CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

Comparative Genomic Analysis of *Akkermansia muciniphila*: Unravelling its Genetic Diversity and Role in Human Health

by

Ibra Khushi

A thesis submitted in partial fulfillment for the degree of Master of Science

in the

Faculty of Health and Life Sciences
Department of Bioinformatics and Biosciences

2025

To my parents,whose sacrifices and encouragement have made this achievement possible.

CERTIFICATE OF APPROVAL

Comparative Genomic Analysis of *Akkermansia muciniphila*: Unravelling its Genetic Diversity and Role in Human Health

by

Ibra Khushi

(MBS233013)

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Shaheen Shahzad	IIU, Islamabad
(b)	Internal Examiner	Dr. Sami Ullah Jan	CUST, Islamabad
(c)	Supervisor	Dr. Syeda Marriam Bakhtiar	CUST, Islamabad

Dr. Syeda Marriam Bakhtiar
Thesis Supervisor
August, 2025

Dr. Syeda Marriam Bakhtiar
Head
Dept. of Bioinfo. & Biosciences
August, 2025

Dr. Sahar Fazal
Dean
Faculty of Health & Life Sciences
August, 2025

Author's Declaration

I, **Ibra Khushi** hereby state that my MS thesis titled "**Comparative Genomic Analysis of *Akkermansia munciniphila*: Unravelling its Genetic Diversity and Role in Human Health**" is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/abroad.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my MS Degree.

(Ibra Khushi)
Registration No: MBS233013

Plagiarism Undertaking

I solemnly declare that research work presented in this thesis titled "**Comparative Genomic Analysis of *Akkermansia muciniphila*: Unravelling its Genetic Diversity and Role in Human Health**" is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS Degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized work.

(Ibra Khushi)

Registration No: MBS233013

Acknowledgement

All praise and gratitude are due to **Allah Almighty**, the Most Compassionate and the Most Merciful, whose infinite blessings, guidance, and mercy have enabled me to complete this thesis. It is through His divine support that I found the strength, patience, and determination to overcome every challenge along the way.

I extend my heartfelt thanks to my **beloved husband, Mobeen Ali**, whose constant support, encouragement, and understanding have been invaluable throughout this journey. His patience, sacrifices, and unwavering belief in me have played a pivotal role in the successful completion of this work.

I am deeply grateful to my **beloved parents and siblings**, who have always been my source of inspiration. Their endless love, prayers, and emotional support have given me the courage to pursue my academic goals with confidence and dedication.

I would also like to express my sincere appreciation to my **in-laws** for their kindness, encouragement, and continuous support during my academic endeavors. Their thoughtful understanding and well wishes have been a great source of comfort.

My warmest thanks go to my **friends and classmates**, whose companionship, shared experiences, and intellectual discussions have enriched my academic journey. Their moral support and moments of laughter made this process more manageable and enjoyable.

I wish to acknowledge and express my sincere appreciation to all my teachers and mentors; I am especially thankful to my **supervisor, Syeda Mariam Bakhtiar**, for their valuable guidance, expert insight, and constructive feedback throughout the research process. Their mentorship has significantly shaped the quality and direction of this thesis.

Finally, I would like to thank the **Department of Bioinformatics and Biosciences**, Faculty of Health and Life Sciences, **Capital University of Science**

and Technology, for providing a conducive learning environment, academic resources, and opportunities that have supported my growth and learning throughout my study. I am grateful to the Dean Faculty of Health and Life Sciences, Dr Sahar Fazal, and her office for facilitation and guidance throughout my journey. I also appreciate the efforts of the Graduate Studies office in streamlining processes for us, enabling the timely completion of tasks.

(Ibra Khushi)

Abstract

The gut microbiota plays a pivotal role in human health, with *Akkermansia muciniphila* emerging as a keystone species due to its mucin-degrading capabilities and probiotic potential. This study conducted a comparative genomic analysis of 280 *Akkermansia muciniphila* strains to delineate genetic diversity, functional adaptations, and safety profiles relevant to probiotic development. Utilizing pangenome and phylogenomic approaches, identification of an open pangenome structure was identified with 227 strains retained after species delimitation (ANI >95%). The core genome revealed conserved pathways for mucin degradation, nutrient transport, and stress response, while accessory genes highlighted metabolic flexibility and ecological adaptations. Antibiotic resistance genes (ARGs) were detected in 14.5% of strains, including lincosamide, tetracycline, and macrolide resistance determinants, with plasmids and rare virulence factors observed in selected isolates. Biosynthetic gene clusters (BGCs) for terpenes and arylpolyenes were widespread, whereas niche-specific clusters (e.g., lassopeptides) were strain-restricted. These findings emphasize the importance of strain-specific evaluation for therapeutic applications, balancing metabolic benefits against risks of ARG dissemination. The study highlights *A. muciniphila*'s potential as a next-generation probiotic while advocating for rigorous genomic screening to ensure safety.

Keywords: Comparative genomics, pangenome analysis, *Akkermansia muciniphila*, antibiotic resistance genes, probiotic potential.

Contents

Author's Declaration iv

Plagiarism Undertaking v

Acknowledgement vi

Abstract viii

List of Figures xi

List of Tables xiii

Abbreviations xiv

1 Introduction 1
- 1.1 Pangenome Analysis . . . 3
  - 1.1.1 Components of Pangenome . . . 4
  - 1.1.2 Types . . . 5
    - 1.1.2.1 Open Pangenome . . . 5
    - 1.1.2.2 Characteristics . . . 5
    - 1.1.2.3 Biological Examples . . . 5
    - 1.1.2.4 Implications . . . 5
    - 1.1.2.5 Closed Pangenome . . . 6
    - 1.1.2.6 Characteristics . . . 6
    - 1.1.2.7 Biological Examples . . . 6
  - 1.1.3 Goal of Pangenome Analysis . . . 6
  - 1.1.4 Comparative Genomics . . . 7
  - 1.1.5 Tools & Applications . . . 8
  - 1.1.6 Pangenome of *Akkermansia muciniphila* . . . 8
- 1.2 Problem Statement . . . 11
- 1.3 Research Objectives . . . 11

2 Literature Review 12
- 2.1 General Characteristics of *Akkermansia muciniphila* . . . 12

2.2 General Characteristics of *Akkermansia muciniphila* . . . 13
2.2.1 Mucin Degradation . . . 13
2.3 Classification of *Akkermansia muciniphila* . . . 14
2.4 Genomic Characteristics of *Akkermansia muciniphila* . . . 15
2.5 Diversity in *Akkermansia muciniphila* . . . 16
2.6 Significance of *Akkermansia muciniphila* . . . 17
2.7 Significance of *Akkermansia muciniphila* as Probiotic . . . 18
2.8 Genome Plasticity . . . 19
2.9 Gap Analysis & Concerns about *A. muciniphila* . . . 21

3 Material and Method **24**
3.1 Genome Retrieval and Selection . . . 25
3.2 Genome Annotation and Average Nucleotide Identity (ANI) Analysis 25
3.3 Pangenome Estimation and Phylogenetic Analysis . . . 26
3.4 Functional Annotation and Enrichment Analysis . . . 27
3.5 Identification of Antibiotic Resistance Genes (ARGs), Virulence Factors (VFs), and Plasmids . . . 27
3.6 Prediction of Secondary Metabolite Biosynthesis Gene Clusters (BGCs) 28

4 Results and Discussion **29**
4.1 General Characteristics of the Genomes . . . 29
4.2 Species Delimitation and Exclusion of Divergent Strains . . . 29
4.3 Pangenome Analysis Reveals Significant Reduction in Gene Pool Following Strain Exclusion . . . 38
4.4 Functional Annotation Reveals Metabolic Robustness . . . 49
4.5 Distribution and Diversity of ARGs in *A. muciniphila* Strains . . . 51
4.6 Limited Detection of Virulence Factors in *A. muciniphila* Strains . 52
4.7 Distribution of Plasmid Replicons . . . 53
4.8 Biosynthetic Gene Cluster (BGC) Detection . . . 53
4.9 Discussion and Interpretation of Results . . . 54

5 Conclusion & Future Recommendations **58**

Bibliography **60**

List of Figures

1.1 Potential significance of *A. muciniphila* as a next-generation probiotic [4] . . . 2
1.2 Methods to increase the abundance of *Akkermansia muciniphila* in Gut [8] . . . 3
1.3 Organization of a pangenome composed of core and dispensable components of the genome [12] . . . 4
1.4 Evolutionary relationships among more than 70 mammalian species inferred from genomic sequence data.Color-coded clades highlight major mammalian lineages, revealing patterns of divergence and common ancestry [6] . . . 7
1.5 Protective role of *Akkermansia muciniphila* in intestinal homeostasis [14] . . . 9
1.6 Role of *Akkermansia muciniphila* in metabolic diseases [17] . . . 9
1.7 Role of Akkermencia Muciniphila in gut health [22] . . . 10

2.1 Microscopic image of *Akkermansia muciniphila* [2] . . . 12
2.2 Interaction at the host mucus layer between secreted mucins and bacteria. Bacteria can bind to mucins through adhesins and/or lectins that recognize O-glycans [15] . . . 13
2.3 Graphical circular map of the chromosome of *Akkermansia muciniphila* KGMB01988. From outside to the center: Genes on forward strand (color by COG categories), Genes on reverse strand (color by COG categories), GC content (black), GC skew (light green/Orange), and gained gene families on forward strand and negative strand after speciation (color by branching times) [2] . . . 16
2.4 Impact of low number of *Akkermansia muciniphila* on gut homeostasis [2] . . . 17
2.5 Health benefits of *Akkermansia muciniphila* [2] . . . 18
2.6 *A. muciniphila*'s function in inflammatory intestinal disorders [26] . . . 19
2.7 Impact of genome plasticity on bacterial pathogenesis . . . 20

3.1 Graphical abstract of pangenome analysis of *Akkermansia muciniphila* [2] . . . 24

4.1 Comparative pangenome analysis represented as pie charts, showing the distribution of core, soft core, shell, and cloud genes in the initial dataset of 280 *A. muciniphila* genomes [49] . . . 39

4.2 Comparative pangenome analysis represented as pie charts, showing the distribution of core, soft core, shell, and cloud genes in the refined dataset of 227 genomes following the exclusion of divergent strains, highlighting the reduction in cloud genes and increase in core genes after removing divergent strains[51] 40
4.3 Cumulative number of genes in the pangenome as a function of added genomes, demonstrating an open pangenome 41
4.4 Number of new genes added with each additional genome, highlighting high initial diversity and ongoing gene acquisition. 41
4.5 Trend of total and conserved genes with increasing genomes, showing a stable yet very small core genome. 42
4.6 Detailed view of the number of conserved genes across all genomes . 42
4.7 Trend of new and unique genes as genomes are added. 43
4.8 Number of unique genes added with each genome, indicating a high degree of accessory gene diversity. 43
4.9 Cumulative number of genes in the pangenome as a function of added genomes, showing a gradual increase indicative of a less pronounced open pangenome compared to the initial dataset. 44
4.10 Number of new genes added with each additional genome, demonstrating a reduced rate of novel gene acquisition. 44
4.11 Trend of total and conserved genes with increasing genomes, revealing a larger core genome and a more stable core gene set. 45
4.12 Detailed view of the number of conserved genes across all genomes, showing minimal variation. 45
4.13 Trend of new and unique genes as genomes are added, showing a lower rate of new gene accumulation and very few unique genes. . . 46
4.14 Number of unique genes added with each genome, highlighting a significant reduction in unique gene content compared to the initial dataset. 46
4.15 Phylogenetic analysis of *A. muciniphila* genomes. Circular phylogram depicting the phylogenetic relationships among 280 bacterial genomes, constructed using accessory gene presence/absence data from Roary pangenome analysis and visualized with iTOL. The red-colored branches highlight the 53 divergent strains that formed a distinct clade, indicating significant genetic dissimilarity from the remaining strains 47
4.16 Phylogenetic analysis of *A. muciniphila* genomes. Circular phylogram constructed using accessory gene presence/absence data from Roary pangenome analysis and visualized with iTOL using the refined dataset of 227 genomes after the exclusion of 53 divergent strains 48
4.17 Functional annotation of *A. muciniphila* genes, highlighting the abundance of genes with unknown functions. 50

5.1 Summary of main objectives of the project and results achieved . . 58

List of Tables

1.1 Comparison of key features between comparative genomics and-pangenome analysis highlighting scope, methods, and functional insights [18] 8

2.2 Classification of *Akkermansia muciniphila*s 14

4.1 Table of ANI results (>95%) used for species boundary determination of 227 selected strains, detailing Query and Reference strain comparisons, ANI values, Fragment alignment, and Alignment Length. 30

Abbreviations

ARGs	Antibiotic-Resistant Genes
BGCs	Biosynthetic Gene Cluster
BLAST	Basic Local Alignment Search Tools
COGs	Cluster of Orthologous Genes
LOS	Lipoologosaccarides
NAFLD	Non-Alcoholic Fatty Liver Disease
PAMPs	Pathogen-Associated Molecular Patterns
SCFA	Short Chain Fatty Acid
SNP	Single Nucleotide Polymorphisms
T2DM	Type 2 diabetes mellitus
TLRs	Toll-like receptors

Chapter 1

Introduction

Gram-negative, oval-shaped, non-motile, and non-spore-forming, *Akkermansia muciniphila* is a bacterium that resides in the mucous layer of the human digestive system. It utilizes specialized enzymes, including mucinases, glycosidases, sulfatases, and sialidases, to utilize mucin as its primary energy source and thrives in anaerobic environments [1]. Although *A. muciniphila* lacks traditional mucus-binding domains, it successfully colonizes the mucosal layer, because of surface lipo-oligosaccharides and the outer membrane protein Amuc_1100 [1, 2].The bacteria *A. muciniphila* demonstrate metabolic flexibility by digesting substitute substrates such as glucose and amino sugars in conditions where mucin is low. Although the bacteria *A. muciniphila* are anaerobic but certain strains show oxygen tolerance, indicating that they can adapt to oxic-anoxic interface of the gut mucosa [1]. In natural gut ecosystems, *A. muciniphila* plays a critical role in gut health by supporting intestinal barrier function and controlling immunological responses. Its ability to degrade mucin promotes homeostasis and preserves the integrity of the mucus layer [1].

The bacteria *A. muciniphila* is actively involved in maintain health of the gut and it is reported that in the case of the metabolic and inflammatory conditions such as obesity, type 2 diabetes, inflammatory bowel disease (IBD), and non-alcoholic fatty liver disease (NAFLD) a significant decrease is observed in levels of *A. muciniphila* [3]. The association with health maintenance initiates an interest in

A. muciniphila as a treatment option as a probiotic against chronic inflammation and metabolic diseases including the immunotherapy in case of colorectal cancer as well as in improvement in cognitive outcomes in neurodegenerative disease [1].

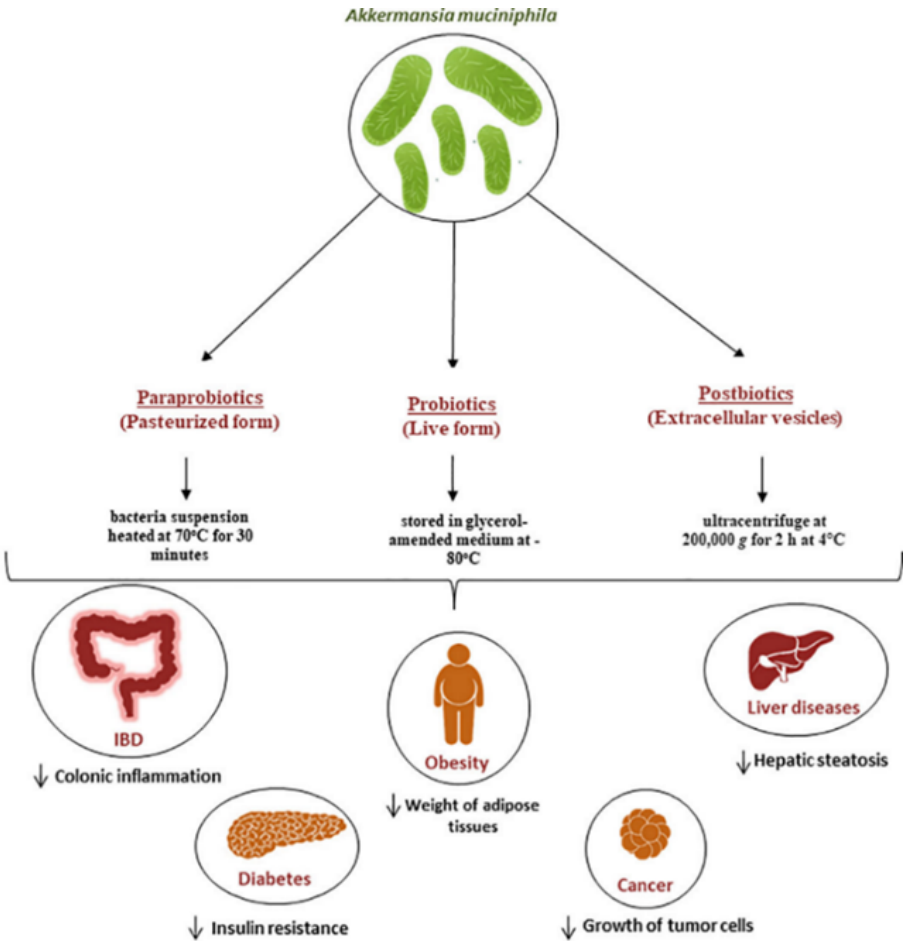

FIGURE 1.1: Potential significance of *A. muciniphila* as a next-generation probiotic [4]

Evolutionary plasticity of *A. muciniphila*, i.e., the ability of this bacterium to modify its phenotype, including the physical traits, behavior, and physiology, in response to environmental changes, is a major observation regarding *A. muciniphila*. Although this change or variation is not a direct outcome of genetic change but the modifications in phenotypes are observed due to gene expression. This evolutionary plasticity indicates an open pangenome, which supports the idea that new genes are added in each strain due to variations in the environment to which they belong. The argument is strongly supported by comparative genomics and pangenomic analysis of *A. muciniphila* [5].

The genome plasticity gene pool comprises both essential genes, like core genes, as well as accessory genes related to the generation of vitamin B12, mucin breakdown, and changeable surface antigens [6, 7]. This genetic adaptability of *A. muciniphila* makes it a potential next generation probiotic with numerous probiotic characteristics, such as immunological regulation and gut barrier preservation, which are supported by strain-level diversity (AmI–AmIV) [7]. But at the same time it raises eyebrows on the safety of the bacterium to be used as a probiotic.

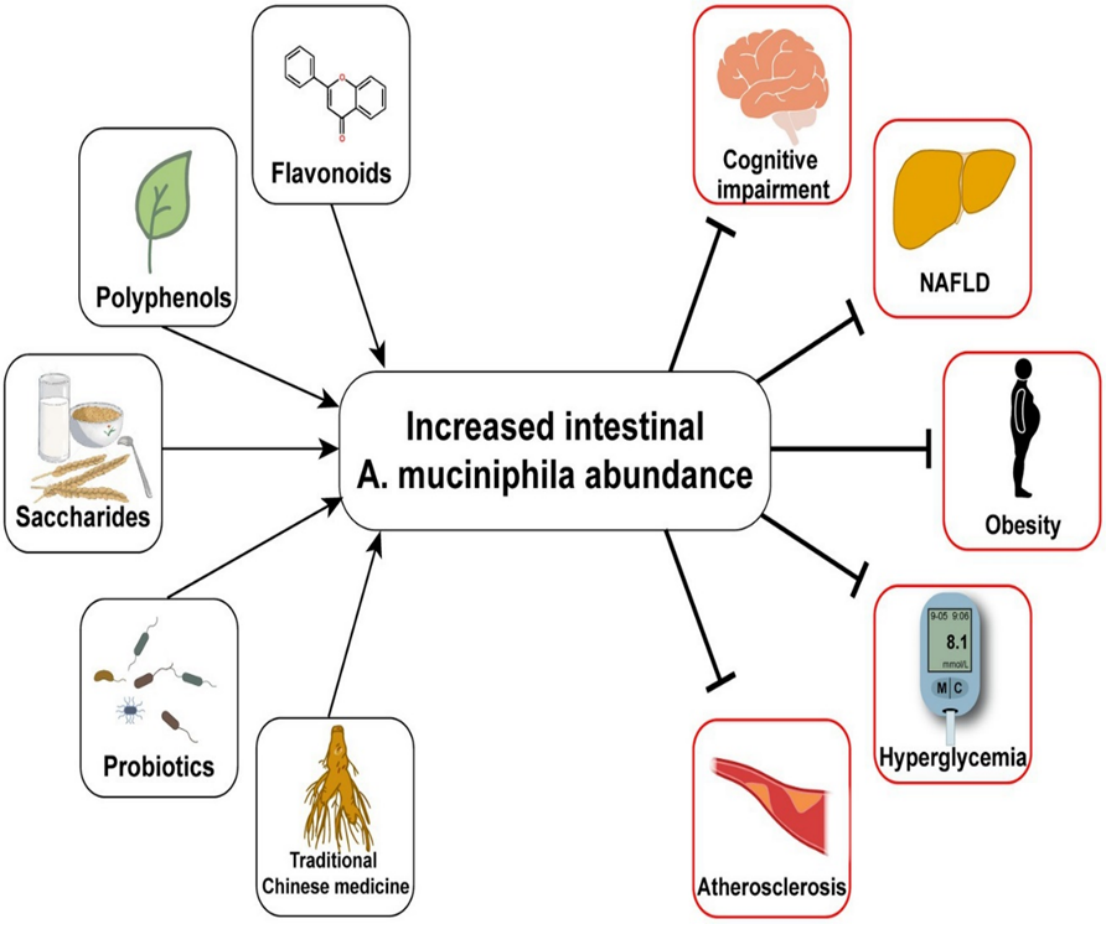

FIGURE 1.2: Methods to increase the abundance of *Akkermansia muciniphila* in Gut [8]

1.1 Pangenome Analysis

The set of orthologous and distinct genes found in a certain group of organisms is known as the pan-genome. The core genome, accessory genome, and genes specific to a species or strain make up the pan-genome. Depending on the Heap

law's alpha value, the pan-genome is categorized as open or closed. In a closed pan-genome, the number of gene families will not significantly increase, but in an open pan-genome, the number of gene families will steadily rise as more genomes are added to the analysis [9]. Compared to conventional genomic research, pangenome analysis has a number of benefits. A pangenome, for instance, can capture greater genetic heterogeneity since it is not constrained by the physical limitations of a single genome [10].

1.1.1 Components of Pangenome

Core and accessory genomes are two subtypes of pan-genomes. A target clade's core genome is a collection of gene families that are present in every genome and are probably necessary for the clade's development or survival. An accessory genome is a group of genes found in one or more, but not all, in a particular clade. To comprehend the differences in the clade's genomes and, consequently, their distinct lifestyles and evolutionary paths, the accessory genome is seen as optional [11].

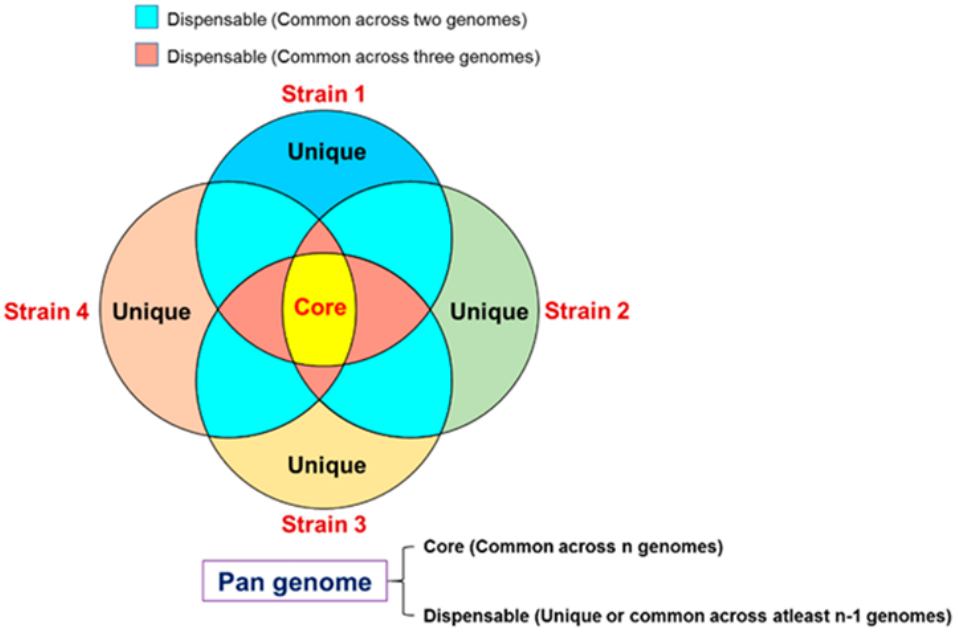

FIGURE 1.3: Organization of a pangenome composed of core and dispensable components of the genome [12]

1.1.2 Types

1.1.2.1 Open Pangenome

An open pangenome is one in which the total number of genes continues to increase with the addition of each new genome, indicating ongoing acquisition of genetic material, often through horizontal gene transfer (HGT) [13].

1.1.2.2 Characteristics

- High genomic plasticity
- High rates of gene acquisition and recombination
- Gene discovery curve does not plateau
- Common in bacteria inhabiting diverse, fluctuating, or competitive environments

1.1.2.3 Biological Examples

- *Escherichia coli*
- *Streptococcus agalactiae*
- *Pseudomonas aeruginosa*

These species encounter various ecological niches and exhibit high HGT rates, supporting a constantly growing gene repertoire [14].

1.1.2.4 Implications

Open pangenomes reveal a species ability to adapt and evolve in response to environmental pressures, antimicrobial exposure, or host defenses [15].

1.1.2.5 Closed Pangenome

A closed pangenome is one in which the total number of genes reaches a saturation point, and adding more genomes yields few or no new genes. This pattern implies limited genetic diversity and low rates of HGT [16].

1.1.2.6 Characteristics

- Low genomic variation
- Large, stable core genome
- Limited gene gain or loss
- Common in clonally reproducing or host-restricted organisms

1.1.2.7 Biological Examples

- *Mycobacterium tuberculosis*
- *Bacillus anthracis*
- Symbionts like *Buchnera aphidicola*s

Such organisms often live in stable environments or are highly specialized, reducing the need for gene acquisition [13].

1.1.3 Goal of Pangenome Analysis

The goal of pangenomes is to fully capture genetic heterogeneity within a collection of organisms, addressing the shortcomings of conventional linear reference genomes. This all-encompassing method improves knowledge of genetic variation and makes more precise genomic analysis possible [17]. By increasing analysis precision and recall, pangenomes improve the detection of large genetic variants, especially in cardiomyopathies. They overcome the disadvantages of standard gene sections, which have trouble handling larger insertions and deletions, and expedite diagnostic procedures while cutting expenses [13].

1.1.4 Comparative Genomics

Comparative genomics is the study of genome content, sequence, and structure across species or strains to infer evolutionary history, detect functionally important genes, and understand genome dynamics. It includes analysis of orthologs, synteny blocks, gene duplications, and structural variations [**?**].

It has been widely used to annotate newly sequenced genomes, understand phylogenetic relationships, and detect evolutionary events such as genome rearrangements or selective pressures [**?**].

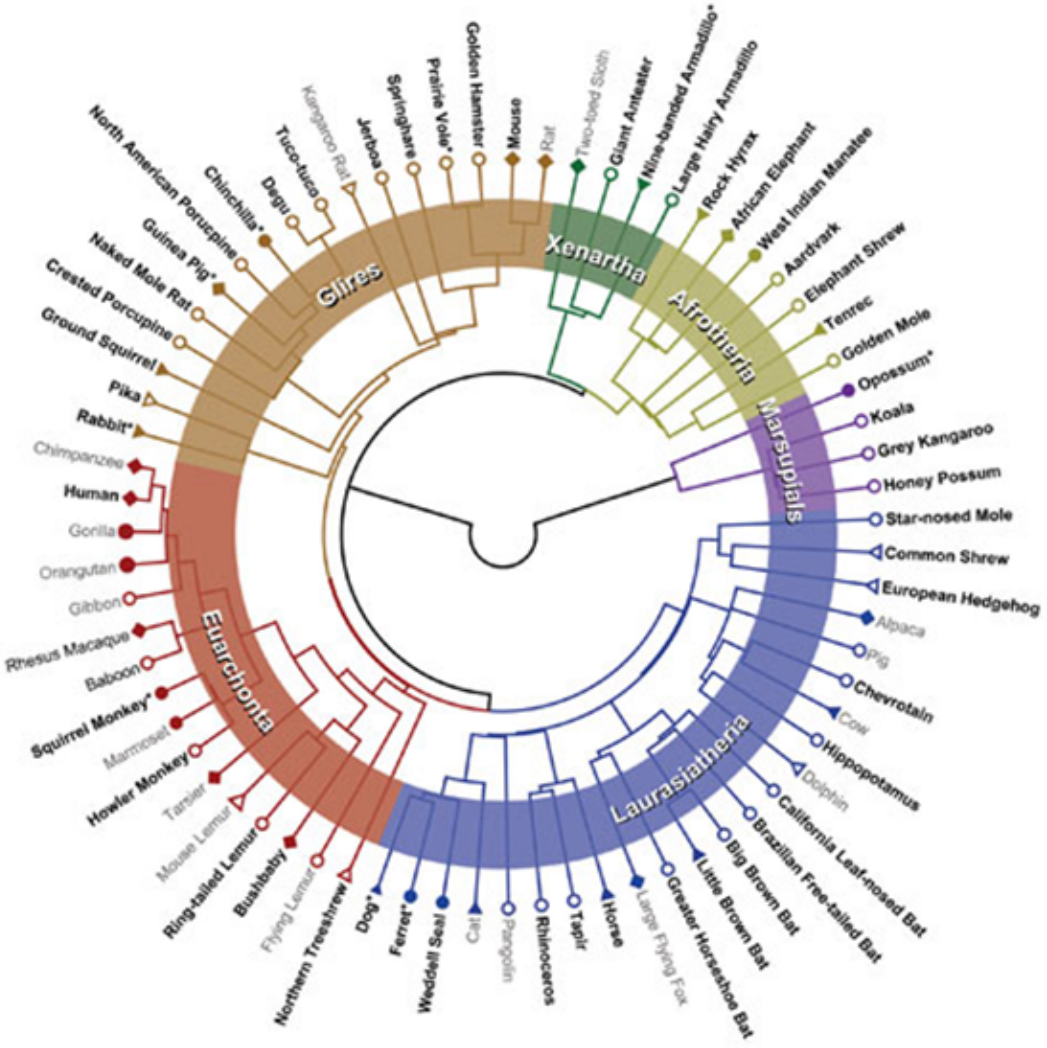

FIGURE 1.4: Evolutionary relationships among more than 70 mammalian species inferred from genomic sequence data.Color-coded clades highlight major mammalian lineages, revealing patterns of divergence and common ancestry [6]

1.1.5 Tools & Applications

Recent tools such as PanACoTA enable large-scale microbial genome comparisons through automated workflows for genome downloading, quality filtering, annotation, and tree construction [18]. These approaches facilitate the reconstruction of species phylogeny and the identification of both core and strain-specific genomic features.

Comparative genomics has also advanced into next-generation microbiology, integrating genome-wide studies with functional characterization [18].

TABLE 1.1: Comparison of key features between comparative genomics and-pangenome analysis highlighting scope, methods, and functional insights [18]

Aspect	**Comparative Genomics**	**Pangenome Analysis**
Scope	Between species or strains	Within species or clades
Primary methods	Alignment, synteny, orthology, evolutionary models	Gene presence – absence, graph-based genome comparison
Variants analyzed	SNPs, CNVs, indels, synteny blocks	Gene gain/loss, presence – absence variation (PAV), SVs
Outcomes	Conserved genes, divergence, phylogenetic relationships	Pangenome curves, core/accessory genome structure
Examples of tools	MUMmer, SyRI, VISTA	PanACoTA, PGGB, Pangraph, minigraph

1.1.6 Pangenome of *Akkermansia muciniphila*

Comparative genomic analysis of several strains provides a comprehensive overview of genetic diversity, evolutionary paths, and strain-specific roles. This analysis provides several classified phylogroups based on substantial genomic diversity among *A. muciniphila* with unique functional characteristics [19].

This approach makes it possible to identify the variable genes associated with the formation of Short Chain Fatty Acids (SCFA), sulfur assimilation, and oxygen tolerance. Additionally, it highlights the mobile genetic components that support

ecological adaptability and host adaptation, such as genomic islands and horizontal gene transfer events [20].

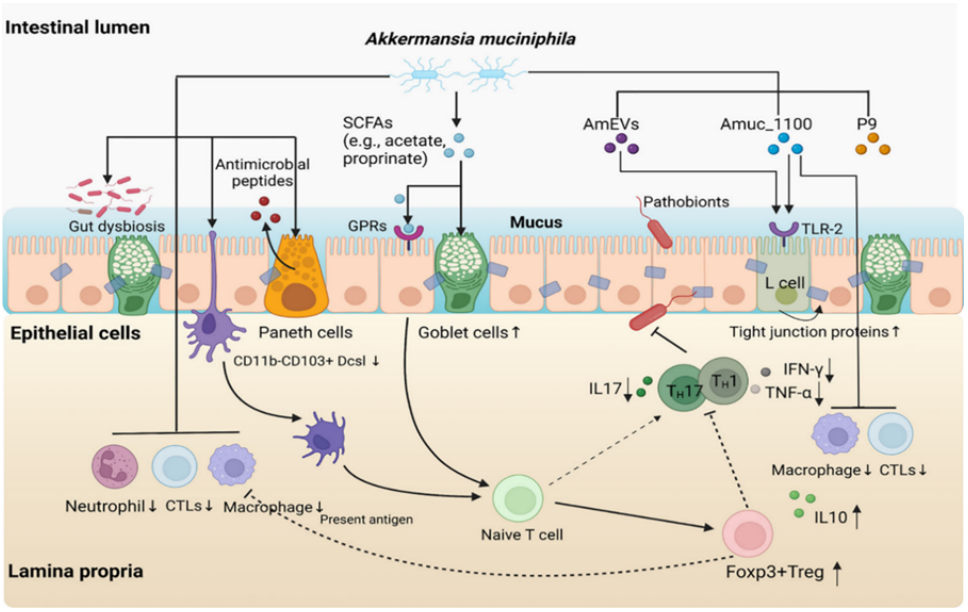

FIGURE 1.5: Protective role of *Akkermansia muciniphila* in intestinal homeostasis [14]

Pangenome Analysis further enhances the results of comparative genomic analysis. Pangenome analysis evaluates the entire gene repertoire of strains of *A. muciniphila*. As discussed above, *A. muciniphila* has an open pangenome, which means that with every new strain, new genes are found, which demonstrates remarkable evolutionary flexibility [5].

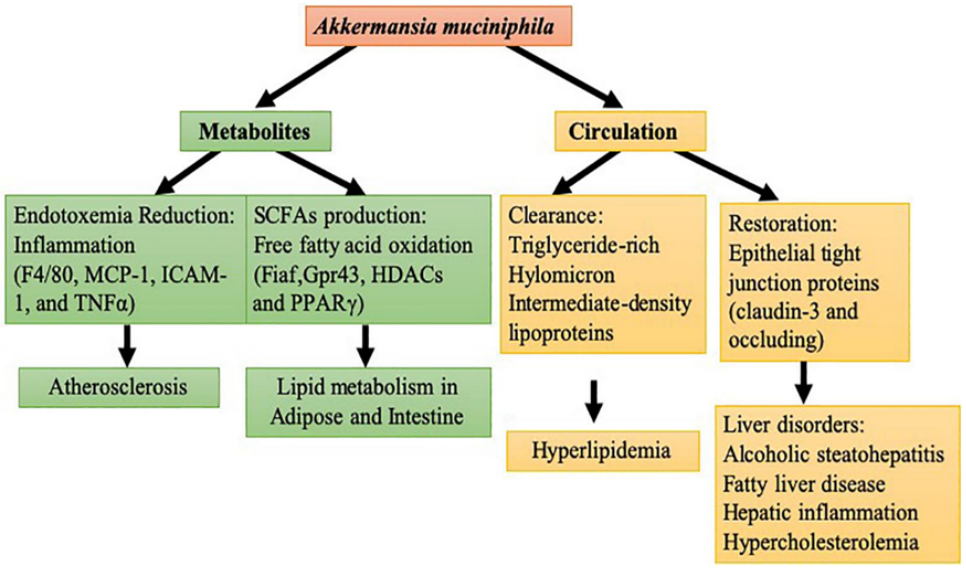

FIGURE 1.6: Role of *Akkermansia muciniphila* in metabolic diseases [17]

The pangenome consists of core genes, which are also considered essential genes present in all the strains, as well as accessory genes that are acquired due to environmental adaptation. Genes encoding for key features such as mucin breakdown, variable surface antigens, and vitamin B12 production are among them [6, 7].

Akkermansia muciniphila is an essential component of intestinal homeostasis due to its mucin-degrading capacity, and is commonly found in the mucus layer of the intestine. Promoting the synthesis of mucins and strengthening tight junction proteins, *A. muciniphila* helps to maintain the integrity of the gut barrier by blocking the entry of inflammatory chemicals and harmful bacteria [21]. Additionally, by encouraging anti-inflammatory responses, especially through the control of regulatory T cells and cytokine profiles, *A. muciniphila* boosts the host immunity [1].

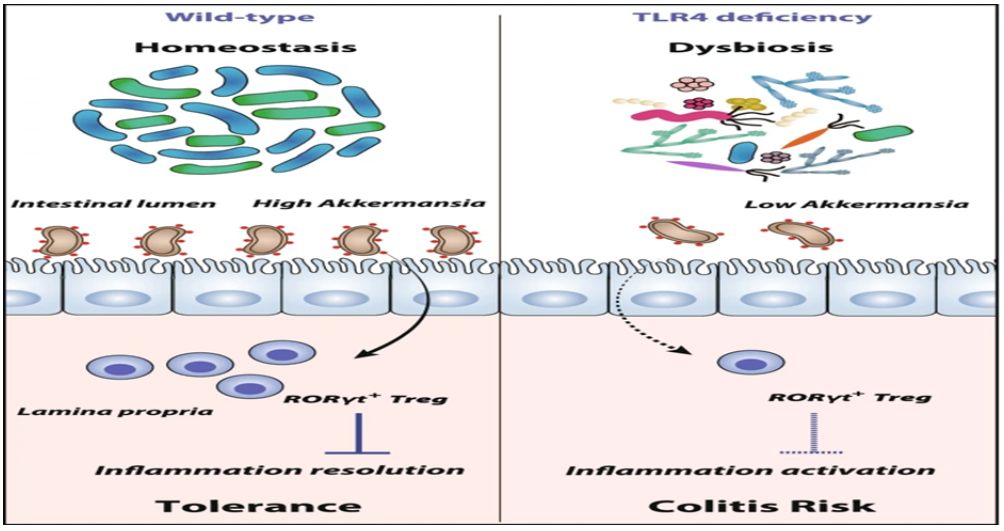

FIGURE 1.7: Role of Akkermencia Muciniphila in gut health [22]

Its protective and health-promoting properties are highlighted by the fact that its abundance has been inversely linked to metabolic illnesses such as obesity, type 2 diabetes, and inflammatory bowel diseases. These observations supports the fact that *A. muciniphila* is a helpful commensal microbe that is necessary for intestinal health.

The capacity of *A. muciniphila* to form biofilms on mucin-rich surfaces, which improves colonization and stress tolerance, demonstrates its extraordinary flexibility

[12]. With the help of outer membrane proteins as Amuc_1100, these biofilms promote immunological regulation and gut adherence without having any harmful effects [23]. Although typically thought to be benign, *A. muciniphila* produces lipooligosaccharides that could cause immunological reactions, and the breakdown of mucin could affect the integrity of the gut when inflammation is present [12]. Long-term safety monitoring is still crucial, although human trials show it is well tolerated in both live and pasteurized versions [8].

1.2 Problem Statement

Akkermansia muciniphila is considered a potential probiotic strain against obesity and other inflammatory and metabolic diseases, but its biofilm production and genome plasticity initiate discussions about its safety to be used as probiotic.

1.3 Research Objectives

This study aims to conduct a detailed comparative genomic analysis of various *Akkermansia muciniphila* strains to highlight the extent of their genetic diversity and to explore the potential links between this diversity and their diverse roles in maintaining human health and contributing to disease pathogenesis.

Specific objectives of the study are given below:

1. To delineate the genetic diversity and population structure of *A. muciniphila*.
2. To characterize its core and accessory genomes, including horizontally acquired genes.
3. To assess strain-specific traits such as antibiotic resistance and virulence potential.
4. To explore secondary metabolites and metabolic pathways, highlighting their probiotic functions.

Chapter 2

Literature Review

This chapter provides a critical review of previous research work on the topic.

2.1 General Characteristics of *Akkermansia muciniphila*

Akkermansia muciniphila is a Gram-negative, oval-shaped, non-motile, and non-spore-forming bacterium, typically measuring between 0.6 and 1.0 μm in diameter. It was first isolated in 2004 from human fecal samples by Muriel Derrien and Willem de Vos at Wageningen University, Netherlands [24]. This species predominantly inhabits the mucus layer of the gastrointestinal tract, particularly near host epithelial cells lining the intestinal mucosa.

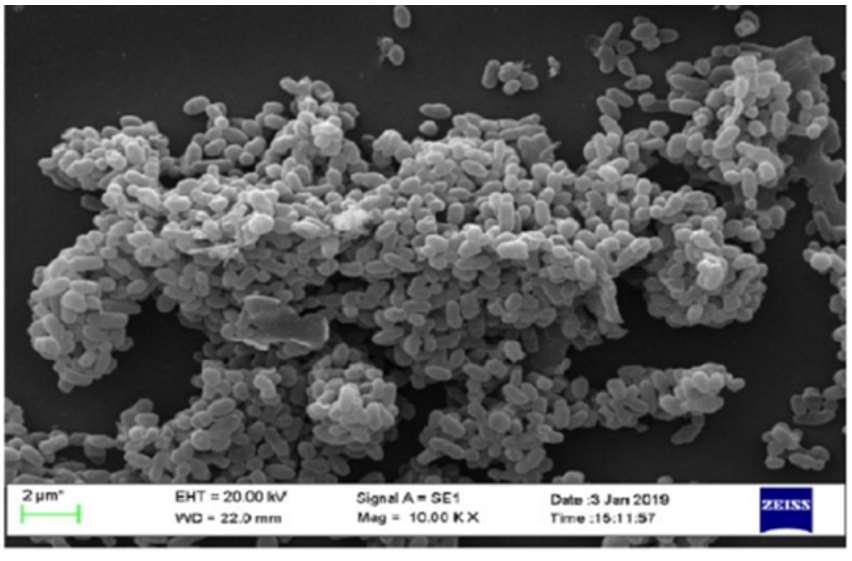

FIGURE 2.1: Microscopic image of *Akkermansia muciniphila* [2]

2.2 General Characteristics of *Akkermansia muciniphila*

Property	Description
Morphology	Oval-shaped, 0.6–1.0 μm diameter
Motility	Non-motile
Spore Formation	Non-spore-forming
Oxygen Requirement	Strictly anaerobic (initially); some strains tolerate low oxygen levels
Primary Energy Source	Mucin
Other Substrates	Glucose, N - acetylglucosamine, N - acetylgalactosamine (with protein source)
Mucinase Activity	Yes, possesses numerous mucinase-encoding genes
Mucus Binding	Lacks canonical mucus-binding domains
Key Surface Molecules	Lipooligosaccharide (LOS), Outer Membrane Protein Amuc_1100

2.2.1 Mucin Degradation

Mucin degradation is the most significant feature of the bacterium. Mucin is a complex glycoprotein that forms a major component of the mucus layer. *A. muciniphila* achieves this via a diverse array of mucinolytic enzymes, including mucinases, glycosidases, sulfatases, and sialidases [24]. These enzymes convert oligosaccharide chains into monosaccharides, acetate, propionate, and butyrate short-chain fatty acids (SCFAs) critical for host gut health and microbial nutrient cycling.

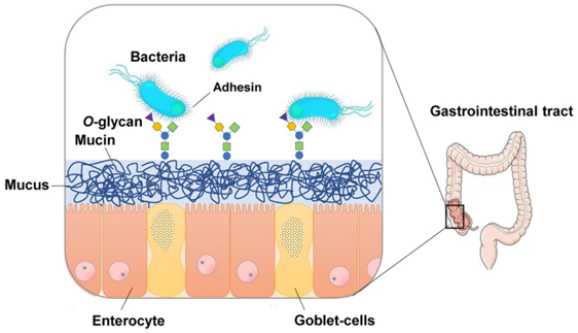

FIGURE 2.2: Interaction at the host mucus layer between secreted mucins and bacteria. Bacteria can bind to mucins through adhesins and/or lectins that recognize O-glycans [15]

Despite its specialization in mucin degradation, *A. muciniphila* lacks canonical mucus-binding domains, suggesting alternative mechanisms for adhering to the mucus layer [2]. It also displays metabolic flexibility, growing on substrates like glucose, N - acetylglucosamine, and N - acetylgalactosamine [23]. Some strains exhibit potential for vitamin B12 biosynthesis [1].

The bacterium interacts with the host immune system via surface molecules. Its lipooligosaccharide (LOS), lacking the O-antigen unit, activates TLR4 and TLR2 signaling pathways [12]. Additionally, the Amuc_1100 outer membrane protein is a potent TLR2 activator [2], indicating a possible role in immune modulation and molecular mimicry.

2.3 Classification of *Akkermansia muciniphila*

A. muciniphila is a critical bacterium present in the gut microbiota and plays a significant role in human health and immunomodulation. *Akkermansia muciniphila* is classified as a Gram-negative, anaerobic, oval-shaped bacterium belonging to the Verrucomicrobia phylum. It is a mucin-degrading bacterium found in the intestinal tract, particularly colonizing the mucus layer of humans and animals.

A. muciniphila is known for its ability to utilize mucin as its primary source of carbon and nitrogen.

TABLE 2.2: Classification of *Akkermansia muciniphila*s

Characteristic	Description
Phylum	Verrucomicrobiota (formerly Verrucomicrobia)
Class	Verrucomicrobiae
Order	Verrucomicrobiales
Family	Akkermansiaceae
Genus	Akkermansia
Species	*Akkermansia muciniphila*

2.4 Genomic Characteristics of *Akkermansia muciniphila*

The genome of a bacterium is characterized by a circular chromosome with no plasmids and possesses 55.8% GC content with unique evolutionary characteristics with high degree of genetic diversity across various strains [1, 19]. Reference strain of *Akkermansia muciniphila* ATCC BAA-835 is 2.6 MB, but the genome size of *A. muciniphila* ranges from approximately 2.6 to 3.0 Mb, reflecting its niche-specific adaptation [1].

The genome encodes for 2176 predicted protein-coding sequences, out of which 65%, i.e., 1408 sequences, have putative functions assigned to them. Most of the CDS or coding sequences are shared with other species of Verrucomicrobia, indicating unique evolutionary features, but high genetic diversity is observed among *A. muciniphila* isolates, even classified as subtypes and distinct clades.

The genome of *A. muciniphila*, due to its open pangenome, is considered to be acquiring new genes by horizontal gene transfer from other bacteria, indicating that *A. muciniphila* possesses a flexible genome.

The bacteria can synthesize various essential amino acids and can metabolize multiple sugars present in mucin. The major categories of protein prodiced aby bacteria are [1]:

Mucin degradation enzymes (e.g., glycoside hydrolases, sialidases, sulfatases)

- SCFA synthesis enzymes (acetate, propionate)
- Adhesion-related proteins
- Vitamin B12 synthesis genes (in some strains)

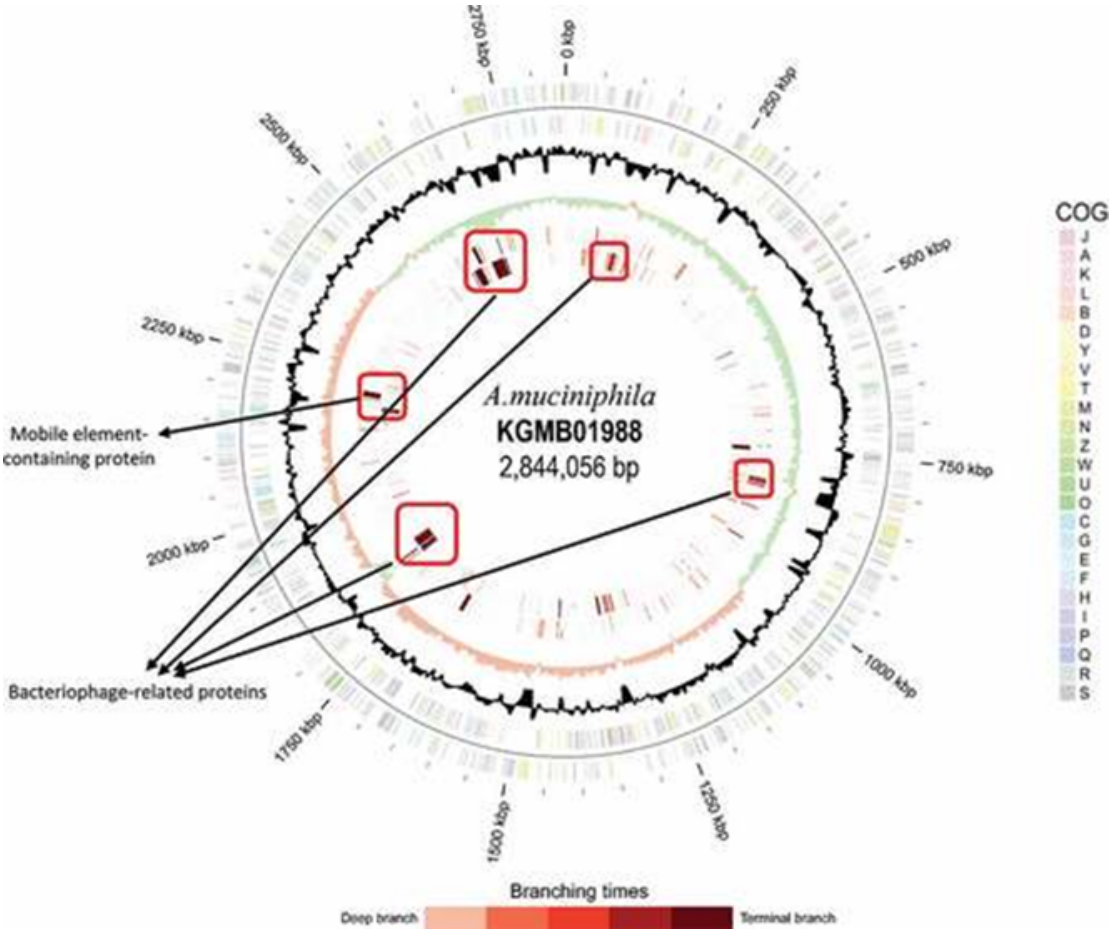

FIGURE 2.3: Graphical circular map of the chromosome of *Akkermansia muciniphila* KGMB01988. From outside to the center: Genes on forward strand (color by COG categories), Genes on reverse strand (color by COG categories), GC content (black), GC skew (light green/Orange), and gained gene families on forward strand and negative strand after speciation (color by branching times) [2]

2.5 Diversity in *Akkermansia muciniphila*

Akkermansia muciniphila is divided into various distinct clades or phylogroups from AmI to AmIV as it exhibits genomic flexibility, hence diversity. Each clade is reported from a single host, mostly, and it rarely happens that two clades co-occur in one host simultaneously. The lack of co-occurrence suggests a competitive dynamic of these phylogroups. Each phylogroup has unique genomic and phenotypic characters based on their metabolic capabilities and host gut environment as well as oxygen sensitivity, details of each phylogroup are as under

- AmI: available as a probiotic and represented by MucT strain
- AmII: exhibits aerotolerance and outcompetes AmI
- AmIII: predominant phylogroup in Chinese population
- AmIV: highly sensitive to oxygen

2.6 Significance of *Akkermansia muciniphila*

The human gut hosts a lot of microbial species that not only reside on the human body but also contribute significantly to maintaining health and avoiding disease. *Akkermansia muciniphila*, due to its mucin degradation properties, plays an important role in maintaining gut health and host homeostasis. It is well reported that the number of *A. muciniphila* is reduced in patients with metabolic disorders and obesity. *A. muciniphila* occupies a crucial ecological niche at the oxic-anoxic interface of the intestinal mucus layer, where it contributes significantly to mucin turnover and nutrient cycling [9]. The species helps maintain gut homeostasis and barrier integrity through mucin degradation, generating SCFAs that benefit both the host and other microbial residents. It supports microbial balance, potentially inhibits pathogen colonization, and may enhance host metabolic health [2].

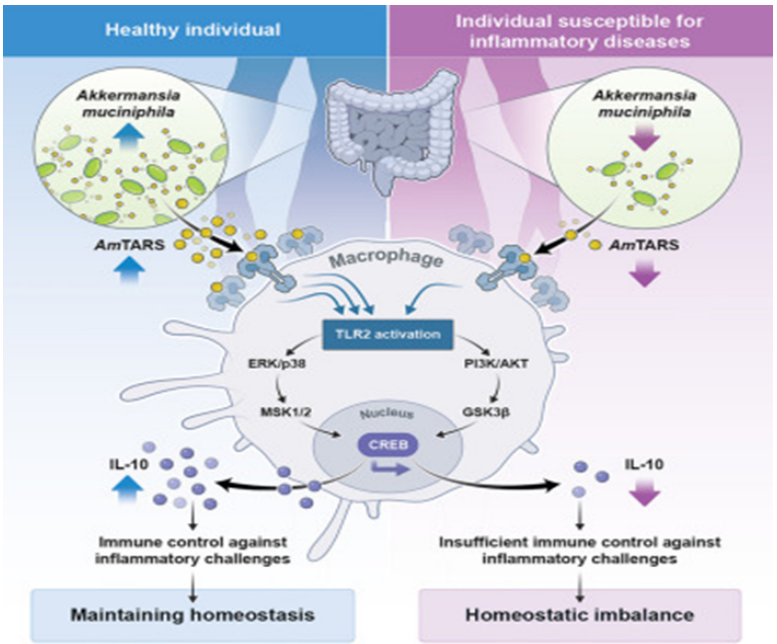

FIGURE 2.4: Impact of low number of *Akkermansia muciniphila* on gut homeostasis [2]

A. muciniphila is also reported to be associated with the enhancement of anti-tumor immunotherapy, inhibition of proinflammatory cytokines, protection of mucin resulting in an intestinal barrier, an increase healthier gut profile by modulating microbial diversity, and enhances glucose metabolism and insulin sensitivity.

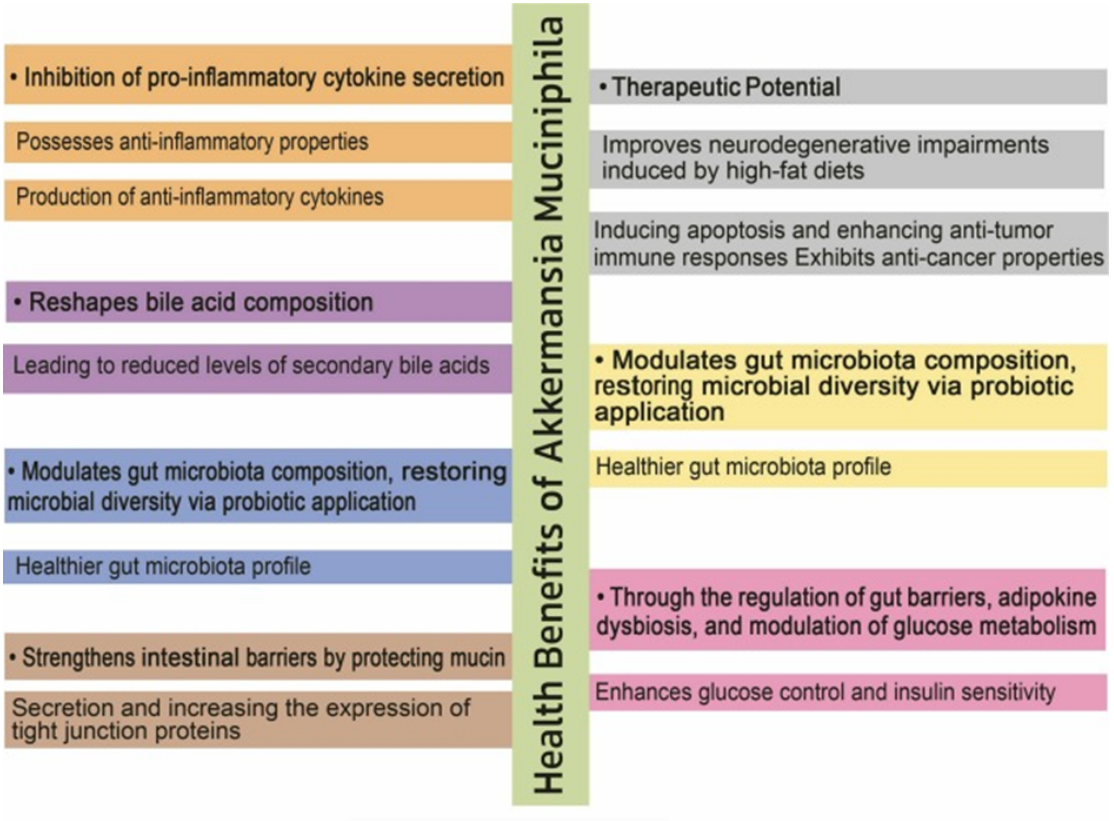

FIGURE 2.5: Health benefits of *Akkermansia muciniphila* [2]

2.7 Significance of *Akkermansia muciniphila* as Probiotic

Though not yet approved as a commercial probiotic, *A. muciniphila* has garnered attention for its health-promoting properties:

- Strengthens the gut barrier by thickening the mucus layer [25]
- Modulates immune responses via TLR signaling [12].
- Exhibits anti-inflammatory properties in obesity, type 2 diabetes, and IBD

- Extracellular vesicles (AmEVs) and the outer membrane protein Amuc_1100 show potential for therapeutic applications, including immune modulation and metabolic benefits [12].

- In case of inflammatory bowel disease, ulcerative colitis, and Crohn's diseaes *A. muciniphila* improves the intestinal mucosal barrier's which includes physical barrier; as well as immune barrier; it stablises the colonisation of healthy bacteria.

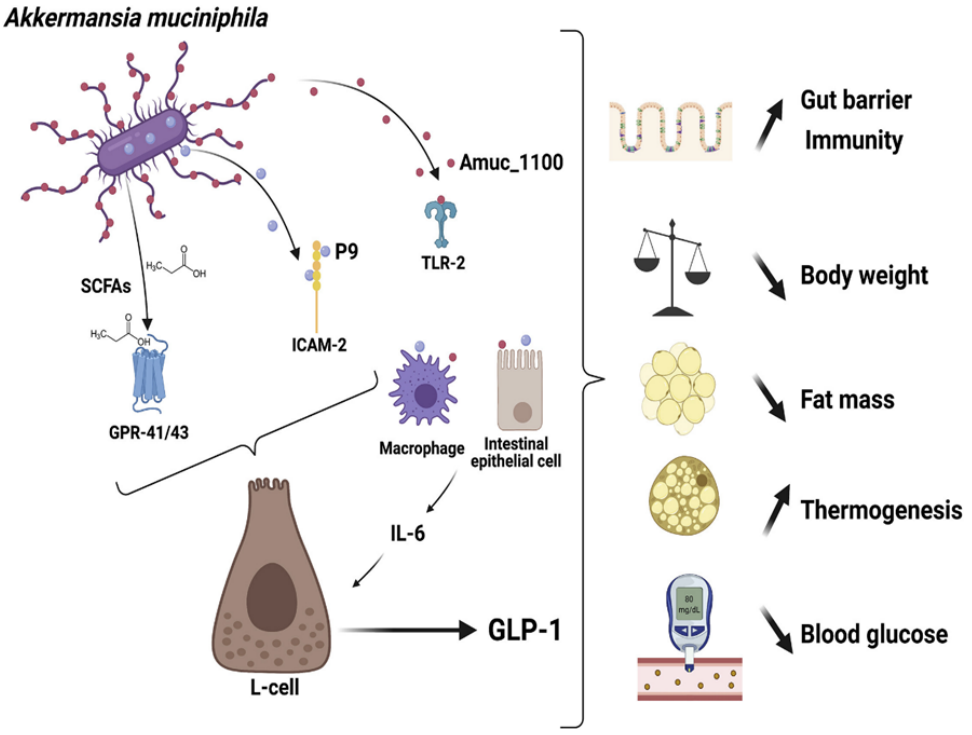

FIGURE 2.6: *A. muciniphila*'s function in inflammatory intestinal disorders [26]

2.8 Genome Plasticity

Bacterial genome is flexible in nature, and show multiple changes during their life time especially when they are living inside host. For survival in the intestinal region, bacterial species encounter various stresses and to survive in this condition they undergo multiple changes in their genome. This adaptability is the key feature contributing in evolution, virulence or pathogenicity or even acquisition of antibiotic resistance. Figure 2.7 summarizes that how bacterial specie can

loose or gain virulent genes contributing in its pathogenicity. Multiple features of bacteria which contribute in genome plasticity include Horizontal Gene Transfer including acquisition of new genes through transformation, transduction and conjugation. Horizontal gene transfer is the quickest way by which bacteria get new genes especially genes associated with antibiotic resistance. Genomic islands and pathogenicity islands are also outcome of horizontal gene transfer. Mobile genetic elements including plasmids, transposons, integerons and insertion sequences also contribute in genomic variations resulting in change in gene expression or altered surface proteins which help in host evasion during infection, one of the important virulence factor. Genome plasticity contributes majorly in antibiotic resistance both addition or shedding of genes, acquisition of virulence factors contributing in evolution, but in general it contributes in making bacteria able to survive in diverse environment.

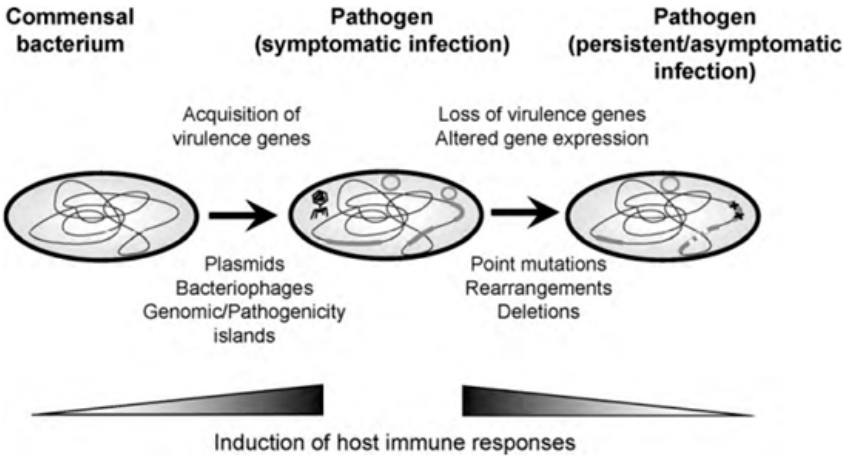

FIGURE 2.7: Impact of genome plasticity on bacterial pathogenesis

Genome plasticity of *A. muciniphila* is evident through the fact that this bacteria has flexible and expanding pangenome, i.e. with each new strain sequenced, new and novel gene families are reported from this bacterial species indicating that there is continuous genome expansion. These new genes mostly lie in category of accessory gene that indicates that genes are acquired as a result of environmental stress. Core gene of this bacteria however shows genes associated with mucin degradation.

Open pangenome suggests functional adaptations in *A. muciniphila* [7] as the bacterial has to survive in multiple and diverse gut environments ranging from human to mice to pig and many more mammals. In addition to variations in gut environment *A. muciniphila* has to cope up with diverse gut microflora, coexisting with which requires *A. muciniphila* to acquire genes or gene clusters/ islands via horizontal gene transfer. The acquired gene cluster or island are rich in mobile genetic elements which are key regions for genetic variations. These unique or accessory genes are the fundamental differences in various phylogroups. Therefore it could be concluded that open pangenome shows that *A. muciniphila* is adaptive enough to survive in gut environment.

2.9 Gap Analysis & Concerns about *A. muciniphila*

Metabolic diseases are one of the most prevalent group of diseases not only in Pakistan but also worldwide whether its Diabetes or nonalcoholic fatty liver disease and more importantly obesity, making these diseases as global health challenges. Pakistan faces a high disease burden because of these diseases due to rapid urbanization and change in food habits. As these diseases have only few solutions including life style changes and potential probiotics. Various bacterial strains [27] are reported to be useful against various disease such as

- Lactobacillus gasseri, Lactobacillus rhamnosus, and Bifidobacterium breve against obesity.
- Lactobacillus casei, Bifidobacterium lactis against Type II diabetes
- Lactobacillus plantarum, Bifidobacterium bifidum, Lactobacillus acidophilus and Lactobacillus reuteri against metabolic syndrome
- Lactobacillus rhamnosus and Bifidobacterium longum against nonalcoholic fatty liver For all these probiotic strains it is observed that although these bacterial species have shown promising results but individual variations

among strains matters. Similarly, not all probiotics work for everyone. Due to these reasons new bacterial species other than lactobacillus were explored and *A. muciniphila* is considered as next generation probiotic as it offers health benefits through different mechanisms.

- A good probiotic bacterial strain requires [24] to be able to Tolerate harsh environment of gut, i.e. the bacterial strain should have acid resistance and stomach has very low pH. In addition resistance to bile salts is also required.there are lot of different enzymes present in gut and bacterial strain should be able to cope with those proteolytic enzymes
- In order to colonize (transient or permanent) and to compete with other bacteria in gut, the bacterial species should be able to adhere with intestinal epithelium or mucosal surfaces
- The bacterial strain should exhibit antimicrobial activity to out compete competitor bacteria by using nutrients and attachements sites as well as production of bacteriocins or organic acids to inhibit growth of competitors.
- Probiotic strain should be able to stimulate immune system for enhancement of innate and adaptive immunity, in addition anti-inflammatory properties should also be exhibited
- GRAS: to be considered as probiotic, bacterial strain should be Generally Recognized As Safe i.e. it should not produce pathogenicity or toxicity.
- The probiotic strain should have some health benefits either help in digestion or metabolisms or reduce toxicity etc.

A. muciniphila is considered to have next generation probiotic potential [28] due to the reason that it imparts health benefits by supporting gut barrier by living in mucus layer and using mucins as food and results in decreasing the chance of gut inflammation and endotoxemia. Clinical trials have shown that pasteurized form of *A. muciniphila* can result in improved insulin sensitivity, reduced body weight and decreased cholesterol and inflammation. *A. muciniphila* is also reported to produce short chain fatty acids [29] which modulate gut health and compete with

pathogens while it can interact with immune cells and metabolic receptors such as TLR2.

As mentioned in last section genome of *A. muciniphila* shows genome plasticity which enables it to survive in mucosal layer and in harsh gut environment but it highlights a concern to use *A. muciniphila* as probiotic as genome plasticity often results in strain level genetic variations which affects probiotic efficacy as well as its immunoregulations thus raining concerns about safety of probiotic. Therefore using *A. muciniphila* as probiotic or in microbiome based therapies it is recommended to closely monitor genomic changes in each strain. The is the reason because of which *A. muciniphila* is not yet approved as probiotic in many regions and long term safety is still under consideration [7].

Chapter 3

Material and Method

This study employed a comparative genomics approach to investigate the genetic diversity and potential functional variations among *Akkermansia muciniphila* strains. The methodology involved several key steps, including genome retrieval and selection, pangenome analysis, and functional annotation.

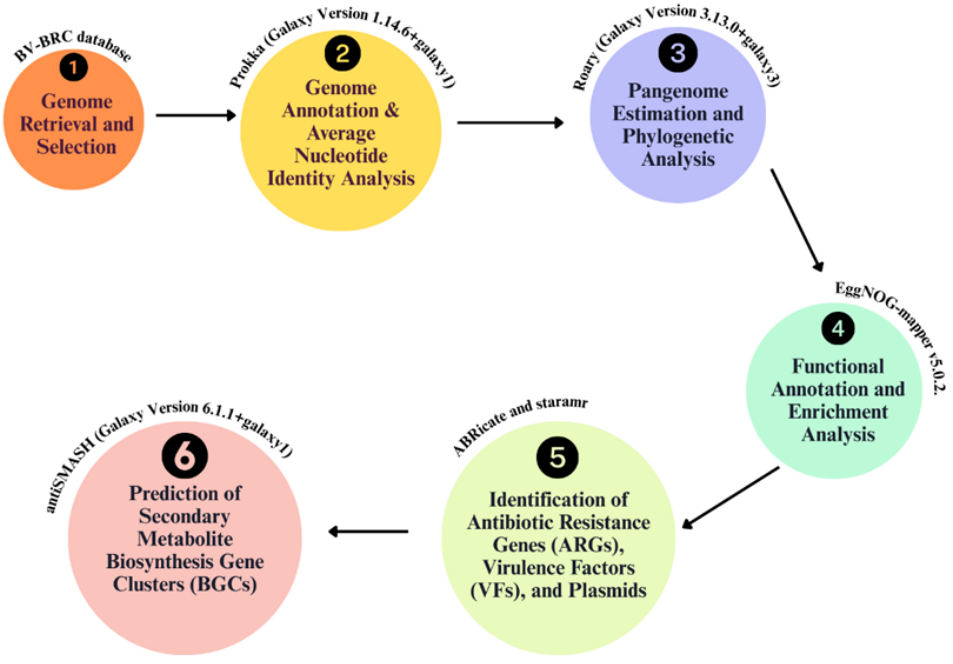

FIGURE 3.1: Graphical abstract of pangenome analysis of *Akkermansia muciniphila* [2]

3.1 Genome Retrieval and Selection

A comprehensive set of 280 *Akkermansia muciniphila* genomes was retrieved from the BV-BRC (Bacterial and Viral Bioinformatics Resource Center) database on March 16, 2025 [27]. This dataset comprised 71 complete genomes, including the reference genome JCM 30893, and 209 Whole Genome Sequencing (WGS) entries. To ensure the relevance of this study to human health, only genomes derived from human hosts were selected. Genomes originating from other hosts, such as black rats, chickens, mice, patents, and pigs, were excluded. Stringent quality control measures were applied during the retrieval process. Only high-quality genomes were included in the dataset, while deprecated and poor-quality genomes were excluded. This selection criterion aimed to maximize the reliability and accuracy of subsequent comparative genomic analyses. The resulting dataset of 280 *A. muciniphila* genomes provided a robust foundation for investigating the species' genetic diversity.

3.2 Genome Annotation and Average Nucleotide Identity (ANI) Analysis

Gene prediction and functional annotation were performed using Prokka (Galaxy Version 1.14.6+galaxy1) [30]. Prokka utilizes Prodigal for the identification of protein-coding regions, followed by functional prediction through similarity searches against various protein and protein domain databases [31]. The annotation process included the identification of RNA genes, specifically tRNA and rRNA, using Barrnap and Aragorn, respectively. Analyses were conducted using default parameters, with a similarity e-value cut-off of 1e-06. The bacterial kingdom and genetic code 11 were specified to ensure accurate annotation for the *A. muciniphila* genomes.

Average Nucleotide Identity (ANI) was calculated using FastANI (Galaxy Version 1.3) to determine the taxonomic relatedness of the retrieved genomes (Jain et

al., 2018). This analysis was performed to validate the species classification and to ensure the selection of genomes within the *A. muciniphila* species boundary. All selected genomes were compared against the reference genome *A. muciniphila* JCM 30893. A standard ANI cut-off of >95% was applied to define the species boundary. Genomes exhibiting ANI values below this threshold were excluded from subsequent analyses, ensuring the inclusion of only actual *A. muciniphila* strains in the comparative genomic study. This selection aimed to capture a broad spectrum of the species' genetic diversity while excluding highly redundant or potentially misclassified genomes.

3.3 Pangenome Estimation and Phylogenetic Analysis

Pangenome analysis was conducted to characterize the genomic diversity of *A. muciniphila*. This analysis was performed on both the initial set of genomes and the shortlisted set of genomes, which were selected based on Average Nucleotide Identity (ANI) and gene presence-absence patterns. Roary (Galaxy Version 3.13.0+galaxy3) was employed for pangenome estimation [32]. The minimum percentage identity for BLASTp was set to 70%, and the percentage of isolates required for a gene to be considered part of the core genome was set to 95%. These parameters were chosen to ensure robust identification of orthologous gene clusters and to define a stringent core genome. The pangenome was determined using a gene-based approach, where genes present in all strains were classified as core genes, genes present in a subset of strains as accessory genes, and genes unique to a single strain as unique genes. The resulting pangenome was further characterized to assess its openness or closedness, indicating the potential for further gene acquisition within the species. To visualize the pangenome structure, a phylogenetic tree was generated and compared against a presence/absence matrix of core and accessory genes using 'roary_plots.py v0.1.0'. This allowed for the assessment of the relationship between genomic variation and gene content. Additionally, the dynamics of the pangenome, specifically the distribution of conserved and unique

genes, were visualized using a custom R script, 'create_pan_genom_plots.R'. This script, which utilizes the R programming language and the ggplot2 library, processed the Roary output files (*.Rtab) to generate plots depicting the variation of the pangenome as genomes were added in random orders. This analysis provided insights into the openness or closedness of the *A. muciniphila* pangenome and the rate of gene acquisition or loss. The core genome multiple sequence alignment was used for the inference of phylogenetics using iTOL v7. This analysis provides insights into the genomic diversity and evolutionary dynamics of *A. muciniphila*.

3.4 Functional Annotation and Enrichment Analysis

The functional annotation of both pan genes was performed using EggNOG-mapper v5.0.2. A minimum e-value threshold of 0.001 was applied to ensure the reliability of the annotation [28]. EggNOG-mapper was specifically utilized to assign genes to functional categories based on the EggNOG database, providing insights into the metabolic pathways, cellular processes, and other functional attributes of the analyzed genomes.

3.5 Identification of Antibiotic Resistance Genes (ARGs), Virulence Factors (VFs), and Plasmids

Antibiotic resistance genes (ARGs), virulence factors (VFs), and plasmids were identified using ABRicate (Galaxy Version 1.0.1) and staramr (Galaxy Version 0.11.0 + galaxy0). These tools utilized the VFDB (Virulence Factors Database),

CARD (Comprehensive Antibiotic Resistance Database), ResFinder, and PlasmidFinder databases for comparative sequence analysis [33–36]. A minimum coverage of 30% and a minimum sequence identity of 75% were applied as thresholds for the identification of these genetic elements.

3.6 Prediction of Secondary Metabolite Biosynthesis Gene Clusters (BGCs)

Antimicrobial peptides and secondary metabolite biosynthesis gene clusters (BGCs) were predicted using antiSMASH (Galaxy Version 6.1.1+galaxy1) [25]. AntiSMASH was employed to identify genomic regions with the potential to produce secondary metabolites such as antibiotics, siderophores, and other bioactive compounds. This analysis contributed to the understanding of the diverse metabolic capabilities of *A. muciniphila* strains.

Chapter 4

Results and Discussion

4.1 General Characteristics of the Genomes

The genomic characteristics of the retrieved 280 *A. muciniphila* strains revealed significant variability. The average genome size was determined to be 2.84 Mb, with observed sizes ranging from 1.3 Mb in strain N1191.18, representing the smallest genome, to 3.3 Mb in strain Akk1863, the largest. The average guanine-cytosine (GC) content was 55.8%, exhibiting a range from 53.12% in strain BIOML-A18 (minimum) to 58.46% in strain CSUN-56 AmIII (maximum). The number of contigs varied substantially, with strain P048L1-35 displaying the highest number at 558. The number of coding sequences (CDS) also showed considerable variation, ranging from 1306 in strain N1191.18 (comprising 13 contigs) to 3336 in strain DFI.6.92 (comprising 462 contigs). The average number of CDS across all strains was calculated to be 2829.

4.2 Species Delimitation and Exclusion of Divergent Strains

Species boundaries were determined using Average Nucleotide Identity (ANI) analysis, with a standard cutoff of >95%. This analysis identified 53 out of 280 strains

exhibiting ANI values below this threshold, indicating potential interspecies divergence (Table 4.1). Subsequent gene presence-absence analysis revealed that these 53 strains formed a distinct, distant cluster, displaying significant genetic dissimilarity from the remaining strains. Consequently, these 53 divergent strains were excluded, resulting in a refined dataset of 227 strains for further analysis. The phylogenetic relationships among the accessory binary genes of the excluded divergent strains are illustrated in Figure 4.1 and Figure 4.2.

TABLE 4.1: Table of ANI results (>95%) used for species boundary determination of 227 selected strains, detailing Query and Reference strain comparisons, ANI values, Fragment alignment, and Alignment Length.

No	**Query**	**Reference**	**ANI**	**Fragments**	**Alignment _Length**
1	40 _2 _59	JCM_30893	96.4741	222	227
2	43 _2 _06	JCM_30893	96.583	321	336
3	GP43	JCM_30893	96.9739	823	927
4	D6 _3 _FMU	JCM_30893	96.9865	824	907
5	CSUN _12 _AmI	JCM_30893	96.9895	832	912
6	H1	JCM_30893	96.9984	827	913
7	H2	JCM_30893	97.0048	841	939
8	GP04	JCM_30893	97.0323	854	935
9	NBRC _115032	JCM_30893	97.0327	848	947
10	Akk0500b	JCM_30893	97.0347	842	929
11	43 _1 _43	JCM_30893	97.0456	837	920
12	Marseille _P5162	JCM_30893	97.0479	833	947
13	N1170 _24	JCM_30893	97.0536	822	895
14	GP03	JCM_30893	97.0561	852	932
15	Map _156 _012	JCM_30893	97.062	740	798
16	Map _3 _010	JCM_30893	97.0658	800	869
17	CBA5201	JCM_30893	97.0801	863	953
18	OMN02 _HAM _0033	JCM_30893	97.0885	858	988

continued on next page

No	Query	Reference	ANI	Fragments	Alignment _Length
19	Akk16115	JCM_30893	97.0969	855	1000
20	DOME _MAG _9332	JCM_30893	97.0978	820	884
21	EB _AMDK _16	JCM_30893	97.1042	834	923
22	Akk1613	JCM_30893	97.1088	856	996
23	C4 _44	JCM_30893	97.1117	814	894
24	B189 _010	JCM_30893	97.1124	807	891
25	Akk1610	JCM_30893	97.1131	854	1000
26	Akk16145	JCM_30893	97.1204	857	1000
27	Akk1616	JCM_30893	97.1216	857	1000
28	EB _AMDK _18	JCM_30893	97.1246	831	923
29	Map _48 _015	JCM_30893	97.1248	806	876
30	GP36	JCM_30893	97.1253	847	913
31	EB _AMDK _17	JCM_30893	97.1264	831	923
32	EB _AMDK _15	JCM_30893	97.1273	831	923
33	39 _2 _104	JCM_30893	97.1286	863	998
34	DFI _6 _92	JCM_30893	97.1396	851	952
35	L2 _040 _365G1 _dasL2 _040 _3651 _dasL2 _040 _365G1 _concoct _4	JCM_30893	97.1439	818	896
36	AK32	JCM_30893	97.1464	875	1001
37	COPD431	JCM_30893	97.1632	833	908
38	Akk2030	JCM_30893	97.164	836	934
39	UBG017	JCM_30893	97.1685	800	949
40	C043 _18	JCM_30893	97.1686	855	969
41	Akk2090	JCM_30893	97.1709	835	934
42	C039 _28	JCM_30893	97.174	861	980

continued on next page

No	Query	Reference	ANI	Fragments	Alignment _Length
43	Akk1990	JCM_30893	97.1744	836	934
44	DFI _6 _69	JCM_30893	97.1779	850	952
45	Map _95 _020	JCM_30893	97.1782	765	857
46	HB03	JCM_30893	97.18	832	902
47	B32 _bin _24	JCM_30893	97.1841	797	879
48	COPD433	JCM_30893	97.1911	799	848
49	39 _1 _31	JCM_30893	97.1987	846	942
50	CSUN _59 _AmI	JCM_30893	97.203	838	955
51	C029 _35	JCM_30893	97.2036	809	892
52	D1169bin22	JCM_30893	97.2071	853	954
53	DFI _5 _32	JCM_30893	97.2128	845	950
54	C047 _04	JCM_30893	97.2142	861	1010
55	D3120bin17	JCM_30893	97.2231	851	965
56	CLA _AP _H31	JCM_30893	97.2249	832	924
57	14 _2 _60	JCM_30893	97.2274	845	959
58	MGS _154	JCM_30893	97.2287	808	871
59	GP21	JCM_30893	97.2299	871	939
60	47 _2 _14	JCM_30893	97.2311	849	923
61	Map _12 _014	JCM_30893	97.2414	836	913
62	B86 _013	JCM_30893	97.2559	740	834
63	Akk2650	JCM_30893	97.2581	855	944
64	Akk0880	JCM_30893	97.2603	858	988
65	COPD435	JCM_30893	97.2621	592	625
66	Akk1756	JCM_30893	97.2642	872	980
67	COPD434	JCM_30893	97.2669	831	913
68	Map _115 _032	JCM_30893	97.2678	789	848
69	Map _111 _039	JCM_30893	97.268	790	854
70	GP05	JCM_30893	97.2727	833	925
71	B210 _019	JCM_30893	97.2826	822	900

continued on next page

No	Query	Reference	ANI	Fragments	Alignment _Length
72	COPD430	JCM_30893	97.2855	854	936
73	225S _18FAA	JCM_30893	97.2858	834	938
74	B14 _bin _17	JCM_30893	97.2877	812	870
75	Akk1683	JCM_30893	97.29	843	999
76	GP39	JCM_30893	97.298	839	949
77	Map _2 _013	JCM_30893	97.2981	805	891
78	Akk1576	JCM_30893	97.3011	859	992
79	Akk1570	JCM_30893	97.3054	856	988
80	N1157 _9	JCM_30893	97.3087	846	916
81	GP41	JCM_30893	97.3094	834	916
82	P002 _10	JCM_30893	97.3126	833	913
83	B9 _maxbin _004	JCM_30893	97.3146	805	893
84	B13 _bin _15	JCM_30893	97.3172	786	843
85	Map _38 _014	JCM_30893	97.3208	829	908
86	CSUN _33 _AmI	JCM_30893	97.3264	848	922
87	Akk14745a	JCM_30893	97.3286	844	932
88	GP06	JCM_30893	97.3295	832	915
89	47 _1 _05	JCM_30893	97.3304	857	941
90	GP38	JCM_30893	97.3326	835	941
91	YGMCC2645	JCM_30893	97.3458	826	907
92	Marseille _P6566	JCM_30893	97.3514	836	907
93	Marseille _P7245	JCM_30893	97.3514	836	907
94	GP29	JCM_30893	97.3583	822	909
95	Akk14745b	JCM_30893	97.3601	844	932
96	KGMB01988	JCM_30893	97.3621	846	948
97	C017 _30	JCM_30893	97.3691	854	951
98	KGMB01990	JCM_30893	97.3695	845	948
99	N1164 _13	JCM_30893	97.3711	830	899
100	KGMB01989	JCM_30893	97.3765	844	948

continued on next page

No	Query	Reference	ANI	Fragments	Alignment _Length
101	KGMB02009	JCM_30893	97.3796	844	948
102	BIOML _A32	JCM_30893	97.3845	832	907
103	Akk13715	JCM_30893	97.3896	838	934
104	Akk1376	JCM_30893	97.3898	839	934
105	Akk1370	JCM_30893	97.3923	837	934
106	BIOML _A2	JCM_30893	97.4092	835	908
107	BIOML _A22	JCM_30893	97.4101	835	906
108	EB _AMDK _14	JCM_30893	97.4199	845	921
109	BIOML _A19	JCM_30893	97.421	831	909
110	B30 _bin _40	JCM_30893	97.4283	679	703
111	BIOML _A33	JCM_30893	97.4305	827	916
112	EB _AMDK _8	JCM_30893	97.431	843	941
113	EB _AMDK _11	JCM_30893	97.4321	844	921
114	BIOML _A15	JCM_30893	97.4335	832	905
115	BIOML _A28	JCM_30893	97.4363	832	909
116	BIOML _A14	JCM_30893	97.4392	830	917
117	BIOML _A20	JCM_30893	97.4462	829	904
118	BIOML _A23	JCM_30893	97.4491	831	906
119	BIOML _A16	JCM_30893	97.4495	830	909
120	BIOML _A40	JCM_30893	97.4513	830	912
121	BIOML _A9	JCM_30893	97.4521	828	931
122	BIOML _A44	JCM_30893	97.4523	825	904
123	BIOML _A41	JCM_30893	97.4531	833	909
124	BIOML _A1	JCM_30893	97.4548	832	905
125	BIOML _A25	JCM_30893	97.462	832	912
126	BIOML _A18	JCM_30893	97.4627	830	1026
127	EB _AMDK _13	JCM_30893	97.4631	839	921
128	BIOML _A39	JCM_30893	97.4639	832	907

continued on next page

No	Query	Reference	ANI	Fragments	Alignment _Length
129	OB21 _FAA _NB _28	JCM_30893	97.4643	833	950
130	BIOML _A26	JCM_30893	97.4649	829	906
131	29 _2 _27	JCM_30893	97.4659	746	811
132	bin _844	JCM_30893	97.4663	845	925
133	EB _AMDK _10	JCM_30893	97.4664	837	921
134	BIOML _A13	JCM_30893	97.4666	830	919
135	BIOML _A38	JCM_30893	97.4669	830	908
136	BIOML _A21	JCM_30893	97.4673	832	983
137	BIOML _A6	JCM_30893	97.4713	832	913
138	BIOML _A42	JCM_30893	97.4763	832	907
139	BIOML _A8	JCM_30893	97.4776	826	906
140	Marseille _P9642	JCM_30893	97.4782	828	896
141	BIOML _A4	JCM_30893	97.4787	831	912
142	BIOML _A37	JCM_30893	97.479	830	909
143	BIOML _A36	JCM_30893	97.4806	826	905
144	BIOML _A30	JCM_30893	97.4808	828	906
145	GP01	JCM_30893	97.4828	833	920
146	EB _AMDK _12	JCM_30893	97.4844	842	921
147	P012L1 _26	JCM_30893	97.4859	758	797
148	BIOML _A43	JCM_30893	97.4868	829	907
149	BIOML _A45	JCM_30893	97.4878	831	908
150	BIOML _A7	JCM_30893	97.4895	831	906
151	17 _1 _17	JCM_30893	97.4907	834	916
152	BIOML _A24	JCM_30893	97.4915	829	906
153	BIOML _A27	JCM_30893	97.4935	829	906
154	BIOML _A11	JCM_30893	97.4964	829	912
155	BIOML _A29	JCM_30893	97.498	826	905
156	BIOML _A31	JCM_30893	97.4983	827	905

continued on next page

No	Query	Reference	ANI	Fragments	Alignment _Length
157	am _0171	JCM_30893	97.4995	823	902
158	EB _AMDK _19	JCM_30893	97.502	829	908
159	DFI _9 _24	JCM_30893	97.5024	806	887
160	BIOML _A34	JCM_30893	97.5033	827	906
161	COPD429	JCM_30893	97.5069	816	878
162	EB _AMDK _22	JCM_30893	97.5077	829	908
163	BIOML _A35	JCM_30893	97.5086	827	906
164	EB _AMDK _21	JCM_30893	97.5094	826	908
165	EB _AMDK _20	JCM_30893	97.5116	827	908
166	BIOML _A3	JCM_30893	97.5137	826	908
167	29 _1 _57	JCM_30893	97.5239	832	905
168	N1191 _18	JCM_30893	97.6271	419	433
169	2218st1 _C6 _22035	JCM_30893	98.0324	811	880
170	2218st1 _G7 _22035	JCM_30893	98.072	796	868
171	2218st1 _A7 _22035	JCM_30893	98.0789	814	886
172	2218st1 _D7 _22035	JCM_30893	98.0948	821	896
173	Marseille _P9184	JCM_30893	98.5315	843	910
174	COPD426	JCM_30893	98.5423	812	877
175	CSUN _7 _AmI	JCM_30893	98.5908	836	933
176	EB _AMDK _7	JCM_30893	98.5989	853	933
177	rmaize _MAXBIN _ _011	JCM_30893	98.6095	876	935
178	GP27	JCM_30893	98.6382	846	901
179	Marseille _Q2586	JCM_30893	98.6394	869	915
180	CHE528 _008	JCM_30893	98.6422	822	873

continued on next page

No	Query	Reference	ANI	Fragments	Alignment _Length
181	GP14	JCM_30893	98.6433	831	895
182	P058 _15	JCM_30893	98.6436	830	914
183	GP28	JCM_30893	98.665	858	901
184	P015L1 _1	JCM_30893	98.6703	827	871
185	OF08 _9	JCM_30893	98.698	619	656
186	Akk2670	JCM_30893	98.7119	864	920
187	N1167 _9	JCM_30893	98.7122	843	906
188	GP02	JCM_30893	98.7167	846	885
189	P059 _25	JCM_30893	98.7258	811	892
190	MRD1520	JCM_30893	98.7397	849	887
191	COPD428	JCM_30893	98.7409	842	881
192	H4	JCM_30893	98.7473	820	859
193	H6	JCM_30893	98.7498	844	890
194	I42XC9	JCM_30893	98.7502	816	852
195	COPD432	JCM_30893	98.7505	841	888
196	Akk1713	JCM_30893	98.7553	860	908
197	H5	JCM_30893	98.7582	822	869
198	DOME _MAG _8722	JCM_30893	98.7593	819	871
199	DSM _22959	JCM_30893	98.7593	853	888
200	Akk0500a	JCM_30893	98.7634	859	908
201	H3	JCM_30893	98.7634	823	876
202	Akk0200	JCM_30893	98.7665	851	887
203	AM06	JCM_30893	98.7679	853	888
204	AMU	JCM_30893	98.768	853	888
205	Akk007	JCM_30893	98.7688	853	888
206	I42XC8	JCM_30893	98.7751	817	849
207	P061 _22	JCM_30893	98.7836	784	839
208	S05B _meta _bin	JCM_30893	98.7991	787	814

continued on next page

No	Query	Reference	ANI	Fragments	Alignment _Length
209	S05 _meta _bin	JCM_30893	98.8084	766	792
210	GP16	JCM_30893	98.8094	850	885
211	P048L1 _35	JCM_30893	98.8188	484	491
212	AKK	JCM_30893	98.8349	863	941
213	S05A _meta _bin	JCM_30893	98.8438	779	813
214	P010L1 _2	JCM_30893	98.8544	686	705
215	P052L1 _43	JCM_30893	98.8699	707	722
216	P037 _81	JCM_30893	98.9227	820	862
217	GP40	JCM_30893	98.9694	884	925
218	GP19	JCM_30893	98.987	880	925
219	VEG01 _HAM	JCM_30893	99.015	874	925
220	GP23	JCM_30893	99.4819	905	950
221	GP35	JCM_30893	99.5369	898	948
222	GP20	JCM_30893	99.5542	905	934
223	NBRC _114322	JCM_30893	99.7536	890	916
224	Akk0096	JCM_30893	99.7801	892	918
225	P007L1 _24	JCM_30893	99.819	872	878
226	JCM_30893	JCM_30893	100	955	958
227	L3 _069 _242G1 _dasL3 _069 _242G1 _maxbin _044s _strain	JCM_30893	99.8926	869	923

4.3 Pangenome Analysis Reveals Significant Reduction in Gene Pool Following Strain Exclusion

Pangenome analysis was initially conducted on the complete dataset of 280 genomes. This analysis revealed a core genome comprising 429 genes present in 95-100% of

the strains. The soft-core genome, defined as genes present in 94-95% of the strains, consisted of 53 genes. The shell genome, encompassing genes present in 15-94% of the strains, contained 2811 genes. The cloud genome, representing genes present in less than 15% of the strains, comprised 21743 genes. The total pangenome size for the 280 genomes was determined to be 25036 genes.

Following the exclusion of 53 divergent strains, a subsequent pangenome analysis was performed on the remaining 227 genomes. This refined analysis demonstrated a substantial shift in pangenome characteristics. The core genome increased to 1237 genes. The soft-core genome decreased to 41 genes. The shell genome was reduced to 1725 genes. Notably, the cloud genome exhibited a significant reduction, decreasing to 10441 genes. The total pangenome size for the 227 genomes was calculated to be 13444 genes. These results indicate a substantial reduction in the accessory gene pool and an increase in core gene conservation following the exclusion of divergent strains, thus, highlighting the importance of accurate species delimitation in pangenome analysis.

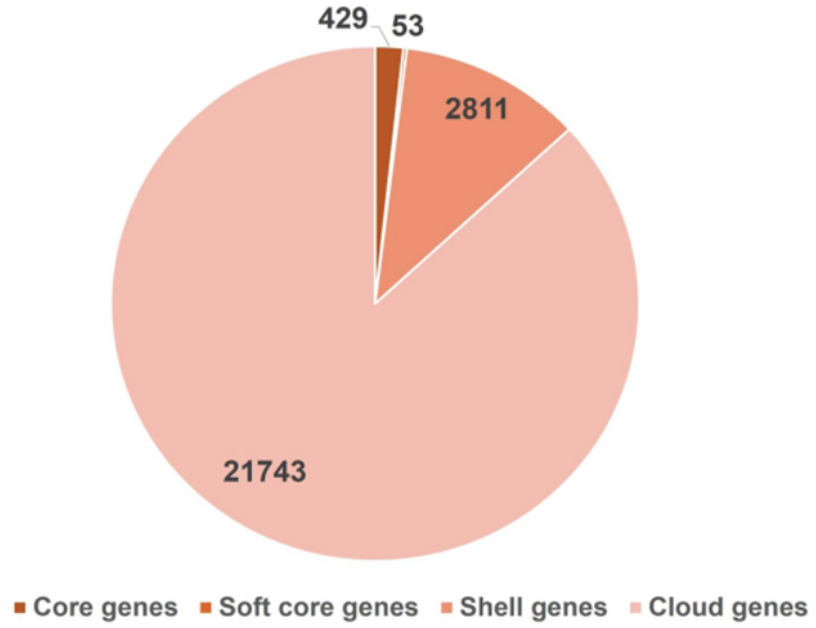

FIGURE 4.1: Comparative pangenome analysis represented as pie charts, showing the distribution of core, soft core, shell, and cloud genes in the initial dataset of 280 *A. muciniphila* genomes [49]

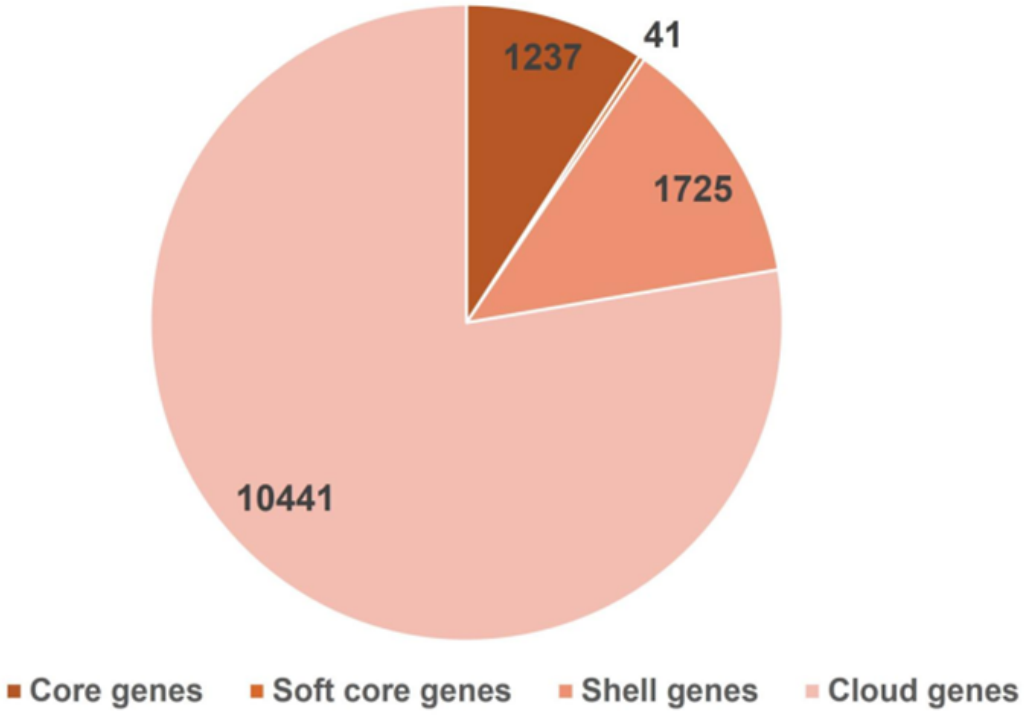

FIGURE 4.2: Comparative pangenome analysis represented as pie charts, showing the distribution of core, soft core, shell, and cloud genes in the refined dataset of 227 genomes following the exclusion of divergent strains, highlighting the reduction in cloud genes and increase in core genes after removing divergent strains[51]

Pangenome analysis of the 280 *A. muciniphila* genomes revealed a dynamic and diverse genomic landscape. The cumulative gene count in the pangenome exhibited a continuous increase with the addition of each genome, as illustrated in Figure 4.2, indicative of an open pangenome. This observation was further supported by the consistent identification of new genes throughout the analysis, particularly prominent during the initial stages, as shown in Figure 4.3, highlighting substantial initial genomic diversity.

In contrast, the number of conserved genes rapidly reached a plateau, as depicted in Figure 4.5 and Figure 4.6, suggesting a relatively stable core genome was established early in the analysis, with a decrease in number as new genomes were added.

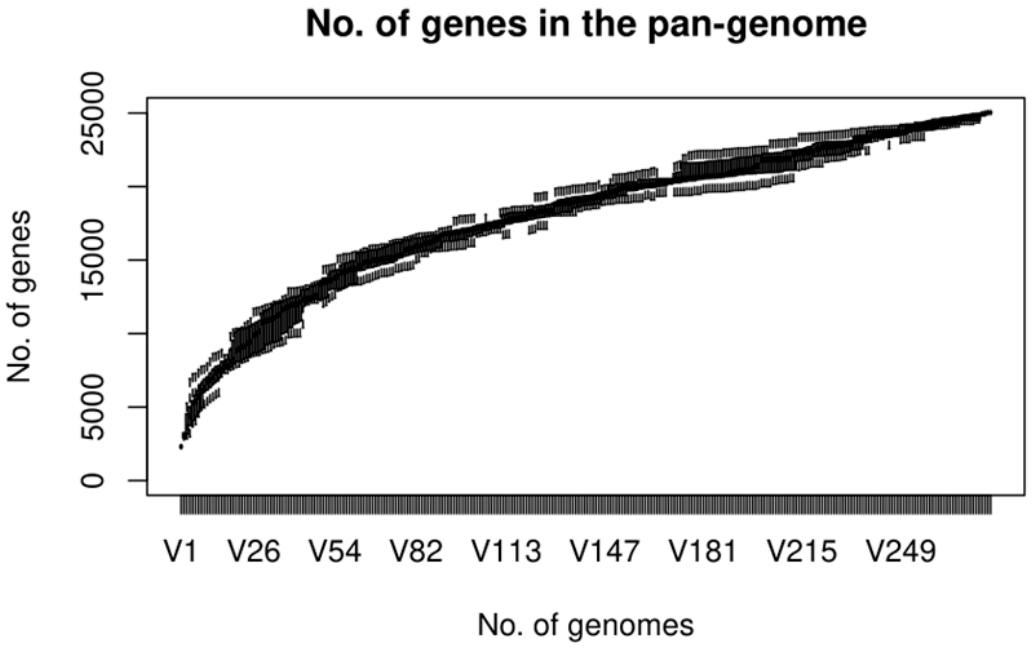

FIGURE 4.3: Cumulative number of genes in the pangenome as a function of added genomes, demonstrating an open pangenome

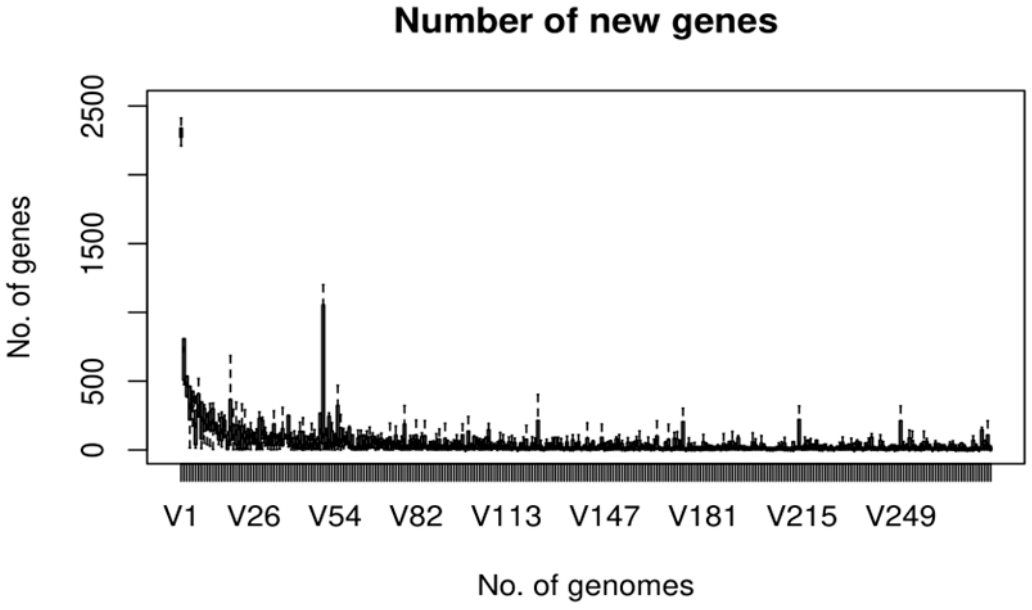

FIGURE 4.4: Number of new genes added with each additional genome, highlighting high initial diversity and ongoing gene acquisition.

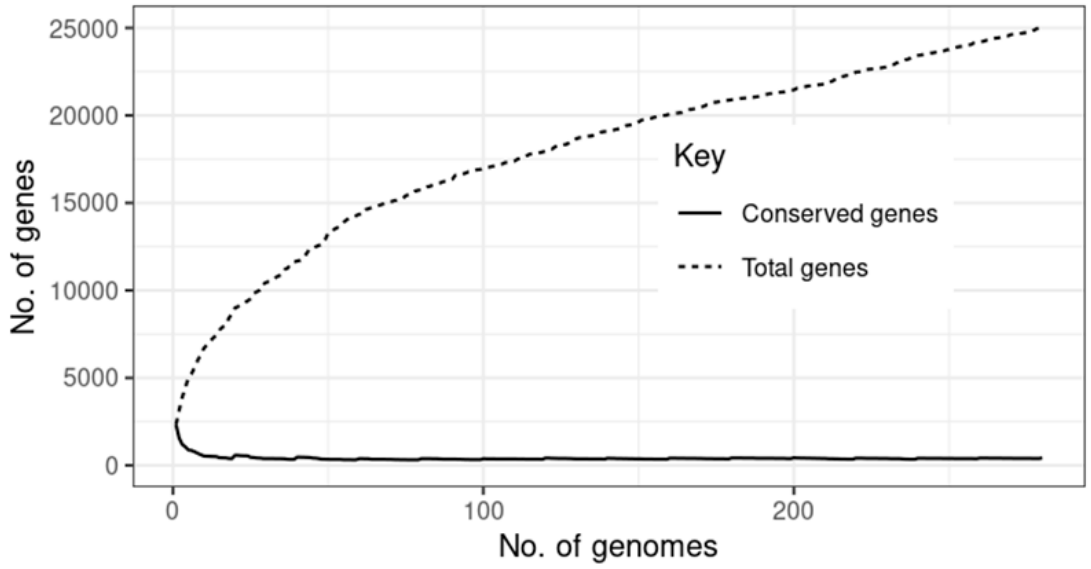

FIGURE 4.5: Trend of total and conserved genes with increasing genomes, showing a stable yet very small core genome.

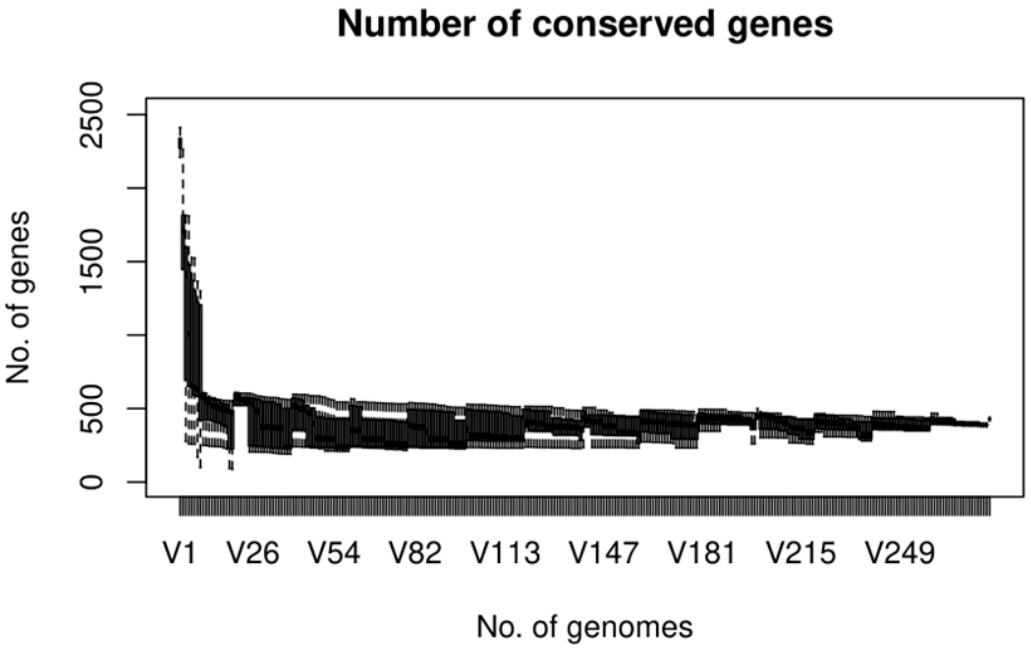

FIGURE 4.6: Detailed view of the number of conserved genes across all genomes

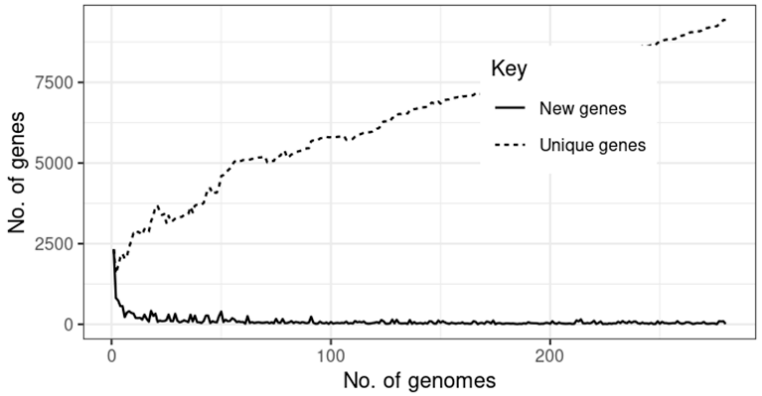

FIGURE 4.7: Trend of new and unique genes as genomes are added.

Furthermore, the analysis of new and unique genes, shown in Figure 4.7, 4.8, demonstrated a persistent influx of unique genes, reinforcing the notion of a highly variable accessory genome. This continuous acquisition of new and unique genes, coupled with the expanding total gene pool, highlighted the adaptability and evolutionary flexibility of this bacterial population.

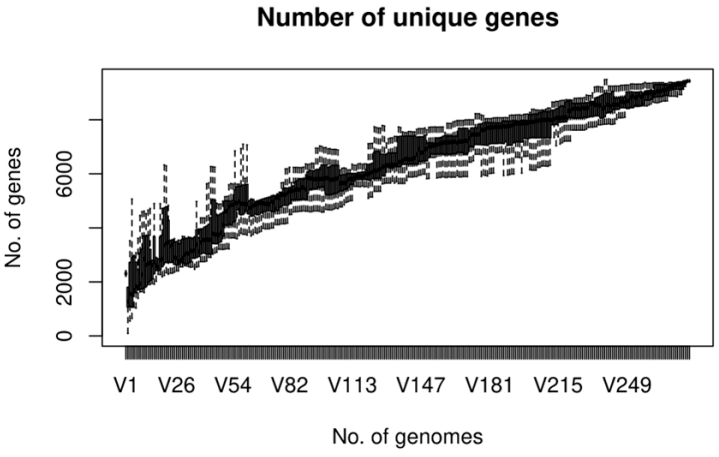

FIGURE 4.8: Number of unique genes added with each genome, indicating a high degree of accessory gene diversity.

The pangenome analysis of the 227 *A. muciniphila* strains revealed a more conserved and less diverse genomic landscape compared to the initial 280-strain dataset. This is a direct consequence of excluding divergent strains, which were likely responsible for a significant portion of the observed genomic variability. The refined dataset highlights a larger core genome and a reduced accessory genome, indicating a more cohesive and closely related group of strains. This suggests that the

227 strains represent a more accurate reflection of the core genomic characteristics of *A. muciniphila*, while the 280 strains captured a broader range of genomic diversity, including potentially distinct subspecies or ecotypes (Figure 4.8).

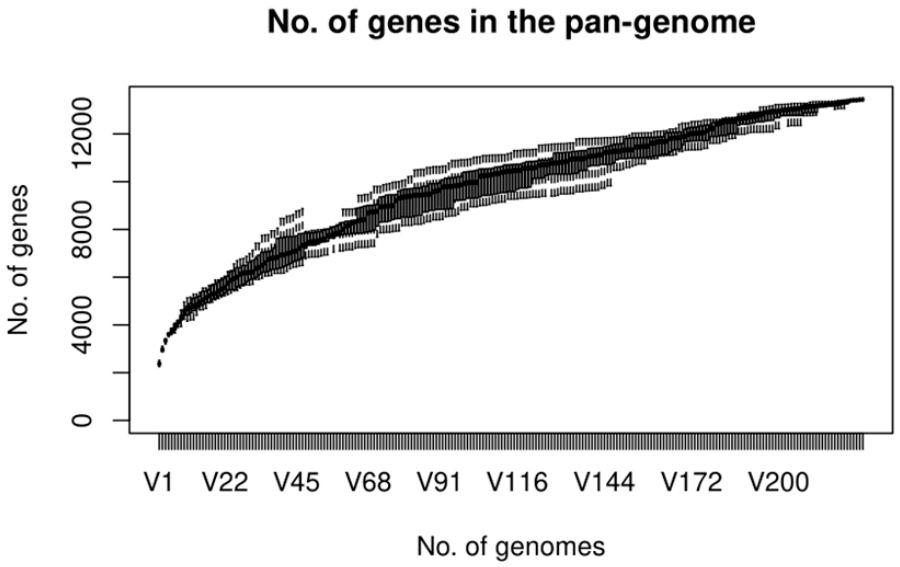

FIGURE 4.9: Cumulative number of genes in the pangenome as a function of added genomes, showing a gradual increase indicative of a less pronounced open pangenome compared to the initial dataset.

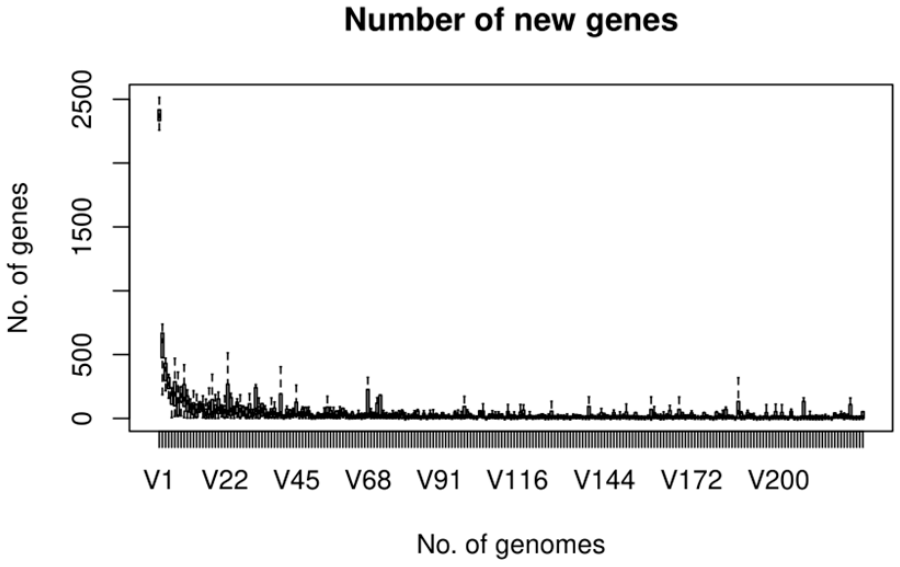

FIGURE 4.10: Number of new genes added with each additional genome, demonstrating a reduced rate of novel gene acquisition.

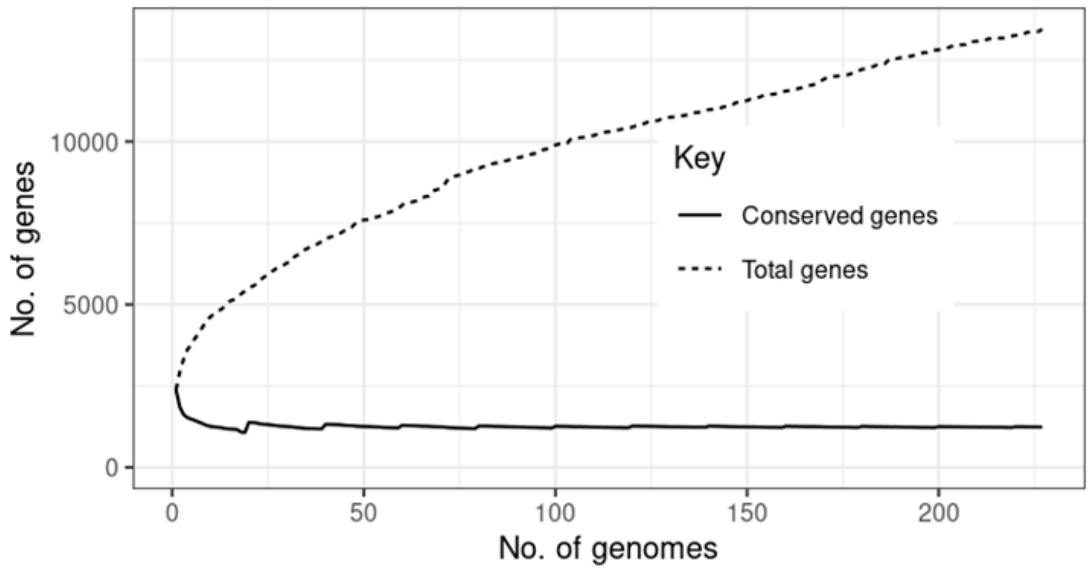

FIGURE 4.11: Trend of total and conserved genes with increasing genomes, revealing a larger core genome and a more stable core gene set.

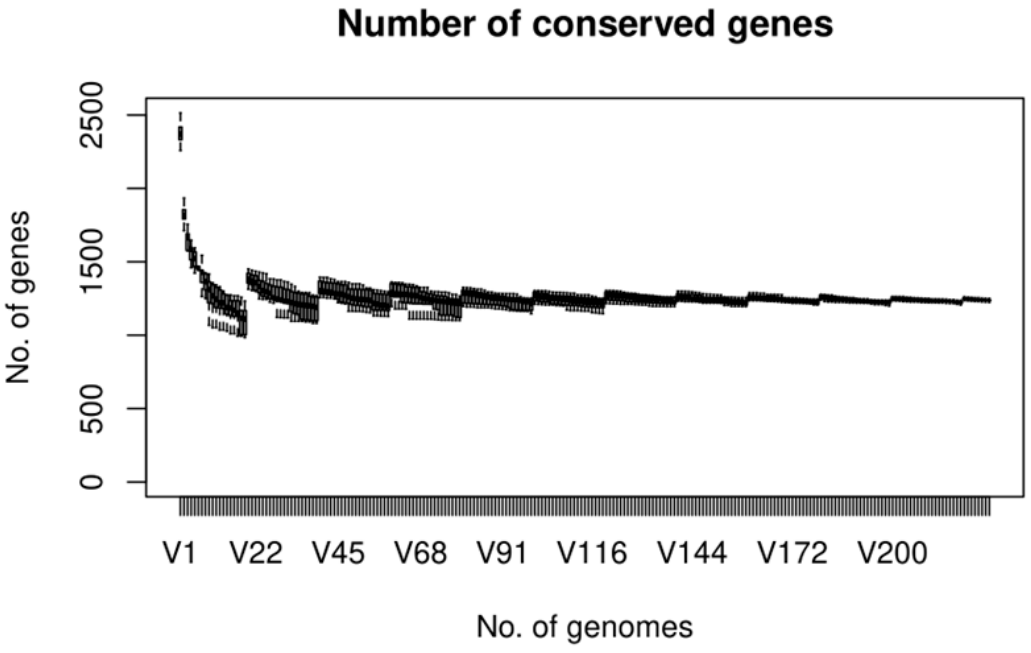

FIGURE 4.12: Detailed view of the number of conserved genes across all genomes, showing minimal variation.

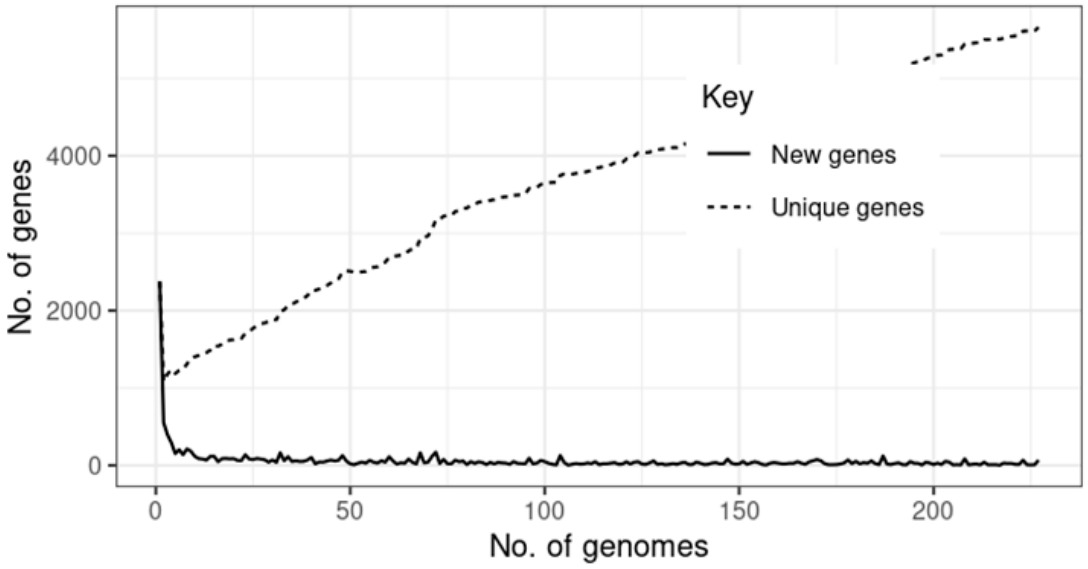

FIGURE 4.13: Trend of new and unique genes as genomes are added, showing a lower rate of new gene accumulation and very few unique genes.

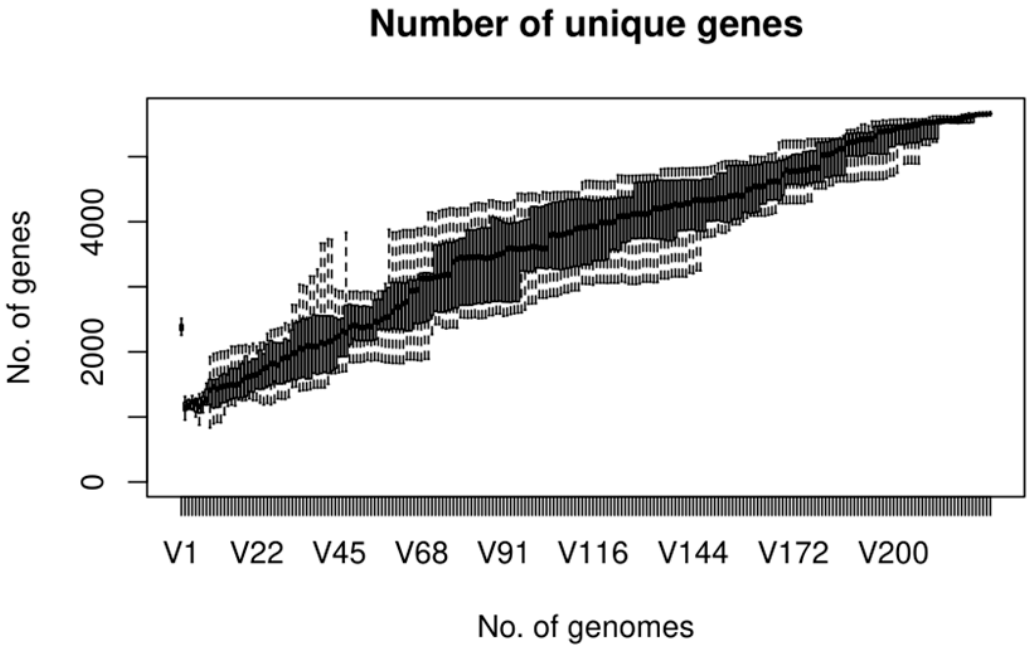

FIGURE 4.14: Number of unique genes added with each genome, highlighting a significant reduction in unique gene content compared to the initial dataset.

Phylogenetic analysis, based on accessory gene presence-absence, revealed a distinct clade comprising 53 strains, visually highlighted in red within the circular phylogram (Figure 4.13 and Figure 4.14). These strains, including CSUN-37

AmIV, Akk2750, and others, formed a well-separated cluster, indicating a significant degree of genetic divergence from the remaining strains. This divergence was supported by substantial branch lengths separating this clade from the rest of the tree.

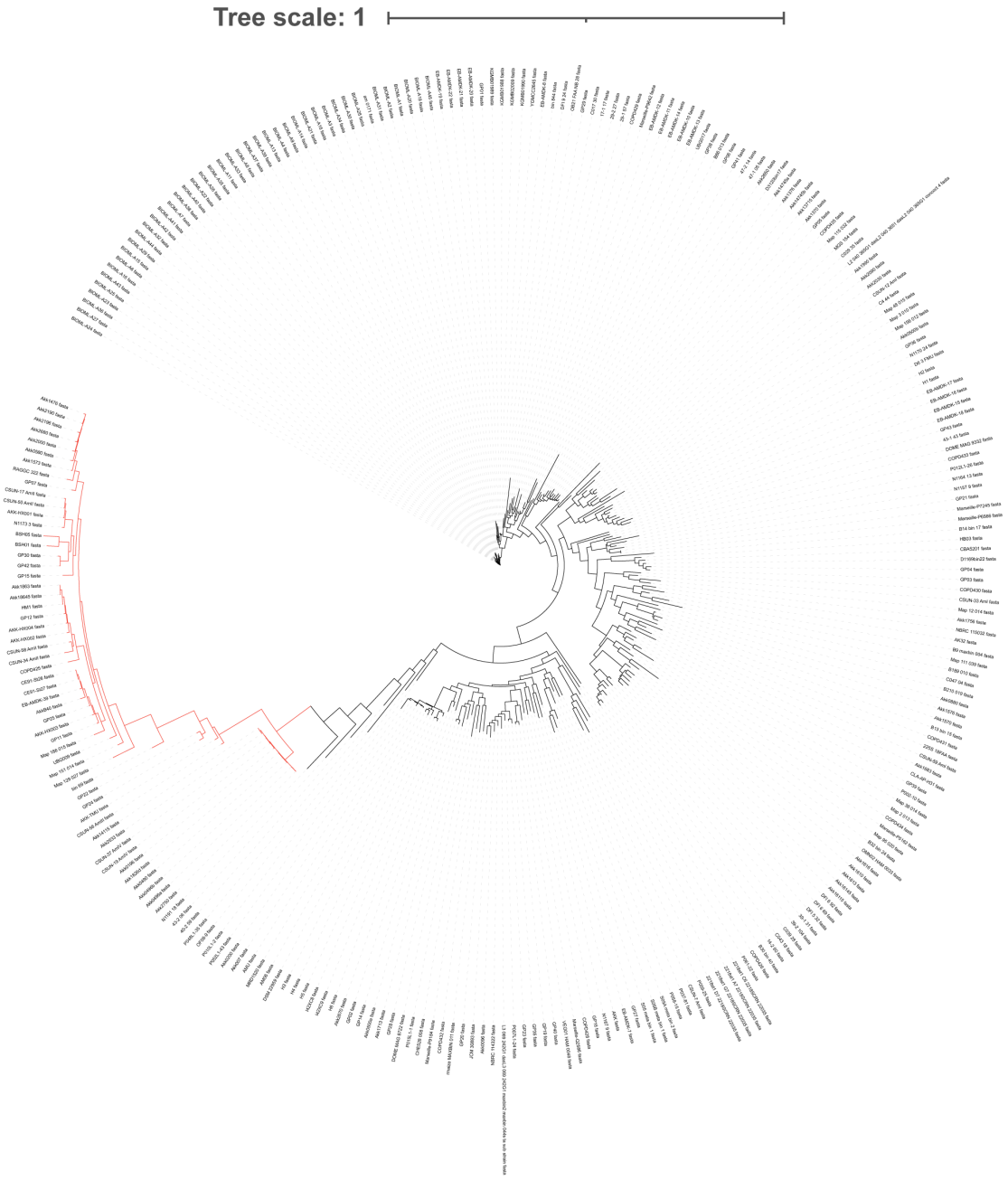

FIGURE 4.15: Phylogenetic analysis of *A. muciniphila* genomes. Circular phylogram depicting the phylogenetic relationships among 280 bacterial genomes, constructed using accessory gene presence/absence data from Roary pangenome analysis and visualized with iTOL. The red-colored branches highlight the 53 divergent strains that formed a distinct clade, indicating significant genetic dissimilarity from the remaining strains

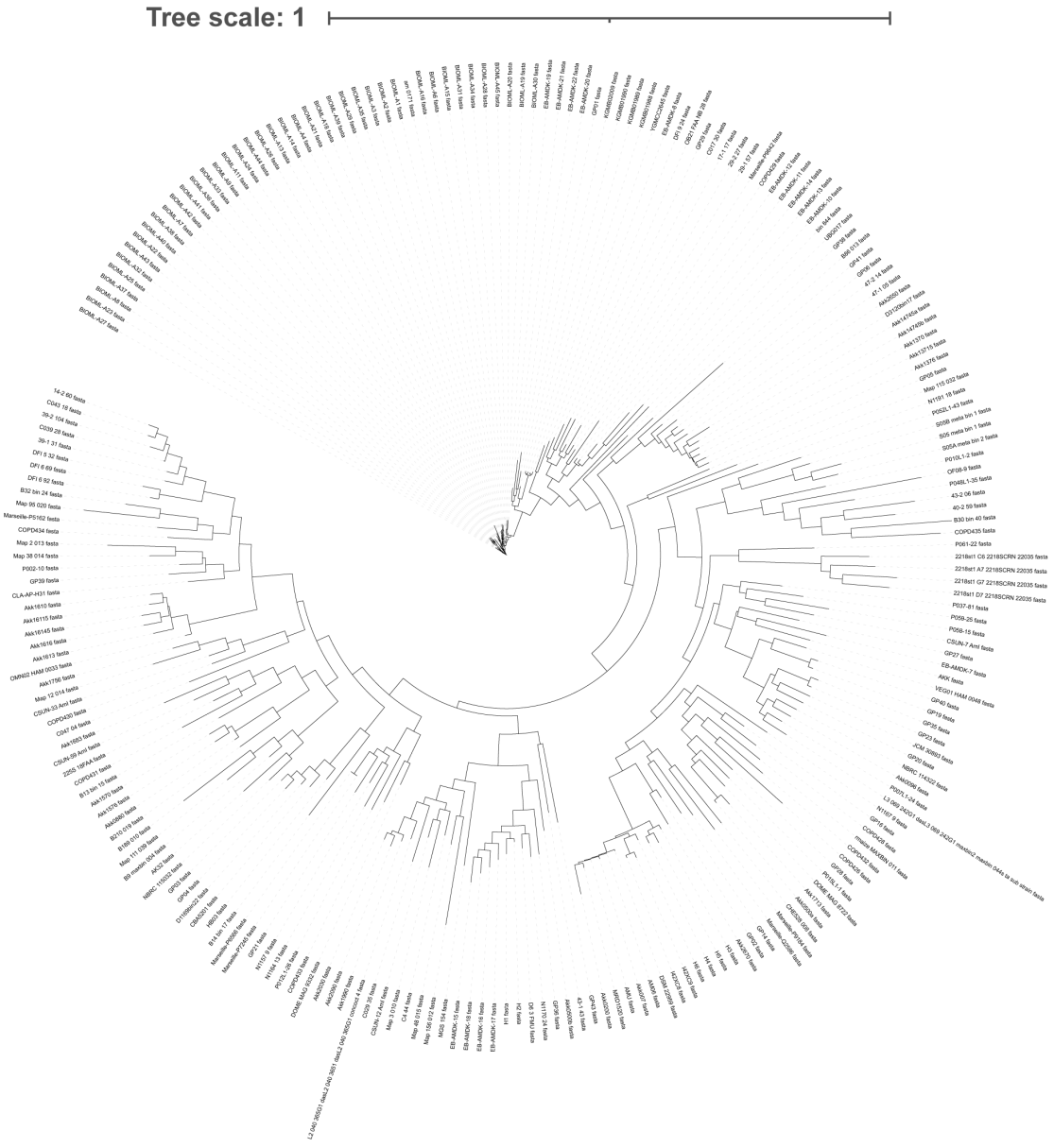

FIGURE 4.16: Phylogenetic analysis of *A. muciniphila* genomes. Circular phylogram constructed using accessory gene presence/absence data from Roary pangenome analysis and visualized with iTOL using the refined dataset of 227 genomes after the exclusion of 53 divergent strains

The formation of this distinct clade, coupled with observed dissimilarities in gene content and ANI values, justified the exclusion of these 53 strains from subsequent analyses. Their removal ensured a more accurate representation of the phylogenetic relationships among the remaining, more closely related strains (Figure 4.13B). This phylogenetic separation suggested the potential for these 53 strains

to represent a distinct species, subspecies, or ecotype, warranting further investigation into their taxonomic classification. The Tree scale of 1 provided a metric for interpreting the extent of evolutionary

4.4 Functional Annotation Reveals Metabolic Robustness

Functional annotation analysis of *A. muciniphila* genes revealed a functional distribution consistent with its role as a gut commensal (Figure 4.15). The largest proportion of genes (2,200) were assigned to "Metabolism and Transport" functions, highlighting the bacterium's capacity for nutrient acquisition and utilization within the gut environment. A significant proportion of genes (1,542) were categorized under "Information, Storage, and Processing," indicating a well-developed system for gene expression and regulation. Cellular processes and signaling, represented by 1,502 genes, further emphasized the bacterium's ability to interact with its environment and respond to stimuli. Notably, a substantial portion of genes (1,317) remained uncharacterized, highlighting the significant knowledge gap regarding the full functional potential of *A. muciniphila*.

The KEGG pathway analysis using KAAS of pangenome of 227 *A. muciniphila* strains revealed a genomically streamlined species exquisitely adapted to thrive in the mucus-rich gut environment, with metabolic capabilities finely tuned for mucin degradation, energy efficiency, and host interaction. Central carbon metabolism is dominated by robust glycolysis (19 genes) and a complete TCA cycle (15 genes), enabling efficient ATP production from mucin-derived sugars like N-acetylglucosamine and galactose. This is complemented by strong pyruvate metabolism (16 genes) and propanoate/butanoate pathways (9-11 genes), which drive the production of short-chain fatty acids (SCFAs) such as propionate and acetate-key metabolites known to enhance gut barrier function and regulate host immunity. The pentose phosphate pathway (8 genes) further underscores metabolic flexibility, balancing NADPH synthesis for redox homeostasis with nucleotide precursor generation.

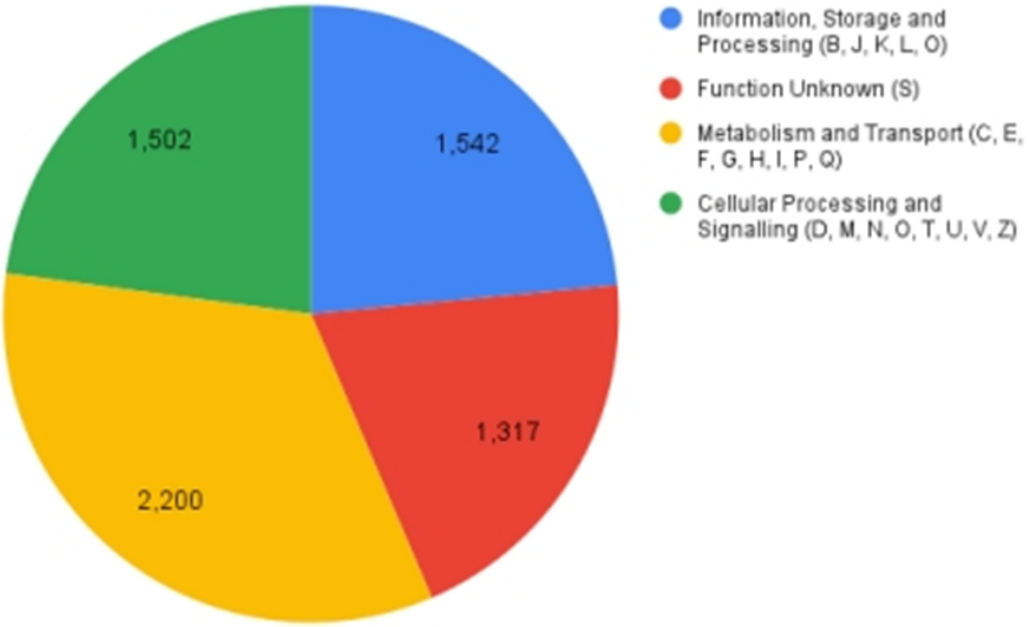

FIGURE 4.17: Functional annotation of *A. muciniphila* genes, highlighting the abundance of genes with unknown functions.

A. muciniphila's specialization in mucin utilization is evident in its arsenal of glycoside hydrolases (e.g., starch/sucrose metabolism: 12 genes) and sulfatases, supported by ABC transporters (29 genes) that facilitate uptake of mucin breakdown products. Pathways for amino sugar metabolism (20 genes) and glycosaminoglycan degradation (4 genes) highlight its evolutionary focus on dismantling complex glycans, while lipopolysaccharide (9 genes) and teichoic acid biosynthesis (4 genes) reflect adaptations to persist within the mucus layer. Strikingly, amino acid metabolism pathways-particularly glycine/serine (18 genes) and cysteine/methionine (18 genes)-suggest nitrogen scavenging from mucin, with glutathione metabolism (1 gene) and arginine biosynthesis (15 genes) further supporting redox balance and immunomodulatory nitric oxide production.

The pangenome also reveals niche-specific adaptations. While terpenoid biosynthesis (10 genes) and ubiquinone pathways (9 genes) likely stabilize membranes under oxidative stress, rare clusters for antibiotic biosynthesis (e.g., streptomycin: 6 genes) and secondary metabolites (e.g., lassopeptides) hint at ecological competition strategies. However, the detection of beta-lactam (14 genes) and vancomycin resistance genes (6 genes) raises safety considerations for probiotic applications,

necessitating stringent strain selection. Vitamin auxotrophy-limited biosynthesis of B6 (3 genes) and biotin (8 genes)-suggests reliance on host/microbiota-derived nutrients, positioning *A. muciniphila* as a metabolic collaborator in the gut ecosystem.

Duplicated pathways (e.g., glycolysis@2x, TCA@2x) indicate genomic redundancy, possibly enhancing metabolic robustness in fluctuating gut conditions. Meanwhile, sulfur metabolism (12 genes) and oxidative phosphorylation (28 genes) underscore adaptations to microaerophilic mucosal niches. The near-absence of xenobiotic degradation pathways aligns with its specialization in host-derived substrates, while folate (13 genes) and cobalamin metabolism (4 genes) suggest roles in host nutrient provisioning.

4.5 Distribution and Diversity of ARGs in *A. muciniphila* Strains

Analysis of the 227 *A. muciniphila* strains revealed the presence of diverse antibiotic resistance genes (ARGs) across 33 distinct genomes. Specifically, resistance determinants were identified for multiple clinically relevant antibiotic classes. Lincosamide resistance genes (lnuC) were detected in strains AKK, Akk2670, GP16, GP20, GP28, Marseille-Q2586, OF08-9, and L3_069_242G1_dasL3_069_242G1 _maxbin2.maxbin.044s_ta_sub_strain.

Resistance to sulfonamides (sul2) and aminoglycosides (APH(3")-Ib, APH(6)-Id) was observed in strains Akk0500b. Akk1756, C047_04, C4_44, GP36, and P002-10.

Tetracycline resistance genes (tetW, tet32, tet(40), tet(W/N/W)) were identified in strains C047_04, GP05, GP14, GP36, GP38, and Map_95_020. Macrolide resistance genes (Mef(En2)) were present in strains GP19, GP27, and GP40. Macrolide - lincosamide - streptogramin B (MLSB) resistance genes (ErmB) were found in strains EB-AMDK-15, EB-AMDK-16, EB-AMDK-17, EB-AMDK-18, KGMB01988, KGMB01989, KGMB01990, KGMB02009, and N1170.24.

The distribution of these ARGs was heterogeneous, with some strains harboring multiple resistance genes and others lacking detectable ARGs. This observed variability suggests differential selective pressures or diverse mechanisms of ARG acquisition among the strains.

Genomic context analysis, including contig coordinates and strand orientation, provided data conducive to investigating the potential for horizontal gene transfer of these ARGs.

Notably, the presence of clinically significant resistance genes, such as ErmB and Mef(En2) (macrolide resistance), and various tetracycline resistance genes, raises concerns regarding the potential dissemination of antibiotic resistance within the gut microbiome.

Consequently, the observed inter-strain variation in ARG presence highlights the necessity for comprehensive monitoring of antibiotic resistance in *A. muciniphila* and detailed investigations into the mechanisms governing ARG acquisition and transfer within this bacterial species.

4.6 Limited Detection of Virulence Factors in *A. muciniphila* Strains

Virulence factor analysis, using the Virulence Factor Database (VFDB), identified virulence-associated genes in only two out of the 227 *A. muciniphila* strains analyzed.

Specifically, Map_156_012 harbored genes encoding EtpA and EtpB, components of the Enteropathogenic Escherichia coli (EPEC) type IV bundle-forming pilus (BFP), known to mediate adhesion.

In contrast, BIOML-A13 exhibited the sat gene. The limited presence of these genes, associated with host interaction, suggests that while most *A. muciniphila* strains analyzed lack known virulence factors, a small subset possesses genes potentially influencing their interaction with host cells.

4.7 Distribution of Plasmid Replicons

Among the 227 *A. muciniphila* strains analyzed, 5 isolates were found to harbor plasmids. Specifically, strains BIOML-A11, BIOML-A13, BIOML-A40, and BIOML-A9 each contained multiple plasmid replicons, including Col(MG828), Col156, Col8282, and ColRNAI. Additionally, strains BIOML-A28 and BIOML-A40 also carried the repUS2 plasmid replicon.

The presence of these plasmids suggests the potential for horizontal gene transfer within these *A. muciniphila* populations, which could contribute to the dissemination of antibiotic resistance genes or other adaptive traits.

The multiple plasmid replicons observed in some isolates indicate the potential for complex plasmid dynamics and interactions within these bacterial genomes.

4.8 Biosynthetic Gene Cluster (BGC) Detection

antiSMASH analysis of the 227 *A. muciniphila* strains revealed a consistent pattern of biosynthetic gene cluster (BGC) distribution, with the majority of strains possessing two terpene BGCs and one arylpolyene BGC.

Terpenes are known for their diverse biological activities, including antimicrobial, antifungal, and signaling functions, potentially contributing to *A. muciniphila*'s interactions within the gut microbiome. Arylpolyenes, often associated with pigments and antioxidants, may play a role in stress response or protection against oxidative damage. Notably, lassopeptide BGCs, associated with the production of ribosomally synthesized and post-translationally modified peptides (RiPPs) with potential antimicrobial or signaling properties, detected in only one strain, BIOML_A9

Similarly, ranthipeptide BGCs, which also produce RiPPs with diverse bioactivities, were identified solely in BIOML_A18. This limited distribution of lassopeptide and ranthipeptide BGCs indicates that these secondary metabolite biosynthetic capabilities are rare within this *A. muciniphila* population.

The predominant presence of terpene and arylpolyene BGCs suggests that these metabolite classes may play significant roles in the ecological or physiological functions of *A. muciniphila*, potentially influencing interactions within the gut microbiome. While the overall BGC profile was largely uniform, the presence of unique BGCs in a few strains highlights potential biosynthetic diversity.

4.9 Discussion and Interpretation of Results

Akkermensia muciniphila from phylum Verrucomicrobia lives in mucus layer of intestine and considered as next generation probiotic against metabolic diseases especially obesity, fatty liver and diabetes [37].

This potential was the motivation to explore the genome of *A. muciniphila* for its safety and genome plasticity. The first step was to explore genomic characteristics, GC content which indicates genome stability and extent of gene expression was explored. GC content ranging between 55-58 indicates that this moderately high content is reason for thermal stability of genome and its adaptability to variable environments [38].

CDS or Coding DNA Sequences is the indication of the genome part which would be actually expressed. Instead of whole genome sequence most of the comparative analysis, phylogenetic analysis and pangenome analysis requires CDS.

When multiple strains of bacterial species are analyzed, a reference strain is selected. Reference strain is the standardized strain which is well characterized and could be used as representative or benchmark for genomic analysis. For *A. muciniphila* MucT or ATCC BAA835 is use as reference strain.

ANI or average Nucleotide Identity is a genomic matric which is used to infer genetic similarity between prokaryotic genomes. When we compare two genomes ANI refers to the average percent identity of homologous genomic regions. The results shown in table 1.1 show ANI between 97-99% that indicates that despite variation, compared genomes should be considered between same species (to be considered same species ANI should between 95-99%) [38].

In this case ANI values indicate genomic divergence between different strains of the bacteria.

As the ANI indicated divergence, Pangenome analysis was performed to understand the full genetic diversity within all reported strains of *A. muciniphila*.

Pangenome provides a comprehensive picture of genetic capabilities, adaptations and evolutionary flexibilities. It gives information about what is shared by all strains and what is different or unique in each strain shedding light on the process of evolution and adaptation. Pangenome analysis also helps to identify genomic clusters and phylogroups.

Major components of pangenome analysis include Core genome which signifies the genes found in all strains indicating that these genes are essential functions, in case *A. muciniphila* mucin degradation genes are part of core genes and present in all species. Soft core genes are the genes which are not essentially present in all but are present in most of the strains. The criteria is that the gen is considered core if it is preset in 100% genomes, soft core if present in 95-99% genomes and if less than 95 then it is considered part of accessory genome [39].

Accessory genome consists of genes found in some species but not all strains and these genes indicate the adaptation due to diverse environment. Unique genes are found in only one strain while cloud genes indicate the genes which are rare or highly variable and indicate genomic instability or plasticity. The shell genome comprises 15-95% genomes and indicate adaptive or niche specific functions.

The pangenome analysis in this study revealed that each strain indicates variations owing to host interactions and health benefits they can impart. *A. muciniphila* shows open pangenome which refers that gene pool kept on expanding which signifies that there is high genetic diversity in this specie which is most oftenly due to genetic variability caused by Horizontal gene transfer, diverse ecological niche and frequent recombination.

Due to variable gut environment the bacteria face pressure for adaptation to different hosts, diets and metabolic versatility. It also indicates that the bacteria has high ecological flexibility [40].

In general it could be concluded that *A. muciniphila* is genetically dynamic, ecologically versatile and continuously evolving.

The diversity among different strains required to understand that how these strains are related. Phylogenetic analysis provides insight into evolutionary relationship based on genetic similarity among strains [41].

In case of *A. muciniphila* it was found that 53 strains indicated significant divergence that they could be considered as separate clade. This analysis indicated that how genes evolved in various strains resulting in functional divergence.

To understand functional divergence amongst various strains and to understand the probiotic potential of each strain, Antibiotic resistance genes were explored.

These genes enable microorganism to survive in the presence of a particular antibiotic. *A. muciniphila* strains indicate the presence of genes against Lincosamide, a commonly used antibiotic against gram positive anaerobes. It is a bacteriostatic antibiotic which inhibit protein synthesis and are used in variety of infection in humans including dental infections, osteomyelitis, pneumonia and colitis.

Some species indicated the presence of resistant genes against sulfonamides which inhibit folic acid synthesis and are used against various Gram positive as well Gram-negative bacteria involved in urinary tract infections, respiratory infections and eye infections.

Tetracycline resistance is one of the common capability found in pathogenic bacteria, probiotics and even in environmental strains. Tetracycline is extensively used bacteriostatic antibiotic used to inhibit bacterial protein synthesis. Macrolides are protein synthesis inhibiting bacteriostatic antimicrobial agents used against variety of Gram negative and Gram positive bacteria as well as atypical pathogens such as Mycoplasma [42].

This observed variation in strains of *A. muciniphila* indicates that differential selective pressures due to variations in host gut environment and exposure to various antimicrobial compounds have resulted in diverse mechanisms for acquisition of antimicrobial resistance genes.

With reference to acquisition of virulence factors by *A. muciniphila*, the bacterial strains were found to lack known virulence factors but few adherence molecules were found which could be due to their mechanism of interaction with host. Virulence factors are the capabilities acquired by bacterial species enabling them to cause disease, these include capabilities to invade host, evade immune system, case cell damage or obtain nutrients from host [43]. *A. muciniphila* express capabilities of adherence to host and enzymes for mucus degradation, which are not classified as typical virulence factors but are essential for colonization

Owing to genome plasticity and open pangenome *A. muciniphila* has four distinct phylogroups which also show functional adaptations [44]. Major functional variations are in accessory genome, while core genome is majorly mucin degrading genes. Accessory genome or variable genes include majorly carbohydrate active enzymes, especially GT4 glycosyltranferases. Surface structures and host interactions i.e. variations in exopolysaccharides contribute in its activities related to immunoregulation and host recognition. In addition its oxygen sensitivity and its ability to outgrow it competitors is also encoded as accessory genome.

Chapter 5

Conclusion & Future Recommendations

This study reveals *Akkermansia muciniphila* as a genetically diverse gut symbiont with a dynamic pangenome shaped by niche-specific adaptations. While its core genome supports mucin degradation and host interactions, strain-level variations in ARGs, plasmids, and secondary metabolites necessitate precision in probiotic development. The discovery of divergent clades suggests unresolved taxonomic complexity, urging re-evaluation of species boundaries. Future work should prioritize functional characterization of unannotated genes, particularly those linked to immunomodulation and gut barrier enhancement, to clarify mechanistic roles.

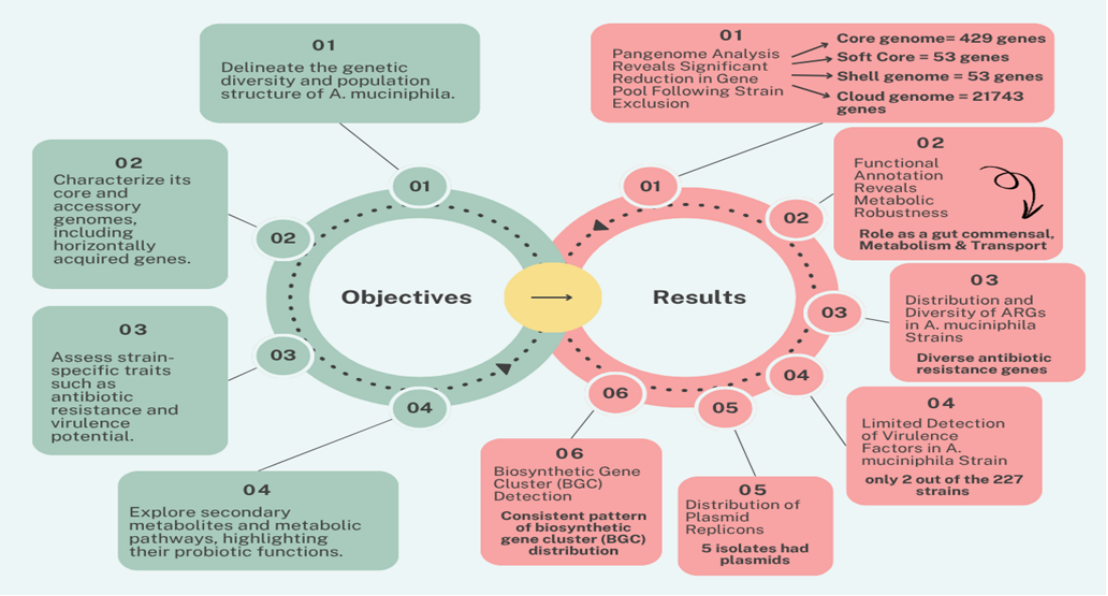

FIGURE 5.1: Summary of main objectives of the project and results achieved

Despite the concerns, the safe strains of *A. muciniphila* are already used in various probiotics such as Pendulum Akkermansia probiotic (Live), Nature Akk11 (live), Akkermensia Co Gut support (pasteurized), Vitamatic (Live) and Codeage synbiotic (live). The feedback of these probiotics included the gut discomfort or no colonization, very high expenses due to high oxygen sensitivity of bacteria which requires oxygen free facilities. Although *A. muciniphila* is considered next generation probiotic, concerns about its mucin degradation are discussed. This capability in one hand is beneficial as thin mucus is better but degradation can cause leaky gut or inflammation in few cases. Although *A. muciniphila* is not known as pathogen but presence of some putative virulence associated genes suggest that it could be opportunistic pathogen. Similar concerns are raised against its immune interactions and stimulation of TLR2 genes. Live strains of *A. muciniphila* are oxygen sensitive and feedback of pasteurized strains is that they do not colonize. Although *A. muciniphila* posses a high potential as probiotic for metabolic health, but concerns related to development, clinical trails and population specific risk assessments cannot be undermined.

In future Clinical trials are critical to validate strain-specific benefits in metabolic and inflammatory disorders, while ecological studies must unravel its interplay with host immunity and microbiota. Addressing ARG transmission risks through horizontal gene transfer and elucidating evolutionary drivers of genomic diversity will further ensure safe therapeutic deployment. By bridging genomics with translational medicine, *A. muciniphila* could emerge as a tailored probiotic, offering novel strategies to combat dysbiosis-linked diseases. Future efforts must focus on functional validation of uncharacterized genes, clinical validation of strain-specific benefits, and ecological studies to unravel its interactions within the gut microbiome, ultimately unlocking its full therapeutic promise.

Bibliography

[1] E. Aja, A. Zeng, W. Gray, K. Connelley, A. Chaganti, and J. P. Jacobs, "Health effects and therapeutic potential of the gut microbe Akkermansia muciniphila," *Nutrients*, vol. 17, no. 3, p. 562, Jan 2025.

[2] C. Depommier, "Supplementation with Akkermansia muciniphila in overweight and obese human volunteers: a proof-of-concept exploratory study," *Nature Medicine*, vol. 25, no. 7, pp. 1096–1103, Jul 2019.

[3] Y. Li, "Exploring impact of gut microbiome on human health and disease: A review," *Theor. Nat. Sci.*, vol. 74, pp. 163–169, 2025.

[4] P. Abondio, E. Cilli, and D. Luiselli, "Human pangenomics: Promises and challenges of a distributed genomic reference," *Reproductive and Developmental Biology*, vol. 13, no. 6, p. 1360, Jun 2023.

[5] Z. Zhang, "The worldview of Akkermansia muciniphila, a bibliometric analysis," *Front. Microbiol.*, vol. 16, 2025.

[6] W. Lu, "Whole-genome sequencing and genomic analysis of four Akkermansia strains newly isolated from human feces," *Front. Microbiol.*, vol. 15, 2024.

[7] Q.-B. Lv, "A thousand metagenome-assembled genomes of Akkermansia reveal phylogroups and geographical and functional variations in the human gut," *Front. Cell. Infect. Microbiol.*, vol. 12, 2022.

[8] J. Huerta-Cepas, "Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper," *Mol. Biol. Evol.*, vol. 34, no. 8, pp. 2115–2122, Apr 2017.

[9] Y. Kim, C. Gu, H. U. Kim, S. Y. Lee, M. E. Vanoni, and P. Palumbo, "Current status of pan-genome analysis for pathogenic bacteria," *Current Opinion in Biotechnology*, 2019.

[10] C. A. Matthews, N. S. Watson-Haigh, R. A. Burton, and A. E. Sheppard, "A gentle introduction to pangenomics," *Briefings in Bioinformatics*, vol. 25, no. 6, Sep 2024.

[11] F. Mazzarotto, "Pangenomes aid accurate detection of large insertion and deletions from gene panel data: the case of cardiomyopathies," *bioRxiv*, Dec 2024.

[12] A. J. Page, "Roary: Rapid large-scale prokaryote pan genome analysis," *Bioinformatics*, vol. 31, no. 22, pp. 3691–3693, 2015.

[13] A. A. Golicz, P. E. Bayer, P. L. Bhalla, and J. Batley, "Pangenomics in plant genomics research: from genes to genomes," *The Plant Genome*, vol. 13, no. 3, p. e20008, 2020.

[14] H. Tettelin, "Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial 'pan-genome'," *Proceedings of the National Academy of Sciences*, vol. 102, no. 39, pp. 13 950–13 955, 2005.

[15] C. M. Kobras, A. K. Fenton, and S. K. Sheppard, "Next generation microbiology: from comparative genomics to gene function," *Genome Biology*, vol. 22, p. 123, Apr 2021.

[16] J. Zanette, J. Pan, and G. Ferreira, "Distinct evolutionary patterns in open and closed pangenomes of aquatic and terrestrial bacterial lineages," *Microbiome*, vol. 11, no. 99, 2023.

[17] R. E. Anderson and J. L. Seifert, "Horizontal gene transfer and the evolution of microbial pangenomes," *Nature Reviews Microbiology*, vol. 19, no. 6, pp. 325–336, 2021.

[18] D. P. Melnyk, "Comparative pangenomics: analysis of 12 microbial pathogen pangenomes reveals conserved global structures of genetic and functional diversity," *BMC Genomics*, vol. 23, p. 7, Jan 2022.

[19] L. Li, "Function and therapeutic prospects of next-generation probiotic Akkermansia muciniphila in infectious diseases," *Front. Microbiol.*, vol. 15, Feb 2024.

[20] A. G. McArthur, "The comprehensive antibiotic resistance database," *Antimicrob. Agents Chemother.*, vol. 57, no. 7, pp. 3348–3357, 2013.

[21] M. Khalil, "Unraveling the role of the human gut microbiome in health and diseases," *Microorganisms*, vol. 12, no. 11, p. 2333, Nov 2024.

[22] Y. Liu, "TLR4 regulates ROR t regulatory t-cell responses and susceptibility to colon inflammation through interaction with Akkermansia muciniphila," *Microbiome*, vol. 10, no. 1, p. 98, 2022.

[23] V. D. Appanna, "Dysbiosis, probiotics, and prebiotics: in diseases and health," in *Springer eBooks*, 2018, pp. 81–122.

[24] P. Derrien, E. E. Vaughan, C. Plugge, and W. M. de Vos, "Akkermansia muciniphila gen. nov., sp. nov., a human intestinal mucin-degrading bacterium," *Int. J. Syst. Evol. Microbiol.*, vol. 54, no. 5, pp. 1469–1476, Sep 2004.

[25] N. Ottman, "Pili-like proteins of Akkermansia muciniphila modulate host immune responses and gut barrier function," *PLoS ONE*, vol. 12, no. 3, p. e0173004, 2017.

[26] R. Iwaza, "Akkermansia muciniphila: The state of the art, 18 years after its first discovery," *Front. Gastroenterol.*, vol. 1, Oct 2022.

[27] J. J. Gillespie, "PATRIC: the comprehensive bacterial bioinformatics resource with a focus on human pathogenic species," *Infect. Immun.*, vol. 79, no. 11, pp. 4286–4298, Nov 2011.

[28] D. Hyatt, "Prodigal: prokaryotic gene recognition and translation initiation site identification," *BMC Bioinformatics*, vol. 11, no. 1, Mar 2010.

[29] M. H. Medema, "antiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal

genome sequences," *Nucleic Acids Res.*, vol. 39, no. suppl. 2, pp. W339–W346, 2011.

[30] Q. Zhang, "Genetic mapping of microbial and host traits reveals production of immunomodulatory lipids by Akkermansia muciniphila in the murine gut," *Nat. Microbiol.*, vol. 8, no. 3, pp. 424–440, 2023.

[31] A. Ioannou, M. D. Berkhout, S. Y. Geerlings, and C. Belzer, "Akkermansia muciniphila: biology, microbial ecology, host interactions and therapeutic potential," *Nat. Rev. Microbiol.*, Oct 2024.

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

[33] V. Bortolaia, "Resfinder 4.0 for predictions of phenotypes from genotypes," *J. Antimicrob. Chemother.*, vol. 75, no. 12, pp. 3491–3500, Jul 2020.

[34] A. Carattoli, "In silico detection and typing of plasmids using plasmidfinder and plasmid multilocus sequence typing," *Antimicrob. Agents Chemother.*, vol. 58, no. 7, pp. 3895–3903, Apr 2014.

[35] L. Chen, "VFDB: a reference database for bacterial virulence factors," *Nucleic Acids Res.*, vol. 33, no. Database issue, pp. D325–D328, Dec 2004.

[36] C. Mo, "The influence of Akkermansia muciniphila on intestinal barrier function," *Gut Pathog.*, vol. 16, no. 1, p. 41, 2024.

[37] M. Kechagia, D. Basoulis, S. Konstantopoulou, D. Dimitriadi, K. Gyftopoulou, N. Skarmoutsou, and E. Fakiri, "Health benefits of probiotics: a review," *ISRN Nutrition*, vol. 2013, p. 481651, 2013.

[38] A. Rodríguez-Gijón, M. Buck, and A. Andersson, "Linking prokaryotic genome size variation to metabolic potential and environment," *ISME Communications*, vol. 3, p. 25, 2023.

[39] H. Tettelin, D. Riley, C. Cattuto, and D. Medini, "Comparative genomics: the bacterial pan-genome," *Current Opinion in Microbiology*, vol. 11, pp. 472–477, 2008.

[40] D. Medini, C. Donati, H. Tettelin, V. Masignani, and R. Rappuoli, "The microbial pan-genome," *Current Opinion in Genetics & Development*, vol. 15, pp. 589–594, 2005.

[41] K. Georgiades and D. Raoult, "Defining pathogenic bacterial species in the genomic era," *Frontiers in Microbiology*, vol. 1, p. 151, 2010.

[42] J. Olivares, A. Bernardini, G. Garcia-Leon, F. Corona, B. Sanchez, and J. Martinez, "The intrinsic resistome of bacterial pathogens," *Frontiers in Microbiology*, vol. 4, p. 103, 2013.

[43] V. Merhej, K. Georgiades, and D. Raoult, "Postgenomic analysis of bacterial pathogens repertoire reveals genome reduction rather than virulence factors," *Briefings in Functional Genomics*, vol. 12, pp. 291–304, 2013.

[44] M. Farré and A. Ruiz-Herrera, "The plasticity of genome architecture," *Genes*, vol. 11, no. 12, p. 1413, 2020.