Universitat Oberta de Catalunya

uoc.edu

In-silico analysis of the resistome of compost

Sandra Wohlmuther
Máster Universitario en Bioinformática y Bioestadística
Área 4. Subárea 6: Microbiología, biotecnología y biología molecular

Consultora: **Paloma Pizarro Tobías**
Profesor responsable de la asignatura: **Antoni Pérez Navarro**

Fecha Entrega
Junio 2021

FICHA DEL TRABAJO FINAL

Título del trabajo:	In-silico analysis of the resistome of compost
Nombre del autor:	Sandra Wohlmuther
Nombre del consultor/a:	Paloma Pizarro Tobías
Nombre del PRA:	Antoni Pérez Navarro
Fecha de entrega (mm/aaaa):	06/2021
Titulación:	Máster Universitario en Bioinformática y Bioestadística
Área del Trabajo Final:	Microbiología, biotecnología y biología molecular
Idioma del trabajo:	Inglés
Número de créditos:	15
Palabras clave	Compost, antibiotic resistance gene, phylogenetic analysis

Resumen del Trabajo:

La diseminación de genes resistentes a los antibióticos (GRAs) constituye un problema global al que, al margen de los tratamientos médicos, contribuye la agricultura. El resultado del compostaje una estrategia de gestión de residuos sólidos, se emplea como fertilizantes de suelos. La literatura muestra que el compostaje puede reducir la presencia de GRAs, pero los resultados son diversos. Este trabajo persigue analizar datos públicos de genomas de completos secuenciados de compost de GRAs y explorar su relación filogenética. Para ello, se emplean datos de 9 experimentos del archivo europeo de nucleótidos. Tras un ensamblaje de-novo, se examinan 4 bases de datos (ARG-annot, CARD, NCBI y Resfinder) para identificar GRAs y se construye un árbol filogenético a partir de sus secuencias de referencia. En total, se identifican 381 GRAs diferentes en 7 conjuntos de datos. Los genes detectados con mayor frecuencia son lnu(C), lnu(D), erm(G), erm(A), mef(A) y tet(X). Sin embargo, estos genes no se encuentran en las muestras de compost de origen conocido, en las que se detectaron los genes aadA1, aadA14, aadA31, aph(3')-Ia, aph(3'')-Ib, aph(6)-Id, blaCARB-8, blaCMY-8, blaTEM-150, blaTEM-171, sul1, sul2, tet(H), tet(W) y vanR-O. Dentro de las 24 clases de medicamentos antimicrobianos, aquellos con mayor presencia en los ARGs son el aminoglucósido, la tetraciclina, el macrólido y los multifármacos (60,8%). En la clasificación taxonómica se identifican como filias dominantes las actinobacterias, bacteroidetes, firmicutes y proteobacterias. El análisis filogenético no resulta concluyente y se encuentran únicamente clústeres parciales de resistencia antimicrobiana.

Abstract:

Dissemination of antibiotic resistant genes (ARG) is a global health concern to which not only medical treatments but also agriculture contributes to. Composting is a waste management strategy for solid organic waste and compost is widely used as soil fertilizer. Studies show that composting can reduce ARGs, but results are inconsistent, influenced by different factors, and increase in abundance also got reported. The objective of this study was to analyze public available whole genome sequencing (WGS) data of compost for ARGs and explore their phylogenetic relationship. WGS data from 9 experiments were obtained from the European nucleotide archive. After de-novo assembly, 4 databases (ARG-annot, CARD, NCBI and Resfinder) were screened to identify ARGs. A phylogenetic tree was built from their reference sequences. In total 381 different ARGs were identified within 7 datasets. The most frequently detected genes were *lnu(C), lnu(D)*, *erm(G), erm(A), mef(A)* and *tet(X).* However within the samples of known origin of finished compost, those genes were not detected; here, *aadA1, aadA14, aadA31, aph(3')-Ia, aph(3'')-Ib, aph(6)-Id*, *blaCARB-8, blaCMY-8, blaTEM-150, blaTEM-171, sul1, sul2, tet(H), tet(W)* and *vanR-O* were found. From the 24 antimicrobial drug classes the main contributors to the resistome were against aminoglycoside, tetracycline, macrolide and multi-drug, accounting for 60.8%. The taxonomic classification identified as dominant phyla overall Actinobacteria, Bacteroidetes, Firmicutes and Proteobacteria. The phylogenetic analysis was not conclusive, only partial cluster for antimicrobial resistance were found.

Table of contents

List of figures

List of tables

1 Summary

Background
Dissemination of antibiotic resistant genes means a global health concern to which not only medical treatments but also agriculture contributes to. Composting has been shown to reduce antibiotic resistance genes (ARGs), but not to eliminate them. Therefore, adjusting composting processes regarding the specific organisms depending on the types of compost could help the management for ARGs.

Methods
Publicly available whole genome sequencing data of compost samples were de-novo assembled and mapped against four different antibiotic resistance gene databases. A taxonomic classification was performed to obtain the microbial community of those samples. Lastly, a phylogenetic analysis was conducted on the reference sequences of the detected ARGs.

Results
ARGs have been found in 7 out of 9 datasets, in total 381 different ARGs were identified. The dominant phyla reported were Actinobacteria, Bacteroidetes, Chloroflexi, Deinococcus Thermus, Firmicutes, Planctomycetes and Proteobacteria. The phylogenetic analysis was not conclusive, only partial cluster for antimicrobial resistance were found. The frequency of the ARGs was not reflected.

Conclusion
ARGs can be found frequently in compost samples. A more meticulous analysis as it was possible within this work is needed, to investigate further the microbial community and the dissemination of ARGs by horizontal gene transfer during the composting process.

Contribution
It was possible to construct a phylogenetic tree from the ARGs detected within the available compost samples.

2 Introduction

2.1 Background and justification

The World Health Organization (WHO) considers antibiotic resistance as one of the greatest threats for global health, food security and development. The so-called antibiotic resistance crisis responds to several factors: overuse and inappropriate prescribing in human Medicine, availability of few new antibiotics, regulatory barriers, and extensive agricultural use [1].
The employment of antibiotics in agriculture and husbandry contributes to the dissemination of antibiotic resistance among microbial population in the environment through several channels. Not only antibiotics are widely used for treating infections in the cattle but also to prevent them in groups of animals. Furthermore, antibiotics are used as a growth promoter for food-producing animals [2]. In the European Union, the latter application is prohibited and there are relevant restrictions on their preventive use, which will be totally forbidden since 2022, that explains the significant reduction in their sales from 2010 to 2018 [3]. The employment of antibiotics in agriculture has also an impact on the environmental microbiome: the urine and stool excreted by livestock, which represents up to 90% of the received intake, is widely dispersed through groundwater and surface runoff and used as fertilizer. Bacteria can acquire antibiotic resistance through mutation, but the dissemination of resistance surges because of antibiotic resistant genes (ARGs) [4]. The growing concern about antibiotic resistance is coupled by the increasing evidence suggesting a non-neglectable possibility for ARGs in agricultural soil can affect the food chain [1, 5, 6, 7].

Composting, decomposition of organic material, has become a waste management strategy for manure and also for solid organic waste in general [8, 9]. As compost contributes to improve the structure of the soil, adding nutrients for plants and beneficial microbes, it is of wide use in gardening and agriculture (particularly, in organic farming, where synthetic fertilizers are not allowed). Depending on the material and the methods applied, there are several types of composting procedures. The most common ones are vessel composting, windrow composting, and aerated static pile composting, all three can be summarized as thermophilic composting as it makes use of thermophilic bacteria, and vermicomposting. Vermicomposting makes use of earthworms that degrade organic matter by feeding on it and it takes between two and three months. Vessel compositing is carried out in an enclosed area (a container, a building or a vessel) and makes use of forced aeration or mechanical turning techniques. Its duration is variable, from 4 to 12 weeks. Windrow composting, which demands a minimum of 15 days, consists in placing raw materials (manure, plants) in long narrow piles, turned regularly to allow aeration. Finally, static composting is a quite traditional method involving passive aeration requiring between 3 and 6 weeks [9, 10, 11, 12].

Studies show that composting can be a suitable method to reduce the concentration of ARGs [13, 14]. Yet, the process of composting itself and the

organic material used in the process can lead to highly variable results, including persistence or increase of the ARG concentration [15, 16].

Metagenomic studies on the dynamics of antibiotics and ARGs abundance during composting, demonstrate a relationship between ARGs and mobile genetic elements (MGE). MGEs mediate horizontal gene transfer, so even if the initial ARG-carrying bacteria diminish during the compost process, the resistance can remain as ARG residue within the compost and may be a risk to public health. Further, one study indicates that the compost resistome displays a varied transcriptional response to composting process, suggesting microbial phylogeny as the key determinant. Monitoring for ARGs in compost seems to be advisable [14, 17, 18, 19, 20].

Nevertheless, the available evidence is far from conclusive and current knowledge of how different types of compost are related to the persistence of ARGs is still incomplete [15, 20]. Therefore, a better understanding of ARGs might improve composting procedures and further be beneficial for adapting measures for the antibiotic resistance dissemination management.

2.2 Objectives

The main aim of this work is to perform an analysis of the resistome of compost in order to establish phylogenetic relationships among antibiotic-resistant genes (ARGs). Particularly, this study tries to explore the phylogenetic relationships between ARGs found in the compost samples.

2.3 Approach and methodology

This research made use of metagenomic sequencing data from the European Nucleotide Archive (ENA) [21]. As they are publicly available data, as long as they are cited in an appropriated way, there are no further ethical or legal concerns. All the analyses were carried out in a personal notebook running under Microsoft Windows 10. Therefore, aiming to avoid downloading and processing the raw data on the notebook - a very time consuming and computationally demanding work unfeasible with the mentioned computer facilities-, the processing of the metagenomic data was performed making use of the web-based platform for bioinformatic analysis Galaxy [22].

For the identification of ARGs a workflow was created applying the tools: fasterq-dump utility of the SRA Toolkit [21], FastQC [23], MEGAHIT [24], ABRicate [25] and staramr [26]. Further a taxonomic classification on the fastq files of the compost samples was done on the Galaxy server using the tools: Kraken2 [27], Convert Kraken [28], and krona [29]. Once ARGs were identified, the phylogenetic analyses was carried out using the free statistical software R and its companion packages [30]. The elaboration of the final document relied on Microsoft Word 2019. The planning of the project was done under GanttProject 3.0 [31]. Further methodological details are provided in Chapter 4.

2.4 Planning of the project

The work unfolds in the following six tasks (in chronological order):

1. Literature Review
2. Database search for metagenomic data of compost samples
3. Identifying ARGs within the metagenomic data
4. Phylogenetic analysis of the ARGs
5. Writing the final document to be submitted
6. Preparing the presentation

Figure 1 displays the timeline created with GanttProject 3.0 that illustrates the different stages of the work from the initial planning to the final presentation Phase 1, Phase 2, Finishing Work, Preparing Presentation, and Defense refer to the course outline. The total time required for completing the work, allocated among the specific tasks of data retrieval, analysis, and writing, is 300 hours. Weekends are marked in light grey and are not included in the estimation of the days needed for fulfilling a task. Furthermore, several tasks were performed in parallel (e.g., identifying ARGs and searching for suitable data or starting the writing process during analysis).

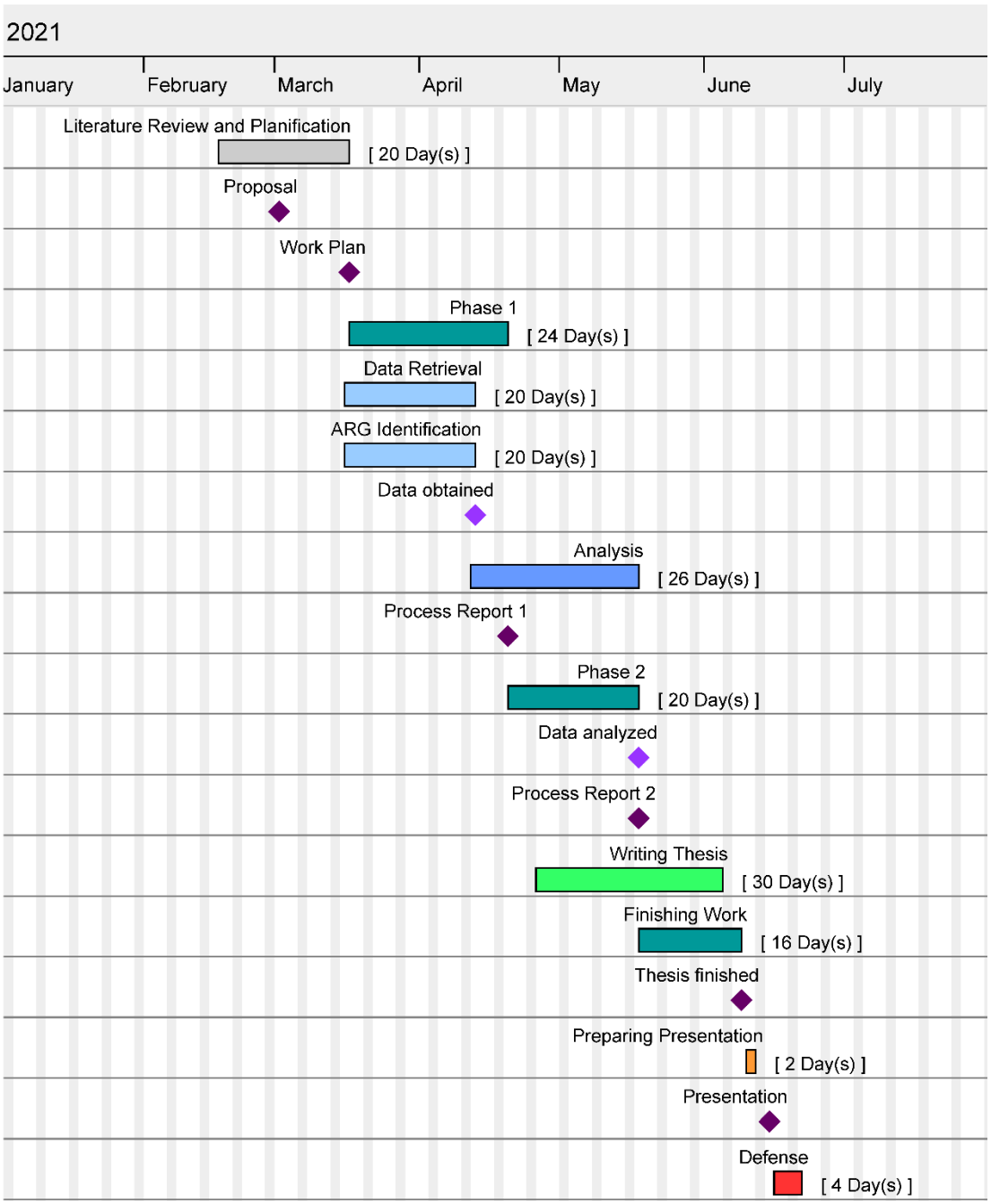

Figure 1. Gantt chart of the project schedule.

As shown in the timeline, milestones of the work in chronological order are the following ones:

1. Proposal: first draft of the planned work, including literature about the topic.
2. Work plan: planification and outline of the steps needed to accomplish the work.

3. Data obtained: obtained the public metagenomic compost data and identified ARGs.
4. Process Report 1: report of the process of the work during phase 1.
5. Data analyzed: completed the phylogenetic analysis of ARGs.
6. Process Report 2: report of the process of the work during phase 2.
7. Thesis finished: completed the writing process of the thesis.
8. Presentation: finished presentation of the work.
9. Defense.

Proposal, Work plan, Process Report 1, Process Report 2, Thesis finished, Presentation, and Defense correspond to the continuous evaluation over the course.

2.5 Brief summary of contributions and outputs

The output generated by this in-silico analysis of compost sequencing data is as follows:

1. FastQC report files
2. MEGAHIT de-novo assembly files
3. Antibiotic resistance genes databases result files
4. Taxonomic classification report files
5. Krona interactive visualization of taxonomic classification
6. R file and RData file for ARG summaries
7. R Script, RData file and all related input/output files of the phylogenetic analysis
8. Supplement.xlsx containing:
    a. List of datasets (ENA projects),
    b. ARG frequency by database
    c. Taxonomic summary files

All files are available in google drive:
https://drive.google.com/drive/folders/1tIjBXZQmE5g1SaKviVEUELLpbWHbUAmt?usp=sharing

Point 1-5 are stored within the folder Galaxy, organized by ENA project number.

2.6 Brief description of the rest of chapters of the project

The rest of the works unfolds in five sections as follows. Chapter 3 summarizes the state of the art and frames the contribution of this work. The fourth chapter describes the main methodological tools used for analyzing the data. The main results of the work are presented in Chapter 5 and a further discussion on their implications and their relationship with the aims of the thesis is performed in the sixth section. The last chapter summarizes the main conclusions of the work, discusses several future research lines, and assess how the planning schedule performs in practical terms.

3 State of the art

There are many studies on how composting affects the abundance of antibiotics and antibiotic resistance genes. Most of them investigated the effect of composting on husbandry manure, lesser on sewage sludge or food waste [32, 15, 19, 7, 33]. Composting experiments that examine the influence of different types of co-substrates are commonly done with rice, wheat or corn straw, composting end-products, sawdust, spent mushroom or cotton stalk, while the most studied antibiotic resistance genes belong to resistance to tetracyclines, sulfonamides, macrolides, fluoroquinolones and beta-lactam. The composting methods are mainly lab-scale experiments of thermophilic composting methods, fewer studies make use of vermicomposting. Both seem to be favorable for reducing the resistome, still many factors influence the outcome, leading to contradictory results [34, 15].

In terms of methodology, studies on ARG abundance are frequently conducted by qPCR, using a set of primer for the genes of interest. For the taxonomic classification of those samples a 16S rRNA amplicon sequencing approach is chosen. When working with 16S rRNA for classification, it is recommended to analyze and interpret those results carefully. The 16S rRNA genes can be affected by horizontal gene transfer (HGT). The existence of bacteria containing two types of rRNA operons or mosaics of sequences from multiple species have been demonstrated [35].

Screening for ARGs with a qPCR allows only for detection of those selected genes. Although of interest is the clinically relevant antimicrobial resistance, this may be a too limiting approach. Microbiomes are complex, non pathogenic bacteria can be carrier of ARGs or genes that potentially might function as resistance genes. A pan-microbial approach, using whole genome sequencing (WGS) for resistome analysis, may be a better choice for the management of ARGs [36, 37]. WGS allows for detection of all known, but also for discovering new antibiotic resistance genes.

The concern of this work is to investigate publicly available WGS data of compost samples for identifying ARGs and examining their phylogenetic relationship.

4 Methodology

4.1 Data Retrieval

The majority of metagenomic studies of compost apply 16S rRNA amplicon sequencing and test for the resistome by quantitative PCR of selected ARGs. Therefore, available data is limited. As source for the datasets the European Nucleotide Archive (ENA) was chosen [21]. The use of the database for metagenomic data MG-RAST resulted not possible as datasets could not be accessed due to accession restrictions or time-out errors [38]. Another reason for using ENA was, that it has a direct interconnectivity with the Galaxy platform.
For selecting the datasets, they had to be specified as WGS in Library Strategy. One exception is PRJNA526758, which is defined as AMPLICON on the ENA website, but in the corresponding publication the ENA Project number refers to the samples of the shotgun sequencing only [13].
This leads to the conclusion, that it is possible that not all projects are defined correctly on the website. It was also observed, that a compost sample is defined as soil metagenome instead of compost metagenome. Additionally, a reference to a publication or further description is missing in many cases. All this made the process of data selection more difficult. In consequence, it is possible that not all suitable datasets were found and selected. The table below shows the selected datasets.

Table 1. List of metagenomic datasets.
The table lists the selected ENA project numbers, the PubMed identifier if a publication is available and the given study title on the ENA server.

ENA Project	PMID	ENA Study Title
PRJNA433771	31563779	Metagenomics of chicken manure composts
PRJNA526758	31884359	Cow manure metagenomics
PRJNA311675	27834174	Metagenomic analysis of a vermicomposting system in Uganda
PRJNA549056[a]	31751342	Shotgun metagenomic sequencing of food waste and compost samples from a Vermont poultry farm
PRJNA41493	-	Composting bioreactor sample metagenome
PRJNA684647	-	Effects of compound microbial inoculants on nitrogen conversion during livestock manure composting process
PRJNA337811	-	Thermal compost microbial communities from rain forest in Puerto Rico metagenome
PRJNA288410	-	Compost bacteria Metagenome
PRJNA329458	-	Compost metagenome

[a] Only compost samples are included.

4.2 Identification of ARGs

The analysis of the metagenomic data was conducted on the European (usegalaxy.eu) and US (usegalaxy.org) server. This allowed maintaining the timeline as the processing times of the different tools on the web-based platform can take varying time from hours to days. The basic workflow with the different applied tools is shown in figure 2.

For characterization of ARGs the sequenced fragments can be directly mapped to databases (read-based profiling). Another approach, the one applied in this work, is *de novo* assembly. With this method longer contigs are created from the sequencing reads. De novo assembly can lead to loss of data and is computationally more demanding, but it is considered to be more accurate in detection of protein-coding genes [39].

In continuation, two different tools were used to map the contigs to four different databases. This allows for a wide search spectrum, identifying as many genes as possible.

The tools are described briefly; if not stated otherwise, the analysis was performed under the standard settings.

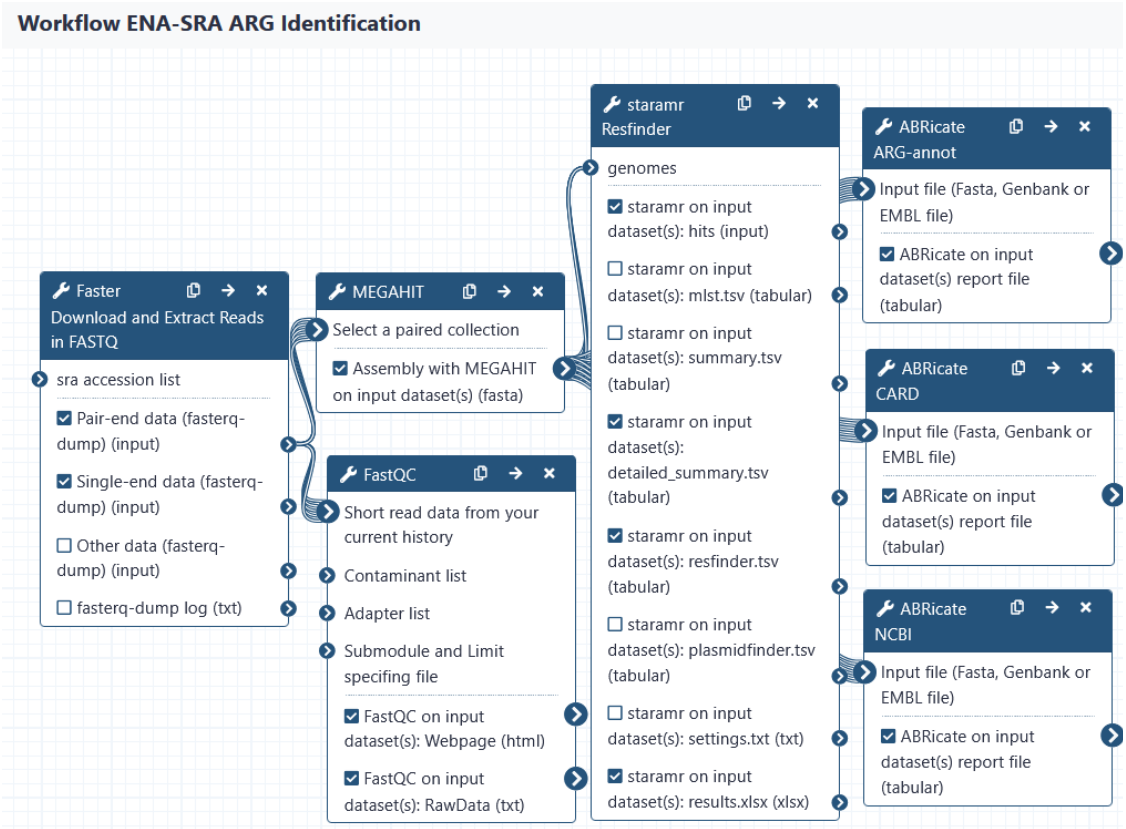

Figure 2. Bioinformatic workflow of ARG identification on Galaxy platform.
The figure shows the tools used for retrieving the fastq files from ENA, contig assembly and antibiotic resistance gene identification.

Faster Download and Extract Reads in FASTQ
Providing the Sample Accession Numbers, the tool extracts the data in fastq format from the Short Read Archive (SRA) at the National Center for Biotechnology Information (NCBI). It is based on the fasterq-dump utility of the SRA Toolkit (sra-tools, Version 2.10.9) [40].

FastQC
The quality of the sequence reads were viewed using FastQC (Version 0.11.8), a well-established tool for evaluation the quality of sequencing data [23].
All created web reports, as well as the raw data, were downloaded and stored.

MEGAHIT Assembly
The fastq files from each ENA project were assembled using MEGAHIT (Version 1.1.3). It is a de-novo assembler for large and complex metagenomics NGS reads with reduced memory usage [24]. This was an important factor, as online tools have limited memory and storage usage. The minimum length of contigs was set to 300. The assembly results were also downloaded and stored.

staramr
staramr (Version 0.7.2) was used for mapping the contigs against the Resfinder database [26, 41]. The *Percent identity threshold* for BLAST was set to 90%.
The files *resfinder.tsv*, detailed *summary.tsv* and *results.xlsx* were downloaded, stored.

ABRicate
ABRicate (Version 1.0.1) was used to screen for ARGs against the ARG-ANNOT, CARD and NCBI Bacterial Antimicrobial Resistance Gene databases [25] [42] [43] [44].
The report files were downloaded and stored.

4.3 Taxonomic classification

Many tools exist for taxonomic classification. Limited to the Galaxy platform and computational restrains, Kraken2 was selected.
Below the workflow on the Galaxy platform is shown for the taxonomic classification of the metagenomic samples.

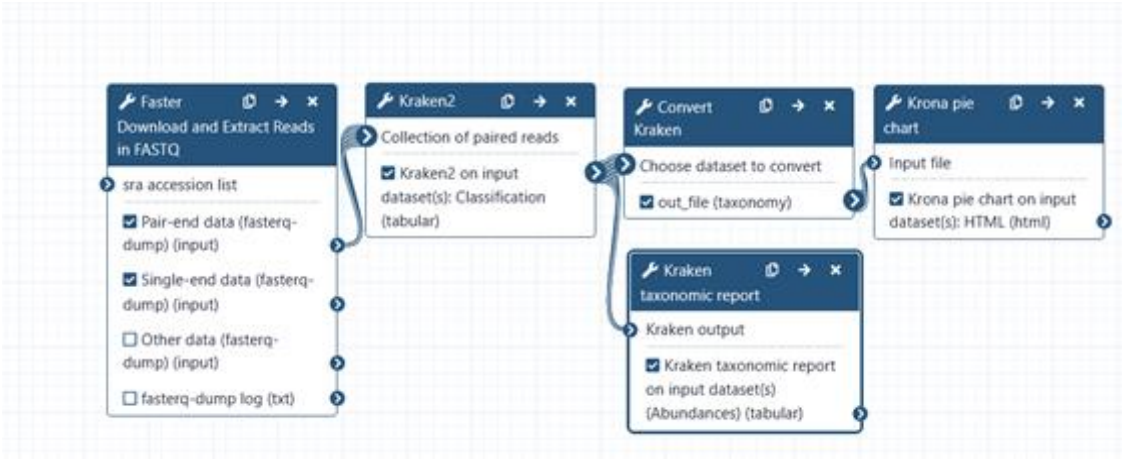

Figure 3. Workflow of taxonomic classification of WGS samples on Galaxy platform. The figure shows the tools used for the taxonomic classification on fastq files.

Kraken2
For the taxonomic classification Kraken2 (Version 2.1.1) was used with the Kraken2 database *Standard created 2020-06-24* [27]. This tool provides a fast classification on the raw reads using a *k-mer* based approach. The generated classification files were downloaded and stored.

Convert Kraken
The classification files were converted using the Convert Kraken tool to Galaxy taxonomy representation (Galaxy Version 1.2), a full representation of NCBI taxonomy. This file was needed for the following visualization with Krona.

Krona pie chart
From the converted taxonomy file a Krona pie chart was created with krona version 2.7.1 on the Europe server and version 2.6.1 on the US server [28]. The newest version did not work on the US server, and with the 2.6 version on the Europe server, problems with the interactive features were found. The resulting interactive html files were downloaded and stored.

Kraken taxonomic report
From the kraken2 classification files of each ENA project a taxonomic report was created, selecting the kraken database archea_2020 on Europe sever and bacterial on the US server (Galaxy Version 0.0.2) [29]. All reports were downloaded and stored.

The files created by the tools on the Galaxy platform can be found within the folder *Galaxy* in google drive.

4.4 Phylogenetic Analysis of ARGs

The phylogenetic analysis was done with the reference sequences of the ARGs. Computational limitation did not allow for analyzing the metagenomic data itself.

The ARG databases use to some degree different gene names and no common identifier. Because of this it was necessary to compare the results for alternative gene names. The obtained unique genes used for further analysis.

For the genes reported by ABRicate screening the NCBI database, the corresponding accession numbers were used for their sequence retrieval [44]. For the genes found only on the other databases, NCBI was used as reference if possible. In case there was no sequence for the gene listed as reference, the sequences were obtained from the CARD database [43]. However, there were some genes were neither on NCBI nor on CARD existed a reference: tetR(G) and vanA-G, that were reported by ARG-ANNOT [42]. But there were also genes reported by CARD, but then could not be found on the website. This was the case with SAT-1, the gene parY was listed as private model on the website and could not be accessed.

The analysis was done under *R* software version 4.0.5 [30]. There exist many packages for phylogenetic analysis. In this work the package *ape* was used [45]. It is one of the core packages for phylogenies in R.

Workflow
First, the NCBI reference sequences were obtained using the *read.GenBank()* function. The remaining sequences from the CARD database were obtained manually and merged with the NCBI fasta files.

For the multiple sequence alignment the *msa* package was used, applying the MUSCLE algorithm [46, 47]. An advantage of this package for multiple sequence alignment is, that it does not need any other external software.

The construction of the tree was done by neighbor joining tree estimation using the *nj()* funtion [48]. A distance based algorithm that is widely used for phylogeny reconstruction. The distance was calculated with the *dist.gene()* function.

Finally, the tree was visualized with the *phydataplot()* function. This function allows to add additional information to the tree. In this case, a barplot was added to show also the frequency in which the genes were detected. The tree was drawn as phylogram for best readability, for the branches the edge length was not used (*use.edge.length=FALSE*).

Bootstrapping was not performed on the phylogeny due to computational limitations.

In google drive the following files are available:
arg-seqs.txt, containing all reference fasta sequences.
card_seqs.txt, containing the reference sequences from CARD.
msaMuscle.txt, multiple sequence alignment file.
tiplabelannot.csv, annotation file for adding additional information of the antimicrobial resistance and the frequency of detection of the genes to the visualization of the tree.
Further all intermediary files, the R script and the .RData file (*phylo.R, pylo.RData*).

5 Results

5.1 Whole genome data of compost

In this work, 9 whole genome sequencing datasets - containing in total 95 samples - were analyzed. The list of the ENA project numbers, their given study title on ENA and their PubMed identification number, if a publication is available, is given in table 1.

For the datasets without a reference to a publication, it can not be determined from which stage of composting the samples were taken. Neither which organic material or method was used.

The origins of *PRJNA288410* and *PRJNA329458* are unknown. As shown in table 1 they are described as compost metagenome but no further details are given.

The data of *PRJNA41493* and *PRJNA337811* consist of one single sample each. The assigned title for PRJNA41493 on the ENA browser is “Composting bioreactor sample metagenome”. The description states that green-waste compost and switchgrass were “incubated under simulated composting conditions in a bioreactor”.
The assigned title for PRJNA337811 is “Thermal compost microbial communities from rain forest in Puerto Rico metagenome”, no further details are given.

PRJNA684647 consists of samples named from S1-S13, with assigned study title “Effects of compound microbial inoculants on nitrogen conversion during livestock manure composting process”. A publication was not found, so no further information is available.

PRJNA433771 has a publication, but the sample names (XC1-XC12) do not allow assigning them to the description given by the authors [49]. The compost samples result from thermophilic composting of chicken manure with rice chaff, taken on day 0, 4, 21 and 28 in triplicate.

Out of the 9 analyzed datasets, 3 have a full description in their corresponding publication: PJRNA311675, PRJNA54056 and PRJNA526758.

PRJNA311675 corresponds to vermiculture of dairy manure and food waste [50]. The sampling was done at the endpoint on day 172. However, it is unclear if a difference exist between the samples named CB1-3 and CV1-3.

PRJNA54056 refers to a study on food waste composting on a poultry farm [32]. Sampling occurred at the beginning of the raw compost, of the unfinished, and of the finished in duplicates. On which days the samples were taken is unknown. The authors compared two composting methods, windrow and vermiculture. The vermiculture composting is done on this farm for commercial purposes. The samples “Worm Castings” refer to this method. The windrow composting was experimental, the samples are named “Compost”.

PRJNA526758 refers to a study on aerobic dairy manure composting, the experiment was performed in summer and winter for investigating seasonal changes [51].

A list of the datasets is given in *Supplemental.xlsx*.

5.2 Identification of ARGs

It was possible to verify ARGs in 7 datasets.
The number of unique ARGs in all datasets found by databases are:

ResFinder: 209
ARG-ANNOT: 223
CARD: 286
NCBI: 236

In total, 381 different antibiotic resistance genes were detected.
The list of those genes with their frequency of detection by database can be found in *Supplemental.xlsx*.

ARGs by dataset
The data PRJNA288410 and PRJNA329458 were negative for ARGs.
The total number of ARGs for the other datasets are shown in table 6.
In continuation a description of the results by datasets are given.

PRJNA311675 consist only of endpoint testing of vermiculture in which 4 ARGs were found (also see table 2).

PRJNA41493 (bioreactor) and PRJNA337811 (rain forest) may also be endpoint testing, as they consist of one single sample.
In the green-waste compost sample from a bioreactor, *bla-TEM16* was detected and in the rain forest compost *ant(9)-Ia*, *rphB* and *rphC*.

The number of ARGs found within the samples that could be clearly assigned to the composting stage are shown in table 2 below.

Table 2. Comparison of varying ARG abundance in composting process and methods.
The table shows the number of unique antibiotic resistance genes of the 3 datasets for which the samples could be related to the sampling points.

ENA Project	Sample	Composting	Material	Stage	Total ARGs
PRJNA549056	Finished Compost	Windrow	Food Waste	Mature	**10**
	Unfinished Compost	Windrow	Food Waste	Unfinished	**18**
	Raw Compost	Windrow	Food Waste	Raw	**5**
	Worm Castings	Vermiculture	Food Waste	Mature	**0**
	Shifted Worm Castings	Vermiculture	Food Waste	Unfinished	**0**
	Top Worm Castings	Vermiculture	Food Waste	Raw	**6**
PRJNA526758	Sum14	Aerated pile	Dairy Manure	Mature	**5**
	Sum3	Aerated pile	Dairy Manure	Unfinished	**15**
	Sum0	Aerated pile	Dairy Manure	Raw	**58**
	Win14	Aerated pile	Dairy Manure	Mature	**2**
	Win6	Aerated pile	Dairy Manure	Unfinished	**9**
	Win3	Aerated pile	Dairy Manure	Unfinished	**20**
	Win0	Aerated pile	Dairy Manure	Raw	**84**
PRJNA311675	CB/CV	Vermiculture	Dairy Manure + Food Waste	Mature	**4**

The commercial vermiculture was positive for ARGs only at the initial layer of the worm casting. At the second sampling of the unfinished vermicompost after shifting, no ARGs were detected. The windrow composting of the same food waste showed the lowest detectable number of ARGs at the beginning, the highest at the second sampling of the unfinished compost. The finished compost had reduced from 18 to 10 ARGs, a higher amount than at the initial stage.

The composting by aerated pile of dairy manure displays a continuous decrease in ARGs from the raw to the mature compost. The experiment conducted during the winter shows a slightly higher number of ARGs.

In continuation, the detected ARGs from those 3 datasets are shown in table 3.

Table 3. List of ARGs found in PRJNA549056, PRJNA526758 and PRJNA311675.
The table shows the types of antibiotic resistance genes detected within the samples of the datasets PRJNA549056 (compost/worm castings), PRJNA526758 (Sum/Win) and PRJNA311675 (CB/CV). The results for the 6 CB/CV samples were taken together as they refer to one endpoint sampling. The samples of compost and worm castings are duplicates analyzed together.

Sample	AR-Genes
Finished Compost	aadA1, aadA14, aadA31, aph(3')-Ia, aph(3'')-Ib, aph(6)-Id, blaCARB-8, sul1, sul2, tet(H)
Unfinished Compost	aadA1, aadA14, aadA27, aadA31, aph(3'')-Ib, aph(6)-Id, mef(A), mexF, msr(D), lnu(B), lnu(C), lsa(E), sul2, tet(L), tet(H), tet(W), tet(37), tet(39)
Raw Compost	aadA1, ant(3'')-Ia, aph(3'')-Ib, aph(6)-Id, sul2
Worm Castings	-
Shifted Worm Castings	-
Top Worm Castings	aadA1, ant(3'')-Ia, aph(3'')-Ib, aph(6)-Id, sul2, tet(X)
Sum14	aph(3'')-Ib, aph(3')-Ia, blaTEM-171, tet(W), vanR-O
Sum3	aadA1, aadA6, ant(6)-Ia, aph(3'')-Ib, aph(6)-Id, aph(3')-Ia, blaTEM-17, mef(A), tet(Q), lnu(D), sul1, sul2, spw, vanR-O, qacH
Sum0	aadA1, aadA6, aadA9, aadK, aadS, aac(3)-IId, aac(6')-Ib', aac(6')-Ib-Hangzhou, aac(6')-Ib-cr, ant(2'')-Ia, ant(3'')-Ia, ant(6)-Ia, aph(3'')-Ib, aph(6)-Id, aph(3')-Ia, aph(3')-IIIa, bla-TEM171, blaOXA-347, cfxA, CpxR, dfrA1, emrE, ermF, floR, lin(A), lsa(E), lnu(A), lnu(B), lnu(C), lnu(D), lnu(G), mef(A), mef(C), mexB, mexK, mexF, mexW, msr(D), mph(E), mph(G), OprN, OpmH, spw, sul1, sul2, tet(A), tet(G), tet(L), tet(M), tet(Q), tet(R), tetR(G), tet(W), tet(X), tet(X4), tet(39), qacH
Win14	blaTEM-150, vanR-O
Win6	aadA1, aadA10, aph(6)-Id, aph(3'')-Ib, blaTEM-150, CpxR, mexK, rpoB2, vanR-O
Win3	aadA1, aadA6, aadA13, aph(6)-Id, aph(3'')-Ib, ant(6)-Ia, blaTEM-150, CpxR, dfrA1, emrE, lnu(C), lnu(D), mexK, mexW, spw, sul1, sul2, tet(W), vanR-O, qacH
Win0	aadA1, aadA5, aadA6, aadA15, aadA27, aadE, aadS, aac(3')-IId, aac(6')-IIa, aac(6')-Ib3, aac(6')-Ib4, aac(6')-Ib9, aac(6')-Ib-cr, aph(3')-Ib, aph(3')-Ia, aph(3')-VI, aph(3'')-Ib, aph(3'')-III, aph(3')-VIa, aph(6)-Id, ant(6)-Ia, blaCARB-5, blaCARB-16, blaOXA-278, blaOXA-285, blaOXA-335, blaOXA-347, blaOXA-646, blaTEM-150, ble-MBL, catB3, catB4, catB8, catB11, catQ, cfxA, cfxA2, dfrA1, dfrA17, ere(D), erm(F), erm(G), floR, lin(A), lnu(A), lnu(B), lnu(C), lnu(D), lnu(G), lnu(F), lsa(E), Mef(En2), mef(A), mef(B), mexW, mph(E), msr(E), msr(D), tet(39), qnrD1, sul1, sul2, sul3, sat2, sat4, tet(A), tetA(P), tet(H), tet(L), tet(M), tet(O), tet(Q), tet(R), tetR(G), tet(W), tet(X), tet(X4), tet(Y), tet(32), tet(36), tet(44), vanR-I, vanR-O, vanZ-F
CB/CV	aph(3'')-Ib, blaCMY-8, sul1, sul2

PRJNA433771, the thermophilic composting of chicken manure, with sampling on 4 stages of the process, shows 6 samples with a number of ARGs above 100. The other part of the samples show a decrease in abundance to the half or less (see table 4). As it is unknown to which stage the samples belong to, it remains unclear if the abundance decreased or increased during the composting process.

Table 4. Variation of ARG abundance in chicken manure compost.
The table shows the variation of detected ARGs of experimental thermophylic chicken manure composting. The samples can not be related to the sampling points. It is known that the sampling was done in triplicate on day 0, 4, 21 and 28.

Project	Sample	Composting	Material	NCBI	Resfinder	CARD	ARG-annot
PRJNA433771	XC12	Thermophilic	Chicken Manure + Rice Chaff	54	38	46	43
	XC11	Thermophilic	Chicken Manure + Rice Chaff	57	51	50	49
	XC10	Thermophilic	Chicken Manure + Rice Chaff	51	38	46	43
	XC9	Thermophilic	Chicken Manure + Rice Chaff	31	28	24	28
	XC8	Thermophilic	Chicken Manure + Rice Chaff	43	41	39	38
	XC7	Thermophilic	Chicken Manure + Rice Chaff	46	37	40	40
	XC6	Thermophilic	Chicken Manure + Rice Chaff	118	120	104	111
	XC5	Thermophilic	Chicken Manure + Rice Chaff	132	125	122	122
	XC4	Thermophilic	Chicken Manure + Rice Chaff	120	112	109	111
	XC3	Thermophilic	Chicken Manure + Rice Chaff	135	127	156	129
	XC2	Thermophilic	Chicken Manure + Rice Chaff	148	131	163	140
	XC1	Thermophilic	Chicken Manure + Rice Chaff	138	133	149	132

The experiment with livestock manure composting (PRJNA684647) has similar number of ARGs in all samples (see table 5).

Table 5. Variation of ARG abundance in experimental livestock manure composting.
The table shows the number of ARGs found on the different databases of all samples of PRJNA684647. It is unknown to what composting process, experiment or testpoints refer to.

Project	Sample	Composting	Material	NCBI	Resfinder	CARD	ARG-annot
PRJNA684647	S13	na	livestock manure	62	58	54	56
	S12	na	livestock manure	50	42	10	49
	S11	na	livestock manure	71	78	60	65
	S2	na	livestock manure	54	65	76	53
	S1	na	livestock manure	54	62	50	52
	S10	na	livestock manure	66	56	57	63
	S9	na	livestock manure	61	50	53	59
	S8	na	livestock manure	73	71	63	69
	S7	na	livestock manure	63	53	55	61
	S6	na	livestock manure	68	71	75	66
	S5	na	livestock manure	86	90	130	88
	S4	na	livestock manure	70	78	100	68
	S3	na	livestock manure	62	63	75	61
	S18	na	livestock manure	92	83	137	88
	S17	na	livestock manure	102	86	101	89
	S16	na	livestock manure	64	61	59	56
	S15	na	livestock manure	112	94	104	99
	S14	na	livestock manure	65	61	56	59

Most frequent ARGs and AMR

The most frequently identified gene is lnu(C), reported 65 times by Resfinder. Above 50 times reported were erm(G) and mef(A), both macrolide resistance genes. Table 6 below shows the genes that were detected more than 30 times by at least on database within all 95 samples.

Table 6. Most frequently detected ARGs.
The table shows those antibiotic resistance genes, that were detected more than 30 times within all 95 samples. The number of detection is given by database. Further the corresponding antimicrobial substance class (AMR) is given.

Gene	Resfinder	NCBI	CARD	ARGannot	AMR
lnu(C)	65	58	58	58	LINCOSAMIDE
erm(G)	25	56	55	56	MACROLIDE
mef(A)	51	35	35	35	MACROLIDE
erm(A)	36	44	40	40	MACROLIDE
lnu(D)	0	43	43	43	LINCOSAMIDE
tet(X)	41	22	11	22	TETRACYCLINE
cfr	39	7	7	7	MULTIDRUG
ant(9)-Ia	20	35	3	35	AMINOGLYCOSIDE
cat	35	0	0	0	CHLORAMPHENICOL
msr(D	18	24	35	24	MULTIDRUG
spw	0	35	0	35	AMINOGLYCOSIDE
tet(L)	35	26	26	26	TETRACYCLINE
aph(3'')-Ib	34	31	31	31	AMINOGLYCOSIDE
floR	25	26	34	34	PHENICOL
aph(6)-Id	33	30	30	30	AMINOGLYCOSIDE
sul2	33	30	30	30	SULFONAMIDE
aadA1	15	30	32	29	AMINOGLYCOSIDE
erm(T)	32	32	32	32	MULTIDRUG
aph(2'')-Ih	9	31	0	9	AMINOGLYCOSIDE
sul1	31	28	28	28	SULFONAMIDE
aac(6')-aph(2'')	27	0	30	30	AMINOGLYCOSIDE
ant(6)-Ia	30	13	12	13	AMINOGLYCOSIDE
aph(3')-IIIa	30	30	30	30	AMINOGLYCOSIDE
erm(F)	30	28	28	28	MACROLIDE
lnu(G)	26	30	30	0	LINCOSAMIDE

A total of 24 ARG types were found in the compost samples (see table 7). Genes encoding for aminoglycoside resistance are the most prominent ones with 23.9% of all detected ARGs. This drug class also has the highest number of resistance genes with 73 (19.2% of total ARGs). Between 14% and 10% are resistance for tetracyclines, macrolides and genes responsible for multi-drug resistance. The resistance genes encoding mechanism against beta-lactam, are the second highest in terms of number of genes, with 64 (16.8% of ARGs), yet their contribution to the resistome is 7.8%. A similar proportion of 6.3% shows the resistance for lincosamide, yet they only make up for 3.4% of the ARGs.

Table 7. Antibiotic resistance prevalence by antimicrobial drug class.
The table shows the antimicrobial drug class, how many of the 381 different antibiotic resistance genes belong to this class in total number and percentage. Further how often a resistance gene for the corresponding AMR class was detected in all 95 samples (Total AMR) and the percentage (AMR %).

Drug Class	Total ARGs	ARG %	Total AMR	AMR %
AMINOGLYCOSIDE	73	19.2	849	23.9
TETRACYCLINE	37	9.7	472	13.3
MACROLIDE	35	9.2	450	12.7
MULTI-DRUG	48	12.6	388	10.9
BETA-LACTAM	64	16.8	277	7.8
CHLORAMPHENICOL	25	6.6	249	7.0
LINCOSAMIDE	13	3.4	223	6.3
TRIMETHOPRIM	17	4.5	153	4.3
PHENICOL	4	1	76	2.1
VANCOMYCIN	9	2.4	76	2.1
SULFONAMIDE	4	1	71	2.0
NUCLEOSIDE	8	2.1	62	1.7
PEPTIDE	9	2.4	32	0.9
BLEOMYCIN	3	0.8	31	0.9
FLUOROQUINOLONE	5	1.3	25	0.7
RIFAMYCIN	4	1	23	0.6
STREPTOGRAMIN	4	1	20	0.6
QUINOLONE	5	1.3	19	0.5
AMINOCOUMARIN	6	1.6	17	0.5
FOSFOMYCIN	3	0.8	14	0.4
DISINFECTANT	1	0.3	12	0.3
STREPTOMYCIN	1	0.3	10	0.3
NITROIMIDAZOLE	2	0.5	5	0.1
TRICLOSAN	1	0.3	1	0.03

5.3 Taxonomic classification

The most common phyla in all compost samples are Actinobacteria, Bacteroidetes, Firmicutes and Proteobacteria in variable distribution.
In table 8 the phyla above 1% of identified bacteria are shown by dataset.

Table 8. Detected number of ARGs and dominant phyla by datasets.
The table shows the phyla for each dataset that compose for <1% of identified bacteria in at least one sample within the ENA project. Further total amount of ARGs of all samples within the dataset.

ENA Project	Total ARGs	Phyla (< 1%)
PRJNA433771	1073	Actinobacteria, Bacteroidetes, Firmicutes, Proteobacteria
PRJNA526758	169	Actinobacteria, Bacteroidetes, Deinococcus Thermus, Chloroflexi, Firmicutes, Proteobacteria, Planctomycetes
PRJNA311675	4[b]	Actinobacteria, Bacteroidetes, Chloroflexi, Firmicutes, Gemmatimonadetes, Proteobacteria, Planctomycetes, Spirochaetes, Tenericutes, Verrucomicrobia
PRJNA549056[a]	31	Actinobacteria, Bacteroidetes, Firmicutes, Proteobacteria, Planctomycetes
PRJNA41493	1[c]	Actinobacteria, Bacteroidetes, Firmicutes, Proteobacteria, Planctomycetes
PRJNA684647	1315	Actinobacteria, Bacteroidetes, Firmicutes, Proteobacteria, Spirochaetes
PRJNA337811	2[c]	Actinobacteria, Bacteroidetes, Chloroflexi, Firmicutes, Proteobacteria
PRJNA288410	-	Actinobacteria, Bacteroidetes, Chloroflexi, Firmicutes, Gemmatimonadetes, Proteobacteria
PRJNA329458	-	Actinobacteria, Bacteroidetes, Deinococcus Thermus, Firmicutes, Proteobacteria, Tenericutes

[a] compost samples only
[b] one testpoint only
[c] single sample

The list of the percentages by samples is given in the appendix. Further, the taxonomic reports of the kraken classification, as well as the interactive krona pie chart visualization are available in google drive.

The dynamics of the composition of the main phyla from the two datasets were the testpoints are known for the samples, is visualized below.

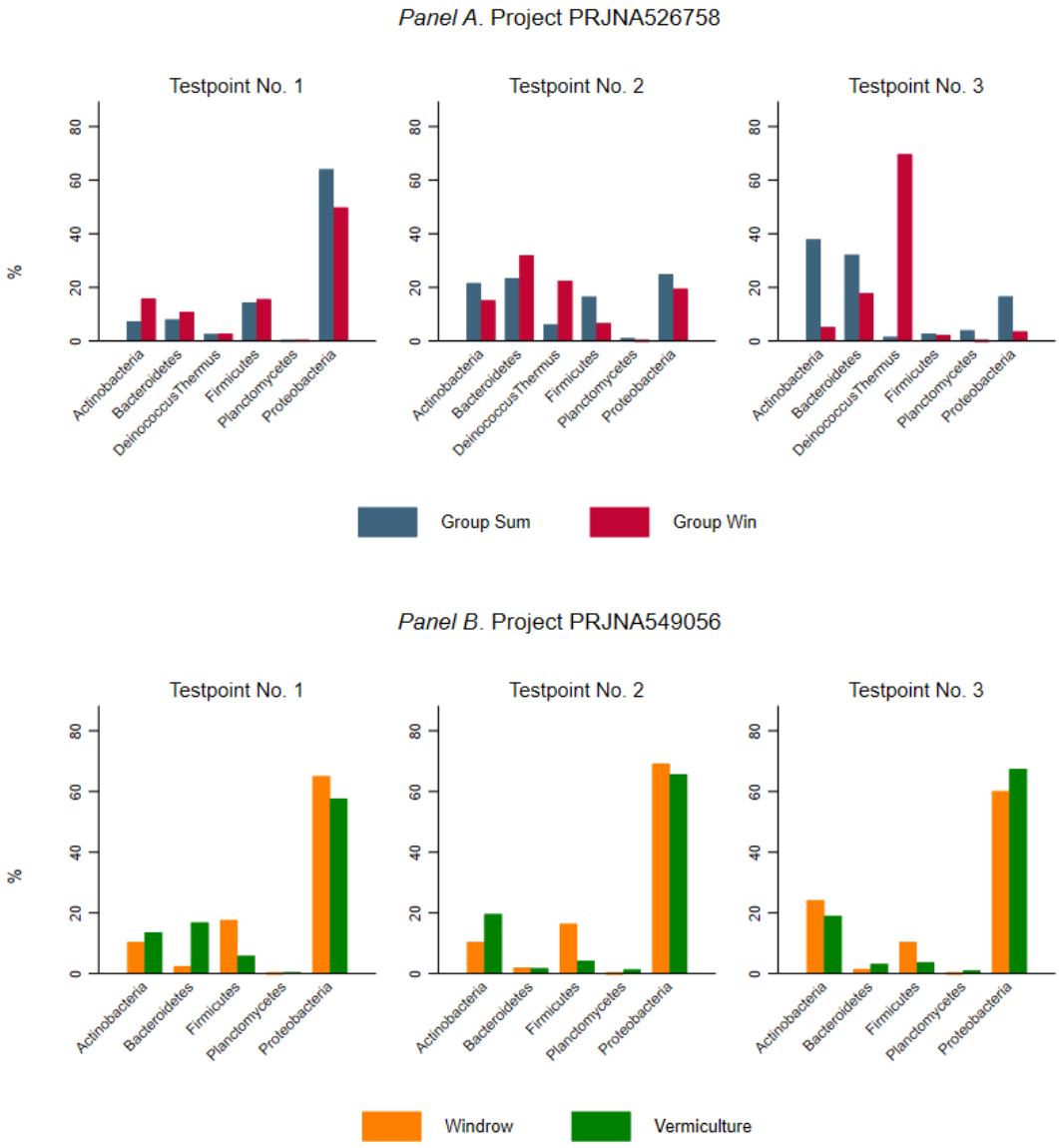

Figure 4. Variation of main phyla during composting process.
Panel A shows the change of the main phyla (<1% of identified bacteria) in aerated pile composting of dairy manure during summer (blue) and winter (red). Testpoint 1 refers to raw compost on day 0, testpoint 2 to day 3 and testpoint 3 to finished compost on day 14.
Panel B shows the change of the main phyla (<1% of identified bacteria) in windrow composting (yellow) and commercial vermiculture (green) of food waste. Testpoint 1 refers to raw compost/first layer vermiculture, testpoint 2 to unfinished compost/after shifting vermiculture and testpoint 3 to finished compost/vermiculure compost.

The main differences in the seasonal aerated pile composting of dairy manure are at the finished compost samples. One can see a dominance of *Deinococcus Thermus* within the finished compost sample taken in winter. The sample taken in the summer experiment shows a higher prevalence of *Actinobacteria, Bacteroidetes* and *Proteobacteria.*

The experiment of food composting by means of windrow or commercial vermiculture displays a higher percentage of *Firmicutes* in windrow through the whole process. The vermiculture has at the beginning a higher composition of Bacteroidetes, but at the finished compost they are similar in both groups.

5.4 Phylogenetic Analysis

The phylogenetic tree created from the reference sequences of the verified antibiotic resistance genes in the compost samples is shown in figure 5.
The gene names are colored by their corresponding antimicrobial substance class. The bars visualize how often the respective genes were reported (maximum count by one database).

The confidence levels for the phylogeny is unknown as it was not possible to carry out bootstrapping.

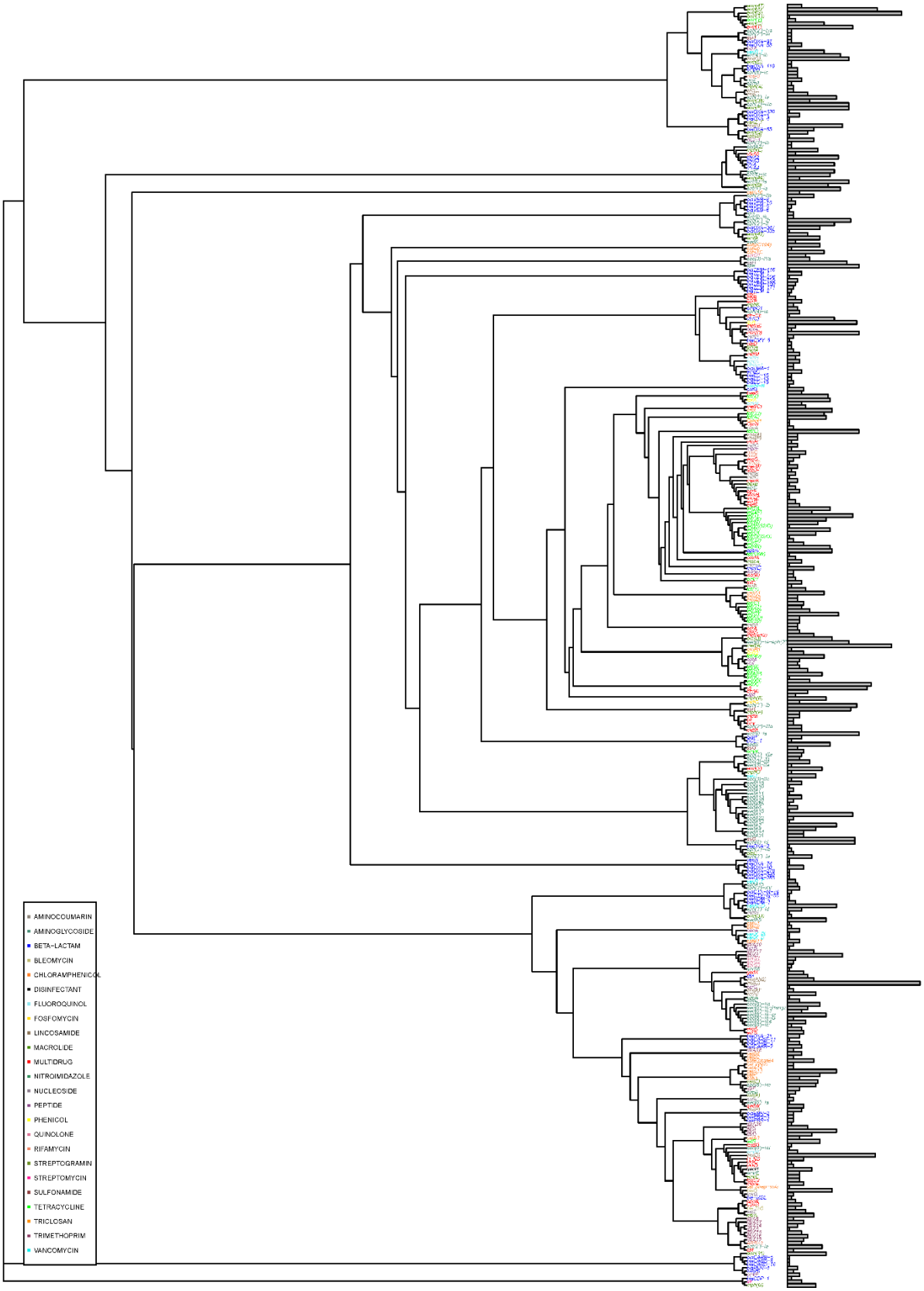

Figure 5. Phylogenetic tree of antibiotic resistance genes.
The figure shows the phylogenetic tree by neighbor joining method of the antibiotic resistance genes reference sequences. The branches of the tree were drawn without using the edge length of the phylogeny. The gene names are colored by antimicrobial substance class. The bars visualize the frequency of their detection in all datasets.

It is visible, that the antibiotic resistance genes for the same antimicrobial substance form only partially clusters. Most of the tetracycline resistance genes can be found in the middle of the tree (see figure 5 and 7).

Yet, if we take for example beta-lactam resistance (colored in dark blue), some of the subclasses like *EBR, TEM* and *VEB* build cluster, but the different types of the *OXA* subclass can be found across the tree (see figure 5, 6 and 8).

In terms of the frequency of a genes detection, having a look at the most common macrolide resistance genes *erm(A), erm(G)* and *mef(A)*, former two are located at the top of the tree, while latter is situated in the middle between the tetracycline (*tet*) clusters (see figure 6 and 7).

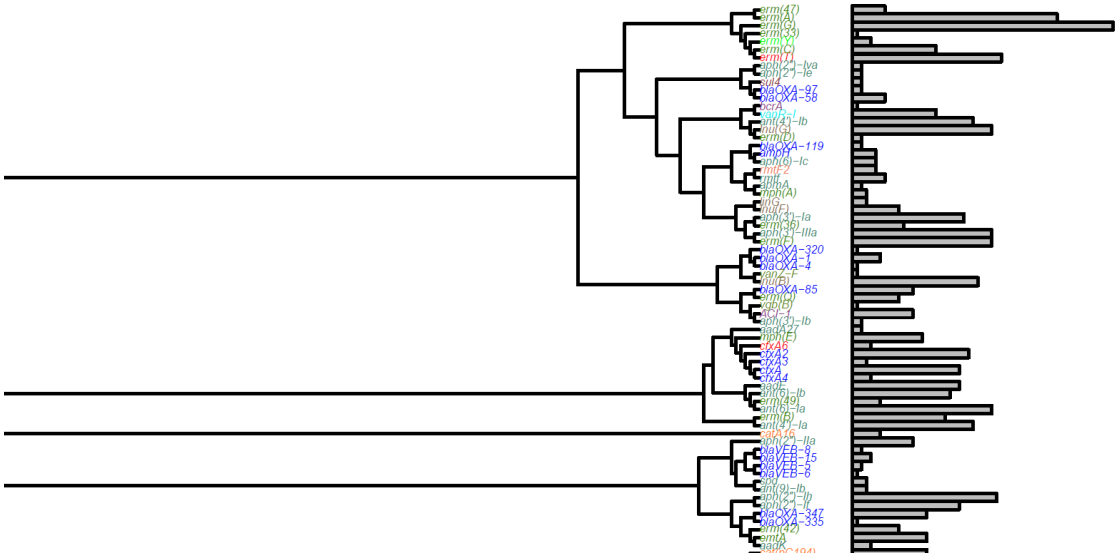

Figure 6. Phylogenetic tree fragement from the top.
The figure shows the first cluster of the phylogenetic tree. The frequently detected genes erm(A) and erm(G) are at the top. Also visible is that genes of OXA subclass are not all grouped together like the VEB subclass.

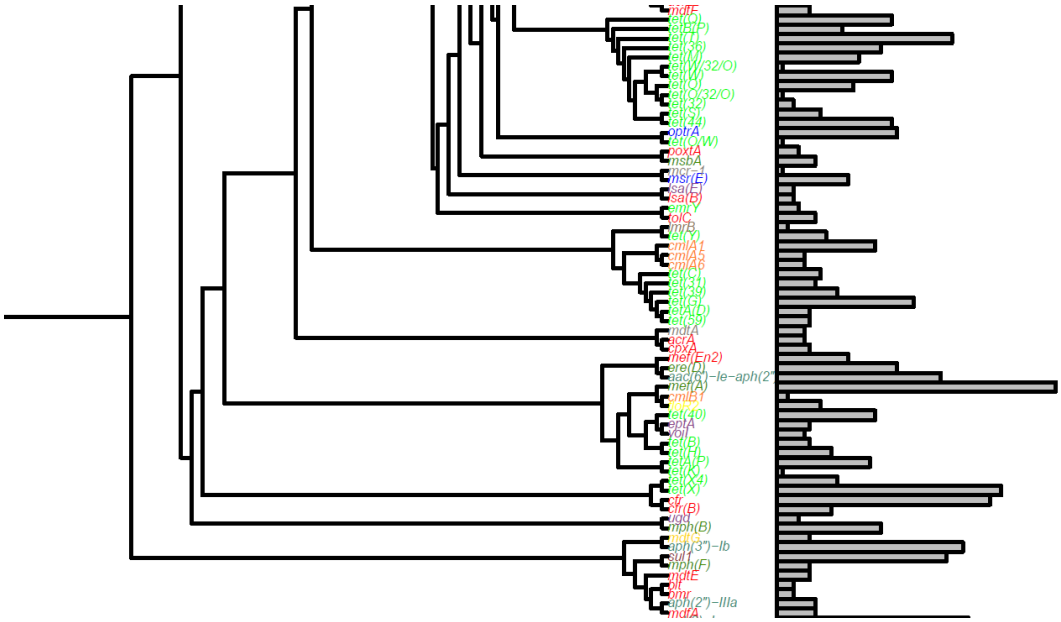

Figure 7. Fragment from the middle of the phylogenetic tree.
The figure shows the third most frequent detected gene mef(A) and clusters of the tetracycline resistance (light green).

The most frequent gene *lnu(C)* is grouped with the less frequent lincosamide resistance genes *lnu(AN2), lnu(A)'* and *lnu(A)*. Yet, the more prevalent gene *lnu(D)*, is situated further down the tree (see figure 8).

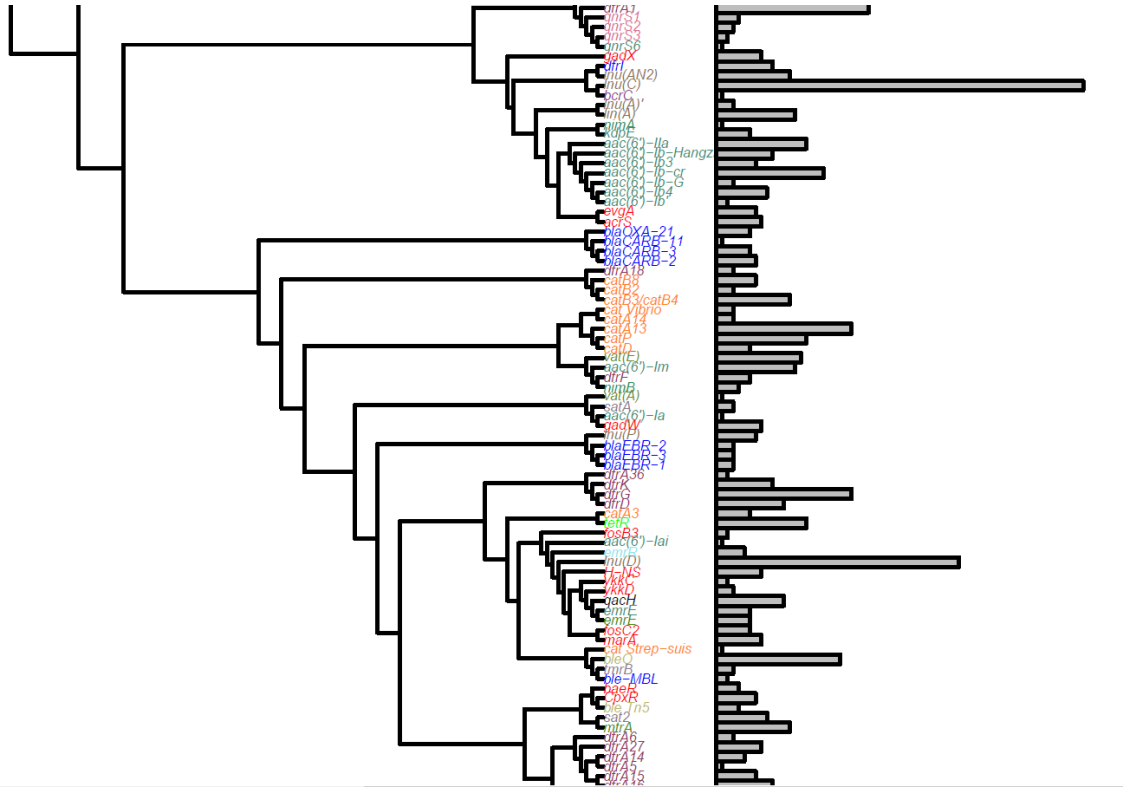

Figure 8. Fragment from the bottom of phylogenetic tree.
The figure shows that the most common lincosamide resistance genes lnu(C) and lnu(D) are not grouped together.

The phylogenetic tree is available in google drive as *phylotree.pdf*.

6 Discussion

Soil is a natural reservoir for antibiotic resistance. However, the wide-spread use of antibiotics in medicine, veterinary medicine and agriculture lead to an exacerbate dissemination of antibiotic resistance genes in the environment [4]. Composting manure or sludge from wastewater treatments can help to reduce those ARGs, but study results are still inconsistent, showing also an increase of total ARGs during the process [15, 34]. Compost is commonly used as a soil fertilizer, not only in agriculture, but also in private households. A better understanding of ARG removal mechanisms during the composting process is needed.

Most studies on this topic were performed screening for ARGs using qPCR, selecting the most clinically relevant genes only.
Here, datasets were obtained from whole genome sequencing of compost that are publicly available on the ENA server. This in-silico analysis aimed to verify the presence of ARGs in compost, obtain the microbial community and explore the phylogeny of the detected ARGs. It needs to be mentioned that, the WGS data differ in sequencing platform, sequencing depths and sample replicates. Additionally there is a lack of information of sample origin and test points. Further it needs to be considered, that the computational limitations – a private notebook running under Microsoft Windows 10 and 8 GB working memory – had an impact on the methodological possibilities.

WGS data
The search had to be limited to the ENA server because of accession problems to the MG-RAST server, but also for avoiding to download and then again upload data to the Galaxy server on which most of the analysis steps were executed.
Sequencing compost samples is more frequently done by 16S rRNA amplicon sequencing, so available WGS data is limited. Further, sharing data as open access still needs some more encouragement. Another difficulty to encounter suitable data is the scarce description or even mislabeling within the database.

In the end, it was possible to obtain data from 9 different experiments - consisting of 95 samples - that could be analyzed. Two datasets from a compost facility, ENA project number PRJNA337761 and PRJNA337762, had to be discarded as the de-novo assembly step was not possible to execute on the Galaxy server due to working memory restrictions.

ARG identification
After de-novo assembly with MEGAHIT the contigs were screened against the databases ARG-ANNOT, CARD and NCBI using the tool ABRicate and against Resfinder database using staramr. The reason for using several databases was that they vary in curated genes and also in which intervals they are updated. The number of different genes found by database are 223, 286, 236 and 209, respectively. Although Resfinder detected fewer, it is a well curated database for acquired resistance, involving HGT events. CARD identified the most as it covers a broad spectrum of ARGs [39]. It came to attention that in many cases multi-drug resistance genes were detected only by CARD. The CARD database was searched with the tool ABRicate. The reported results included the gene *SAT-1*,

which does not exist on the CARD website, and *Streptomyces_rishiriensis_parY*. Latter is marked as a private model on the website with the note that it is currently not used. This leads to the conclusion that a screening with their own tool Resistance Gene Identifier (RGI) may lead to slightly different results than the one obtained in this work.

The different databases also use their own identifiers and to some degree the gene names differ. This made it more laborious to unify the obtained results. The list of those unified results is shown in the Supplemental.xlsx. In total 381 different ARGs were found.

As mentioned before, only 9 datasets were analyzed and for most of them, information about their origin or sample points are missing. Only for 3 experiments (PRJNA549056, PRJNA526758 and PRJNA311675) a full description including assignment of the samples to the sample points was available. For PRJNA433771 a publication exists, yet due to a missing description, the samples can not be related to the testing points. This lack of sufficient data on different types of organic material and composting method, as well as, that most of the data is derived from different types of experimental settings, does not allow a statistical comparison for drawing conclusions.
Nevertheless, the results observed are concordant with what is described in literature.

The experiment on thermophilic composting of chicken manure (PRJNA433771) showed the highest number of ARGs by individual samples (see table 4). This is coherent with chicken manure having a higher ARG abundance than cattle manure [52, 7, 14].

The experiment of vermicomposting dairy manure with food waste, shows a similar number of ARGs at the endpoint compared to an experiment of aerated pile composting of dairy manure (see table 2). However, the vermiculture data only consists of endpoint testing, it remains unknown if the initial abundance was also comparable.

The food waste composted by windrow method (PRJNA549056) shows an increase of ARGs during the process, while the vermiculture for commercial purpose had eliminated all ARGs already at the test point of the unfinished stage. A study on food composting, using aerated static pile method, also observed an increase in total ARG during the process, suggesting that food waste could be an important reservoir of ARGs [53]. In terms of vermiculture, it was demonstrated that earthworms have an significant effect on the bacterial community and inhibit the growth of various human pathogenic bacteria [54].

In continuation the ARGs present in the samples that refer to finished compost are shown grouped by antimicrobial substance class.

Aminoglycoside: *aadA1, aadA14, aadA31, aph(3')-Ia, aph(3")-Ib, aph(6)-Id*
Beta-Lactam: *blaCARB-8, blaCMY-8, blaTEM-150, blaTEM-171*
Sulfonamide: *sul1, sul2*
Tetracycline: *tet(H), tet(W)*

Vancomycin: *vanR-O*

The data from two ENA projects consist of one sample each. The green-waste compost from a bioreactor and a compost from a rain forest. In those samples the genes *bla-TEM16* (beta-lactam) and *ant(9)-Ia* (aminoglycoside), *rphB*, *rphC* (rifamycin) respectively, were detected.

Interestingly, taking all samples together, *blaCARB-8* and *blaCMY-8* got reported only this one time. It needs to be mentioned, that the counts of the ARGs refer to unique ARG by sample. That means, that if a gene was reported several times within one contig, it was count only once.
The genes *aadA1, ant(9)-Ia, aph(3'')-Ib, aph(6)-Id, sul1, sul2, tet(W)* and *vanR-O* were more commonly detected (more than 30 times within 95 samples) within the analyzed data. Yet the most frequently detected genes *lnu(C), lnu(D)* (lincosamides), *erm(G), erm(A), mef(A)* (macrolides) and *tet(X)* were not found in those endpoint samples.

Genes responsible for resistance to aminoglycosides and beta-lactams are the mayor contributors to the 381 different ARGs, with 73 and 64 types each, they make up for 35.96% of ARGs. In this study, they add up to 31.7% of the compost resistome. While the aminoglycoside resistances are a little higher than the number of ARGs, the opposite is the case for beta-lactam resistance. The ARGs are 16.8% of the different genes, yet in the resistome proportion is only 7.8%. This suggests a lesser persistence of beta-lactam resistance genes
The opposite is observed for lincosamide resistance. The ARGs *lnu(C)* and *lnu(D)* were often found in the samples (65 and 43 times). This increases their contribution to the resistome with 6.3%, although the number of ARGs belonging to this group are only 13, corresponding to 3.4% of the different genes. Apart from the drug class lincosamide, also for macrolide and tetracycline a higher contribution to the resistome than to the number of ARGs is observed.
The resistance genes against tetracyclines, macrolides and multi-drugs contribute with 13.3%, 12.7% and 10.9% correspondingly.
In case of sulfonamide, the resistance against this drug class consist of 1% of the detected ARGs (sul1-4) but contribute 2% to the resistome.
This may indicate that ARGs encoding for resistance mechanisms against those drug classes could be more persistent within compost.

Looking at the occurrence of ARG in livestock waste, the drug classes tetracycline, sulfonamide, macrolide and beta-lactam are most frequently detected, with tetracyclines and sulfonamides seemingly being the most persistent ones [7, 55]. So the corresponding genes can be expected also within the finished manure compost, which is coherent with this study results.

Taxonomic classification

The main phyla detected in all samples are *Actinobacteria, Bacteroidetes, Firmicutes* and *Proteobacteria*.
The composting of food waste or added food waste, seems to show a higher prevalence of Proteobacteria. A high percentage of Proteobacteria was also

found in the compost sample from the rain forest were the organic material is unknown.
One study found higher proportions of Acidobacteria and Planctomycetes in vermicomposting [54]. Planctomycetes was also a little higher in this study in vermicompost, but the percentage was in general low. Acidobacteria were not found to be above 1% (see also Supplement.xlsx).

The most striking variation in all samples was found at the aerated pile composting experiment conducted once during summer and winter. The main difference in variation of the dominant phyla between the seasons is *Deinococcus Thermus* (see also figure 4, panel A). At testpoint 2 at day 3 accounting for 6.12% during summer and 22.52% in winter. This difference augmented in the finished compost. While during the summer the contribution of this phylum decreased to 1.51%, during winter it increased further to 69.65%. This was the only dataset with such high percentage of this phyla. In most of the samples it was below 1%. The ARGs diminished in both piles during the composting process, from 58 to 5 and 84 to 2 correspondingly (see table 2).

A summarization of the taxonomic classification below phylum level was not conducted. Neither a test for correlation analysis between the microbiome and resistome. Therefore a more detailed analysis cannot be given here.

Phylogenetic Analysis

The phylogenetic analysis of the antibiotic resistance genes was performed with the reference sequences from the NCBI or CARD database. Unfortunately, as metagenomic data is computationally intensive, examining the phylogeny of the ARG-carrying microbes within the WGS compost samples was unfeasible.
Further, it needs to be emphasized that the confidence level of the obtained phylogenetic tree remains unknown as the testing was also too computational intensive for the notebook used in this study.

Within those limitations, the phylogenetic tree of the ARGs was visualized adding information on the antimicrobial drug class and the frequency of detection to obtain a general impression of their phylogeny (see figure 5).

The contribution of the ARGs to the compost resistome are not reflected in the cluster of the tree.

In terms of antimicrobial resistance, some partial clusters are visible. Most of the tetracycline resistance genes (*tet*) are grouped in proximity in three different cluster (see also figure 7).
In comparison, within the beta-lactam resistance genes the subclasses *EBR*, *TEM* and *VEB* build cluster, yet the genes belonging to the *OXA* subclass can be found across the tree (see figure 5, 6 and 8).

One study on the diversity of tetracycline resistance genes indicated that mobile tetracycline resistance genes originate, depending on the resistance mechanism, from different phyla [56]. Further visualization that takes into consideration the

phyla and drug resistance mechanism may provide a clearer picture. However, in this study the available time was not sufficient for investigating the phylogeny in more detail.

7 Conclusions

7.1 Conclusions

The purpose of this work was to study the resistome of compost samples. Particularly, to explore the phylogenetic relationship between the detected ARGs.

Availability of publicly shared WGS data and computational constraints are the inherent limitations to this study.

Nevertheless, it was possible to analyze data from 9 different experiments. In 7 datasets antibiotic resistance genes were detected. A total of 381 different ARGs were identified on 4 different databases. A phylogenetic tree was built from 365 genes, from which 282 have a corresponding reference sequence listed on NCBI and 83 reference sequences were obtained from the CARD database. For the remaining genes a reference sequence was not found. The 4 ARG databases use to some degree different gene names, thus the results had to be combined manually. This could be a source for mistakes and the possibility exists that not all gene synonyms were identified as such. Working with databases and the tools connected to them, also means that the results are influenced by the curation of the database and the maintenance of both. In the case of the CARD database, screened with the tool ABRicate, genes got reported that on the CARD website returned either ‘no results’ or were marked as a private model and are therefore unavailable. As a phylogenetic analysis of the WGS data was not possible within this work, the tree of the reference sequences reflects only a general overview of the antibacterial resistance genes. In the visualization step, information about the corresponding antimicrobial drug class and the frequency of their occurrence within the compost samples was added. Some partial clusters are visible with regard to the antimicrobial drug class. The frequency of detection of the ARGs is not reflected. The available time for the analysis did not allow for further visualizations, taking into consideration the phyla and drug resistance mechanism, which may provide a clearer picture. It also needs to be mentioned, that the confidence level of the phylogeny could not be tested as the working memory resulted to be insufficient for this task.

As for the working process itself, at first glance the task of identification of ARGs within WGS data and subsequently phylogenetic analysis seemed to be straight forward. While researching the topic of the resistome in compost and soil itself and the bioinformatical methods for analyzing metagenomic data, it became soon apparent, that the complexity of this study was somewhat underestimated.
It was clear from the beginning that the available data will be limited and that due to the data size most of the processing needed to be done on the Galaxy platform. Still, the search of data resulted more complicated simply because of a lack of description given for the few datasets available. Although the Galaxy platform

serves its purpose quite well, the processing times are somewhat unpredictable, the tools that can be used for the analysis are naturally limited and may not always be updated.

As for the complexity of the topic itself, the available time for conducting this study was not sufficient for an extensive analysis of the resistome. The planning focused on the identification of the ARGs and the phylogenetic analysis. During the work process it became evident, that the microbial community needs a deeper analysis as what was possible here. Although examining the mobile genetic elements was not part of this work, they are needed to be investigated, as they are linked to the distribution of resistances between bacteria. And the ability of horizontal gene transfer influences the phylogeny.

Taking all together, further analysis of the resistome with the adequate computational infrastructure is advisable.

7.2 Future research lines

There are several analysis that still remain open.
To start with, a network analysis for investigating the co-occurrence among ARGs was not performed. The dissemination of ARGs occurs under the influence of MGEs, so they should be analyzed within the samples. A joined analysis, taking the microbial community into consideration, should be performed to characterize horizontal gene transfer among compost bacteria.
A phylogenetic analysis performed directly with the WGS data could give a better picture of the variation among the composting process.
Also a phylogenetic analysis by resistance mechanism and bacterial taxonomy could add more information to the phylogeny.

Lastly, it would be preferable, if data sharing would be more the praxis. This would allow for more comparisons of compost types by material and method.

7.3 Assessment of planning

The milestones of this study were obtaining WGS data of compost samples, identifying ARGs within those samples and conduct the resistome analysis on them. The timetable for obtaining the data and identifying the genes was adhered to. What took up more time than expected was to unify the results from the 4 different databases as they use to some degree different gene names. Also, it was necessary to obtain a part of the ARG reference sequences manually from the CARD database. This was unforeseen and needed additional time. In the end, the main aim of obtaining a phylogenetic tree of the ARGs was fulfilled within the timeline. However, adjustments or further analysis relating ARGs and microbial community and was not possible.

The selection of the methodology was greatly influenced by the computational limitations inherent to metagenomic data. Selecting the data from ENA and the use of the Galaxy platform for the main bioinformatics workflow was the only

feasible possibility. Even so, two datasets needed to be excluded from this study as it was not possible to process them on the server due to working memory limitations.
The plan to perform the phylogenetic analysis with the R software also proved to be correct, as online tools for multiple sequence alignments and phylogenetic tree construction are limited in the number of sequences possible. Here more than 300 sequences were analyzed, a number that exceeds those established limits for the online tools.

To sum up, it was possible to carry out the study within the set timeline. Nevertheless, the availability of WGS data and facing the problem of computational limitations for processing the data proved to be a shortcoming of this work and prevented a more meticulous analysis of the resistome.

8 Glossary

ARG	-	Antibiotic Resistance Gene
ARG-ANNOT	-	Antibiotic Resistance Gene-ANNOTation Database
BLAST	-	Basic Local Alignment Search Tool
CARD	-	Comprehensive Antibiotic Resistance Database
ENA	-	European Nucleotide Archive
HGT	-	Horizontal Gene Transfer
MGE	-	Mobile Genetic Element
MUSCLE	-	MUltiple Sequence Comparison by Log- Expectation
NCBI	-	National Center for Biotechnology Information
NGS	-	Next Generation Sequencing
PCR	-	Polymerase Chain Reaction
qPCR	-	quantitative Polymerase Chain Reaction
rRNA	-	ribosomal Ribonucleic Acid
RGI	-	Resistance Gene Identifier
SRA	-	Short Read Archive
US	-	United States
WGS	-	Whole Genome Sequencing

9 References

[1] C. L. Ventola, «The antibiotic resistance crisis,» *P&T,* vol. 40, nº 277-283, p. 4, 2015.

[2] J. Muurinen, J. Richert, C. L. Wickware, B. Richert y T. A. Johnson, «Swine growth promotion with antibiotics or alternatives can increase antibiotic resistance gene mobility potential.,» *Scientific Reports,* nº 11, p. 5485, 2021.

[3] European Medicines Agency, «Sales of veterinary antimicrobial agents in 31 countries in 2018,» European Medicines Agency, Amsteram, 2020.

[4] E. Cytryn, Z. Markiewicz and M. Popowska, “Antibiotics and antibiotics resistance genes dissemination in soils,” in *Antibiotics and antibiotics resistance genes in soils*, vol. 51, M. Z. Hashmi, V. Strezov and A. Varma, Eds., Cham, Springer, 2017, p. 151–190.

[5] B. Bojarski, B. Kot y M. Witeska, «Antibacterials in aquatic environment and their toxicity to fish,» *Pharmaceuticals (Basel),* vol. 13, nº 8, p. 189, 2020.

[6] G. K. P. Guron, G. Arango-Argoty, L. Zhang, A. Pruden y M. A. Ponder, «Effects of dairy manure-based amendments and soil texture on lettuce- and radish-associated microbioota and resistomes,» *mSphere,* vol. 4, nº 3, pp. e00239-19, 2019.

[7] Y. He, Q. Yuan, J. Mathieu, L. Stadler, N. Senehi, R. Sun y P. J. Alvarez, «Antibiotic resistance gene from livestock waste: occurance, dissemination, and treatment,» *npj Clean Water,* vol. 3, nº 4, 2020.

[8] M. Petruzzello, «Compost,» de *Encylopedia Britannica Online*, 2021.

[9] M. S. Ayilara, O. S. Olanrewaju, O. O. Babalola y O. Odeyemi, «Waste management through composting: challenges and potentials,» *Sustainability,* vol. 12, nº 11, p. 4456, 2020.

[10] D. A. Neher, T. R. Reicht, S. T. Bates, J. W. Leff y N. Fierer, «Changes in bacterial and fungal cimmunities across compost recipes, preparation methods, and composting thimes,» *PLOS ONE,* vol. 8, nº 11, p. e79512, 2013.

[11] R. V. R. R. N. Misra y H. Hiraoka, «On-farm composting methods,» Food and Agriculture Organization of the United Nations, Rome, 2003.

[12] S. S. Gonawala y H. Jardosh, «Organic waste in composting: A brief review,» *International Journal of Current Engineering and Technology,* vol. 8, nº 1, pp. 836-838, 2018.

[13] M. Gou, H. WeiHu, Y.-J. Zhang, W. Jun-Tao, H. Hayden, Y.-Q. Tang and J.-Z. He, "Aerobic composting reduces antibiotic resistance genes in cattle manure and the resistome dissemination in agricultural soils," *Science of The Total Environment,* vol. 612, pp. 1600-1610, 2018.

[14] W.-Y. Xie, X.-P. Wang, Q. Li, L.-H. Wu, Q.-R. Shen y F.-J. Zhao, «Changes in antibiotic concentrations and antibiotic resistome during commercial composting of animal manures,» *Environmental Pollution,* vol. 219, p. 182–190, 2016.

[15] A. Ezzariai, M. Hafidi, A. Khadra, Q. Aemig, L. El Fels, M. Barret, G. Merlina, D. Patureau and E. Pinelli, "Human and veterinary antibiotics during composting of sludge or manure: Global perspectives on persistence, degradation, and resistance genes," *Journal of Hazardous Materials,* vol. 359, pp. 465-481, 2018.

[16] X. Quian, W. Sun, X.-J. Wang, Y.-J. Zhang, M.-L. Duan, H.-C. Li and R.-R. Zhang, "Reducing antibiotic resistance genes, integrons, and pathogens in dairy manure by continuous thermophilic composting," *Biosource Technology,* vol. 220, pp. 425-432, 2016.

[17] M. Zhang, L.-Y. He, Y.-S. Liu, J.-L. Zhao, J.-N. Zhang, J. Chen, Q.-Q. Zhang and G.-G. Ying, "Variation of antibiotic resistome during commercial livestock manure composting," *Environment International,* vol. 136, p. 105458, 2020.

[18] C. Wang, D. Dong, P. J. Strong, W. Zhu, Z. Ma, Y. Qin and W. Wu, "Microbial phylogeny determines transcriptional response of resistome to dynamic composting processes," *Microbiome,* vol. 5, p. 103, 2017.

[19] J.-Q. Su, B. Wei, W.-Y. Ou-Yang, F.-Y. Huang, Y. Zhao, X. Hui-Juan and Y.-G. Zhu, "Antibiotic Resistome and Its Association with Bacterial Communities during Sewage Sludge Composting," *Environmental Science & Technology,* vol. 49, no. 12, p. 7356–7363, 2015.

[20] F. N. Ezugworie, igbokwe y C. O. Onwosi, «Proliferation of antibiotic-resistant microorganisms and associated genes during composting: An overview of the potential impacts on public health, management and future,» *Science of The Total Environment,* vol. 784, p. 147191, 2021.

[21] R. Leinonen, R. Akhtar, E. Birney, L. Bower, A. Cerdeno-Tárraga, Y. Cheng, I. Cleland, N. Faruque, N. Goodgame, R. Gibson, G. Hoad and M. Jang, "The European Nucleotide Archive," *Nucleic Acids Research,* vol. 39, no. suppl. 1, p. D28–D31, 2011.

[22] V. Jalili, E. Afgan, Q. Gu, D. Clements, D. Blankenberg, J. Goecks, J. Taylor and A. Nekrutenko, "The Galaxy platform for accessible, reproducible and

collaborative biomedical analyses: 2020 update," *Nucleic Acids Research,* vol. 48, no. W1, p. W395–W402, 2020.

[23] S. Andrews, "FastQC a quality control tool for high throughput sequence data," 2019. [Online]. Available: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

[24] D. Li, C.-M. Liu, R. Luo, K. Sadakane and T.-W. Lam, "MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph," *Bioinformatics,* vol. 31, no. 10, p. 1674–1676, 2015.

[25] T. Seemann, "ABRicate: mass screening of contigs for antiobiotic resistance genes," 2016. [Online]. Available: https://github.com/tseemann/abricate.

[26] A. Petkau, "staramr," 2018. [Online]. Available: https://github.com/phac-nml/staramr.

[27] D. E. Wood and S. L. Salzberg, "Kraken: ultrafast metagenomic sequence classification using exact alignments," *Genome Biology,* vol. 15, p. R46, 2014.

[28] B. D. Ondov, N. H. Bergman and A. M. Phillippy, "Interactive metagenomic visualization in a Web browser," *BMC Bioinformatics,* vol. 12, p. 385, 2011.

[29] C. Schoch, S. Ciufo, M. Domrachev, C. Hotton, S. Kannan, R. Khovanskaya and D. Leipe, "NCBI Taxonomy: a comprehensive update on curation, resources and tools," *Database,* vol. 2020, p. baaa062, 2020.

[30] R Core Team, "R: A language and environment for statistical computing," R Foundation for Statistical Computing, 2021. [Online]. Available: https://www.r-project.org/.

[31] BarD Software s.r.o, «GanttProject,» BarD Software s.r.o, 2020. [En línea]. Available: https://www.ganttproject.biz/.

[32] K. Eckstrom y J. W. Barlow, «Resistome metagenomics from plate to farm: The resistiome and microbial composition during food waste feeding and composting on a Vermont poultry farm,» *PLOS ONE,* nº 14, p. e0219807, 2019.

[33] G. Cui, F. Lü, H. Zhang, L. Shao y P. Hea, «Critical insight into the fate of antibiotic resistance genes during biological treatment of typical biowastes,» *Bioresource Technology,* nº 317, 2020.

[34] R. Coa, J. Wang, W. Ben y Z. Qiang, «The profile of antibiotic resistance genes in pig manure composting shaped by composting stage: Mesophilic-thermophilic and cooling-maturation stages,» *Chemosphere,* nº 250, p. 126181, 2020.

[35] K. Kitahara y K. Miyazaki, «Revisiting bacterial phylogeny,» *Mobile Genetic Elements,* vol. 3, nº 1, 2013.

[36] G. D. Wright, «The antibiotic resistome: the nexus of chemical and genetic diverisity,» *Nat Rev Microbiol,* nº 5, p. 175–186, 2007.

[37] C. U. Köser, M. J. Ellington y S. J. Peacock, «Whole-genome sequencing to control antimicrobial resistance,» *Trends in Genetics,* vol. 30, nº 9, pp. 401-407, 2014.

[38] F. Meyer, S. Bagchi, S. Chaterji, W. Gerlach, A. Grama, T. Harrison, T. Paczian, W. L. Trimble and A. Wilke, "MG-RAST version 4—lessons learned

from a decade of low-budget ultra-high-throughput metagenome analysis," *Briefings in Bioinformatics,* vol. 20, no. 4, p. 1151–1159, 2019.

[39] C. L. Gupta, R. K. Tiwari y E. Cytryn, «Platforms for elucidating antibiotic resistance in single genomes and complex metagenomes,» *Environment International,* nº 138, 2020.

[40] R. Leinonen, H. Sugawara and M. S. , "The Sequence Read Archive," *Nucleic Acids Research,* vol. 39, no. suppl. 1, p. D19–D21, 2010.

[41] V. Bortolaia, R. Kaas, E. Ruppe, C. Marilyn, S. Roberts, V. Schwarz, A. Cattoir, Philippon, A. Rosa L Allesoe, A. Rebelo, L. Ferrer Florensa, T. Fagelhauer, B. Chakraborty, G. Neumann, J. Werner, K. Bender, M. Stingl, J. Nguyen, B. Coppens, S. Xavier, H. Malhotra-Kumar, M. Westh, Pinholt, F. Muna, Anjum, A. Nicholas, I. Duggett, S. Kempf, S. Nykäsenoja, K. Olkkola, A. Wieczorek, L. Amaro and Clemente, "ResFinder 4.0 for predictions of phenotypes from genotypes," *Journal of Antimicrobial Chemotherapy,* vol. 75, no. 12, p. 3491–3500, 2020.

[42] S. Gupta, B. Padmanabhan, S. Diene, R. Lopez-Rojas, M. Kempf, L. Landraud and J. Rolain, "ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes," *Antimicrobial agents and chemotherapy,* vol. 58, no. 1, p. 212–220, 2014.

[43] B. P. Alcock, A. R. Raphenya, T. T. Y. Lau, K. K. Tsang, M. Bouchard, A. Edalatmand, W. Huynh, A.-L. V. Nguyen, A. A. Cheng, S. Liu, S. Y. Min, A. Miroshnichenko, H.-K. Tran and Werfa, "CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database," *Nucleic Acids Research,* vol. 48, no. D1, p. D517–D525, 2020.

[44] National Center for Biotechnology Information, "National Database of Antibiotic Resistant Organisms (NDARO)," National Center for Biotechnology Information, 2021. [Online]. Available: https://www.ncbi.nlm.nih.gov/pathogens/antimicrobial-resistance/.

[45] E. Paradis y K. Schliep, «ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R.,» *Bioinformatics,* nº 35, pp. 526-528, 2019.

[46] U. Bodenhofer, E. Bonatesta, C. Horejš-Kainrath y S. Hochreiter, «msa: an R package for multiple sequence alignment,» *Bioinformatics,* nº 31(24), pp. 3997-9, 2015.

[47] Edgar, R. C., "MUSCLE: multiple sequence alignment with high accuracy and high throughput.," *Nucleic Acids Res.,* vol. 32, p. 1792–1797, 2004.

[48] N. Saitou y M. Nei, «The neighbor-joining method: a new method for reconstructing phylogenetic trees.,» *Molecular Biology and Evolution,* vol. 4, nº 4, p. 406–425, 1987.

[49] W. Y. Xie, X. Zou, D. Y. Liu, Q. Li, Q. Shen y F. J. Zhao, «Dynamics of metal(loid) resistance genes driven by succession of bacterial community during manure composting,» *Environmental Pollution,* nº 255, pp. 113-276, 2019.

[50] A.-L. Blomström, C. Lalander, A. J. Komakech, B. Vinneras y S. Boqvist, «A metagenomic analysis displays the diverse microbial community of a vermicomposting system in Uganda,» *Infection Ecology & Epidemology,* nº 6, p. 32453, 2016.

[51] L. Zhang, L. Li, G. Sha, C. Liu, Z. Wang y L. Wang, «Aerobic composting as an effective cow manure management strategy for reducing dissemination of antibiotic resistance genes: An integrated meta-omics study,» *Journal of Hazardous Materials,* nº 368, p. 121895, 2020.

[52] Y. Xu, H. Li, R. Shi, J. Lv, B. Li, F. Yang, X. Zheng y J. Xu, «Antibiotic resistance genes in different animal manures and their derived organic fertilizer,» *Environmental Science Europe,* nº 102, 2020.

[53] H. Liao, V.-P. Friman, S. Geisen, Q. Zhao, P. Cui, X. Lu, Z. Chen, Z. Yu y S. Zhou, «Horizontal gene transfer and shifts in linked bacterial community composition are associated with maintenance of antibiotic resistance genes during food waste composting,» *Science of The Total Environment,* nº 660, pp. 841-850, 2019.

[54] B. Lv, M. Xing y J. Yang, «Exploring the effects of earthworms on bacterial profiles during vermicomposting process of sewage sludge and cattle dung with high-throughput sequencing,» *Environmental Science and Pollution Research,* nº 25, p. 12528–12537, 2018.

[55] X. Qian, W. Sun, J. Gu, X.-J. Wang, J.-J. Sun, Y.-N. Yin y M.-L. Duan, «Variable effects of oxytetracycline on antibiotic resistance gene abundance and the bacterial community during aerobic composting of cow manure,» *Journal of Hazardous Materials,* nº 315, pp. 61-69, 2016.

[56] F. Berglund, M.-E. Böhm, A. Martinsson, S. Ebmeyer, T. Österlund, A. Johnning, J. D. Larsson y E. Kristiansson, «Comprehensive screening of genomic and metagenomic data reveals a large diversity of tetracycline resistance genes,» *Microbial Genomics,* nº 6, 2020.

[57] B. J. Koch, B. A. Hungate and L. B. Price, “Food-animal production and the spread of antibiotic resistance: the role of ecology,” *Frontiers in Ecology and the Environment,* vol. 15, no. 6, p. 309–318, 2017.

[58] J. L. Martínez, “Environmental pollution by antibiotics and by antibiotic resistance determinants,” *Environmental Pollution,* vol. 157, no. 11, p. 2893–3902, 2009.

[59] A. Van Der Walt, M. Van Goethem, J. Ramond, T. Makhalanyane, O. Reva and D. Cowan, “Assembling metagenomes, one community at a time,” *BMC Genomics,* vol. 18, p. 521, 2017.

[60] G. D. Wright, “Antibiotic resistance in the environment: a link to the clinic?,” *Current Opinion in Microbiology,* vol. 13, no. 5, p. 589–594, 2010.

[61] C. Tyrrell , C. M. Burgess , F. P. Brennan and F. Walsh, “Antibiotic resistance in grass and soil,” *Biochemical Society Transactions,* vol. 47, no. 1, p. 477–486, 2019.

[62] P. W. Harrison, A. Ahamed, R. Aslam, B. T. F. Alako, J. Burgin, N. Buso, M. Courtot, J. Fan, D. H. M. Gupta, S. Holt, T. Ibrahim, E. Ivanov, S. Jayathilaka y V. Kadhirvelu, «The European Nucleotide Archive in 2020,» *Nucleic Acids Research,* vol. 49, nº D1, p. D82–D85, 2021.

[63] W.-Y. Xie, Q. Shen y F. Zhao, «Antibiotics and antibiotic resistance from animal manures to soil: a review,» *European Journal of Soil Science,* nº 69, pp. 181-195, 2018.

Annexes

Table 9. Percentage of the main phyla (threshold 1%).

ENA Project	Sample	Actinobacteria	Bacteroidetes	Chloroflexi	Deinococcus Thermus	Firmicutes	Gemmatimonadetes	Planctomycetes	Proteobacteria	Spirochaetes	Tenericutes	Verrucomicrobia
PRJNA288410	SRR2080282	67.79	1.99	1.97	0.05	23.51	0.82	0.08	3.43	-	-	0.08
	SRR2080283	62.44	8.61	1.73	0.05	7.86	2.67	0.08	15.31	-	-	0.05
	SRR2080284	50.12	15.60	1.26	0.03	10.64	-	-	22.18	-	-	-
	SRR2080285	39.20	30.13	0.09	-	21.54	0.09	0.06	8.80	-	-	-
	SRR2080286	24.66	50.17	0.09	-	3.14	0.06	0.09	21.47	-	-	0.06
	SRR2080287	42.61	23.43	0.19	0.04	16.19	0.04	-	17.43	-	-	-
	SRR2080288	1.25	2.74	0.45	-	92.93	0.06	0.18	1.70	0.06	0.18	-
	SRR2080289	48.11	0.04	0.40	-	50.83	0.26	-	0.26	-	-	-
	SRR2080290	7.31	0.67	0.03	-	90.32	0.03	0.03	1.52	-	-	-
	SRR2080291	79.28	0.08	0.03	-	20.18	0.05	-	0.32	-	-	-
	SRR2080292	6.34	4.00	0.02	0.07	73.00	-	0.09	16.34	-	0.05	-
	SRR2080293	71.06	0.32	0.03	-	26.02	0.14	-	2.40	-	0.03	-
PRJNA311675	CB1	8.18	5.25	0.07	0.06	7.24	0.01	1.56	75.59	0.06	0.04	0.15
	CB2	10.53	4.16	0.13	0.10	7.46	0.07	1.96	73.26	0.06	0.04	0.16
	CB3	9.80	4.00	0.56	0.17	3.83	0.10	6.37	69.91	0.13	0.06	0.44
	CV1	3.50	38.99	0.13	0.04	14.57	0.26	0.48	31.22	3.84	1.00	1.22
	CV2	3.71	29.00	0.48	0.13	16.31	0.98	0.39	34.38	6.23	2.23	1.23
	CV3	3.04	22.50	1.82	0.09	15.05	1.28	0.80	26.04	17.21	0.58	1.27
PRJNA329458	SRR3932001	16.55	40.50	0.15	3.88	11.15	0.06	0.13	27.20	0.03	0.09	0.02
	SRR3932002	7.80	54.02	0.06	1.07	4.73	0.03	0.09	32.03	0.00	0.05	0.01
	SRR3932003	18.08	33.72	0.33	3.19	24.60	0.55	0.34	18.88	-	0.09	0.03
	SRR3932004	23.42	14.29	0.18	0.52	49.40	0.12	0.17	11.64	0.02	0.04	0.04
	SRR3932005	12.69	6.71	0.08	0.59	68.79	0.13	0.26	10.59	-	0.04	0.03
	SRR3932006	17.73	7.67	0.26	0.38	56.66	0.14	0.13	16.72	0.00	0.05	0.06
	SRR3932007	7.95	2.41	0.05	0.35	85.08	0.01	0.03	4.00	0.00	0.00	0.01
	SRR3932008	2.87	5.97	0.08	0.32	86.34	0.03	0.02	3.68	0.00	0.01	0.01
	SRR3932009	10.98	15.23	0.02	0.70	59.18	0.00	0.02	13.49	-	0.05	0.01
	SRR3932010	13.11	46.87	0.12	0.30	8.47	0.06	0.11	30.56	0.00	0.14	0.02
	SRR3932011	10.87	47.20	0.09	0.15	22.56	0.01	0.05	18.83	-		-
	SRR3932012	37.44	21.10	0.10	0.74	15.03	0.06	0.21	24.73	0.03	0.28	-
	SRR3932013	13.24	11.02	0.16	1.34	52.74	0.14	0.22	20.31	0.12	0.37	0.04
	SRR3932014	9.25	6.01	0.09	1.63	70.82	0.22	0.27	11.16	0.01	0.25	-
	SRR3932015	8.92	8.86	0.13	0.53	75.18	0.03	0.16	6.06	-	0.01	0.02
	SRR3932016	3.88	5.49	0.08	0.25	85.49	0.01	0.04	4.10	-	-	0.01
	SRR3932017	5.62	19.06	-	0.20	64.97	-	0.02	9.93	0.00	0.03	0.00
	SRR3932023	16.79	34.86	0.10	0.23	18.53	0.03	0.03	29.13	0.01	0.09	0.03
	SRR3932030	12.80	4.03	0.34	0.31	60.78	0.03	0.58	20.61	0.01	0.02	0.01
	SRR3932031	14.23	0.36	0.38	0.11	67.05	0.01	0.07	16.14	0.05	-	-
	SRR3932032	17.73	0.82	0.24	0.21	66.71	0.00	0.09	8.68	0.00	2.47	0.00
	SRR3932033	15.17	0.52	0.32	0.05	63.12	0.05	0.03	17.22	0.01	0.09	-
	SRR3932034	18.04	7.08	0.40	0.47	54.13	0.02	0.04	17.64	0.01	0.01	0.00
	SRR3932035	7.61	2.02	0.23	0.24	75.60	0.03	0.12	11.43	0.00	0.16	0.02
	SRR3932036	1.94	40.40	0.09	1.54	14.34	0.05	0.13	40.37	0.01	0.95	0.03
PRJNA337811	rain forest	9.46	2.61	5.07	0.10	6.48	0.01	0.05	74.06	0.05	0.01	0.03
PRJNA41493	bioreactor	22.32	7.05	0.77	0.35	9.59	0.20	1.30	52.02	0.06	0.02	0.22
PRJNA433771	XC12	45.52	0.84	0.12	0.18	41.42	0.03	0.17	9.67	0.03	0.06	0.04
	XC11	51.30	1.53	0.14	0.18	26.49	0.04	0.20	17.48	0.03	0.07	0.04
	XC10	53.30	1.00	0.09	0.26	31.28	0.03	0.19	11.42	0.03	0.07	0.05
	XC9	45.27	0.97	0.06	0.15	41.60	0.02	0.20	8.53	0.06	0.08	0.05
	XC8	27.63	1.43	0.07	0.12	57.45	0.02	0.23	9.22	0.09	0.14	0.05
	XC7	38.96	1.18	0.07	0.14	46.21	0.02	0.22	9.97	0.07	0.11	0.05
	XC6	9.89	9.29	0.07	0.12	46.24	0.02	0.15	28.76	0.11	0.75	0.03
	XC5	8.13	8.25	0.05	0.07	56.86	0.01	0.10	23.06	0.07	0.54	0.02
	XC4	8.63	8.40	0.06	0.08	56.28	0.01	0.11	22.75	0.09	0.45	0.02
	XC3	6.76	25.31	0.03	0.03	11.96	0.00	0.07	52.34	0.07	0.30	0.03
	XC2	6.08	28.83	0.03	0.03	13.83	0.00	0.06	48.16	0.07	0.42	0.03
	XC1	6.03	27.31	0.03	0.03	12.96	0.01	0.07	50.43	0.07	0.40	0.03
PRJNA526758	Sum14	37.91	32.17	1.90	1.51	2.80	0.22	3.96	16.60	0.02	0.01	0.05
	Sum3	21.57	23.43	1.21	6.12	16.56	0.31	1.07	24.92	0.09	0.07	0.05
	Sum0	7.28	8.01	0.31	2.58	14.34	0.08	0.22	64.03	0.16	0.24	0.06
	Win14	5.19	17.81	0.14	69.65	2.27	0.06	0.10	3.52	0.01	0.00	0.01
	Win6	31.34	38.64	0.69	10.43	3.32	0.24	0.18	11.94	0.02	0.00	0.02
	Win3	15.16	31.97	0.82	22.52	6.63	0.14	0.16	19.55	0.04	0.02	0.03
	Win0	15.77	10.91	0.58	2.75	15.58	0.13	0.26	49.81	0.37	0.38	0.10
PRJNA549056	Finished Compost 1	24.12	1.40	0.81	0.13	10.47	0.04	0.08	60.10	0.02	0.02	0.02
	Finished Compost 2	23.40	1.41	0.81	0.13	10.30	0.04	0.08	61.13	0.02	0.02	0.01
	Raw Compost 1	10.63	2.44	0.55	0.12	17.73	0.12	0.09	64.64	0.05	0.03	0.03
	Raw Compost 2	10.17	2.37	0.51	0.12	17.57	0.11	0.09	65.39	0.06	0.03	0.03
	Shifted Worm Castings 1	19.81	1.73	0.40	0.31	4.24	0.35	1.29	65.50	0.06	0.04	0.16
	Shifted Worm Castings 2	19.36	1.75	0.40	0.30	4.28	0.36	1.33	65.91	0.06	0.04	0.15
	Top Worm Castings 2	13.43	16.59	0.49	0.20	5.75	0.04	0.33	58.11	0.04	0.04	0.09
	Top Worm Castings 1	13.73	17.05	0.49	0.19	5.87	0.03	0.31	57.13	0.04	0.03	0.09
	Worm Castings 2	18.92	3.13	0.34	0.23	3.64	0.26	0.98	67.61	0.04	0.03	0.09
	Worm Castings 1	19.11	3.17	0.36	0.23	3.66	0.26	0.96	67.33	0.04	0.03	0.10
	Unfinished Compost 2	10.37	2.00	0.15	0.05	16.30	0.01	0.04	69.45	0.02	0.03	0.01
	Unfinished Compost 1	10.56	2.02	0.16	0.06	16.52	0.01	0.04	68.92	0.02	0.03	0.01
PRJNA684647	S13	90.56	0.07	0.00	0.01	6.59	0.00	0.01	2.43	0.04	0.01	0.00
	S12	3.09	2.52	0.06	0.08	77.72	0.01	0.07	6.58	4.52	0.45	0.06
	S11	4.83	3.98	0.11	0.15	67.26	0.02	0.11	10.51	4.83	0.84	0.10
	S2	5.68	7.50	0.11	0.02	61.17	0.01	0.19	13.78	3.95	0.40	0.10
	S1	3.67	13.71	0.11	0.02	33.05	0.01	0.24	35.33	6.13	0.52	0.14
	S10	3.55	2.49	0.07	0.10	77.75	0.01	0.07	6.91	3.39	0.50	0.07
	S9	4.59	3.68	0.11	0.13	58.55	0.01	0.09	8.42	17.72	0.59	0.09
	S8	3.99	2.92	0.09	0.11	73.24	0.01	0.08	8.79	4.23	0.66	0.08
	S7	4.58	2.73	0.07	0.09	76.07	0.01	0.07	7.37	3.01	0.63	0.07
	S6	4.07	8.80	0.11	0.13	50.14	0.02	0.12	19.19	8.79	1.13	0.11
	S5	2.82	6.65	0.07	0.10	35.62	0.01	0.08	46.53	2.63	0.52	0.08
	S4	3.26	8.83	0.08	0.10	57.37	0.01	0.11	16.44	6.74	0.71	0.09
	S3	4.63	26.12	0.17	0.17	37.68	0.02	0.14	14.09	9.53	0.29	0.16
	S18	63.89	12.47	0.01	0.02	12.34	0.00	0.01	10.38	0.01	0.02	0.00
	S17	70.82	0.27	0.00	0.01	24.12	0.00	0.01	4.20	0.05	0.02	0.00
	S16	94.46	0.53	0.00	0.01	3.50	0.00	0.01	1.20	0.02	0.02	0.00
	S15	47.39	0.72	0.04	0.13	15.58	0.15	0.10	33.27	0.05	0.04	0.02