CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

Metagenomic Profiling of Diversity and Community Structure in the Termite Gut Microbiome

by

Bisma Anam

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

2026

CERTIFICATE OF APPROVAL

Metagenomic Profiling of Diversity and Community Structure in the Termite Gut Microbiome

by

Bisma Anam

(MBS243008)

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Zoya Khalid	CUI, Islamabad
(b)	Internal Examiner	Dr. Sami Ullah Jan	CUST, Islamabad

Dr. Sahar Fazal
Thesis Supervisor
July, 2026

Dr. Syeda Marriam Bakhtiar
Head
Dept. of Bioinfo. and Biosciences
July, 2026

Dr. Sahar Fazal
Dean
Faculty of Health and Life Sciences
July, 2026

Author's Declaration

I, **Bisma Anam** hereby state that my MS thesis titled "**Metagenomic Profiling of Diversity and Community Structure in the Termite Gut Microbiome**" 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.

(Bisma Anam)

Registration No: MBS243008

Plagiarism Undertaking

I solemnly declare that research work presented in this thesis titled "**Metagenomic Profiling of Diversity and Community Structure in the Termite Gut Microbiome**" 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.

(Bisma Anam)
Registration No: MBS243008

Acknowledgement

All praise and gratitude are due to Almighty Allah, the Most Beneficent and the Most Merciful, whose blessings and guidance enabled me to complete this research work successfully.

I would like to express my sincere gratitude to my respected supervisor, Dr. Sahar Fazal for their continuous guidance, valuable suggestions, and unwavering support throughout the course of this study. Their encouragement and insightful feedback greatly contributed to the completion of this thesis.

I am also thankful to all the faculty members especially Dr. Marriam Bakhtiar, Head of Department, Department of Biosciences and Bioinformatics, CUST for providing me with knowledge, resources, and a conducive learning environment.

I extend my appreciation to my friends who supported me during this journey, whether through academic discussions or moral encouragement.

My deepest thanks go to my beloved parents and family for their endless love, prayers, and sacrifices. Their constant support and motivation have been the foundation of my success.

Finally, I acknowledge all those who directly or indirectly contributed to the completion of this research work.

(Bisma Anam)

Abstract

The termite gut harbours one of the most phylogenetically rich and functionally specialised microbial ecosystems on Earth, enabling wood-feeding termites to thrive on an otherwise recalcitrant, nitrogen-deficient lignocellulosic diet. The present study characterised the bacterial community structure and metabolic potential of the termite gut microbiome through a comprehensive metagenomic workflow encompassing DNA quality assessment, sequencing quality control, de novo assembly, gene prediction, taxonomic profiling, and multi-level functional annotation. High-molecular-weight genomic DNA was extracted and verified by fluorometric quantification (3.2 ng/µL; 50 ng total) and agarose gel electrophoresis, confirming template integrity suitable for downstream library preparation. Illumina sequencing generated 136,742 high-quality reads spanning 37.5 Mbp, with mean Phred scores of Q35.4-Q41.9 across all read positions. De novo assembly using MEGAHIT yielded 10 contigs (total size 3,881 bp; N50 = 406 bp), with extreme coverage heterogeneity (1-777×) reflecting the dominance of a small number of highly abundant taxa within a diverse, low-coverage community. Gene prediction by FragGeneScan identified 127,623 genes across 116,517 sequences (85.21% prediction rate), with all predicted proteins ranging from 21 to 93 amino acids in length, consistent with the short-read constraint of the dataset. Taxonomic profiling through OTU clustering at 97% similarity and multi-level barplot analysis revealed a community dominated by the phylum Bacillota (∼45%), followed by Pseudomonadota (∼15%) and Spirochaetota (∼13%). At the genus level, *Treponema* (12%) and *Bacillus* (9%) were prominent, alongside specialised termite gut symbionts including *Fibrobacter intestinalis, Endomicrobium proavitum, Treponema isoptericolens, Treponema primitia*, and *Fretibacterium fastidiosum.* Alpha diversity analysis revealed 578 observed species, a Shannon diversity index of 4.55, a Simpson's index of 0.037, and complete Good's coverage (1.000), confirming high community richness, evenness, and adequate sequencing depth. Phylogenetic analysis identified 19 major bacterial phyla spanning all major functional guilds of the termite hindgut. Functional annotation using eggNOG-Mapper v2 annotated 54.3% of 77,600 query sequences, with COG category J (Translation,

Ribosomal Structure and Biogenesis) dominating assigned functions (56.1%) and 43.9% assigned to Category S (Function Unknown), reflecting the prevalence of novel, uncharacterised lineages. KEGG pathway mapping linked all annotated sequences to the Ribosome pathway (ko03010) via ribosomal protein S5 (K02950). CAZyme analysis identified four families from 56 annotated proteins, with GH113 (beta-mannanases and beta-mannosidases) overwhelmingly dominant at 87.5%, confirming active hemicellulose degradation targeting beta-1,4-mannose linkages in wood-derived mannans. Glycosyltransferase families GT2, GT106, and GT3 indicated biosynthetic activities, including exopolysaccharide, surface glycan, and glycogen synthesis. MetaCyc functional pathway profiling via PICRUSt2 predicted 380 individual pathways across 8 Level-1 superclasses and 39 Level-2 categories. Biosynthesis pathways dominated at 51.3% of total predicted abundance, followed by Superpathways (27.6%) and Generation of Precursor Metabolites and Energy (9.5%). At Level 2, Nucleoside and Nucleotide Biosynthesis (12.8%) and Amino Acid Biosynthesis (12.3%) collectively accounted for over 25% of all predicted functional capacity. These findings establish a comprehensive metagenomic baseline for the termite gut microbiome, demonstrating a phylogenetically diverse, anabolically dominant, and functionally integrated community with significant potential for lignocellulose bioconversion, nitrogen cycling, and nutritional mutualism with the termite host.

Keywords: Termite Gut Microbiome, Metagenomics, Taxonomic Profiling, Functional Annotation, CAZymes, Lignocellulose Degradation, Microbial Diversity.

Contents

List of Figures

List of Tables

Abbreviations

AAs	Auxiliary Activities
CBMs	Carbohydrate-Binding Modules
CEs	Carbohydrate Esterases
eggNOG	evolutionary genealogy of genes:
	Non-supervised Orthologous Groups
FASTQ	FASTA with Quality Scores
	(sequence and quality file format)
FastQC	Fast Quality Control
GC	Guanine-Cytosine (content)
GHs	Glycoside Hydrolases
GTs	Glycosyltransferases
IIPSP3	Integrated Microbial Genomes &
	Microbiomes Pipeline, version 3 (IMG/M Pipeline v3)
KEGG	Kyoto Encyclopedia of Genes and Genomes
MEGAHIT	MEtaGenome Assembler based on de Bruijn graph
	using HIT (Highly-efficient Iterative de Bruijn graph assembler)
nifH	Nitrogen Fixation Protein H Gene
	(encodes the dinitrogenase reductase subunit of nitrogenase)
OTU	Operational Taxonomic Units
PLs	Polysaccharide Lyases
QUAST	Quality Assessment Tool for Genome Assemblies
rRNA	Ribosomal Ribonucleic Acid

Chapter 1

Introduction

Various wood-consuming invertebrates, including termites, beetles, and wood borers, harbor diverse microbial communities, including bacteria, archaea, and protists. Recent studies have identified symbiotic bacteria within this system that exhibit significant lignocellulose-degrading capabilities. These microbes contribute to the partial breakdown of plant materials, enhancing nutrient availability and improving in vitro digestibility, making them promising candidates for the biological pretreatment of lignocellulosic biomass in animal feed production [1]. Structurally, lignocellulose constitutes a heterogeneous macromolecular matrix in which cellulose microfibrils, hemicellulose, and lignin are interlaced with trace accessory components [1]. The distinct chemical complexity of each constituent polymer collectively imparts a high degree of structural recalcitrance to the composite material. Natural biodegradation of this resilient substrate necessitates a synergistic mechanism involving localized chemical alterations alongside a comprehensive enzymatic repertoire, specifically comprising cellulases, hemicellulases, and oxidative lignin-modifying enzymes. Translating this deconstruction process to an industrial bioconversion scale demands the destabilization of this biomass matrix while maintaining strict economic constraints. At present, navigating the trade-off between efficient matrix disruption and processing cost-effectiveness remains a definitive techno-economic bottleneck, hindering the commercial viability of current large-scale biorefineries [2].

The degradation of lignocellulosic biomass is a fundamental ecological process that sustains nutrient cycling and energy flow within terrestrial ecosystems. Among the various organisms involved in this process, termites stand out as one of the most efficient natural decomposers of plant material. Their remarkable ability to digest lignocellulose that is a complex matrix composed of cellulose, hemicellulose, and lignin which has positioned them as key contributors to global carbon turnover [3]. This ability is not solely attributed to the termites themselves but largely depends on a highly specialized and symbiotic gut microbiome that performs the majority of the biochemical transformations required for biomass degradation.

Wood-feeding (xylophagous) and plant-feeding (phytophagous) termites are well-known examples of herbivorous insects. Classified under the order Isoptera, they have evolved a highly specialized gut environment that enables the efficient breakdown of lignocellulosic plant material, particularly within tropical and subtropical habitats [4].

The termite gut represents a unique and complex microenvironment that harbors a dense and diverse consortium of microorganisms which includes bacteria, archaea, and eukaryotic protists. These microorganisms form a highly structured and cooperative community that enables the efficient breakdown of recalcitrant plant polymers into simpler compounds such as sugars, short-chain fatty acids, and gases [5]. The microbiota of gut and the termite shows the effective examples of mutualism in nature, where both partners benefit from metabolic interdependence and nutrient exchange. The association between termites and their gut microorganisms represents a highly specialized form of insect–microbe symbiosis, characterized by mutualistic interactions that support host nutrition and microbial survival. Termites also have an endogenous enzyme production system; however, it is generally considered less effective because of its lower enzymatic activity and comparatively smaller contribution to biomass hydrolysis than that of gut microsymbionts [6].

Wood-feeding termites are among the most efficient organisms in nature for the degradation of lignocellulosic biomass. Their gut environment, which supports microorganisms adapted to highly specialized and often anaerobic conditions, serves

as a rich source of diverse bacteria with potential applications in bioethanol production [7]. Termites play a fundamental role in the decomposition of lignocellulose and thereby contribute significantly to carbon cycling in tropical and subtropical ecosystems. In addition, facultative anaerobic microbes help maintain low-oxygen conditions by consuming residual oxygen, thereby supporting optimal anaerobic metabolism within the gut system [8].

The termite gut represents a highly specialized microhabitat that supports complex and diverse microbial consortia essential for wood digestion and nutrient recycling. Several cellulolytic bacterial strains isolated from termite guts have shown promising potential for enhancing lignocellulosic pretreatment processes, which are critical in the conversion of biomass into bioenergy. Consequently, termite-associated microbiomes are increasingly recognized as valuable sources of novel cellulolytic microbes and enzymes with significant industrial relevance for sustainable biofuel production and biomass utilization [9].

The termite gut microbiome depicts one of the most efficient natural bioreactors, where microbial communities interact synergistically to degrade lignocellulose. These microorganisms produce a variety of carbohydrate-active enzymes (CAZymes) which includes cellulases, hemicellulases, and lignin-degrading enzymes that helps in the breakdown of plant biomass. Such enzymatic capabilities have attracted significant interest for applications in biofuel production, waste management, and industrial biotechnology [10].

Intestinal symbiont distribution and community profile within the termite gut operate under strict compartmentalization, heavily influenced by host evolutionary lineage, trophic habits, and environmental pressures. This dense, diverse internal microflora is crucial for driving the catabolism of complex lignocellulosic substrates. From an anatomical perspective, the alimentary canal is segmented into three primary physiological regions: the stomatodeum which is known as foregut, mesenteron known as midgut, and proctodeum commonly known as hindgut [11]. In the anterior region, the stomatodeum features an expanded crop that narrows into the gastric caeca. Branching off from these caecal structures and terminating at the origin of the Malpighian tubules is the mesenteron, which presents as a

slender, tubular tract of constant internal diameter. Serving as the primary site for metabolic activity, the proctodeum occupies the largest volume of the gut; within advanced termite lineages, it undergoes further distinct structural partitioning into five specialized zones designated as P1 through P5, corresponding to the ileum, enteric valve, paunch, colon, and rectum, respectively.

Studies have identified dominant microbial phyla such as Firmicutes, Bacteroidetes, Spirochaetes, and Proteobacteria, along with methanogenic archaea involved in methane production. These microbes contribute to key metabolic pathways including fermentation, hydrogen production, and methanogenesis, forming a tightly regulated ecological network [2]. In addition to lignocellulose which are degrading microsymbionts, the termite gut also contains fungi, yeasts, and bacteria that have possible applications in industrial biofuel production, including biodiesel, bioethanol, and biohydrogen, either through fermentation of saccharified biomass or through the synthesis of bioproducts that can be converted into the desired fuels.

Traditional microbiological methods have limitations in studying such complex microbial communities because a significant proportion of microorganisms cannot be cultured under laboratory conditions. Metagenomics is a culture independent procedure that permits the isolation and analysis of genetic material obtained from environmental samples which are helpful in offering detailed insights into the taxonomic diversity and functional capabilities of microbial communities [6].

Advancements in next-generation sequencing technologies, particularly Illumina sequencing, have enabled high-throughput generation of metagenomic data with high accuracy and depth. These technologies produce large volumes of short sequencing reads that require robust bioinformatics tools for processing and analysis. Computational workflows involve several key steps, including quality control, assembly, gene prediction, and taxonomic classification [8].

The termite gut is considered one of the most efficient natural micro-ecosystems for lignocellulose degradation, supported by a complex and highly specialized symbiotic community of bacteria, archaea, and eukaryotic microorganisms. Although

remarkable progress has been made in understanding termite gut physiology and microbiology as much of the taxonomic and functional diversity of its microbial community remains unexplored, especially in higher termites. This thesis seeks to fill this knowledge gap through a comprehensive metagenomic analysis of the gut microbiome of selected termite species, with emphasis on identifying novel lignocellulolytic enzymes, microbial interactions, and metabolic pathways responsible for the exceptional biomass conversion efficiency of these insects. By combining high-throughput sequencing technologies, advanced bioinformatics tools, and functional validation approaches, this study aims to enhance our understanding of termite gut symbiosis while investigating its biotechnological potential for sustainable biofuel production and biorefinery applications. Ultimately, the findings will contribute to ecological knowledge and support the development of bio-based strategies for the valorization of lignocellulosic biomass [7].

1.1 Problem Statement

The termite gut microbiome is a highly complex and largely unexplored microbial system responsible for efficient lignocellulose degradation. Traditional methods are insufficient to fully characterize its diversity and functional roles. However, despite advances in sequencing technologies, comprehensive metagenomic analysis of the microbial communities, functional genes, and enzymatic pathways responsible for this breakdown remains limited. Current knowledge is restricted to a few termite species and culture-dependent methods. A metagenomic approach is therefore essential to fully decode the lignocellulose degrading potential of the gut termite microbiome.

1.2 Aim and Objectives

1.2.1 Aim

To analyze the diversity and community structure of the termite gut microbiome using metagenomic sequencing and bioinformatics tools.

1.2.2 Objectives

i. To characterize the taxonomic diversity of bacterial communities in the termite gut using high-throughput 16S rRNA gene sequencing.

ii. To perform functional annotation for the identification of lignocellulose-degrading genes, enzymes and pathways.

iii. To predict the cellulose / lignocellulose-degrading enzymes from the meta genome of termite gut microbiome.

1.3 Gap Analysis

Significant advances have been made in the characterization of termite gut microbiomes globally; the bacterial communities inhabiting the hindguts of termite populations indigenous to Khyber Pakhtunkhwa (KPK), Pakistan remain entirely uncharacterized at the metagenomic level. The vast majority of data on termite gut microbiomes is available from sub-Saharan Africa, South and Southeast Asia, and the Americas [12]. No study to date has applied high-throughput metagenomic sequencing to characterize the taxonomic diversity and functional gene content of cellulose-degrading bacteria in KPK termites, leaving their phylogenetic affiliations, relative abundances, and CAZyme-encoding potential entirely unknown. Termites from this climatically and ecologically distinct region may harbour novel cellulolytic bacterial lineages absent from currently described termite gut microbiomes, with significant implications for the discovery of novel cellulases applicable to sustainable biofuel production and industrial bioprocessing [13].

Chapter 2

Literature Review

2.1 Classification of Termites

Termites have traditionally been recognized as eusocial insects belonging to the order Isoptera. Although it is widely accepted that termites are closely related to Blattodea and share common ancestry with net-winged insects, their exact phylogenetic placement remains under discussion. Similar to Hymenoptera, termites exhibit advanced social organization, and molecular studies indicate that they form a monophyletic lineage closely associated with Cryptocercus cockroaches in fig 2.1 [14]. In the early years of 2009, termite species were classified by taxonomists into seven groups representing their diverse evolutionary lineages. Later, American researchers David Grimaldi, Michael Engel, and Kumar Krishna revised the classification system, reorganizing termites into nine groups after 2009. These families include Hodotermitidae, Kalotermitidae, Archotermopsidae, Rhinotermitidae, Mastotermitidae, Serritermitidae, Termitidae, Stolotermitidae, and Stylotermitidae [15]. The order Isoptera, divided into lower and higher termites and comprising four families and four genera, included a total of 473 species, many of which were highly destructive. Among these, five species were reported from Southern China [16].

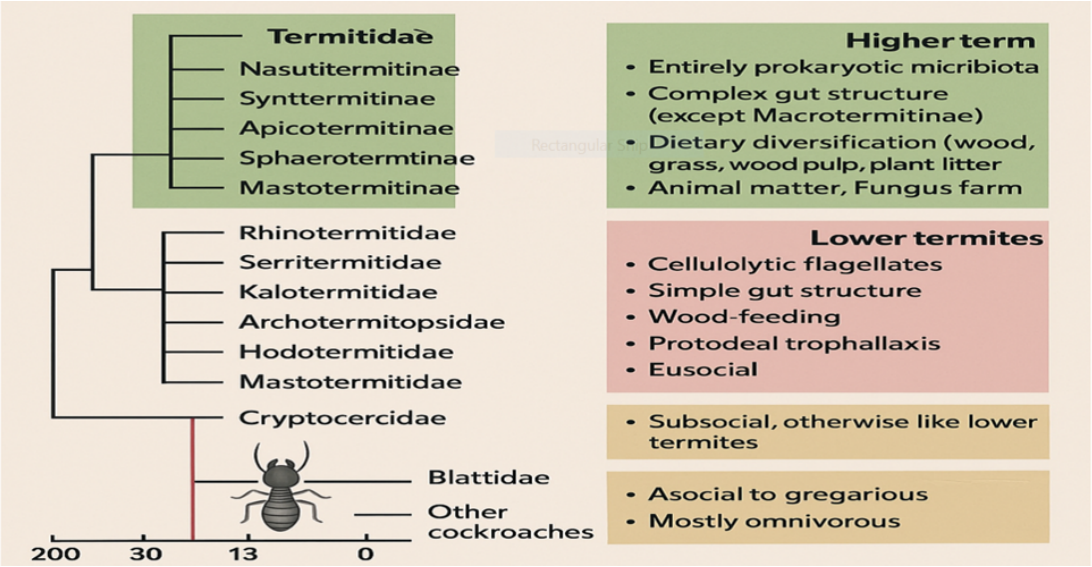

FIGURE 2.1: Termites Phylogeny showing its evolutionary journey [14].

2.2 Ecology of Termites

The ecology of pests, including termites, is concerned with understanding their behavior and interactions with the surrounding environment and other organisms. Termites are generally unable to survive in cold or freezing climates because of their delicate cuticle. Based on habitat and feeding preferences, termites are broadly classified into three ecological groups: subterranean, drywood, and dampwood termites, each showing distinct ecological characteristics and behavior [17].

Cryptotermes brevis, a drywood termite, is recognized as an invasive species in Australia. Both drywood and dampwood termites possess chewing mouthparts and mainly feed on dead plant material such as leaf litter, timber structures, soil-associated organic matter, crops, forests, and plantations. Termites can damage a wide range of materials, including buildings, valuable documents, artwork, books, flooring, carpets, and clothing. Subterranean termites, which live in organized underground colonies, usually attack dead wood and avoid healthy living trees. Termites are mostly found in tropical and subtropical regions, mainly between approximately 50° north and 50° south latitudes, where they contribute significantly to ecosystem biomass [18].

Termites are highly skilled builders, constructing nests and mounds of various sizes and shapes to accommodate the entire colony. Arboreal termites build nests on trees, whereas ground-dwelling species construct elaborate earthen mounds. Tree nests help conserve water by regulating condensation and provide protection for colony members. Deep inside these nests, nursery chambers shelter eggs and early larval stages, where workers provide continuous care. The sophisticated mound structure allows termites to remain active above ground. Many mounds have narrow ends oriented toward the sun during the hottest part of the day, which promotes the upward movement of warm air and improves ventilation within the underground tunnel system. This air circulation is especially important for species that cultivate fungal gardens or must maintain brood temperature within a narrow range ($\pm 1°C$), as it helps regulate internal nest conditions [18].

In tropical savannas, termite mounds vary greatly in size and form. Some species build exceptionally large mounds that may reach heights of up to 9 meters, particularly in wooded areas where conical mounds are common. However, most mounds in such habitats are typically 2-3 meters tall. Their structures may include sculptured hard-soil mounds, irregular domes, grass-covered cones, woody shrub-covered mounds, or combinations of these forms [19].

2.3 Termites Distribution

Termites can be found all over the world, but they are most common in tropical rainforests near the equator, in both the Northern and Southern Hemispheres [20]. If you look at the Eastern Hemisphere, you'll see that it has a lot more different types of termites than the Northern Hemisphere, especially in areas that are farther away from the equator. In fact, termites in the Eastern Hemisphere have been recorded at elevations of up to 2000 m in mountainous regions, where their distribution is more extensive than in the Western Hemisphere [21]. Overall, termites are found on every continent except Antarctica, with their populations concentrated in tropical, subtropical, and warm climatic zones. They thrive particularly well in warm, humid lowlands and coastal environments, with species diversity

generally declining toward higher latitudes. Variations in termite diversity and distribution are evident not only between continents but also among countries. For example, Europe hosts about ten species, whereas North America has approximately 50 species [18]. South America exhibits exceptionally high diversity, with more than 400 recorded species. In Asia, particularly in China, around 435 species have been identified, mainly in tropical and subtropical regions south of the Yangtze River. Of the roughly 3000 described termite species worldwide, about 1000 are found in Africa, reflecting a wide ecological distribution and diversity of mound-building forms [22].

In Ethiopia's central rift valley, seven termite genera *Macrotermes, Microtermes, Odontotermes, Amitermes, Angulitermes, Microcerotermes,* and *Trinervitermes* have been identified based on ecological distribution and species composition. Among these, *Trinervitermes* and *Angulitermes* are relatively rare and confined to specific habitats, whereas the remaining genera are more widely distributed across the region [23]. Termites in these areas were sampled using maize stalk baiting and standardized belt transects across different land-use systems, including farmlands, rangelands, and protected areas. These land-use types vary in the degree of human disturbance, with protected areas being the least disturbed, followed by rangelands and farmlands. Such differences suggest that agricultural practices and livestock grazing may negatively affect certain termite species, possibly by reducing the availability of plant materials such as grass and straw that serve as food sources [24].

The subterranean termite *Reticulitermes chinensis* is a notable species widely distributed across several regions of China, including Beijing, Shaanxi, Tianjin, Shanxi, Chongqing, Huanggang, Changsha, and the Yangtze River basin [25]. This species is considered a significant pest due to its destructive impact on trees, wooden structures, and plant vascular tissues [26]. Another important species in China is *Reticulitermes aculabialis*, which also contributes to environmental and structural damage [27]. In contrast, in regions such as North America and the United Kingdom, soil fauna studies often focus more on earthworms due to the relatively lower prominence of termites and ants. However, in countries like

India, South America, and across Africa, extensive research has highlighted the significant role of termites in influencing soil properties [28]. In Africa, much of this research has concentrated on fungus-growing termites, particularly species of *Macrotermes*, which construct characteristic mounds—organisms not found in Australia. Australia, nevertheless, supports approximately 150 termite species across 25 genera, reflecting its own distinct termite diversity [29].

2.4 Evolution of Symbiotic Digestion

Now it is widely accepted that termites that are originated from within a cockroach lineage, evolved from a subsocial, wood-feeding ancestor. Robust molecular phylogenetic analyses place termites as a monophyletic group nested within Blattodea, with their closest living relatives being members of the Cryptocercidae. Fossil records combined with molecular clock estimates suggest that this divergence occurred during the Middle Jurassic period, approximately 150-170 million years ago (Figure 2.1) [30]. This evolutionary transition marks a significant shift from solitary or subsocial behavior toward the highly organized eusocial systems that characterize modern termite colonies.

A pivotal innovation in termite evolution was the establishment of a mutualistic symbiosis with cellulolytic flagellate protists in the hindgut of early termites. This association, inherited from a common ancestor shared with Cryptocercidae, enabled efficient degradation of lignocellulosic biomass, particularly cellulose and hemicellulose, which are otherwise recalcitrant to digestion. In addition to protists, a diverse community of bacteria and archaea co-evolved within the termite gut, forming a complex, multi-partner symbiotic system that supports nutrient acquisition, carbon cycling, and energy metabolism [6]. This tripartite interaction significantly enhanced the ecological adaptability of early termites and allowed them to exploit wood as a primary resource.

However, a major evolutionary shift occurred with the emergence of the higher termite family Termitidae during the early Cenozoic era, around 50 million years ago

[31]. Unlike lower termites, members of Termitidae lost their symbiotic cellulolytic flagellates and instead developed a gut microbiome composed almost entirely of prokaryotes, particularly specialized bacterial lineages. This transition was accompanied by substantial morphological and physiological modifications of the gut, including compartmentalization and the formation of distinct microhabitats that support diverse microbial functions.

The loss of protists did not reduce digestive efficiency; rather, it was compensated by the expansion and functional diversification of bacterial communities capable of lignocellulose degradation, acetogenesis, and nitrogen fixation. Furthermore, higher termites exhibit a remarkable diversification in feeding strategies, including wood-feeding, soil-feeding, humus-feeding, and fungus cultivation (e.g., symbiosis with *Termitomyces* fungi in fungus-growing termites). These innovations enabled them to colonize a wide range of ecological niches and significantly contributed to their evolutionary radiation and global ecological success [31].

2.4.1 Termite Gut Habitat

The termite gut functions as a highly specialized micro-ecosystem composed of diverse microhabitats that vary significantly in their physical and biological conditions. These variations are influenced both by the structural organization of the gut and by the metabolic activities of the termite and its associated microorganisms. Gradients in factors such as oxygen levels, pH, and redox potential create distinct ecological niches that support a wide range of microbial communities. Over evolutionary time, the complexity and diversity of these niches have undergone substantial changes, reflecting adaptations in both the host and its symbionts [31, 32].

2.4.2 Gut Structure

Termites digestive system closely resembles with that of cockroaches (Figure 2.2), highlighting their shared evolutionary origin. Food enters through the mouth and

is transported into the crop via the foregut, where it mixes with salivary secretions. It then moves to the gizzard, where mechanical digestion occurs, before entering the midgut. In the midgut, digestive enzymes break down food, and nutrients are absorbed through the gut lining and gastric caeca when present.

Undigested material passes into the hindgut (proctodeum), which is divided into several compartments: the ileum (P1), enteric valve (P2), colon (P3 and P4), and rectum (P5). This region contains the majority of the gut microbiota and serves as the main site for microbial fermentation and nutrient recycling. Water and ions are reabsorbed in the hindgut before waste is excreted as feces [33].

Compared to cockroaches, termites show distinct structural adaptations, including a reduced crop, a shortened midgut, and reduced or absent gastric caeca. In lower termites, the anterior hindgut region (P3) is enlarged into a single, large fermentation chamber. In higher termites (excluding the Macrotermitinae), the hindgut is further elongated and subdivided into multiple specialized compartments, supporting diverse microbial processes [34].

Although bacteria are present throughout the digestive tract in both termites and cockroaches, the foregut and midgut are mainly dominated by host-produced digestive enzymes such as proteases, lysozymes, and chitinases. Due to the rapid passage of food through these regions, extensive microbial fermentation is limited. However, the presence of metabolites such as lactate and short-chain fatty acids indicates that some microbial activity does occur [5, 31].

2.4.3 Microhabitats

The rapid movement of digestive material through the termite gut creates challenges for microbial survival, requiring microorganisms either to move actively or attach to solid food particles to avoid being flushed out. In lower termites, gut flagellates adapt by exhibiting strong motility or by attaching to the gut lining using specialized structures. Since these flagellates occupy a large portion of

the hindgut, many prokaryotes are closely associated with them, living on their surfaces, within their cytoplasm, or even inside their nuclei.

In addition, social behaviors such as trophallaxis (the exchange of gut contents among colony members) help maintain and transmit these microbial communities. This process ensures the continuity of symbiotic microorganisms across generations and preserves the stability and functionality of the gut ecosystem beyond the lifespan of individual termites [35].

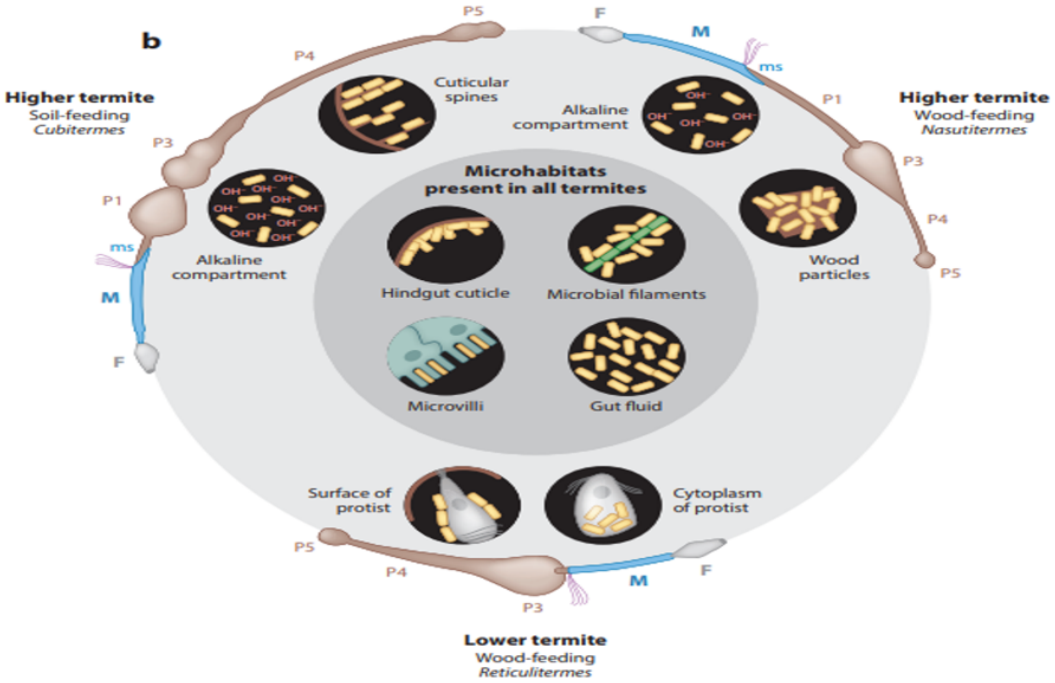

FIGURE 2.2: The anatomy of termite guts.

The hindgut is notably elongated and may be further divided into a mixed segment and several proctodeal compartments (P1-P5), creating additional environments for microbial colonization [30]. The inside of your gut isn't usually a great place for most bacteria to thrive, but there are some exceptions. For example, certain types of spiral-shaped bacteria can move quickly and stay in one spot, which helps them survive. This is interesting because, in some insects like termites that eat wood, the solid bits of wood stay in their gut for a longer time than the liquid around it. Something similar happens in termites that eat soil - the tiny clay particles that are rich in organic matter stick around longer than the bigger sand particles [31].

It's pretty interesting to think about how the tiny microbes inside the guts of certain termites play a huge role in helping them digest the wood they eat. Scientists have found that about a third of the microbial biomass in the hindgut of these wood-feeding termites is actually stuck to the fibrous material they're digesting. They've also discovered that bacterial cells are attached to the tiny little fingers of the termite's gut lining, and even in the space between the gut lining and the outer layer of the gut [5]. The gut itself provides a big surface area for these microbes to colonize, and it's often covered in a thick layer of biofilm. Some termites even have little spines in their gut that help the microbes attach. But here's the thing: when termites shed their skin, they also shed their entire gut lining, which means the microbes have to start all over again. This process, called ecdysis, is like a reset button for the termite's gut microbiota, and it's still not totally clear how they manage to re-establish their microbial communities after each molt [36].

2.5 Functional Niches

The symbiotic degradation of lignocellulose in the hindgut of termite mostly produces short-chain fatty acids and microbial biomass. These products obtained through fermentation are absorbed through the hindgut epithelium and serve as a major energy source for the termite, while the microbial biomass itself also contributes significantly to host nutrition. Although the particular roles of many microbial taxa are still not fully understood, extensive research has enabled the identification of key functional niches within the termite hindgut that are conserved across different feeding groups (Figure 2.3) [32].

2.5.1 Polymer Degradation

A central function of the hindgut microbiota is the breakdown of complex, fibrous plant polymers. In lower termites, flagellate protists play a dominant role by producing a wide array of glycoside hydrolases that facilitate efficient degradation of ingested wood. These enzymes include cellulases such as endoglucanases and

exoglucanases, as well as hemicellulases like xylanases, arabinosidases, and mannosidases [9]. Studies on *Zootermopsis angusticollis* have further demonstrated that the protist community is a major contributor to chitinase activity within the hindgut [37]. Wood degradation largely occurs within the digestive vacuoles of these flagellates, effectively isolating the process from the surrounding gut fluid and potentially restricting bacterial access to released sugars. However, recent metabolomic studies indicate that bacteria found in the hindgut of lower termites also contribute significantly to the breakdown of cellodextrins [6].

In higher termites, the absence of flagellates has driven the evolution of alternative mechanisms for lignocellulose degradation. One notable adaptation is observed in the Macrotermitinae, which form a symbiotic association with lignin-degrading basidiomycete fungi (*Termitomyces spp.*). Unlike gut-resident symbionts, these fungi are cultivated externally in specialized structures known as fungus combs within termite nests. Termites feed on mature portions of these combs, which consist of partially degraded lignocellulose combined with fungal biomass. Variations in comb composition among different termite genera likely contribute to differences in their associated bacterial communities [29].

Bacterial populations in the termite gut, particularly members of the Bacteroidetes phylum, possess a diverse repertoire of glycoside hydrolase enzymes that enable the degradation of complex polysaccharides from both plant and fungal cell walls. Other higher termite groups have evolved specialized feeding strategies that rely on partially decomposed organic matter. This process, known as humification, involves the gradual transformation of materials such as herbivore dung and decaying plant debris, resulting in reduced cellulose content and enrichment of recalcitrant polysaccharides and nitrogen-rich microbial compounds. The composition and distribution of glycoside hydrolase enzyme families which differ distinctly between wood-feeding and dung-feeding termites, reflecting adaptations to their specific diets [2].

Recent studies on *Nasutitermes* species have identified cellulolytic activity associated with fiber-bound microbial communities, that are particularly members of the Fibrobacteres, Spirochaetes, and TG3 phyla. These findings help to explain

the high abundance of cellulase genes which are detected in hindgut in metagenomic datasets, many of which are attributed to Fibrobacteres and Spirochaetes, highlighting their significant role in lignocellulose degradation [2].

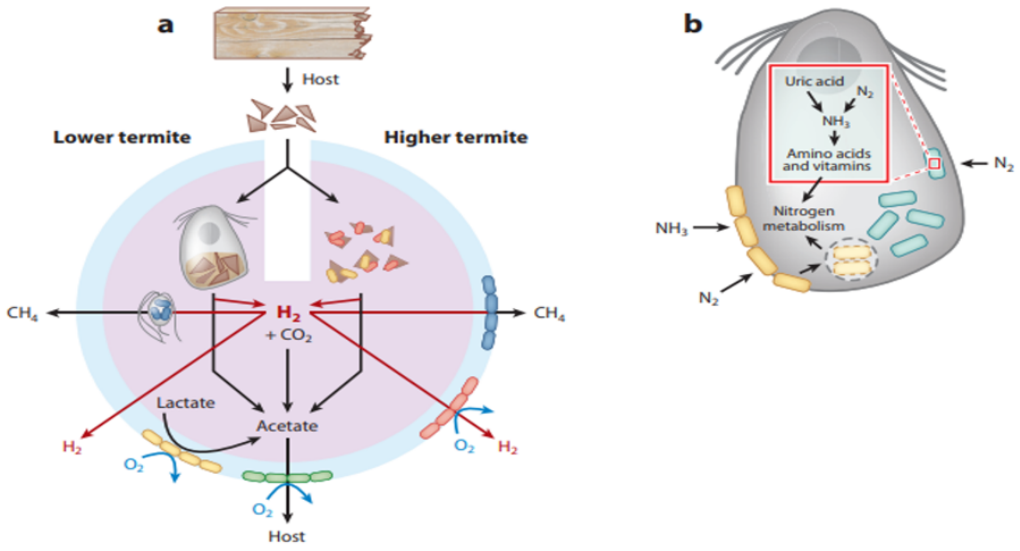

FIGURE 2.3: (a) The process of breaking down wood-derived polysaccharides into acetate and carbon dioxide (b) Symbiotic microorganisms play an important role in the nitrogen cycle of the flagellate host [31].

2.5.2 Hydrogen Metabolism

Hydrogen serves as an important intermediate during fermentation and often accumulates to high concentrations, as illustrated in (Figure 2.3a). In lower termites, it is produced as a byproduct when cellulolytic flagellates degrade polysaccharides into acetate and carbon dioxide [32]. In contrast, the exact microorganisms responsible for initiating fermentation in higher termites are still not clearly identified [31]. In wood-feeding termites, most of the hydrogen generated in the gut is consumed in the reduction of CO_2 to form additional acetate via reductive acetogenesis [5].

When it comes to the microbiota of gut termites, hydrogenotrophic methanogenesis is a common metabolic process. However, its overall impact is relatively small

in termites that feed on wood. This is probably because there isn't enough hydrogen available for methanogens, which are mostly found in the outer areas of the hindgut. On the other hand, termites that grow fungi or feed on soil tend to produce more methane, and the reason for this is still not fully understood. One possible explanation is that the community of methanogens is different in these termites, with a higher proportion of strictly methylotrophic lineages. Alternatively, the structure of the gut might be arranged in a way that facilitates the transfer of reducing equals between different sections [6]. In some species of lower termites that produce more methane, methanogenic archaea are often found in close association with flagellates in the central gut region, which is rich in hydrogen [38].

2.5.3 Nitrogen Metabolism

The low nitrogen content of lignocellulosic material presents a major limitation for wood-feeding termites. However, their hindgut microbial community plays a critical role in nitrogen fixation, recycling, and enrichment [32]. Fermentation by gut microbes produces metabolites that are readily absorbed through the host's gut epithelium. Meanwhile, microbial biomass becomes available for digestion in the midgut only after being transferred via proctodeal trophallaxis. In termites that consume soil or dung, the degradation of nitrogen-rich materials during digestion often leads to a net production of ammonia [31].

Termites primarily fulfil their requirements for essential amino acids and vitamins by digesting microbial biomass in the midgut. The presence of diverse nifH gene homologs key markers of nitrogen fixation highlight the strong diazotrophic potential within their gut microbiota [2]. In lower termites, nitrogen fixation and enrichment are largely influenced by symbiotic flagellates, as shown in Figure 2.3b. However, the specific microorganisms responsible for the high nitrogen fixation rates observed in wood-feeding higher termites remain unclear [39].

2.6 Coevolutionary Patterns

Initial studies on bacterial diversity in termite guts pointed toward a potential coevolutionary relationship, observing that closely related termite species even when geographically distant harbored similar groups of genetically related gut bacteria [40].

With the initiation of high-throughput sequencing technologies enabling detailed taxonomic profiling, strong evidence has emerged supporting coevolution between termite hosts and their gut microbiota across a wide range of species. While certain microbial lineages reflect patterns consistent with symbiont-host cospeciation, others appear to be more strongly influenced by ecological changes and shifts in microhabitats that occurred during the host's evolutionary history [41].

2.6.1 Evidence for Cospeciation

The gut microbiota of termites exhibits a high degree of consistency both within a single colony and among different colonies of the same species [36].

This uniformity suggests that proctodeal trophallaxis plays a crucial role in preserving a stable microbial community and ensuring reliable transmission of symbionts across generations factors that can promote cospeciation over time. However, definitive evidence of cocladogenesis with termite hosts has so far been demonstrated only for "*Candidatus Azobacteroides pseudotrichonymphae*," an intracellular symbiont of *Pseudotrichonympha* flagellates. These flagellates themselves undergo cospeciation with termites of the family Rhinotermitidae, forming a termite-specific lineage of bacterial associates [42].

In the case of the endosymbiont "*Candidatus Endomicrobium trichonymphae*" and the ectosymbiont "*Candidatus Armantifilum devescovinae*," the phylogenetic relationships of the bacterial lineages closely mirror those of their respective flagellate hosts (Figure 2.4b).

However, their evolutionary congruence with termite hosts appears to be disrupted, possibly due to horizontal transfer of flagellate symbionts among different termite species [43].

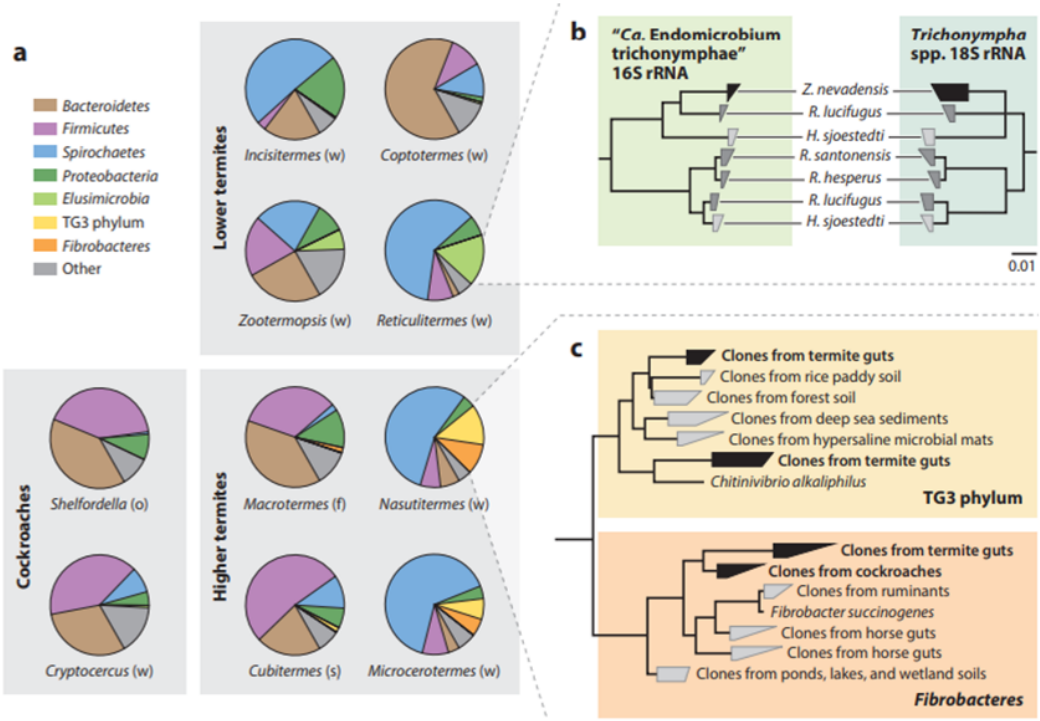

FIGURE 2.4: Microbial Diversity in Termite and Cockroach Digestive Tracts (a) Broad-Scale Phylum Shift Across Key Insect Branches [44]. (b) A comparison of small-subunit rRNA phylogenetic trees. [32].

2.6.2 Habitat-Specific Lineages

Not all flagellate-associated symbionts follow clear cospeciation patterns. For example, "*Candidatus Desulfovibrio trichonymphae*" is found in some, but not all, *Trichonympha* species, suggesting acquisition through non-vertical (environmental or horizontal) transmission [43]. Comparable trends have been reported in other microbial groups, including members of *Elusimicrobia*, *Spirochaetes*, and *Bacteroidetes* [44].

In many instances, some closely related bacteria inhabit the surfaces or cytoplasm of flagellates that are themselves not closely related. Conversely, similar ecological

niches may also be occupied by entirely unrelated symbionts that perform equivalent functional roles, such as nitrogen fixation or the synthesis of amino acids and vitamins [43]. These observations highlight the dynamic nature of microhabitats within the termite gut, which can shift throughout host evolution.

In higher termites, the loss of flagellates and the appearance of new ecological niches for cellulolytic bacteria help explain the decline of flagellate-associated symbionts and the rise of bacterial lineages involved in fiber degradation, particularly in wood-feeding species [31]. Termite-specific clades within the *Fibrobacteres* and TG3 phyla (Figure 2.1c) are often associated with particular termite genera, although they do not consistently demonstrate strict cospeciation. Their close phylogenetic relationship with microbial clones found in leaf-feeding cockroaches suggests that ecological niche selection plays a major role in shaping their distribution [2].

2.7 The Termite Hindgut as a Natural Bioreactor for Plant Biomass

The breakdown of complex lignocellulosic materials by termites relies on a specialized internal ecosystem of symbiotic microorganisms which is acquired through evolutionary development. This digestive process occurs across three primary anatomical segments of the insect's gut: the stomatodeum, the mesenteron, and the proctodeum [33]. Initially mechanical handling and storage happen within the stomatodeum, which is characterized by an expanded anterior crop that leads into the posterior gastric caeca. Moving downstream, the mesenteron functions as a narrow, tubular conduit of uniform diameter, originating at these caecal projections and terminating where the Malpighian tubules insert. Finally, the ingested material enters the proctodeum, the most spacious section of the intestinal tract. Within phylogenetically advanced termite groups, this massive chamber undergoes elaborate subdivision into a sequence of five distinct functional microenvironments: the P1 ileum, P2 enteric valve, P3 paunch, P4 colon, and P5 rectum (Fig. 2.5).

Termites are broadly classified into two groups: lower and higher termites. This classification is primarily based on the presence or absence of flagellate protists in the hindgut [30]. Lower termites harbor flagellate protists along with other microsymbionts, which assist in the hydrolysis of cellulose-rich. In contrast, higher termites lack these protists but possess a more diverse community of prokaryotic symbionts. The evolutionary loss of cellulolytic flagellates in higher termites, combined with the acquisition of specialized bacterial communities, has led to more advanced lignocellulose degradation strategies [29–32].

This transition is also associated with significant differences in gut physiology between two groups (Fig. 2.5). Additionally, intermediate or transitional taxa exist, displaying characteristics of both lower and higher termites [44, 45].

During feeding, termites first mechanically break down ingested biomass using their mandibles. This mastication process reduces large plant material into smaller particles, facilitating the initial disruption of lignocellulosic structures [38, 39]. Further mechanical processing occurs in the foregut, where particles are milled down to sizes of approximately 10-20 μm [5, 33].

This reduction increases the surface area available for enzymatic and microbial activity. The finely processed material then moves to the midgut and hindgut, where most of the biochemical degradation takes place.

The foregut and midgut, which are relatively oxygen-rich, are primarily involved in modifying lignin and hemicellulose components of plant biomass [33, 39]. In contrast, the hindgut provides an anaerobic environment where the majority of cellulose degradation occurs [33, 46].

The total transit time of food through the termite gut is approximately 24 hours, with retention times of 1-2 hours in the foregut, 1-1.5 hours in the midgut, and 14-20 hours in the hindgut [47].

The hindgut hosts the highest density and diversity of microorganisms, whereas the foregut and midgut contain comparatively fewer microbes [5, 48]. This extended retention time and favorable anaerobic conditions make the hindgut the

primary site for microbial activity and symbiosis.

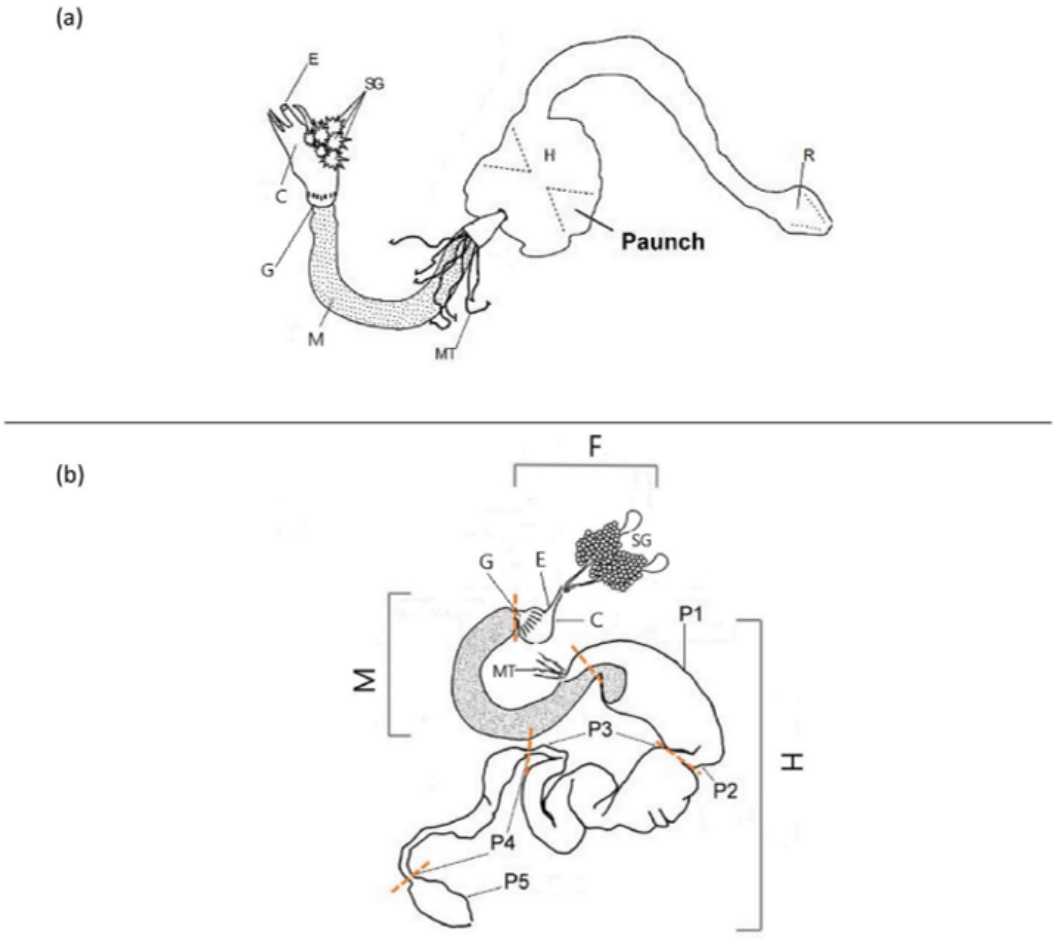

FIGURE 2.5: A diagramatic representation of the gut of (a) *Reticulitermes flavipes* which is lower termite and (b) *Labiotermes labralis* which is higher termite.

During digestion, gut microsymbionts produce a wide range of lignocellulolytic enzymes that enable efficient breakdown of plant biomass. Most of these microorganisms are not transient but are specialized symbionts that have co-evolved with their termite hosts, often showing host-specific community structures at the genus level [49]. Due to their collective activity, termites are capable of degrading approximately 74–99% of cellulose and 65–87% of hemicellulose within 24 hours. This efficiency surpasses that of many other biological systems, including ruminants feeding on less lignified grasses and isolated bacterial or fungal cultures. Consequently, termite gut microbiota are now recognized as a valuable and promising source of lignocellulose-degrading enzymes for biotechnological applications.

2.8 Diversity of Termite Gut Microsymbionts

For over thirty years, studies have shown that the termite gut contains a highly specialized community of symbiotic microbes [49]. This internal ecosystem includes a complex network of bacteria, archaea, and eukaryotes that work together to break down tough wood fibers into digestible nutrients and short-chain fatty acids. Bacteria are the most abundant group in both lower and higher termites. Genetic and microscopic data show that bacteria make up over 90% of the gut population. The remaining minor groups include fungi (1.7%), protists (1.6%), archaea (0.5%), plants/algae (0.2%), and viruses (0.2%) [50]. Termites acquire these vital microbes in two ways: vertically from their parents, and horizontally through social interactions like oral or anal feeding (trophallaxis) and environmental exposure [51].

In lower, wood-feeding termites, anaerobic protists from the phyla Preaxostyla and Parabasalia are the primary drivers of wood digestion, a trait they share with wood-eating cockroaches [52]. These protists thrive in oxygen-free zones of the gut, where they use specialized enzymes to break down cellulose. Because these single-celled organisms are large and highly abundant (ranging from 10^3 to 10^7 cells per gut), they fill over 90% of the hindgut paunch (the P3 compartment). Some of the large protists even harbor smaller bacteria inside themselves to help digest fiber more efficiently. Most of these protists are strictly endemic to termite and cockroach guts and cannot survive in the outside environment. Specific species like *Holomastigotes elongatum*, *Dinenympha gracilis Trichonympha agilis* and *Spirotrichonympha* are found exclusively in Reticulitermes termites and closely related wood-boring roaches [53].

Despite their apparent importance, the exact contribution of these flagellates to lignocellulose digestion remains debated. Some studies suggest that removing protists leads to termite mortality, while others report survival even in their absence [54]. However, only one study reported that termites continued to survive without

flagellates. Since the study did not provide enough details about colony condition, the method used to remove flagellates, or the long-term effects, the findings remain uncertain and the issue is still not fully understood.

In addition to protists, termite guts also contain yeasts from genera such as *Sporothrix*, *Candida*, *Pichia*, and *Debaryomyces*, identified in species like *Neotermes jouteli*, *Mastotermes darwiniensis*, *Reticulitermes santonensis*, and *Zootermopsis spp.*. Novel hemicellulose-degrading yeasts, including *Papiliotrema odontotermitis* and *Sugiyamaella mastotermitis*, have also been isolated from termite guts [55]. Although their cellulolytic activity is relatively low, these yeasts contribute significantly to the breakdown of aromatic compounds such as lignin. Additionally, some yeast species produce valuable bioproducts with potential industrial applications.

The secretory capacity of filamentous fungi represents a critical, albeit low - abundance, component of the termite intestinal tract. Despite their low numerical representation relative to other commensal microbes, these fungi express an extensive repertoire of lignocellulolytic enzymes that significantly drives the breakdown of complex plant fibers. This functional relevance is demonstrated by specific isolates, such as *Aspergillus sp.* A1 from *Bulbitermes sp.*, as well as *Aspergillus nomius* and *Trichoderma harzianum* characterized within *Odontotermes obesus* hosts [56].

By contrast, the structural deconstruction of lignocellulose is primarily driven by the bacterial domain, which exhibits the highest levels of both taxonomic diversity and total cell density within the gut. Surveys of the termite intestinal microbiome consistently identify several major bacterial phyla Spirochaetes, Firmicutes, Proteobacteria, Fibrobacteres (alongside the allied candidate phylum TG3), Actinobacteria, Bacteroidetes, Planctomycetes, Cyanobacteria, and Acidobacteria [48]. Among these taxa, the relative abundance of Spirochaetes closely correlates with host diet; they typically predominate in wood and grass feeding termites but face sharp declines in lineages sustained by humus or cultivated fungi [57].

These dietary dependencies manifest as highly variable population profiles across different termite species. For example, inside the hindgut of the grass- and wood-consuming species *Cortaritermes fulviceps*, Spirochaetes account for approximately 47% of the total microbial community, whereas Proteobacteria and Firmicutes represent secondary fractions at roughly 17% each, followed by minor populations of Fibrobacteres (2%) and Bacteroidetes (5%). While a nearly identical community architecture characterizes *Nasutitermes aquilinus*, distinct shifts in dominance occur in other hosts. In *Cornitermes pugnax*, *Trinervitermes trinervoides*, and *Pseudacanthotermes militaris*, the typical dominance of Spirochaetes is absent, with Firmicutes, Bacteroidetes, and Fibrobacteres emerging as the baseline dominant phyla instead [58].

Overall, *Spirochaetes*, *Firmicutes* and *Fibrobacteres* are generally considered the most prominent bacterial groups in termite guts. Their relative abundance varies depending on termite species, habitat, and the type of lignocellulosic material consumed. Importantly, these dominant bacterial phyla are key producers of lignocellulolytic enzymes essential for efficient biomass degradation within the termite gut ecosystem.

2.9 Gut Microbiome of Termite

Numerous microbial strains are isolated from termite guts display a wide range of specialized metabolic activities, including methanogenesis, acetogenesis, nitrogen fixation, fermentation, and oxygen reduction [59].

2.9.1 Protists in the Termite Gut

Protists are exclusively found in the hindgut of lower termites, where they play a vital role in the degradation of lignocellulosic material. In contrast, higher termites rely mainly on host-derived enzymes and bacterial communities for plant biomass digestion, and protists have not been reported in these species. The first

protozoan in termites was described by Lespes in 1856 [60], and later studies by Cleveland demonstrated that termites cannot survive for long without their gut flagellates [61].

In terms of diversity, termite gut protists are primarily represented by members of the the order *Oxymonadida* of phylum *Preaxostyla* and the phylum *Parabasalia* and [62]. Subterranean termites generally host a more diverse protist community than wood-feeding species. The composition of these protists varies among termite groups; for example, termites in the subfamilies *Heterotermitinae* and *Coptotermitinae* harbor greater flagellate diversity compared to those in the family *Rhinotermitidae*. Meanwhile, members of the family *Stolotermitidae* are known to contain only a few oxymonad flagellates, such as *Oxynympha loricate*, *Termitimonas travisi* and *Saccinobaculus*-like species.

Methanogenic prokaryotes are also associated with these protists, existing either as endosymbionts within their cytoplasm or nucleus or as ectosymbionts attached to their surfaces. Functionally, flagellates are essential for termite nutrition and development. They degrade complex polymers like cellulose, hemicellulose, and chitin, and contribute to metabolic processes such as fermentation, acetogenesis, and the production of hydrogen and carbon dioxide—important intermediates in biofuel-related pathways. Termites lacking flagellates are unable to efficiently digest wood, which is naturally poor in nitrogen and nutrients. Within these symbiotic systems, glucose released by flagellates is metabolized into pyruvate, which in parabasalid flagellates is further processed in specialized organelles known as hydrogenosomes [63].

2.9.2 Bacteria in the Gut Termite

The gut microbiota of cockroaches and termites is believed to have evolved from a common ancestral origin. In higher termites, bacteria are the primary drivers of digestion, as protists are absent. The termite gut contains approximately 15 bacterial phyla, reflecting a level of microbial diversity that often exceeds that found in many other animal digestive systems [64].

In lower termites, bacteria typically form symbiotic associations with flagellates, while in higher termites they are often linked with fungal partners, especially members of the *Basidiomycota* and *Ascomycota*. The dominant bacterial groups include *Bacteroidetes*, *Spirochetes*, *Proteobacteria*, and *Elusimicrobia*. Among these, *Spirochetes* are particularly diverse and, in some species such as *Microcerotermes* and *Nasutitermes*, can account for more than half of the population of bacterial gut. In higher termites feeding on wood or grass, *Spirochetes*, *Fibrobacter*, and members of the candidate phylum TG3 are abundant, whereas *Firmicutes*, *Bacteroidetes*, and *Proteobacteria* dominate in soil-, humus-, and fungus-feeding termites [65].

Although *Spirochetes* are largely absent in fungus-growing termites, they play key roles in other higher termites by contributing to fiber degradation, fermentation, homoacetogenesis, and nitrogen fixation [65]. Studies on *Reticulitermes lucifugus* have identified major bacterial groups such as *Firmicutes*, *Proteobacteria*, *Spirochetes*, *Bacteroidetes*, and candidate phyla like TG1, including cellulolytic aerobic bacteria [45]. In this genus, *Clostridia* and lactic acid bacteria are involved in sugar metabolism, acetogenesis, and fermentation processes. Additionally, *Proteobacteria* help regulate hydrogen and carbon dioxide levels in the hindgut, while *Desulfovibrio* species contribute significantly to nitrogen fixation across different termite species [66].

2.9.3 Archaea in the Termite Gut

Archaeal communities in termite guts vary between lower and higher termites. In lower termites, archaea are predominantly represented by members of the order *Methanobacteriales*, especially the genus *Methanobrevibacter*. In higher termites, archaeal diversity is broader, including groups such as *Methanosarcinales* (e.g., *Methanomicrococcus*), *Methanomicrobiales*, *Methanomassiliicoccales*, as well as non-methanogenic taxa like *Thermoplasmatales* and representatives of the phylum *Thaumarchaeota*.

Several archaeal species have been identified in termite guts, including *Methanobrevibacter filiformis*, *M. curvatus, M. smithii, M. cuticularis, M. arboriphilicus,* and *M. bryantii*, along with *Natronococcus amyloticus, Methanospirillum hungatei, and Methanomicrococcus blatticola.* Species of Methanosarcina such as *M. vacuolata* and *M. barkeri* are also present. *Methanogenic archaea* produce methane primarily from CO_2 and H_2 or from substrates like methanol, whereas non-methanogenic groups such as *Thermoplasmatales* lack this capability [67].

Methane production is generally higher in higher termites, likely due to their more complex gut structure and greater availability of methanogenic substrates. Archaeal diversity is closely linked to the activity of bacteria and protists, which generate these substrates. Consequently, increased substrate availability leads to higher archaeal diversity and enhanced methane emissions, making termites a notable natural source of greenhouse gases [67].

2.9.4 Bacteriophages in Termites

Recent studies have revealed the presence of bacteriophages in termite guts, an area previously unexplored. These viruses are thought to enter the gut through ingested material such as wood and may be internalized by flagellates. Within these hosts, bacterial endosymbionts serve as targets for phage infection. Phages can initiate lytic cycles, leading to bacterial cell lysis and contributing to nutrient recycling within the gut ecosystem. The dense and structured microbial environment of the termite gut provides favorable conditions for phage proliferation. To date, bacteriophages have mainly been reported in worker castes of higher termites [68].

Identified examples include phage CVT22 (family *Podoviridae*), which infects *Citrobacter* species in Formosan subterranean termites, and newly discovered phages such as Tyrion and Arya that target *Enterobacter* species. Additionally, several novel viral groups have been reported, including members of the order *Caudovirales*, the family *Microviridae*, and circular viruses from the family *Genomoviridae*

(TaCV-1 to TaCV-4). Ongoing research continues to explore viral diversity in termite guts, particularly their interactions with microbial communities and their ecological roles. Environmental bacteriophages also influence gut bacteria, and mechanisms such as superinfection immunity may regulate phage–phage interactions within this complex system [69].

2.10 Lignocellulolytic Enzymes from Termite Gut Microsymbionts

The enzymatic machinery responsible for processing plant biomass is unified under the Carbohydrate-Active Enzymes (CAZymes) nomenclature. This comprehensive grouping consists of carbohydrate esterases (CEs), glycoside hydrolases (GHs), auxiliary activity enzymes (AAs), polysaccharide lyases (PLs), non-catalytic carbohydrate - binding modules (CBMs) and glycosyltransferases (GTs). Automatically, GTs function to assemble glycosidic bonds by pairing monosaccharides with lipid or protein acceptors, whereas PLs specialize in the non-hydrolytic cleavage of plant pectate. Efficiency is enhanced by CBMs, which do not perform catalysis themselves but instead maximize enzyme-substrate proximity by binding directly to insoluble cell-wall surfaces. These units are supported by AAs, which modify the surrounding tissue geometry to expose embedded structural carbohydrates to incoming GH, PL, and CE complexes [70].

Deconstruction of the main structural cell-wall polysaccharides specifically cellulose and hemicellulose is driven by a cooperative dynamic between GHs and CEs. The former group hydrolyzes and rearranges internal glycosidic linkages, while the latter removes the ester-linked side chains that stabilize the broader polysaccharide networks. Lignin, however, requires an entirely different approach due to its non-sugar, polyphenolic structure. Its breakdown is dictated by specialized ligninases categorized as redox-active AAs, which chemically alter the phenolic infrastructure to unmask shielded hemicellulose and cellulose fibers for subsequent enzymatic digestion.

Comparative metagenomic profiling demonstrates that the abundance and distribution of these cooperative CAZyme pools are directly shaped by the host insect's nutritional niche. Termite species that target raw vegetable matter or wood carry a significantly enriched repertoire of these carbohydrate-degrading genes relative to geophagous lineages [71]. This genomic disparity underscores a clear evolutionary optimization of the intestinal microbiota toward high-lignocellulose diets, contrasting sharply with soil-feeding species that exploit humified, nitrogen-dense substrates containing low concentrations of complex structural plant fibers.

2.10.1 Cellulases

Cellulose, the most abundant plant polysaccharide, that is composed of long chains of glucose units which is linked by β-1,4-glycosidic bonds. Its tightly packed crystalline structure makes it highly resistant to degradation. However, cellulases can effectively hydrolyze these linkages. Rather than being a single enzyme, cellulase represents a complex system of three main glycoside hydrolases: exoglucanases, endoglucanases, and β-glucosidases. Together, these enzymes are used to break down cellulose into glucose monomers [72].

Termite gut studies have identified that numerous microsymbionts are capable of producing cellulolytic enzymes in species such as, *Coptotermes*, *Trinervitermes*, *Reticulitermes* and *Odontotermes* [72].

2.10.2 Endo-1,4-β-glucanase

Endoglucanases cleave the internal bonds within cellulose chains, particularly in amorphous regions, thereby exposing chain ends for further degradation [32]. In termite guts, these enzymes are predominantly produced by bacteria, which thrive in cellulose-rich and moist environments. Several bacterial strains—such as, *Stenotrophomonas maltophilia, Lysinibacillus fusiformis, Paenibacillus lactis,* and *Bacillus cereus*—have been isolated from termite guts with notable endoglucanase

activity [73]. Among these, S. maltophilia showed significant activity, while *Bacillus subtilis* from *Reticulitermes labralis* demonstrated one of the highest recorded activities.

Bacterial endoglucanases often exhibit high specific activity. For example, *Bacillus sp.* CF96 displayed strong dual endo- and exoglucanase activity compared to yeast-derived enzymes, which generally show lower efficiency. Some bacteria also form close symbiotic associations with gut flagellates, either as internal or surface-associated partners, contributing to cellulose breakdown and nutrient acquisition.

Fungi within the termite gut also produce endoglucanases. Species such as *Sarocladium kiliense* and *Trichoderma virens* have demonstrated enzyme activities comparable to bacterial strains. Under solid-state fermentation conditions, fungal strains like *Aspergillus tubingensis* exhibit exceptionally high enzyme productivity, surpassing bacterial counterparts. This is likely due to their adaptation to low-moisture, solid environments. Although yeasts such as *Meyerozyma caribbica* have been identified as endoglucanase producers, their activity is generally lower, suggesting a lesser role in cellulose hydrolysis compared to bacteria and filamentous fungi.

Interestingly, enzyme activities from termite gut microbes are often comparable to or even exceed those from other environments such as soil, aquatic systems, or ruminant guts. This highlights the gut termite as a valuable source of efficient cellulolytic enzymes with potential industrial applications.

2.10.3 Exo-1,4-β-glucanase

Exoglucanases act on the ends of cellulose chains, results in releasing cellobiose units in a stepwise manner [1]. Unlike endoglucanases, they possess a tunnel-shaped active site that enables progressive cleavage of cellulose.

Various bacteria isolated from termite guts including *Bacillus sp., Pseudomonas sp.*, and *Chryseobacterium rhizoplanae* exhibit strong exoglucanase activity. Some enzymes, such as those from *Bacillus sp.* CF96, demonstrate dual functionality,

acting as both endo- and exoglucanases with high stability under diverse conditions [74].

Fungi such as *S. kiliense* and *T. virens* also produce exoglucanases, though often at lower levels compared to strains from other environments. This may reflect ecological adaptation, where fungi rely more on bacterial partners for certain stages of cellulose degradation. Overall, bacterial exoglucanases appear to play a dominant role in cellulose depolymerization within the termite gut.

2.10.4 β-Glucosidase

β-Glucosidases is used to catalyze the final step of cellulose degradation by converting cellobiose and short oligosaccharides into glucose [72]. Their activity is essential to prevent the accumulation of inhibitory intermediates and to ensure efficient biomass conversion.

Yeasts such as *Meyerozyma caribbica* and *M. guilliermondii* have been identified in termite guts with β-glucosidase activity, although their enzyme levels are relatively modest. Filamentous fungi, including *S. kiliense* and *T. virens*, are also effective producers, with some strains showing high productivity under solid-state fermentation conditions.

Bacterial species such as *Spirochaeta coccoides*, *Pseudomonas*, *Chryseobacterium*, and *Brevibacillus* have demonstrated strong β-glucosidase activity, often comparable to fungal enzymes [73]. These microorganisms likely act synergistically to optimize cellulose degradation. Notably, a novel enzyme with both β-glucosidase and β-xylosidase activity has been identified from *Enterobacter* species in termite guts, highlighting its potential for industrial applications in converting agricultural waste into fermentable sugars for biofuel production.

Overall, the coordinated action of cellulolytic enzymes from diverse termite gut microsymbionts highlights their efficiency in lignocellulose degradation and their significant potential for biotechnological exploitation.

TABLE 2.1: β-Glucosidase activities of microorganisms inhabiting the gut of diverse termite species.

Microorganism	**Termite species**	**Microbial species**	**Enzyme activity (U/ mL)**	**Reference**
Fungi	*Bulbitermes sp.*	*Aspergillus sp.*	22.97a	[74]
	NR	*A.tubingensis*	142.37a	[75]
	R. santonensis	*Sarocladium kiliense*	0.1	[76]
	R. santonensis	*Trichoderma virens CTGxAviL*	0.38	[76]
Bacteria	*Bulbitermes sp.*	*Brevibacillus sp.*	5.45a	[74]
	NR	*Cellulomonas sp.*	0.35	[75]
	NR	*Micrococcus sp.*	0.28	[75]
	Odontotermes hiananensis	*Bacillus subtilis K21*	0.01	[77]

NR – Not Reported. a U/g dry substrate.

2.10.5 Hemicellulases

Hemicellulose is the second most common polysaccharide found in plant cell wall [32]. It is a highly complex and varied polymer made of five-carbon pentose sugars (mostly xylose and arabinose), six-carbon hexose sugars (like galactose, mannose, and glucose), and acetyl groups. The exact type of hemicellulose depends on the plant type: glucuronoxylan is the main form in hardwoods, glucomannan dominates in softwoods, and arabinoxylan is the primary structure in grasses [32].

Because xylan is the main backbone of most hemicelluloses, breaking it down is a vital step in digesting plant biomass. This xylan backbone consists of β-1,4-linked D-xylose units attached to various side chains, such as acetyl, glucuronosyl, and arabinofuranosyl groups. Because the structure is so complex, a single enzyme cannot break it down completely; instead, it requires a coordinated group of

multiple enzymes working together. The xylanolytic system includes endo-1,4-β-D-xylanases as the primary enzymes, along with several accessory enzymes such as α-D-glucuronidases, β-D-xylosidases, ferulic acid esterases , acetylxylan esterases, p-coumaric acid esterases and α-L-arabinofuranosidases.

2.10.6 Xylanases

Because xylan forms the backbone of hemicellulose, xylanases play a central role in its depolymerization by cleaving the main chain and releasing xylose-rich oligosaccharides. In the termite gut, bacteria, yeasts, and filamentous fungi that collectively produce xylanolytic enzymes that act synergistically to degrade hemicellulose.

Genes encoding xylanases have been identified in the gut microbiota of termites such as *Cortaritermes* and *Nasutitermes*. Further studies have isolated xylanase-producing bacteria from several termite species, including members of Firmicutes, Spirochaetes, Actinobacteria, and Bacteroidetes [70]. Among these, *Paenibacillus macerans* IIPSP3 demonstrated particularly high enzyme activity, with strong volumetric and specific xylanase production. Another notable example is *Clostridium* sp. X53, which showed exceptionally high xylanase activity, surpassing that of many microbes from other environments such as ruminant and insect guts.

Bacterial xylanases are especially valuable for industrial applications due to their stability across a wide range of temperatures and pH levels, as well as their relatively low molecular weight. Additionally, bacteria are easier to culture, genetically manipulate, and harvest, making them highly suitable for large-scale enzyme production [78].

Filamentous fungi also contribute significantly to xylanase production. For instance, *Trichoderma virens* CTGxAviL, isolated from the gut of *Reticulitermes santonensis*, exhibited high enzyme productivity. Although fungal xylanase production is generally higher than that of other microbial groups, certain bacterial strains such as *Clostridium* sp. X53 can outperform fungal counterparts. This

highlights the potential of termite gut bacteria as a source of highly efficient lignocellulolytic enzymes for industrial use.

TABLE 2.2: Xylanase activities of microbial strains from the termite gut

Microorganism	**Termite Family / Species**	**Microbial Species**	**Enzyme Activity (U/mL)**	**Ref**
Yeast	*Coptotermes formosanus*	*Vanrija humicola* SSA-1544	6.06	[79]
	Odontotermes javanicus	*Hannaella pagnoccae* STAG 1.14	1.17	[80]
	Odontotermes javanicus	*Pseudozyma hubeiensis* STAG 1.7	1.31	[80]
	Reticulitermes chinensis	*Candida pseudorhagii* SSA-1542T	1.73	[79]
Bacteria	*Cryptotermes brevis*	*Bacillus* sp. BMP01	0.21	[81]
	Macrotermes barneyi	*Bacillus stratosphericus* BCMC2	0.25	[82]
	Macrotermes sp.	*Klebsiella spallanzanii* MA4MC	2.99	[83]
	Microcerotermes sp.	*Streptomyces* sp. MS-S2	6.56	[81]
	NR	*Clostridium* sp. X53	1252	[79]
	Reticulitermes santonensis	*Bacillus subtilis* ABGx	44.3	[84]
	Reticulitermes santonensis	*Sarocladium kiliense* CTGxxyl	40.8	[76]
	Termitidae	*Paenibacillus macerans* IIPSP3	120	[78]
Fungi	*Bulbitermes* sp.	*Aspergillus* sp. A1	96.82a	[74]
	NR	*Aspergillus tubingensis* M7	8188.00a	[75]
	Reticulitermes santonensis	*Trichoderma virens* CTGxAviL	426	[76]

In contrast, yeasts in the termite gut typically exhibit lower xylanolytic activity. Early studies reported modest enzyme production from yeasts such as *Candida*, *Pichia*, *Sporothrix*, and *Debaryomyces*. More recent work has identified additional yeast species from termite genera like *Coptotermes*, *Reticulitermes*, and *Odontotermes*, but their xylanase activity remains comparatively low. Even the highest reported activity, from *Vanrija humicola*, is still lower than that of bacterial and fungal producers [85]. These findings suggest that termites rely primarily on bacterial and fungal symbionts rather than yeasts for efficient xylan degradation, although yeasts may still play roles in biofuel-related processes.

2.10.7 Accessory Enzymes for Hemicellulose Hydrolysis

Xylanases alone are insufficient for complete hemicellulose breakdown. Efficient degradation requires the involvement of accessory enzymes that remove side chains and modify the structure of xylan. These enzymes including α - L - arabinofuranosidases, β-xylosidases, α-glucuronidases, and various hemicellulolytic esterases that work synergistically with xylanases to release xylose and xylooligosaccharides from hemicellulose polymers.

Together, the coordinated activity of xylanases and accessory enzymes ensures effective depolymerization of hemicellulose within the termite gut ecosystem.

Research on the specific types of live, growable termite gut microbes that generate xylan-degrading accessory enzymes is still quite limited [73]. Only a small number of these microbial strains have been successfully grown in laboratories. One example is the bacterium Bacillus licheniformis A11, isolated from the digestive tract of *Odontotermes hainanensis*, which yields β-xylosidase at concentrations of roughly 4 U/mL inside the cell and 11 U/mL outside the cell. Additionally, yeasts belonging to the genera *Candida*, *Pichia*, and *Sporothrix* have been retrieved from *Mastotermes darwiniensis, Reticulitermes santonensis,* and *Zootermopsis nevadensis*; these fungi produce arabinofuranosidase, though their overall output remains minimal. Currently, no laboratory has successfully isolated a live termite gut microbe capable of producing hemicellulolytic esterases, glucuronidases, or galactosidases.

To bypass these culturing difficulties, scientists utilize multiomic sequencing techniques to examine the genetic profile of the entire gut microbiome simultaneously. This DNA and protein analysis has successfully identified hidden genes for these missing accessory enzymes inside the gut communities of *Coptotermes gestroi*, *Pseudacanthotermes militaris*, *Cortaritermes sp.*, *Nasutitermes aquilinus*, and *Reticulitermes flavipes* [48]. By extracting these specific genetic codes and putting them into standard laboratory host organisms, researchers have been able to manufacture the enzymes artificially. Once expressed, these recombinant proteins show high functional performance, reaching peak activities of 72.27 U/mg for xylosidase, 22.85 U/mg for arabinofuranosidase, 67.20 U/mg for feruloyl esterases, 11.60 U/mg for mannanase, and 92 U/mg for mannosidase.

However, translating these findings into large-scale enzyme production remains challenging. Molecular cloning approaches often face limitations such as enzyme toxicity to host cells, poor solubility, or loss of enzyme functionality. Additional complications include unintended gene expression in host organisms and improper protein folding, which can hinder enzyme secretion and overall efficiency.

TABLE 2.3: Ligninase-producing microsymbionts from the guts of diverse termite species

Microorganism	**Termite Species**	**Microbial Species**	**Enzyme Activity (U/mL)**	**Ref**
		Laccase		
Bacteria	*Amitermes hastatus*	*Streptomyces sp. MV32*	5.66	[85]
Bacteria	*Bulbitermes sp.*	*Bacillus sp. B2*	71.18a	[74]
Bacteria	*Coptotermes curvignathus*	*Lysinibacillus sp.*	0.07	[86]
Bacteria	*Cryptotermes brevis*	*Ochrobactrum oryzae BMP03*	8	[81]
Bacteria	*Macrotermes barneyi*	*Bacillus stratosphericus BCMC2*	0.001	[82]
Bacteria	*Microtermes pakistanicus*	*Pseudocitrobacter anthropi MP-4*	0.04	[87]

Table 2.3 continued from previous page

Microorganism	Termite Species	Microbial Species	Enzyme Activity (U/mL)	Ref
Bacteria	*Microcerotermes sp.*	*Streptomyces sp. MS-S2*	0.05	[88]
Fungi	*Bulbitermes sp.*	*Aspergillus sp. A1*	69.44a	[74]
Fungi	*Odontotermes obesus*	*Aspergillus nomius MN700028*	0.04	[89]
		Manganese peroxidase		
Yeast	*Coptotermes formosanus*	*Meyerozyma caribbica SSA1654*	27	[90]
Fungi	*Bulbitermes sp.*	*Aspergillus sp. A1*	43.18a	[74]
Fungi	*Odontotermes obesus*	*Aspergillus nomius MN700028*	0.22	[89]
Bacteria	*Bulbitermes sp.*	*Bacillus sp. B1*	47.73a	[74]
Bacteria	*Coptotermes curvignathus*	*Lysinibacillus sp.*	0.08	[86]
Bacteria	*Microcerotermes sp.*	*Streptomyces sp. MS-S2*	0.005	[88]
		Lignin peroxidase		
Bacteria	*Bulbitermes sp.*	*Brevibacillus sp. Br3*	648.60a	[74]
Bacteria	*Cryptotermes brevis*	*Ochrobactrum oryzae BMP03*	14	[81]
Bacteria	*Coptotermes curvignathus*	*Lysinibacillus sp.*	0.26	[86]
Bacteria	*Microtermes pakistanicus*	*Pseudocitrobacter anthropi MP-4*	0.09	[87]
Bacteria	*Reticulitermes chinensis*	*Enterobacter hormaechei PY12*	0.28	[91]
Bacteria	*Reticulitermes chinensis Snyder*	*Bacillus licheniformis MX5*	0.26	[91]
Fungi	*Bulbitermes sp.*	*Aspergillus sp. A1*	729.12a	[74]
Fungi	*Odontotermes obesus*	*Aspergillus nomius (MN700028)*	10.39	[89]

a U/g dry substrate

2.11 Use of Termite Gut Microbiome and Enzyme - Based Systems for Biofuel Production

The remarkable ability of termites to efficiently degrade lignocellulosic biomass that often surpassing other herbivores such as ruminants that has long attracted scientific interest. This efficiency is largely due to their gut microsymbionts, which secrete a diverse array of enzymes that work synergistically to break down complex plant materials found in agricultural residues and pulp waste. Through this process, lignocellulose is converted into simple sugar units (Table 2.3), which can then be utilized by the microbes for metabolic activities. These metabolic pathways can directly yield biofuels or produce intermediate compounds, such as intracellular lipids, that serve as precursors for non-fermentative biofuels. Key biofuels associated with termite gut systems include bioethanol, biomethane, biodiesel, and biohydrogen.

2.11.1 Production of Bioethanol from Termite Gut Microsymbionts

Bioethanol is a widely used liquid biofuel produced through the fermentation of sugars. Due to its high oxygen content (∼35%) and compatibility with existing engines, it contributes to reduced greenhouse gas emissions and represents about 62% of global biofuel production [92]. Traditionally, Saccharomyces cerevisiae has been selected for ethanol fermentation. However, its limitations such as sensitivity to high temperatures, intolerance to high ethanol concentrations, and inability to ferment pentose sugars restrict its industrial efficiency.

Termite gut microsymbionts offer promising alternatives. Some microorganisms isolated from termite guts possess dual capabilities: they can both hydrolyze lignocellulose and ferment the resulting sugars. For example, *Candida pseudorhagii* from *Reticulitermes chinensis* can ferment xylose and produce ethanol with notable efficiency, while also exhibiting xylan-degrading activity. Similarly, *Candida*

tropicalis from *Coptotermes heimi* has demonstrated the ability to convert biomass hydrolysates into ethanol.

In addition to yeasts, certain bacteria such as *Spirochaeta coccoides* can ferment sugars into ethanol while producing enzymes involved in lignocellulose breakdown. More advanced approaches involve microbial consortia derived from termite guts. One such consortium, consisting of bacterial and yeast species, was capable of simultaneous biomass saccharification and ethanol production, achieving high enzyme activity and ethanol yields. These findings suggest that termite-derived microbial systems could overcome many limitations of conventional fermentation processes.

2.11.2 Production of Biomethane from Termite Gut Microsymbionts

Biomethane is produced by anaerobic digestion of organic materials and serves as an important renewable energy source, despite its role as a potent greenhouse gas. In termite guts, lignocellulose-derived sugars are fermented into intermediates such as carbon dioxide, hydrogen, and acetate. While acetate is used by the host, hydrogen and carbon dioxide are converted into methane by methanogenic archaea[78].

These methanogens are typically anaerobic and inhabit the hindgut, particularly the anoxic P3 compartment. Lower termites primarily host *Methanobrevibacter* species, whereas higher termites contain a broader diversity of methanogens, including members of *Methanosarcinales*, *Methanomicrobiales*, and *Methanomassiliicoccales*. In some lower termites, methanogens form symbiotic associations with flagellates, enhancing methane production efficiency.

The dominant pathway for methane formation in termite guts is hydrogenotrophic methanogenesis, where hydrogen and carbon dioxide serve as substrates. Although alternative pathways such as acetoclastic and methylotrophic methanogenesis have been observed, they are less common. The coordinated interaction

between fermenting microbes and methanogens enables efficient biomass conversion into methane. However, for industrial applications, competition for hydrogen among different microbial groups must be managed to optimize methane yields.

2.11.3 Production of Biohydrogen from Termite Gut Microsymbionts

Biohydrogen is considered a clean and efficient energy carrier due to its high energy density and environmentally friendly combustion, which produces only water. In termite guts, hydrogen is generated as an intermediate during the fermentation of polysaccharides into acetate and carbon dioxide. This process creates localized hydrogen-rich zones, particularly in the hindgut [85].

Numerous studies have identified hydrogen-producing anaerobic bacteria in termite guts. For instance, *Clostridium* species have demonstrated high hydrogen production rates alongside strong xylan-degrading activity. Other microbes, such as *Enterobacter* cloacae isolated from termite guts, have also shown significant hydrogen production potential. Measurements within termite hindguts have revealed elevated hydrogen concentrations, further supporting their role as efficient biological hydrogen reactors.

However, hydrogen production is often limited by competing processes within the gut. Hydrogen-consuming microbes, such as acetogens and methanogens, convert hydrogen into acetate and methane, reducing overall hydrogen yield. For industrial applications, suppressing these competing pathways and continuously removing hydrogen and carbon dioxide can significantly enhance production efficiency.

2.11.4 Production of Biodiesel from Termite Gut Microsymbionts

The process of making biodiesel depends on chemical reactions called the esterification of free fatty acids and the transesterification of triglycerides, which are

activated using specific catalysts. Traditionally, standard plant-derived vegetable oils have been the primary choice for this fuel due to their ready availability. Despite this widespread use, crop-based oils face major long-term sustainability issues because land crops grow slowly and require enormous amounts of farming acreage to produce sufficient yields.

As a result, fast-growing single-celled microbes have become a highly attractive alternative for fuel production. These tiny organisms multiply rapidly and can accumulate massive amounts of lipids within their cells, which are called single-cell oils (SCOs). Because these microbial fats are loaded with triglycerides and have a fatty acid composition that looks almost identical to traditional vegetable oils, they provide an ideal, highly sustainable raw material for producing high-quality biodiesel.

The termite gut hosts sebaceous yeasts capable of accumulating more than 20% of their dry cell weight as intracellular lipids, particularly under stress conditions such as excess carbon availability and nutrient limitation. These yeasts possess metabolic pathways that support SCO production, which can subsequently be utilized in biodiesel-generating reactions.

Although yeast presence in termite guts has long been recognized, their potential for SCO production and biodiesel applications remains underexplored. Recent research has identified multiple yeast strains from termite species, with some demonstrating significant lipid accumulation. For example, certain strains have shown lipid contents exceeding 20%, highlighting their potential as biodiesel feedstock. Among these, some strains exhibit particularly high lipid yields, suggesting strong industrial relevance.

Despite these promising findings, the commercial application of termite gut-derived yeasts for biodiesel production is still at an early stage. Further exploration is needed to identify more efficient strains, optimize lipid production, and evaluate the quality and performance of the resulting biodiesel under industrial conditions.

Interestingly, termites themselves store lipids in specialized fat bodies, which function as energy reserver during non-feeding periods. This indicates that termite

biomass could theoretically serve as a direct source of oils for biodiesel production. Some studies have demonstrated the feasibility of producing biodiesel from termite body oils with favorable properties such as low viscosity and efficient combustion.

However, using whole termites as a biodiesel source is impractical on an industrial scale. The enormous number of termites required, along with the ecological and ethical concerns associated with large-scale harvesting, makes this approach unsustainable. Such practices could disrupt ecosystems and negatively impact biodiversity.

Chapter 3

Methodology

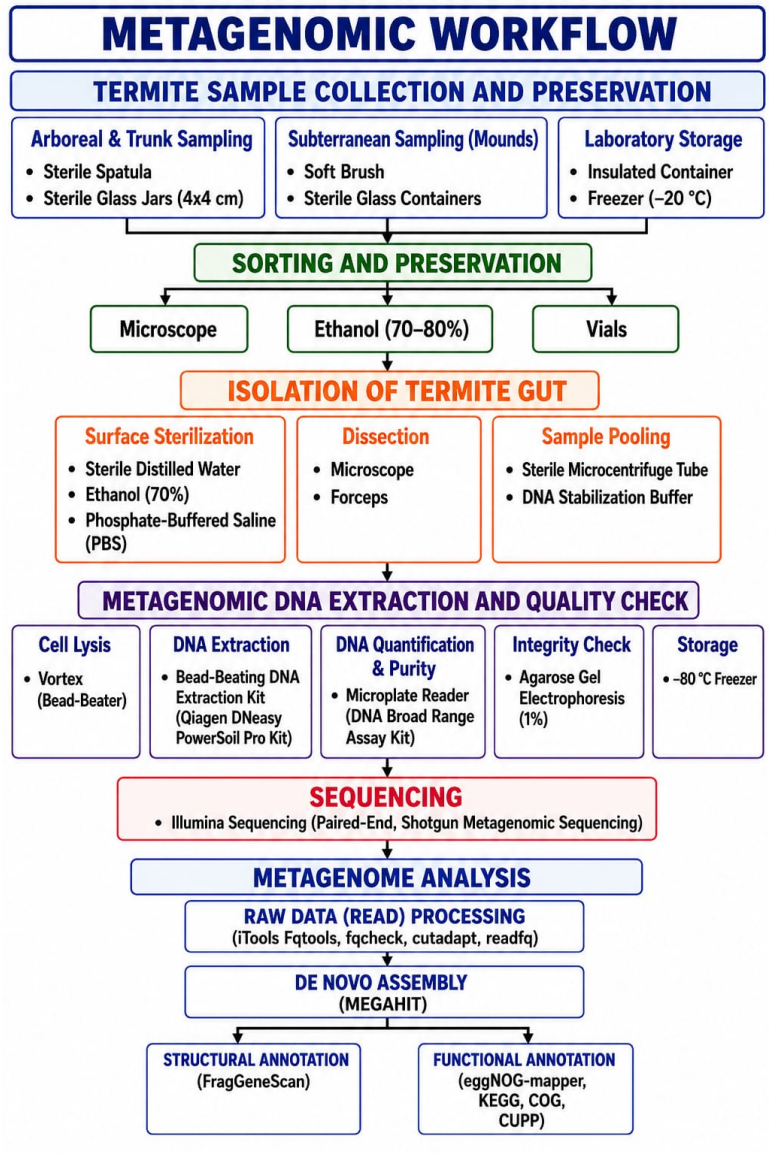

FIGURE 3.1: Workflow of Metagenomic analysis of Termite Gut

3.1 Termite Sample Collection and Preservation

Termite samples were collected from two locations within the Mardan district of KPK province of Pakistan. Sampling was based on collection from the nesting and foraging behaviors of the local termite populations.

3.1.1 Arboreal and Trunk Sampling

The bark of infested trees were scratched with the spatula to collect the foraging termites, and runway termites were collected directly into sterile, glass jars (4×4 cm) by placing the jars on the scratched infested area to avoid crushing the specimens.

Sampling of Subterranean termites from mounds was performed by gently digging sections of soil mounds, transferring active termites into separate sterile glass containers using soft brush.

3.1.2 Laboratory Storage

The jars with sample termites were sealed with lids to maintain, and kept inside the insulated containers, and immediately transported to the laboratory. The samples designated for gut microbiome analysis were transferred to a freezer maintained at -20 ^{o}C to preserve the structural integrity of the microbial DNA.

3.2 Sorting and Preservation

The composite samples were sorted under a microscope to isolate the soldier and worker caste, which possess the diagnostic morphometric traits required for species-level identification. Sorted specimens were preserved in vials that contain 70-80

3.3 Isolation of Termite Gut

A thorough surface-sterilization and dissection protocol was performed to isolate the gut microbiome without external environmental contamination.

3.3.1 Surface Sterilization

The active worker termites were selected for dissection. The termite specimens were first washed with sterile distilled water, later on immersed in 70% ethanol for 60 seconds to remove external surface contaminants, and after that the samples were rinsed three times in sterile phosphate-buffered saline (PBS).

3.3.2 Dissection

The dissections of termites were performed under a microscope under sterile conditions. The posterior end of the abdomen was gently and carefully pulled using forceps, to extract the intact intestinal tract.

3.3.3 Sample Pooling

The gut extracts from 30 to 50 worker termites per sampling were pooled into a sterile microcentrifuge tube containing DNA stabilization buffer to ensure a sufficient yield of microbial DNA.

3.4 Metagenomic DNA Extraction and Quality Check

For extracting total genomic DNA from the pooled termite gut samples, a commercially available bead-beating DNA extraction kit (e.g., Qiagen DNeasy PowerSoil Pro Kit) optimized for tough-to-lyse environmental and host samples was used.

3.4.1 Cell Lysis

The sample was vortexed for mechanical homogenization to efficiently disrupt both gram-positive and gram-negative bacterial cell walls, followed by chemical lysis.

3.4.2 DNA Quantification and Purity

The quantification of the extracted metagenomic DNA was performed by fluorometric quantification. The sample was analyzed using a Microplate Reader equipped with a DNA Broad Range (BR) assay kit.

The verification for DNA structural integrity and the absence of significant shearing was done by running the extracts on a 1% agarose gel electrophoresis system. The high-molecular-weight bands indicating high-integrity samples were preserved at -80℃ until they were dispatched for commercial Illumina sequencing.

3.5 Sequencing

The commercial sequencing services were utilized for Illumina sequencing technology, shotgun metagenomic sequencing of termite gut microbial DNA. To generate high-quality reads that are suitable for downstream metagenomic analysis, paired-end sequencing was performed. The sequencing output was obtained in FASTQ format containing forward (R1) and reverse (R2) reads.

3.6 Metagenome Analysis

Whole metagenome analysis refers to the identification and characterization of microbial communities, functional genes, and metabolic pathways present in the termite gut ecosystem through high-throughput sequencing technologies. The analysis involves multiple computational and bioinformatics approaches including

sequence quality assessment, host DNA removal, metagenome assembly, taxonomic profiling, gene prediction, and functional annotation.

3.6.1 Raw Data Filtering

3.6.1.1 iTools Fqtools

It acts as the primary data handling framework and used to analyze raw sequencing streams, manage file formats, and execute structured, multi-step filtering rules. iTools Fqtools allows for the execution of specific read-length thresholds post-truncation, without compromising computational efficiency.

3.6.1.2 fqcheck

A quality assessment tool used statistical profiles of raw and processed FASTQ files. It analyzes sequencing datasets for calculation of base composition distributions, detection of potential biases, and calculate cumulative Phred quality scores (Q-scores) across all read cycles.

This empirical data is required for sliding-window truncation and to confirm that the clean reads meet the criteria of targeted quality thresholds.

3.6.1.3 cutadapt

It inspects every read for user-defined nucleotide sequences to identify, trim, and discard synthetic adapter molecules, and primers. It regulates adapter trimming by enforcing a minimal 15-base overlap while accommodating up to three mismatches, ensuring that targeted sequences are completely eliminated without misidentifying true biological variations.

3.6.1.4 readfq

It is used for scanning sequence to detect structural anomalies, such as ambiguous base designations (N bases) or low-complexity homopolymer repeats (e.g., 10 consecutive identical bases). It ensures the quality of the sequence with biologically relevant sequence fragments.

3.6.2 Galaxy Platform

Galaxy Project is a web-based bioinformatics platform that enables the analysis of high-throughput sequencing data through an easy graphical interface without requiring advanced programming skills. The Galaxy platform integrates several metagenomic tools in a reproducible workflow system that allows researchers to process raw sequencing data into biologically meaningful results.

In this study, Galaxy was used for sequence quality assessment, metagenome assembly, taxonomic profiling, gene prediction, functional annotation, and diversity analysis of termite gut microbial communities.

3.6.2.1 FastQC

FastQC is widely used bioinformatics tool designed to evaluate the quality of raw sequencing reads generated through high-throughput sequencing technologies. It generates detailed reports containing information regarding per-base sequence quality, sequence duplication levels, adapter contamination, GC content distribution, and overrepresented sequences.

The quality assessment step is essential for detecting sequencing errors and technical biases in FASTQ files before downstream analyses such as assembly and annotation. High-quality reads improve the reliability and accuracy of metagenomic interpretation [92].

3.6.2.2 MEGAHIT

MEGAHIT is an ultra-fast metagenomic assembler designed for assembling short sequencing reads into longer contigs. It utilizes succinct de Bruijn graph algorithms to efficiently assemble large and complex metagenomic datasets.

In this study, MEGAHIT was used to reconstruct microbial genomic fragments from termite gut sequencing reads. The assembled contigs were further utilized for gene prediction and functional annotation [93].

3.6.2.3 FragGenScan

An advanced *ab initio* gene prediction tool for identifying protein-coding regions within short, fragmented, and error-prone sequencing reads typical of shotgun metagenomic datasets. It dynamically integrates context-specific codon usage statistics, start/stop codon frequencies, and platform-specific sequencing error models. FragGeneScan successfully identifies partial or truncated genes disrupted by read boundaries and accommodates sequencing frameshifts without relying on de novo genome assembly. In this pipeline, the software scans individual environmental reads to generate precise nucleotide (.ffn) and translated amino acid (.faa) outputs of putative coding sequences, ensuring the robust recovery of both known and novel functional genes from low-abundance organisms within the community [94] (https://doi.org/10.1093/nar/gkq747)

3.6.3 Taxonomic Profiling of the Termite Gut Microbiome

3.6.3.1 USEARCH

USEARCH (v7.0.1090) is used for downstream sequence processing, including the dereplication of unique sequence reads, high-precision de novo or reference-based chimera detection, and operational taxonomic unit (OTU) clustering based on user-defined identity thresholds (e.g., 97% or 99% similarity). By grouping homologous sequences, USEARCH establishes highly accurate representative sequences

essential for definitive taxonomic assignment and microbial diversity profiling. For 16S rDNA and ITS sequences, chimeras in OTU are screened and filtered by mapping to gold database(v20110519), UNITE(v20140703) respectively.

3.6.3.2 RDP Classifier

Taxonomic analysis of OTU representative sequences was carried out by RDP classifier Bayesian algorithm used to identify the composition of microbial structure. Abundances of species on seven levels (Phylum, Class, Order, Family, Genus, Species) was calculated after annotation.

3.6.3.3 Species GraPhlAn Map

GraPhlAn is a software tool used for producing high-quality circular representations of taxonomic and phylogenetic trees. GraPhlAn usually focuses on concise, integrative, and informative phylogenetically and taxonomically driven investigation.

Software: GraPhlAn (https://huttenhower.sph.harvard.edu/graphlan)

3.6.3.4 FastTree

FastTree (version 2.1.3) (http://www.microbesonline.org/fasttree/Species) is an open-source, high-throughput bioinformatics tool specifically optimized for handling massive sequence alignments based on maximum-likelihood (ML).

FastTree is uniquely suited for large-scale metagenomic datasets, enabling the robust phylogenetic placement of thousands of uncultured operational taxonomic units (OTUs) across complex evolutionary lineages.

3.6.4 Diversity Analysis QIIME

QIIME 2 (Quantitative Insights Into Microbial Ecology) is a next-generation, open-source bioinformatics platform for analyzing microbiome and metagenomic sequencing data. It allows researchers to take raw sequencing reads and perform quality filtering, taxonomic classification, and ecological diversity analyses (alpha and beta diversity). Featuring a modular plugin architecture and automatic data provenance tracking for strict reproducibility, QIIME 2 can be run via a command-line interface, Python APIs, or integrated into graphical platforms like Galaxy.

3.6.5 Functional Annotation Databases

3.6.5.1 eggNOG-mapper

It utilizes optimized search algorithms (such as HMMER or Diamond) to align query protein sequences to the evolutionary genealogy of genes: Non-supervised Orthologous Groups (eggNOG) database. By transferring functional annotations explicitly from precisely identified orthologs rather than close paralogs, the tool significantly increases annotation accuracy. In this pipeline, the software assigns comprehensive functional descriptors to predicted genes, mapping sequences directly to standardized Gene Ontology (GO) terms, Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic pathways, and Clusters of Orthologous Groups (COG) functional categories, thereby providing a robust biological overview of the metabolic potential within the studied community. eggNOG-mapper was used for orthologous gene classification, Gene Ontology assignment, and functional prediction of microbial proteins [95].

3.6.5.2 Kyoto Encyclopedia of Genes and Genomes Database

Functional interpretation of the predicted metagenomic content was performed using the KEGG database hierarchy. Predicted genes were mapped to KEGG

Orthology (KO) identifiers to reconstruct higher-order molecular interaction networks. By organizing individual KOs into standardized metabolic pathways, modules, and BRITE functional hierarchies, this analysis provides a macro-level overview of the community's metabolic architecture.

3.6.5.3 Clusters of Orthologous Groups

The COG database is a phylogenetic system that classifies proteins from sequenced genomes based on evolutionary relationships. It groups orthologs genes in different species that evolved from a common ancestor, which typically share the same function. This allows researchers to annotate new sequences, predict gene functions, and compare functional categories (like metabolism or cell division) across diverse organisms.

3.6.5.4 CUPP

Conserved Unique Peptide Patterns (CUPP) is alignment - free bioinformatic method designed for the rapid clustering and precise functional annotation of Carbohydrate-Active Enzymes (CAZymes). It works on the principle of peptide pattern recognition, by scanning query proteins for short, evolutionarily conserved amino acid motifs and evaluating the exact spacing between adjacent conserved residues.

By mapping sequence entries directly to highly resolved motif models, the platform facilitates automated, rapid genome-wide annotation of CAZyme designations and precise Enzyme Commission (EC) numbers, offering enhanced accuracy in distinguishing distinct substrate specificities within complex environmental metagenomes.

Carbohydrate-Active Enzymes were identified that were involved in lignocellulose degradation including cellulases, hemicellulases, and glycoside hydrolases [96].

3.6.5.5 PICRUSt2

Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt2) was utilized. This computational framework overcomes the limitations of standard marker-gene surveys by phylogenetically placing study sequences (16S rRNA) into a reference tree containing tens of thousands of fully sequenced genomes. By leveraging hidden state prediction (HSP) algorithms, PICRUSt2 accurately estimates the copy number of unobserved gene families—such as Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthologs (KOs) and Enzyme Commission (EC) numbers for each taxon. These predicted gene abundances are subsequently integrated with sample-specific community profiles to reconstruct higher-level metabolic pathways, typically mapped against the MetaCyc or KEGG databases.

3.6.5.6 MetaCyc Database

Downstream annotations were performed against the MetaCyc database to predict and map the functional metabolic potential of the identified microbial community. MetaCyc a database of experimentally revealed, reactions, metabolic pathways, enzymes, and substrate interactions across all domains of life. By aligning the predicted open reading frames (ORFs) against this resource through computational workflows like Humann or MinPath, individual microbial genes were clustered into complete, biologically meaningful metabolic networks. This database framework enabled the reconstruction of complex primary and secondary metabolic pathways, such as carbohydrate catabolism, nitrogen cycling, and xenobiotic degradation, providing a direct functional link between the taxonomic profile of the microbiome and its active metabolic niche.

Chapter 4

Results

4.1 Metagenomic DNA Quality Check

The extraction of genomic DNA yielded purified and intact DNA templates from all processed samples. The fluorometric verification using a Microplate Reader provided with a DNA Broad Range (BR) assay kit determined the concentration of the sample to be 3.2 ng/μL. A final volume of 16μL containing a total mass of 0.05 μg (50ng) of genomic DNA was utilized for downstream application.

DNA integrity was assed with Agarose gel electrophoresis (1% w/v). The gel revealed well-defined, and high-molecular-weight bands for all samples, matching the expected migration pattern relative to the molecular weight DNA ladder (Figure 4.1). The absence of smearing or low-molecular-weight RNA bands confirms that the extracted genomic DNA was intact, undegraded, and qualified for downstream PCR amplification and library construction (Figure 4.1).

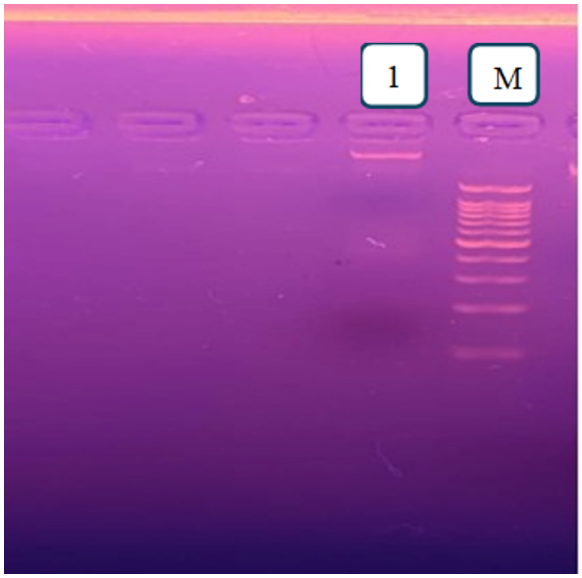

FIGURE 4.1: Agarose gel electrophoresis (1% w/v) assessing the structural integrity of isolated genomic DNA. Lane M: ladder (kb); Lane 1: Representative genomic DNA samples.

4.2 Metagenome Analysis

4.2.1 Raw Data Filtering

Quality control (QC) and filtering pipeline were used to ensure downstream analytical accuracy by subjecting the Raw sequencing reads to generate clean reads. The data filtering protocol was applied using iTools Fqtools, fqcheck (v0.25), cutadapt (v2.6), and readfq (v1.0) according to the following criteria:

i. Quality-Based Truncation was performed where the reads at the terminal end with an average Phred quality score dropped below 20. Following truncation, any read retaining less than 75 of its original length was entirely discarded from the dataset. A sliding window approach of 25 base pairs (bp) was used to evaluate the reads.

ii. Adapter Contamination Removal: Reads with a minimum overlap of 15 bases with known adapters were removed, with a maximum threshold of 3 base mismatches [97].

iii. Ambiguous Base Elimination: All reads containing ambiguous or undetermined bases (designated as "N") were filtered out.

4.3 Quality Assessment of Raw Reads

Raw sequencing data were evaluated for quality using FastQC (Galaxy). The analysis was performed to evaluate read quality prior to downstream processing. The key findings are summarized (Table: 4.1)

TABLE 4.1: Basic Statistics of the Sequence Quality Check

Parameters	**Value**
File Name	Concatenate datasets
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	136742
Total Bases	37.5 Mbp
Sequences flagged as poor quality	0
Sequence length	190-297
%GC	53

The dataset contained a total of 136,742 sequences, representing 37.5 Megabase pairs (Mbp) of sequence data. No sequences were flagged as poor quality, which indicates that the data is suitable for further analysis. Sequence lengths ranged from 190 to 297 bp, and the overall GC content was 53%, composition of mixed bacterial communities typically found in termite gut metagenomes and can be expected depending on the microbial community or genome being studied.

4.3.1 Per Base Sequence Quality

The per-base sequence quality analysis revealed that the data cleared FastQC's quality threshold check. The mean Phred quality score across all read positions

ranged from Q35.4 to Q41.9, with an overall average quality of Q39.8 across the entire read length. These values remained consistently and substantially above both the Q20 threshold (≥99% base-calling accuracy) and the more stringent Q30 threshold (≥99.9% base-calling accuracy) throughout all positions, including toward the 3' end of the reads.

Quality scores were consistently high across the majority of read positions, with mean Phred scores well above Q28 throughout most of the read length, showing high quality of data (Figure 4.2).

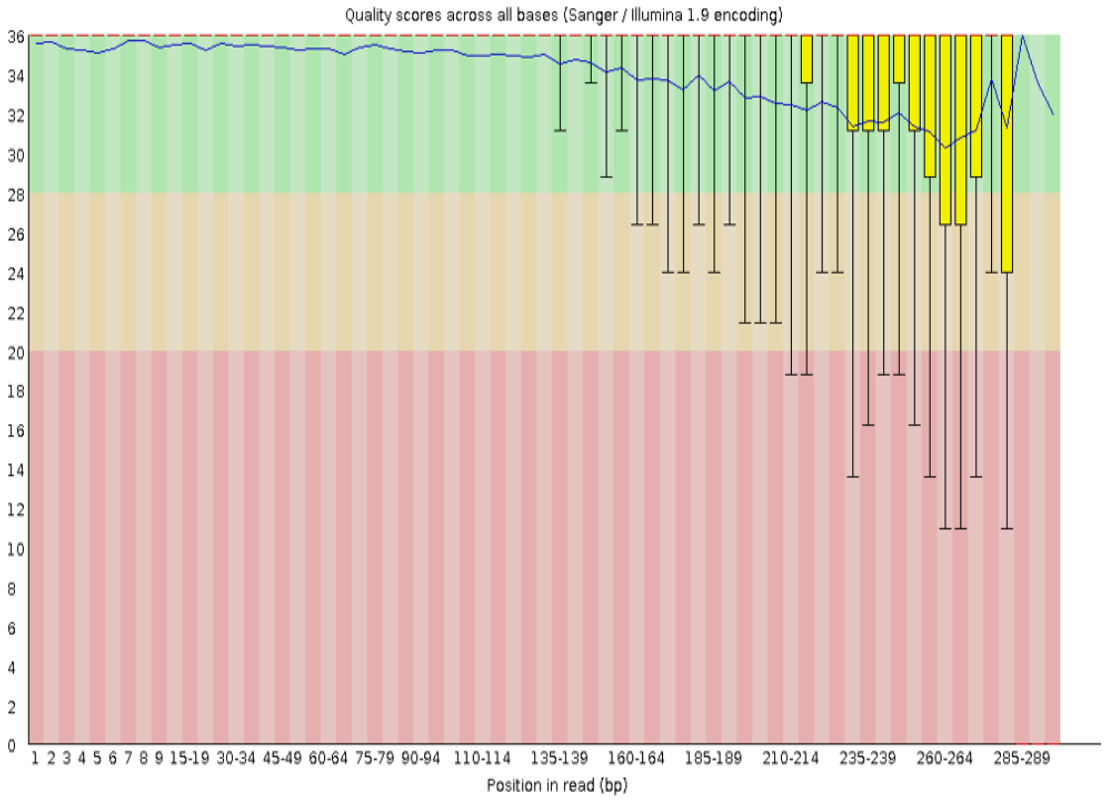

FIGURE 4.2: Per base sequence quality score across all bases.

4.3.2 Per Base N Content

The per base N content confirm negligible ambiguous base calls and a uniform read length distribution at 300 bp across all read positions, which confirms the quality of the sequence generated by modern Illumina platforms, where the base caller can reliably assign nucleotide identities even in lower-quality regions (Figure 4.3).

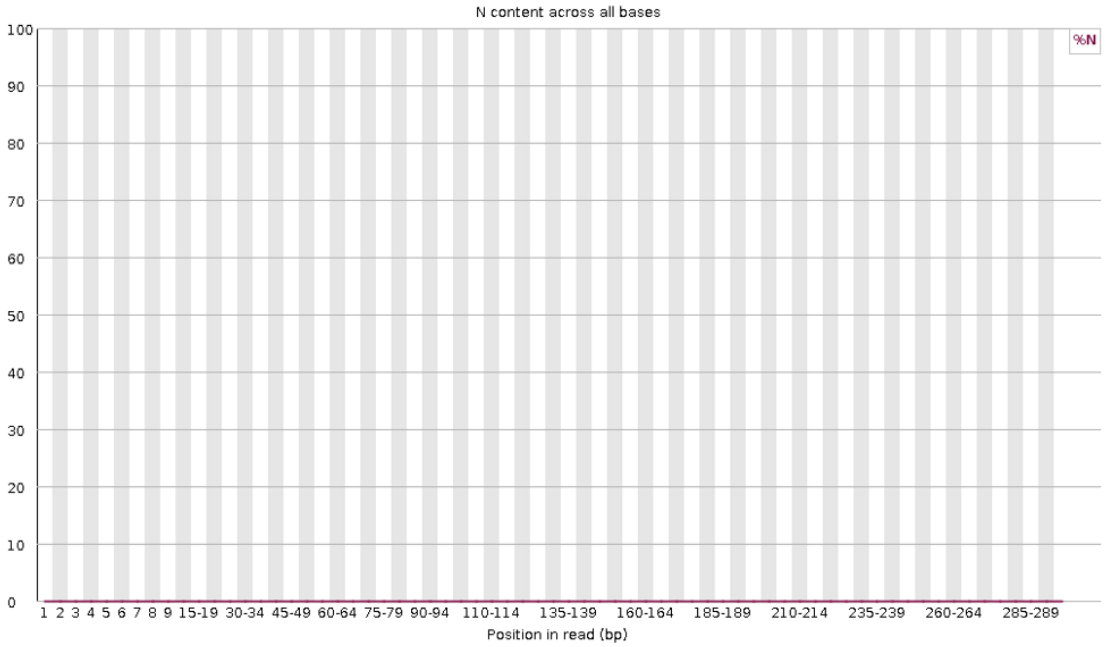

FIGURE 4.3: Per base N contents across all bases.

4.3.3 Adapter Contents

Figure 4.4 shows that no adapter contamination was detected in the dataset. This suggests that the sequenced reads do not require adapter trimming prior to downstream analysis.

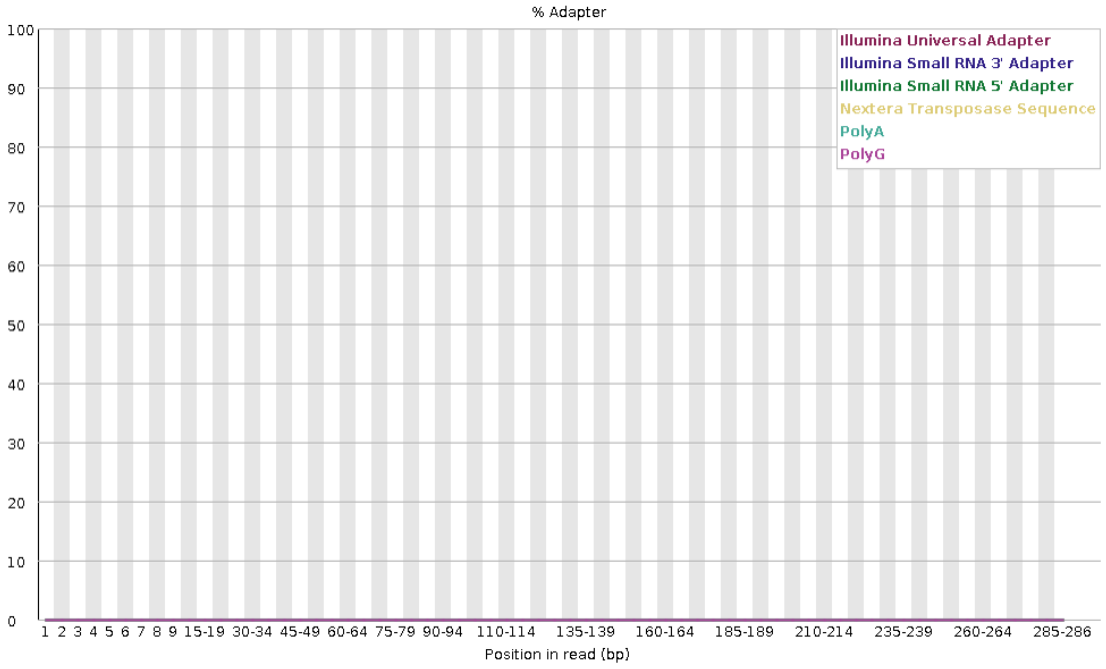

FIGURE 4.4: Adapter Contents for the Sequence for FastQC.

4.4 Metagenome *De Novo* Assembly

The metagenome de novo assembly produced by using MEGAHIT revealed 10 contigs with a total assembly size of 3,881 base pairs (bp) with the contig lengths ranging from 318 to 430 bp. The assembly statistics are summarized in table 4.2:

TABLE 4.2: The Summary of assembly statistics produced by MEGAHIT

Parameter	**Value**
Total Number of Contigs	10
Total Assembly Size	3,881 bp
Minimum Contig Length	318 bp
Maximum Contig Length	430 bp
Mean Contig Length	388.1 bp
N50	406 bp
k-mer size used	k99
Assembler	MEGAHIT (Galaxy)

The *de novo* assembly yielded an N50 value of 406 bp, meaning that half of the total assembled sequence length is contained within contigs of 406 bp or longer. This low N50 value indicates a highly fragmented assembly. Such fragmentation is typical for diverse environmental metagenomes, where low-abundance species are difficult to fully assemble due to incomplete sequence coverage.

4.4.1 Per-Contig Assembly Statistics

Table 4.3 shows the individual contig-level statistics, which include length, sequencing coverage, and GC content. These parameters are crucial for differentiating microbial taxa, identifying high-abundance organisms, and assessing assembly quality. Two contigs sequence shows highest coverage in a dataset, k99_6 with Coverage 777x, Length: 418 bp, GC: 54.1%. whereas k99_3 with Coverage: 479x, Length: 404 bp, GC: 47.5% is the second most abundant contig but with lower GC-content of 47.5%. The high revealed that they are obtained from the dominant organism in the termite gut microbiome.

TABLE 4.3: Per-Contig Assembly Statistics

Contig ID	**Length (bp)**	**Coverage (x)**	**GC Content (%)**
k99_0	341	4	57.8
k99_8	386	2	54.4
k99_7	407	13	51.4
k99_1	342	1	55.6
k99_5	429	1	52.4
k99_9	430	1	56
k99_2	406	1	57.6
k99_6	418	777	54.1
k99_4	318	1	54.1
k99_3	404	479	47.5

The moderate coverage is shown by the contigs k99_7 (coverage: 13x, length: 407 bp, GC: 51.4%), k99_0 (coverage: 4x, length: 341 bp, GC: 57.8%) and k99_8 (coverage: 2x,length: 386 bp, GC: 54.4%). These contigs represent organisms with relatively low abundance.

While the remaining five contigs (k99_1, k99_2, k99_4, k99_5, k99_9) show a coverage of 1x, with GC contents 55.6%, 52.4%, 56%, 57.6% and 54.1% respectively represent rare taxa within the community of termite gut microbiome.

4.5 Gene Prediction

Out of 136,742 input sequences, 116,517 (85.21%) yielded predicted genes, whereas 20,225 (14.79%) produced no prediction. A total of 127,623 genes were predicted across the dataset (Table 4.4).

TABLE 4.4: Summary of FragGenScan statistics for gene prediction

Parameters	**Value**
Total input sequences	136,742
Sequences with predicted genes	116,517 (85.21%)
Sequences without predicted genes	20,225 (14.79%)

Table 4.4 continued from previous page

Parameters	**Value**
Total predicted genes	127,623
Average genes per sequence (gene-containing)	1.1
Maximum genes on a single sequence	4

4.5.1 Gene Length

Gene lengths (nucleotides) ranged from a minimum of 67 bp to a maximum of 281 bp (Table 4.5). The length of all predicted genes was less than 300 bp threshold confirms that these are short-read sequences.

TABLE 4.5: Summary of the Length distribution and percentage allocation of predicted genes from FragGeneScan

Length Category	**Count**	**Percentage**
<200 bp	63,687	49.90%
200 - 281 bp	63,936	50.10%
>300 bp	0	0.00%

The distribution and structural metrics of the coding regions identified by FragGeneScan, were visualized using a high-density heatmap (Figure 4.5). The coordinate profile demonstrated a high level of uniformity in data density across all evaluated open reading frames (ORFs) on the Y-axis.

The noticeable color differentiation between Column 2 and Column 3 reveals the predicted open reading frames (ORFs). Column 2 is dominated by shades of blue, which represent the initial upstream 5' start regions of the genes where the relative nucleotide count is low. In contrast, Column 3 transitions into a dense, deep red profile, representing the downstream 3' terminal ends where the cumulative nucleotide length of the predicted genes peaks toward the maximum threshold (>250 bases). Furthermore, Column 6 presents solid dark blue stripe across every single row indicates the highly structurally clean dataset.

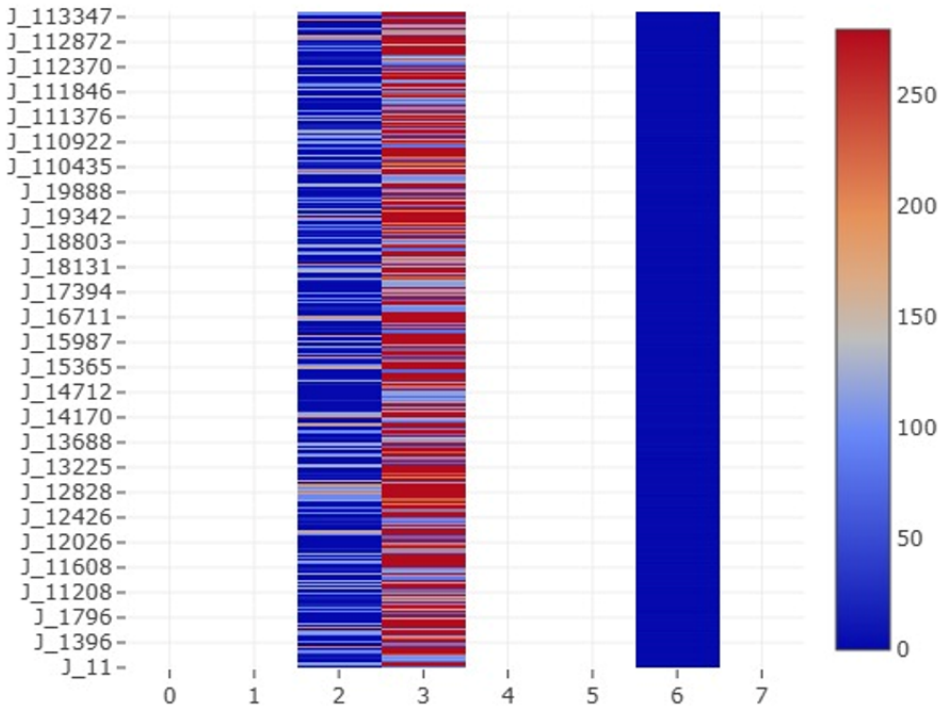

FIGURE 4.5: Heatmap characterization of coding sequence boundaries across predicted genes. (Y-axis represent individual predicted genes, open reading frames (ORFs) and X-Axis columns correspond directly to the data fields containing features such as gene start positions, end positions, strand direction, reading frame (1, 2, or 3), and prediction scores.

4.5.2 Per Sequence Gene Density

The table 4.6 shows the high proportion of sequences carrying multiple genes i.e., 44.2%. The top gene-dense contigs include J_12057, J_13639, J_14516, J_15666, and J_15945, each carrying four predicted genes.

TABLE 4.6: The breakdown of gene distribution between single- and multi-gene sequences

Category	**Count**	**Percentage**
Sequences with 1 gene	14,527	12.50%
Sequences with 2+ genes	51,474	44.20%
Maximum genes on one sequence	4	-

4.5.3 Protein Length

All 127,623 predicted proteins range between 21 and 93 amino acids in length, which is entirely expected when gene prediction is performed on short sequencing reads (Table 4.7). Moreover, 100% of the predicted proteins fall under the 100 amino acid threshold (127,623 peptides), reflecting a population dominated by truncated or partial open reading frames.

TABLE 4.7: Summary statistics of translated protein lengths predicted by FragGeneScan

Parameters	**Value**
Minimum protein length	21 amino acids
Maximum protein length	93 amino acids
Mean protein length	62.9 amino acids
Median protein length	66 amino acids
Proteins <100 aa	127,623 (100%)

4.5.4 Orientation of Strands

Of the 127,623 predicted genes, 94,427 (74.0%) are on the forward strand and 33,196 (26.0%) are on the reverse strand (Table 4.8). The columns no 6 of heatmap (figure 4.5) with solid dark blue for every single gene with a uniform matrix column indicates the similar trends showing the grouping of a massive block of these genes by their shared forward orientation.

TABLE 4.8: Distribution and relative abundance of predicted genes across coding strands.

Strand	**Gene Count**	**Percentage**
Forward (+)	94,427	74.00%
Reverse (-)	33,196	26.00%

4.5.5 Open Reading Frames

The Table 4.9 shows the distribution of the reading frames with strong biological association toward Frame 1, which is the dominant reading frame containing 59.6% of the total predicted open reading frames (ORFs). The remaining sequences are distributed nearly equally between Frame 0 (21.2%) and Frame 2 (19.2%).

TABLE 4.9: Distribution and frequency of predicted genes across reading frames

Reading Frame	**Gene Count**	**Percentage**
Frame 0	27,104	21.20%
Frame 1 (dominant)	76,023	59.60%
Frame 2	24,496	19.20%

4.6 Taxonomic Profiling of the Termite Gut Microbiome

4.6.1 Operational Taxonomic Units Clustering

To quantify the abundance of bacteria at every level in each sample, it is necessary to cluster the sequences into OTU with 97% similarity, which acts as a standard for taxa across the seven taxonomic ranks.

The data was clustered into Operational Taxonomic Units (OTUs) to characterize the taxonomic diversity and relative abundance of the bacterial community (Table 4.10).

All recovered OTUs were assigned to the domain Bacteria, revealing a community dominated by the phylum Bacillota (formerly Firmicutes), with lower-abundance represented by phyla Bacteroidota, Spirochaetota, Cyanobacteriota, and Acidobacteriota. The abundance distribution revealed highly dominant bacterial taxa within the sample that is Otu522 which shows highest abundance in the entire dataset with 70 reads. This dominant OTU was tracked down to the

genus level and classified as Acetivibrio within the family Oscillospiraceae (Order Eubacteriales, Class Clostridia).

TABLE 4.10: Taxonomic abundance of major bacterial OTUs

#OTUId	**Abundance**	**Taxonomy**
Otu573	2	Bacteria;Bacillota ; Clostridia ; Eubacteriales ; Oscillospiraceae
Otu486	5	Bacteria ; Bacillota
Otu487	3	Bacteria ; Bacillota ; Clostridia ; Eubacteriales ; Oscillospiraceae
Otu488	2	Bacteria ; Cyanobacteriota ; Cyanophyceae
Otu489	2	Bacteria ; Bacillota ; Clostridia ; Eubacteriales ; Oscillospiraceae
Otu510	3	Bacteria
Otu529	4	Bacteria ; Bacillota ; Clostridia ; Eubacteriales ; Oscillospiraceae
Otu528	3	Bacteria ; Bacillota ; Clostridia ; Eubacteriales ; Oscillospiraceae ; Monoglobus ; Monoglobus_pectinilyticus
Otu523	8	Bacteria ; Bacteroidota ; Bacteroidia ; Bacteroidales
Otu522	70	Bacteria ; Bacillota ; Clostridia ; Eubacteriales ; Oscillospiraceae ; Acetivibrio
Otu521	8	Bacteria ; Bacillota ; Clostridia ; Eubacteriales ; Oscillospiraceae ; Lawsonibacter ; Lawsonibacter_asaccharolyticus
Otu520	5	Bacteria ; Bacillota ; Clostridia ; Eubacteriales ; Oscillospiraceae
Otu527	3	Bacteria ; Bacillota ; Clostridia ; Eubacteriales ; Oscillospiraceae
Otu526	4	Bacteria ; Bacillota ; Clostridia ; Eubacteriales ; Christensenellaceae ; Gehongia ; Gehongia_tenuis
Otu525	9	Bacteria ; Spirochaetota ; Spirochaetia ; Spirochaetales ; Treponemataceae ; Treponema
Otu524	2	Bacteria ; CandidatusSaccharibacteria ; Unclassified ; Unclassified ; Unclassified ; Saccharibacteria
Otu318	9	Bacteria
Otu319	7	Bacteria ; Bacillota ; Clostridia ; Eubacteriales
Otu310	13	Bacteria ; Acidobacteriota ; Acidobacteria_Gp16 ; Unclassified ; Unclassified ; Gp16
Otu311	13	Bacteria ; Bacillota ; Negativicutes

Other moderately abundant sequences included Otu310 (13 reads), assigned to the phylum Acidobacteriota (Gp16), and Otu311 (13 reads), classified within the class Negativicutes of the phylum Bacillota. The rest of the core community

consisted of a diverse group of low-abundance OTUs belonging to the family Oscillospiraceae. However, Otu510, Otu318, and Otu486 were only tracked to the kingdom or phylum level. Moreover, several specialized anaerobic and hydrolytic taxa down to the genus and species levels were identified (Otu525 identified as Treponema, Otu521 identified as *Lawsonibacter asaccharolyticus*, Otu526 as *Gehongia tenuis* and Otu528 as *Monoglobus pectinilyticus*).

4.6.2 OTU Rank Curve

OTU rank curve are used to visualize the diversity of the species in the samples. The relative abundance of each OTU is calculated, and species were ranked by their relative abundance in descending manner. The graph (Figure 4.5) below shows the OTU rank on X-axis and relative abundance of OTU on Y-axis. Species richness can be seen as the number of differential species. It is visible that the wider the curve, the richer the species. Whereas slope in the graph tell about the species evenness and the flat curve indicates high evenness of species in given sample.

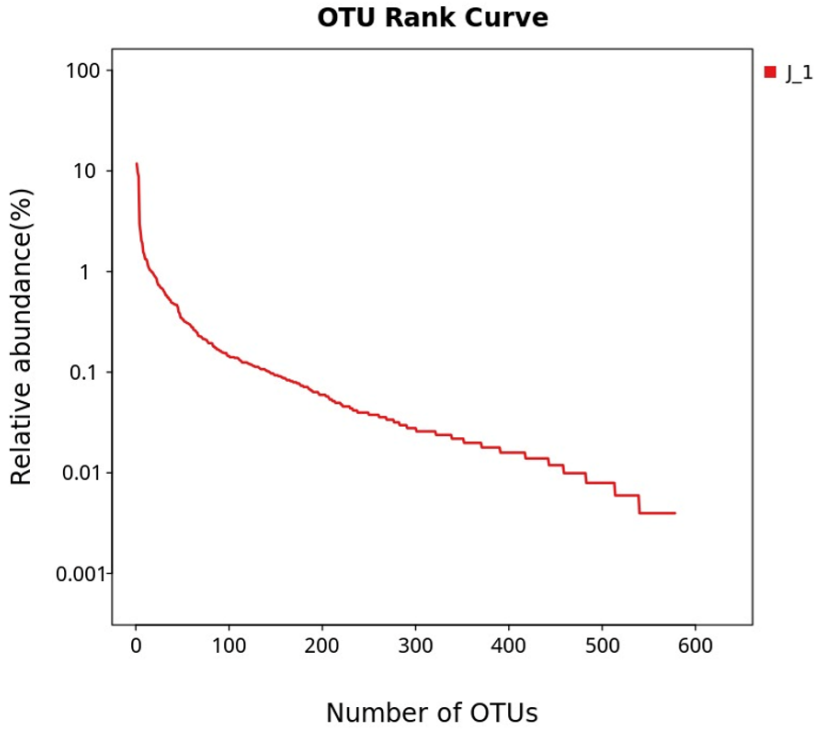

FIGURE 4.6: OTU rank-abundance curve of the microbial community

Curve Slope referring to community evenness revealed a rapid drop at the initial high-rank (ranks 1 to 50), where the relative abundance drops sharply from approximately 11% down to 0.5%. This steep vertical drop indicates low evenness among the top taxa, reflecting that a small, highly specialized group of dominant bacterial species occupies the community.

The horizontal length of the curve indicates the community richness. The curve shifts into a prolonged slop with horizontal extension that crosses the rank 500, dropping down to a relative abundance threshold of 0.004%. This extensive long tail represents an enormous fraction of low-abundance operational taxonomic units, showing high overall species richness within the sample.

The stair-like pattern at the end of the curve (ranks 300–500+) refers to single-count or low-count sequences that have low relative abundance.

4.6.3 Taxonomic Composition at Different Levels

The species abundance barplot shows the composition and proportion of species in each sample/group. Species were classified into 'others' if their relative abundance was less than 0.5%.

4.6.3.1 Phylum

The bar plot (Figure 4.7) shows relative abundance at phylum-level which represents Bacillota (∼45%) as the dominant phylum, followed by Pseudomonadota (∼15%) and Spirochaetota (∼13%). Bacteroidota, Fibrobacterota, Campylobacterota, Acidobacteriota, Actinomycetota, Candidatus Saccharibacteria, SR1, Elusimicrobiota, Deferribacterota, Cyanobacteriota, Mycoplasmatota, Planctomycetota, Fusobacteriota, Verrucomicrobiota, and Chloroflexota make up the remaining diversity in progressively smaller proportions, with "Other" and Synergistota contributing modest shares (∼6–8% combined).

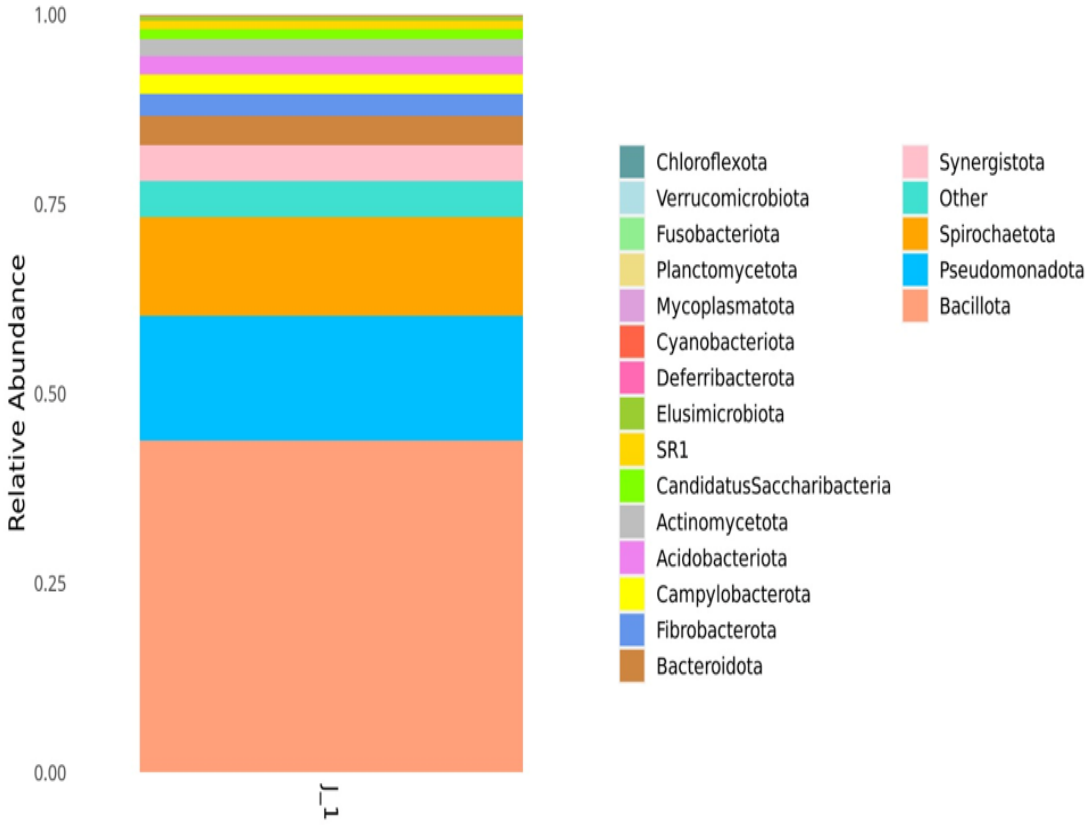

FIGURE 4.7: Relative abundance of bacterial taxa at the Phylum level in the termite gut microbiome

4.6.3.2 Class

The class Clostridia (∼32%) was the dominat in the sample followed by Spirochaetia (∼13%), "Other" classes (∼13%), Bacilli (∼11%), and Alphaproteobacteria (∼11%). The remaining classes include the Synergistia, Deltaproteobacteria, Campylobacteria, Betaproteobacteria, Actinobacteria, Acidobacteria_Gp18, Fibrobacteria, Chitinispirillia, Holophagae, Endomicrobiia, and Coriobacteriia each contribute smaller proportions (<5%) (Figure 4.8).

The dominance of Clostridia and Spirochaetia is consistent with the expected fiber - fermenting and symbiotic gut flora typical of wood-feeding termites.

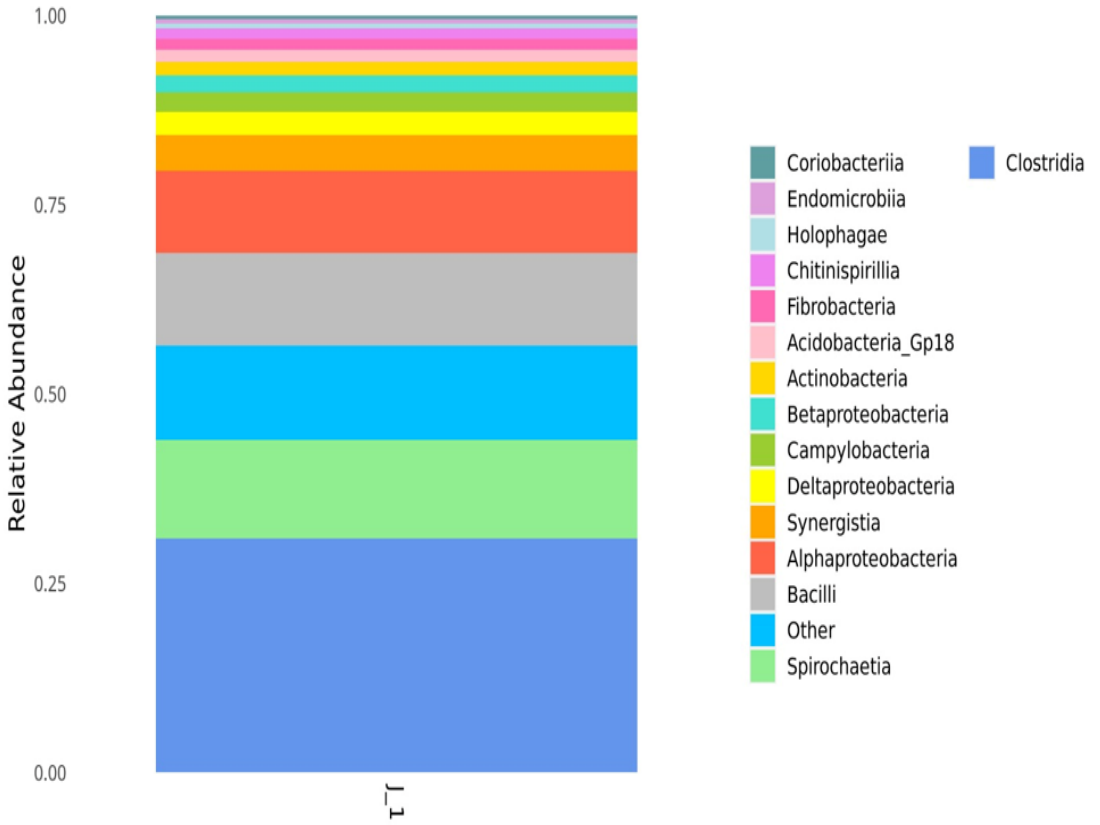

FIGURE 4.8: Relative abundance of bacterial taxa at the Class level in the termite gut microbiome

4.6.3.3 Order

The bar plot (Figure 4.9) shows relative abundance at order-level, with Eubacteriales as the single most abundant order (∼32%) which is followed by "Other" orders (∼17%) and Spirochaetales (∼15%) while both as together are accounting for the majority of the community.

Caryophanales (∼12%) and Rickettsiales (∼10%) form moderate contributions, while Synergistales, Campylobacterales, Lactobacillales, Desulfobacterales, Fibrobacterales, Chitinispirillales, Neisseriales, Desulfovibrionales, Micrococcales, Holophagales, Streptomycetales, Endomicrobiales, and Eggerthellales each represent minor fractions.

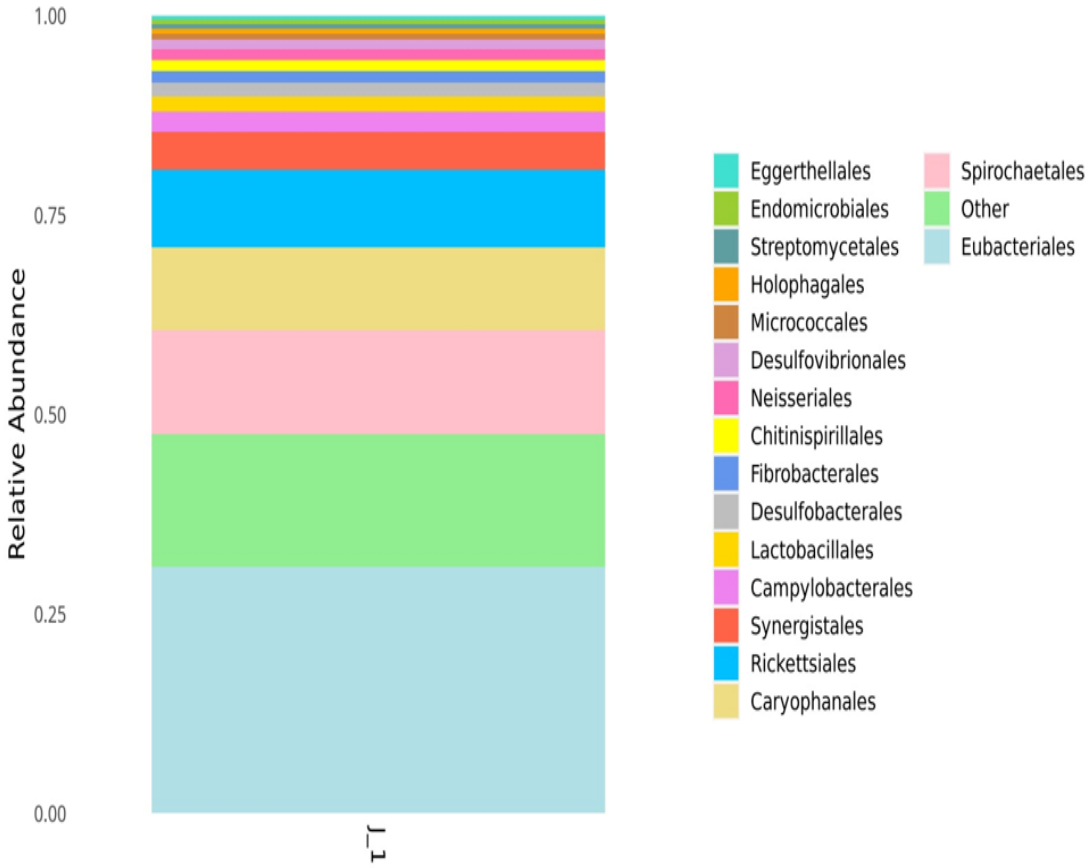

FIGURE 4.9: Relative abundance of bacterial taxa at the Order level in the termite gut microbiome

4.6.3.4 Family

The relative abundance at Family level revealed that Other category (∼39%) represents high family-level diversity. Treponemataceae (∼12%), Anaplasmataceae (∼11%), Bacillaceae (∼12%), and Oscillospiraceae (∼10%) were the next abundant families followed by smaller contributions from Synergistaceae, Sulfurospirillaceae, Desulfobacteraceae, Eubacteriales, Paenibacillaceae, Fibrobacteraceae, Streptococcaceae, Chitinispirillaceae, Neisseriaceae, Lachnospiraceae, Desulfovibrionaceae, Holophagaceae, Streptomycetaceae, Endomicrobiaceae, and Eggerthellaceae (Figure 4.10).

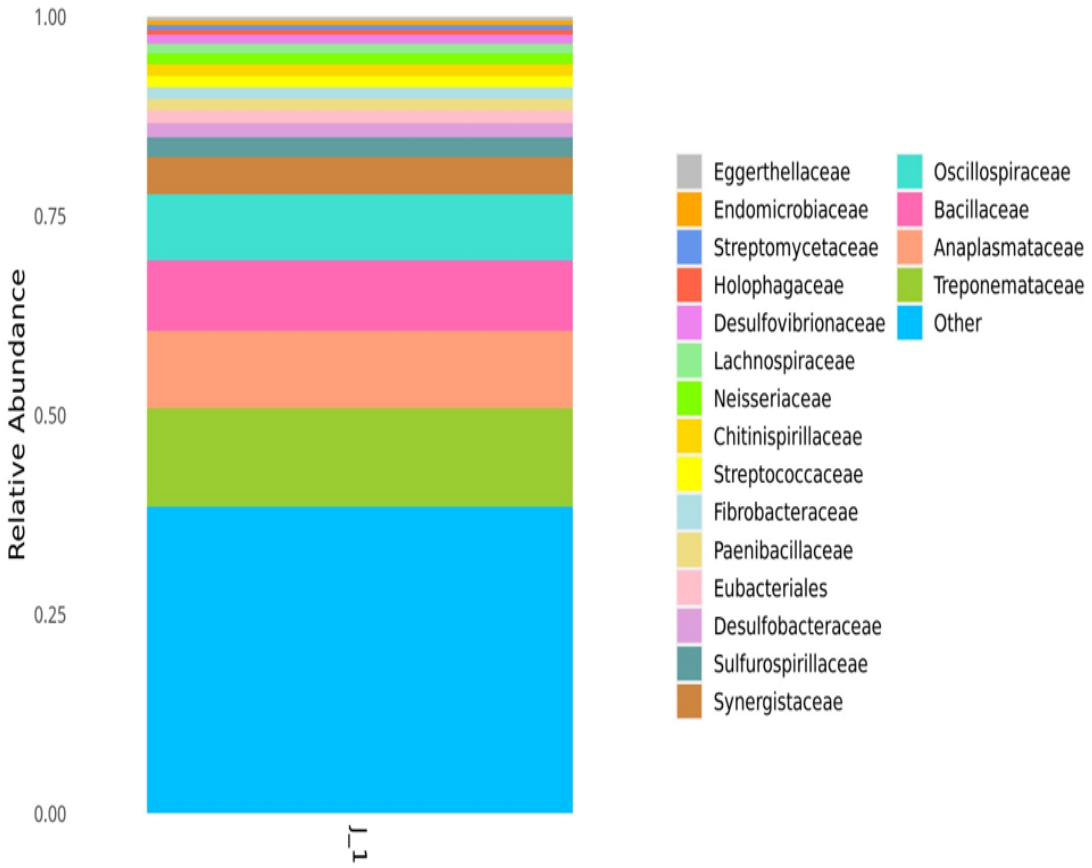

FIGURE 4.10: Relative abundance of bacterial taxa at the Family level in the termite gut microbiome.

4.6.3.5 Genus

A significant proportion of the total sequences (50%) was assigned to the "Other" category, reflecting a highly diverse community structure comprised of low - abundance taxa (Figure 4.11). Among the other genera, Treponema (12%) and Ehrlichia (11%) emerged as the dominant taxonomic groups within the microbiome. These were followed by the moderately abundant genera Bacillus (9%) and Sulfurospirillum (3%). The remaining community consisted of a diverse low - abundance genera, including Fretibacterium, Gp18, Paenibacillus, Lactococcus, Chitinispirillum, Saccharibacteria, SR1, Fibrobacter, Aminipila, Acetivibrio, Cupidesulfovibrio, Cloacibacillus, Geothrix, Streptomyces, and Endomicrobium.

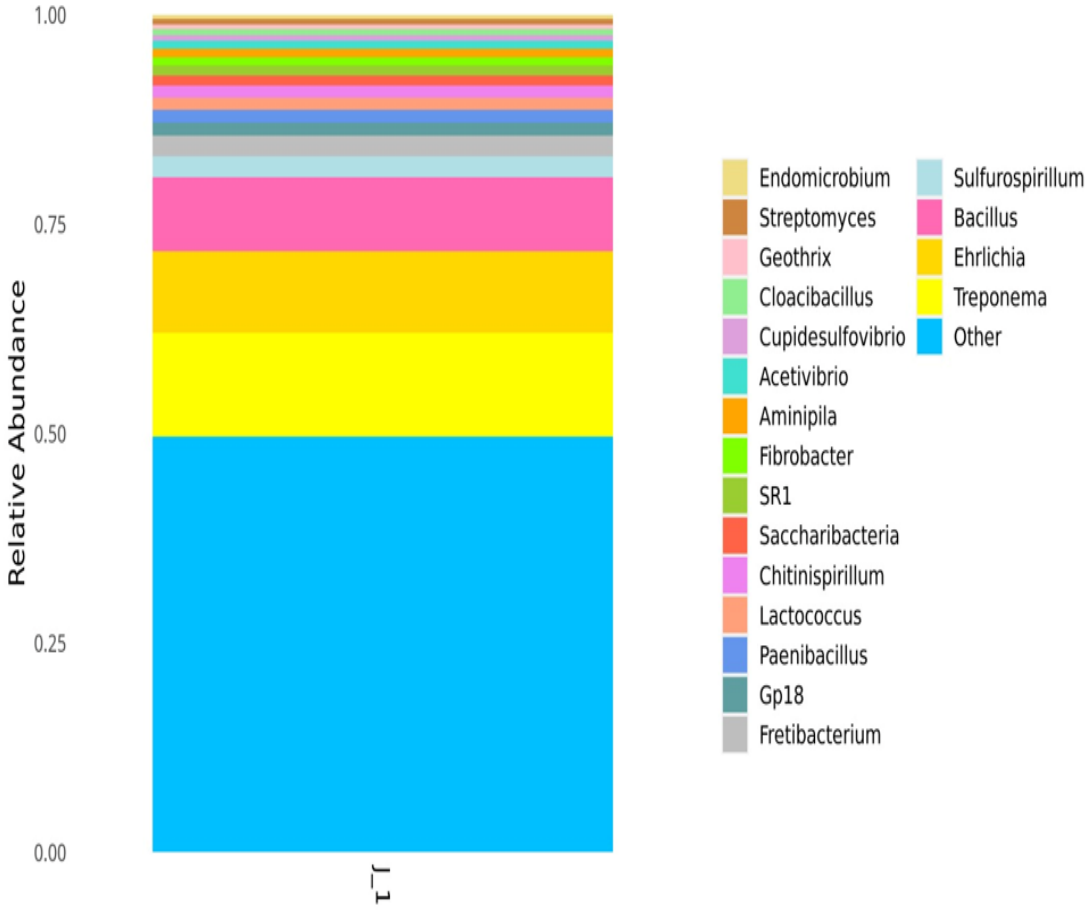

FIGURE 4.11: Relative abundance of bacterial taxa at the Genus level in the termite gut microbiome.

4.6.3.6 Species

The majority of the sequence reads (69%) were grouped into the "Other" category. The other resolved taxa, *Ehrlichia ewingii* (9%) and *Bacillus cereus* (8%) emerged as the co-dominant species within the community profile. The smaller, low-frequency contributions included the *Treponema isoptericolens, Fretibacterium fastidiosum, Paenibacillus popilliae, Chitinispirillum alkaliphilum, Fibrobacter intestinalis, Aminipila terrae, Treponema primitia, Geothrix fermentans,* and *Endomicrobium proavitum* (Figure 4.12).

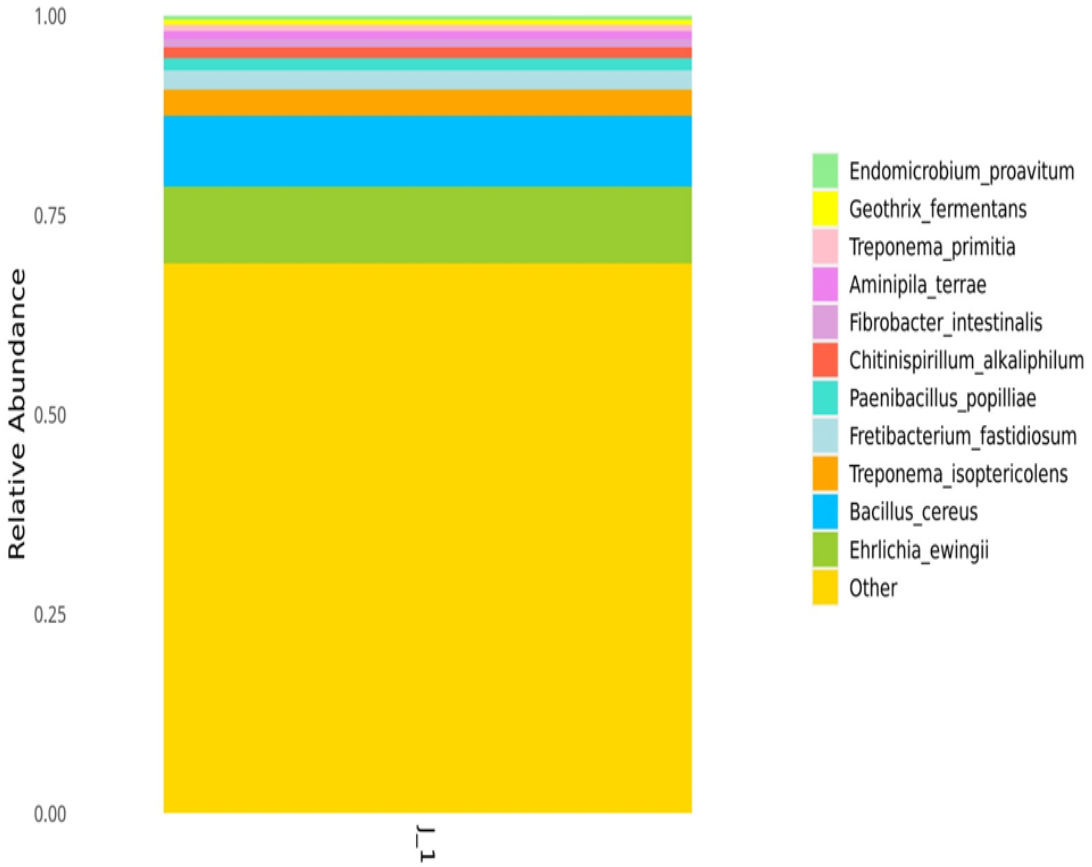

FIGURE 4.12: Relative abundance of bacterial taxa at the Species level in the termite gut microbiome.

4.6.4 Taxonomic Tree

Circular phylogenetic tree (Figure 4.13) revealed the evolutionary relationships among major bacterial lineages. The tree displays branches radiating from a central divergence point and is color-coded according to phylum-level taxonomy (legend at right). Major clades are highlighted by background shading, including Bacillota (purple, lower left), Spirochaetota (green/teal, upper left), and Alphaproteobacteria within Pseudomonadota (blue, right).

Key taxa of interest are labeled as follows: *A; Bacillaceae* , *B; Bacillus* , *C; Anaplasmataceae* , *D; Ehrlichia* , *E; Treponemataceae* , and *F; Treponema*. Additional annotations indicate higher taxonomic groups such as *Bacilli*, *Caryophanales*, *Rickettsiales*, and *Alphaproteobacteria*. This representation provides

a comprehensive overview of the phylogenetic placement of the study organisms within the broader bacterial domain, highlighting deep evolutionary divergences between Gram-positive Bacillota, spirochetes, and proteobacterial lineages. Branch lengths and node support (where indicated) reflect the degree of evolutionary divergence and robustness of the inferred relationships.

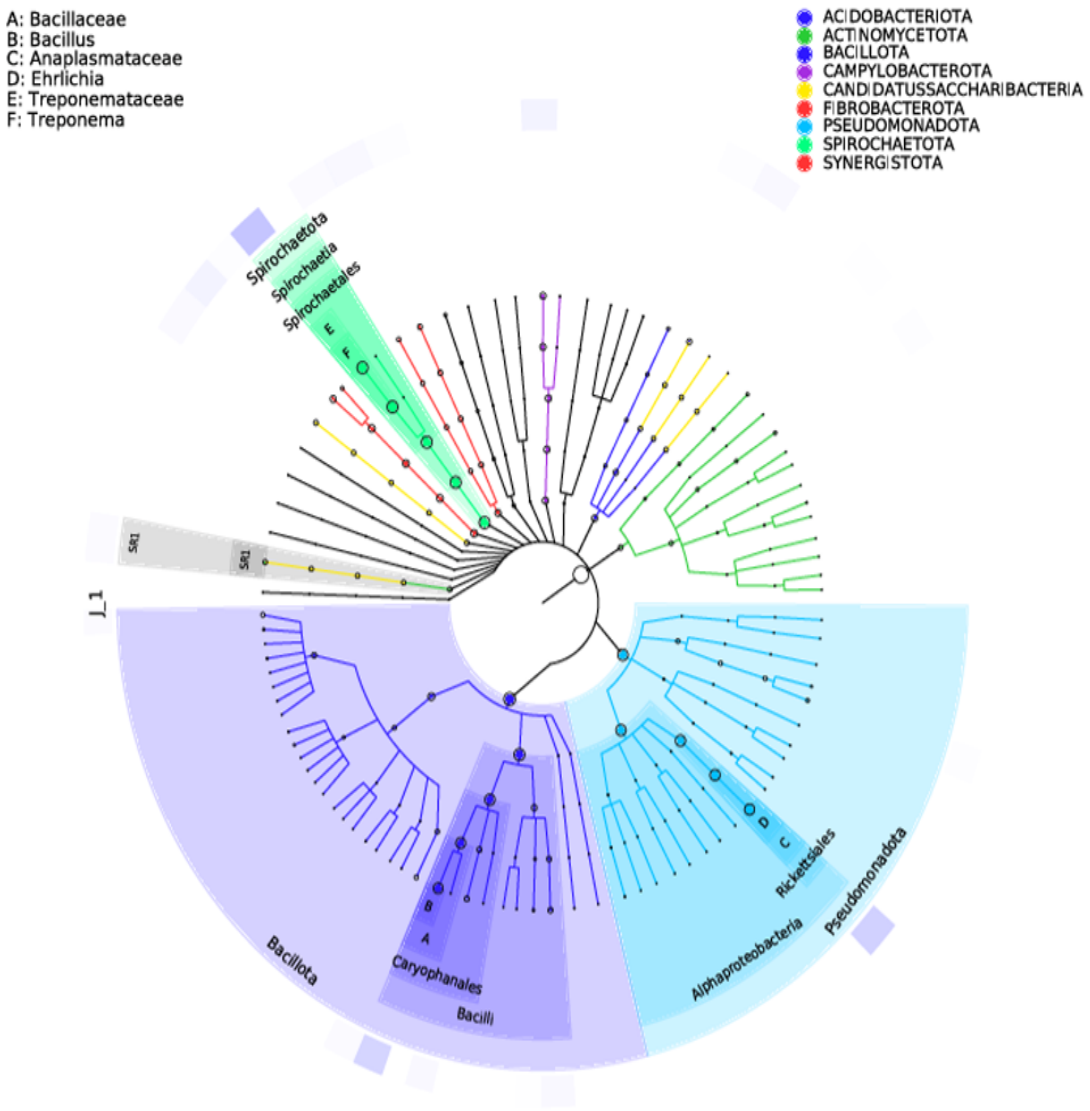

FIGURE 4.13: Circular phylogenetic tree generated using GraPhlAn

Fig 4.13 illustrates the taxonomic composition and relative abundance of microbial communities identified in the sample, colour-coded at the phylum level. Node size reflects relative abundance. Labelled taxa (A–F) indicate the most abundant genera detected.

4.6.5 Phylogenetic Analysis

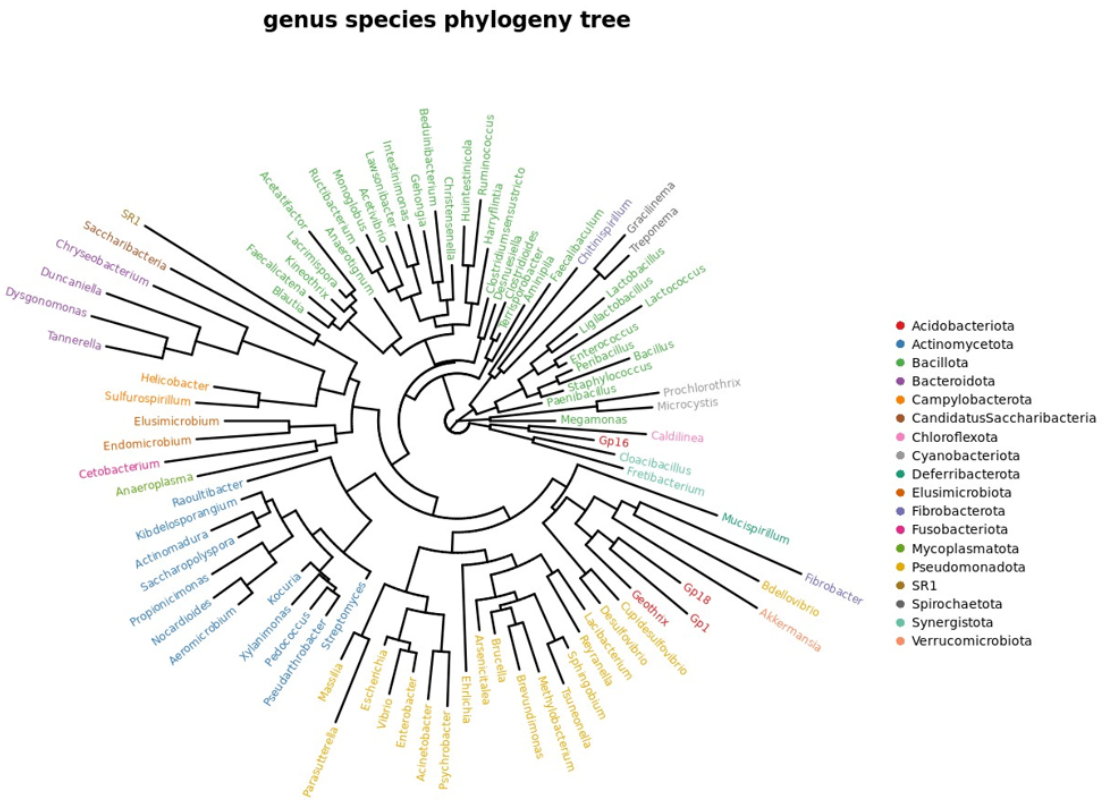

FIGURE 4.14: Circular genus-level phylogenetic tree depicting the bacterial diversity and community structure of the termite gut metagenome.

A genus-level phylogenetic tree analysis revealed a highly diverse bacterial community spanning at 19 major phyla, representing the complex symbiotic architecture. The phylum Bacillota (formerly Firmicutes) forms one of the most prominent clades, comprising diverse anaerobic and facultative anaerobic genera including *Blautia*, *Ruminococcus*, *Clostridium*, *Lactobacillus*, *Bacillus*, and *Paenibacillus*.

Pseudomonadota (Proteobacteria) represent a metabolically versatile group, encompassing α-, β-, and γ-proteobacteria such as *Escherichia*, *Enterobacter*, *Acinetobacter*, *Ehrlichia*, and *Methylobacterium*.

Actinomycetota appear as a dense, interconnected clade presenting genera such as *Streptomyces*, *Xylanimonas*, *Pseudarthrobacter*, and *Kocuria*, Similarly, Bacteroidota form a highly branched cluster dominated by *Dysgonomonas*, *Duncaniella*, and *Tannerella*.

Several specialized lineages includes *Spirochaetota* (*Treponema* and *Gracilinema*), *Fibrobacterota* (*Fibrobacter*), *Elusimicrobiota* (*Endomicrobium* and *Elusimicrobium*), *Verrucomicrobiota* (*Akkermansia*), *Campylobacterota* (*Helicobacter* and *Sulfurospirillum*), *Synergistota*, *Fusobacteriota*, *Acidobacteriota*, *Cyanobacteriota*, and candidate divisions such as SR1 and *Candidatus Saccharibacteria*, This multiphyla association demonstrates a highly coordinated evolutionary approach.

4.7 Diversity Analysis

Alpha diversity refers to the analysis of species diversity in a sample and is measured by observed species index, chao index, ace index, shannon index, simpson index and Good-coverage index.

The alpha diversity analysis of the termite gut metagenome revealed a bacterial community characterized by moderate to high species richness and good diversity, with excellent sequencing coverage. For the analyzed sample, a total of 578 observed species (Sobs) were detected. The Chao1 and ACE richness estimators both yielded identical values of 578, indicating that the observed richness closely matches the estimated total richness representing the entire bacterial diversity present in the sample, with very few undetected rare taxa.

TABLE 4.11: Alpha diversity metrics of the bacterial community in the termite gut metagenome. Sobs = observed species richness; Chao and ACE = estimated species richness; Shannon and Simpson = diversity indices; Coverage = Good's coverage

Sample name	**sobs**	**chao**	**ace**	**shannon**	**simpson**	**coverage**
	578	578	578	4.552141	0.036658	1

The Shannon diversity index of 4.55 reflects a relatively high level of species diversity and evenness within the community. In microbial ecology, Shannon values typically range from 2–4 in many host-associated microbiomes ; a value above 4.5 indicates a complex and well-balanced microbial assemblage.

Simpson's index measures the dominance of the most abundant species ; values closer to 0 indicate that no single taxon overwhelmingly dominates the community, while values approaching 1 suggest low diversity with strong dominance. The Simpson's index value of 0.0367 is notably low, further confirming high diversity. The low Simpson value demonstrates that the bacterial community is composed of many taxa with relatively balanced abundances.

The Good's coverage of 1.000 (or 100%) indicates excellent sequencing depth and sampling completeness. This means that the generated sequence data adequately represents the microbial community, and additional sequencing would likely yield very few new bacterial taxa.

4.8 Functional Annotation

4.8.1 eggNOG-Mapper

Functional annotation is a critical step in metagenomic analysis, transforming assembled or predicted gene sequences into biologically interpretable information. eggNOG-Mapper (version 2) is a widely used bioinformatics tool that performs orthology-based functional annotation by mapping query sequences against the eggNOG database of orthologous groups derived from thousands of organisms.

4.8.1.1 Seed Ortholog Identification

Seed orthologs file contains the raw alignment results from DIAMOND-based homology search, which contains 77,600 query sequences that represent predicted protein-coding genes from the assembled metagenome (Table 4.12).

An identity of 75% suggests that the matching organisms are closely related at the family-to-genus level to the reference sequences in eggNOG, which is appropriate for inferring shared function through orthology. The mean query coverage of 92.0%

is particularly strong, indicating that nearly the full length of query sequences was aligned to their orthologs in the majority of cases.

TABLE 4.12: Seed Ortholog Mapping generated by eggNOG Mapper

Parameter	**Value**
Total Query Sequences	77,600
Sequences with Seed Ortholog Hit ($e \leq 1\times10^{-3}$)	42,164 (54.3%)
Sequences without Hit	35,436 (45.7%)
Mean Bit Score	48.1
Minimum Bit Score	23.1
Maximum Bit Score	181
Mean Percent Identity	75.30%
Mean Query Coverage	92.00%

The relatively low mean bit score (48.1) reflects the prevalence of short query sequences and borderline matches common in metagenomics, but is acceptable given the query coverage and identity values.

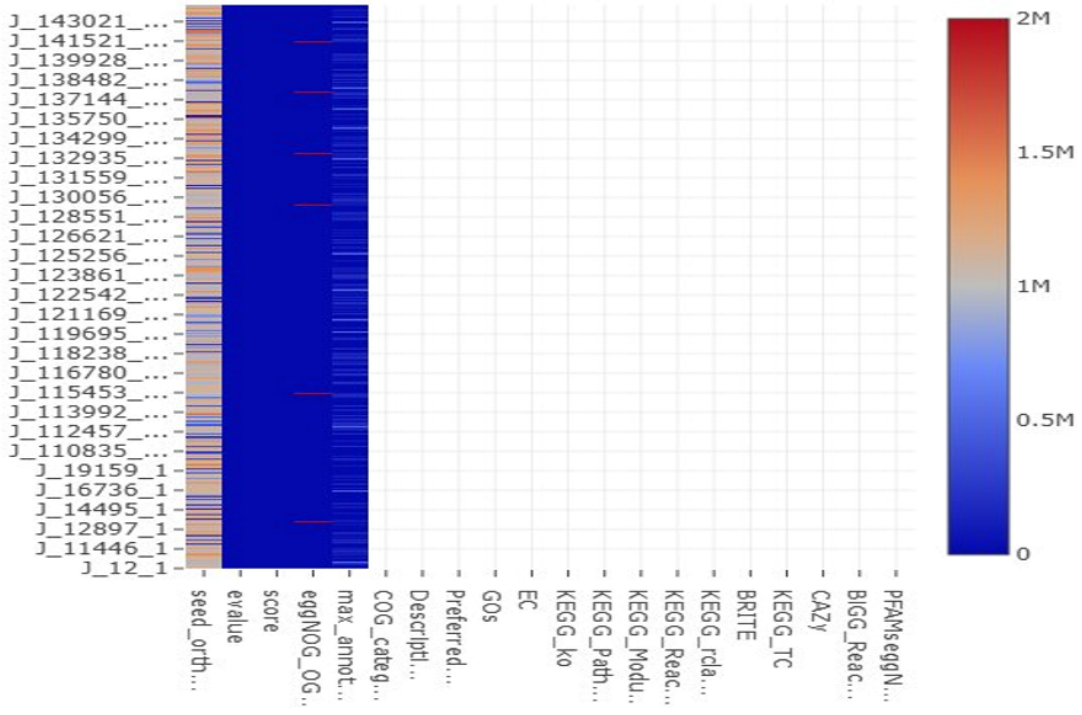

FIGURE 4.15: Heatmap visualization of the seed orthologs file showing alignment metrics across all query sequences (y-axis). Functional annotation fields extracted from eggNOG-mapper (x-axis)

The first column in figure 4.15 represents the Orthologous Seed Diversity. This column displays a highly variable, multi-colored band composed of alternating shades of orange, light grey, and blue. The broad spectrum of colors confirms that the predicted genes map back to a wide, taxonomically diverse pool of reference organisms rather than a single dominant species.

Statistical Significance and Alignment Quality were presented in columns 2–5. In this visualization, dark blue represents values close to zero. For the 'evalue' column, this solid blue line is a strong biological indicator that the sequence alignments achieved a high level of statistical significance, with E-values zero. This confirms highly reliable matches between query gene fragments and known reference proteins.

The red stripes point directly to specific, highly conserved orthologous gene families that were successfully resolved within the query sequence. This complete absence of color in rest of the columns indicates missing data fields where functional terms could not be mapped.

The reasons for unassigned functional categories in the heatmap can be due to the presence of lineages unique to this ecosystem, such as Spirochaetota, Elusimicrobia, and Endomicrobiales, that are poorly represented in public reference databases or the presence of viral or mobile genetic elements, including prophages and transposons, which lack corresponding entries within standard functional repositories.

4.8.1.2 eggNOG-mapper Functional Annotation

Functional annotation using eggNOG-mapper resulted in a total of 42,164 successfully annotated gene sequences (seed orthologs).

While all annotated sequences received eggNOG orthologous group (OG) assignments, only a small fraction were linked to broader functional categories: 264 sequences (0.6%) were assigned COG functional categories and functional descriptions, and 148 sequences (0.4%) received KEGG Orthology (KO) and Gene Ontology (GO) terms (Table 4.13).

TABLE 4.13: Functional Annotation Summary of Predicted Genes

Annotation Category	**No. of Entries**	**% of Annotated**
Total annotated entries (seed orthologs)	42,164	100.00%
With eggNOG OG assignment	42,164	100.00%
With COG functional category	264	0.60%
With KEGG KO assignment	148	0.40%
With GO term assignment	148	0.40%
With functional Description	264	0.60%
With CAZy family assignment	0	0.00%

The heatmap (Figure 4.16) reveals that the termite gut metagenome contains a large number of genes that were successfully annotated using the eggNOG database. The colored vertical stripes on the left represent sequences that received strong, high-quality functional annotations (high similarity and coverage to known orthologs).

These red-to-orange regions indicate genes with strong functional matches. Many sequences confirm that a substantial portion of the microbial genes in the termite gut are functionally recognizable and linked to known metabolic processes.

The large dark blue area on the right shows that for most sequences, certain alignment metrics (such as specific coordinate positions or lower-confidence matches) are minimal or absent. This is expected in metagenomic data, where many genes are partial, novel, or only distantly related to reference databases.

4.8.1.3 Taxonomic Composition from eggNOG Orthologous Groups

Functional annotation using eggNOG-mapper revealed the taxonomic profile of the termite gut metagenome at the ortholog level (Table 4.14). The bacterial community is heavily dominated by members of the phylum Firmicutes, with Bacilli representing the largest fraction (13,472 entries, 32.0%), followed by Clostridia (8,973 entries, 21.3%). Together, these two classes account for over 53% of all annotated orthologs, underscoring their central role in the termite hindgut as key fermenters, lignocellulose degraders, and short-chain fatty acid producers.

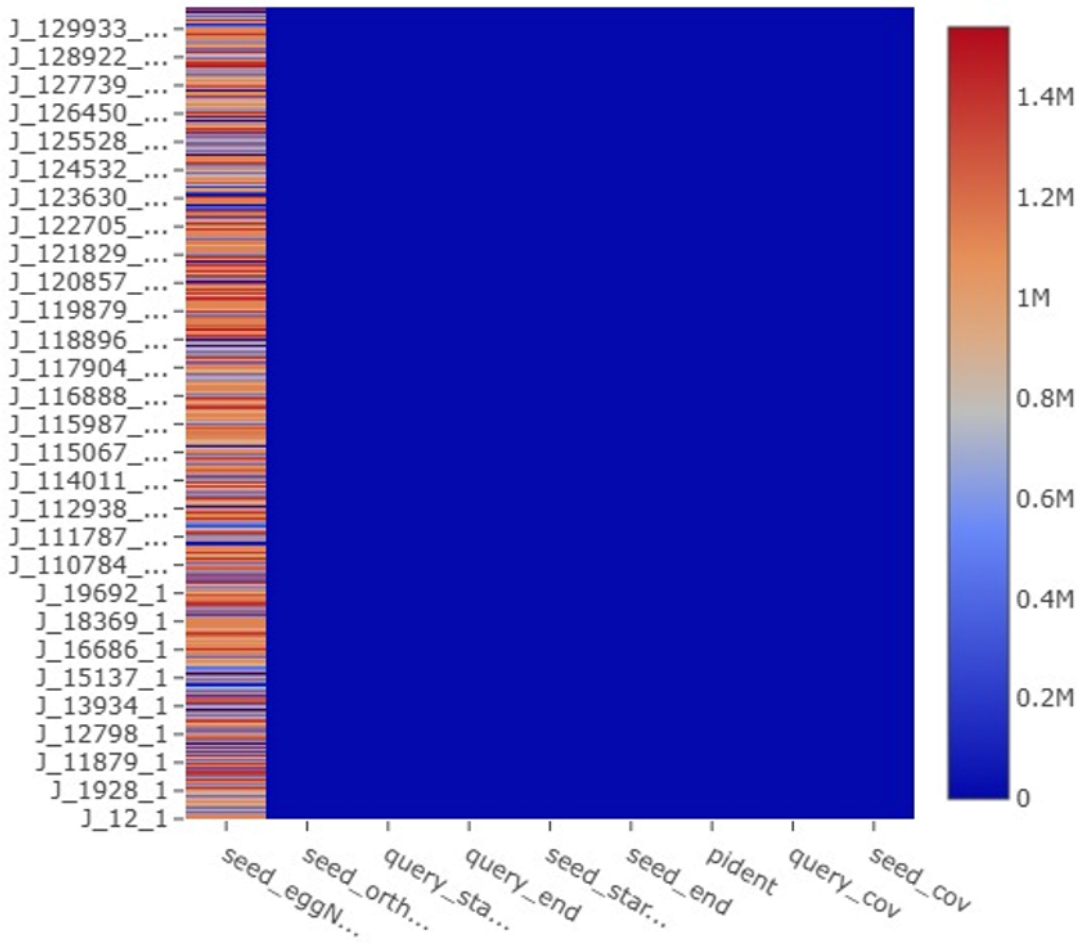

FIGURE 4.16: Heatmap of the annotations file, displaying all functional annotation columns (seed_ortholog, evalue, score, eggNOG_OGs, max_annot_lvl, COG_category, Description, Preferred_name, GOs, EC, KEGG_ko, KEGG_Pathway, KEGG_Module, KEGG_Reaction, KEGG_rclass, BRITE, KEGG_TC, CAZy, BiGG_Reaction, PFAMs).

Bacteroidetes constitute the third most abundant group (3,913 entries, 9.3%), primarily contributing polysaccharide-degrading and nitrogen-recycling functions. Other notable groups include Gammaproteobacteria (6.0%), Deltaproteobacteria (5.8%), and broad-level Firmicutes (5.6%).

A substantial proportion of sequences (7.2%) could only be assigned to the general "Bacteria" category, reflecting the presence of divergent or novel microbial lineages typical of termite gut symbionts. Minor contributions from Actinobacteria, Alphaproteobacteria, Betaproteobacteria, and Cyanobacteria (likely dietary

contaminants) further illustrate the metabolic diversity and environmental interactions within the gut ecosystem. This taxonomic distribution aligns with the phylogenetic tree.

TABLE 4.14: Taxonomic Composition from eggNOG OG Assignments

Taxonomic Group	**Count**	**% Annotated**	**Ecological Significance in Termite Gut**
Bacilli	13,472	32.00%	Bacillus spp., Paenibacillaceae - sporulating, aerotolerant; linked to lignocellulose degradation and nitrogen cycling
Clostridia	8,973	21.30%	Clostridiaceae, Ruminococcaceae - strictly anaerobic fermenters; key acetate/H_2 producers in hindgut
Bacteroidetes	3,913	9.30%	Flavobacteriia, Chryseobacterium, Flavobacterium - polysaccharide degraders and organic nitrogen recyclers
Bacteria (broad)	3,042	7.20%	Unresolved bacterial sequences; high sequence divergence from reference databases
Gamma proteobacteria	2,533	6.00%	Vibrionales, Pasteurellales - facultative anaerobes; secondary community members
Cyano bacteria	2,455	5.80%	Synechococcus spp. - likely dietary contaminants from plant material ingested by termites
Delta proteobacteria	2,454	5.80%	Desulfobacterales - sulfate-reducing bacteria; contribute to redox cycling in gut
Firmicutes (broad)	2,377	5.60%	Erysipelotrichia and unclassified; fermenters and mucosal associates
Erysipelotrichia	771	1.80%	Emerging gut-associated class; fermentative, lipid metabolism roles reported
Beta proteobacteria	707	1.70%	Diverse metabolic roles; some members involved in N_2 fixation
Actinobacteria	601	1.40%	High-GC bacteria; secondary metabolite biosynthesis and chitin degradation
Alpha proteobacteria	364	0.90%	Nitrogen fixers (Rhizobiales-related); secondary symbionts
Eukaryota / Viridiplantae	295	0.70%	Plant-derived sequences from termite diet (lignocellulosic substrate)

4.8.1.4 COG Category Analysis

Of the 264 sequences assigned a COG functional category, only two distinct functional classes were represented in the dataset (Table 4.15).

TABLE 4.15: COG functional classification of annotated metagenomic sequences.

COG	Category Name	Count	% of COG - annotated
J	Translation, Ribosomal Structure and Biogenesis	148	56.10%
S	Function Unknown	116	43.90%

The COG category J (Translation, Ribosomal Structure and Biogenesis) was the dominant one in this annotation. This category refers to ribosomal proteins, translation factors, aminoacyl-tRNA synthetases, and other core translational machinery. These genes are conserved across the bacterial domain. Their detection confirms that the dataset contains bacterial genomic content from diverse members of the termite gut microbiome.

The 43.9% of COG-annotated sequences classified as category S (Function Unknown) is consistent with the observation that a large fraction of proteins encoded by microorganisms represent uncharacterized proteins. This is particularly true in the termite gut, where the majority of microbial lineages remain uncultivated and their functional gene space is largely unexplored.

4.8.1.5 KEGG Pathway and Ortholog Analysis

Only 148 sequences (0.4% of annotated) received KEGG KO assignments, all mapping to a single KO entry:

- KO: K02950 - Small subunit ribosomal protein S5 (rpsE)
- KEGG Pathway: ko03010 / map03010 - Ribosome

This assignment to ribosomal protein S5 (rpsE) and the ribosome KEGG pathway supports the COG J finding. The ribosomal small subunit protein S5 is one of the most conserved proteins in bacteria and a canonical marker for phylogenetic studies.

4.8.1.6 CAZyme Family Annotation

Annotation with eggNOG Mapper does not reveal any CAZymes within the metagenomic data analyzed. To identify the presence of these enzymes data was subjected to CUPP analysis. Four CAZyme families were detected from the 56 annotated proteins. The most dominant family was the glycoside hydrolase family, and the least represented was from three glycosyltransferase families

TABLE 4.16: List of CAZyme families predicted from the termite gut metagenome

CAZyme Family	**Class**	**No. of Hits**	**% of Total**	**CUPP Group**
GH113	Glycoside Hydrolase	49	87.50%	GH113:4.2
GT2	Glycosyltransferase	4	7.10%	GT2:454.1 / GT2:2037.2 / GT2:1825.1
GT106	Glycosyltransferase	2	3.60%	GT106:9.1
GT3	Glycosyltransferase	1	1.80%	GT3:4.1

The presence of CAZyme families like carbohydrate - degrading (GH) and carbohydrate - synthesising (GT) enzyme indicates the double metabolic role of gut microbiome, which is helpful in breaking down dietary polysaccharides.

4.8.2 Glycosyltransferase Families and Predicted Biological Roles

4.8.2.1 GH113

a) Mannan-Degrading Glycoside Hydrolases

GH113 enzymes are beta-mannanases and beta-mannosidases enzymes that cleave beta-1,4-mannose linkages in mannan-containing polysaccharides. GH113 is the overwhelmingly dominant CAZyme family detected in this dataset that accounting for 49 of the 56 annotated hits (87.5%). All 49 GH113 hits belong to the same CUPP group (GH113:4.2) and branch (GH113*0) which indicates that they share a single conserved peptide pattern characteristic of this enzyme family (Table 4.17).

b) Glycosyltransferases - GT2, GT3, and GT106

Seven hits (12.5% of total) belong to glycosyltransferase families. The enzymes that catalyse the formation of glycosidic bonds, help in building polysaccharide chains rather than degrading them.

4.8.2.2 GT2

GT2 is one of the largest and most functionally diverse glycosyltransferase families in the CAZy database. Members of this family catalyse the transfer of sugar pieces from nucleotide-sugar donors to acceptor molecules, which results in producing a wide range of polysaccharide linkages.

Three distinct GT2 CUPP subgroups were detected: GT2:454.1 (1 hit, score 1.099), GT2:2037.2 (2 hits, score 0.751 each), and GT2:1825.1 (1 hit, score 0.892). The GT2:454.1 hit has the highest CUPP score in the entire dataset (1.099) which indicates that it is the strongest pattern match of any annotated protein.

4.8.2.3 GT106

GT106 is a less well-characterised glycosyltransferase family that is associated with the synthesis of specific glycan linkages in bacterial surface polysaccharides. Both GT106 hits belong to the same CUPP group (GT106:9.1) and also carry identical scores (0.662), suggesting they derive from the same gene or closely related variants.

4.8.2.4 GT3

GT3 encodes glycogen synthase activity that is specifically the synthesis of alpha-1,4 and alpha-1,6 glucan chains which form glycogen granules. The detection of a GT3 hit in the termite gut community is biologically meaningful. This single GT3 hit scored 0.227 which is the lowest score in the dataset and covered only 13% of the reference domain, so it should be treated as a tentative annotation pending further confirmation.

TABLE 4.17: Glycosyltransferase family distribution, CUPP group assignments, and predicted biological functions

Family	**CUPP Group**	**Hits**	**Score**	**Coverage**	**Biological Role**
GT2	GT2:454.1	1	1.099	14%	Exopolysaccharide / cell wall biosynthesis
GT2	GT2:2037.22		0.751	14%	Exopolysaccharide / cell wall biosynthesis
GT2	GT2:1825.11		0.892	12%	Exopolysaccharide / cell wall biosynthesis
GT106	GT106:9.1	2	0.662	13%	Bacterial surface glycan synthesis
GT3	GT3:4.1	1	0.227	13%	Glycogen biosynthesis (carbon storage)

4.8.3 Functional Profiling of Metabolic Pathways via MetaCyc

To predict the metabolic potential of the microbial community, PICRUSt2 functional annotation against the MetaCyc database was performed, that analyzed the pathways at Levels 1, 2, and 3. The overall functional landscape of the sample is presented in figure 4.13.

MetaCyc pathway profile obtained includes primary and secondary metabolism and related metabolites, reactions, enzymes and genes. Pathways were classified

according to their biological functions and the types of metabolites they produce or consume.

TABLE 4.18: MetaCyc Level 1 and Level 2 functional pathway distribution predicted by PICRUSt in termite gut metagenome sample. Abundance values represent predicted gene copy numbers. Percentages are relative to total predicted abundance within each level.

Pathway Category	**Abundance (J_1)**	**Relative (%)**
Level 1 - Broad Superclasses		
Biosynthesis	4,381,129	51.30%
Superpathways	2,361,093	27.60%
Generation of Precursor Metabolite and Energy	812,052	9.50%
Degradation/Utilization/Assimilation	676,042	7.90%
Metabolic Clusters / Glycan / Other	310,341	3.60%
Detoxification	12,479	0.10%
Level 2 - Top Functional Categories (excluding 'Other')		
Nucleoside and Nucleotide Biosynthesis	1,105,031	12.80%
Amino Acid Biosynthesis	1,065,489	12.30%
Cofactor, Prosthetic Group, and Vitamin Biosynthesis	870,414	10.10%
Fatty Acid and Lipid Biosynthesis	445,553	5.20%
Carbohydrate Biosynthesis	340,374	3.90%
Fermentation	283,373	3.30%
Cell Structure Biosynthesis	264,749	3.10%
TCA Cycle	162,818	1.90%
Carbohydrate Degradation	134,696	1.60%
Glycolysis	122,823	1.40%

Functional pathway profiling predicted a total of 380 individual pathways across eight broad Level 1 superclasses and 39 Level 2 categories in sample (Table 4.18). Biosynthesis pathways dominated the community's functional repertoire, accounting for 51.3% of total predicted abundance, followed by Superpathways (27.6%) and Generation of Precursor Metabolites and Energy (9.5%) at Level 1.

Degradation, Utilization, and Assimilation pathways contributed 7.9%, while Detoxification remained minimal at 0.1%, indicating that the community is primarily oriented toward anabolic rather than catabolic or defensive functions under the sampled conditions.

At Level 2 (Figure 4.17), the most abundant defined pathway categories were Nucleoside and Nucleotide Biosynthesis (12.8%) and Amino Acid Biosynthesis (12.3%), collectively accounting for over 25% of all predicted functional capacity. This dual dominance reflects the high anabolic demand of the termite gut microbiome for nitrogen-containing biomolecules in the context of a nitrogen-poor, lignocellulosic diet. Cofactor, Prosthetic Group, and Vitamin Biosynthesis ranked third (10.1%), underscoring the critical role of coenzyme availability across the diverse enzymatic network of the community. Fatty Acid and Lipid Biosynthesis (5.2%), Carbohydrate Biosynthesis (3.9%), Fermentation (3.3%), and Cell Structure Biosynthesis (3.1%) further confirmed active membrane biogenesis, cell wall synthesis, and anaerobic energy metabolism consistent with the microoxic to anoxic conditions of the hindgut. The TCA cycle (1.9%), Glycolysis (1.4%), and Carbohydrate Degradation (1.6%) together reflect a functional central metabolic core supporting both energy generation and biosynthetic precursor supply.

At the highest level (Level 3), the leading individual pathways were Aerobic Respiration I via cytochrome c (1.02%), Gondoate Biosynthesis anaerobic (0.96%), Pentose Phosphate Pathway non-oxidative branch (0.95%), and cis-Vaccenate Biosynthesis (0.94%), alongside multiple nucleotide and amino acid biosynthesis superpathways each contributing 0.83-0.89%. The presence of aerobic respiration as the top-ranked specific pathway indicates the existence of microoxic niches within the gut where oxidative phosphorylation is active, while the prominence of anaerobic fatty acid biosynthesis and fermentation pathways confirms the predominantly anaerobic character of the broader community. Overall, the MetaCyc profile of sample J_1 reveals a metabolically diverse and anabolically dominant microbiome that combines active biosynthesis of essential biomolecules with versatile central carbon metabolism, consistent with a nutritional mutualist role in supporting the nitrogen and energy requirements of the termite host.

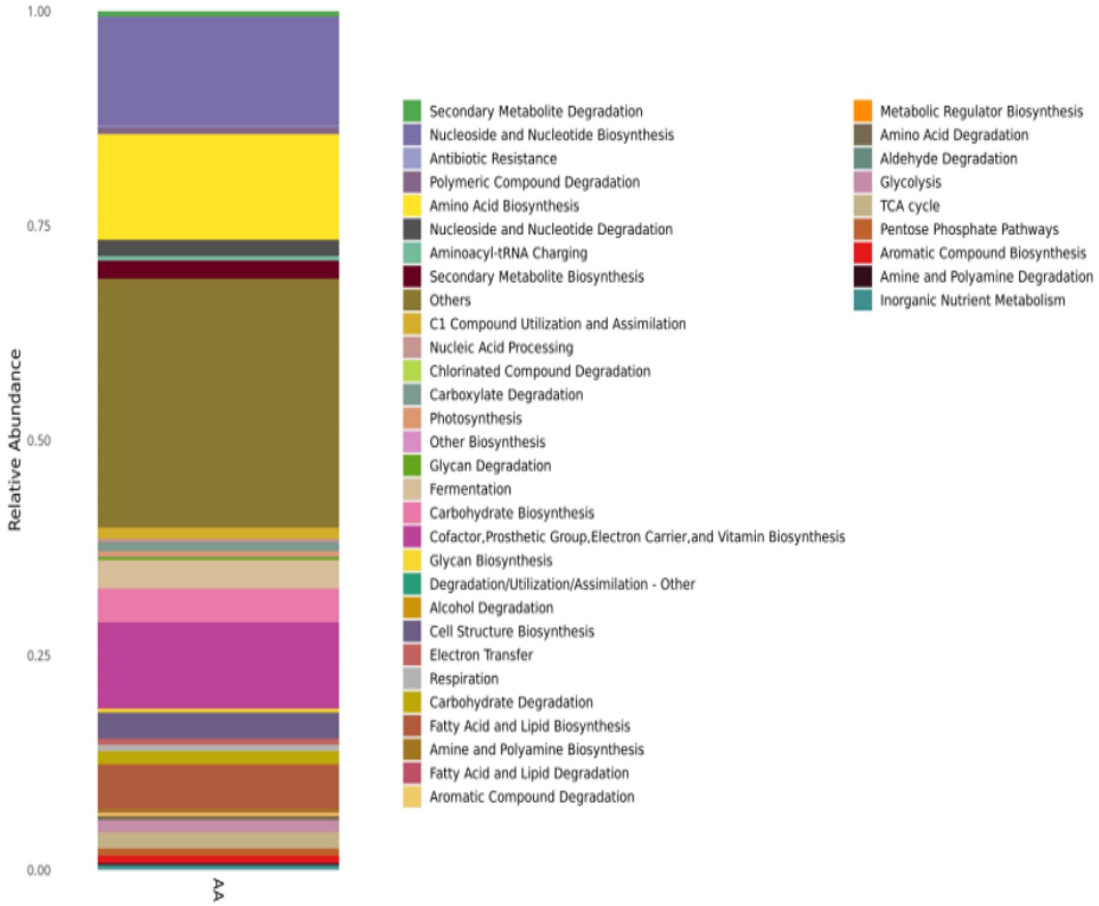

FIGURE 4.17: Predicted metabolic pathway profiles of the microbial community at MetaCyc Level-2

Chapter 5

Discussion

The multi-level analysis conducted in this study presents detailed insight of the termite gut microbiome as a functionally integrated, phylogenetically diverse, and anabolically active ecosystem.

At the phylum level, the termite gut microbiome in this study was dominated by Bacillota (formerly Firmicutes, ∼45%), followed by Pseudomonadota (∼15%) and Spirochaetota (∼13%). This tripartite dominance is highly consistent with the established paradigm for wood-feeding lower termite gut microbiomes, where Firmicutes, Proteobacteria, and Spirochaetes collectively comprise the majority of the bacterial community [98, 99]. Bacillota dominance in the termite hindgut reflects the ecological niche's selective environment, which favours anaerobic fermenters capable of producing acetate, propionate, and butyrate from lignocellulosic substrates as primary metabolites for host nutrition [100]. The detection of Spirochaetota at ∼13% relative abundance is particularly significant from an ecological perspective. Spirochaetes of the genus *Treponema* are well-documented as dominant and functionally critical members of lower termite hindguts, contributing to acetogenesis (the Wood-Ljungdahl pathway), H_2/CO_2 fixation, and N_2 fixation, processes that are essential for the termite's carbon and nitrogen economy [101, 102]. The *Treponema*-dominated Spirochaetota clade identified in this

study is phylogenetically consistent with the *Treponema isoptericolens* and *Treponema primitia* species detected at the species level, both of which are recognised as specialist gut symbionts of wood-feeding termites [103].

The presence of Bacteroidota (∼moderate abundance) is expected in this ecosystem, as Bacteroidales are established contributors to polysaccharide hydrolysis and nitrogen recycling in the termite hindgut [104]. The detection of Fibrobacterota, Elusimicrobiota, and Candidatus Saccharibacteria, at low relative abundances, is noteworthy as these lineages are disproportionately represented in insect gut environments and are thought to play specialised roles in cellulose degradation (Fibrobacterota) and mutualistic parasitism of bacterial hosts (Saccharibacteria)[105, 106].

Treponema (12%) and Ehrlichia (11%) emerged as the most abundant resolved genera, followed by Bacillus (9%) and Sulfurospirillum (3%). While the dominance of *Treponema* is fully consistent with the termite gut ecosystem as discussed above. The high relative abundance of Ehrlichia needs to be consider specifically. Ehrlichia is an obligate intracellular pathogen of vertebrates transmitted by ticks and is not typically a member of the termite gut microbiome [107]. Its detection at high abundance in this dataset, with the co-dominant species identified as Ehrlichia ewingii (9% at species level), is most likely attributable to database-mediated misclassification, where 16S rRNA sequences from Anaplasmataceae-affiliated gut bacteria with sequence similarity to Ehrlichia reference sequences in the database are being incorrectly assigned. This phenomenon is well-documented in microbiome studies that rely on short 16S amplicons for classification, where the limited taxonomic resolution of short variable regions can produce spurious genus-level assignments [108, 109].

The detection of Bacillus cereus as a co-dominant species (8%) similarly needs special consideration. While Bacillus species are genuine members of termite gut communities and have been associated with cellulose degradation, nitrogen fixation, and lignocellulose pre-treatment activities [110], the identification of B. cereus specifically may reflect the broader Bacillus cereus sensu lato complex assignment common in 16S-based classification, where taxonomic resolution at the

species level is insufficient to distinguish between closely related Bacillus lineages. The genuine gut-associated Bacilli are more likely to belong to aerotolerant anaerobic lineages such as Paenibacillus spp., which were also detected in this dataset [111].

Among the reliably identified specialised gut symbionts, the detection of *Fibrobacter intestinalis*, *Endomicrobium proavitum*, *Treponema isoptericolens, Treponema primitia,* and *Fretibacterium fastidiosum* is of considerable biological significance. These species represent highly specialised, often uncultivated termite gut symbionts whose ecology and function have been characterised primarily through metagenomics and single-cell genomics [112]. *Fibrobacter intestinalis* is a cellulose-degrading specialist closely related to the ruminant cellulose degrader *F. succinogenes* and is well-established as a key lignocellulose-deconstructor in lower termite hindguts [113]. *Endomicrobium proavitum* (phylum Elusimicrobiota) is an obligate intracellular endosymbiont of termite gut flagellates, representing one of the most intimate levels of symbiosis in this system [114].

The alpha diversity analysis revealed 578 observed species (Sobs), with Chao1 and ACE estimators both returning identical values of 578, indicating near-complete sampling of the community with a Good's coverage of 1.000 (100%). The Shannon diversity index of 4.55 reflects high species diversity and evenness, substantially exceeding the typical range of 2–4 reported for many host-associated gut microbiomes [115]. The Simpson's index of 0.0367 confirms that no single taxon overwhelmingly dominates the community, a value close to 0 indicates high species evenness and true diversity rather than numerical dominance by one or a few taxa [116].

These diversity metrics are consistent with published reports for lower termite gut microbiomes, which are recognised as among the most diverse microbial ecosystems per unit volume on Earth [117].

The genus-level phylogenetic tree and GraPhlAn circular map revealed a bacterial community spanning 19 major phyla, consistent with the established understanding that termite gut microbiomes harbour extraordinary phylogenetic diversity,

including multiple lineages that are poorly represented or entirely absent from other ecosystems [118, 119].

The prominent Bacillota clade, encompassing diverse anaerobic fermenters including *Blautia*, *Ruminococcus*, *Clostridium*, *Lactobacillus*, *Bacillus*, and *Paenibacillus*, reflects the multi-functional anabolic and catabolic roles of this phylum in the termite hindgut, ranging from primary cellulose fermentation to nitrogen fixation and hydrogen metabolism [100, 120].

The presence of *Pseudomonadota* forming a large metabolically versatile clade encompassing Alpha-, Beta-, and Gammaproteobacteria including *Ehrlichia*, *Methylobacterium*, *Escherichia*, and *Acinetobacter* reflects the known metabolic versatility of Proteobacteria in gut environments. *Methylobacterium* are genuine termite gut associates known to participate in C1 compound metabolism and nitrogen fixation [121].

Actinomycetota, appearing as a dense interconnected clade with genera such as *Streptomyces*, *Xylanimonas*, and *Pseudarthrobacter*, are known contributors to secondary metabolite biosynthesis and chitin degradation in termite gut and nest environments [122]. Similarly, Bacteroidota, dominated by *Dysgonomonas*, *Duncaniella*, and *Tannerella*, confirm the role of Bacteroidales as polysaccharide-degrading specialists that complement the fermentative activities of Bacillota in the hindgut [104].

Among the 264 sequences receiving COG functional category assignments, COG category J (Translation, Ribosomal Structure and Biogenesis) dominated at 56.1%, with 43.9% assigned to COG category S (Function Unknown). The strong dominance of COG J reflects a well-recognised pattern in metagenomic datasets: ribosomal proteins are among the most highly conserved, universally distributed, and deeply characterised genes in bacteria, making them preferentially detectable even in divergent community members with otherwise low database representation [123]. These findings are congruent with the KEGG assignment, where all 148 annotated sequences mapped exclusively to K02950 (ribosomal protein S5, rpsE) in the Ribosome pathway (ko03010).

The high proportion of COG S (Function Unknown) sequences (43.9%) is biologically significant in the context of the termite gut, where the majority of microbial taxa are uncultivated and their functional gene space remains largely unexplored. This pattern is consistent with published COG analyses of other complex environmental metagenomes, where Category S consistently constitutes 20–50% of annotated sequences [124]. These sequences represent candidates for novel enzyme discovery, particularly in the context of lignocellulose degradation and nitrogen cycling activities unique to the termite gut ecosystem.

The detection of mannan-degrading GH113 enzymes in the termite gut microbiome is consistent with studies demonstrating that wood-feeding termites consume substrates rich in beta-mannans, and that their gut symbionts are equipped with a diversity of hemicellulase activities [125]. Previous metagenomic studies of lower termite hindguts have documented the presence of multiple GH families targeting hemicellulose, including GH26 (mannanases), GH36 (alpha-galactosidases), GH43, and GH10 (xylanases), alongside cellulose-degrading GH families such as GH5, GH7, and GH45 [126].

The three detected glycosyltransferase families (GT2, GT106, GT3) represent biosynthetic rather than degradative carbohydrate activities. GT2 is one of the largest and most phylogenetically widespread GT families, catalysing the biosynthesis of diverse exopolysaccharides and cell wall components across bacteria [127]. The detection of GT106 associated with bacterial surface glycan synthesis and GT3 associated with glycogen biosynthesis (carbon storage) points to active investment by the microbial community in both biofilm formation and intracellular energy storage strategies, consistent with the intermittent substrate availability experienced by gut microorganisms [128].

The PICRUSt2-based MetaCyc functional pathway prediction revealed a metabolic landscape dominated by biosynthetic pathways, with Biosynthesis accounting for 51.3% of total predicted pathway abundance at Level 1. This anabolic dominance is consistent with the ecological role of the termite gut microbiome as a nutritional mutualist, in which the microbial community invests heavily in synthesising

essential biomolecules — including amino acids, nucleotides, lipids, and cofactors that are limiting in the nitrogen-poor, carbon-rich wood-based diet of the host.

At Level 2, the co-dominance of Nucleoside and Nucleotide Biosynthesis (12.8%) and Amino Acid Biosynthesis (12.3%) is ecologically interpretable in the context of the termite nitrogen paradox, where the host and its gut symbionts must collectively synthesise all essential amino acids from a diet with a C:N ratio that can exceed 300:1 in dry wood [129].

The high predicted abundance of amino acid biosynthesis pathways, particularly for branched-chain amino acids (BCAAs: leucine, isoleucine, valine), arginine, histidine, and lysine identified at Level 3, is consistent with the documented role of termite gut spirochaetes, firmicutes, and N_2-fixing diazotrophs in supplying fixed nitrogen and biosynthesising amino acids that are provisioned to the host through a nutritional mutualism [101].

Cofactor, Prosthetic Group, and Vitamin Biosynthesis pathways (10.1%) reflect the critical demand for coenzymes such as cobalamin (vitamin B12), folate, biotin, thiamine, and flavins across the diverse enzymatic activities present in the gut microbiome. This finding is consistent with the established nutritional mutualism in termites, where gut bacteria provision the host with B vitamins that are absent from the woody substrate [130]. The high predicted abundance of cofactor biosynthesis pathways is particularly relevant in light of the phylogenetic composition of the community, as spirochaetes (specifically *Treponema spp.*) are known producers of corrinoid cofactors (B12 and related compounds) that support acetogenic Wood-Ljungdahl metabolism [131].

The strong representation of Fermentation pathways (3.3%) alongside aerobic respiration as the top-ranked Level 3 pathway (Aerobic Respiration I via cytochrome c, 1.02%) indicates metabolic stratification within the gut, with different microbial guilds occupying distinct redox niches. The termite hindgut is known to harbour a radial oxygen gradient, with peripheral microoxic zones supporting aerobic and microaerophilic bacteria while the central lumen remains strictly anoxic and supports diverse fermentative and acetogenic metabolisms [132].

The prominence of fatty acid biosynthesis pathways at Level 3, including Gondoate Biosynthesis (anaerobic, 0.96%) and cis-Vaccenate Biosynthesis (0.94%), reflects active membrane lipid biosynthesis for maintaining cell membrane integrity and fluidity under the variable pH and redox conditions of the termite hindgut [133].

Chapter 6

Conclusion and Future Work

6.1 Conclusion

This study demonstrates that the termite gut microbiome is a phylogenetically diverse, functionally specialized, and anabolically driven microbial ecosystem uniquely adapted to thrive on a recalcitrant, nitrogen-poor, lignocellulosic substrate. The integration of metagenomics with multi-level taxonomic profiling, phylogenetic analysis, and functional annotation with eggNOG, CAZyme, KEGG, and MetaCyc frameworks has yielded a comprehensive baseline characterization of this ecologically important symbiosis.

De novo assembly using MEGAHIT yielded 10 contigs (total size 3,881 bp; N50 = 406 bp), a highly fragmented output that is consistent with the extreme phylogenetic diversity and uneven abundance distribution characteristic of complex environmental metagenomes. Gene prediction using FragGeneScan achieved an 85.21% prediction rate across 136,742 input sequences, yielding 127,623 genes with predicted protein lengths of 21–93 amino acids. The strong forward-strand orientation bias (74.0%) and dominance of Frame 1 ORFs (59.6%) are indicative of a compact, high-coding-density prokaryotic gene space, while the high proportion of multi-gene sequences (44.2%) reflects the prevalence of bacterial operon structures within the dataset.

Taxonomic profiling at six hierarchical levels from phylum through species revealed a bacterial community dominated by Bacillota (∼45%), Pseudomonadota (∼15%), and Spirochaetota (∼13%), a composition fully consistent with established models of the lower termite hindgut microbiome.

At the genus level, Treponema (12%) emerged as the most reliably identified specialist gut symbiont, consistent with its established roles in acetogenesis, hydrogen metabolism, and nitrogen fixation in the termite hindgut.

The presence of highly specialised uncultivated symbionts including *Fibrobacter intestinalis, Endomicrobium proavitum, Treponema isoptericolens, Treponema primitia*, and *Fretibacterium fastidiosum* underscores the ecological specificity and evolutionary integration of the gut microbial community with its termite host.

Alpha diversity analysis confirmed high species richness (Sobs = 578), a Shannon diversity index of 4.55, a near-zero Simpson's index (0.037), and complete Good's coverage (1.000). These values collectively demonstrate that the termite gut microbiome is a highly diverse, relatively even, and thoroughly sampled bacterial community, validating the downstream functional predictions derived from this dataset.

Phylogenetic reconstruction identified 19 major bacterial phyla, confirming the extraordinary taxonomic breadth of the termite gut ecosystem and revealing deep evolutionary divergences between the major functional.

Functional annotation using eggNOG-Mapper v2 annotated 54.3% of 77,600 query sequences, with COG category J (Translation, Ribosomal Structure and Biogenesis) comprising 56.1% of COG-assigned sequences, consistent with the universal conservation and high abundance of ribosomal proteins across all bacterial taxa. The dominance of COG category S (Function Unknown, 43.9%) reflects the substantial 'microbial dark matter' component of the termite gut, comprising divergent and novel lineages not yet characterised in public databases. KEGG pathway mapping confirmed all 148 KO-annotated sequences map to ribosomal protein S5 (K02950) within the Ribosome pathway (ko03010).

CAZyme analysis identified 56 annotated carbohydrate-active enzymes from four families, with GH113 (beta-mannanases/beta-mannosidases targeting beta-1, 4-mannose linkages) overwhelmingly dominant at 87.5%. This finding directly implicates hemicellulose degradation specifically mannan hydrolysis as the primary enzymatic polysaccharide-processing activity captured in the dataset, consistent with the wood-based diet of the host termite.

MetaCyc functional pathway profiling via PICRUSt2 predicted 380 individual pathways across 8 Level-1 superclasses and 39 Level-2 categories. Biosynthesis pathways dominated at 51.3% of total predicted abundance, driven primarily by Nucleoside and Nucleotide Biosynthesis (12.8%) and Amino Acid Biosynthesis (12.3%) at Level 2, reflecting the community's high investment in nitrogen-containing biomolecule synthesis under a nitrogen-limited diet.

Fermentation (3.3%) and Aerobic Respiration I via cytochrome c (top-ranked Level-3 pathway at 1.02%) together confirm the existence of both strictly anoxic and microoxic metabolic niches within the termite hindgut. The overall MetaCyc profile characterizes the termite gut microbiome as an anabolically dominant, metabolically stratified, and functionally integrated ecosystem that simultaneously supports lignocellulose bioconversion, energy generation, nitrogen cycling, and nutritional provisioning to the termite host. The findings collectively establish a detailed baseline for understanding the structural and functional organisation of the termite gut microbiome as a model lignocellulose-degrading and nutrient-cycling ecosystem.

6.2 Future Recommendations

The findings of this study, while comprehensive within the scope of an amplicon-based metagenomic workflow, highlight several important limitations and open research questions that should be addressed in future investigations. The following recommendations are proposed to extend, validate, and build upon the results obtained:

In summary, the findings of this study establish a solid and comprehensive metagenomic foundation for future investigations of the termite gut microbiome. The priority recommendations for immediate follow-up studies are:

(i) Whole-genome shotgun metagenomics and long-read sequencing to overcome the assembly and gene-length limitations of the present short-read dataset;

(ii) MAG binning to obtain genome-resolved functional annotation for dominant community members including *Acetivibrio*, *Treponema*, and *Fibrobacter* spp.

(iii) Heterologous expression and biochemical characterisation of the GH113 mannanases identified as the dominant CAZyme activity;

(iv) Metatranscriptomics to validate MetaCyc pathway activity predictions in vivo; and

(v) $^{15}N_2$ isotope labelling to quantify nitrogen fixation and transfer in the termite gut directly. Together, these future directions will transform the descriptive metagenomic baseline presented here into a mechanistic, experimentally validated understanding of one of the most ecologically important and biotechnologically promising microbial ecosystems on Earth.

Bibliography

[1] A. T. Lima, G. de Oliveira, A. A. Souza, A. C. Lopes, R. H. Santana, P. Istvan, B. F. Quirino, J. Barbosa, S. de Freitas, A. V. Garay, and R. H. Krüger, "Termites and their gut microbiome in animal nutrition: Advances and biotechnological applications," *Anim. Nutr.*, vol. 23, pp. 527–534, Aug 2025.

[2] R. Xie, B. Danso, and J. Sun, "Metagenomics analysis of bacterial community structure from wood- and soil-feeding termites: Metabolic pathways and functional structures toward the degradation of lignocellulose and recalcitrant compounds," *Front. Microbiol.*, vol. 15, p. 1424982, 2024.

[3] G. Korsa, A. Beyene, and A. Ayele, "Bacterial diversity from soil-feeding termite gut and their potential application," *Ann. Microbiol.*, vol. 73, p. 38, 2023.

[4] A. Arab, L. C. Paulino, and M. S. L. Rolim, "The gut microbiome and lignocellulose digestion in constrictotermes cyphergaster," *Diversity*, vol. 16, no. 10, p. 623, 2024.

[5] M. A. Dar, R. Xie, and L. Jing, "Elucidating the structure and composition of bacterial symbionts in the gut regions of wood-feeding termite coptotermes formosanus and their functional profile toward lignocellulolytic systems," *Front. Microbiol.*, vol. 15, 2024.

[6] J. F. M. Salgado, V. Hervé, and M. A. G. Vera, "Unveiling lignocellulolytic potential: A genomic exploration of bacterial lineages within the termite gut," *Microbiome*, vol. 12, p. 201, 2024.

[7] R. Xie, B. Danso, and J. Sun, "Biorefinery and bioremediation strategies for efficient management of recalcitrant pollutants using termites as an obscure yet promising source of bacterial gut symbionts: A review," *Insects*, vol. 15, no. 11, p. 908, 2024.

[8] R. Xie, H. M. Zabed, S. Ali, D. Zhu, and J. Sun, "Termite microbial symbiosis as a model for innovative design of lignocellulosic future biorefinery: Current paradigms and future perspectives," *Biomass*, vol. 4, pp. 180–201, 2024.

[9] Z. Zhang, K. Wang, and C. Zou, "Comparison of microbial diversity and carbohydrate-active enzymes in the hindgut of two wood-feeding termites," *BMC Microbiol.*, vol. 24, p. 470, 2024.

[10] L. Wang, Y. Huang, and M. Poulsen, "Termites and subsocial roaches inherited many bacterial-borne carbohydrate-active enzymes (cazymes) from their common ancestor," *Commun. Biol.*, vol. 7, 2024.

[11] M. Marynowska, X. Goux, and D. Sillam-Dussès, "Compositional and functional characterisation of biomass-degrading microbial communities in guts of plant fibre- and soil-feeding higher termites," *Microbiome*, vol. 8, p. 96, 2020.

[12] M. Romero Victorica, M. A. Soria, R. A. Batista-García, J. A. Ceja-Navarro, S. Vikram, M. Ortiz, O. Ontañon, S. Ghio, L. Martínez-Ávila, O. J. Quintero García *et al.*, "Neotropical termite microbiomes as sources of novel plant cell wall degrading enzymes," *Scientific reports*, vol. 10, no. 1, p. 3864, 2020.

[13] M. Zaman, I. A. Khan, A. Usman, and A.-U.-R. Saljoqi, "Species diversity and damage incidence of termites (isoptera) in different agro-ecological zones of khyber pakhtunkhwa, pakistan." *Sarhad Journal of Agriculture*, vol. 38, no. 2, p. 518, 2022.

[14] M. S. Engel, C. Y. Liang, D. A. Grimaldi, and K. Krishna, "Termites (isoptera): their phylogeny, classification, and rise to ecological dominance," *Am. Mus. Novit.*, vol. 2023, no. 4017, pp. 1–46, 2023.

[15] W. Deng, K. Drager, X.-R. Li, W. Han, M. Djernæs, Y. Wang, J. Martin, J. Schwartz, Z. Zhang, T. G. L. Kovacs, N. Lo, S. Hellemans, Y.-L. Che, T. Bourguignon, Z. Wang, and D. Evangelista, "Blattodea systematics in the genomic age," *Insect Syst. Divers.*, vol. 10, no. 1, p. ixaf042, Jan 2026.

[16] R. Xie, B. Danso, J. Sun, R. Al-Tohamy, M. A. Khalil, M. Schagerl, and S. S. Ali, "Metagenomics analysis of bacterial community structure from wood- and soil-feeding termites: metabolic pathways and functional structures toward the degradation of lignocellulose and recalcitrant compounds," *Front. Microbiol.*, vol. 15, pp. 1–10, 2024.

[17] S. Muarrif, S. Samadi, J. Jauharlina, and Syaukani, "Description of the termite genus prohamitermes (termitinae) with new characters from the leuser ecosystem, sumatra, indonesia," *Biodiversitas*, vol. 24, no. 11, pp. 6119–6125, 2023.

[18] P. Eggleton, "The ecology of termites and their role in ecosystem functioning," *Curr. Opin. Insect Sci.*, vol. 48, pp. 1–8, 2021.

[19] H. Ci, C. Guo, B.-L. Sai, B. Tuo, W.-W. Zhao, H.-T. Qin, T. Zhang, E.-R. Yan, and J. H. C. Cornelissen, "The wood economics spectrum modulates the positive effects of termite foraging intensity on deadwood invertebrate diversity," *Funct. Ecol.*, vol. 39, no. 8, pp. 1–14, 2025.

[20] S. Liu, S. Xia, D. Wu, J. E. Behm, Y. Meng, H. Yuan, P. Wen, A. C. Hughes, and X. Yang, "Understanding global and regional patterns of termite diversity and regional functional traits," *iScience*, vol. 25, no. 12, p. 105538, Dec 2022.

[21] P. Duquesne and C. Fournier, "Urban and agricultural areas under threat of the termite pest genus heterotermes: Insights from species distribution modelling and phylogeny," 2024.

[22] Anonymous, "Historical and present-day records on african termites: Revisiting the diversity and distribution of macrotermes in namibia after six decades," *Insectes Soc.*, vol. 72, pp. 1–18, 2025.

[23] M. Arumugam, S. Sundararaju, S. Jagadesan, P. Gunasekaran, and J. Rajendhran, "Metagenomic analysis of microbial community affiliated with termitarium reveals high lignocellulolytic potential," *Curr. Microbiol.*, vol. 78, pp. 1551–1565, 2021.

[24] D. M. Tadesse, B. G. Debelo, and T. M. Negeri, "Effects of land-use systems on termite diversity and abundance in ethiopian agroecosystems," *Environ. Monit. Assess.*, vol. 196, no. 2, p. 136, 2024.

[25] Z. Khan, M. S. Khan, S. Bawazeer, N. Bawazeer, Suleman, M. Irfan, A. Rauf, and X.-H. Su, "A comprehensive review on the documented characteristics of four reticulitermes termites (rhinotermitidae, blattodea) of china," *Braz. J. Biol.*, vol. 84, p. e256354, 2024.

[26] F. Ahmad, H. Fouad, S.-Y. Liang, Y. Hu, and J.-C. Mo, "Termites and chinese agricultural system: Applications and advances in integrated termite management and chemical control," *Insect Sci.*, vol. 28, no. 1, pp. 2–20, Feb 2021.

[27] S. Zhao, Y. Wang, B. Bie, L. Xu, and H. Bai, "Study on the genetic diversity of reticulitermes aculabialis," in *Proc. 2nd Int. Conf. Mater. Chem. Environ. Prot.*, Sanya, China, Nov 2018, pp. 255–259.

[28] P. Jouquet, S. Traoré, C. Choosai, C. Hartmann, and D. Bignell, "Influence of termites on ecosystem functioning and soil dynamics: recent advances," *Biol. Rev.*, vol. 97, no. 6, pp. 2143–2160, 2022.

[29] P. A. V. Borges, S. Bettencourt, D. Vargas, R. Medeiros, J. Melo, and A. Rodrigues, "A 13-year termite (insecta, blattodea) monitoring programme in the azores: Dataset and findings," *Biodivers. Data J.*, vol. 13, p. e164834, 2025.

[30] L. Gray and M. M. Webster, "False alarms and information transmission in grouping animals," *Biol. Rev.*, vol. 98, no. 3, pp. 833–848, Jun 2023.

[31] A. Brune and M. Dietrich, "The gut microbiota of termites: digesting the diversity in the light of ecology and evolution," *Annu. Rev. Microbiol.*, vol. 76, pp. 499–522, 2022.

[32] A. Brune, "Symbiotic digestion of lignocellulose in termite guts," *Nat. Rev. Microbiol.*, vol. 22, no. 3, pp. 163–179, 2024.

[33] M. Marynowska, X. Goux, and D. Sillam-Dussès, "Compositional and functional characterisation of biomass-degrading microbial communities in guts of plant fibre- and soil-feeding higher termites," *Microbiome*, vol. 8, p. 96, 2020.

[34] B. J. Enagbonma, R. Mmushi, and O. O. Babalola, "Biotechnological utilization: The potential role of the termite gut symbiotic microbiome," *Symbiosis*, vol. 95, no. 3, pp. 307–316, Apr 2025.

[35] J. Arora, A. Buček, S. Hellemans, T. Beránková, J. R. Arias, B. L. Fisher, C. Clitheroe, A. Brune, Y. Kinjo, J. Šobotník, and T. Bourguignon, "Evidence of cospeciation between termites and their gut bacteria on a geological time scale," *Proc. R. Soc. B*, vol. 290, p. 20230619, 2023.

[36] C. A. Nalepa, "What kills the hindgut flagellates of lower termites during the host molting cycle?" *Microorganisms*, vol. 5, no. 4, p. 82, 2017.

[37] S. G. Aguilar, J. Shevat, D. E. Jasso-Selles, K. L. Swichtenberg, C. D. Vecco-Giove, J. Šobotník, D. Sillam-Dussès, F. De Martini, and G. H. Gile, "Distinct patterns of co-evolution among protist symbionts of neoisoptera termites," *Diversity*, vol. 17, no. 8, p. 537, Jul 2025.

[38] H. Q. Loh, V. Hervé, and A. Brune, "Metabolic potential for reductive acetogenesis and a novel energy-converting [nife] hydrogenase in bathyarchaeia from termite guts: A genome-centric analysis," *Front. Microbiol.*, vol. 11, p. 635786, Jan 2021.

[39] L. Newton, E. Tolman, M. Kohli, and J. L. Ware, "Evolution of odonata: genomic insights," *Curr. Opin. Insect Sci.*, vol. 58, p. 101073, 2023.

[40] J. Schierstaedt, S. Jechalke, J. Nesme, K. Neuhaus, S. J. Sørensen, R. Grosch, and K. Smalla, "Salmonella persistence in soil depends on reciprocal interactions with indigenous microorganisms," *Environ. Microbiol.*, vol. 22, no. 7, pp. 2639–2652, Jul 2020.

[41] J. Arora, A. Buček, S. Hellemans, T. Beránková, J. R. Arias, B. L. Fisher, C. Clitheroe, and A. Brune, "Evidence of cospeciation between termites and their gut bacteria on a geological time scale," *Proc. R. Soc. B Biol. Sci.*, vol. 290, no. 2003, p. 20222411, Jul 2023.

[42] G. H. Gile, "Protist symbionts of termites: Diversity, distribution, and coevolution," *Biol. Rev.*, vol. 99, no. 2, pp. 622–652, Apr 2024.

[43] K. L. Swichtenberg, D. E. Jasso-Selles, L. A. Nguyen, S. G. Aguilar, K. B. McAnally, and S. J. Taerum, "Anacanthotermes symbionts clarify the evolutionary history of termite-symbiotic parabasalia," *Front. Ecol. Evol.*, vol. 14, p. 1615161, 2026.

[44] S. G. Aguilar, J. Shevat, D. E. Jasso-Selles, K. L. Swichtenberg, C. D. Vecco-Giove, J. Šobotník, D. Sillam-Dussès, F. De Martini, and G. H. Gile, "Distinct patterns of co-evolution among protist symbionts of neoisoptera termites," *Diversity*, vol. 17, no. 8, p. 537, Aug 2025.

[45] B. J. Enagbonma, R. Mmushi, and O. O. Babalola, "Biotechnological utilization: The potential role of the termite gut symbiotic microbiome," *Symbiosis*, vol. 95, no. 3, pp. 307–316, Apr 2025.

[46] E. T. Sero, A. Amobonye, P. Bhagwat, N. K. Singh, J. C. de Carvalho, K. Permaul, and S. Pillai, "Exploring the termite gut as a hub of industrially important microbes and enzymes for biofuel production," *Biomass Bioenergy*, 2025.

[47] M. A. Benjamino and A. Graf, "Gut transit dynamics and microbial activity in termites," *Appl. Environ. Microbiol.*, vol. 87, no. 18, 2021.

[48] D. E. Tokuda and N. Watanabe, "Hidden cellulases in termite gut systems," *Front. Microbiol.*, vol. 12, p. 641953, 2021.

[49] S. Mikaelyan, T. Strassert, and A. Brune, "The fibre-associated microbiome in termite digestion," *ISME J.*, vol. 15, pp. 172–184, 2021.

[50] S. Huang, "Comparative metagenomics reveals microbial diversity in termite guts," *Front. Microbiol.*, vol. 13, 2022.

[51] A. Benjamino and M. Graf, "Transmission mechanisms of insect gut microbiota," *ISME Commun.*, vol. 2, p. 43, 2022.

[52] T. Inoue and T. Noda, "Protists in termite and cockroach digestion," *Protist*, vol. 171, no. 4, 2020.

[53] T. Ikeda-Ohtsubo, "Unique protistan lineages associated with termites," *Environ. Microbiol.*, vol. 23, pp. 541–557, 2021.

[54] H. König and A. Varma, *Intestinal Microorganisms of Termites and Other Invertebrates*. Berlin, Germany: Springer, 2006.

[55] S. Urbina, "Novel xylose-utilizing yeasts isolated from termites," *J. Fungi*, vol. 7, p. 401, 2021.

[56] N. Mathew, "Lignocellulose-degrading fungi from termite guts," *Front. Bioeng. Biotechnol.*, vol. 10, 2022.

[57] S. Mikaelyan, "Diet is the primary determinant of bacterial community structure in termites," *Microbiome*, vol. 5, p. 81, 2017.

[58] T. Rahman, "Functional bacterial shifts across termite species," *Environ. Microbiol. Rep.*, vol. 15, 2023.

[59] S. Benjamino and A. Graf, "Characterization of microbial metabolic functions in termite gut ecosystems through metagenomics and metatranscriptomics," *Front. Microbiol.*, vol. 13, pp. 1–15, 2022.

[60] G. H. Gile, "Protist symbionts of termites: diversity, distribution, and coevolution," *Biol. Rev.*, vol. 99, no. 2, pp. 622–652, Apr 2024.

[61] L. R. Cleveland, "Symbiosis between termites and their intestinal protozoa," *Biol. Bull.*, vol. 46, pp. 177–227, 1924.

[62] M. Noda, S. Kitade, and T. Inoue, "Phylogenetic diversity of termite gut protists and their ecological roles," *Microorganisms*, vol. 8, no. 12, pp. 1–19, 2020.

[63] M. He, "Hydrogen metabolism and cellulose degradation by parabasalid protists in lower termites," *ISME J.*, vol. 16, pp. 1007–1019, 2022.

[64] A. Arora and S. Sharma, "Bacterial community structure in termite gut microbiota revealed by high-throughput sequencing," *Front. Ecol. Evol.*, vol. 9, 2021.

[65] A. Brune, "Symbiotic digestion of lignocellulose in termite guts," *Nat. Rev. Microbiol.*, vol. 22, pp. 157–173, 2024.

[66] T. Otani, "Functional roles of bacterial communities in termite digestion and nitrogen cycling," *mSystems*, vol. 7, no. 4, 2022.

[67] S. Abdul Rahman, "Archaeal diversity and methane metabolism in termite gut microbiomes," *Microbiome*, vol. 10, pp. 1–18, 2022.

[68] M. Grigorescu, "Viral ecology and bacteriophage diversity in termite gut microbiomes," *Viruses*, vol. 14, no. 11, 2022.

[69] Y. Chen, "Interactions between bacteriophages and microbial communities in insect digestive systems," *Front. Cell. Infect. Microbiol.*, vol. 13, 2023.

[70] B. Henrissat and P. M. Coutinho, "Carbohydrate-active enzymes and their applications in biomass conversion," *Curr. Opin. Biotechnol.*, vol. 73, pp. 108–118, 2022.

[71] J. Li, "Comparative metagenomics of lignocellulose-degrading enzymes across termite feeding groups," *Biotechnol. Biofuels*, vol. 15, p. 221, 2022.

[72] S. N. Yadav, "Cellulolytic microorganisms from termite gut and their biotechnological applications," *Bioresour. Technol. Rep.*, vol. 19, 2022.

[73] M. A. Khan, "Isolation and characterization of endoglucanase-producing bacteria from termite guts," *J. Appl. Microbiol.*, vol. 132, no. 2, pp. 1134–1147, 2022.

[74] L. Zhao, "Characterization of exoglucanase-producing bacteria isolated from termite gut systems," *Biotechnol. Lett.*, vol. 45, pp. 1105–1116, 2023.

[75] S. Saxena, J. Bahadur, and A. Varma, "Production and localisation of carboxymethylcellulase, xylanase and β-glucosidase from cellulomonas and micrococcus spp," *Appl. Microbiol. Biotechnol.*, vol. 34, pp. 668–670, 1991.

[76] C. Tarayre, J. Bauwens, C. Brasseur, C. Mattéotti, C. Millet, P. A. Guiot, J. Destain, M. Vandenbol, D. Portetelle, E. De Pauw, E. Haubruge, F. Francis, and P. Thonart, "Isolation and cultivation of xylanolytic and cellulolytic sarocladium kiliense and trichoderma virens from the gut of the termite reticulitermes santonensis," *Environ. Sci. Pollut. Res.*, vol. 22, pp. 4369–4382, 2015.

[77] Z. Bashir, V. K. Kondapalli, N. Adlakha, A. Sharma, R. K. Bhatnagar, G. Chandel, and S. S. Yazdani, "Diversity and functional significance of cellulolytic microbes living in termite, pill-bug and stem-borer guts," *Sci. Rep.*, vol. 3, p. 2558, 2013.

[78] S. S. Ali, R. Al-Tohamy, J. Sun, J. Wu, and L. Huizi, "Screening and construction of a novel microbial consortium ssa-6 enriched from the gut symbionts of wood-feeding termite, coptotermes formosanus and its biomass-based biorefineries," *Fuel*, vol. 236, pp. 1128–1145, 2019.

[79] S. S. Ali, J. Wu, R. Xie, F. Zhou, and J. Sun, "Screening and characterizing of xylanolytic and xylose-fermenting yeasts isolated from the wood-feeding termite, reticulitermes chinensis," *PLoS One*, vol. 12, p. e0181141, 2017.

[80] S. Tiwari, R. Avchar, R. Arora, V. Lanjekar, P. K. Dhakephalkar, S. S. Dagar, and A. Baghela, "Xylanolytic and ethanologenic potential of gut associated yeasts from different species of termites from india," *Mycology*, vol. 48, pp. 501–511, 2020.

[81] G. M. Kalu-Uka, S. Kumar, A. C. Kalu-Uka, S. Vikram, O. O. Okorafor, M. Kigozi, G. O. Ihekweme, and A. P. Onwualu, "Prospects for biodiesel

production from macrotermes nigeriensis: process optimization and characterization of biodiesel properties," *Biomass Bioenergy*, vol. 146, p. 105980, 2021.

[82] Y. Hirakata, M. Hatamoto, M. Oshiki, T. Watari, N. Araki, and T. Yamaguchi, "Food selectivity of anaerobic protists and direct evidence for methane production using carbon from prey bacteria by endosymbiotic methanogen," *ISME J.*, vol. 14, pp. 1873–1885, 2020.

[83] W. A. Abdel Wahab, F. A. Mostafa, S. A. Ahmed, and S. A. A. Saleh, "Statistical optimization of enzyme cocktail production using jew's mallow stalks residues by a new isolate aspergillus flavus b2 via statistical strategy and enzymes characterization," *J. Biotechnol.*, vol. 367, pp. 89–97, 2023.

[84] C. Liang, Q. Wang, W. Wang, C. S. K. Lin, Y. Hu, and W. Qi, "Enhancement of an efficient enzyme cocktail from penicillium consortium on biodegradation of pretreated poplar," *Chem. Eng. J.*, vol. 452, p. 139352, 2023.

[85] P. S. Oliveira, A. Brune, and C. Dietrich, "Comparative analysis of hemicellulose degradation by microbial symbionts from termite guts," *Front. Microbiol.*, vol. 14, pp. 1–18, 2023.

[86] R. K. Sharma, M. Gupta, and S. Singh, "Termite gut microorganisms and lignocellulosic bioconversion for sustainable bioethanol production," *Renew. Sustain. Energy Rev.*, vol. 181, p. 113367, 2023.

[87] S. Abdul Rahman, K. Ohkuma, and A. Brune, "Methanogenic archaea and methane production pathways in termite gut ecosystems," *Microbiome*, vol. 10, pp. 1–17, 2022.

[88] T. Inoue, M. Noda, and K. Ohkuma, "Hydrogen metabolism and biohydrogen production potential in termite gut microbial communities," *Front. Energy Res.*, vol. 11, p. 1189012, 2023.

[89] W. A. Abdel Wahab, F. A. Mostafa, S. A. Ahmed, and S. A. A. Saleh, "Statistical optimization of enzyme cocktail production using jew's mallow

stalks residues by a new isolate aspergillus flavus b2 via statistical strategy and enzymes characterization," *J. Biotechnol.*, vol. 367, pp. 89–97, 2023.

[90] D. Bueno, C. de Freitas, and M. Brienzo, "Enzymatic cocktail formulation for xylan hydrolysis into xylose and xylooligosaccharides," *Molecules*, vol. 28, p. 624, 2023.

[91] S. S. Ali, R. Al-Tohamy, E. Koutra, A. H. El-Naggar, M. Kornaros, and J. Sun, "Valorizing lignin-like dyes and textile dyeing wastewater by a newly constructed lipid-producing and lignin modifying oleaginous yeast consortium valued for biodiesel and bioremediation," *J. Hazard. Mater.*, vol. 403, p. 123575, 2021.

[92] S. Andrews, "Fastqc: A quality control tool for high throughput sequence data," Cambridge, UK, 2010.

[93] D. Li, C. M. Liu, R. Luo, K. Sadakane, and T. W. Lam, "Megahit: An ultra-fast single-node solution for large and complex metagenomics assembly," *Bioinformatics*, vol. 31, no. 10, pp. 1674–1676, 2015.

[94] F. Van der Jeugt, P. Dawyndt, and B. Mesuere, "Fraggenescanrs: faster gene prediction for short reads," *BMC Bioinf.*, vol. 23, p. 198, 2022.

[95] J. Cantalapiedra, "eggnog-mapper v2: Functional annotation, orthology assignments, and domain prediction at metagenomic scale," *Mol. Biol. Evol.*, vol. 38, no. 12, pp. 5825–5829, 2021.

[96] V. Lombard, H. Golaconda Ramulu, E. Drula, P. Coutinho, and B. Henrissat, "The carbohydrate-active enzymes database (cazy) in 2021," *Nucleic Acids Res.*, vol. 49, no. D1, pp. D490–D495, 2021.

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

[98] Y. Hongoh, "Toward the functional analysis of uncultivable, symbiotic microorganisms in the termite gut," *Cell. Mol. Life Sci.*, vol. 68, no. 8, pp. 1311–1325, 2011.

[99] A. Brune and C. Dietrich, "The gut microbiota of termites: Digesting the diversity in the light of ecology and evolution," *Annu. Rev. Microbiol.*, vol. 69, pp. 145–166, 2015.

[100] J. A. Breznak and A. Brune, "Role of microorganisms in the digestion of lignocellulose by termites," *Annu. Rev. Entomol.*, vol. 39, pp. 453–487, 1994.

[101] M. Pester and A. Brune, "Hydrogen is the central free intermediate during lignocellulose degradation by termite gut symbionts," *ISME J.*, vol. 1, no. 6, pp. 551–565, 2007.

[102] J. R. Leadbetter, T. M. Schmidt, J. R. Graber, and J. A. Breznak, "Acetogenesis from h2 plus co2 by spirochetes from termite guts," *Science*, vol. 283, no. 5402, pp. 686–689, 1999.

[103] T. C. Lilburn, "Nitrogen fixation by symbiotic and free-living spirochetes," *Science*, vol. 292, no. 5526, pp. 2495–2498, 2001.

[104] J. R. Graber and J. A. Breznak, "Physiology and nutrition of treponema primitia, an h2/co2-acetogenic spirochete from termite hindguts," *Appl. Environ. Microbiol.*, vol. 70, no. 3, pp. 1307–1314, 2004.

[105] F. Thomas, J.-H. Hehemann, E. Rebuffet, M. Czjzek, and G. Michel, "Environmental and gut bacteroidetes: The food connection," *Front. Microbiol.*, vol. 2, p. 93, 2011.

[106] M. A. Yamin, "Cellulose metabolism by the flagellate trichonympha from a termite is independent of endosymbiotic bacteria," *Science*, vol. 211, no. 4477, pp. 58–59, 1981.

[107] P. Cao, "Xenobiotic degradation and phage invasion dynamics in human gut microbiomes," *Cell Host Microbe*, vol. 29, no. 8, p. 1292, 2021.

[108] J. S. Dumler, "Reorganization of genera in the families rickettsiaceae and anaplasmataceae in the order rickettsiales: Unification of some species of ehrlichia with anaplasma," *Int. J. Syst. Evol. Microbiol.*, vol. 51, no. 6, pp. 2145–2165, 2001.

[109] J. J. Werner, "Impact of training sets on classification of high-throughput bacterial 16s rrna gene surveys," *ISME J.*, vol. 6, no. 1, pp. 94–103, 2012.

[110] R. Poretsky, "Strengths and limitations of 16s rrna gene amplicon sequencing in revealing temporal microbial community dynamics," *PLoS ONE*, vol. 9, no. 4, p. e93827, 2014.

[111] X.-F. Huang, "Bacillus-based bioinoculants in agricultural ecosystems: A review," *Biomass Bioenergy*, vol. 35, no. 9, pp. 3813–3825, 2011.

[112] Y. Xiao, "Paenibacillus and related genera in the termite gut environment," *Insect Sci.*, vol. 19, no. 4, pp. 469–485, 2012.

[113] M. Ohkuma and A. Brune, "Diversity, structure, and evolution of the termite gut microbial community," in *Biology of Termites: A Modern Synthesis*. Dordrecht, The Netherlands: Springer, 2011, pp. 413–438.

[114] U. Stingl, "'endomicrobia': Cytoplasmic symbionts of termite gut protozoa form a separate phylum of prokaryotes," *Appl. Environ. Microbiol.*, vol. 71, no. 3, pp. 1473–1479, 2005.

[115] E. Ransom-Jones, D. L. Jones, A. J. McCarthy, and J. E. McDonald, "The fibrobacteres: An important phylum of cellulose-degrading bacteria," *Microb. Ecol.*, vol. 63, no. 2, pp. 267–281, 2012.

[116] O. Geissinger, "The ultramicrobacterium 'elusimicrobium minutum' gen. nov., sp. nov., the first cultivated representative of the termite group 1 phylum," *Appl. Environ. Microbiol.*, vol. 75, no. 9, pp. 2831–2840, 2009.

[117] C. A. Lozupone, "Diversity, stability and resilience of the human gut microbiota," *Nature*, vol. 489, pp. 220–230, 2012.

[118] E. H. Simpson, "Measurement of diversity," *Nature*, vol. 163, p. 688, 1949.

[119] A. Brune, "Termite guts: The world's smallest bioreactors," *Trends Biotechnol.*, vol. 16, no. 1, pp. 16–21, 1998.

[120] M. L. Sogin, "Microbial diversity in the deep sea and the underexplored 'rare biosphere'," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 103, no. 32, pp. 12 115–12 120, 2006.

[121] P. Yilmaz, "The silva and 'all-species living tree project (ltp)' taxonomic frameworks," *Nucleic Acids Res.*, vol. 42, no. D1, pp. D643–D648, 2014.

[122] M. Ohkuma, "Termite symbiotic systems: Efficient bio-recycling of lignocellulose," *Appl. Microbiol. Biotechnol.*, vol. 61, no. 1, pp. 1–9, 2003.

[123] R. E. Hungate, *The Rumen and Its Microbes*. New York, NY, USA: Academic Press, 1966.

[124] A. Sy, "Methylobacterium spp. as plant epiphytes and termite associates," *Int. J. Syst. Evol. Microbiol.*, vol. 51, no. 5, pp. 1729–1739, 2001.

[125] R. L. Tatusov, "The cog database: New developments in phylogenetic classification of proteins from complete genomes," *Nucleic Acids Res.*, vol. 29, no. 1, pp. 22–28, 2001.

[126] M. Y. Galperin, "Expanded microbial genome coverage and improved protein family annotation in the cog database," *Nucleic Acids Res.*, vol. 49, no. D1, pp. D274–D281, 2021.

[127] S. G. Tringe, "Comparative metagenomics of microbial communities," *Science*, vol. 308, no. 5721, pp. 554–557, 2005.

[128] L. R. S. Moreira and E. X. F. Filho, "An overview of mannan structure and mannan-degrading enzyme systems," *Appl. Microbiol. Biotechnol.*, vol. 79, no. 2, pp. 165–178, 2008.

[129] G. Tokuda, "Major alteration of the expression site of endogenous cellulases in members of an apical termite lineage," *Mol. Ecol.*, vol. 21, no. 13, pp. 3219–3233, 2012.

[130] F. Warnecke, "Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite," *Nature*, vol. 450, pp. 560–565, 2007.

[131] P. M. Coutinho, "An evolving hierarchical family classification for glycosyltransferases," *J. Mol. Biol.*, vol. 328, no. 2, pp. 307–317, 2003.

[132] J. Preiss, "Bacterial glycogen synthesis and its regulation," *Annu. Rev. Microbiol.*, vol. 38, pp. 419–458, 1984.

[133] B. L. Cantarel, "The carbohydrate-active enzymes database (cazy): An expert resource for glycogenomics," *Nucleic Acids Res.*, vol. 37, no. D1, pp. D233–D238, 2009.