UNIVERSITÀ DEGLI STUDI DI PADOVA

DEPARTMENT OF COMPARATIVE BIOMEDICINE AND FOOD SCIENCE

Master's Degree in Biotechnologies for Food Science

MICROBIOME DYNAMICS IN NEWBORN SWINE: COMPREHENSION METAGENOMICS ANALYSIS FROM DAY 1 TO DAY 10 AFTER PIGLETS ' BIRTH.

Supervisor

Massimo Bellato, Ph.D.

Prof. Babara Di Camillo

Co-supervisor

Prof. Luca Bargelloni

Graduating Student

Daodu, Adedamola Godwin

Student Number: 2049661

ACADEMIC YEAR 2023/2024

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Bacterial infections are among the most frequent causes of death in farms. Among others, *Clostridium difficile* emerges to be one of the worst pathogens leading to premature death in newborn piglets in swine farms, causing severe diarrhea. This infection is often managed through therapies that inhibit the toxins' effects while not removing the bacterial pathogen from the swine intestine. This research, in addition to investigating the abundance of *C. difficile* in the sows' and piglets' microbiomes in a farm adopting antibiotic vaccine to neutralize the effect of its toxins, delves into the evolution of microorganisms in piglets from 24 to 240 hours after birth. Additionally, a comparative analysis of the microbial composition between piglets and sows at 24,96, and 240 hours is conducted.

This study undertakes a thorough microbiome analysis of pig fecal samples, focusing on sow-piglet pairs from distinct groups to count for 10 mother sows and 3 piglets each. Samples are collected at pivotal time points: 24, 96, and 240 hours after birth, to reveal insights into the microbial landscape. Covariates such as time, diarrhea occurrence, sex, category (sow or piglet sibling), gestation, family, and room are included in the metadata. Sows that are primiparous and pluriparous are indicated, as the formers have been exposed to the gestational area of the farm and where birthing takes place only once.

The research employs a robust pipeline, including sample collection, sequencing, and QIIME2-based pre-processing of data. Diversity analyses, encompassing different indices, are adopted to explore microbial compositional variations at different taxonomic levels; finally, differential abundance analysis is performed using the MAASLIN2 package.

Results from Alpha diversity analysis unveil significant associations, with time, category, and at the genus level - diarrheic phenotype. The Beta diversity analysis shows significant microbial distinctions emerging between sows and piglets, with piglets exhibiting a pronounced microbial compositional shift at 24 hours post-birth. Surprisingly, no significant differences arise between microbial compositions at 96 hours after the birth of piglets, indicating stability during this critical developmental phase. At the genus level, there is a noteworthy convergence of piglet microbial compositions towards sow patterns, particularly evident at 96 hours post-birth. However, by 240 hours, a subtle but discernible separation merges again. Similar trends are observed at the species level, although at a less pronounced level. Contrastingly, ASV analysis shows a distinct separation at all considered time points, highlighting a unique microbial pattern that deviates from the observed trends at the genus and species levels.

Differential abundance analysis reveals the absence of specific *C. difficile*-related scenarios but highlights the differential presence of other *Clostridium* genera, with *Clostridium perfringens* significantly abundant in samples 24 hours from birth. Additionally, *Bacteroides ovatus* exhibits a threefold increase in diarrheic samples compared to non-diarrheic samples but with a non-significant Q-value. This study contributes valuable insights into swine microbiome dynamics, shedding light on microbial variations, stability, and evolution of new-born piglets and sows across 24,96 and 240 hours after birth.

Summary

1. INTRODUCTION

Overview of *Clostridium difficile* in pig farm

Background

Clostridium difficile, commonly known as *C. difficile*, is a gram-positive, anaerobic bacterium that holds significant clinical implications. As elucidated by previous studies, the organism is distinguished by its remarkable capacity to produce spores, which enhances its environmental resistance (Gil et al., 2017; Shen, 2020). This spore-forming ability, as noted by several researchers, contributes to the enduring nature of *C. difficile* in various settings. According to Napolitano and Edmiston (2017), the bacterium is predominantly linked to gastrointestinal infections, especially antibiotic-associated diarrhea, and false membrane colitis in humans. This assertion underscores the clinical relevance of *C. difficile* and its propensity to cause substantial morbidity. Furthermore, as underscored by Tijerina-Rodríguez et al. (2019), the virulence of *C. difficile* lies in the formation of toxins. Notably, toxins A and B emerge as the primary virulence factors responsible for the clinical manifestations of the associated diseases. This perspective sheds light on the intricate molecular mechanisms employed by the bacterium to induce pathological outcomes in the host. Therefore, the multi-dynamic nature of *C. difficile* manifested by its spore-forming ability and toxin formation, highlights its importance with regard to gastrointestinal infections.

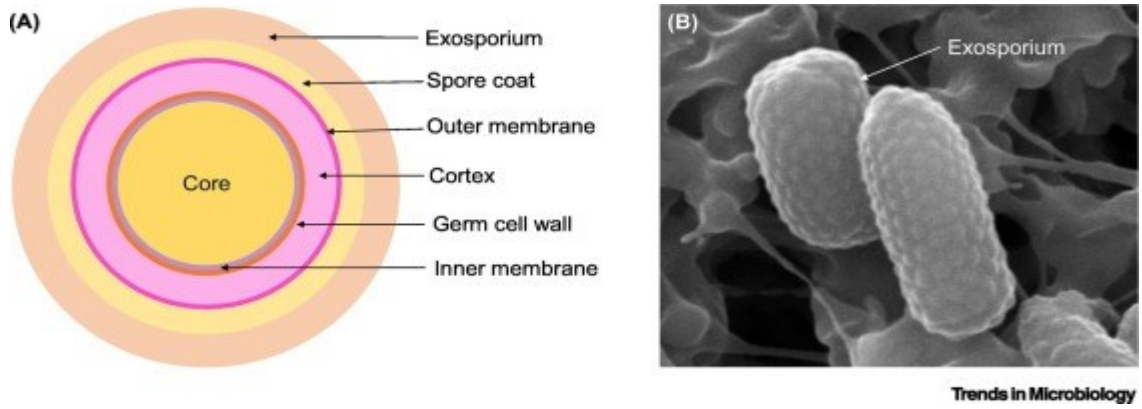


Figure 1.1: *Clostridioides difficile* Spore Structure (Adapted from Lawler, Lambert and Worthington, 2020)

Infections caused by *C. difficile* are a major concern to both veterinary and human medicine. *C. difficile* is known as a potential pathogen in veterinary medicine that affects several animals with pigs being notably susceptible (Grześkowiak et al., 2019). Spigaglia et al. (2023) stated that having an insight into the role of *C.*

difficile in veterinary medicine is vital not just for the health of the animal but also due to its zoonotic potential, stating instances where infections were transferred from animals to human. This emphasises the interconnectedness of health concerns across species boundaries. In addition, Tschudin-Sutter et al. (2018) posit that *C. difficile* is the leading cause of infection in healthcare settings. This has not only resulted in an up rise in morbidity and death rate but has also increased healthcare cost. The effect of *C. difficile* goes beyond veterinary medicine as it affects the cost of health by human as such resulting in multidimensional tactics in addressing and reducing its effect.

Importance of *C. difficile* in pig farm

C. difficile is a prevalent pathogen in pig farms globally, it finds pigs as natural reservoirs, where it dwells and shed and plays a major role in environmental contamination as highlighted by Keessen et al. (2013). The occurrence of the bacterium is aggravated by intensive pig farming system approaches such as high-density housing and consistent usage of antibiotics are employed to create an atmosphere favourable to the persistence and spread of *C. difficile* (Martin, Monaghan and Wilcox, 2016). Moreover, as highlighted by Maes et al. (2019), the use of intensive farming methods not only allows the nourishment of the pathogen but also promotes its transmission. The increased prevalence of *C. difficile* among pigs has elevated serious concerns about the possible transmission of the bacterium within the farm environment, posing risks to both livestock and humans. This intricate relationship between farming practices and the prevalence of *C. difficile* requires a critical examination of the related environmental and public health implications.

C. difficile, a bacterium linked with countless of health issues in pigs, has been implicated in several adverse consequences. Researchers have methodically detected that its clinical manifestation in pigs involves debilitating symptoms like diarrhoea, reduced feeding efficiency, and, in severe cases, death (Flock, 2017; Luppi et al., 2023). In addition, scholars such as Hain-Saunders et al. (2022), and Baines and Wilcox (2015) have shrewdly inferred that the economic effects of *C. difficile* range beyond its direct effect on animal health. It adversely affects the growth rate of pigs and increase cost linked to veterinary care and treatment. Moreover, Gwenzi et al. (2023) stated that the insidious cycle continues as farmers, challenged with *C. difficile* infections, frequently resort to antibiotics for treatment. This dependence on antibiotics not only worsens the problem but also contributes to the worrisome movement of antimicrobial resistance. Thus, the

unified nature of health, economic, and environmental factors demand an inclusive approach to lessen the far-reaching consequences of *C. difficile* in pig farming.

The zoonotic potential of *C. difficile* emphasises an extra layer of significance in its presence among pig farmers. Current studies by Lim, Knight and Riley (2020) and Grześkowiak et al. (2019) have strongly proposed that the main reservoir of *C. difficile* is the gastrointestinal tract of animals, with pigs being a prominent host. This increases concerns about the transmission of the bacterium to humans, particularly those in close contact with infected animals. Consequently, individuals, such as veterinarians, farmers, and those involved in pork production, are stated to be at a greater risk of contracting *C. difficile* infections. The complex nature of the pathogen's transmission underlines the need for comprehensive preventive tactics within the agricultural and veterinary sectors.

C. difficile: Microbiology and Pathogenesis

Microbial characteristics of C. difficile

C. difficile is a gram-positive, and anaerobic bacterium with distinctive structure. Cairns (2018) opined that *C. difficile* is rod-shaped and ranges from 2 to 8 micrometres. Al-Hinai, Jones and Papoutsakis (2015) stated that *C. difficile* forms characteristic endospore which are called spores under specific conditions. Awad *et al.* (2014) stated that these spores have strong exterior layer that protects against harsh environmental conditions like heat and radiation. In addition, Schnizlein and Young (2022) posits that the capability to produce spores enables *C. difficile* to persist in the environment for extended periods which contributes to its transmission as well as survival. An important feature of *C. difficile* lifecycle is sporulation. DuPont (2014) stated that whenever the bacterium faces unfavourable conditions like inadequate nutrient and exposure to antibiotics, sporulation occurs. However, Barra-Carrasco and Paredes-Sabja (2014) stated that spores are highly resistant to environmental stresses allowing them to withstand adverse conditions and facilitating their persistence. Furthermore, Barra-Carrasco and Paredes-Sabja (2014) also disclosed that this resistance contributes to the challenges linked with eradicating and controlling *C. difficile* in both agricultural environments and healthcare settings. In summary, the unique characteristics of *C. difficile,* with its rod-shaped structure, spore formation, and the resilience of these spores, underline the challenges linked with managing and eradicating this bacterium in numerous settings.

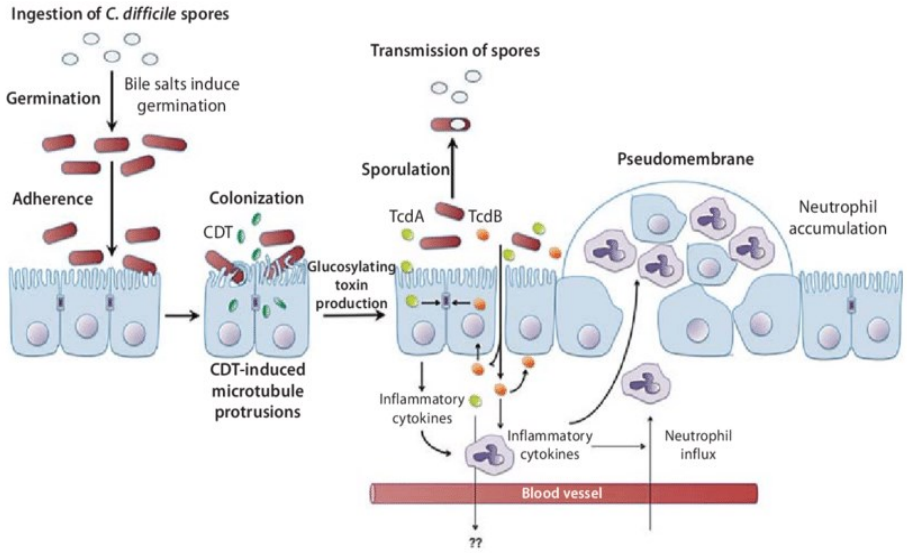


Figure 1.2: Pathogenesis of *C. difficile* (Shen, 2012).

C. difficile toxin production

C. difficile generates toxins which play a crucial role in pathogenesis of linked disease. Aktories, Schwan and Jank (2017) opined that the two key toxins formed by *C. difficile* are designated as toxin A (TcdA) and toxin B (TcdB). While Gao *et al.* (2020) noted that these toxins are large and complex proteins that possess overlapping functions, Di Bella et al. (2016) reported that both toxin A and toxin B damage the intestinal epithelial cells of the host resulting to *C. difficile* infection. In addition, Di Bella et al. (2016) also posits that both toxins have similar mechanism as they target the host cells in the gastrointestinal tract. Wang et al. (2018) elaborated that glucosyltransferase modifies and disrupts the function of the small GTPases within the host cell. This alteration leads to the breakdown of the cytoskeleton, leading to death of cells and damage to the colon mucosal lining. Furthermore, Popoff (2018) highlighted that toxins stimulate the release of pro-inflammatory mediators which contributes to the characteristic inflammation linked to *C. difficile* infections. This interconnected cascade of events highlights the complex and destructive nature of *C. difficile* toxins in the pathogenesis of the related disease.

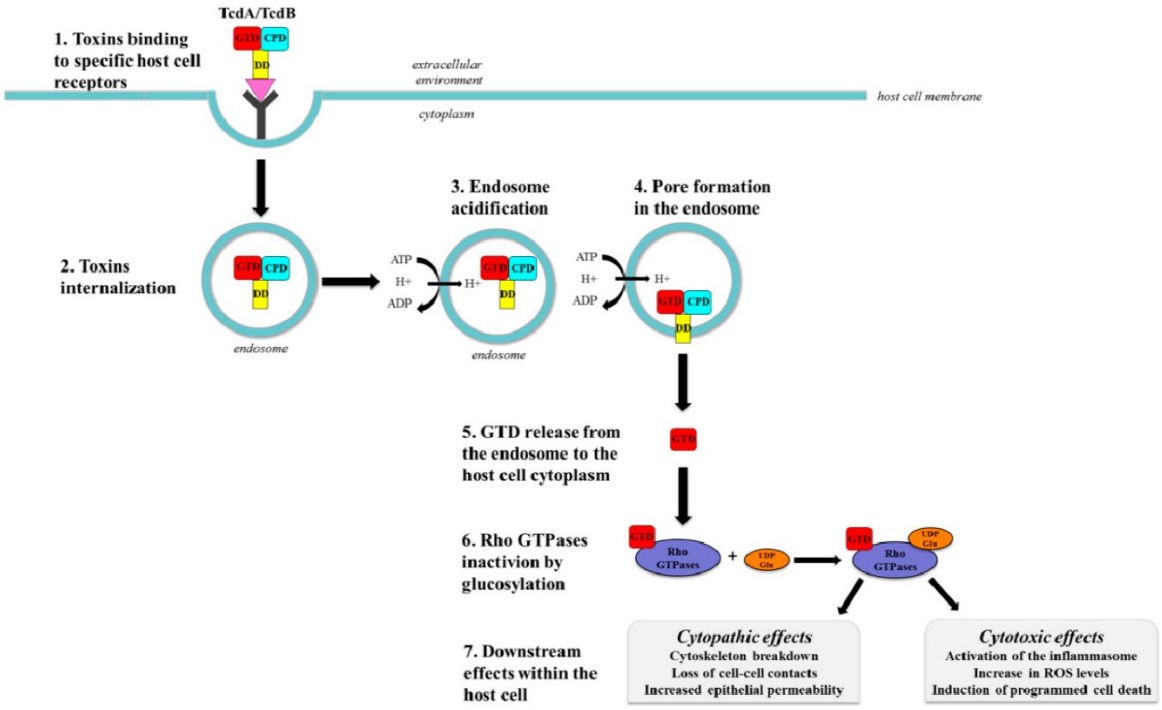


Figure 1.3: Mechanisms of Action of TcdA and TcdB (Adapted from Di Bella et al., 2016)

Factors affecting the development of disease.

Numerous factors affect the development of *C. difficile*-related diseases. One pivotal part, as underlined by Lim, Knight and Riley (2020) and Spigaglia (2015), is the extensive use of antibiotics. Antibiotics interrupt the mild balance of the usual gut microbiota, generating an environment favourable to C. difficile colonisation. This illuminates a key predisposing factor that sets the stage for the subsequent cascade of events resulting in infection. Moreover, a broader perspective by Gebreyes et al. (2020) revealed that additional risk factors linked to the onset of *C. difficile* infection are underlying health conditions and the inevitable march of age. This highlights the multifaceted nature of the variables contributing to the vulnerability of individuals to *C. difficile*-related diseases.

Lee, Chico and Renshaw (2017) investigate into the complex nature of host-pathogen collaboration, it becomes obvious that the interrelationship between the immune response of the host and the virulence factors of *C. difficile* is a nuanced factor in the disease's development. This perspective sheds light on the complex web of factors influencing the pathophysiology of *C. difficile*-related diseases, moving beyond mere predispositions to discover the natural interrelationships between the host and the pathogen. The seriousness of *C. difficile* infections varies, and ranges from asymptomatic colonization to mild diarrhea,

and, in severe cases, toxic megacolon. Therefore, the multidimensional nature of *C. difficile*-related diseases requires an inclusive method to understand the intricate interplay of factors prompting their development.

Prevalence and epidemiology of *C. difficile* in Pig Farms

C. difficile exhibits a vast distribution all over the globe and the prevalence varies across various area. Alves et al. (2022) affirmed that the prevalence of *C. difficile* in pig farms is affected by geographical factors such as climate and the structure of the pig farm. Maes et al. (2019) opined that some regions tend to face a high rate of *C. difficile* prevalence as a result of the variation in pig farming, climatic conditions, and antimicrobial usage. However, Spigaglia (2015) stated that the prevalence of *C. difficile* in pig farms is closely associated with farm management practices. Gebreyes et al. (2020) stated that intensive farming systems with high animal density, frequent movement of animals, and limited biosecurity measures contributes to the fast spread of *C. difficile*. Furthermore, the implementation of effective management practices such as adequate sanitation and biosecurity tactics are vital in reducing the risk of *C. difficile* transmission in farms.

The usage of antibiotics in pig farm is an important factor that affects the prevalence of *C. difficile*. Moono (2017) noted that antibiotics affect the normal balance of the microbiota of the gut and establish a conducive environment for *C. difficile* colonisation and overgrowth. In addition, Andrés-Lasheras et al. (2016) opined that selective pressure exerted by antibiotic usage leads to the emergence of antibiotic-resistant strains of *C. difficile*. It is crucial to monitor and regulate the usage of antibiotics to minimise the risk of *C. difficile* infections in addressing the wide concern of antimicrobial resistance. Keessen et al. (2013) stated that numerous factors contribute to the prevalence of *C. difficile* in pig farms. The persistence of *C. difficile* spores in the environment, particularly in faeces and contaminated surfaces enables the transmission among animals.

Accurate surveillance and monitoring are vital in assessing the prevalence of *C. difficile* in pig farms and the implementation of timely intervention. Gebreyes et al. (2014) noted that the diagnostic approach plays crucial role in identifying animals that are infected and in monitoring contaminated environments. Goldenberg and French (2011) stated that the most common diagnostic techniques are polymerase chain reaction (PCR) and immunoassays (EIAs). According to Pallis et al. (2013) PCR allows the detection of *C.*

difficile deoxyribonucleic acid (DNA) in fecal samples. It is however important to note that the integration of these diagnostic tools into routine surveillance programmes allows early detection and intervention, thereby lowering the effect of C. difficile on pigs.

Impact of *C. difficile* on Pig Health and Productivity

Clinical manifestation of *C. difficile*

Infections caused by *C. difficile* is often shown by clinical signs particularly diarrhoea and other gastrointestinal symptoms. Moono (2017) opined that diarrhoea is the most common C. difficile and that pigs that are affected exhibit watery faeces. Navarre and Pugh (2002) noted that gastrointestinal symptoms include colic, alteration of appetite and abdominal discomfort. The severity and duration of diarrhoea vary with certain pigs experience acute diarrhoea. In addition, Chowdhury et al. (2016) stated that the seriousness of C. difficile-related diseases in pigs range from mild to severe and the outcomes depend on numerous factors such as the strain of C. difficile as well as the overall well-being of the pig. Furthermore, Napolitano and Edmiston (2017) posit that severe cases of *C. difficile* leads to a more significant health concern like dehydration and weight loss, Napolitano and Edmiston (2017) also stated that in extreme cases it leads to death. Jurburg et al. (2019) opined that pigs with compromised immune system are more susceptible to severe outcomes.

Economic implication of C. difficile

C. difficile contributes to increasing death rate in infected pigs. Czepiel et al. (2019) opined that serious cases of *C. difficile* infection lead to dehydration, and systemic complications result in death. Mortality rates differ depending on the virulence of *C. difficile* strain, health status, and age of pigs (Grześkowiak et al., 2019). The rising death rates not only affect the welfare but also pose economic challenges to pig farmers resulting in financial loss as a result of the decline in production and rise in cost linked to the care and treatment of pigs. *C. difficile* infection adversely affects the growth rate and efficient feeding of pigs (Grześkowiak et al., 2019). Kim et al. (2012) stated that diarrhea and gastrointestinal imbalance result in poor nutrient absorption which leads to weight loss and affects growth. The economic consequences of reduced growth rates extend beyond immediate losses as delayed growth affects the overall efficiency of pork production.

Metagenomics

Metagenomics is the direct genetic analysis of genomes contained in an environmental sample. Kumar et al. (2020) stated that this field began with the cloning of environmental DNA, followed by functional expression screening, this was then followed by a direct random shotgun sequencing of environmental DNA. Kirubakaran et al. (2020) opined that this field of genomics bypasses the need for the isolation and cultivation of individual microbial species. In addition, Ravin, Mardanov, and Skryabin (2015) noted that metagenomic comprises collective genomes of all microorganisms with an environmental sample. Furthermore, Alves et al. (2018) posit that metagenomics goes beyond the study of an individual organism as it allows the analyses of genetic components from an entire microbial ecosystem simultaneously.

Metagenomics originates from the cloning of environmental DNA (Garlapati et al., 2019; Kumar et al., 2020), and has undergone a transformative evolution in its methodologies. The initial phase, marked by the cloning approach, laid the groundwork for subsequent advancements (Oulas et al., 2015). However, Denman and McSweeney (2015) stated metagenomics was the advent of functional expression screening that heralded a significant paradigm shift, propelling metagenomics into a new era characterized by a more targeted exploration of the functional potential encoded in environmental genomes. In addition, Creer et al. (2016) stated that the progression from functional expression screening effortlessly changed into the era of direct random shotgun sequencing of environmental DNA, a methodological leap that highlighted the commitment to capturing the entirety of genetic information within complex microbial communities. This transition was not merely a technical shift but signified a profound conceptual departure from traditional genomics approaches.

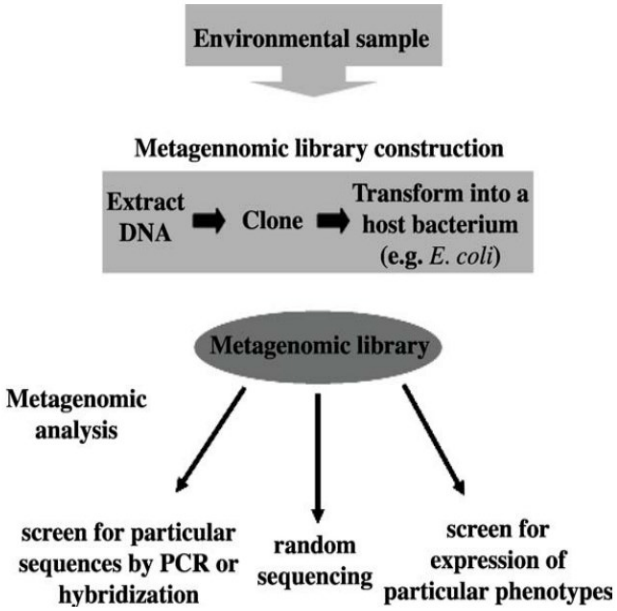


Figure 1.4: Analysis of metagenomics from an environment's microbial population (Adapted from Riesenfeld, Schloss and Handelsman, 2004)

The foundational principles of metagenomics, as highlighted by Garza and Dutilh (2015), challenge the conventional need for the isolation and cultivation of individual microbial species. This departure from isolation-centric methodologies has broad implications, liberating researchers from the constraints imposed by culture-based techniques and enabling the exploration of microbial diversity that was once inaccessible (Pascoal, Costa and Magalhães (2020); Garza and Dutilh, 2015). Zhang et al. (2021) noted that a pivotal aspect of metagenomics lies in its comprehensive nature, encapsulating the collective genomes of all microorganisms within an environmental sample. This holistic approach, acknowledges the interconnectedness and interdependence of microbial communities, emphasizing the need for a collective understanding rather than a reductionist focus on isolated entities.

According to various studies, microbial communities represent a diverse and complex tapestry of microscopic life inhabiting various environments (Gupta, Gupta and Singh, 2016; Prasad et al., 2023). Gupta, Gupta and Singh (2016) added that these communities incorporate bacteria, viruses, fungi, and archaea, creating intricate ecosystems with critical consequences for nutrition cycles, environmental stability, and human health. The abundance and diversity of microbial life play pivotal roles in shaping these ecosystems. In addition, Fadiji and Babalola (2020) deduced that the study of microbial communities via

metagenomics involves the analysis of genetic material that is directly extracted from samples in the environment, and this gives insight into the genomes of these complex ecosystems.

Pérez-Cobas, Gomez-Valero and Buchrieser (2020) stated that metagenomics involves the analysis of genetic material directly extracted from environmental samples, offering a glimpse into the genomes of these complex ecosystems. Contrary to traditional genomics, which focuses on individual organisms and the isolation of a single species' genome, metagenomics captures the collective DNA of multiple species in a given environment (Robbins, Krishtalka and Wooley, 2016). Furthermore, Rashid and Stingl (2015) pointed out that traditional genomics follows a straightforward approach, extracting DNA from pure cultures. In contrast, metagenomics poses unique challenges due to the extraction of DNA from a diverse mixture of microorganisms, necessitating the use of specialized techniques. This distinction highlights the evolving nature of genomic studies and the increasing complexity involved in unravelling the intricacies of microbial communities.

Pipeline of a metagenomic study

Sample Collection

During the first stages of a metagenomic investigation, researchers collect samples from the intended environment, which might include a variety of habitats such as soil, water, or the human stomach. Numerous microorganisms, such as bacteria, fungi, viruses, and archaea, are present in these samples and are all essential to the operation of the ecosystem (Gilbert et al., 2014). The next step in determining the genetic diversity and functional potential of these microbial communities is to isolate genetic material from them using DNA extraction techniques (Caporaso et al., 2011).

DNA Isolation, PCR and Sequencing

After the isolation of DNA, it is amplified using the polymerase chain reaction (PCR) with primers that are intended to target conserved portions of the 16S ribosomal RNA (rRNA) gene, which is a genetic marker that is found in bacteria and archaea globally. By using this amplification technique, researchers can produce DNA sequences that provide information about the taxonomic makeup of the microbial community in the original sample (Gilbert et al., 2014). Then, significant sequence data is produced using high throughput

sequencing platforms, including Illumina sequencing, allowing for a thorough investigation of microbial diversity and community structure (Caporaso et al., 2011).

Pre-processing and Downstream Analysis

After sequencing is finished, preprocessing procedures are applied to raw sequence data to guarantee data dependability and quality. In order to mitigate any biases and errors induced during the sequencing process, these preprocessing processes include quality filtering and the elimination of sequencing artefacts (Schloss et al., 2011). Afterwards, sequence analysis, taxonomy categorization, and evaluation of microbial diversity and composition are performed using bioinformatics tools and pipelines, including QIIME2 or mothur (Caporaso et al., 2011). To maintain the integrity and reproducibility of their results, researchers follow best practices and established protocols throughout the entire pipeline. Metagenomic studies provide priceless insights into the composition, dynamics, and function of microbial communities in a variety of habitats by employing this methodical approach (Gilbert et al., 2014; Schloss et al., 2011).

Unlocking the Secrets of Microbial Communities

According to Miguel et al. (2019) metagenomics is a revolutionary field at the crossroads of molecular biology and bioinformatics, Miguel et al. (2019) also noted that it acts as the key in unravelling microbial tapestry. From the initial sampling of environmental niches to the intricate analyses of genetic data, metagenomics employs a sophisticated array of methods and techniques that collectively propel us into the heart of microbial mysteries. Denman and McSweeney (2015) noted that metagenomic began with the meticulous collection and processing of environmental samples. Environmental sampling serves as an art, strategically choosing habitats known for their microbial richness (Niegowska et al., 2021). Considerations span spatial diversity within an environment and temporal variations, ensuring a holistic representation. The collection process becomes an intricate dance, capturing the microbial nuances embedded in soil, water, or the human gut. Following this, the extraction of DNA emerges as a vital step, presenting its own set of challenges. The diverse array of microorganisms, each with distinct cell types and sizes, demands adaptable DNA extraction methods (Lasken and McLean, 2014). Traditional techniques such as phenol-chloroform extraction coexist with contemporary commercial kits, all geared towards maximizing DNA yield and purity. This extraction is the gateway, allowing us access to the genetic blueprints of entire microbial communities.

Kumari et al. (2017) stated that within the realm of sequencing technologies lies metagenomics which differentiates itself from Next-Generation Sequencing (NGS) and Third-Generation Sequencing. Kumari et al. (2017) deduced that NGS illustrated by Illumina and 454 sequencing, offers unparalleled high throughput, enabling the parallel sequencing of millions of DNA fragments while Wee et al. (2018) stated that Third-Generation Sequencing on the other hand represented by PacBio and Oxford Nanopore, introduces a revolutionary paradigm with significantly longer reads. The metagenomic voyage, guided by these sequencing technologies, ventures into uncharted territories, promising a deeper understanding of microbial diversity.

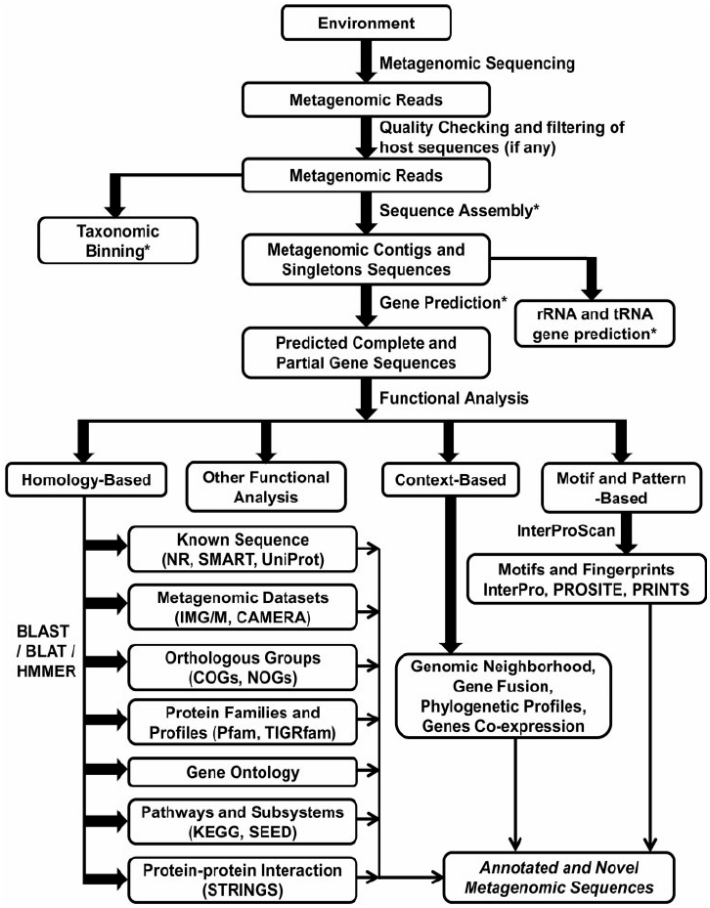


Figure 1.5: Flow chart for the analysis of a metagenome from sequencing to functional annotation (Adapted from Prakash and Taylor, 2012)

Applications of Metagenomics

Metagenomics has altered the knowledge of soil and water ecosystems. Several studies have investigated the soil microbiome by revealing a tapestry that is rich with microorganisms that alter the cycle of nutrients, the health of the soil, and the functionality of the ecosystem (Akhtar, Gulab, and Ghazanfar, 2023; Timmis and Ramos, 2021). Bhargava et al. (2019) noted that metagenomic analysis discloses the range of bacteria, fungi,

viruses, and archaea in the soil, this study also deduced the roles of microorganisms in supporting the growth of the plant and their relationship. Bhargava et al. (2019) added that this has boosted the implication of metagenomics in sustainable agriculture thereby creating a means of attaching microbial communities to improve soil fertility. In addition, Abia et al. (2018) affirmed that water a crucial component of the ecosystem harbors numerous microbial communities essential for preserving the quality of water.

According to Gomaa (2020), the human gut houses numerous microorganisms called the gut microbiomes. Malan-Muller et al. (2018) noted that metagenomics changes knowledge of the world about its effect on metabolism, digestion, and mental health. Insights from gut microbiome studies have implications for personalized medicine, as variations in microbial composition influence individual responses to drugs and susceptibility to diseases (Mesnage et al., 2018; Malan-Muller et al., 2018). In addition, Hellmann et al. (2021) noted that metagenomics extends its reach to the skin, Hellmann et al. (2021) also stated that the microbiome of the skin is vital in sustaining the health of the skin and in preventing infections.

2. RESEARCH AIMS AND OBJECTIVES

The primary objective of this research is to examine the presence and differential abundance of *C. difficile* in the gastrointestinal tracts of both sows and piglets within a sampled pig farm with past cases of *C.difficile* infections and with saws treated via a vaccine against *C. difficile* toxins. Given the association between *C. difficile* and diarrhea in piglets reported by Squire and Riley (2012), this investigation aims to provide insights into the prevalence and distribution of the bacterium within the examined population. By assessing the abundance in both sows and piglets, the research seeks to contribute valuable information regarding the potential sources and transmission dynamics of *C. difficile* on the farm.

A second key goal is to explore the complex interactions between *C. difficile* and other bacteria residing in the bowels of swine. Understanding the synergistic or antagonistic relationships between *C. difficile* and the broader microbial community is crucial for unraveling the intricate dynamics that influence the prevalence and impact of *C. difficile* infections. This objective aims to enhance our comprehension of how various microorganisms interact within the swine gut, shedding light on potential factors influencing the persistence or mitigation of *C. difficile*.

Additionally, this research aims to conduct a comparative longitudinal analysis of the microbial composition in piglets and sows at distinct time points post-birth: 24, 96, and 240 hours. This temporal exploration seeks to uncover potential shifts in the microbial communities of piglets and sows during crucial developmental phases. By investigating these specific time points, the research aims to discern patterns in the microbial composition, providing valuable insights into the dynamic evolution of the gut microbiota within the early postnatal period.

3. ANALYSIS PIPELINE

Data acquisition

Sampling

Sampling for this study involved a comprehensive investigation across 10 distinct pig families, each comprising a mother sow and three piglets. The families are categorized, considering various parameters encapsulated in the metadata. Sample names were encoded as S{1-10}_P{0-3}_T{0-2}. Here, 'S' denoted the "cage," 'P' represented the piglet number (0 indicating the mother sow), and 'T' signified the time point of the sample collection (0 for approximately 24 hours from birth, 1 for around 96 hours, and 2 for approximately 240 hours).

The numerical variable 'sow' was employed to uniquely identify each family through the mother sow's ID. Families were further divided into three rooms, denoted by the numerical variable 'room,' where piglets were preserved from birth to the weaning stage.

The categorical variable 'sex' delineates the gender of the animals, while 'diarrhea' provided insights into the health status, with options ‘yes’, 'no’, or ‘hematic’. The latter was particularly crucial for piglets with minimal content in their rectums at 24 hours, making the rectal walls slightly more susceptible to swab-induced scratches. The numerical variable 'neigh' served to identify the neighborhood family, essential for understanding shared water bowls between two families.

For the identification of mothers, a Boolean variable 'is_sow' was introduced. Additionally, several numeric variables offered a comprehensive perspective on the reproductive and survival dynamics within each family. 'Gestations' denoted the number of times the sow had been pregnant, 'nest' indicated the total number of piglets born by each sow, 'alive' and 'dead' represented the counts of piglets that survived and perished after birth, 'transferred' captured the number of piglets moved to another sow, 'uw_el' accounted for piglets eliminated from the nest due to underweight issues, and 'weaned' indicated the number of piglets that survived until the weaning stage.

The DNA extraction, amplification, library construction, and sequencing processes were conducted by BMR Genomics, based in Padua, Italy. As part of their commitment to quality, the BMR Genomics laboratory

adheres to a comprehensive Quality Management System compliant with the requirements of the international standard UNI EN ISO 9001:2015.

The DNA extraction process involved the supply of sampling kits, and extraction was performed using the Qiagen method. Subsequent steps included the amplification of DNA using modified primers with universal tails, following a specific protocol. The amplified DNA was purified using magnetic beads and Thermolabile Exonuclease I (NEB). In a second PCR step, Illumina Nextera XT Indexes were bound to the universal tails. The samples were then normalized and multiplexed. The resulting library was uploaded to the Miseq platform for sequencing, employing a 300PE (paired end) strategy. Finally, the quality of the sequencing run was assessed, and the reads were delivered in the fastq format.

Amplification and sequencing

Tailed primers for amplifying the 16S rRNA regions (V3 to V4) were employed, as per the method described by Takahashi *et al.* (2014). The forward primer Pro341F and reverse primer Pro805R were designed with universal tails to facilitate subsequent steps in the sequencing process.

The sequencing of the V3–V4 hypervariable region of the bacterial 16S rRNA gene was conducted using Illumina MiSeq V2 chemistry, generating paired-end reads of 250 base pairs each. Sequencing libraries adhered to the Illumina 16S Metagenomic Sequencing Library Preparation Guide (Part #15044223 Rev. B) and employed the Nextera XT Index Kit. Post-sequencing.

Pre-processing

The pre-processing phase of the data involved the utilization of QIIME2, a robust microbiome data science platform. This platform is recognized for its reproducibility, interactivity, scalability, and extensibility in microbiome data analysis. This comprehensive data processing pipeline aligns with the principles of QIIME2 and relevant literature (Bolyen *et al.,* 2019; Martin, 2011), ensuring accuracy and reliability in subsequent microbiome analyses.

Creation of input data and Importing of Data in QIIME2

Generate Manifest S1_1.The R script R is used to create a manifest file, which is an essential part of the QIIME2 analysis pipeline. This file acts as a reference, providing all the pertinent information regarding the

file paths for each sample's forward and reverse sequencing reads. The sample data is arranged systematically thanks to the tabular format it uses, which includes the columns "sample-id," "forward-absolute-filepath," and "reverse-absolute-filepath." Another R script in the workflow, S1_2_GenerateMetadata.R, is responsible for creating metadata.tsv, a format that is compatible with QIIME2, from the original metadata file. To do this, add a label to the first row of the table for the matching column data type. After that, the metadata.tsv file is a crucial part of QIIME2's import and analysis process. A key component of the QIIME2 analytic pipeline's importation and visualization stages is a bash script called S2_Demux.sh. It enables the conversion of .fastq files into a QIIME2 artifact (.qza data format) by utilizing the qiime tools import mechanism. The manifest.tsv file, which contains file path information, the fastq manifest format (PairedEndFastqManifestPhred33V2), that indicates paired-end reads with a PHRED offset for a quality score of 33, and the type of sample data (SampleData[PairedEndSequencesWithQuality]) are the input parameters. Furthermore, the script uses the metadata plugin to create metadata.qzv file. Next, visualizations are produced using the Demux plugin, which also produces interactive positional quality charts and offers insights into counts per sample for all samples. QIIME2 View is used to interactively explore the data owing to the resulting .qzv files.

Demultiplexing

Demultiplexing is a critical process that separates pooled sequencing data into individual samples, ensuring accurate downstream analyses. The demultiplexing method employed in this study aligns with the established framework described by Bolyen *et al.* in their comprehensive work on QIIME2 (2019). Their approach provides a standardized and reliable means of handling microbiome data, enhancing the accuracy and reproducibility of subsequent analyses.

Demultiplexed Sequence Counts Summary is reported in Table 1.0, while a boxplot overview of the sequencing per base position quality check is reported in Figure 1.4 for forward and reverse primers (panels A. and B. respectively)

	forward reads	reverse reads
Minimum	45991	45991
Median	71013.5	71013.5
Mean	72808.958333	72808.958333
Maximum	137157	137157
Total	8737075	8737075

Table 3.1 Demultiplexed sequence count summary.

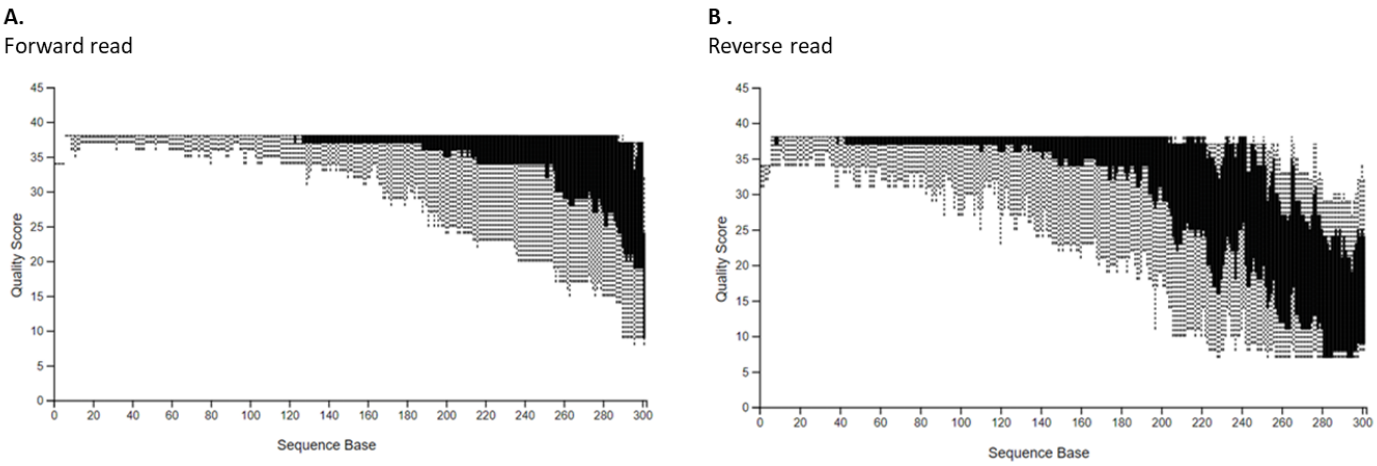


Figure 3.1: Interactive quality plot of the forward (panel A.) and reverse (panel B.) sequence base.

The plot was generated using a random sampling of 10000 out of 8737075 sequences without replacement. The minimum sequence length identified during subsampling was 301 bases. Outlier quality scores are not shown in box plots for clarity.

These plots were generated using a random sampling of 10000 out of 8737075 sequences without replacement. The minimum sequence length identified during subsampling was 301 bases. Outlier quality scores are not shown in box plots for clarity.

Trimming (Primers and adapters removal)

S3_TrimmingPrimer.sh is a bash script that uses the cutadapt plugin to remove primers from files. Additionally, a corresponding .qzv file was developed, mirroring the method used in the previous stage, to aid in data interpretation and produce high-quality charts for visualization on the Qiime2 viewer. Using the -p-front-f option for forward reads and the -p-front-r parameter for reverse reads, the primer sequences that are specifically targeted for removal are defined. It is important to note that the primer sequences contain

wildcard characters that follow the IUPAC nucleotide codes, including N, V, H, and W. The --p-match-read-wildcards option is mentioned to guarantee that certain characters are interpreted correctly during matching. Stressing a cautious cutting strategy, the script imposes stringent requirements: During primer matching, no base insertions or deletions are permitted. reads without recognized primer sequences are rejected (--p-discard-untrimmed), a sufficient overlap of nucleotides in the primer sequence is required (--p-overlap), and a zero-error rate in alignment is maintained (--p-error-rate). Furthermore, reads that have a length of less than 200 nucleotides are not included (--p-minimum-length).

The trimmed Sequence Counts Summary is reported in Table 1.1. The resulting quality check is reported in figure 3.2, with results for forward primer sequencing in panel A. and for reverse primer in panel B.

	forward reads	reverse reads
Minimum	39756	39756
Median	61078.5	61078.5
Mean	62355.158333	62355.158333
Maximum	107687	107687
Total	7482619	7482619

Table 3.2 Summarized trimmed sequence counts.

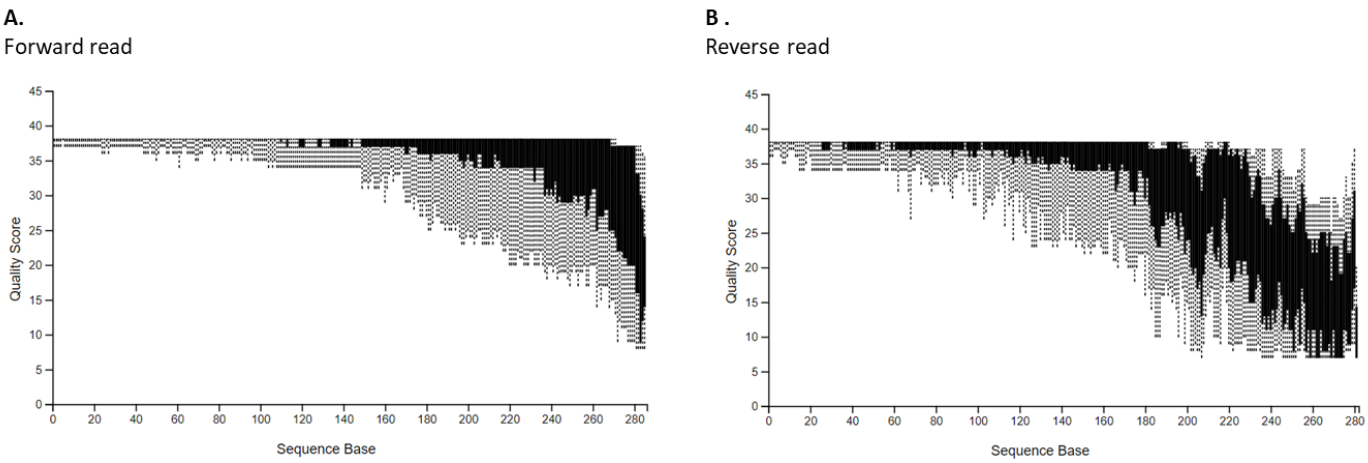


Figure 3.2 Trimmed sequence of the forward (panel A.) and reverse (panel B.) sequence base.

The plot at position 242 was generated using a random sampling of 10000 out of 7482619 sequences without replacement. The minimum sequence length identified during subsampling was 281 bases. Outlier quality scores are not shown in box plots for clarity.

These plots were generated using a random sampling of 10000 out of 7482619 sequences without replacement. The minimum sequence length identified during subsampling was 276 bases. Outlier quality scores are not shown in box plots for clarity.

Denoising and Imputation

The fourth step encompassed denoising and quality filtering using the bash file S4_Denoising.sh. DADA2, a robust algorithm for Amplicon Sequence Variant (ASV) reconstruction, was utilized. Default DADA2 plugin parameters were employed, with truncation lengths based on quality plots. Outputs included ASV table, ASV sequences, and denoising statistics (feature_table.qza, feature_sequences.qza, denoising_stats.qza). Imputation was performed using the mbimpute method and R script imputation.R, resulting in feature_table_imp.qza. Visualization artifacts (feature_table.qzv and denoising_stats.qzv) were also generated.

Imputation addresses the prevalence of zero counts in microbiome datasets which poses a challenge for meaningful downstream analyses and result interpretation. In response to this challenge, Jiang et al. (2020) introduced the mbimpute method, a specialized imputation technique tailored for microbiome data. Addressing the issue of false zeros, mbimpute strategically incorporates information from akin samples, taxa with similarities, and, optionally, covariates and taxon phylogeny. Demonstrated effectiveness in mitigating spurious zeros enhances the accuracy of subsequent analyses, including differential abundance testing and community ordination.

Taxonomy Classification

In the process of taxonomy classification (S5_TaxonomyClassification.sh), a bash script was utilized to assign each Amplicon Sequence Variant (ASV) to its corresponding bacterial taxonomy. The ASVs were compared against the GreenGenes reference database, and a naive Bayes classifier was employed, which had been trained on reference sequences clustered at 99% similarity (gg_13_8_otus/rep_set/99_otus.fasta) along

with their respective taxonomies (gg_13_8_otus/taxonomy/99_otu_taxonomy.txt). This reference data was sourced from the GreenGenes database v13.8 ftp://greengenes_release/gg_13_5/gg_13_8_otus.tar.gz.

The naive Bayes approach was used, which is predicated on the independence of features within the same class and is based on the Bayes theorem. Crucially, the qiime tools import method was used to import the reference sequences and taxonomies into QIIME2. Notably, two distinct semantic types were used: FeatureData[Taxonomy] for classification and FeatureData[Sequence] for sequences. Since the taxonomy containing greengenes file did not have a header, the import was done with the HeaderlessTSVTaxonomyFormat argument. The naive Bayes classifier was trained only on the target region, concentrating on sequences of interest between the primers 314F and 805R, as suggested by Werner et al. To remove sequences that were either too short or too lengthy in comparison to the ASVs, the primer sequences (--p-f-primer and --p-r-primer) as well as the --p-min-length and --p-max-length parameters were taken into consideration while using the feature-classifier extract-reads method. The feature-classifier fit-classifier-naive-bayes approach was used to train the classifier. After using the feature-classifier classify-sklearn to the classifier.qza artifact, the ASVs were classified, yielding the taxonomy table (taxonomy.qza). In the end, a tabular version of the taxonomy table was produced using the metadata plugin.

Phylogenetic Tree

The bash script S6_PhylogeneticTree.sh is dedicated to constructing a phylogenetic tree for ASVs, offering valuable insights for diversity analysis. This tree elucidates kinship relationships based on sequence similarities. Two primary computational methods for phylogenetic tree reconstruction exist: de novo or by integrating representative sequences into a reference tree. In this case, we employed a hybrid approach facilitated by the SEPP (SATé-Enabled Phylogenetic Placement) tool. SATé (Simultaneous Alignment and Tree Estimation) addresses complex alignments where accurate sequence alignment is challenging without the concurrent availability of the tree.

The SEPP algorithm iteratively optimizes both sequence alignment estimates and tree structures. For ASVs absent from the reference tree, a similar iteration determines their intermediate positions in successive sub-trees, converging to a definitive positioning. The Qiime 2 implementation utilizes the fragment-insertion sepp method, with sepp-refs-gg-13-8.qza containing the reference tree skeleton. This artifact is built from

GreenGenes sequences clustered at 99% similarity. The resulting tree.qza is crucial for diversity analysis. Additionally, the tree_placements.qza file details the intermediate positions of each ASV within the algorithm.

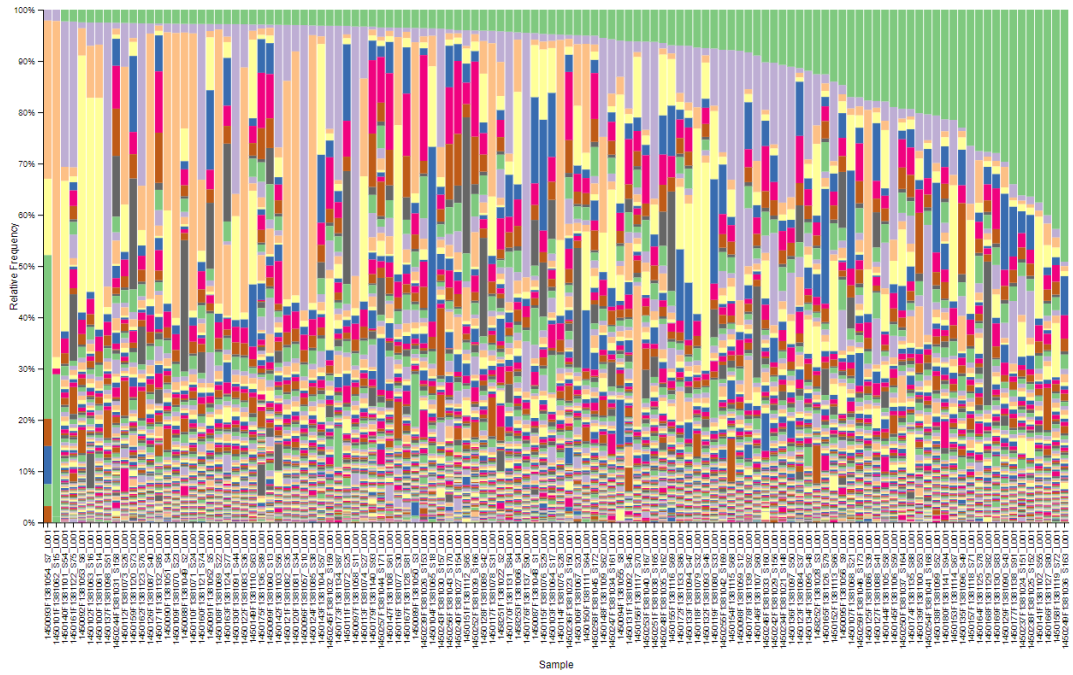


Figure 3.3 Bar plot showing the frequency of the different taxonomic features in each sample.

Taxonomic classification and data normalization.

The bash script S7_CollapseFilterNormalize.sh orchestrates several crucial steps in the analysis pipeline: collapsing ASVs at specified taxonomic levels, eliminating zero features, normalizing sample profiles, and importing the normalized data into Qiime2. The taxonomic levels of species and genus were chosen in addition to ASVs for comprehensive analysis. The taxa collapse plugin is employed to group features at the genus and species levels, determined by the --p-level parameter (set as 7 for species and 6 for genus).

For each taxonomic level, subject profiles undergo normalization using the GMPR method, based on the Geometric Mean of Pairwise Ratios. GMPR resolves challenges posed by zero frequencies, common in metagenomic data, by computing the median of non-zero cell ratios for each subject pair. The size factor (SF), representing the relative library size, is obtained as the geometric mean of these ratios. Subsequently, subject taxonomic profiles are scaled based on their corresponding SF.

Internally, the script calls the GMPRnorm.R script, utilizing the qiime2R package for abundance table reading, GMPR normalization in R, and generation of the .biom file for normalized data. This .biom file is then converted back to .qza using qiime tools import.

GMPR Normalization in Microbiome Sequencing Data Analysis

Normalization is a pivotal step in microbiome sequencing data analysis, crucial for addressing the inherent variability in library sizes across samples. Traditional RNA-Seq normalization methods face challenges when applied to microbiome data, characterized by a multitude of zeros. The high frequency of zeros makes these methods prone to instability. In response to this issue, the Geometric Mean of Pairwise Ratios (GMPR) normalization method has been proposed, offering a robust solution.

The GMPR method involves a two-step process, introducing a novel approach to normalization inspired by the DESeq2 normalization steps. First, the method calculates rij, the median count ratio of nonzero counts between samples, using the formula:

$$rij = \text{median}(ckj/cki)$$

Here, k ranges from 1 to the total number of Operational Taxonomic Units (OTUs), and c_{ki} and c_{kj} represent the non-zero counts of the kth OTU in samples i and j, respectively.

Subsequently, the size factor (si) for a given sample i is calculated using the geometric mean of the previously obtained rij values:

$$si = \text{geometric\_mean}(rij)$$

The resulting vector of GMPR size factors provides an effective means of normalizing microbiome data, accounting for library size variations while mitigating the challenges posed by excessive zeros. This innovative approach ensures stability and reliability in downstream analyses, making GMPR a valuable tool in microbiome research (Li Chen *et al.,* 2018).

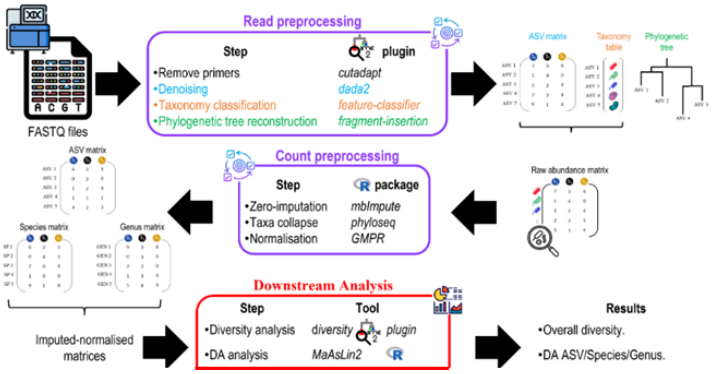


Figure 3.4 Overview of Bioinformatic Pipeline (Adapted from Bellato 2023)

Post-processing.

Post-processing analysis, which consists of diversity analysis (both Alpha and Beta diversity), as well as differential abundance analysis, was conducted utilizing the R programming language.

Diversity analysis

Alpha Analysis

Alpha diversity is a vital analysis, that offers insights into the diversity within individual samples considering factors like evenness and richness (Magurran, A. E. 2004). The flexible R package vegan, which offers a comprehensive range of tools for ecological and diversity analyses, was used in this study to perform alpha diversity analysis. The differences in alpha diversity between samples were statistically evaluated using the Kruskal-Wallis test. Boxplots were employed for visualization. (Oksanen et al., 2007). vegan: Community Ecology Package. R package version 2.5-6. https://CRAN.R-project.org/package=vegan)

A wide range of measures were employed in the alpha diversity analysis across all the genus, species, and ASV taxonomic levels to capture various aspects of the organization of the microbial community. Shannon's diversity, Simpson's diversity, Richness, Pielou's evenness, and Faith's phylogenetic diversity were all the metrics considered. Faith's Phylogenetic Diversity metric was used only at the ASV (amplicon sequence variant) level, which offers a more nuanced view of the evolutionary relationships between microbial taxa.

Shannon Index (H)

The Shannon index measures a community's evenness and richness. It offers a more thorough measurement of variety by accounting for the abundance of every species. A community with a higher Shannon diversity

has a greater number of species and a more uniform distribution of individuals within those species (Magurran,1988).

$$H = -\sum_{i=1}^{S} p_i \cdot \ln(p_i)$$

Simpson Index (D)

Dominance in a community is emphasized by the Simpson Index. It calculates the likelihood that two randomly chosen members of the sample are members of the same species. A more diversified group is shown by higher Simpson diversity, which additionally indicates less influence by a single species (Simpson 1949).

$$D = \frac{1}{\sum_{i=1}^{S} p_i^2} \text{ (Alternatively, } D = 1 - \sum_{i=1}^{S} p_i^2)$$

Faith's Phylogenetic Diversity (PD)

Faith's PD considers the links between species through evolution. It calculates the phylogenetic tree's total branch length, which links every species within a community.

Greater species richness and the distinct evolutionary history of the community are both indicated by higher PD, which denotes greater evolutionary diversity (Faith,1992).

$$PD = \sum_{i=1}^{S} d_i,$$

where di is the branch length connecting species i to its nearest ancestor.

Pielou's Evenness (J)

Pielou's evenness (J index) quantifies the degree of dispersion of individuals within a species within a community. It shows how fairly the distribution of species abundance is distributed. Low evenness denotes domination by a small number of species, and high Pielou's evenness shows a more even distribution of individuals among species (Pielou 1966).

$$J = \frac{H}{\log(S)},$$

where H is the Shannon diversity index and S is the species richness

Observed Features (Richness)

To put it simply, richness is the total number of distinct species or traits that exist in a community. More species are indicated by higher richness, which reflects the diversity of organisms within a community (Hurlbert 1971).

S, the total count of observed species/features in a sample

Kruskal-Wallis Statistical Test

The significance of variations in alpha diversity indices: Shannon, Simpson, Pielou's evenness, Richness, and Faith phylogenetic diversity, among different samples and groups was evaluated using the Kruskal-Wallis statistical test. This non-parametric test applies to this specific kind of data and was used at the three taxonomic levels considered as it compares multiple independent groups. The p-values derived from the Kruskal-Wallis test information regarding the presence of statistically significant differences in diversity metrics between the groups. In terms of alpha diversity, a low p-value suggests that at least one group differs from the others significantly. Understanding the differences in species richness, evenness, and phylogenetic diversity is essential for clarifying microbial communities' dynamics and ecological structure in each sample among different covariates. The statistical significance was set at a conventional threshold of $p < 0.05$ (Kruskal and Wallis 1952).

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1)$$

,

N is the total number of observations across all groups,

K is the number of groups

Ri is the sum of ranks for group *i*

ni is the number of observations in group *i*

The test statistic *H* follows a chi-squared distribution with *K - 1* degrees of freedom. The null hypothesis assumes that the populations from which the samples are drawn have identical distributions, and a low p-value indicates rejection of this hypothesis, suggesting at least one group significantly differs from the others.

Beta Analysis

Beta diversity analysis provides insights into the diversity between different samples, emphasizing the compositional dissimilarity across microbial communities (Lozupone and Knight 2005). In this study, beta diversity was assessed using the versatile R package vegan (Oksanen et al., 2007), specifically designed for ecological and diversity analyses. The analysis involved the calculation of dissimilarity matrices using Bray-Curtis, and Jaccard metrics. These matrices were further visualized using principal coordinates analysis (PCoA) to represent the relationships between samples in a low-dimensional space.

Principal Coordinate Analysis (PCoA) (classical scaling) on a given distance matrix D can be computed more easily with the help of the function pcoa, which is used in ecological and diversity investigations. Pairwise distances between samples are converted into a collection of orthogonal axes, or primary coordinates, using the PCoA technique, which reveals the underlying structure of the data. Gower first presented this method in 1966, and it provides a geometrically intuitive depiction of the data's dissimilarity patterns. The relationships between samples based on their dissimilarity profile can be seen and understood according to the computed main coordinates. The pcoa function, in particular, applies two corrective techniques to solve issues related to negative eigenvalues in the analysis, improving the robustness and dependability of the findings.

Bray Curtis Dissimilarity Matrix

In ecology, the Bray-Curtis dissimilarity is a commonly employed metric that measures the compositional dissimilarity of two distinct locations or samples by considering the relative abundances of several species. Because it considers both the existence and abundance of species, this dissimilarity index is appropriate for research on community ecology. Whereas a larger number denotes greater dissimilarity between samples, a lower Bray-Curtis dissimilarity shows greater similarity between samples (Bray and Curtis 1957).

The mathematical formula for Bray- Curtis dissimilarity is given by:

$$1 - \frac{2c}{(a+b)},$$

Where a and b represent the total abundances of species in the two samples, and c represents the sum of the minimum abundances for each shared species.

Jaccard Dissimilarity Matrix

A typical metric in ecological and biological investigations to measure the degree of dissimilarity between two sets, highlighting the presence or lack of shared features, is the Jaccard dissimilarity matrix. Evaluating compositional changes between samples is particularly useful in community ecology. By dividing the total number of unique elements in each set by the total number of elements in both sets together, the Jaccard dissimilarity index is computed (Magurran 2004).

Mathematically, it can be expressed as:

$$J(A, B) = |A \cap B| / |A \cup B|,$$

where A and B represent two sets being compared. A Jaccard dissimilarity value of 0 indicates complete similarity, while a value of 1 signifies complete dissimilarity.

Differential Abundance Analysis

MaAsLin2 package on the R environment was employed for this analysis. MaAsLin2 is a technique intended to find relationships in large-scale population research between complicated metadata and microbiome meta-omics traits. To customize the analysis to the particulars of your study, this software provides a range of analysis techniques, including filtering, normalization, and transform options. It also supports numerous variables and repeated measures. Input data and metadata files in a tab-delimited format are needed to use MaAsLin2. After that, the output is sent to a specified folder. The criteria, which include minimum abundance, prevalence, and variance, can be customized by users to meet the specific needs of their studies. The software uses the q-value, or significance threshold, to find relationships. The correction method used for the DA analysis is the Benjamini-Hochberg method.

A key feature of MaAsLin2 is its ability to handle the normalization and transformation of data, allowing users to choose the most suitable methods for their datasets. Additionally, the tool supports various analysis methods and provides options for both fixed and random effects in the model. Corrections for computing q-values, such as standardization and z-score application, further enhance the robustness of the analysis. The software also offers visualization options, including heatmaps and scatter plots, to aid in interpreting the significant associations.

Notably, MaAsLin2 supports parallel processing with the option to specify the number of R processes to run concurrently. Users can choose to save full model outputs and customize the reference factor for variables with more than two levels. In summary, MaAsLin2 provides a user-friendly platform for comprehensive association analysis in microbiome studies, offering flexibility and customization to meet the unique needs of different research projects (Mallick *et al.,*2021).

Benjamini-Hochberg (BH) correction method

The Benjamini-Hochberg (BH) correction method is a widely utilized statistical technique for controlling the false discovery rate (FDR) in multiple hypothesis testing within scientific research. This method aims to mitigate the risk of erroneously identifying statistically significant results when conducting numerous comparisons simultaneously. By adjusting p-values, the BH correction strikes a balance between identifying true positives and minimizing false positives. BH correction method offers practicality and it enhances the reliability of findings in studies with multiple comparisons Benjamini, Y., & Hochberg, Y. (1995).

4. RESULTS AND DISCUSSION

Data Overview

In this study, a total of 40 pigs were considered, consisting of 10 sows and 30 piglets. The piglets were organized into families, each comprising one mother and three piglets. Among the piglets, there were 16 males and 14 females. All 40 animals underwent testing at three different time points: 24 hours, 96 hours, and 240 hours. For piglets, the testing times were after birth. This resulted in a total of 120 samples collected. All piglet mothers were negative for diarrhea however, a small subset of piglets (5 females and 1 male) tested positive for diarrhea. Notably, all four positive females exhibited signs of diarrhea at the 96-hour mark, with no overlap of positive piglets between different testing times. Moreover, at 96 hours, all male piglets were negative for diarrhea, and the sole positive male was identified at the 240-hour mark. All piglets showed signs of hematic content at the 24-hour time point. It is not surprising for piglets to exhibit minimal rectal content during this early stage, reflecting a physiological characteristic typical of their developmental phase shortly after birth.

				Time		
Covariate	**Levels**	**Gender**	**#**	**24**	**96**	**240**
Diarrhea	yes	Male	1	0	0	1
		Female	4	0	4	0
	no	Male	31	0	16	15
		Female	54	10	20	24
	hematic	Male	16	16	0	0
		Female	14	14	0	0

Table 4.1 General distribution of data.

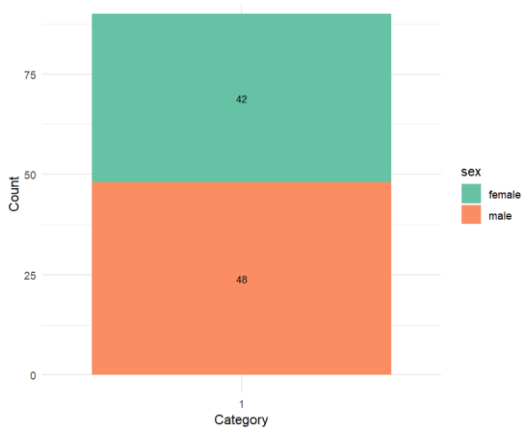


Figure 4.1 shows the distribution of samples based on the sex of the piglets (without sows).

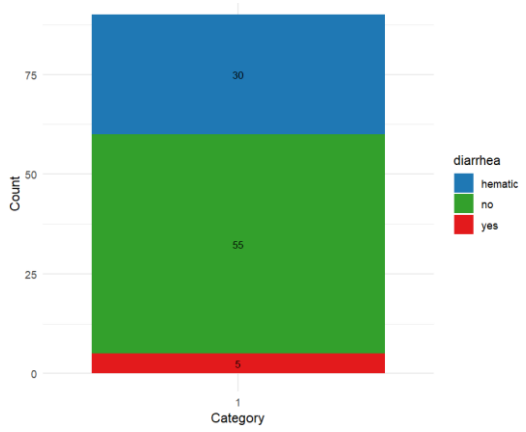


Figure 4.2 shows the distribution of samples collected based on the diarrheic status of the piglets (without sows).

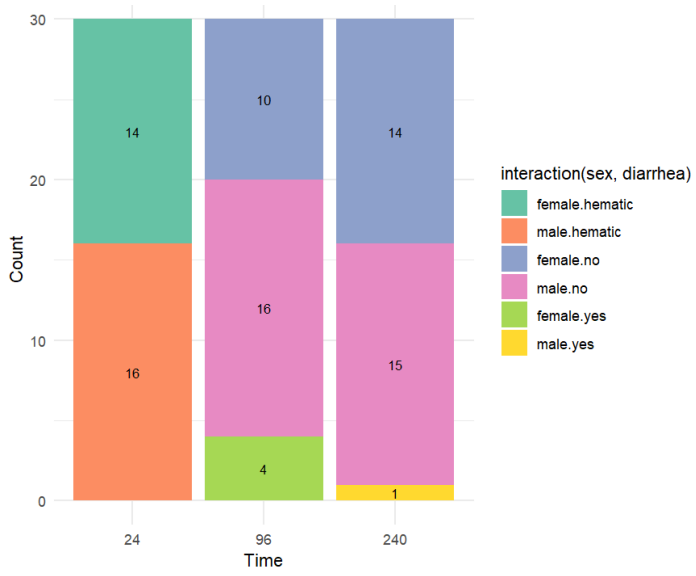


Figure 4.3 Distribution of samples collected based on the diarrheic status and the sex of the piglets (without sows) with time.

Diversity analysis

Alpha Diversity Analysis

The alpha diversity analysis revealed that diarrhea is significant across at the genus, species and ASV taxonomic levels for most of the diversity indices. This indicates that the microbial diversity in the gut microbiota of diarrheal piglets differs from that of healthy piglets, suggesting changes in both composition and richness.

All taxonomic levels and diversity indices show that sex is a significant factor, except Pielou evenness at the species level. There are clear differences between the microbial composition and diversity of males and females, and the lack of significance for Pielou's evenness at the species level indicates that the evenness of the microbial communities is similar in both genders.

Except for richness at the ASV level, time indicates that there are temporal changes in the structure and diversity of microbial communities, and are significant across all diversity indices and taxonomic levels. This suggests that although total diversity and composition vary over time, the quantity of distinct Amplicon Sequence Variants (ASVs) does not change much. Changes in the distribution and quantity of already-existing ASVs are more closely linked to the temporal dynamics in gut microbiota than are the introductions of completely new microbial variations.

The category, which labels samples as siblings of piglets or as sows, is significant for all taxonomic levels and diversity indices. This highlights the notable distinctions in the gut microbiota between sows and their piglet siblings by emphasizing the significant variations in microbial diversity and composition between sow and piglet samples. Table 1.3 below shows the Kruskal-Wallis’s test done for the alpha diversity indexes on all samples at different taxonomic levels. Figures in red indicate insignificance in the relevant covariates considered.

	Genus Taxonomic Level				Species Taxonomic Level				ASV Taxonomic Level			
Covariate	**Shannon**	**Simpson**	**Pielou**	**Richness**	**Shannon**	**Simpson**	**Pielou**	**Richness**	**Shannon**	**Simpson**	**Pielou**	**Richness**
Diarrhea	2.46E-09	1.78E-08	1.10E-06	3.70E-05	3.48E-09	1.95E-09	3.45E-06	3.64E-05	6.38E-10	4.76E-11	4.28E-07	0.001648769
Sex	0.01923464	0.01923464	0.028693	1.11E-05	0.03170544	0.01763981	0.079826044	6.03E-05	0.02768	0.03086	0.03968258	0.000285828
Time	7.90E-05	1.30E-03	0.000298	0.3030253	2.07E-07	1.45E-06	3.44E-06	0.04811085	4.20E-05	8.55E-08	2.84E-07	0.05761195
Category	1.88E-11	1.28E-12	1.58E-07	6.59E-15	1.99E-07	7.13E-09	1.80E-03	7.43E-12	8.84E-09	1.62E-08	3.38E+05	3.63E-12
Room	0.730067	0.743424	0.2637106	0.6193903	0.7461528	0.8420008	0.2593462	0.7651181	0.72871	0.63355	0.315465	0.7160923
Alive	0.4771554	0.4914044	0.1983525	0.7683149	0.4524807	5.24E-01	0.1562268	0.8734865	0.62881	0.782	0.320942	0.8058703
Transferred	0.2384378	0.2202027	0.1152118	0.567879	0.3325835	0.379665	0.161185	0.4185836	0.43957	0.497739	0.1942135	0.6254032
Dead	0.550232	0.5268012	0.2554651	0.4126006	0.315968	0.3789265	0.102962	0.7188179	0.603145	0.82027	0.2631124	0.4462988
Gestations	0.477352	0.4053442	0.2376488	0.3664579	0.6388908	0.6458074	0.2726073	0.2031348	0.68531	0.70692	0.4244554	0.6007767
Neigh	0.569398	0.5464761	0.2565271	0.6634289	0.7055423	0.7894326	0.3744638	0.565347	0.80301	0.82579	0.4746943	0.8515482
Weaned	0.7877792	0.6594835	0.709047	0.2409145	0.7530594	0.6839456	0.630157	0.2016978	0.7877792	0.6058457	0.5577493	0.7810955

Table 4.2: Kruskal-Wallis's test done for the alpha diversity indexes on all samples at different taxonomic levels.

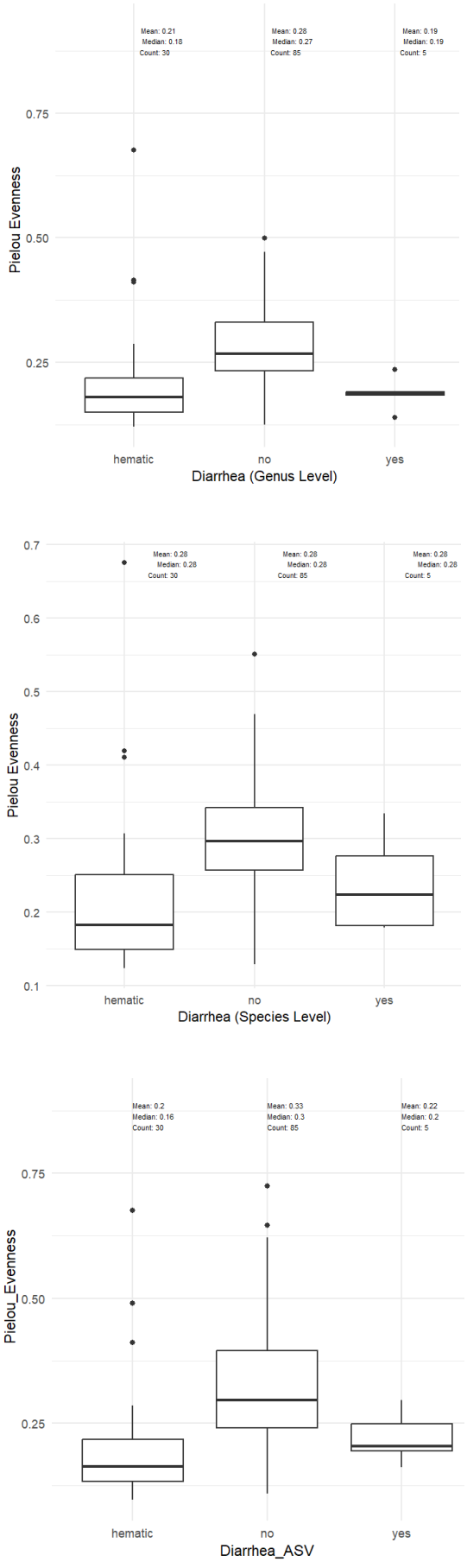


Figure 4.4: Pielou's Evenness box plot for Diarrhea

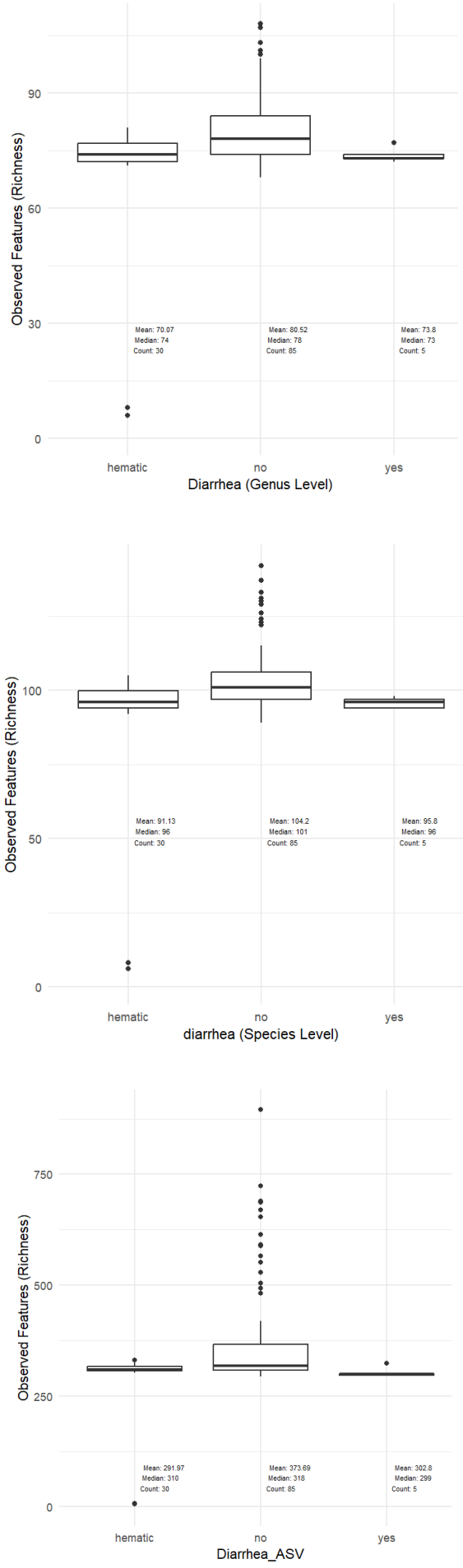


Figure 4.5: Observed features (richness) box plot for Diarrhea.

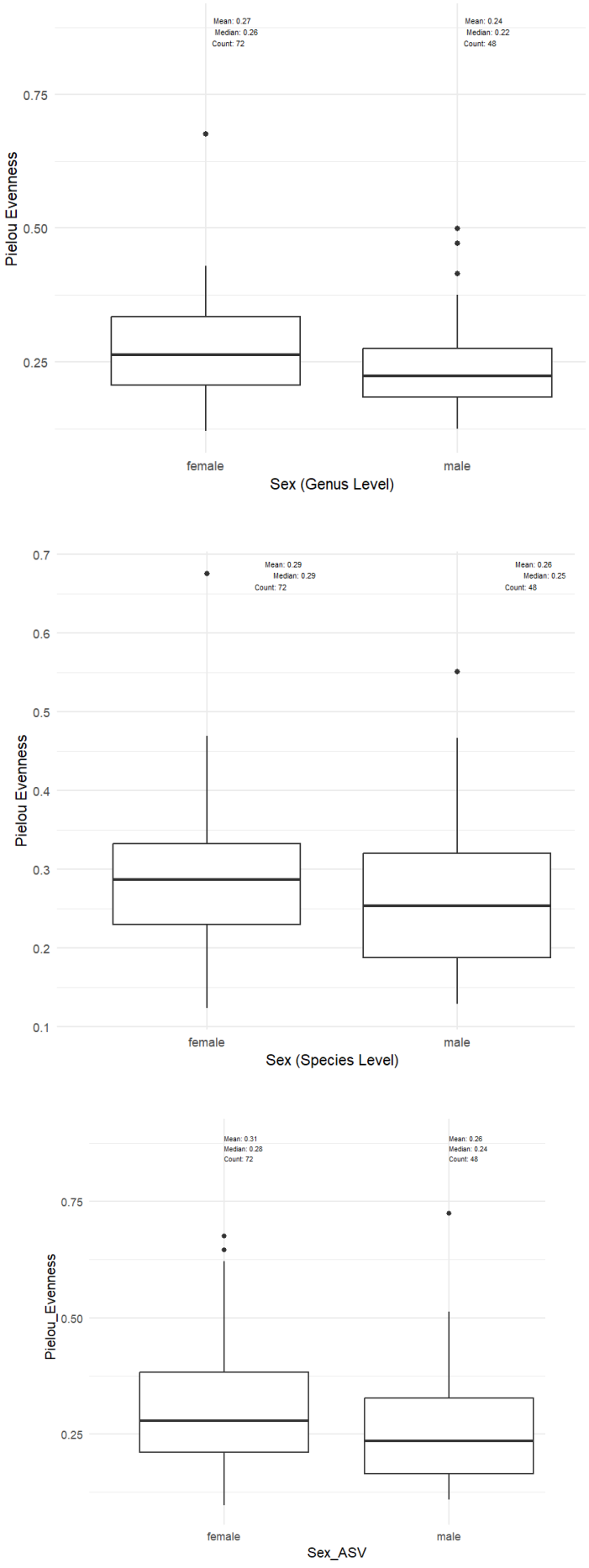
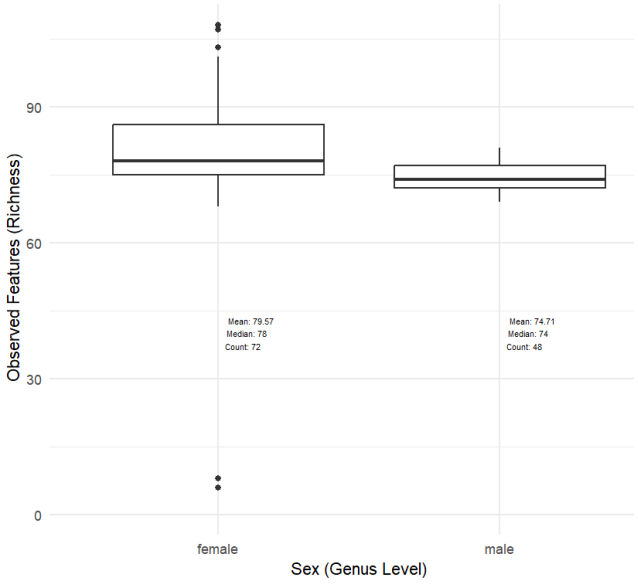
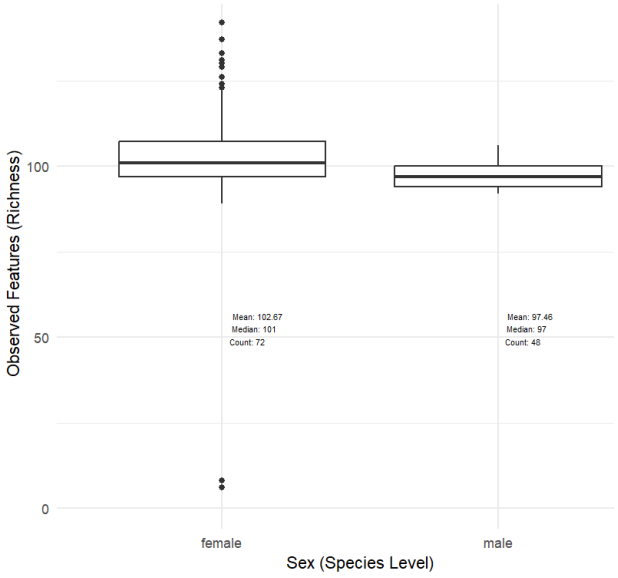


Figure 4.6: Pielou's Evenness box plot for sex





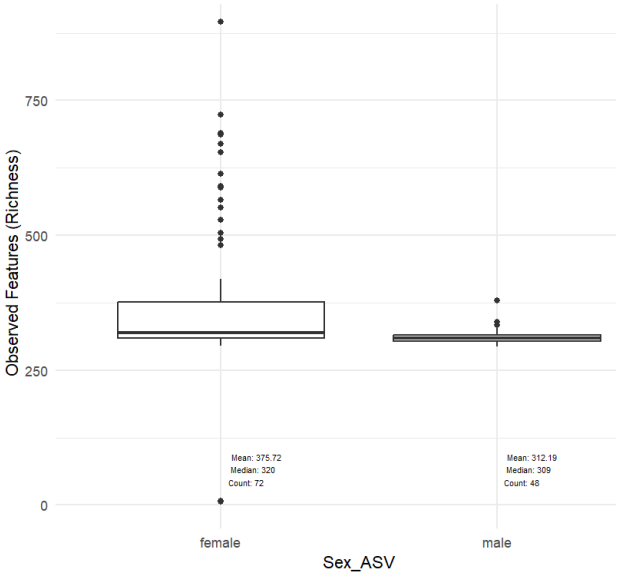


Figure 4.7: Observed features (richness) box plot for sex.

However, it was necessary to re-evaluate the data because sow samples could have had an impact on our conclusions, especially on the results from the sex since all sows are females and there is a huge difference in the microbial composition of sows and piglets. We were able to eliminate any possible confounding effects from sow samples and concentrate exclusively on piglet samples due to this tactical change. The diarrhea variable was also re-grouped into two categories: diarrhoeic and non-diarrheic (combining hematic and no). Redefining the diarrhea variable allowed us to improve the accuracy of our research and make sure that the piglets' circumstances were the main cause of the patterns in microbial diversity and composition that we saw. The objective of identifying and comprehending elements particularly related to piglet health and gut microbiota dynamics is in line with this methodical approach. Table 1.4 shows the result of the Kruskal-Wallis's test done for the alpha diversity indexes on samples from only piglet and diarrhea organized into two levels (diarrheic and non-diarrheic) at different taxonomic levels.

	Genus Taxonomic Level				Species Taxonomic Level				ASV Taxonomic Level				
Covariate	**Shannon**	**Simpson**	**Pielou**	**Richness**	**Shannon**	**Simpson**	**Pielou**	**Richness**	**Faith**	**Shannon**	**Simpson**	**Pielou**	**Richness**
Diarrhea	**0.039885**	**0.038598**	**0.03738**	0.08096	0.250499	0.434559	0.334406	0.054408	0.270932	0.209769	0.29041	0.290417	0.01481
Sex	0.274941	0.190126	0.691886	0.281	0.674076	0.639008	0.839766	0.27179	0.24417	0.502043	0.486716	0.974192	0.670904
Time	5.65E-06	1.29E-04	2.14E-04	1.59E-01	3.64E-07	3.87E-07	2.37E-05	6.28E-01	5.45E-04	9.66E-08	1.17E-08	4.74E-06	5.54E-02
Room	0.760305	0.702833	0.417431	0.494651	0.870581	0.897785	0.610121	8.12E-01	6.65E-01	0.858372	0.795765	0.525538	0.716092
Alive	2.62E-01	1.86E-01	1.12E-01	5.29E-01	4.74E-01	4.48E-01	1.67E-01	0.912514	0.494742	5.85E-01	7.29E-01	3.04E-01	6.88E-01
Transferred	0.112927	0.074382	0.07261	0.015589	0.336126	0.300636	0.01724	3.56E-01	4.16E-01	0.305151	0.432623	0.252826	0.599689
Dead	5.70E-01	3.82E-01	2.00E-01	3.02E-01	3.68E-01	3.62E-01	8.81E-02	6.61E-01	5.18E-01	0.766473	8.70E-01	3.67E-01	4.69E-01
Gestations	0.246909	0.547186	0.205819	0.302688	0.368215	0.368215	0.442813	3.31E-01	7.83E-01	0.369181	0.547186	0.388399	0.451673
Neigh	3.65E-01	2.55E-01	1.97E-01	3.55E-01	6.36E-01	6.36E-01	3.71E-01	2.03E-01	6.74E-01	0.71973	7.72E-01	5.74E-01	6.43E-01
Weaned	0.484812	0.448486	0.611516	0.097115	0.658387	0.581347	0.581347	6.68E-01	4.33E-01	0.59322	0.425178	0.633696	0.6831

Table 4.3: Kruskal-Wallis's test done for the alpha diversity indexes on samples from only piglets(without sows) and diarrhea organized into two levels (diarrheic and non-diarrheic) at different taxonomic levels. Figures in red indicate insignificance in the covariate focused on (diarrhea and sex in this case).

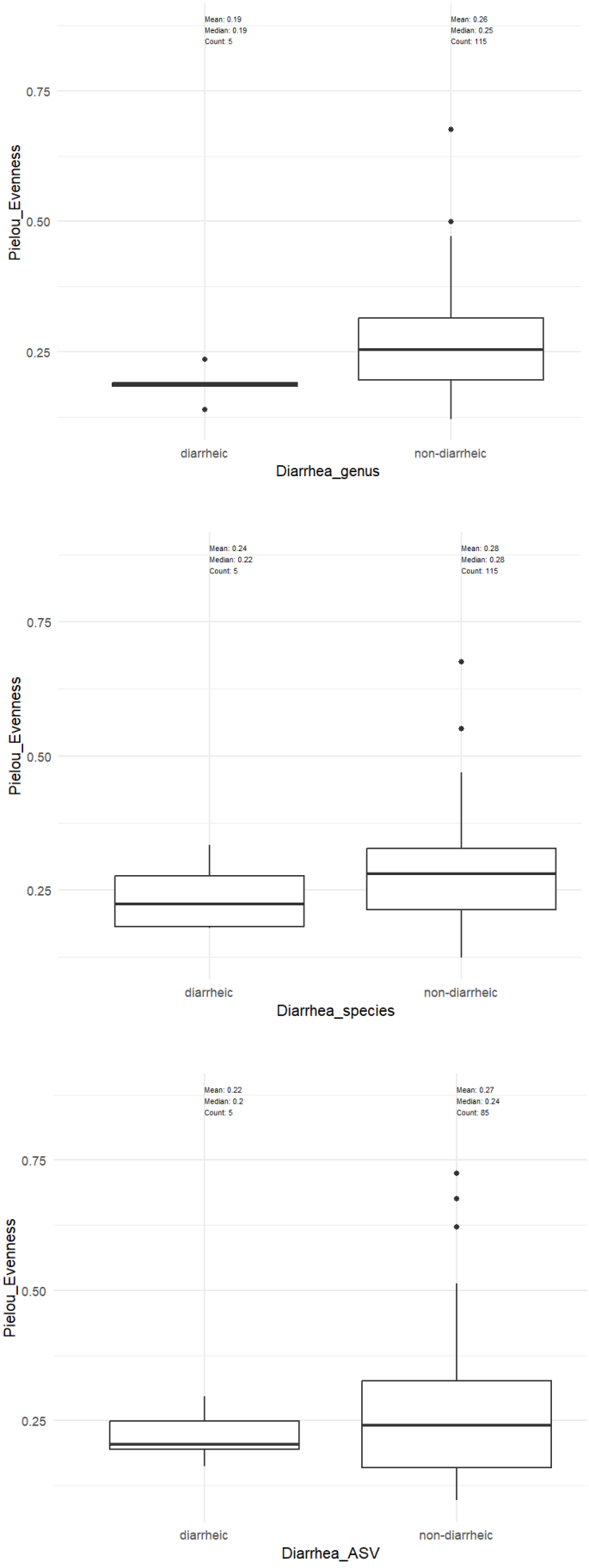
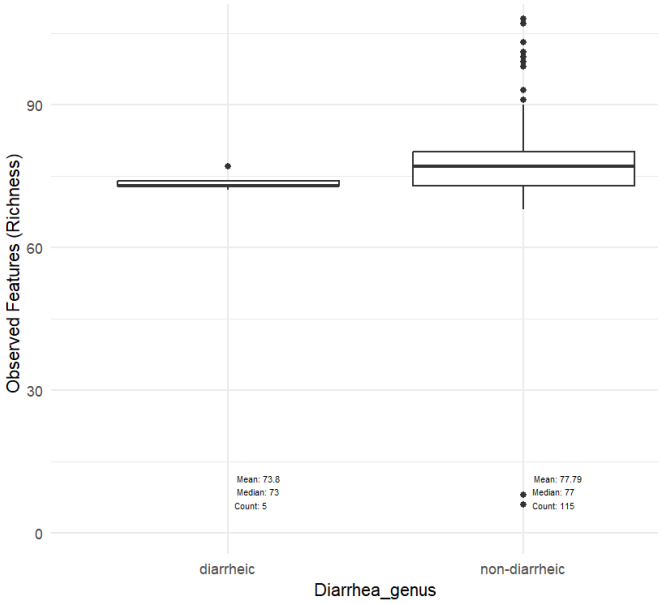
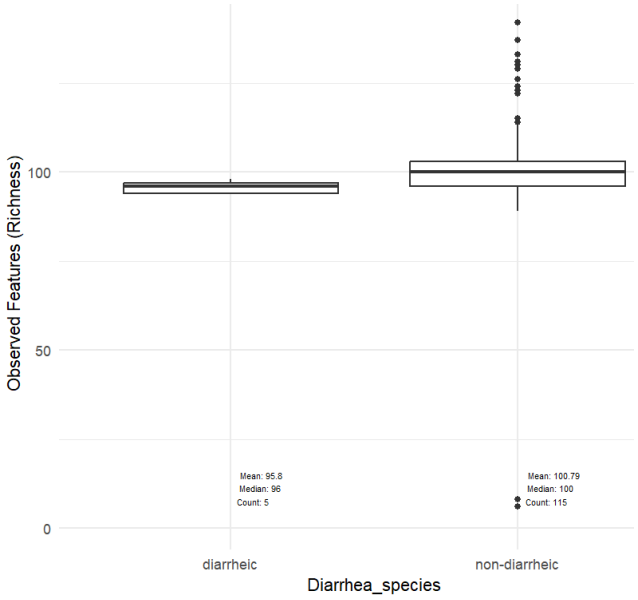


Figure 4.8: Pielou's Evenness box plot for diarrhea after considering hematic samples as non-diarrheic.





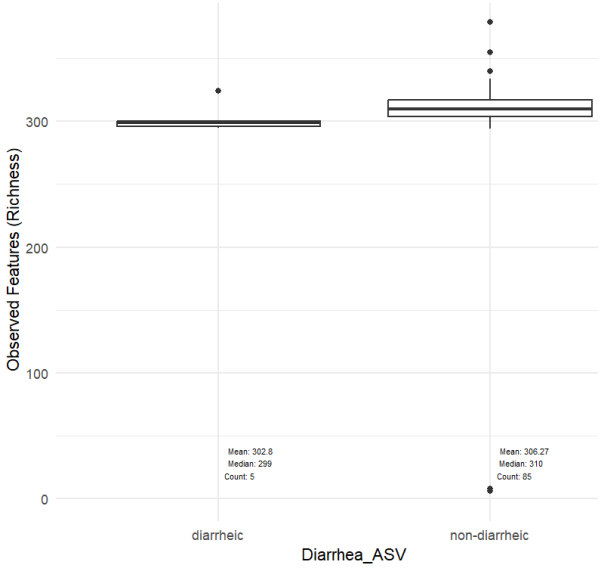


Figure 4.9: Observed features (richness) box plot for diarrhea after considering hematic samples as non-diarrheic.

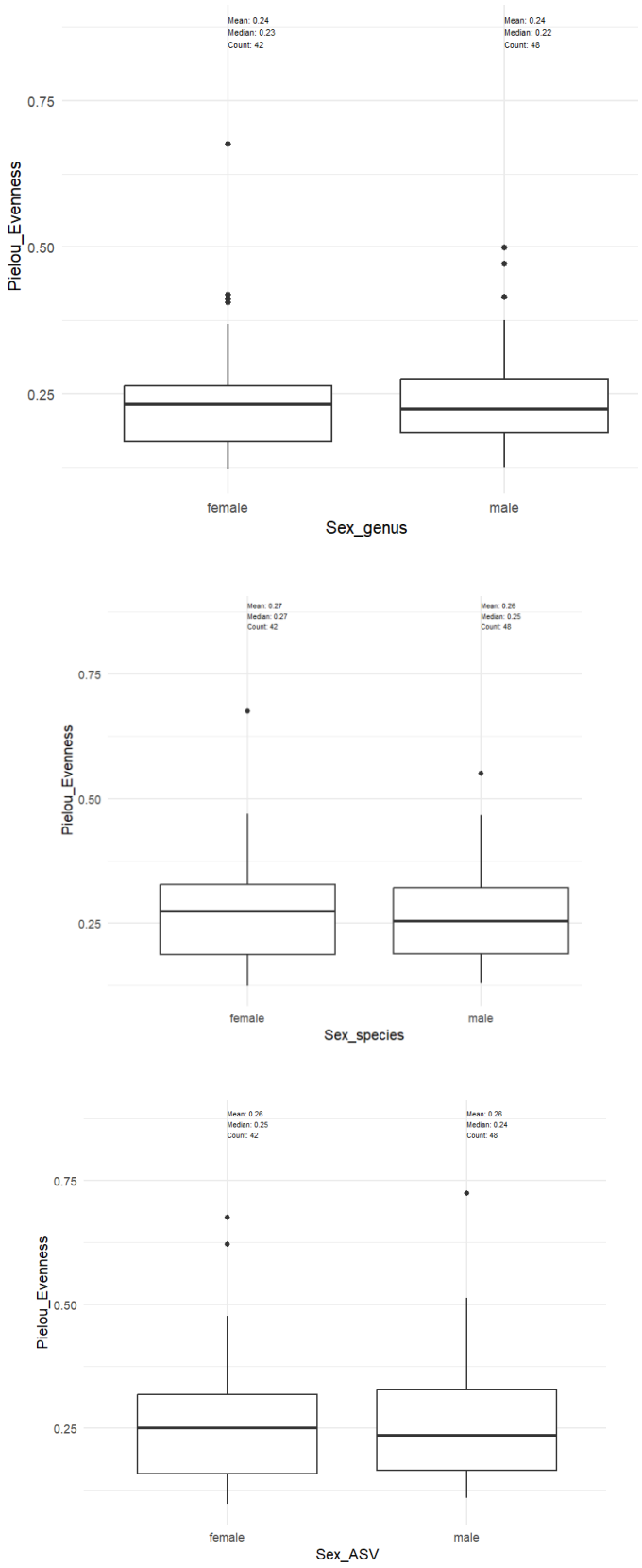


Figure 4.10: Pielou's Evenness box plot for samples from piglets only

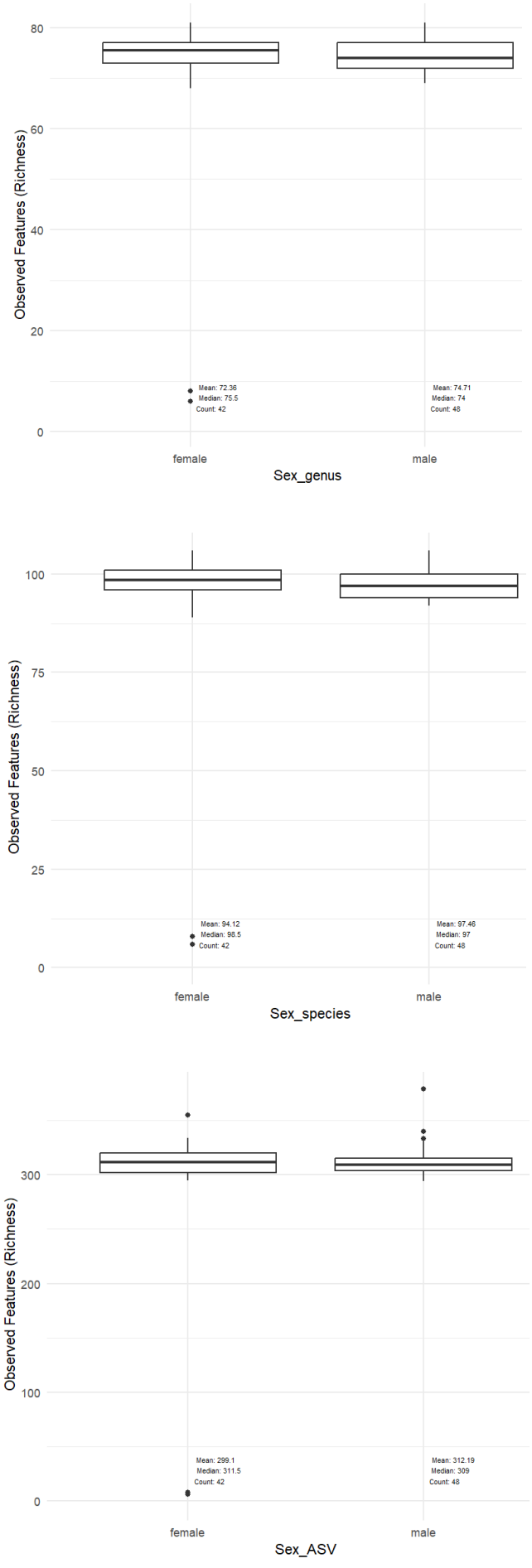


Figure 4.11: Observed features (richness) box plot for sex considering samples from Piglets only.

Diarrhea shows significance at the genus level across all diversity indexes, except for richness. Diarrhea was insignificant at the species and ASV taxonomic levels. This persistence at the genus level implies a consistent divergence in the gut microbiota of piglets with diarrhea compared to those without diarrhea at the genus level.

The variable of sex has become statistically insignificant across all taxonomic levels and diversity indexes. This alteration underscores the influence of sow samples on the initial analysis, revealing that the microbial differences between male and female piglets lose statistical significance when sows' samples are excluded. This adjustment emphasizes that the gut microbiota of sows significantly differs from that of piglets, impacting the initial association observed between sex and microbial composition.

Time, and diarrhea (only at the genus level) were found to be significant based on the results of the alpha diversity analysis. The impact of these two factors on changes in the microbial community was then the particular focus of the beta diversity analysis.

Beta Diversity Analysis

Two-dimensional PCoA visualizations was plotted employing Bray-Curtis and Jacccard distances. The result shows how time and diarrhea affect the microbial dynamics in the swine. The resulting plots are shown in Figures 4.12 and 4.13. These graphics provide a thorough understanding of the complex interactions between microbial communities by employing Bray Curtis and Jaccard dissimilarity metrics, respectively. The results obtained from these two dissimilarity matrices showed no significant variations.

At the considered taxonomic levels, it is evident that a microbial composition is formed as soon as 24 hours after birth, highlighting a more distinct microbial profiles at 24 hours post-birth. Microbial communities exhibit a tendency towards convergence as time increases to 96 and 240 hours, which is noteworthy as it implies stability of the gut microbiota.

However, as the gut microbiota of sows is thought to be more stable than that of piglets, incorporating sow samples in this analysis may have an impact on the orientation of clusters and result in an inaccurate conclusion. To grasp the scenario better, the sows in the graphics need to be highlighted. The outcome of this is displayed in figures 4.14 and 4.15.

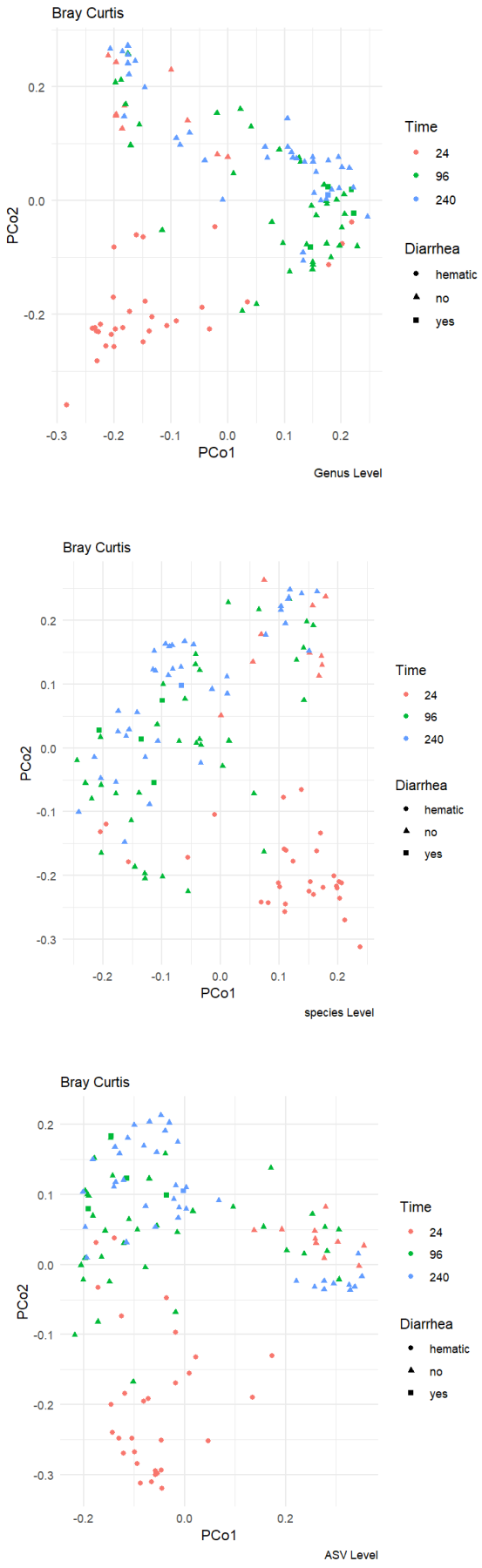


Figure 4.12: 2D PcoA visualization of Bray Curtis's dissimilarity metric showing the microbial dynamics between samples considering time and diarrhea.

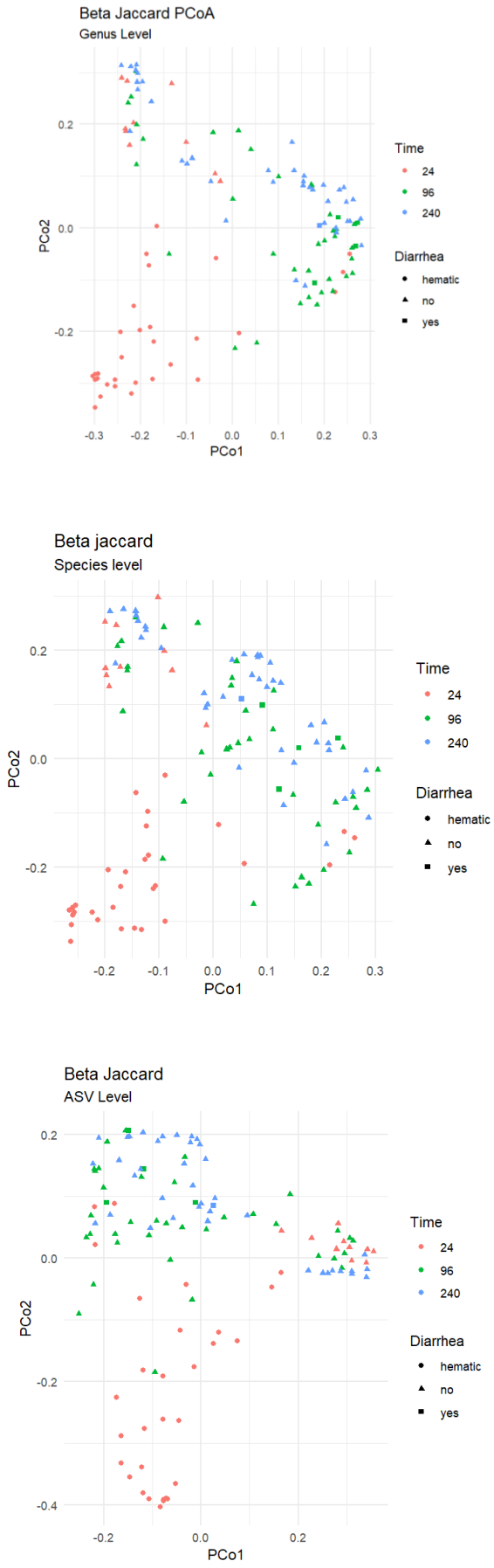


Figure 4.13: 2D PcoA visualization of Jaccard dissimilarity metric showing the microbial dynamics between samples considering time and diarrhea.

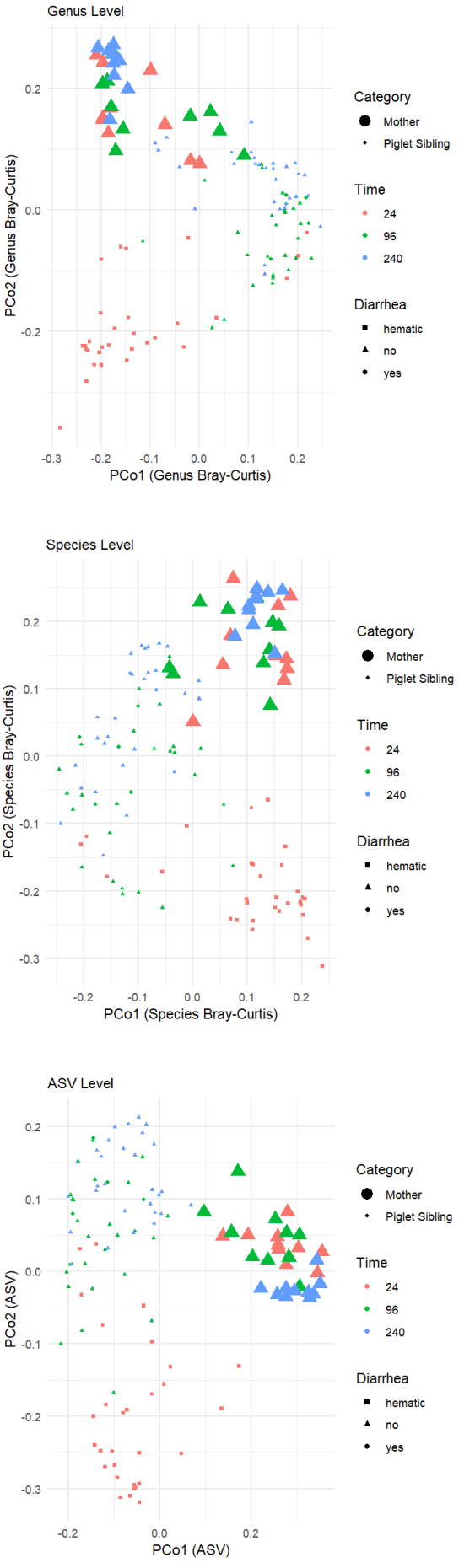


Figure 4.14: 2D PcoA visualization of Bray Curtis dissimilarity metric showing the microbial dynamics between samples considering time and diarrhea with sows accentuated.

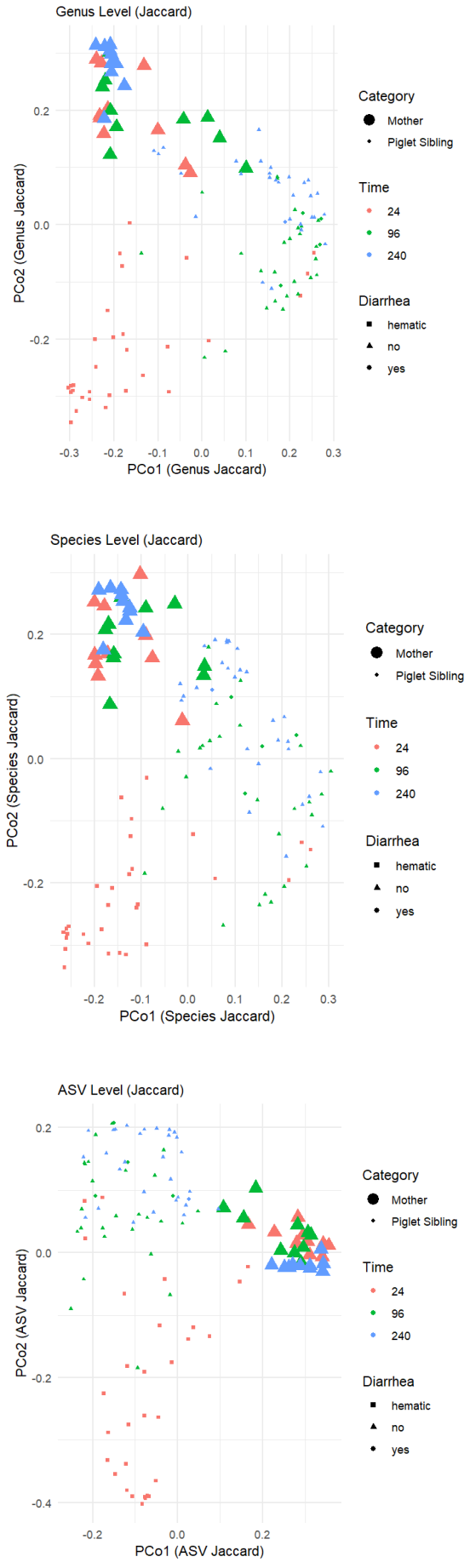


Figure 4.15: 2D PcoA visualization of Jaccard dissimilarity metric showing the microbial dynamics between samples considering time and diarrhea with sows accentuated.

Upon closer examination at the 24-hour mark, the clusters of bacteria in the piglets' guts and those in their sows' guts are noticeably different at first. This discrepancy implies that piglets rapidly form their distinct microbial communities, potentially impacted by early exposure and colonization mechanisms. However, at 96 and 240 hours post-birth, an interesting image emerges, especially at the genus and species levels. The mothers' and the piglets' microbial compositions here show a small overlap, suggesting that the microbial structures of piglets have evolved over time. This finding points to the possibility of shared environmental impacts or microbiological traits being transferred from mothers to piglets. Importantly, the gut microbial makeup of the sows remains more stable over time compared to the changes observed in piglets. This highlights the consistency of the maternal microbial reservoir. Exploring these time-related and taxonomic details gives us valuable insights into the complex relationship between the gut microbiota of sows and piglets with time.

A 3D PCoA graphic as seen in figures 4.16-4.18 was created with each time split to see the dynamics of what occurs at each time in a 3D view in order to examine the interaction between the gut microbiota of sows and piglets over time in more detail across the different taxonomic levels.

3D PCoA bray Curtis Genus Taxonomic level

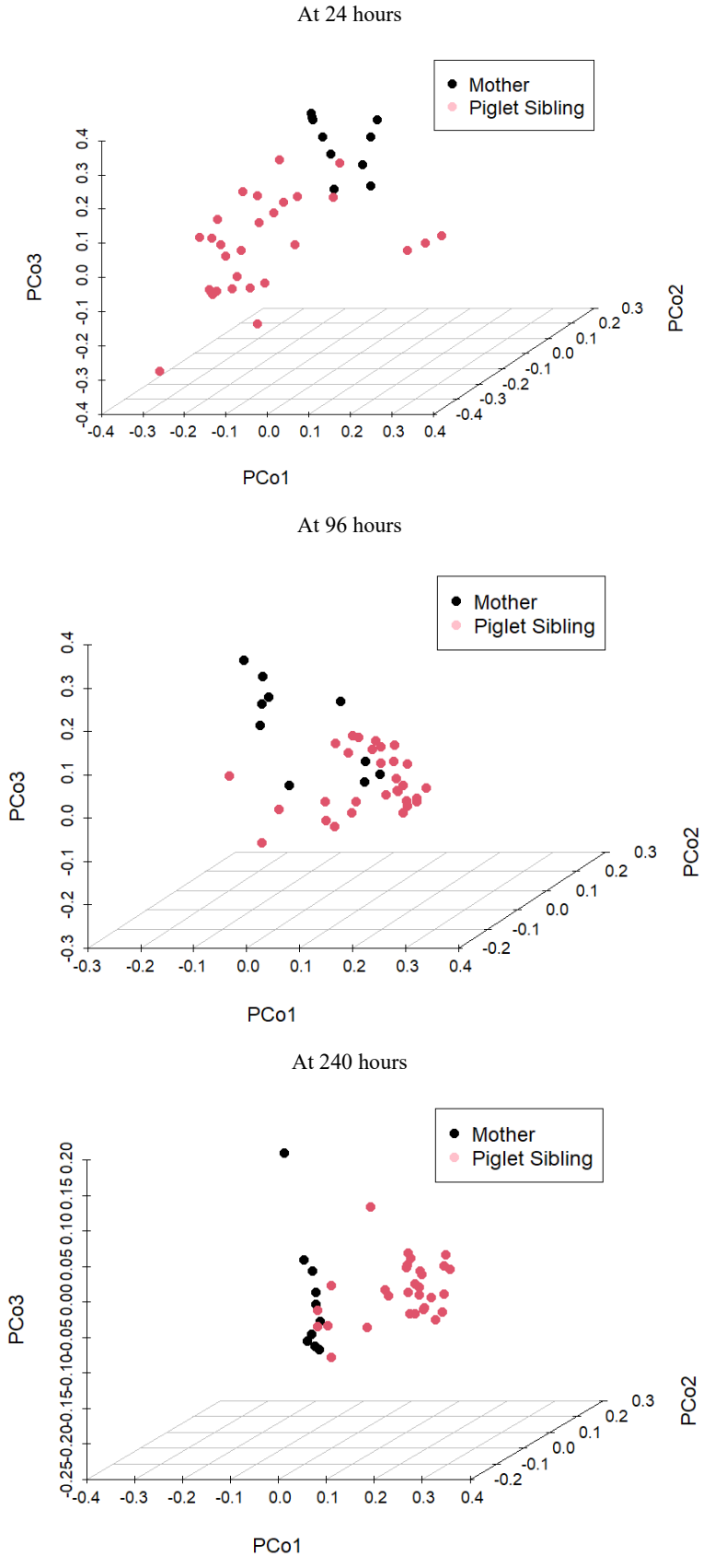
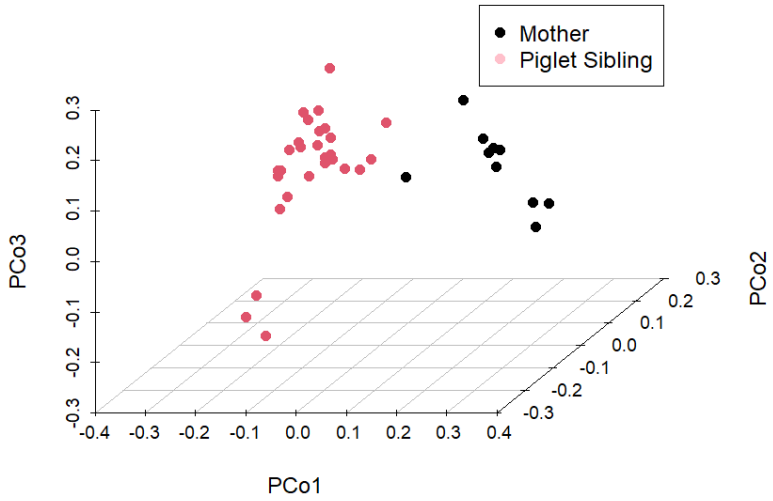


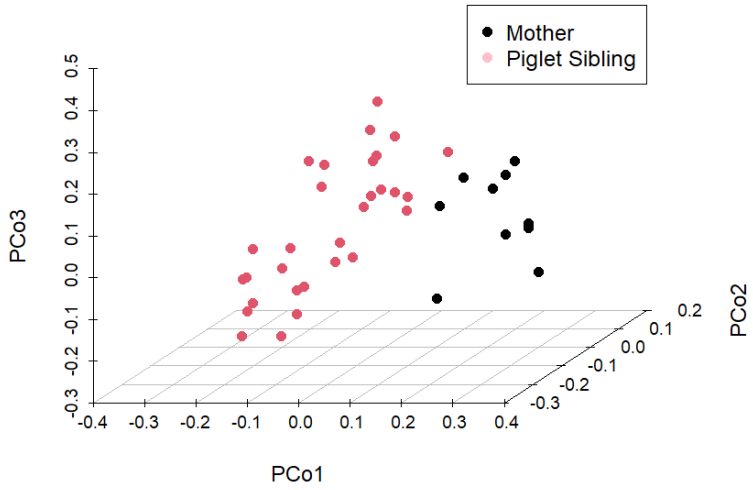
Figure 4.16: 3D PcoA visualization of Bray Curtis dissimilarity metric at the genus level showing the microbial dynamics between sows and Piglets.

3D PCoA bray Curtis Species Taxonomic level

At 24 hours



At 96 hours



At 240 hours

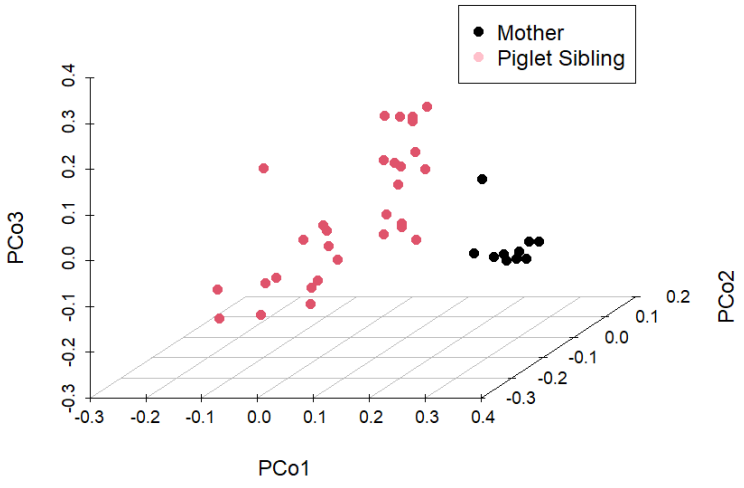


Figure 4.17: 3D PcoA visualization of Bray Curtis dissimilarity metric at the species level showing the microbial dynamics between sows and Piglets.

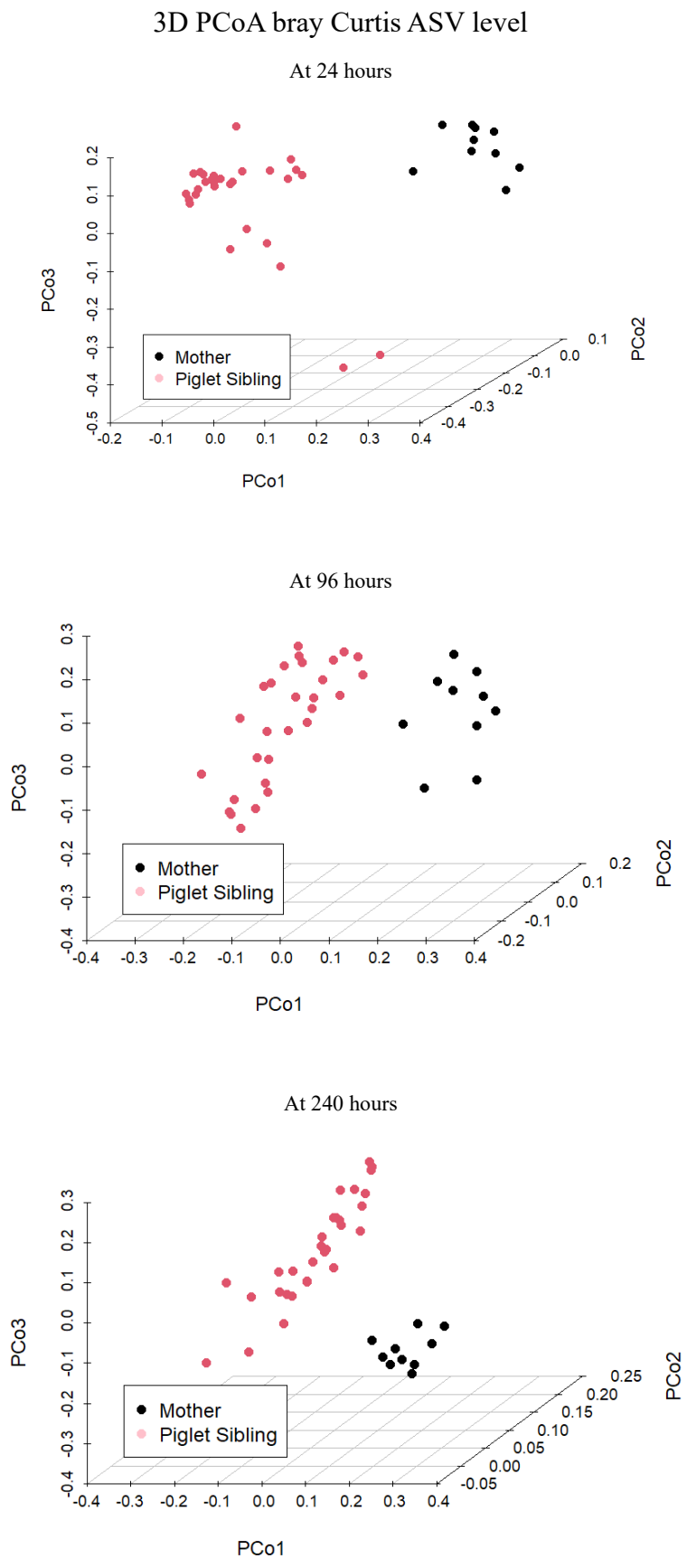


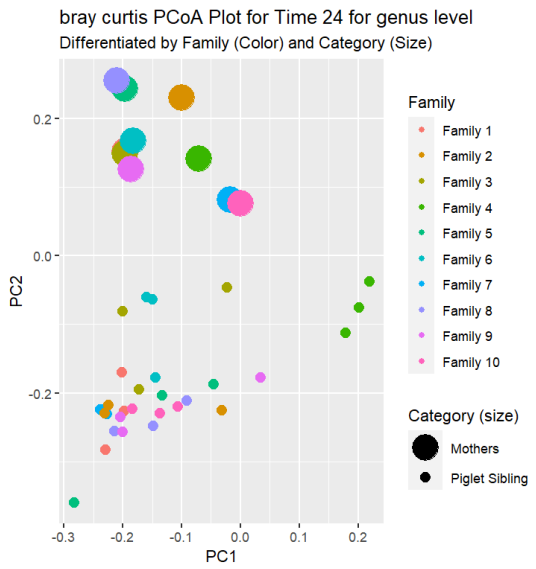
Figure 4.18: 3D PcoA visualization of Bray Curtis dissimilarity metric at the ASV level showing the microbial dynamics between sows and Piglets.

Interestingly, a clear shift toward the microbial composition of sows at the genus level was detected, which is most noticeable at the 96-hour mark. This implies that piglets and their mothers engage in dynamic interaction or possible sharing of microbiological characteristics, which may be impacted by variables such as nursing or close physical contact. However, at 240 hours, a faint but noticeable separation reappears, mirroring the situation seen at 24 hours.

When considering the species level, we see something similar happening, but with less movement and interaction. This might point to a more limited impact or targeted transfer of particular microbiological taxa between sows and piglets. A completely different situation emerges at the ASV level, demonstrating a constant and clear differentiation between sows' and piglets' microbial dynamics throughout the different sampling times.

Next, an additional examination is conducted to determine whether there is an association between the microbial features of sows and piglets belonging to the same family at different times. Different colors were used to symbolize each family in this 2D PCoA visualization.

Genus Taxonomic Level



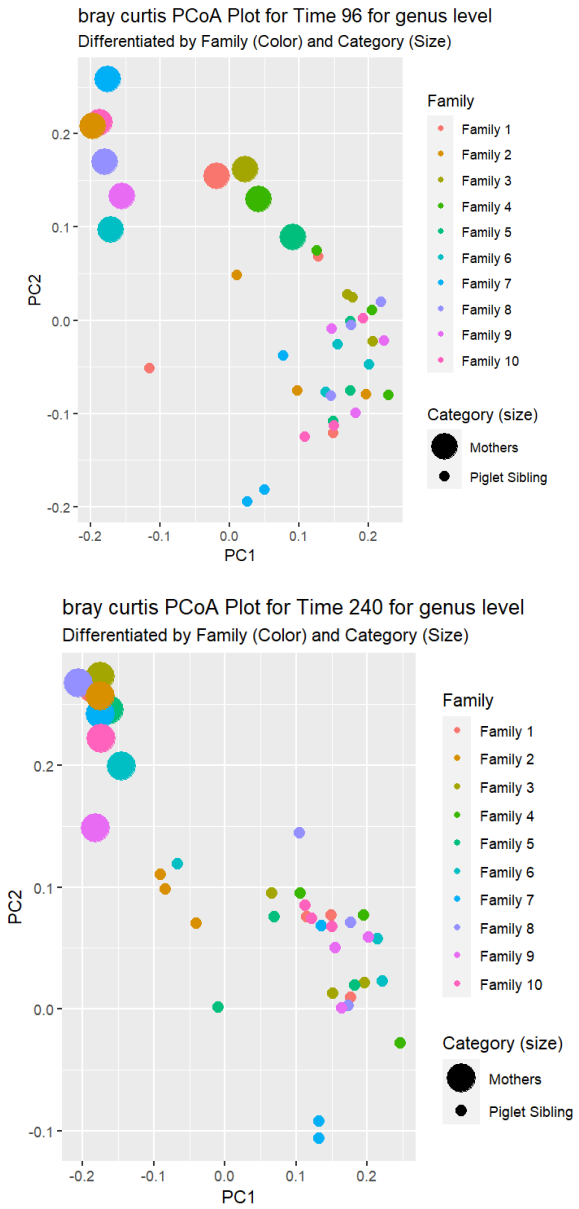


Figure 4.19: 2D PcoA visualization of Bray Curtis dissimilarity metric at the genus level showing the microbial dynamics between sows and Piglets of the same family.

Species Taxonomic Level

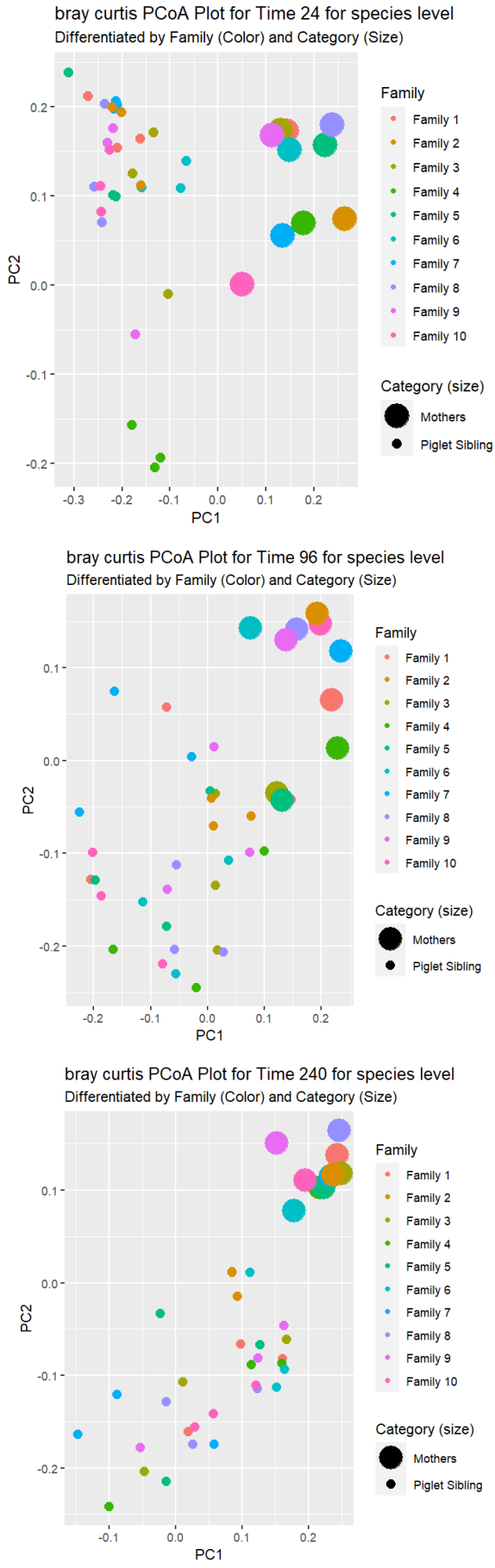


Figure 4.20: 2D PcoA visualization of Bray Curtis dissimilarity metric at the species level showing the microbial dynamics between sows and Piglets of the same family.

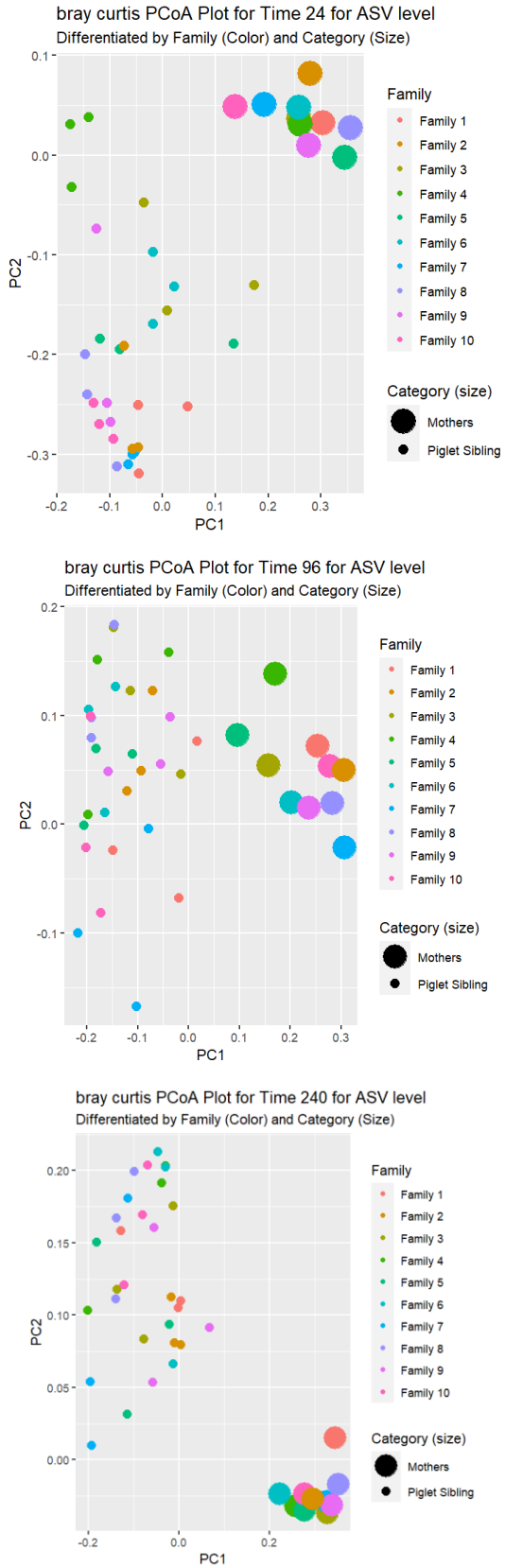


Figure 4.21: 2D PcoA visualization of Bray Curtis dissimilarity metric at the ASV level showing the microbial dynamics between sows and Piglets of the same family.

Sows and piglets belonging to the same family do not share clear microbial associations. The sow samples were separated for additional analysis to see whether piglets from the same sow or family shows any significant similarity of microbial traits.

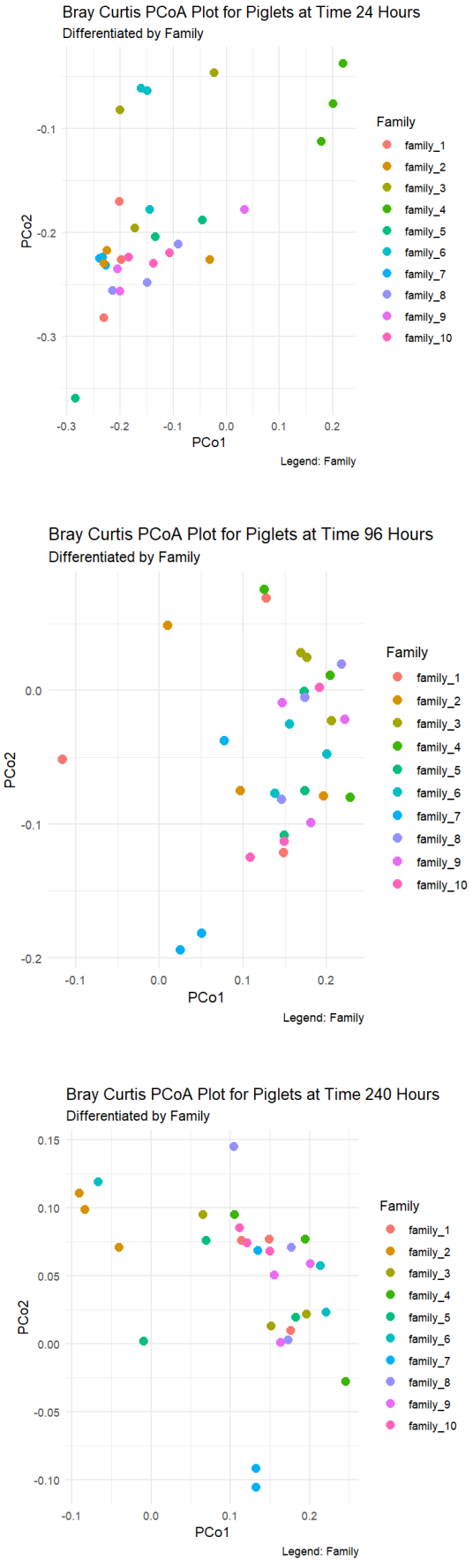


Figure 4.22: 2D PcoA visualization of Bray Curtis dissimilarity metric at the genus level showing the microbial dynamics between Piglets of the same family.

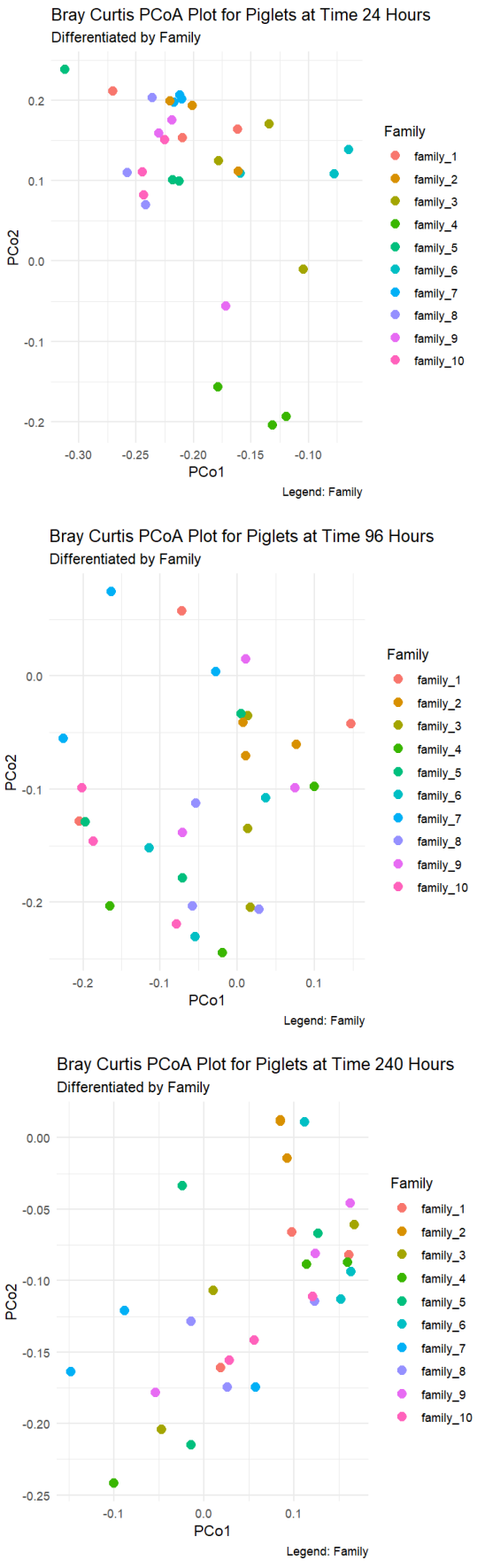


Figure 4.23: 2D PcoA visualization of Bray Curtis dissimilarity metric at the species level showing the microbial dynamics between Piglets of the same family.

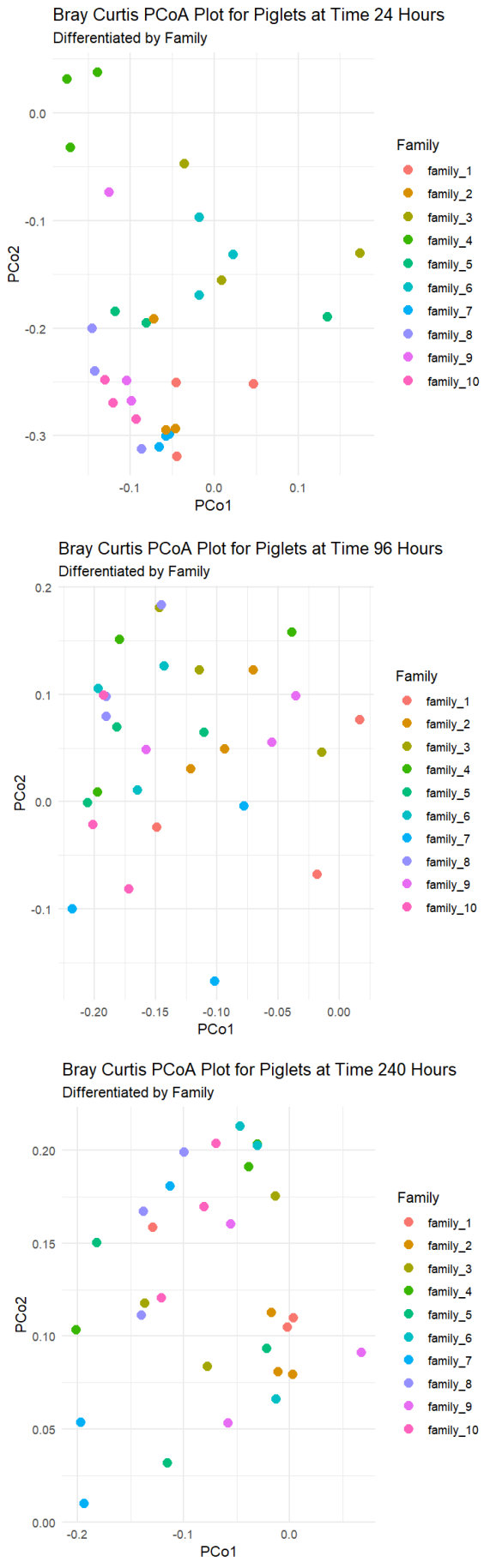


Figure 4.24: 2D PcoA visualization of Bray Curtis dissimilarity metric at the ASV level showing the microbial dynamics between Piglets of the same family.

There are no clear associations or relationships between the microbial compositions of piglets from the same family. The absence of a distinct microbiological connection among piglet samples is indeed surprising. Given that piglets from the same sow share the same room, it would be reasonable to expect some degree of microbial sharing between them. However, the study results defy this expectation, suggesting that the microbial communities in piglets are less interconnected than anticipated. Moreover, the insignificance of the living space factor (room) across all alpha diversity indexes supports this outcome, emphasizing the consistency of the findings. This discovery raises intriguing questions about the factors influencing microbial transmission in this particular piglet-sow environment. Consequently, further investigation is needed to delve into the dynamics of microbial interactions within this specific context.

Differential Abundance Analysis

Differential abundance (DA) analysis focused exclusively on the species level to attain a detailed understanding of microbial communities, surpassing the broader taxonomic resolutions such as genus. To enhance interpretability, time was transformed into a categorical variable, setting 24 hours as the reference point for a clearer heatmap presentation.

Upon analysis, a notable trend emerged as seen in Figure 4.25, indicating that the majority of significantly altered bacteria exhibited increased abundance from 24 hours to 240 hours. However, exceptions were observed in the cases of *C. perfringes* and *Parvimonas*, which displayed a decline in abundance from 96 to 240 hours concerning the reference at 24 hours.

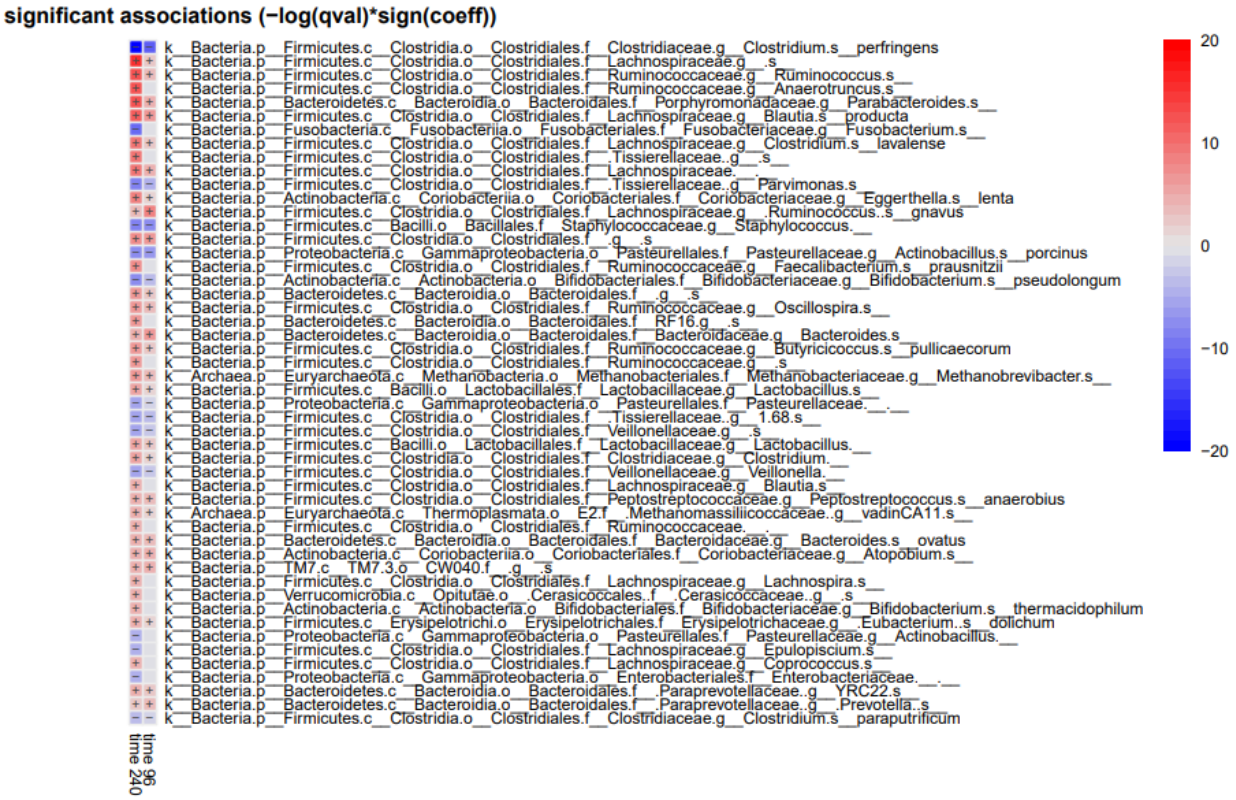


Figure 4.25: Heat map showing the DA of Piglets & Sows at different times with 24 hours as a reference.

Feature	metadata	value	coef	stderr	N	N.not.0	pval	qval
Clostridium perfringens	time	time	-1.1393	0.1504	120	120	8.95E-12	1.27E-09
Lachnospiraceae	time	time	0.59823	0.0886	120	118	5.72E-10	4.06E-08
Anaerotruncus	time	time	0.888076	0.1413	120	20	5.73E-09	2.71E-07
Ruminococcus	time	time	0.492662	0.0839	120	118	4.08E-08	1.45E-06
Fusobacterium	time	time	-0.78482	0.1407	120	120	1.58E-07	4.25E-06
Parabacteroides	time	time	0.675239	0.1217	120	118	1.80E-07	4.25E-06
Tissierellaceae	time	time	0.226249	0.0447	120	65	1.55E-06	3.15E-05
Clostridium lavalense	time	time	1.059733	0.2121	120	26	2.05E-06	3.64E-05
Blautia producta	time	time	0.678316	0.1398	120	103	3.76E-06	5.94E-05
Eggerthella lenta	time	time	0.793603	0.1659	120	22	5.01E-06	7.12E-05
Faecalibacterium prausnitzii	time	time	0.608328	0.1304	120	16	8.20E-06	0.000106

Table 4.4: Top significant features that are differentially abundant in Piglets & Sows at different times with 24 hours as a reference.

Interestingly, Fusobacterium was absent at 96 hours but reappeared at 240 hours but at a diminished abundance compared to the reference at 24 hours. This temporal pattern suggests a dynamic microbial response over the tested time intervals. The decrease in abundance of certain bacteria could potentially be linked to shifts in the microbial community's composition or adaptations to the evolving gut environment.

Furthermore, another analysis DA analysis performed was to check the general differentially abundant microbial taxa between sows and piglets without reference to any time. The table below (table 4.5) shows the results obtained.

The majority of the most significant differentially abundant taxa are less abundant in the piglets except *Clostridium aldenense, Fusobacteriaceae, and Enterobacteriaceae. Campylobacteraceae* campylobacter is the most significant differentially abundant taxa.

Table 4.6 shows the result of another important DA analysis done was on piglet samples only across different sampling times using the 24-hour mark as a reference. Figure 4.26 and Table 4.6 shows the results. The differential abundance (DA) study reveals *C. perfringens* as a major finding. It shows a considerable decline in abundance beginning 24 hours post-birth and continuing to drop at 96 and 240 hours. Notably, *C. perfringens* consistently emerges as the most differentially abundant characteristic, demonstrating a drop in abundance post the 24-hour sample time, when comparing piglets and sows at different time points with 24 hours as a reference.

Feature	Metadata	value	coef	stderr	N	N.not.0	pval	qval
Campylobacteraceae Campylobacter	Category	Piglet Sibling	-2.8177	0.162	120	118	2.53E-34	3.60E-32
Paraprevotellaceae	Category	Piglet Sibling	-5.15479	0.3605	120	26	1.61E-27	1.14E-25
Clostridium aldenense	Category	Piglet Sibling	1.33647	0.0977	120	101	4.27E-26	2.02E-24
Coriobacteriaceae	Category	Piglet Sibling	-4.0402	0.3206	120	39	1.37E-23	4.88E-22
Fusobacteriaceae	Category	Piglet Sibling	4.169793	0.3403	120	98	9.09E-23	2.58E-21
Spirochaetaceae	Category	Piglet Sibling	-3.19615	0.2876	120	20	4.54E-20	9.20E-19
Enterobacteriaceae	Category	Piglet Sibling	3.306633	0.3061	120	114	2.50E-19	4.44E-18
Pyramidobacter	Category	Piglet Sibling	-2.27897	0.232	120	19	5.24E-17	8.27E-16
Bulleidia	Category	Piglet Sibling	-2.21402	0.2306	120	17	1.77E-16	2.51E-15

Table 4.5: Top significant features that are differentially abundant in Piglets & Sows with no reference to time.

Furthermore, an analysis done on piglets' samples alone as shown in the figure 4.26 again indicates that *C. perfringes* is the most significant differential abundant feature. The first four of the five cases of diarrheal samples happened at 96 hours, whereas the fifth sample wasn't seen until 240 hours. According to this, we can hypothesize that number of diarrhea cases decrease as *C. perfringens* abundance decreases from 96 to 240 hours.

This remark is supported by a review of the literature. Necrotic enteritis is the term for *C. perfringens* associated with diarrhea in pigs. Jacobson (2022) claims that diarrhea, tiredness, and slowed development rates are the normal symptoms of this sickness in young pigs. In their elaboration on the pathophysiology, Mehdizadeh-Gohari et al. (2021) emphasize C. perfringens' ability to produce toxins, specifically the NetB toxin in type A strains. This toxin destroys the lining of the stomach, interfering with normal gut function. As noted by Lee and Lillehoj (2022), this disruption leads to the development of necrotic lesions and inflammation, manifesting in diarrhea and other clinical signs.

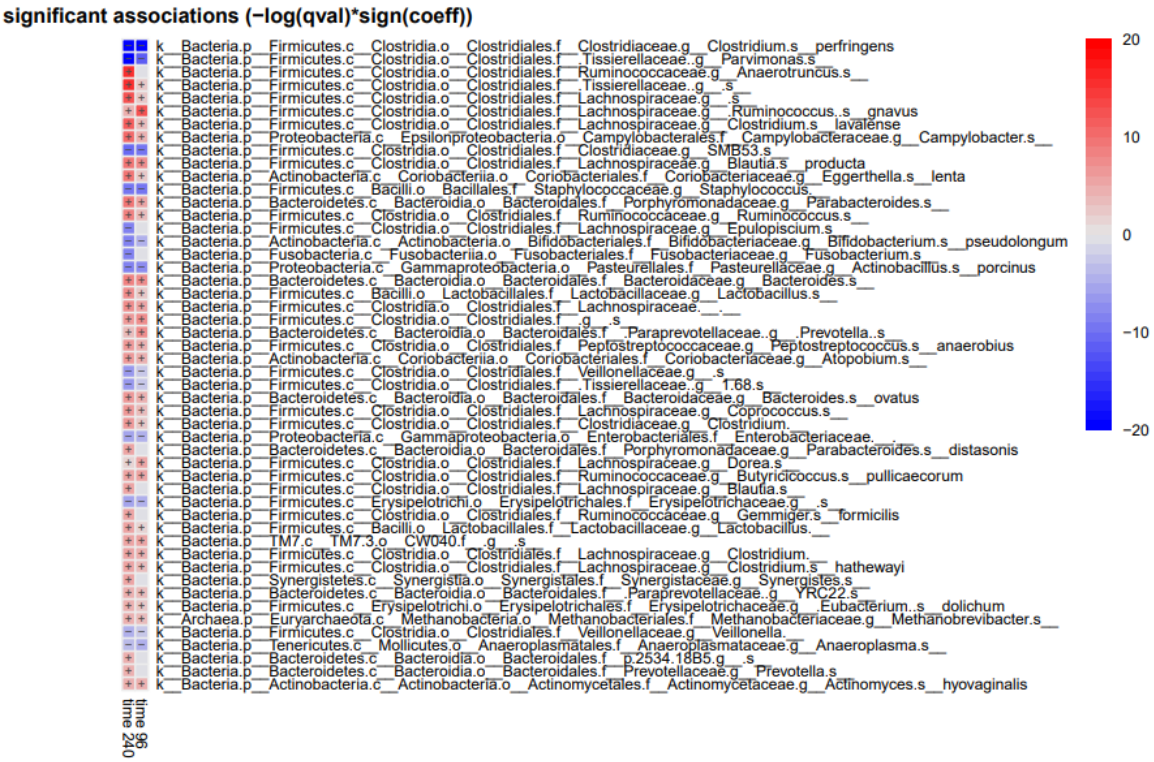


Figure 4.26: Heat map showing the DA of only Piglets at different times with 24 hours as a reference.

FEATURE	metadata	value	coef	stderr	N	N.not.0	pval	qval
C. perfringens	time	240	-3.911	0.3288	90	90	6.20E-20	1.46E-17
C. perfringens	time	96	-2.5013	0.3288	90	90	3.09E-11	3.65E-09
Parvimonas	time	240	-1.2513	0.1674	90	84	5.65E-11	4.44E-09
Anaerotruncus	time	240	2.6844	0.4095	90	20	3.79E-09	2.24E-07
Tissierellaceae	time	240	0.7172	0.111	90	35	5.76E-09	2.72E-07
Lachnospiraceae	time	240	1.5746	0.2729	90	88	1.20E-07	4.04E-06
Ruminococcus gnavus	time	96	1.9408	0.3347	90	88	1.05E-07	4.04E-06
Clostridium lavalense	time	240	3.4723	0.6316	90	26	3.79E-07	1.12E-05
Parvimonas	time	96	-0.9105	0.1674	90	84	4.81E-07	1.26E-05
Campylobacter	time	240	0.2845	0.0532	90	88	7.18E-07	1.70E-05

Table 4.6: Top significant features that are differentially abundant only in piglet samples across the different times with reference to the 24 hours.

Finally, the microbial differential abundance between diarrheic and non-diarrheic samples was examined. Table 4.7 shows the outcome. Contrary to expectations, *C. perfringens* is not part of the significant

differential abundant taxa. The claim that the decrease in abundance of *C. perfringens* from the 24-hour to 240-hour sampling times might be associated with the occurrence of diarrhea is questionable. Furthermore, none of the featured taxa in this particular DA analysis has a significant q-value.

Feature	metadata	value	coef	stderr	N	N.not.0	pval	qval
Bacteroides ovatus	diarrhea_status	diarrheic	2.604367	0.809517	90	88	0.00181	0.167614
Erysipelotrichaceae	diarrhea_status	diarrheic	-0.68509	0.223109	90	56	0.00284	0.167614
Bacteroides	diarrhea_status	diarrheic	1.654933	0.591164	90	88	0.00629	0.247425

Table 4.7: Top significant features that are differentially abundant in piglet diarrheic and non-diarrheic samples.

5. CONCLUSIONS

The primary aim of this research is to comprehend the intricacies of *C. difficile* infection, particularly how other gut microorganisms interact synergistically or antagonistically with this infectious bacterium. Surprisingly, the examination of the pig farm revealed the absence of *C. difficile*. Initially, there was suspicion regarding the differential abundance of *C. perfringens* at various sampling times, assuming a potential link to the diarrheic status of piglets. This because *Clostridium perfringens* has been linked to diarrhea in pigs. According to Jacobson (2022), the disease causes diarrhoea, tiredness, and a drop in growth rates in young piglets. Number of diarrheic piglets reduces from four at 96 hours after birth of piglets to one at 240 hours after birth. This reduction in number of diarrheic piglets was thought to be related to the decrease in the differential abundance of *C. perfringens*. However, DA analysis between diarrheic piglets and non-diarrheic piglets indicated that this might not be the case. This is because we would expect to observe a notable change in abundance of *C. perfringens* between the two categories. This was not the case as the most significant differential abundant species was *Bacteroides ovatus* and has an insignificant Q-value.

Interestingly, the beta analysis unveiled a noteworthy shift in the microbial composition of piglets, both at the genus and species levels, towards that of the sow from 24 hours to 240 hours post-birth. However, this trend experiences a deviation, marked by a noticeable separation at the 240-hour mark. Additionally, it is evident that the gut microbiota of sows remains relatively more stable compared to that of piglets. Some of the most significant genera that are differentially abundant between sows are piglets are *Campylobacter, Clostridium aldenese,* and *Paraprevotellaceae.* Of this list *C. aldenese* was the more abundant in piglets while the others were less abundant. Zhang *et al.* (2018) propose that *C. aldenense* holds promise in enhancing short-chain fatty acid (SCFA) production within the pig gut, thereby fostering intestinal health through mechanisms such as pH reduction, inhibition of pathogenic bacteria, and provision of nourishment to gut cells. Their study was conducted on weaning pigs fed varied fiber sources, suggests that inclusion of *C. aldenense* in diets may confer these beneficial effects. Additionally, it is posited that *C. aldenense* could outcompete and suppress detrimental bacteria in the gut, potentially further bolstering gut health (Li *et al*., 2014).

For future improvement, this experiment could benefit from sampling a pig farm confirmed to have a *C. difficile* infection. This would require establishing a connection and trust between pig farmers and scientists. Furthermore, extending the sampling duration to include more days, with samples collected on a 24-hour basis, could provide a broader understanding of the dynamic evolution of the gut microbiota in piglets and the influential role of sows in shaping these dynamics.

REFERENCES

1. Abia, A.L.K., Alisoltani, A., Keshri, J. and Ubomba-Jaswa, E. (2018). Metagenomic analysis of the bacterial communities and their functional profiles in water and sediments of the Apies River, South Africa, as a function of land use. *Science of The Total Environment*, 616-617, pp.326–334. DOI: https://doi.org/10.1016/j.scitotenv.2017.10.322.

2. Akhtar, M., Gulab, M. and Ghazanfar, M. (2023). Assessing Soil Health and Fertility through Microbial Analysis and Nutrient Profiling Implications for Sustainable Agriculture. *Innovative Research in Applied, Biological and Chemical Sciences*, [online] 1(1), pp.29–42. Available at: https://irabcs.com/ojs/article/view/14 [Accessed 14 Jan. 2024].

3. Aktories, K., Schwan, C. and Jank, T. (2017). Clostridium difficile Toxin Biology. *Annual Review of Microbiology*, 71(1), pp.281–307. DOI: https://doi.org/10.1146/annurev-micro-090816-093458.

4. Al-Hinai, M.A., Jones, S.W. and Papoutsakis, E.T. (2015). The Clostridium Sporulation Programs: Diversity and Preservation of Endospore Differentiation. *Microbiology and Molecular Biology Reviews : MMBR*, [online] 79(1), pp.19–37. DOI: https://doi.org/10.1128/MMBR.00025-14.

5. Alves, F., Nunes, A., Castro, R., Sequeira, A., Moreira, O., Matias, R., Rodrigues, J.C., Silveira, L., Gomes, J.P. and Oleastro, M. (2022). Assessment of the transmission dynamics of Clostridioides difficile in a farm environment reveals the presence of a new toxigenic strain connected to swine production. *Frontiers in Microbiology*, 13, p.858310.

6. Alves, L. de F., Westmann, C.A., Lovate, G.L., de Siqueira, G.M.V., Borelli, T.C. and Guazzaroni, M.-E. (2018). *Metagenomic Approaches for Understanding New Concepts in Microbial Science*. [online] International Journal of Genomics. Available at: https://www.hindawi.com/journals/ijg/2018/2312987/ [Accessed 14 Jan. 2024].

7. Andrés-Lasheras, S.R., Bolea, Rosa.C., Mainar-JaimE, E., Kuijper, E., Sevilla, E., Martín-Burriel, I. and Chirino-Trejo, M. (2016). Presence of clostridium *difficile* in pig faecal samples and wild animal species associated with pig farms. *Journal of Applied Microbiology*, [online] 122(2), pp.462–472. DOI: https://doi.org/10.1111/jam.13343.

8. Awad, M.M., Johanesen, P.A., Carter, G.P., Rose, E. and Lyras, D. (2014). Clostridium difficile virulence factors: Insights into an anaerobic spore-forming pathogen. *Gut Microbes*, 5(5), pp.579–593. DOI: https://doi.org/10.4161/19490976.2014.969632.

9. Baines, S. and Wilcox, M. (2015). Antimicrobial Resistance and Reduced Susceptibility in Clostridium difficile: Potential Consequences for Induction, Treatment, and Recurrence of C. difficile Infection. *Antibiotics*, 4(3), pp.267–298. DOI: https://doi.org/10.3390/antibiotics4030267.

10. Barra-Carrasco, J. and Paredes-Sabja, D. (2014). Clostridium difficile spores: a major threat to the hospital environment. *Future Microbiology*, 9(4), pp.475–486. DOI: https://doi.org/10.2217/fmb.14.2.

11. Bellato, M., Cappellato, M., Longhin, F., Del Vecchio, C., Brancaccio, G., Cattelan, A.M., Brun, P., Salaris, C., Castagliuolo, I. and Di Camillo, B. (2023). Uncover a microbiota signature of upper respiratory tract in patients with SARS-CoV-2 + . *Scientific Reports,* 13, 16867. DOI: https://doi.org/10.1038/s41598-023-43040-x

12. Bhargava, P., Khan, M., Verma, A., Singh, A., Singh, S., Siddharth Vats and Goel, R. (2019). Metagenomics as a Tool to Explore New Insights from Plant-Microbe Interface. *Plant microbe interface*, pp.271–289. DOI: https://doi.org/10.1007/978-3-030-19831-2_12.

13. Cairns, M.D. (2018). *Evolution of Clostridium difficile*. [online] discovery.ucl.ac.uk. Available at: https://discovery.ucl.ac.uk/id/eprint/10060221/ [Accessed 11 Jan. 2024].

14. Chowdhury, P.R., DeMaere, M.Z., Chapman, T.A., Worden, P., Charles, I.G., Darling, A.E. and Djordjevic, S.P. (2016). Comparative genomic analysis of toxin-negative strains of Clostridium difficile from humans and animals with symptoms of gastrointestinal disease. *BMC Microbiology*. DOI: https://doi.org/10.1186/s12866-016-0653-3.

15. Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W.K., Potter, C. and Bik, H.M. (2016). The ecologist's field guide to sequence-based identification of biodiversity. *Methods in Ecology and Evolution*, 7(9), pp.1008–1018. DOI: https://doi.org/10.1111/2041-210x.12574.

16. Czepiel, J., Dróżdż, M., Pituch, H., Kuijper, E.J., Perucki, W., Mielimonka, A., Goldman, S., Wultańska, D., Garlicki, A. and Biesiada, G. (2019). Clostridium difficile infection: review. *European Journal of Clinical Microbiology & Infectious Diseases*, [online] 38(7), pp.1211–1221. DOI: https://doi.org/10.1007/s10096-019-03539-6.

17. Denman, S.E. and McSweeney, C.S. (2015). The Early Impact of Genomics and Metagenomics on Ruminal Microbiology. *Annual Review of Animal Biosciences*, 3(1), pp.447–465. DOI: https://doi.org/10.1146/annurev-animal-022114-110705.

18. Di Bella, S., Ascenzi, P., Siarakas, S., Petrosillo, N. and di Masi, A. (2016). Clostridium difficile Toxins A and B: Insights into Pathogenic Properties and Extraintestinal Effects. *Toxins*, 8(5), p.134. DOI: https://doi.org/10.3390/toxins8050134.

19. DuPont, H.L. (2014). Challenges and opportunities in the management of Clostridium *difficile* infection. *Expert Review of Gastroenterology & Hepatology*, 8(8), pp.863–874. doi https://doi.org/10.1586/17474124.2014.939630.

20. Fadiji, A.E. and Babalola, O.O. (2020). Metagenomics methods for the study of plant-associated microbial communities: A review. *Journal of Microbiological Methods*, 170, p.105860. DOI: https://doi.org/10.1016/j.mimet.2020.105860.

21. Flock, G. (2017). Clostridium difficile: A Study on its Potential as a Food-borne Pathogen and Strategies for Controlling its Transmission. *Doctoral Dissertations*. [online] Available at: https://opencommons.uconn.edu/dissertations/1584/ [Accessed 11 Jan. 2024].

22. Gao, Y., Gao, W., Cheng, J., Ma, L. and Su, J. (2020). Identification of the role of toxin B in the virulence of Clostridioides difficile based on integrated bioinformatics analyses. *International Microbiology*, 23(4), pp.575–587. DOI: https://doi.org/10.1007/s10123-020-00128-y.

23. Garlapati, D., Charankumar, B., Ramu, K., Madeswaran, P. and Ramana Murthy, M.V. (2019). A review on the applications and recent advances in environmental DNA (eDNA) metagenomics. *Reviews in Environmental Science and Bio/Technology*, 18(3), pp.389–411. DOI: https://doi.org/10.1007/s11157-019-09501-4.

24. Garza, D.R. and Dutilh, B.E. (2015). From cultured to uncultured genome sequences: metagenomics and modeling microbial ecosystems. *Cellular and Molecular Life Sciences*, 72(22), pp.4287–4308. DOI: https://doi.org/10.1007/s00018-015-2004-1.

25. Gebreyes, W.A., Dupouy-Camet, J., Newport, M.J., Oliveira, C.J.B., Schlesinger, L.S., Saif, Y.M., Kariuki, S., Saif, L.J., Saville, W., Wittum, T., Hoet, A., Quessy, S., Kazwala, R., Tekola, B., Shryock, T., Bisesi, M., Patchanee, P., Boonmar, S. and King, L.J. (2014). The Global One Health Paradigm: Challenges and Opportunities for Tackling Infectious Diseases at the Human, Animal, and Environment Interface in Low-Resource Settings. *PLoS Neglected Tropical Diseases*, 8(11), p.e3257. DOI: https://doi.org/10.1371/journal.pntd.0003257.

26. Gebreyes, W.A., Jackwood, D., de Oliveira, C.J.B., Lee, C.-W., Hoet, A.E. and Thakur, S. (2020). Molecular Epidemiology of Infectious Zoonotic and Livestock Diseases. *Microbiology Spectrum*, 8(2). doi https://doi.org/10.1128/microbiolspec.ame-0011-2019.

27. Gil, F., Lagos-Moraga, S., Calderón-Romero, P., Pizarro-Guajardo, M. and Paredes-Sabja, D. (2017). Updates on Clostridium difficile spore biology. *Anaerobe*, 45, pp.3–9. DOI: https://doi.org/10.1016/j.anaerobe.2017.02.018.

28. Goldenberg, S.D. and French, G.L. (2011). Diagnostic testing for Clostridium difficile: a comprehensive survey of laboratories in England. *Journal of Hospital Infection*, 79(1), pp.4–7. DOI: https://doi.org/10.1016/j.jhin.2011.03.030.

29. Gomaa, E.Z. (2020). Human Gut microbiota/microbiome in Health and diseases: a Review. *Antonie van Leeuwenhoek*, 113(12). DOI: https://doi.org/10.1007/s10482-020-01474-7.

30. Grześkowiak, Ł.M., Pieper, R., Huynh, H.A., Cutting, S.M., Vahjen, W. and Zentek, J. (2019). Impact of early-life events on the susceptibility to Clostridium difficile colonization and infection in the offspring of the pig. *Gut Microbes*, 10(2), pp.251–259. doi https://doi.org/10.1080/19490976.2018.1518554.

31. Gupta, A., Gupta, R. and Singh, R.L. (2016). Microbes and Environment. *Principles and Applications of Environmental Biotechnology for a Sustainable Future*, pp.43–84. DOI: https://doi.org/10.1007/978-981-10-1866-4_3.

32. Gwenzi, W., Ngaza, N., Marumure, J., Makuvara, Z., Muteveri, M., Nyambiya, I., Musvuugwa, T. and Chaukura, N. (2023). Occurrence and Health Risks of Antibiotic Resistance in African Aquatic Systems. *Springer eBooks*, pp.107–159. DOI: https://doi.org/10.1007/978-3-031-23796-6_7.

33. Hain-Saunders, N.M.R., Knight, D.R., Bruce, M. and Riley, T.V. (2022). *Clostridioides difficile* infection and One Health: an equine perspective. *Environmental Microbiology*, 24(3), pp.985–997. DOI: https://doi.org/10.1111/1462-2920.15898.

34. Hellmann, K., Tuura, C., Fish, J.P., Patel, J.M. and Robinson, D.A. (2021). Viability-Resolved Metagenomics Reveals Antagonistic Colonization Dynamics of Staphylococcus epidermidis Strains on Preterm Infant Skin. *mSphere*, 6(5). DOI: https://doi.org/10.1128/msphere.00538-21.

35. Jacobson, M. (2022). On the Infectious Causes of Neonatal Piglet Diarrhoea—A Review. Veterinary Sciences, 9(8), p.422. DOI: https://doi.org/10.3390/vetsci9080422.

36. Jurburg, S.D., Cornelissen, J.J., De Boer, P., Smits, M.A. and Rebel, J.M. (2019). Successional dynamics in the gut microbiome determine the success of Clostridium difficile infection in adult pig models. *Frontiers in cellular and infection microbiology*, 9, p.271.

37. Keessen , E.C., Harmanus , C., Dohmen , W. and Lipman, J.A. (2013). *Clostridium difficile* Infection Associated with Pig Farms. *Emerging infectious diseases*, 19(6), pp.1032–1034. DOI: https://doi.org/10.3201/eid1906.121645.

38. Kim, J.C., Hansen, C.F., Mullan, B.P. and Pluske, J.R. (2012). Nutrition and pathology of weaner pigs: Nutritional strategies to support barrier function in the gastrointestinal tract. *Animal Feed Science and Technology*, 173(1-2), pp.3–16. doi https://doi.org/10.1016/j.anifeedsci.2011.12.022.

39. Kirubakaran, R., ArulJothi, K.N., Revathi, S., Shameem, N. and Parray, J.A. (2020). Emerging priorities for microbial metagenome research. *Bioresource Technology Reports*, [online] 11, p.100485. DOI: https://doi.org/10.1016/j.biteb.2020.100485.

40. Kumar, V., Thakur, I.S., Singh, A.K. and Shah, M.P. (2020). *Chapter 8 - Application of metagenomics in remediation of contaminated sites and environmental restoration*. [online] ScienceDirect. Available at: https://www.sciencedirect.com/science/article/pii/B9780128198605000080 [Accessed 14 Jan. 2024].

41. Kumari, A., Sumer, S., Jalan, B., Nongbri, P.L. and Laskar, M.A. (2017). Impact of Next-Generation Sequencing Technology in Plant–Microbe Interaction Study. *Microbial Applications Vol.1*, pp.269–294. DOI: https://doi.org/10.1007/978-3-319-52666-9_13.

42. Lasken, R.S. and McLean, J.S. (2014). Recent advances in genomic DNA sequencing of microbial species from single cells. *Nature reviews. Genetics*, [online] 15(9), pp.577–584. DOI: https://doi.org/10.1038/nrg3785.

43. Lawler, A.J., Lambert, P.A. and Worthington, T. (2020). A Revised Understanding of Clostridioides difficile Spore Germination. *Trends in Microbiology*, 28(9), pp.744–752. DOI: https://doi.org/10.1016/j.tim.2020.03.004.

44. Lee, J.A., Chico, T.J.A. and Renshaw, S.A. (2017). The triune of the intestinal microbiome, genetics and inflammatory status and its impact on lower gastrointestinal anastomosis healing. *FEBS Journal*, [online] 285(7), pp.1212–1225. Available at: https://eprints.whiterose.ac.uk/125093/ [Accessed 11 Jan. 2024].

45. Lee, K.W. and Lillehoj, H.S. (2022). Role of Clostridium perfringens Necrotic Enteritis B-like Toxin in Disease Pathogenesis. Vaccines, [online] 10(1), p.61. DOI: https://doi.org/10.3390/vaccines10010061.

46. Lim, S.C., Knight, D.R. and Riley, T.V. (2020). Clostridium difficile and One Health. *Clinical Microbiology and Infection*, 26(7), pp.857–863. DOI: https://doi.org/10.1016/j.cmi.2019.10.023.

47. Luppi, A., D'Annunzio, G., Torreggiani, C. and Martelli, P. (2023). Diagnostic Approach to Enteric Disorders in Pigs. *Animals*, 13(3), p.338. DOI: https://doi.org/10.3390/ani13030338.

48. Maes, D.G.D., Dewulf, J., Piñeiro, C., Edwards, S. and Kyriazakis, I. (2019). A critical reflection on intensive pork production with an emphasis on animal health and welfare. *Journal of Animal Science*, 98. DOI: https://doi.org/10.1093/jas/skz362.

49. Malan-Muller, S., Valles-Colomer, M., Raes, J., Lowry, C.A., Seedat, S. and Hemmings, S.M.J. (2018). The Gut Microbiome and Mental Health: Implications for Anxiety- and Trauma-Related Disorders. *OMICS: A Journal of Integrative Biology*, 22(2), pp.90–107. DOI: https://doi.org/10.1089/omi.2017.0077.

50. Martin, J.S.H., Monaghan, T.M. and Wilcox, M.H. (2016). Clostridium difficile infection: epidemiology, diagnosis and understanding transmission. *Nature Reviews Gastroenterology & Hepatology*, 13(4), pp.206–216. DOI: https://doi.org/10.1038/nrgastro.2016.25.

51. Mehdizadeh-Gohari, I., Navarro, M.A., Li, J., Shrestha, A., Uzal, F. and McClane, B.A. (2021). Pathogenicity and virulence of Clostridium perfringens. Virulence, 12(1), pp.723–753. DOI: https://doi.org/10.1080/21505594.2021.1886777.

52. Mesnage, R., Antoniou, M.N., Tsoukalas, D., Goulielmos, G.N. and Tsatsakis, A. (2018). Gut microbiome metagenomics to understand how xenobiotics impact human health. *Current Opinion in Toxicology*, 11-12, pp.51–58. DOI: https://doi.org/10.1016/j.cotox.2019.02.002.

53. Miguel, L., Borelli, G., Camargo, A.P., Alexandrino, M., de-Ferraz, S.M.F., Fiamenghi, M.B., José, J., Mofatto, L.S., Nagamatsu, S.T., Persinoti, G.F., Silva, N.V., de, A., Amarante, G. and Carazzolle, M.F. (2019). Bioinformatics applied to biotechnology: A review towards bioenergy research. *Biomass and Bioenergy*, 123, pp.195–224. DOI: https://doi.org/10.1016/j.biombioe.2019.02.016.

54. Moono, P. (2017). *Aspects of Clostridium difficile infection in pigs*. [online] the UWA Profiles and Research Repository. Available at: https://research-repository.uwa.edu.au/en/publications/aspects-of-clostridium-difficile-infection-in-pigs [Accessed 11 Jan. 2024].

55. Napolitano, L.M. and Edmiston, C.E. (2017). Clostridium difficile disease: Diagnosis, pathogenesis, and treatment update. *Surgery*, 162(2), pp.325–348. DOI: https://doi.org/10.1016/j.surg.2017.01.018.

56. Navarre, C.B. and Pugh, D.G. (2002). Diseases of the Gastrointestinal System. *Elsevier eBooks*, [online] pp.69–105. DOI: https://doi.org/10.1016/b0-72-169052-1/50006-5.

57. Niegowska, M., Sanseverino, I., Navarro, A. and Lettieri, T. (2021). Knowledge gaps in the assessment of antimicrobial resistance in surface waters. *FEMS Microbiology Ecology*, 97(11). DOI: https://doi.org/10.1093/femsec/fiab140.

58. Oulas, A., Pavloudi, C., Polymenakou, P., Pavlopoulos, G.A., Papanikolaou, N., Kotoulas, G., Arvanitidis, C. and Iliopoulos, I. (2015). Metagenomics: Tools and Insights for Analyzing Next-Generation Sequencing Data Derived from Biodiversity Studies. *Bioinformatics and Biology Insights*, [online] 9, pp.75–88. DOI: https://doi.org/10.4137/BBI.S12462.

59. Pallis, A., Jazayeri, J., Ward, P., Dimovski, K. and Svobodova, S. (2013). Rapid detection of Clostridium difficile toxins from stool samples using real-time multiplex PCR. *Journal of Medical Microbiology*, 62(9), pp.1350–1356. DOI: https://doi.org/10.1099/jmm.0.058339-0.

60. Palmer, J., Hough, R., West, H.M. and Avery, L.M. (2019). A review of the abundance, behavior, and detection of clostridial pathogens in agricultural soils. European Journal of Soil Science, 70(4), pp.911–929. DOI: https://doi.org/10.1111/ejss.12847.

61. Pascoal, F., Costa, R. and Magalhães, C. (2020). The microbial rare biosphere: current concepts, methods, and ecological principles. *FEMS Microbiology Ecology*, 97(1). DOI: https://doi.org/10.1093/femsec/fiaa227.

62. Pérez-Cobas, A.E., Gomez-Valero, L. and Buchrieser, C. (2020). Metagenomic approaches in microbial ecology: an update on whole-genome and marker gene sequencing analyses. *Microbial Genomics*, 6(8). DOI: https://doi.org/10.1099/mgen.0.000409.

63. Popoff, M.R. (2018). Clostridium difficile and Clostridium sordellii toxins, proinflammatory versus anti-inflammatory response. *Toxicon*, 149, pp.54–64. DOI: https://doi.org/10.1016/j.toxicon.2017.11.003.

64. Posthaus, H., Kittl, S., Tarek, B. and Bruggisser, J. (2020). Clostridium perfringens type C necrotic enteritis in pigs: diagnosis, pathogenesis, and prevention. Journal of Veterinary Diagnostic

Investigation: Official Publication of the American Association of Veterinary Laboratory Diagnosticians, Inc, [online] 32(2), pp.203–212. DOI: https://doi.org/10.1177/1040638719900180.

65. Prakash, T. and Taylor, T.D. (2012). Functional assignment of metagenomic data: challenges and applications. *Briefings in Bioinformatics*, 13(6), pp.711–727. DOI: https://doi.org/10.1093/bib/bbs033.

66. Prasad, M., Brar, B., Kiran, B. and Singh, N. (2023). Emerging Microbial Technologies. *Indian Journal of Microbiology*, 63(3), pp.231–234. DOI: https://doi.org/10.1007/s12088-023-01103-7.

67. Rashid, M. and Stingl, U. (2015). Contemporary molecular tools in microbial ecology and their application to advancing biotechnology. *Biotechnology Advances*, 33(8), pp.1755–1773. DOI: https://doi.org/10.1016/j.biotechadv.2015.09.005.

68. Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., ... & Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences, 108(Supplement 1), 4516-4522.

69. Gilbert, J. A., Jansson, J. K., & Knight, R. (2014). The Earth Microbiome Project: successes and aspirations. BMC biology, 12(1), 69.

70. Schloss, P. D., Gevers, D., & Westcott, S. L. (2011). Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PloS one, 6(12), e27310.

71. Ravin, N.V., Mardanov, A.V. and Skryabin, K.G. (2015). Metagenomics as a tool for the investigation of uncultured microorganisms. *Russian Journal of Genetics*, 51(5), pp.431–439. DOI: https://doi.org/10.1134/s1022795415050063.

72. Riesenfeld, C.S., Schloss, P.D. and Handelsman, J. (2004). Metagenomics: Genomic Analysis of Microbial Communities. *Annual Review of Genetics*, 38(1), pp.525–552. DOI: https://doi.org/10.1146/annurev.genet.38.072902.091216.

73. Robbins, R.J., Krishtalka, L. and Wooley, J.C. (2016). Advances in biodiversity: metagenomics and the unveiling of biological dark matter. *Standards in Genomic Sciences*, 11(1). DOI: https://doi.org/10.1186/s40793-016-0180-8.

74. Schnizlein, M.K. and Young, V.B. (2022). Capturing the environment of the Clostridioides difficile infection cycle. *Nature Reviews Gastroenterology & Hepatology*, [online] 19(8), pp.508–520. DOI: https://doi.org/10.1038/s41575-022-00610-0.

75. Shen, A. (2012). *Clostridium difficile* Toxins: Mediators of Inflammation. *Journal of Innate Immunity*, 4(2), pp.149–158. DOI: https://doi.org/10.1159/000332946.

76. Shen, A. (2020). Clostridioides difficile Spore Formation and Germination: New Insights and Opportunities for Intervention. *Annual Review of Microbiology*, 74(1), pp.545–566. DOI: https://doi.org/10.1146/annurev-micro-011320-011321.

77. Shrestha, A., Uzal, F.A. and McClane, B.A. (2018). Enterotoxic Clostridia: Clostridium perfringens Enteric Diseases. Microbiology Spectrum, [online] 6(5). DOI: https://doi.org/10.1128/microbiolspec.gpp3-0003-2017.

78. Spigaglia, P. (2015). Recent advances in the understanding of antibiotic resistance in Clostridium difficile infection. *Therapeutic Advances in Infectious Disease*, [online] 3(1), pp.23–42. DOI: https://doi.org/10.1177/2049936115622891.

79. Spigaglia, P., Barbanti, F., Faccini, S., Vescovi, M., Criscuolo, E.M., Ceruti, R., Gaspano, C. and Rosignoli, C. (2023). Clostridioides difficile in Pigs and Dairy Cattle in Northern Italy: Prevalence, Characterization and Comparison between Animal and Human Strains. *Microorganisms*, [online] 11(7), p.1738. DOI: https://doi.org/10.3390/microorganisms11071738.

80. Tijerina-Rodríguez, L., Villarreal-Treviño, L., Morfín-Otero, R., Camacho-Ortíz, A. and Garza-González, E. (2019). Virulence Factors of Clostridioides (Clostridium) difficile Linked to Recurrent Infections. *Canadian Journal of Infectious Diseases and Medical Microbiology*, 2019, pp.1–7. DOI: https://doi.org/10.1155/2019/7127850.

81. Timmis, K. and Ramos, J.L. (2021). The soil crisis: the need to treat as a global health problem and the pivotal role of microbes in prophylaxis and therapy. *Microbial Biotechnology*, 14(3), pp.769–797. DOI: https://doi.org/10.1111/1751-7915.13771.

82. Tschudin-Sutter, S., Kuijper, E.J., Durovic, A., Vehreschild, M.J.G.T., Barbut, F., Eckert, C., Fitzpatrick, F., Hell, M., Norèn, T., O'Driscoll, J., Coia, J., Gastmeier, P., von Müller, L., Wilcox, M.H., Widmer, A.F., Allerberger, F., Cornely, O.A., Delmée, M., Olesen, B. and van Broeck, J. (2018). Guidance document for prevention of Clostridium difficile infection in acute healthcare settings. *Clinical Microbiology and Infection*, [online] 24(10), pp.1051–1054. DOI: https://doi.org/10.1016/j.cmi.2018.02.020.

83. Wang, Z., McCloskey, A., Cheng, S., Wu, M., Xue, C., Yu, Z., Fu, J., Liu, Y., Luo, Z.-Q. and Liu, X. (2018). Regulation of the small GTPase Rab1 function by a bacterial glucosyltransferase. *Cell Discovery*, 4(1). DOI: https://doi.org/10.1038/s41421-018-0055-9.

84. Wee, Y., Bhyan, S.B., Liu, Y., Lu, J., Li, X. and Zhao, M. (2018). The bioinformatics tools for genome assembly and analysis are based on third-generation sequencing. *Briefings in Functional Genomics*, [online] 18(1), pp.1–12. DOI: https://doi.org/10.1093/bfgp/ely037.

85. Zhang, L., Chen, F., Zeng, Z., Xu, M., Sun, F., Yang, L., Bi, X., Lin, Y., Gao, Y., Hao, H., Yi, W., Li, M. and Xie, Y. (2021). Advances in Metagenomics and Its Application in Environmental Microorganisms. *Frontiers in Microbiology*, 12. DOI: https://doi.org/10.3389/fmicb.2021.766364.

86. Squire, M. M., & Riley, T. V. (2012). Clostridium difficile Infection in Humans and Piglets: A 'One Health' Opportunity. In J. Mackenzie, M. Jeggo, P. Daszak, & J. Richt (Eds.), One Health: The Human-Animal-Environment Interfaces in Emerging Infectious Diseases (pp. 237–251). Current Topics in Microbiology and Immunology, 365. Springer. https://doi.org/10.1007/82_2012_237

87. Takahashi, S., Tomita, J., Nishioka, K., Hisada, T., & Nishijima, M. (2014). Development of a prokaryotic universal primer for simultaneous analysis of Bacteria and Archaea using next-generation sequencing. PLoS ONE, 9(8), e105592. https://doi.org/10.1371/journal.pone.0105592

88. Bolyen, E., Rideout, J. R., Dillon, M. R., et al. (2019). Reproducible, interactive, scalable, and extensible microbiome data science using QIIME 2. Nature Biotechnology, 37(8), 852-857. https://doi.org/10.1038/s41587-019-0209-9

89. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. Journal, 17(1), 10. https://doi.org/10.14806/ej.17.1.200

90. Jiang, R., Li, W. V., & Li, J. J. (2020). mbImpute: an accurate and robust imputation method for microbiome data. Genome Biology, 21(1), 1-12. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-021-02400-4.

91. Werner, M. C., et al. (2012). A novel method for classifying metagenomic sequences using a naive Bayes classifier. BMC Bioinformatics, 13(1), 1.

92. Chen, L., Reeve, J., Zhang, L., Huang, S., Wang, X., & Chen, J. (2018). GMPR: A robust normalization method for zero-inflated count data with application to microbiome sequencing data. PeerJ, 6, e4600. https://doi.org/10.7717/peerj.4600

93. Latham, R. E. (2004). [Review of Measuring Biological Diversity, by A. E. Magurran]. Journal of Vegetation Science, 15(6), 854–856. http://www.jstor.org/stable/4096686

94. Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Stevens, M. H. H., Oksanen, M. J., & Suggests, M. (2007). The vegan package. Commun Ecol Pack, 10, 631–637.

95. Magurran, A. E. (1988). Ecological diversity and its measurement. Springer.

96. Simpson, E. H. (1949). Measurement of diversity. Nature, 163(4148), 688.

97. Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. Biological Conservation, 61(1), 1-10.

98. Pielou, E. C. (1966). The measurement of diversity in different types of biological collections. Journal of Theoretical Biology, 13, 131-144.

99. Hurlbert, S. H. (1971). The non-concept of species diversity: A critique and alternative parameters. Ecology, 52(4), 577-586.

100. Yu Y, Zhang M, Liu J, Xu B, Yang J, Wang N, Yan S, Wang F, He X, Ji G, Li Q, Miao L. Long Non-coding RNA PVT1 Promotes Cell Proliferation and Migration by Silencing ANGPTL4 Expression in Cholangiocarcinoma. Mol Ther Nucleic Acids. 2018 Dec 7;13:503-513. doi: 10.1016/j.omtn.2018.10.001. Epub 2018 Oct 10. PMID: 30388624; PMCID: PMC6205330.

101. Henderson SE, Lowe JR, Tudares MA, Gold MS, Almarza AJ. Temporomandibular joint fibrocartilage degeneration from unilateral dental splints. Arch Oral Biol. 2015 Jan;60(1):1-11. doi: 10.1016/j.archoralbio.2014.08.022. Epub 2014 Sep 6. PMID: 25247778; PMCID: PMC4252753.