ESCI upf.
School of International Studies

MEMÒRIA DEL TREBALL DE FI DE GRAU DEL GRAU (ESCI-UPF)

Unravelling the Nasal Microbiome of Piglets Through Whole Genome Metagenomics Sequencing

AUTOR/A: Oriol Castellano Aguilera	**NIA:** 106470
GRAU: Bachelor's Degree in Bioinformatics	
CURS ACADÈMIC: 3r	
DATA: 18 de juny de 2024	
TUTOR/S: Florencia Correa Fiz	

FULL DE RESUM DEL TREBALL DE FI DE GRAU DEL BDBI (ESCI-UPF)

TÍTOL DEL PROJECTE: Unravelling the Nasal Microbiome of Piglets Through Whole Genome Metagenomics Sequencing	
AUTOR/A: Oriol Castellano Aguilera	**NIA:** 106470
CURS ACADÈMIC: 3r	
DATA: 18 de juny del 2024	
TUTOR/S: Florencia Correa Fiz	

PARAULES CLAU (mínim 3)

- Català: Metagenòmica, microbioma, viroma, probiòtics, secretoma, garrins.
- Castellà: Metagenómica, microbioma, viroma, probióticos, secretoma, lechones.
- Anglès: Metagenomics, microbiome, virome, probiotics, secretome, piglets.

RESUM DEL PROJECTE (extensió màxima: 100 paraules per llengua)

- Català:

El microbioma nasal dels garrins, que inclou bacteris, arquees, fongs i virus, impacta significativament en la seva salut i resistència a malalties. El seqüenciat metagenòmic del genoma complet de mostres nasals de porcellons sotmesos a diversos tractaments, com ara probiòtics i antibiòtics, va revelar canvis significatius en la composició de la comunitat microbiana amb el temps. L'anàlisi feta amb l'eina MetaWRAP i eines com Kraken2, PROKKA i SignalP va identificar canvis notables en taxons bacterians, arqueus, fúngics i virals, predominant els bacteriòfags en el viroma. Es van observar diferències significatives en la diversitat microbiana entre els tractaments i els dies. L'estudi també va destacar perfils secretòmics diferents entre els grups, suggerint un paper en la salut i resistència a malalties dels porcellons.

- Castellà:

El microbioma nasal de los lechones, que incluye bacterias, arqueas, hongos y virus, impacta significativamente en su salud y resistencia a enfermedades. El secuenciado metagenómico completo del genoma de disparo de muestras nasales de lechones sometidos a diversos tratamientos, como probióticos y antibióticos, reveló cambios significativos en la composición de la comunidad microbiana con el tiempo. El análisis utilizando el paquete MetaWRAP identificó cambios notables en taxones bacterianos, arqueas, fúngicos y virales, predominando los bacteriófagos en el viroma. Se observaron diferencias significativas en la diversidad microbiana entre los tratamientos y los días. El estudio también destacó perfiles secretómicos distintos entre los grupos, sugiriendo un papel en la salud y resistencia a enfermedades de los lechones.

- Anglès:

The nasal microbiome of piglets, including bacteria, archaea, fungi, and viruses, significantly impacts their health and disease resistance. Whole-genome shotgun metagenomic sequencing of nasal samples from piglets subjected to various treatments, such as probiotics and antibiotics, revealed significant shifts in microbial community composition over time. Analysis using the MetaWRAP pipeline identified notable changes in bacterial, archaeal, fungal, and viral taxa, with bacteriophages predominating the virome. Significant differences in microbial diversity were observed across treatments and days. The study also highlighted distinct secretome profiles among groups, suggesting a role in piglet health and disease resistance.

Unravelling the Nasal Microbiome of Piglets Through Whole Genome Metagenomics Sequencing

Oriol Castellano Aguilera
Bachelor's Degree in Bioinformatics (UPF-UPC-UB-UAB)

Scientific director: Florencia Correa Fiz
IRTA-CReSA
June 2024

Abstract

The nasal microbiome of piglets, comprising bacteria, archaea, fungi, and viruses, plays a pivotal role in their health and disease resilience. In swine production, microbial imbalances can lead to diseases such as polyserositis, often managed through antibiotic treatments which can perturb the microbiota. Whole-genome shotgun metagenomic sequencing was employed on nasal samples collected at different time points from different groups of piglets subjected to varying treatments, including the inoculation of probiotics, and antibiotics, to unravel the microbiome of piglet nasal cavities, encompassing bacteria, archaea, fungi, and, for the first time, the virome.

Samples were analyzed using the MetaWRAP pipeline. The process encompassed quality control and preprocessing and read assembly. Taxonomic profiling and diversity analysis were performed using Kraken 2, complemented with further statistical analyses. Metagenomic binning, quantification, and classification of the metagenome-assembled genomes were conducted, finally, functional annotation of genomes and proteins was performed using PROKKA and SignalP, respectively.

The results revealed significant differences in microbial community composition between days, with notable shifts in bacterial, archaea, fungal, and viral taxa. Bacteriophages, particularly from the Caudoviricetes class, were predominant in the virome. Significant differences in alpha and beta diversity were observed between different treatments and sampling days, indicating the impact of environmental factors and treatment regimens on the microbiome.

Additionally, the research aimed to evaluate differences in the bacterial secretome among piglets to understand its potential role in shaping piglet health and disease resistance. The study found distinct secretome profiles among groups, relating them to the secretome of key microbial colonizers, with variations in secreted protein abundance between days and groups.

Keywords: *Metagenomics, microbiome, virome, probiotics, secretome, piglet.*

ESCI upf.
School of International Studies
IRTA CReSA
Centre de Recerca en Sanitat Animal
World Organisation for Animal Health
Founded as OIE
Collaborating Center of the WOAH for Research and control of emerging and re-emerging pig diseases in Europe

1 Introduction

Comprehensive research history has shown that only a small proportion of microorganisms are directly associated with pathogenicity [1]. The microbiome, comprises bacteria, archaea, fungi, and viruses [2]. Given this, microbiota comprises all living members forming the microbiome [3]. The profound influence of the microbiome extends beyond the mere presence of these microorganisms; influences the well-being of organisms, leaving an indelible mark on their metabolism, and immune systems, thus sculpting the delicate equilibrium of life.

As bacteria are the most abundant component of the microbiota, the vast majority of studies over the last decades have focused primarily on the composition of bacterial microbiota and its effects [4].

Within this microscopic landscape, the virome emerges. The term "virome" comes from "virus" and "genome" and is used to describe viral shotgun metagenomes consisting of a collection of nucleic acids associated with this particular ecosystem [1].

Characterizing viral communities remains a challenge. The primary hurdle lies in the absence of a universal gene shared across all viral genomes [5], a feature that distinguishes viruses from other microorganisms. In contrast to the methods employed for bacterial ribosomal DNA profiling, which are applicable across a wide spectrum of microorganisms, the lack of a common genetic marker prevents the development of analogous approaches for viruses. This impediment amplifies the intricacy of deciphering the virome. Compounding this challenge, together with the low amount of both free and cellular DNA, makes deciphering the virome more complex. The elusive nature of viral genetic material poses a significant obstacle in the quest to understand the microscopic world fully. Bacteriophages, also known as phages, are viruses that infect and replicate within bacteria [6], come in all shapes and sizes with genomes consisting of either double-stranded or single-stranded DNA or RNA [7]. These viral entities, which constitute the most prevalent biological entities on Earth [8], with an estimated extension of 10^{31} total phage particles [9], transcend their conventional association with pathogenicity [10]. Instead, they contribute significantly to the complexity of microbial interactions. Recent studies have showcased the potential of bacteriophages in disease control, where phages are explored for their application in phage therapy as a promising alternative to traditional antibiotics.

Previous studies have given profitable bits of knowledge into the particular microbial constituents of the pig nasal composition, using 16S amplicon sequencing and whole genome shotgun (WGS) metagenomics, essentially being focused on the bacteria and archaea present in the nasal cavity of piglets [11]. Otherwise, most previous studies of the microbiome in pigs, are carried out in the gut [12] or fecal samples [13].

In the context of swine production stages, can be divided into 4 phases; gestation, farrowing (birth to weaning), nursery, and growing-finishing. Weaning is an urgent stage in the advancement of pigs, marking the transition from dependence on maternal support to free reinforcement. This period is characterized by notable physiological and natural changes, with suggestions for pig health, resilient advancement, and overall strength. The nasal microbiome, associated with the respiratory system, acts as the first line of defense of the organism [14], undergoes notable changes during this transition, as weaning can be a stress-inducing stage with potential repercussions on the pig's health [11]. Since the advent of antimicrobial agents, bacterial diseases have traditionally been managed through their application. The high efficacy of antimicrobials has led to a diminished emphasis on the development and utilization of vaccines against these infections [15]. Antibiotics have been used routinely in farm animal production since the 1950s to treat, control, prevent disease and to increase productivity [16]; however, the widespread and indiscriminate use of antibiotics for managing bacterial diseases, often in the form of metaphylaxis, has raised concerns due to the emergence of antimicrobial-resistant bacteria.

As an alternative strategy to enhance the well-being of piglets and consequently reduce the reliance on antibiotics in swine production, a potential approach involves employing microorganisms designed to confer advantages to the host and thereby exclude pathogens. While probiotics are commonly employed in human health practices, their utilization in pigs is less prevalent, primarily focusing on the gut microbiota [17] [18]. In addition to their established role in gut health, probiotics, defined as "live microorganisms which, when administered in adequate amounts, confer a health benefit on the host" [19], show promise in controlling pathogens in swine. Probiotics can confer advantages by enhancing host immunity excluding pathogens and providing several physiological, functional, nutritional, and immunological advantages [20]. Notably, the exploration of probiotics in the respiratory tract has been limited to a few studies where their effectiveness against infections associated with the respiratory system was demonstrated [21] [22] [23] [24].

The secretome encompasses all secreted proteins

and secretory machinery of microorganisms. It constitutes a vital aspect of microbial function and interaction with their environment. Microbial secretome proteins, which encompass receptors, transporters, enzymes, toxins, and virulence factors, are essential for various cellular processes such as nutrient acquisition, communication, and defense mechanisms [25]. The bacterial secretome plays crucial roles in adaptation, virulence, pathogenesis, and host interactions [25]. The study of bacterial secretomes has led to the identification of novel antigens and biomarkers, facilitating early diagnosis and the development of effective vaccines. Despite significant progress in understanding bacterial secretomes, there are still gaps in knowledge.

Next-generation sequencing permits DNA sequencing providing high-throughput, adaptability and speed. Within the scene of microbial investigation, advances have driven to a more profound understanding of complex environments.

The rise of metagenomics has impacted microbial investigation, giving a more comprehensive approach by sequencing all the genetic material within a sample. Metagenomics includes the coordinate sequencing of DNA extracted from environmental samples, eliminating the requirement for culturing individual species. Thus, aims to be especially significant in examining complex biological systems where a considerable parcel of microbial differences may be unculturable. On the other hand, 16S rDNA sequencing focuses specifically on the hypervariable regions of the 16S rDNA gene, found in all Bacteria and Archaea, which are used to identify and classify taxa based on their genetic differences, providing insights into the diversity and relative abundance of bacterial species within a sample.

1.1 Objectives

This study aims to investigate the comprehensive microbiome of piglet nasal cavities, including bacteria, archaea, fungi, and, for the first time, the virome, through WGS metagenomic sequencing of samples collected on days 21 and 49. The samples come from groups of piglets differentiated based on the treatment they have received, capturing the impact of post-weaning, probiotics, and antibiotics. The study will compare whether there are differences in the taxonomic diversity of the samples between or within groups and days.

Additionally, the research aims to evaluate differences in the bacterial secretome among piglets to understand its potential role in shaping piglet health and disease resistance and compare these findings with the secretome profiles of identified microbial colonizers. The study will also assess the effectiveness and correlation of the taxonomy found through a previous study that used 16S amplicon sequencing versus WGS metagenomics squencing to analyze these complex microbial communities.

2 Materials and Methods

2.1 Design

A study was previously performed in a 2-site pig commercial farm with recurrent respiratory problems. During the first three weeks of life, piglets experienced the period of lactation in the first facility; after weaning, piglets were transferred to the nursery, where they stayed until nine weeks of age [11].

Piglets were categorized based on their maternal treatment, piglets born to non-treated sows served as the control group and remained untreated (NTS-NTP). At birth (D0), piglets born to treated sows remained untreated (TS-NTP), while piglets born to non-treated sows were subjected to intramuscular treatment with 1 mg of crystalline ceftiofur (NTS-TP). Furthermore, to explore the impact of early colonization with specific bacterial strains on piglet nasal microbiota, piglets were inoculated on D0 with a nasal spray containing a cocktail of five bacterial strains belonging to distinct species: *Vagococcus lutrae*, *Streptococcus pluranimalium*, *Moraxella pluranimalium*, *Rothia nasimurium*, and *Glaesserella parasuis*. The inoculated piglets were non-treated and derived from two groups: piglets born to ceftiofur-treated sows (TS-IP), and piglets born to non-treated sows (NTS-IP). Table 1 provides a representation of the categorized groups. During the nursery phase, any animals displaying clinical signs were treated with antibiotics (1mL/10kg of Gentamox ®; 150 mg of amoxicillin, and 40 mg of gentamicin per mL) and were subsequently excluded from the study.

2.2 Sample collection

Nasal swabs were collected from piglets at different time points as part of the study by M. Blanco et al. [11], where DNA extracted from the nasal swabs at weaning (D21) and at the end of the nursery period (D49) was employed for both 16S rDNA sequencing and WGS metagenomic analysis, using Illumina Novaseq 6000 technology.

In this study, building on the samples collected in the previous study, the genomic data underwent analysis employing various tools within the metaWRAP pipeline [26].

Table 1: **Representation of the groups.**

Sow treatment	Piglet treatment	Group
Treated Sow (TS)	Non-treated piglet (NTP)	TS-NTP
	Inoculated piglet (IP)	TS-IP
Non-treated Sow (NTS)	Inoculated piglet (IP)	NTS-IP
	Treated piglet (TP)	NTS-TP
	Non-treated piglet (NTP)	NTS-NTP*

*Control group.

2.3 Quality control and processing

Quality control and preprocessing were conducted on the raw sequence reads to ensure the integrity and accuracy of the data. Trim-Galore [27], which integrates Cutadapt [28] and FastQC [29], was employed to trim and preprocess the sequences, removing low-quality bases and adapter sequences. Additionally, bmtagger [30] was employed alongside the *Sus scrofa* reference genome v11.1 [31] to systematically eliminate any reads corresponding to the host genome. Quality reports were generated for each sequenced sample to ensure the integrity and accuracy of the data.

2.4 Read assembly

The assembly of high-quality, preprocessed reads was executed utilizing MetaSPAdes [32]. The resultant contigs emerged as integral elements for further exploration. These contigs, akin to genetic building blocks, encapsulate the genetic material retrieved from our metagenomic samples.

2.5 Taxonomic profiling and diversity analysis

Taxonomic profiling of the metagenome-assembled contigs was performed using Kraken 2 [33]. Krona [34] was used to visualize and generate plots of all the taxa found in the metagenomic samples. Alpha and beta diversity analyses were performed to assess differences between groups after weaning (D21) and the final extraction day (D49), and also inspect the diversity within each day. Kruskal-Wallis test [35], was performed to evaluate the potentially significant impact of distinct treatments on alpha diversity indices across piglet groups. PERMANOVA was employed to examine the dissimilarity in microbial communities between samples, considering both within-group and between-group variations.

2.6 Metagenomic fragment binning and quantification

Binning is the process of grouping together fragments of sequenced DNA that are believed to originate from the same genome, thus reconstructing the genomes of individual microbial species from metagenomic data. In this study, the assembly was binned with the metagenomic binning software metaBAT2 [36]. Following binning, the quality of the resulting metagenome-assembled genomes (MAGs) was evaluated using CheckM [37] to ensure their completeness and reduce contamination. Abundance quantification of the MAGs was performed using Salmon [38] to determine the distribution and relative abundance of the reconstructed genomes across metagenomic samples. Taxonomic classification provided information on the identity of the organisms represented by the MAGs.

2.7 Classification and annotation of MAGSs for protein prediction

Functional annotation of genomes and proteins was conducted using PROKKA [39] to explore potential roles and metabolic capabilities within the bacterial community. Subsequently, SignalP [40] was employed to predict secreted proteins

2.8 Comparative Analysis: 16S rDNA vs. Metagenomics

Microbial community profiles from 16S rDNA gene sequencing, based on a previously published study [11], were compared with those obtained from WGS metagenomic sequencing. Linear modeling was employed to identify differences between the methods.

2.9 Data Processing and R analysis

Subsequent analyses of the outputs, data preparation, table creation, and plotting were conducted using RStudio [41]. The following packages were utilized: vegan [42], ggplot2 [43], dplyr [44], tidyr [45], and stringr [46].

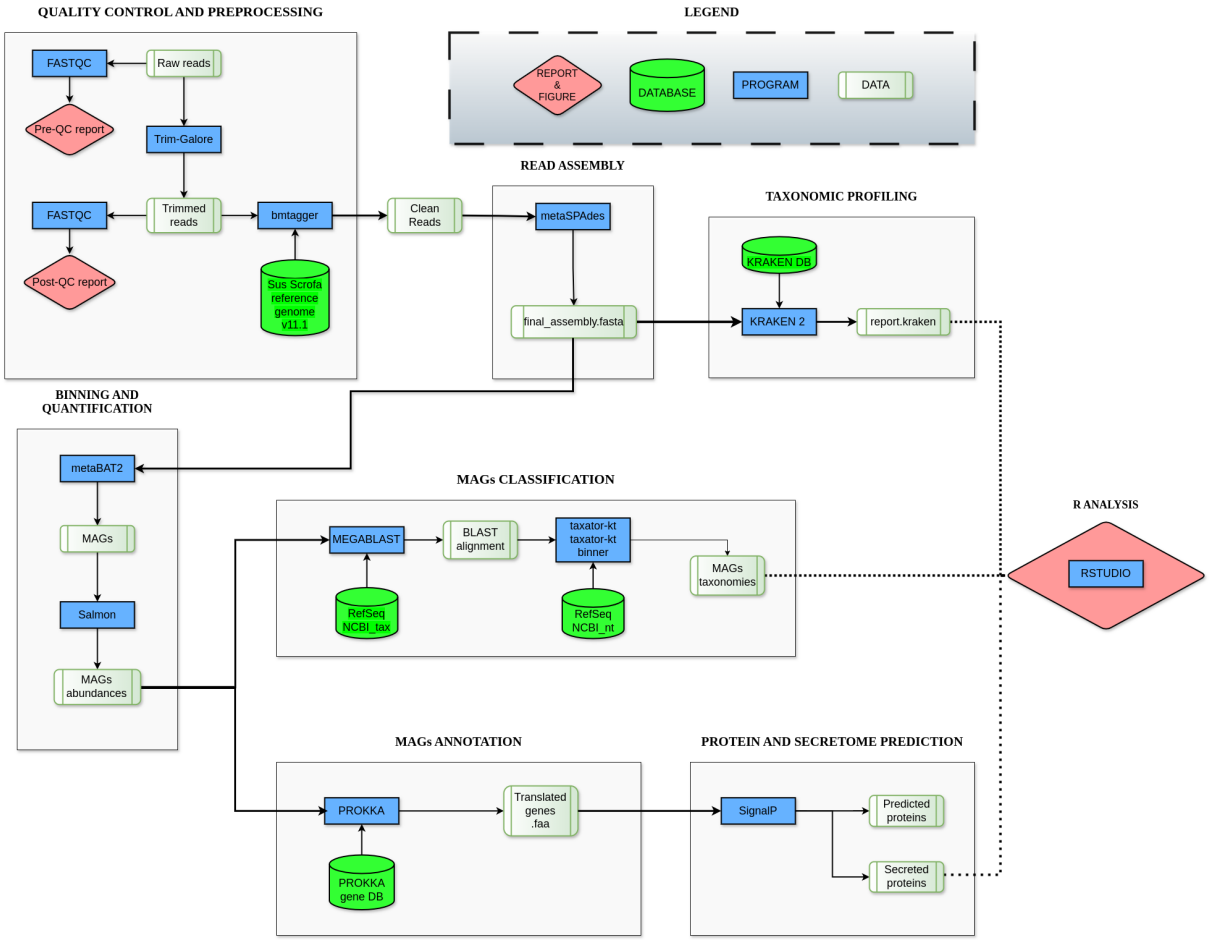


Figure 1: **Overview of the workflow used in the metagenomic analysis**. Detailing programs, databases, data, reports, and key steps, including previous mentioned steps, as quality control, assembly, taxonomic profiling, metagenomic fragment binning, MAGs quantification, classification, annotation for protein prediction, finishing with data analysis in RStudio.

3 Results and discussion

3.1 Dataset description

Using sequences obtained from the study by *M. Blanco et al. 2023* [11], 48 nasal swab samples from five different groups of piglets (see Table 1) were analyzed. Quality filtering and preprocessing steps were applied to the samples, resulting in clean reads that excluded *Sus scrofa* reference genome v11.1. The subsequent assembly of these reads produced metagenome-assembled contigs suitable for taxonomic profiling.

3.2 Taxonomic profiling of metagenome-assembled contigs

Taxonomic profiling of each metagenome-assembled contig was done using Kraken 2, providing comprehensive reports detailing the classification of taxa present in the metagenomic samples. These reports were essential in providing a detailed breakdown of the microbial taxa identified, along with their corresponding abundance levels. The output files, which included classified reads and their taxonomic labels, were subsequently imported into RStudio for further processing and analysis. Figure 2 represents the taxonomic profile of one sample, which followed the patterns of all data, just to show that Bacteria was the predominant taxa found in the samples, representing a mean percentage of 56.47% across them, followed by fungi, which represented the 1.86%, and viruses and archaea which represented the 1.32% and 1.26%, respectively.

Figure 2: **KRONA plot generated from sample F112 taken on D21**. The representation visually organizes the microbial diversity detected. It provides a hierarchical view of taxonomic classifications, aiding in understanding the distribution and prevalence of different microbial groups present in the sample. Segments are color-coded and nested, starting from the broadest classification (superkingdom) down to finer taxonomic levels (family).

3.3 Virome

Viruses are obligate intracellular parasites, relying on host cells for replication [47]. Their abundance is inherently limited by the availability and susceptibility of host cells within the microbiota. Bacteriophages were the predominant viruses identified in the samples (Supplementary Figure 1), with the class *Caudoviricetes* notably comprising almost 84% of all viruses detected (Supplementary Figure 2). Bacteria and phages have a long history of coevolution, leading to the development of diverse mechanisms for phages to counteract bacterial resistance. These mechanisms often involve minimal fitness costs [48]. Furthermore, individual phages can harbor multiple resistance systems, which together provide robust defenses, allowing specific bacterial clones to thrive in phage-rich environments [48]. Thus, the high prevalence of bacteriophages can be explained by the substantial abundance of bacteria in the piglet nasal microbiota.

Alpha diversity, measured by the Shannon index, was done to asses possible variations within and between different piglet groups. The absence of substantial differences is visually evident in Figure 3. Statistical validation through the Kruskal-Wallis test (p-value = 0.2661) indicated no statistically significant variations in alpha diversity among the studied groups. Additionally, pairwise comparisons between groups did not reveal any statistically significant differences, suggesting a level of homogeneity in alpha diversity metrics, emphasizing the consistency in the viral species composition within each group.

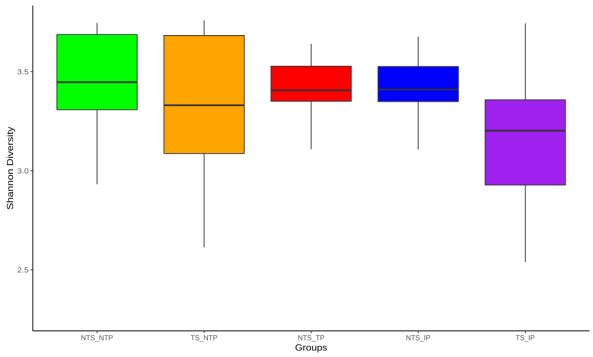


Figure 3: **Virome alpha diversity boxplot of the groups under study**. The boxplot summarizes the diversity across samples based on the Shannon index, indicating the distribution and variability of species diversity within each group.

Beta diversity, examined through Bray-Curtis (BC) dissimilarity, focused on shared abundances of viruses between samples. Figure 4 showcase principal coordinates analysis (PCoA) representations for D21 and D49, respectively. In agreement with alpha diversity, PERMANOVA indicated no statistically significant differences in viral composition between groups on both days (p-value = 0.718 and 0.085 for days 21 and 49, respectively). However, there were significant differences in beta diversity between days (p-value = 0.001), suggesting distinct viral community composition after weaning (D21) and after nursery (D49), that can be attributed to changes in environmental exposures, immune maturation of piglets [11], or alterations in microbial interactions within the nasal microbiome over time [49] [50].

3.4 Archaea and Fungi

Archaea constituted the least abundant taxa within the nasal microbiome of piglets. Alpha diversity (Shannon) revealed statistically significant differences of TS-IP with NTS-TP and NTS-IP (*Supplementary Figure 4*). Treatment of the sow likely influenced the initial microbiota transferred to the piglets [51], creating a unique microbial environment that affected archaea diversity. Piglet inoculation may have introduced specific microorganisms that have interacted differently with archaea, leading to variations in their community composition. The combination of sow treatment and piglet inoculation may have created a distinct microenvironment [52] that has supported or suppressed certain archaeal taxa.

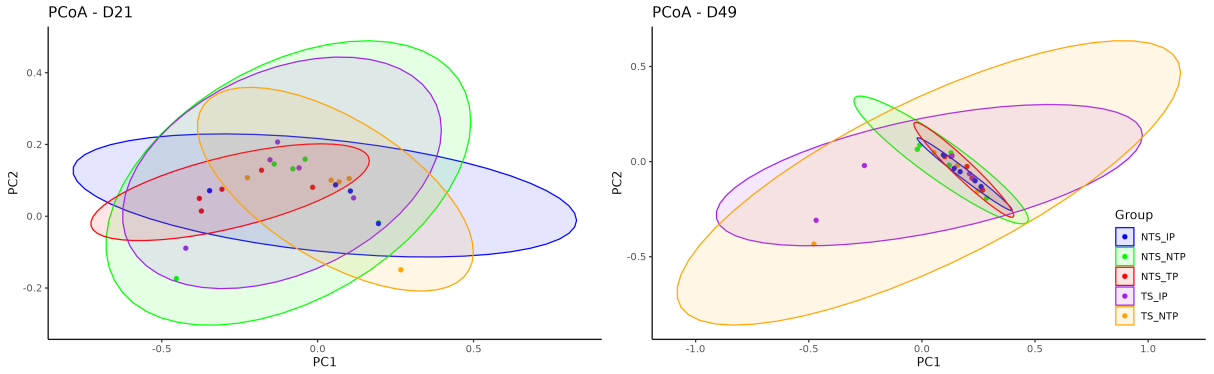
Different sow and piglet treatments have altered the fungal diversity and abundance. Fungi revealed statistically significant differences in alpha diversity. TS-IP significantly differed from NTS-IP and NTS-TP; and the control group (NTS-NTP) showed significant differences compared to NTS-IP (Supplementary Figure 6).

In beta diversity analysis (BC), both archaea and fungi showed similar patterns (Supplementary Figures 5 and 7). There were no significant differences in community structures between groups, indicating a high degree of similarity in microbial composition regardless of treatment. otherwise. However, significant differences were observed in the composition between sampling days

3.5 Bacteria

The bacterial composition within the nasal microbiota of piglets mirrored the results previously published by *M. Blanco et al. (2023)* [11], confirming a similarly rich and diverse array of bacterial communities. At D21, the predominant genera across all groups were *Glaesserella*, *Moraxella*, *Bergeyella*, *Mycoplasma*, *Streptococcus*, *Pasteurella*, *Neisseria*, *Mannheimia*, *Lactobacillus*, and *Acinetobacter*. By D49, species typically associated with the gut microbiota, which were absent at weaning, became prominent. Significant variations in microbial diversity and composition were found between D21 and D49 (p-value = 0.001) (Supplementary Figure 10).

Although no significant differences in Shannon diversity were found between groups based on the Kruskal-Wallis test (p-value = 0.723) (Supplementary Figure 9). However, pairwise tests indicated that the TS-IP group differed statistically from the others. Furthermore, no significant differences in bacterial compositions between groups at D21 (p-value = 0.106) and D49 (p-value = 0.16) were observed (*Supplemetary Figure 10*). These findings are consistent with previous studies, from a predominance of *Firmicutes*, *Proteobacteria*, and *Bacteroidetes*, considered as the most prevalent organisms found in the nasal cavity [53], which noted



(a) PCoA at D21

(b) PCoA at D49

Figure 4: **Principal Coordinate Analysis (PCoA) plots of viral communities on different days.** Plots illustrate the variation in viral community structure at two different time points: D21 (left) and D49 (right). Each point represents a sample, colored by group membership, with its position reflecting the similarity to other samples based on viral community composition. Ellipses represented the 95% confidence interval.

a transition at weaning, to a microbiota with significant contributions from *Clostridiales* and other gut-associated taxa by D49. This shift was likely driven by the change from milk to solid feed and increased exposure to environmental microbes in the post-weaning barn. Higher animal density in the barn likely exacerbates air quality, contributing to nasal inflammation and allowing the presence of anaerobic taxa. This mirrors human studies, where increased anaerobic bacteria in the sinuses correlate with inflammation and chronic diseases [11].

3.6 Secretome

The MAGs abundances were initially analyzed to understand the distribution of microbial species across different samples. Figure 5 represents the abundance patterns of MAGs at different time points (D21 and D49), highlighting the distinct clustering of samples by days based on their abundance profiles. This can be associated with the previous taxonomic results, which showed distinct patterns at the end of the nursery phase compared to after weaning, indicating that the composition and diversity of the microbiome changed longitudinally [54].

MAGs abundances (Supplementary Figure 11) were compared with the predicted secreted proteins to assess the relationship between microbial community composition and functional potential. This step highlighted three of them with notable prevalence and higher abundance than others, suggesting their potential dominance within the sampled ecosystem. These findings align with taxonomic classification, which revealed diverse taxonomic compositions across MAGs. Notably, one of them, identified as *G. parasuis*, displayed a significant abundance, indicating its potential role as a key colonizer in the studied environment. Furthermore, taxonomic profiling unveiled the presence of a MAGs classified as a virus, characterized by the taxonomic hierarchy: *Caudoviricetes*, *Caudovirales*, *Myoviridae*. This viral class emerged as the most prevalent virome taxonomic group, suggesting its importance in shaping the bacterial composition [55] within the ecosystem. MAGs abundances were compared with the number of predicted secreted proteins showing no significant differences. Analysis of secreted proteins across samples unveiled several observations. Despite overall consistency in predicted protein abundance across MAGs, significant differences were observed in the frequency of secreted proteins between days ($p = 0.024$) (Supplementary Figure 12), with D49 exhibiting higher frequencies compared to D21 (Supplementary Figure 15). Additionally, significant differences between groups were observed at D49, highlighting temporal variations in group-specific protein secretion patterns. Among the identified colonizers, *S. pluranimalium*, *M. pluranimalium*, and *G. parasuis* emerged as prominent taxa, indicating their ecological significance within the microbial community. Although *Rothia* anticipated as a potential colonizer, its absence in the MAGs taxonomic profiles (binning) indicates a different ecological niche or a lower abundance within the sampled environment.

An in-depth analysis of the predicted secreted proteins from the colonizers was performed, taking as input their genomes. *S. pluranimalium*, *M. pluranimalium*, and *G. parasuis*, revealed distinct

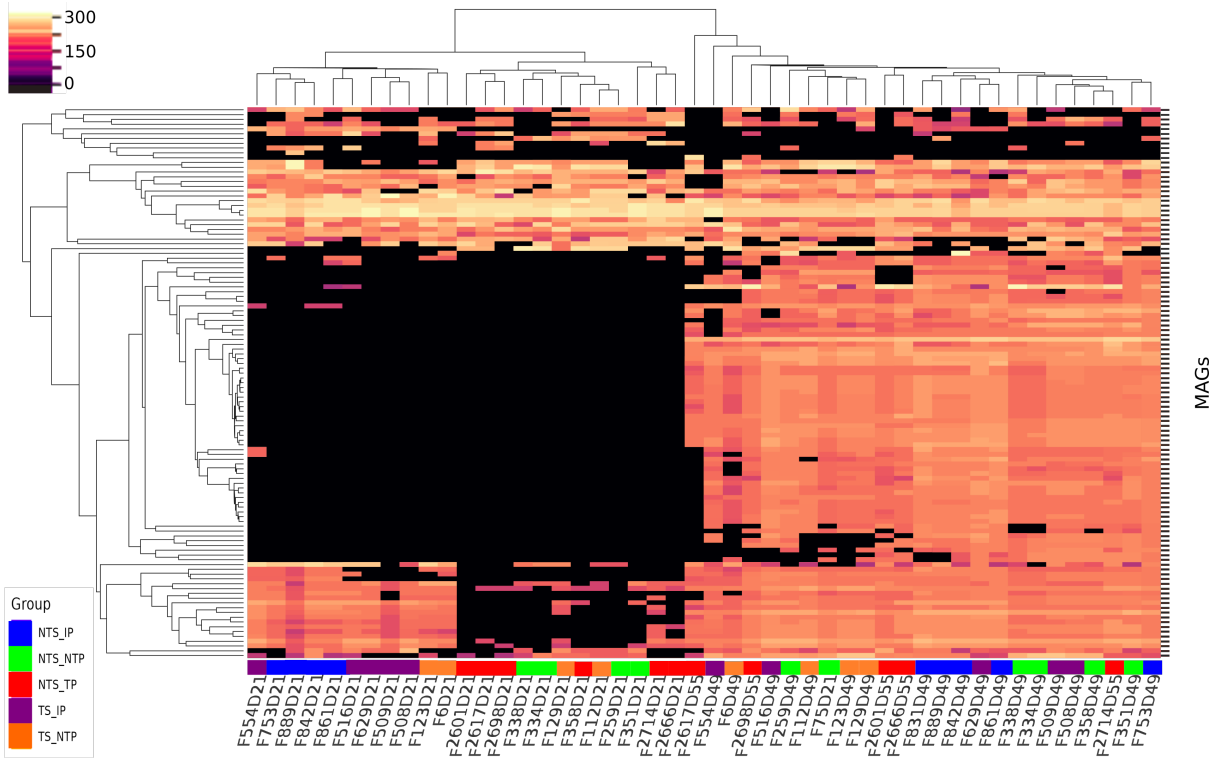


Figure 5: **MAGs abundance heatmap across samples, groups, and days**. Each cell shows the abundance of a MAG in a sample, with color intensity indicating abundance levels. Samples on the x-axis are color-coded by group and clustered by MAG abundance profiles, forming a dendrogram above. MAGs on the y-axis are also clustered based on their profiles, with a dendrogram on the left.

protein secretion profiles. Notably, *R. nasimurium* had the lowest number of secreted proteins among the identified colonizers (Supplementary Figure 13), suggesting a less active role in host interactions or ecological processes within the nasal microbiome. On the other hand, *G. parasuis* exhibited the highest number of secreted proteins, followed by *M. pluranimalium* and *S. pluranimalium*. This higher number of secreted proteins in *G. parasuis* implies a more complex interaction with the host and other microbial community members [56], potentially contributing to its colonization and persistence.

A comparison of the differential secreted proteins between the control group and the NTS-IP group was performed to identify differences and patterns, excluding the effects of antibiotics. We identified 134 proteins that were secreted in the NTS-IP group but not in the control group (see Figure 6). Among these, 17 new proteins were attributed to the colonizers. One notable secreted protein identified was the ferrichrome outer membrane transporter/phage receptor, involved in the transport of ferrichrome (a siderophore that binds iron) across the outer membrane of bacteria and also acting as a receptor for bacteriophages. This protein was secreted by *G. parasuis*, which includes species known to be pathogenic, particularly in animals, which causes Glässer's disease in pigs [57]. Further analysis is necessary to assess the pathogenic potential of the proteins and understand their roles and interactions. Ensuring that the probiotics used in the study are well-characterized and verified to be free of known pathogens is crucial, as contamination or misidentification can lead to unexpected results. Conducting pathogenicity assays, both *in vivo* and *in vitro*, will help ascertain if the presence of this protein correlates with any pathogenic effects. Moreover, analyzing the host's immune response will indicate if there is any sign of infection or adverse reaction in the inoculated group.

3.7 Comparative analysis of 16S rDNA and Metagenomic taxonomic profiling

16S rDNA gene sequencing data revealed a diverse microbial community across the samples. The dominant phyla identified included *Firmicutes*, *Bacteroidetes*, and *Proteobacteria*. At the genus level,

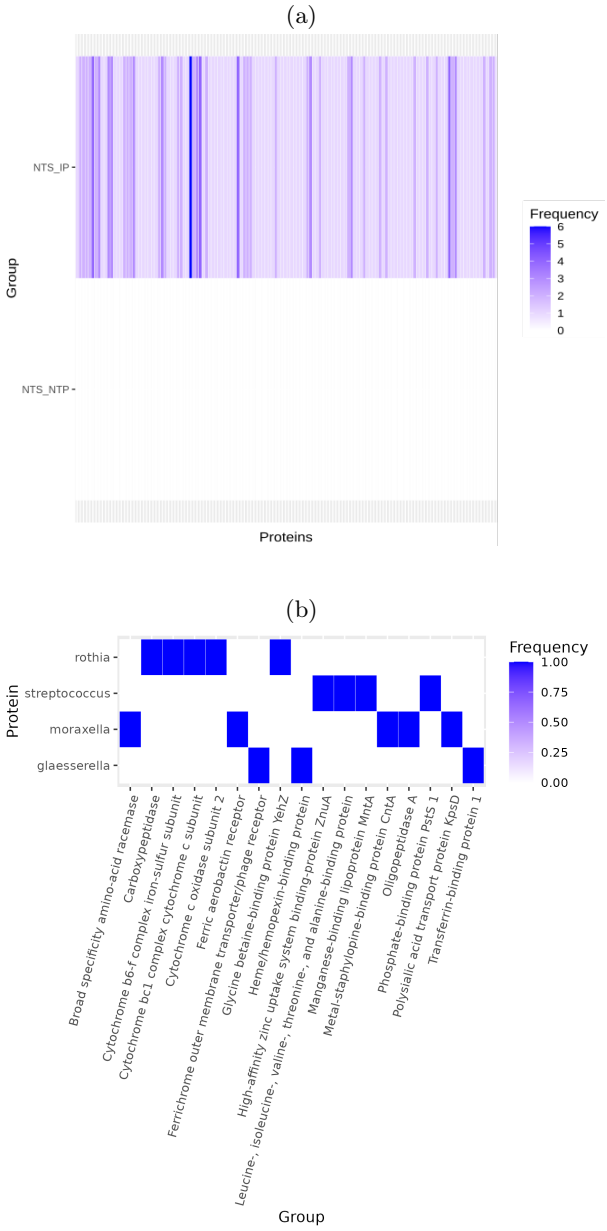


Figure 6: **Differentially predicted secreted proteins heatmaps.** 6a Heatmap of differentially secreted proteins in the NTS-IP group not predicted in the control group (NTS-NTP). 6b Heatmap of differentially secreted proteins in the NTS-IP group not predicted in the control group (NTS-NTP) secreted by colonizers

key microbial taxa such as *Lactobacillus*, *Bacteroides*, and *Escherichia* were prevalent [11].

However, metagenomic sequencing provided a more comprehensive view of the microbial community structure, capturing not only the taxonomic composition but also the functional potential of the microbiome [58]. The metagenomic data corroborated the dominance of *Firmicutes*, *Bacteroidetes*, and *Proteobacteria*, but also revealed additional taxa and functional insights that were not detected by 16S sequencing alone. These findings align with previous studies that highlight the limitations of 16S sequencing in detecting the full breadth of microbial diversity [59].

For instance, the study demonstrated that shotgun sequencing identifies more taxa than 16S sequencing. Specifically, a larger number of genera were detected by shotgun sequencing that were missed by 16S sequencing. This discrepancy is particularly evident in Figure 7, a correlation plot comparing the microbial profiles obtained from 16S rDNA gene sequencing and metagenomic sequencing, the majority of the abundances were clustered near the origin, indicating a general agreement between the taxonomic abundances found in both methods, this clustering near the origin explains that both 16S rDNA gene sequencing and metagenomics were able to identify many low-abundant taxa. However, the data predominantly aligned along the top of the linear regression line, indicating that metagenomics often detects higher abundances of certain taxa compared to 16S sequencing, highlighting the complementary nature of these two methods.

Despite this, the overall correlation between the methods was low. This low correlation can be explained by the inherent limitations of 16S sequencing, which is restricted to specific regions of the rDNA gene, whereas metagenomics sequences the entire genome, providing a more comprehensive view. While 16S sequencing is efficient and cost-effective for high-level microbial community profiling, metagenomics provides a deeper, more detailed insight, into taxa and functional profiling.

Furthermore, 16S sequencing often fails to resolve taxa at the genus or species level, a limitation less pronounced in metagenomics due to its broader sequencing approach. Metagenomics is less prone to amplification biases [60], which can underrepresent certain taxa and affect the accuracy of taxonomic assignments. This difference in taxonomic resolution significantly impacts the correlation between the results obtained by these two methods. Metagenomics, with its ability to provide detailed taxonomic information, may detect taxa that 16S sequencing misses or classifies differently.

Consequently, observed taxonomic abundances and distributions from the two methods may not align perfectly, resulting in lower correlation scores when comparing their results.

Therefore, for a comprehensive understanding of microbial community structure and function, metagenomics presents a compelling choice over 16S sequencing, particularly in studies requiring detailed taxonomic resolution, functional insights, and the detection of novel microbial species. It's important to note that 16S rDNA amplicon sequencing is widely used in the field of microorganisms. It has a lower cost and is the preferred omics method for microbial community analysis [61]. However, its reliance on specific gene regions and susceptibility to amplification biases can limit its ability to fully capture microbial diversity and functional potential compared to metagenomics. While 16S sequencing remains invaluable for initial surveys and large-scale community profiling, metagenomics offers unparalleled depth and accuracy in characterizing microbial communities across various environments. Finally, targeted amplicon NGS requires a hypothesis about which organism group (bacterial or fungal) is suspected, to ensure that appropriate amplification targets are chosen. Additionally, viruses as a group lack universal amplification targets analogous to the 16S or internal transcribed spacer regions in bacteria and fungi, although some conserved regions exist within viral families [60].

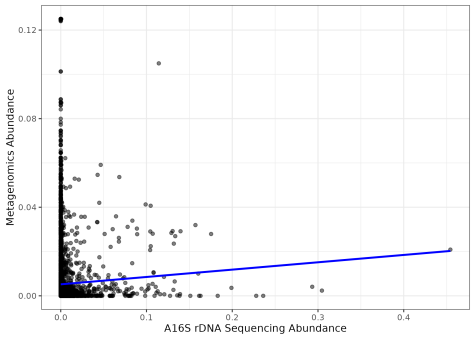


Figure 7: **Correlation plot of 16S rDNA vs WGS metagenomics abundances.** Shows the relationship between the relative abundances of taxa obtained from 16S rDNA gene amplicon sequencing and WGS metagenomic sequencing. The blue line is fitted to the data points, representing the best-fit linear regression line.

4 Conclusions

This study analyzed the dynamic nature of the piglet nasal microbiome and its response to developmental and environmental factors. We demonstrated significant temporal shifts in mirobiome community composition and secreted protein patterns between weaning (D21) and post-nursery (D49) periods. Probiotics suggested a potential non-invasive alternative to antibiotics for enhancing pig health.

5 Future directions

Further investigation is essential to uncovering the specific interactions of secreted proteins identified in this study. Future research should prioritize assessing the pathogenic potential associated with these proteins through comprehensive characterization, including robust pathogenicity assays integrating metatranscriptomics and metaproteomics. Detailed expression analysis of key proteins, such as the ferrichrome outer membrane transporter/phage receptor in *G. parasuis*, will shed light on their functional roles within the nasal microbiome. Understanding the impact of these proteins on host immune responses is critical, requiring comprehensive monitoring and metaproteomic profiling to decipher their modulation of immune pathways and health outcomes. Moreover, conducting in vivo and in vitro approaches will be essential for validating these findings and elucidating their mechanisms of action. Additionally, genomic sequencing of probiotics and other bacterial taxa within nasal microbiome samples will provide insights into virulence factors and microbial dynamics, enhancing our understanding of their functional implications in the microbiome ecosystem.

Github

The code related to this project can be found on our GitHub repository: here.

Supplementary Material

Supplementary material to this project is available through the following link.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Florencia Correa Fiz, for her invaluable guidance and support throughout this project.

References

[1] G. Berg, D. Rybakova, D. Fischer, *et al.*, *Microbiome definition re-visited: Old concepts and new challenges. microbiome 8: 103*, 2020.

[2] J. R. Marchesi and J. Ravel, *The vocabulary of microbiome research: A proposal*, 2015.

[3] N. Aggarwal, S. Kitano, G. R. Y. Puah, S. Kittelmann, I. Y. Hwang, and M. W. Chang, "Microbiome and human health: Current understanding, engineering, and enabling technologies," *Chemical Reviews*, vol. 123, no. 1, pp. 31–72, 2022.

[4] M. Matijašić, T. Meštrović, H. Čipčić Paljetak, M. Perić, A. Barešić, and D. Verbanac, "Gut microbiota beyond bacteria—mycobiome, virome, archaeome, and eukaryotic parasites in ibd," *International journal of molecular sciences*, vol. 21, no. 8, p. 2668, 2020.

[5] V. Khot, M. Strous, and A. K. Hawley, "Computational approaches in viral ecology," *Computational and Structural Biotechnology Journal*, vol. 18, pp. 1605–1612, 2020.

[6] L. M. Kasman and L. D. Porter, "Bacteriophages," in *StatPearls [Internet]*, StatPearls Publishing, 2022.

[7] C. L. Hsu, Y. Duan, D. E. Fouts, and B. Schnabl, "Intestinal virome and therapeutic potential of bacteriophages in liver disease," *Journal of hepatology*, vol. 75, no. 6, pp. 1465–1475, 2021.

[8] H. W. Virgin, "The virome in mammalian physiology and disease," *Cell*, vol. 157, no. 1, pp. 142–150, 2014.

[9] R. W. Hendrix, G. F. Hatfull, M. E. Ford, M. C. Smith, and R. N. Burns, "Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage," in *Horizontal gene transfer*, Elsevier, 2002, pp. 133–VI.

[10] E. V. Koonin, "The wonder world of microbial viruses," *Expert review of anti-infective therapy*, vol. 8, no. 10, pp. 1097–1099, 2010.

[11] M. Blanco-Fuertes, M. Sibila, G. Franzo, *et al.*, "Ceftiofur treatment of sows results in long-term alterations in the nasal microbiota of the offspring that can be ameliorated by inoculation of nasal colonizers," *Animal Microbiome*, vol. 5, no. 1, p. 53, 2023.

[12] R. Vasquez, J. K. Oh, J. H. Song, and D.-K. Kang, "Gut microbiome-produced metabolites in pigs: A review on their biological functions and the influence of probiotics," *Journal of Animal Science and Technology*, vol. 64, no. 4, p. 671, 2022.

[13] M. Billaud, Q. Lamy-Besnier, J. Lossouarn, *et al.*, "Analysis of viromes and microbiomes from pig fecal samples reveals that phages and prophages rarely carry antibiotic resistance genes," *ISME Communications*, vol. 1, no. 1, p. 55, 2021.

[14] J. Yuk, M. M. H. Akash, A. Chakraborty, S. Basu, L. P. Chamorro, and S. Jung, "Morphology of pig nasal structure and modulation of airflow and basic thermal conditioning," *Integrative and comparative biology*, icad005, 2023.

[15] M. Costa-Hurtado, E. Barba-Vidal, J. Maldonado, and V. Aragon, "Update on glässer's disease: How to control the disease under restrictive use of antimicrobials," *Veterinary microbiology*, vol. 242, p. 108 595, 2020.

[16] A. Lekagul, V. Tangcharoensathien, A. Mills, J. Rushton, and S. Yeung, "How antibiotics are used in pig farming: A mixed-methods study of pig farmers, feed mills and veterinarians in thailand," *BMJ global health*, vol. 5, no. 2, e001918, 2020.

[17] D. Shin, S. Y. Chang, P. Bogere, *et al.*, "Beneficial roles of probiotics on the modulation of gut microbiota and immune response in pigs," *PloS one*, vol. 14, no. 8, e0220843, 2019.

[18] L. Hansen, T. Cieplak, B. Nielsen, Y. Zhang, C. Lauridsen, and N. Canibe, "Screening of probiotic candidates in a simulated piglet small intestine in vitro model," *FEMS Microbiology Letters*, vol. 368, no. 7, fnab045, 2021.

[19] C. Hill, F. Guarner, G. Reid, *et al.*, "Expert consensus document: The international scientific association for probiotics and prebiotics consensus statement on the scope and appropriate use of the term probiotic," *Nature reviews Gastroenterology & hepatology*, 2014.

[20] S. Wahab, D. Almaghaslah, S. E. Mahmood, *et al.*, "Pharmacological efficacy of probiotics in respiratory viral infections: A comprehensive review," *Journal of Personalized Medicine*, vol. 12, no. 8, p. 1292, 2022.

[21] S. Al-Romaih, O. Harati, L. E. Mfuna, *et al.*, "Response to intranasal lactococcus lactis w136 probiotic supplementation in refractory crs is associated with modulation of non-type 2 inflammation and epithelial regeneration," *Frontiers in Allergy*, vol. 4, p. 1 046 684, 2023.

[22] S. Luan, Y. Sun, Y. Wang, R. Sa, and H. Zhang, "Bacillus amyloliquefaciens spray improves the growth performance, immune status, and respiratory mucosal barrier in broiler chickens," *Poultry science*, vol. 98, no. 3, pp. 1403–1409, 2019.

[23] A. Mårtensson, F. U. Nordström, C. Cervin-Hoberg, *et al.*, "Nasal administration of a probiotic assemblage in allergic rhinitis: A randomised placebo-controlled crossover trial," *Clinical & Experimental Allergy*, vol. 52, no. 6, pp. 774–783, 2022.

[24] L. Niittynen, A. Pitkäranta, and R. Korpela, "Probiotics and otitis media in children," *International journal of pediatric otorhinolaryngology*, vol. 76, no. 4, pp. 465–470, 2012.

[25] D. Gagic, M. Ciric, W. X. Wen, F. Ng, and J. Rakonjac, "Exploring the secretomes of microbes and microbial communities using filamentous phage display," *Frontiers in microbiology*, vol. 7, p. 169 717, 2016.

[26] G. V. Uritskiy, J. DiRuggiero, and J. Taylor, "Metawrap—a flexible pipeline for genome-resolved metagenomic data analysis," *Microbiome*, vol. 6, pp. 1–13, 2018.

[27] F. Krueger, "Trim galore!: A wrapper around cutadapt and fastqc to consistently apply adapter and quality trimming to fastq files, with extra functionality for rrbs data," *Babraham Institute*, 2015.

[28] M. Martin, "Cutadapt removes adapter sequences from high-throughput sequencing reads," *EMBnet. journal*, vol. 17, no. 1, pp. 10–12, 2011.

[29] A. S., "Fastqc: A quality control tool for high throughput sequence data.," 2010. [Online]. Available: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

[30] K. Rotmistrovsky and R. Agarwala, "Bmtagger: Best match tagger for removing human reads from metagenomics datasets," *NCBI/NLM, National Institutes of Health*, 2011.

[31] A. Warr, N. Affara, B. Aken, *et al.*, "An improved pig reference genome sequence to enable pig genetics and genomics research," *Gigascience*, vol. 9, no. 6, giaa051, 2020.

[32] S. Nurk, D. Meleshko, A. Korobeynikov, and P. A. Pevzner, "Metaspades: A new versatile metagenomic assembler," *Genome research*, vol. 27, no. 5, pp. 824–834, 2017.

[33] D. E. Wood, J. Lu, and B. Langmead, "Improved metagenomic analysis with kraken 2," *Genome biology*, vol. 20, pp. 1–13, 2019.

[34] B. D. Ondov, N. H. Bergman, and A. M. Phillippy, "Interactive metagenomic visualization in a web browser," *BMC bioinformatics*, vol. 12, pp. 1–10, 2011.

[35] P. E. McKight and J. Najab, "Kruskal-wallis test," *The corsini encyclopedia of psychology*, pp. 1–1, 2010.

[36] D. D. Kang, F. Li, E. Kirton, *et al.*, "Metabat 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies," *PeerJ*, vol. 7, e7359, 2019.

[37] D. H. Parks, M. Imelfort, C. T. Skennerton, P. Hugenholtz, and G. W. Tyson, "Checkm: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes," *Genome research*, vol. 25, no. 7, pp. 1043–1055, 2015.

[38] R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, and C. Kingsford, "Salmon provides fast and bias-aware quantification of transcript expression," *Nature methods*, vol. 14, no. 4, pp. 417–419, 2017.

[39] T. Seemann, "Prokka: Rapid prokaryotic genome annotation," *Bioinformatics*, vol. 30, no. 14, pp. 2068–2069, 2014.

[40] F. Teufel, J. J. Almagro Armenteros, A. R. Johansen, *et al.*, "Signalp 6.0 predicts all five types of signal peptides using protein language models," *Nature biotechnology*, vol. 40, no. 7, pp. 1023–1025, 2022.

[41] RStudio Team, *Rstudio: Integrated development environment for r*, RStudio, PBC., Boston, MA, 2020. [Online]. Available: http://www.rstudio.com/.

[42] J. Oksanen, R. Kindt, P. Legendre, *et al.*, *Vegan: Community ecology package*, R package version 1.15-1, 2008. [Online]. Available: http://cran.r-project.org/,%20http://vegan.r-forge.r-project.org/.

[43] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016, ISBN: 978-3-319-24277-4. [Online]. Available: https://ggplot2.tidyverse.org.

[44] H. Wickham, R. François, L. Henry, and K. Müller, *Dplyr: A grammar of data manipulation*, R package version 0.7.6, 2018. [Online]. Available: https://CRAN.R-project.org/package=dplyr.

[45] H. Wickham, D. Vaughan, and M. Girlich, *Tidyr: Tidy messy data*, R package version 1.3.1, https://github.com/tidyverse/tidyr, 2024. [Online]. Available: https://tidyr.tidyverse.org.

[46] H. Wickham, *Stringr: Simple, consistent wrappers for common string operations*, R package version 1.2.0, 2017. [Online]. Available: https://CRAN.R-project.org/package=stringr.

[47] W. Summers, "Virus infection," *Encyclopedia of Microbiology*, p. 546, 2009.

[48] A. D. Teklemariam, R. R. Al-Hindi, I. Qadri, *et al.*, "The battle between bacteria and bacteriophages: A conundrum to their immune system," *Antibiotics*, vol. 12, no. 2, p. 381, 2023.

[49] H. J. Kim, J.-H. Kim, S.-A. Han, and W. Kim, "Compositional alterations of the nasal microbiome and staphylococcus aureus–characterized dysbiosis in the nasal mucosa of patients with allergic rhinitis," *Clinical and Experimental Otorhinolaryngology*, vol. 15, no. 4, p. 335, 2022.

[50] K. Law, B. Lozinski, I. Torres, *et al.*, "Disinfection of maternal environments is associated with piglet microbiome composition from birth to weaning," *Msphere*, vol. 6, no. 5, pp. 10–1128, 2021.

[51] J.-A. Lim, J. Cha, S. Choi, J.-H. Kim, and D. Kim, "Early colonization of the intestinal microbiome of neonatal piglets is influenced by the maternal microbiome," *Animals*, vol. 13, no. 21, p. 3378, 2023.

[52] M. Blanco-Fuertes, F. Correa-Fiz, S. López-Serrano, M. Sibila, and V. Aragon, "Sow vaccination against virulent glaesserella parasuis shapes the nasal microbiota of their offspring," *Scientific reports*, vol. 12, no. 1, p. 3357, 2022.

[53] C. M. Bassis, A. L. Tang, V. B. Young, and M. A. Pynnonen, "The nasal cavity microbiota of healthy adults," *Microbiome*, vol. 2, pp. 1–5, 2014.

[54] N. Habibi, A. S. Mustafa, and M. W. Khan, "Composition of nasal bacterial community and its seasonal variation in health care workers stationed in a clinical research laboratory," *PloS one*, vol. 16, no. 11, e0260314, 2021.

[55] S. S. Alkhalil, "The role of bacteriophages in shaping bacterial composition and diversity in the human gut," *Frontiers in Microbiology*, vol. 14, p. 1 232 413, 2023.

[56] Y. S. Mahmmod, F. Correa-Fiz, and V. Aragon, "Variations in association of nasal microbiota with virulent and non-virulent strains of glaesserella (haemophilus) parasuis in weaning piglets," *Veterinary research*, vol. 51, no. 1, p. 7, 2020.

[57] A. E. Watt, G. F. Browning, A. R. Legione, *et al.*, "A novel glaesserella sp. isolated from pigs with severe respiratory infections has a mosaic genome with virulence factors putatively acquired by horizontal transfer," *Applied and environmental microbiology*, vol. 84, no. 11, e00092–18, 2018.

[58] M. I. Smith, W. Turpin, A. D. Tyler, M. S. Silverberg, and K. Croitoru, "Microbiome analysis–from technical advances to biological relevance," *F1000prime reports*, vol. 6, 2014.

[59] F. Durazzi, C. Sala, G. Castellani, G. Manfreda, D. Remondini, and A. De Cesare, "Comparison between 16s rrna and shotgun sequencing data for the taxonomic characterization of the gut microbiota," *Scientific reports*, vol. 11, no. 1, p. 3030, 2021.

[60] A. Dulanto Chiang and J. P. Dekker, "From the pipeline to the bedside: Advances and challenges in clinical metagenomics," *The Journal of infectious diseases*, vol. 221, no. Supplement_3, S331–S340, 2020.

[61] A. Klindworth, E. Pruesse, T. Schweer, *et al.*, "Evaluation of general 16s ribosomal rna gene pcr primers for classical and next-generation sequencing-based diversity studies," *Nucleic acids research*, vol. 41, no. 1, e1–e1, 2013.