Medicines Agency. (2014). TERAPIANBEFALING: BRUK AV ANTIBAKTERIELLE MIDLER TIL HUND OG KATT. https://legemiddelverket.no/Documents/Veterinærmedisin/Terapianbefalinger/Terapianbefaling_Antibakterielle%20midler%20hund%20og%20katt_2014_N.pdf
Norwegian University of Life Sciences. (2021). *Om HUNT*. Retrieved 25.02 from https://www.nmbu.no/prosjekter/hunt/om
Norwegian Veterinary Institute. (2021). *HUNT Én helse*. Retrieved 15.02 from https://www.vetinst.no/forskning-innovasjon/pagaende-forskningsprosjekter/hunt-en-helse
Oxford Nanopore Technologies. (2022). *Oxford Nanopore Technologies*. Retrieved 15.04 from https://nanoporetech.com

PACBIO. (2022). *HOW IT WORKS*. Retrieved 15.04 from https://www.pacb.com/technology/hifi-sequencing/how-it-works/
Pang, Z., Raudonis, R., Glick, B. R., Lin, T. J., & Cheng, Z. Y. (2019). Antibiotic resistance in Pseudomonas aeruginosa: mechanisms and alternative therapeutic strategies. *Biotechnology Advances*, *37*(1), 177-192. https://doi.org/10.1016/j.biotechadv.2018.11.013
Perry, J., Waglechner, N., & Wright, G. (2016). The Prehistory of Antibiotic Resistance. *Cold Spring Harbor Perspectives in Medicine*, *6*(6), Article a025197. https://doi.org/10.1101/cshperspect.a025197
Piddock, L. J. V. (2006). Multidrug-resistance efflux pumps - not just for resistance. *Nature Reviews Microbiology*, *4*(8), 629-636. https://doi.org/10.1038/nrmicro1464
Pilla, R., & Suchodolski, J. S. (2020). The Role of the Canine Gut Microbiome and Metabolome in Health and Gastrointestinal Disease. *Frontiers in Veterinary Science*, *6*, Article 498. https://doi.org/10.3389/fvets.2019.00498
Polikanov, Y. S., Aleksashin, N. A., Beckert, B., & Wilson, D. N. (2018). The Mechanisms of Action of Ribosome-Targeting Peptide Antibiotics. *Frontiers in Molecular Biosciences*, *5*, Article 48. https://doi.org/10.3389/fmolb.2018.00048
Rajan, S. K., Lindqvist, M., Brummer, R. J., Schoultz, I., & Repsilber, D. (2019). Phylogenetic microbiota profiling in fecal samples depends on combination of sequencing depth and choice of NGS analysis method. *Plos One*, *14*(9), Article e0222171. https://doi.org/10.1371/journal.pone.0222171
Rinninella, E., Raoul, P., Cintoni, M., Franceschi, F., Miggiano, G. A. D., Gasbarrini, A., & Mele, M. C. (2019). What is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases. *Microorganisms*, *7*(1), Article 14. https://doi.org/10.3390/microorganisms7010014
Røken, M., Forfang, K., Wasteson, Y., Haaland, A. H., Eiken, H. G., Hagen, S. B., & Bjelland, A. M. (2022). Antimicrobial resistance- Do we share more than companionship with our dogs? In *Journal of applied microbiology*.
Rowe, W. P. M., & Winn, M. D. (2018). Indexed variation graphs for efficient and accurate resistome profiling. *Bioinformatics*, *34*(21), 3601-3608. https://doi.org/10.1093/bioinformatics/bty387
Sayers, E. W., Beck, J., Bolton, E. E., Bourexis, D., Brister, J. R., Canese, K., Comeau, D. C., Funk, K., Kim, S., Klimke, W., Marchler-Bauer, A., Landrum, M., Lathrop, S., Lu, Z., Madden, T. L., O'Leary, N., Phan, L., Rangwala, S. H., Schneider, V. A., . . . Sherry, S. T. (2020). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *49*(D1), D10-D17. https://doi.org/10.1093/nar/gkaa892
Schrader, C., Schielke, A., Ellerbroek, L., & Johne, R. (2012). PCR inhibitors - occurrence, properties and removal. *Journal of applied microbiology*, *113*(5), 1014-1026. https://doi.org/10.1111/j.1365-2672.2012.05384.x
Shebl, H. R., Zaki, W. K., Saleh, A. N., & Salam, S. A. A. (2020). Prevalence of MecC Gene Among Methicillin Resistant Staphylococcus aureus isolated from Patients in Ain-Shams University Hospital. *Journal of Pure and Applied Microbiology*, *14*(4), 2807-2813, Article 6619. https://doi.org/10.22207/jpam.14.4.56
Sikri, N., Dalal, S., & Taneja, R. (2018). EFFLUX PUMPS: AN OVERVIEW. *International Journal of Pharmaceutical Sciences and Research*, *9*(3), 854-861. https://doi.org/10.13040/ijpsr.0975-8232.9(3).854-61
Strommenger, B., Kettlitz, C., Werner, G., & Witte, W. (2003). Multiplex PCR assay for simultaneous detection of nine clinically relevant antibiotic resistance genes in

Staphylococcus aureus. *Journal of Clinical Microbiology*, *41*(9), 4089-4094. https://doi.org/10.1128/jcm.41.9.4089-4094.2003
The Norwegian Vetarinary Institute. (2020). *Overvåkingsprogrammet for antibiotikaresistens* Retrieved 08.05 from https://www.vetinst.no/overvaking/antibiotikaresistens-norm-vet
Thomson, P., Garcia, P., Miles, J., Isla, D., Yanez, C., Santibanez, R., Nunez, A., Flores-Yanez, C., del Rio, C., & Cuadra, F. (2022). Isolation and Identification of Staphylococcus Species Obtained from Healthy Companion Animals and Humans. *Veterinary Sciences*, *9*(2), Article 79. https://doi.org/10.3390/vetsci9020079
Villa, T. G., Feijoo-Siota, L., A., S.-P., Rama, J. R., & Sieiro, C. (2019). Horizontal Gene Transfer in Bacteria, an Overview of the Mechanisms Involved. In *Horizontal Gene Transfer* (pp. 3–76). Springer, Cham. https://doi.org/https://doi.org/10.1007/978-3-030-21862-1_1
Waseem, H., Jameel, S., Ali, J., Rehman, H. S. U., Tauseef, I., Farooq, U., Jamal, A., & Ali, M. I. (2019). Contributions and Challenges of High Throughput qPCR for Determining Antimicrobial Resistance in the Environment: A Critical Review. *Molecules*, *24*(1), Article 163. https://doi.org/10.3390/molecules24010163
Willis, A. D. (2019). Rarefaction, Alpha Diversity, and Statistics [Perspective]. *Frontiers in Microbiology*, *10*. https://doi.org/10.3389/fmicb.2019.02407
World Health Organization. (2017). *One Health*. Retrieved 28.03 from https://www.who.int/news-room/questions-and-answers/item/one-health
World Health Organization. (2021). *Antimicrobial resistance*. Retrieved 28.03 from https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance
Wright, G. D. (2005). Bacterial resistance to antibiotics: Enzymatic degradation and modification. *Advanced Drug Delivery Reviews*, *57*(10), 1451-1470. https://doi.org/10.1016/j.addr.2005.04.002
Wright, G. D. (2010). The antibiotic resistome. *Expert Opinion on Drug Discovery*, *5*(8), 779-788. https://doi.org/10.1517/17460441.2010.497535
Zaheer, R., Noyes, N., Polo, R. O., Cook, S. R., Marinier, E., Van Domselaar, G., Belk, K. E., Morley, P. S., & McAllister, T. A. (2018). Impact of sequencing depth on the characterization of the microbiome and resistome. *Scientific Reports*, *8*, Article 5890. https://doi.org/10.1038/s41598-018-24280-8
Zhang, M., Wei, M., Dong, Z. G., Duan, H. B., Mao, S., Feng, S. L., Li, W. Q., Sun, Z. P., Li, J. W., Yan, K. L., Liu, H., Meng, X. P., & Ge, H. X. (2019). Fecal DNA isolation and degradation in clam Cyclina sinensis: noninvasive DNA isolation for conservation and genetic assessment. *Bmc Biotechnology*, *19*(1), Article 99. https://doi.org/10.1186/s12896-019-0595-6
Zhang, S., & Chen, D. C. (2019). Facing a new challenge: the adverse effects of antibiotics on gut microbiota and host immunity. *Chinese Medical Journal*, *132*(10), 1135-1138. https://doi.org/10.1097/cm9.0000000000000245
Zhao, R. X., Feng, J., Liu, J., Fu, W. J., Li, X. Y., & Li, B. (2019). Deciphering of microbial community and antibiotic resistance genes in activated sludge reactors under high selective pressure of different antibiotics. *Water Research*, *151*, 388-402. https://doi.org/10.1016/j.watres.2018.12.034
ZYMO RESEARCH. ZymoBIOMICSTM Microbial Community Standard II (Log Distribution). In Z. RESEARCH (Ed.).

Appendix

APPENDIX 1: FastQC script on raw reads

```
#!/bin/bash
#SBATCH --account=nn9305k
#SBATCH --time=16:00:0        ## Total time requested

##memory specs
#SBATCH --mem=100G            ## How much memory  the job needs
#SBATCH --partition=bigmem
#SBATCH --cpus-per-task=16    ## Total number of CPUs, max on SAGA is 40
#SBATCH --job-name=fastqc


#loading a module, use version number
module load FastQC/0.11.9-Java-11

# input
files=/cluster/projects/nn9305k/active/okarlsen/hunt_analysis/raw_data/*.g
z

# outdir
outdir=/cluster/projects/nn9305k/active/okarlsen/hunt_analysis/fastqc_raw

# Running fastqc, include threads: --threads <tall>
fastqc --threads 8 -o $outdir $files

echo -e "\n This is the end.\n"
```

APPENDIX 2: FastQC script on trimmed reads

```
#!/bin/bash
#SBATCH --account=nn9305k
#SBATCH --time=6:00:0          ## Total time requested

##memory specs
#SBATCH --mem=100G             ## How much memory the job needs
#SBATCH --partition=bigmem
#SBATCH --cpus-per-task=16     ## Total number of CPUs, max on SAGA is 40
#SBATCH --job-name=fastqc

#loading a module, use version number
module load FastQC/0.11.9-Java-11

# input
files=/cluster/projects/nn9305k/active/okarlsen/hunt_analysis/trimgalore/*
.fq.gz

# outdir
outdir=/cluster/projects/nn9305k/active/okarlsen/hunt_analysis/fastqc_trim
med

# Running fastqc, include threads: --threads <tall>
fastqc --threads 10 -o $outdir $files

echo -e "\n This is the end.\n"
```

APPENDIX 3: MultiQC script on raw reads

```
#!/bin/bash
#SBATCH --account=nn9305k
#SBATCH --time=1:00:0        ## Total time requested

##memory specs
#SBATCH --mem=20G            ## How much memory the job needs
#SBATCH --cpus-per-task=2    ## Total number of cpus, max on SAGA is 40
#SBATCH --job-name=multiqc

# Activate conda
source /cluster/projects/nn9305k/src/miniconda/etc/profile.d/conda.sh

# Activate multiqc
conda activate bifrost

# Input data path
path=/cluster/projects/nn9305k/active/okarlsen/hunt_analysis/fastqc_raw
outdir=/cluster/projects/nn9305k/active/okarlsen/hunt_analysis/multiqc_raw

# Run command
multiqc -o $outdir $path
```

APPENDIX 4: MultiQC script on trimmed reads

```
#!/bin/bash
#SBATCH --account=nn9305k
#SBATCH --time=1:00:0       ## Total time requested

##memory specs
#SBATCH --mem=20G           ## How much memory the job needs
#SBATCH --cpus-per-task=2   ## Total number of CPUs, max on SAGA is 40
#SBATCH --job-name=multiqc

# Activate conda
source /cluster/projects/nn9305k/src/miniconda/etc/profile.d/conda.sh

# Activate multiqc
conda activate bifrost

# Input data path
path=/cluster/projects/nn9305k/active/okarlsen/hunt_analysis/fastqc_trimme
d
outdir=/cluster/projects/nn9305k/active/okarlsen/hunt_analysis/multiqc_tri
mmed

# Run command
multiqc -o $outdir $path
```

APPENDIX 5: Trim Galore script on raw reads

```
#!/bin/bash
#SBATCH --account=nn9305k
#SBATCH --time=5:00:0          ## Total time requested

##memory specs
#SBATCH --mem=40G              ## How much memory the job needs
#SBATCH --partition=bigmem
#SBATCH --cpus-per-task=8      ## Total number of CPUs, max on SAGA is 40
#SBATCH --job-name=trimgalore

# Array (R1 and R2 in 1 array)
#SBATCH --array=0-95

# Input and output
input=/cluster/projects/nn9305k/active/okarlsen/hunt_analysis/raw_data
output=/cluster/projects/nn9305k/active/okarlsen/hunt_analysis/trimgalore

cd $input
files=($(ls *1.fq.gz))
R1=${files[$SLURM_ARRAY_TASK_ID]}
R2=${R1%%1.fq.gz}2.fq.gz

# Activate conda
source /cluster/projects/nn9305k/src/miniconda/etc/profile.d/conda.sh

# Activate Trimgalore
conda activate Trimgalore

# Run program
trim_galore -o $output --paired --quality 15 $R1 $R2

echo -e "\n This is the end.\n"
```

APPENDIX 6: AMRPlusPlus NextFlow script on trimmed reads

```
# Nextflow command ran in screen from amrplusplus_v2 folder
nextflow run main_AmrPlusPlus_v2.nf -profile singularity_slurm --reads "/c
luster/projects/nn9305k/active/okarlsen/hunt_analysis/amr_data/*_{1,2}.fq.
gz" --amr "/cluster/projects/nn9305k/src/amrplusplus_v2/data/amr/megares_d
atabase_v1.02.fasta" --output "/cluster/projects/nn9305k/active/okarlsen/h
unt_analysis/amrplusplus" -work-dir $USERWORK/amrplusplus
```

APPENDIX 7: R script for calculating beta diversity (Jaccard distance)

```
# Calculate Jaccard distances

dist <- vegdist(all_data, method = "jaccard", binary = TRUE)

nmds <- metaMDS(dist, distance = "jaccard", k = 2, try = 500,
autotransform = FALSE)
```

Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway