
# Exploring temporal changes in the malting barley seed microbiome with meta-omics to understand nitrogen content effects.

by

**Kalonji Abondance Tshisekedi**

**(1423566)**

**Thesis**

Submitted in fulfilment of the requirements for the degree

**Philosophiae Doctor**

in

**Molecular and Cell Biology**

in the Faculty of Science, University of the Witwatersrand, Johannesburg, South Africa

Supervisor: Dr. Angela Botes

Co- Supervisor: Prof. Pieter De Maayer

October 2024

**Declaration**

I, Kalonji Abondance Tshisekedi (1423566), am a student registered for the degree of Philosophiae Doctor by Thesis in the academic year 2021-2024.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where I have explicitly indicated otherwise.
- I have not submitted this work before for any other degree or examination at this or any other University.
- The information used in the Dissertation HAS NOT been obtained by me while employed by, or working under the aegis of, any person or organisation other than the University.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is in not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

NRF Declaration

The financial assistance of the National Research Foundation (NRF) and Anheuser-Busch InBev SA/NV (AB InBev) towards this research is hereby acknowledged.

Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be contributed to the NRF and/or AB InBev.

## DEDICATION

*This thesis is dedicated to my parents.*

**Chers Papa et Maman, les mots ne seront jamais suffisants pour dire merci. Votre amour et soutien ont sculpté ma personne. Chaque jour, je m'efforce d'être le reflet de votre bonté.**

## ACKNOWLEDGEMENTS

I am deeply grateful for the paths that have led me to this thesis. This journey would not have been possible without the grace of God, who has blessed me with countless opportunities.

I would like to extend my heartfelt thanks to Professor Rob Veale, who took a chance on me. This decision has left an indelible mark, and I am forever grateful.

To my supervisor, Dr. Angela Botes, and my co-supervisor, Prof. Pieter De Maayer, I could not have asked for better mentors. Thank you for your trust and confidence in me, for challenging me to step out of my comfort zone, and for steering me away from the danger zone whenever I veered off course. I hope to continue learning from you.

I am immensely grateful to Anheuser-Busch InBev and the National Research Foundation (NRF) for their generous funding and support, which were instrumental in the completion of this research.

To the past and present members of the Botes Lab, thank you for your assistance, encouragement, and for always checking up on me.

To Nondumiso, thank you for believing in me, supporting me through your prayers, and most importantly, for becoming my home.

To Jouissance, my little sister, all these years I thought I was taking care of you, but it turns out you were the one taking care of me. I wouldn't be here without your support.

To my friend, Uncle Graham Hewett. Your consistent check-ins over the last decade have meant the world to me.

To the Weru family, thank you for being my home away from home. I think of Clare and David often, and may their souls continue to rest in peace.

To all my friends, thank you for always being there for me.

To my siblings, thank you for your unwavering support.

Lastly, to my parents, thank you for never giving up on me and for being the best parents I could have ever asked for.

## Research outputs

### Journal Article

**Tshisekedi**, K.A., De Maayer, P. & Botes, A. Metagenomic sequencing and reconstruction of 82 microbial genomes from barley seed communities. Sci Data 11, 484 (2024). https://doi.org/10.1038/s41597-024-03332-x

### Plenary talks

**Tshisekedi KA,** De Maayer P, Botes A (2023). Elucidating Nitrogen Cycling and Microbial Adaptation in Malting Barley Grain: A Meta-Omics Approach. South African Society for Microbiology (SASM) conference. Stellenbosch, September 2023. **Award:** Best Oral Presentation 2nd place.

### Conference outputs – Oral presentations

**Tshisekedi, K.A**., De Maayer P., and Botes A (2023). Unveiling the role of nitrogen in shaping microbial communities and malting quality in barley grain through meta-omics analysis. International Symposium on Microbial Ecology (ISME) #19. Cape Town. 18th Aug 2024 – 23rd Aug 2024 **Award:** African Travel Grants.

### Conference outputs – Poster presentations

1. **Tshisekedi, K.A**., De Maayer P., and Botes A (2023). Exploring Nitrogen-Induced Changes in Barley Microbial Communities through Genome-resolved metagenomics. FEMS Conference on Microbiology 2023, 8th -13th July 2023.
2. **Tshisekedi, K.A**., De Maayer P., and Botes A (2023). Exploring Nitrogen-Induced Changes in Barley Microbial Communities through Genome-resolved metagenomics. Ubuntu Proteomics Summer School 2023. 19th February 2023. **Award:** Best poster

## Table of Contents

## List of figures

## List of tables

## List of abbreviations

AA: Auxiliary Activities

AB InBev: Anheuser-Busch InBev

AI: Artificial intelligence

ANI: Average Nucleotide Identity

ANOSIM: ANalysis Of SIMilarities

BLASTP: Basic Local Alignment Search Tool for Proteins

CAZymes: Carbohydrate-Active Enzymes

CBM: Carbohydrate-Binding Modules

CDS: Coding DNA Sequences

CE: Carbohydrate Esterases

COGs: Clusters of Orthologous Groups

CSIR: Council for Scientific and Industrial Research

DDA: Data-Dependent Acquisition

DNA: Deoxyribonucleic Acid

dsDNA: Double-Stranded DNA

DTT: Dithiothreitol

E.C.: Enzyme Commission numbers

FDR: False Discovery Rate

FHB: *Fusarium* Head Blight

GH: Glycoside Hydrolases

GO: Gene Ontology

GT: Glycosyl Transferases

GTDB-Tk: Genome Taxonomy Database Toolkit

HH: Harvest, High grain nitrogen content

HL: Harvest, Low grain nitrogen content

KEGG: Kyoto Encyclopedia of Genes and Genomes

KO: KEGG Orthology terms

LAB: Lactic Acid Bacteria

LCA: Lowest Common Ancestor

LC-MS: Liquid Chromatography-Mass Spectrometry

MAGs: Metagenome-Assembled Genomes

Mb: Megabases

mDNA: Metagenomic DNA

MIMAG: Minimum Information about a Metagenome-Assembled Genome

MPA: MetaPhlAn Profile Analysis

mRNA: Messenger RNA

NCBI: National Center for Biotechnology Information

NGS: Next-Generation Sequencing

NMDS: Non-metric Multidimensional Scaling

OMPs: Outer Membrane Proteins

ORFs: Open Reading Frames

PAGE: Polyacrylamide Gel Electrophoresis

PBS: Phosphate-Buffered Saline

PCoA: Principal Coordinates Analysis

PCR: Polymerase Chain Reaction

PHI: Pathogen Host Interaction database

PL: Polysaccharide Lyases

RNA: Ribonucleic Acid

ROS: Reactive Oxygen Species

rRNA: ribosomal RNA

S3H: 3 months storage, High grain nitrogen content

S3L: 3 months storage, Low grain nitrogen content

S6H: 6 months storage, High grain nitrogen content

S6L: 6 months storage, Low grain nitrogen content

S9H: 9 months storage, High grain nitrogen content

S9L: 9 months storage, Low grain nitrogen content

SCOs: Single-Copy Orthogroups

SDS: Sodium Dodecyl Sulphate

tRNA: Transfer RNA

## Abstract

Barley (*Hordeum vulgare* L.) is a critical cereal crop, particularly in beer production, where it plays a significant role in the economy, especially in South Africa. Despite its importance, the barley seed microbiome, which affects seed storage and quality, is not well understood. This research addresses two key questions: (1) how microbial composition and function evolve during storage and (2) how the inherent nitrogen content of the grain affects these dynamics. Using metagenomic and metaproteomic approaches, eight barley samples from the Kadie cultivar, stored for various durations (harvest, three, six, and nine months) with high and low nitrogen content, were analysed. Metagenomic sequencing revealed a predominance of bacterial sequences and minimal fungal presence, with storage time having a greater impact on microbial diversity than nitrogen content. However, specific bacterial genera such as *Erwinia*, *Pantoea*, *Pseudomonas*, and *Stenotrophomonas* showed nitrogen-dependent prevalence. Metagenome-assembled genomes (MAGs) were reconstructed, representing 26 bacterial genera, with minimal shared orthologues, highlighting taxonomic diversity. Functional analysis identified key metabolic pathways and carbohydrate-active enzymes (CAZymes) essential for microbial adaptation during storage. Metaproteomic analysis further showed the active expression of proteins related to nutrient transport and stress response, indicating functional changes over storage time. Overall, this research enhances the understanding of the barley seed microbiome, providing valuable insights into storage practices that could improve brewing quality and agricultural sustainability.

## Preface

Barley (*Hordeum vulgare* L.) ranks as the fourth most widely cultivated cereal crop globally. Characterised by its short growth period and robust adaptability, barley is predominantly used in beer production, boosting the global economy through revenue generation and employment. In South Africa, the bulk of the barley harvest supports the brewing sector, making a substantial contribution to the nation's gross domestic product (GDP). Despite its economic relevance, the barley microbiome is the least explored among cereal crops. Two pivotal unresolved questions were addressed in this study using metagenomic and metaproteomic approaches 1) how the microbial composition and functional capabilities change with prolonged storage and 2) how the microbiome is influenced by the inherent nitrogen content of the barley grain. The microbiome of eight barley samples from the Kadie cultivar (obtained from Anheuser-Busch InBev storage facilities in the Western Cape Province, South Africa) were analysed. These samples were derived at four distinct time points, at harvest and after three, six and nine months of silo storage. Further, at each time point, one barley batch each with low or high grain nitrogen content were analysed.

**Chapter 1** provides a review of the existing literature related to beer production, focusing on the role of barley as a crucial ingredient. It begins with an overview of the beer brewing process and explains why barley is particularly suited for this purpose. The chapter also discusses the impact of grain nitrogen content on beer quality. Special attention is given to the barley seed microbiome, examining its characteristics both before and during storage. The dynamic changes in the microbiome throughout these stages are explored to highlight their implications for barley preservation.

**Chapter 2** presents the metagenomic sequencing of the barley seed microbiomes, which was conducted using the Illumina short read sequencing. Approximately 365.3 million reads were derived across all eight samples. Initial taxonomic classification of these metagenomic reads revealed a predominance of bacterial sequences, with only a minimal presence of fungal reads. Diversity analysis suggested that the duration of storage, rather than the nitrogen content of the grain, was the primary factor influencing microbial diversity associated with the barley seed.

**Chapter 3** describes the reconstruction of 82 high quality metagenome-assembled genomes (MAGs). These MAGs encompass four bacterial phyla and 26 genera, which form part of the core microbiome of the barley seeds. Comparative genomic analysis revealed a limited number

of shared single-copy orthologues across the MAGs, with only 66 common orthogroups detected. Additionally, a survey of plant pathogenicity-related proteins highlighted several MAGs that possess a substantial proportion of these proteins, offering insights into potential plant-microbe interactions.

In **Chapter 4** the metagenomes were assembled and annotated using a variety of *in silico* tools. While metagenomics can provide insights into the functional potential of the microbial community associated with barley seeds, it does not reflect their actual activity. Hence, metaproteomic analysis was undertaken. The findings of this thesis enrich our understanding of the barley seed microbiome and lay a foundation for future meta-omics studies of seed microbiomes.

**Chapter 5** synthesises the findings of the previous chapters, focusing on how the microbial dynamics of barley seeds are influenced by nitrogen content and storage duration. While the diversity analysis showed that storage time had a more significant impact than nitrogen content on microbial composition, nitrogen still played a role in shaping specific community structures. This chapter discusses the broader implications of these findings, particularly for storage practices and brewing quality. Additionally, it highlights the distinct behaviour of the six-month low nitrogen (S6L) sample, which presented unique assembly metrics and challenges for interpretation. Chapter 5 concludes by suggesting future research directions, emphasising the importance of biological replication, deeper investigation into microbial interactions, and the application of advanced techniques such as metametabolomics and machine learning to better understand and optimise storage conditions for barley seeds.

# Chapter 1 Literature review

## 1.1 What is beer?

Beer is deeply ingrained in human culture and ranks as the third most favoured drink globally, following water and tea (Bamforth, 2017; Collins *et al.*, 2023). Its popularity can be attributed to its distinct taste, aroma, and appearance, all of which result from a brewing process with ancient origins, dating back more than 8,000 years (Hornsey, 2003).

## 1.2 Overview of the beer making process

This process begins with the selection of grains, most notably barley, because of its high enzyme content (Magliano *et al.*, 2014); however, wheat, rye, oats, and corn, can also be used, adding distinct flavours and qualities to the beer (Wunderlich and Back, 2009). Once selected, the grain is malted, which involves soaking it in water, allowing it to germinate, and then drying it in a kiln (Figure 1). The malted grain is then mashed, a process where it is soaked in hot water to facilitate the enzymatic breakdown of starches into sugars (Schwarz and Li, 2010). The resulting liquid, known as wort, is boiled with hops for flavour before cooling (Briggs, 1998). Finally, yeast is introduced to the wort for fermentation, which converts the sugars into alcohol and carbon dioxide ($CO_2$), resulting in beer (Boulton, 2006). The quality of the grain, water, hops, and yeast, as well as any additional additives known as adjuncts, all affect the taste, aroma, and appearance of beer (Schwarz and Li, 2010). This process, although seemingly straightforward in concept, requires accuracy and attention to detail in order produce a product of high quality.

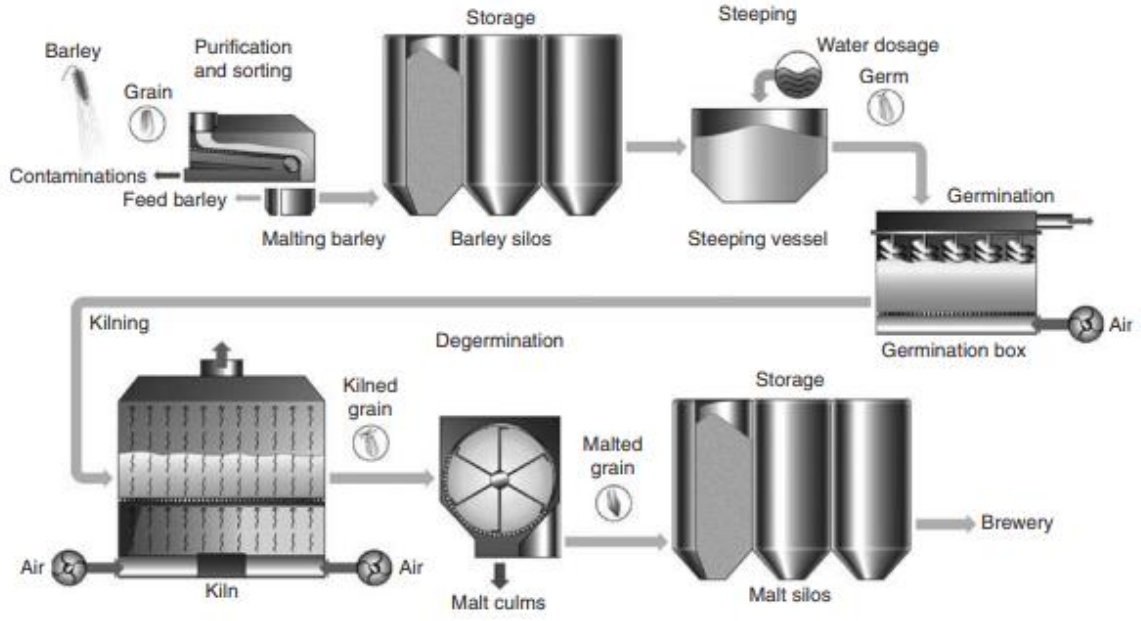


Figure 1.1 Diagram depicting the beer making process adapted from Wunderlich & Back (2009).

### 1.2.1 The selection and grading of grains

Prior to malting, high quality grains such as barley, rice, maize, oats and wheat must be selected (Wunderlich and Back, 2009). This is achieved by assessing properties of the grain such as husk condition, odour, and uniformity (Burger and LaBerge, 1985). Additionally, researchers conduct germination energy (GE) and germinative capacity (GC) tests to ensure the grains are suitable for use (Bera *et al.*, 2018). The GE test quantifies the percentage of seeds that begin to germinate within the first three days, specifically looking for the emergence of the radicle or the initial root. This test serves as a precursor to the GC test, which further evaluates the viability of the seeds (Bera *et al.*, 2018). In the GC test, a hydrogen peroxide solution is used to determine whether non-germinating seeds are dormant or have sustained damage, by assessing their response to oxidative stress (Bera *et al.*, 2018; Rani and Bhardwaj, 2021). For optimal results, GC should reach 95% in preliminary assessments and 96% following cleaning and storage (Bera *et al.*, 2018; Wunderlich and Back, 2009). Suitable grains are then preserved in refrigerated silos for four to eight weeks. During this dormancy period, the grains are protected from premature stem germination (Shu *et al.*, 2015).

### 1.2.2 The malting processes

Malting involves converting barley or other cereal grains into an important ingredient for brewing (Burger and LaBerge, 1985). This process consists of three essential stages: steeping, germination, and kilning, with each contributing uniquely to the formation of the final malt product (Rani and Bhardwaj, 2021).

#### 1.2.2.1 Steeping and Germination

The steeping process involves a series of water soakings spanning 48 to 96 hours, followed by an air rest period (Brookes *et al.*, 1976). During this rest period, water is drained and $CO_2$ is released through aeration (Rani and Bhardwaj, 2021). The goal of steeping is to increase the moisture content of the grain to a range between 42 and 48%, while retaining seed viability (Brookes *et al.*, 1976). The development of the coleorhiza, a protective sheath for the emerging root, after successful steeping indicates that the grain is ready for germination (Schwarz and Li, 2010). The germination process begins at approximately 35% moisture content, but greater steep moisture levels are often required to achieve consistent enzyme diffusion throughout the endosperm, resulting in uniform alteration of the endosperm components (Schwarz and Li, 2010).

Germination is the process by which a seed sprouts and grows into a seedling (Domin *et al.*, 2020). It begins with the seed absorbing water, which causes it to swell and become soft (Shu *et al.*, 2015). This activates the metabolic activity of the seed, which had previously been dormant or reduced (Shu *et al.*, 2015). Germination changes the endosperm, promoting enzyme activity that converts nutrients for the embryo, while helping to limit water loss through respiration (Ahmed *et al.*, 2018). This stage is characterised by embryo development, usually observed through rootlet growth and an increase in shoot length (Shu *et al.*, 2015). During the germination phase, soaked grains are spread over the floor and must be flipped over on a frequent basis to remain loose (Ahmed *et al.*, 2018; Burger and LaBerge, 1985). This is a critical step, as it allows for adequate airflow and prevents matting and excessive heat production by the grains (Ahmed *et al.*, 2018; Shu *et al.*, 2015). Germination takes three to six days in aerobic and humid conditions at 16 - 20°C, depending on the technique and raw material used (Poutanen, 2020).

During germination, gibberellic acid (GA) penetrates the aleurone layer, triggering the production of hydrolytic enzymes that break down polysaccharides and proteins in the

endosperm (Kumar *et al.*, 2013; Armitage *et al.*, 2004; Schwarz, 2020). This breakdown allows starch-converting enzymes, primarily α-amylase and β-amylase, to access the endosperm (Kumar *et al.*, 2013). β-amylase converts starch into fermentable sugars like maltose, which later form alcohol and $CO_2$, while α-amylase produces non-fermentable sugars such as maltodextrins, contributing to the texture and malty flavour of beer (Hudson, 1986; Bera *et al.*, 2018; Rani and Bhardwaj, 2021).

Concurrently, proteases produced break down 40 to 50% of the proteins into peptides and free amino acids, producing free amino nitrogen (FAN), important for yeast metabolism during fermentation (Fox, 2010; Kumar *et al.*, 2013). These nutrients are transferred to the embryo via the scutellum, providing energy and building blocks for growth (Fox, 2010). Grain modification is considered optimal when the shoot reaches three-quarters of its length, marking readiness for the kilning phase (Schwarz and Li, 2010).

#### 1.2.2.2 Kilning and Storage of malt

Kilning consists of three stages: free drying (50-60°C), falling rate (65-75°C), and curing (above 80°C for up to two hours), reducing grain moisture from 44 to 4-5% while preserving enzymatic activity and developing the malt's aroma and colour (Schwarz and Li, 2010; Rani and Bhardwaj, 2021).This process also removes unwanted flavours and prevents enzyme denaturation (Hudson, 1986). After kilning, rootlets (culms) are removed, and the malt is matured for at least three weeks to ensure even moisture distribution (Briggs *et al.*, 1981). Proper control of moisture and temperature during storage prevents microbial growth and enzyme degradation (Rani and Bhardwaj, 2021; Guine and Correia, 2013).

### 1.2.3 The brewing processes

The brewing process of beer comprises three main phases: milling and mashing the malted grains to convert starches into fermentable sugars, followed by lautering, boiling the wort with hops, and fermenting with yeast to produce alcohol and $CO_2$ (Briggs *et al.*, 1981). The final steps involve maturation to enhance flavour, filtration to remove solids, and carbonation before packaging in bottles, cans, or kegs. Each phase contributes to the distinct flavour of beer (Stewart *et al.*, 2017).

#### 1.2.3.1 Milling and mashing

Milling is the process of crushing dried malted grains into a coarse powder called grist, which is crucial for the subsequent stages of brewing, such as mashing and lautering, as it affects the quality of the final beer (Briggs *et al.*, 1981; Stewart *et al.*, 2017). A finer milling increases the surface area, facilitating enzymatic reactions and the dissolution of ingredients (Stewart *et al.*, 2017). During mashing, the grist is mixed with warm water, initiating the conversion of starch into simple sugars and proteins into peptides, a process influenced by pH and temperature control (Mallett, 2014). Starch breakdown involves gelatinisation, liquefaction, and saccharification, converting starch into fermentable sugars, primarily maltose, while proteolysis breaks down proteins into amino acids, which are vital for flavour and foam stability (Schwarz and Li, 2010; Briggs, 1998). The mashing process culminates in the formation of wort, a liquid containing sugars, proteins, and minerals essential for fermentation, which is prepared in a mash tun, where water aids in enzymatic reactions and the extraction of key organic molecules (Schwarz and Li, 2010).

#### 1.2.3.2 Lautering, wort boiling and fermentation

The lautering process separates the liquid wort from the solid grist using a lauter tun, where the grain husk acts as a natural filter (Briggs *et al.*, 1981). After draining the wort, the grist is rinsed with hot water to extract as much sugar as possible (Schwarz and Li, 2010; Stewart *et al.*, 2017). The wort is then boiled to halt enzymatic activity and extract bitterness from hops, which add flavour, aroma, and preservative qualities (Briggs *et al.*, 1981; Stewart *et al.*, 2017). After boiling, the wort is cooled and prepared for fermentation, with yeast converting the sugars into alcohol and $CO_2$ over 14 days (Briggs *et al.*, 2004; Stewart *et al.*, 2017). Top fermentation, using *Saccharomyces cerevisiae*, produces ales with a rich taste, while bottom fermentation with *Saccharomyces pastorianus* results in lagers with a cleaner, crisper flavour (Bamforth, 2017). Throughout the process, acids are added to lower the pH and prevent bacterial growth, ensuring a stable final product (Bamforth, 2017).

#### 1.2.3.3 Maturation, filtration and packaging

Following fermentation, the freshly generated “green” beer goes through a maturation and carbonation phase (Briggs *et al.*, 1981). The beer is stored at 0°C for several weeks or months, dependent on the beer type. It is during this period that yeast, resin, proteins and other unwanted components are filtered out, resulting in a cleaner beer (Stewart *et al.*, 2017). This maturation

further improves flavour and aroma through the formation of esters (Wunderlich and Back, 2009). The desired level of carbonation is then introduced by infusing $CO_2$ into the beverage (Schwarz and Li, 2010). The beer is pasteurised at 60°C to stop yeast activity and eliminate any other microorganisms in the beer, consequently increasing shelf life (Stewart *et al.*, 2017). Finally, the beer is packaged in clean and pasteurised containers (Schwarz and Li, 2010).

## 1.3 The microbiology of beer making

Most of the steps involved in making beer are designed to stop the growth of undesirable microorganisms (Bamforth, 2017). Inevitably, microbial contamination in beer production can occur at any stage of the supply chain, from the field to the manufacturing plant to the point of consumption (Van Nierop *et al.*, 2006). Generally, microbial activity is known to enhance many sensory characteristics that define high-quality beer (Bamforth, 2017). However, at certain stages, the presence of specific microorganisms may be undesirable and can adversely affect the quality of the beer (Bokulich and Bamforth, 2013). This is alarming for the brewing industry since microbial spoilage causes an estimated economic loss of millions of US dollars each year (Bokulich and Bamforth, 2013).

### 1.3.1 Microbial impact on malt and its quality

The malting phase creates an environment conducive to microbial growth due to several factors (Briggs, 1998; Doran and Briggs, 1993; Kelly and Briggs, 1992). For instance, malting involves soaking barley, which substantially increases its moisture content, creating ideal conditions for microbial proliferation (Armitage *et al.*, 2004). Additionally, the availability of nutrients in barley, or any other grains used for brewing, supports this growth (Bamforth, 2017). The sugars, proteins, and minerals released during the malting process provide a rich food source for microorganisms (Agribook, 2021). This stage is also characterised by several temperature-controlled steps, which may promote microbial growth (Bamforth, 2017). The pH levels during malting are typically neutral to slightly alkaline, optimal for many types of microorganisms (Bamforth, 2017). Furthermore, the malting process includes aerating the grains, supplying the oxygen necessary for aerobic microorganisms (Bokulich and Bamforth, 2013).

In general, microbial growth during germination negatively affects malt quality (Bokulich and Bamforth, 2013). For instance, the microorganism on barley grain surfaces can compete for oxygen with the embryo, slowing germination, and lowering rootlet growth and α-amylase activity (Briggs, 1998; Doran and Briggs, 1993; Kelly and Briggs, 1992). Furthermore, certain bacteria and fungi present in barley are capable of producing the plant hormone indole-3-acetate, GA and abscisic acid, which may alter germination and enzyme synthesis (Tuomi *et al.*, 1995).

The proliferation of the fungal population, particularly *Fusarium* species, is promoted by moisture conditions during the steeping and germination stages (Boeira *et al.*, 1999). However, this population declines as the malt undergoes drying during the kilning phase (Van Nierop *et al.*, 2006). Specific fungal genera, including *Aspergillus*, *Cladosporium*, and *Penicillium*, exhibit resistance to elevated temperatures however, and may persist in their growth processes on malt (Gyllang *et al.*, 1977; Van Nierop *et al.*, 2006).

Similarly, the bacterial population grows considerably during the malting process when compared to the community found in the dried seed (pre-malting seed) (Flannigan, 2003). This increase is dominated by the genus *Pseudomonas* and lactic acid bacteria (LAB), specifically those of the genus *Lactobacilli* (Van Nierop *et al.*, 2006). The growth is predominantly observed from the steeping phase to the completion of the germination process, and then it subsequently drops dramatically during the kilning stage (Bokulich and Bamforth, 2013; Van Nierop *et al.*, 2006). However, certain bacteria may persist on the final malt. This includes members of the genera *Lactobacilli* and *Pediococci*, which are recognised to play are role in beer spoilage (Booysen *et al.*, 2002; O'Sullivan *et al.*, 1999).

### 1.3.2 Microbial influence on wort quality

Microbial populations decline during the mashing stage, although thermotolerant microorganisms, especially LAB can survive (O'Sullivan *et al.*, 1999). The presence of bacteria during the mashing phase can provide benefits. For example, the use of LAB to acidify the mash improves nitrogen content, and stability of the foam, in addition to enhancing the colour and character of the beer (Lowe *et al.*, 2005). However, bacterial proliferation can also present major obstacles during prolonged mashing periods (Lowe *et al.*, 2005). For instance, *Bacillus* species can cause over-acidification and the production of nitrosamines by reducing nitrate to nitrite (Calderbank and Hammond, 1989; Smith *et al.*, 1992). Additionally, the

growth of *Clostridium* in the mash or wort can result in elevated levels of butyric acid, imparting a cheese-like aroma to the beer (Hawthorne *et al.*, 1991). Bacterial overgrowth on malt can also impede mash filtering through the production of dextrans (Haikara and Home, 1991). Dextrans are long branching polysaccharides that increase the viscosity of the mash, making it difficult to filter (Laus *et al.*, 2022). Furthermore, fungi on malt may produce β-glucanases and xylanases, which reduce wort viscosity and increase mash filtration (Laitila *et al.*, 2007). This decrease in viscosity negatively affects beer foam quality (Evans *et al.*, 1999).

Following the mashing process, the wort is boiled for an extended period, essentially sterilising it (O'Sullivan *et al.*, 1999). However, its high nutritional content makes it susceptible to opportunistic spoiling microorganisms unless fermentation is started immediately. Gram-negative enterobacteria, such as members of the genus *Citrobacter*, *Escherichia*, *Klebsiella*, *Obesumbacterium* and *Pantoea* are responsible in wort spoiling (Laitila *et al.*, 2018). These bacteria are known to produce large amounts of 2,3-butanediol, organic acids, and dimethyl sulphide in the wort, which can give the beer an unpleasant fruity odour (Laitila *et al.*, 2018; Priest *et al.*, 1974). *Enterobacteria* also inhibit the growth of *Saccharomyces* (Bokulich and Bamforth, 2013; Laitila *et al.*, 2018).

### 1.3.3 Impact of microbial influences on fermentation

Microorganisms on the barley husk produce enzymes that break down the cell walls into various components, some of which are then used in the brewing process (van Nierop *et al.*, 2004). This enzymatic activity results in the generation of various breakdown products, including acidic polysaccharides of substantial molecular weight, which are hypothesised to enhance the cross-linking of lectin-like proteins on yeast cell surfaces, thereby forming significant clusters or flocs of yeast cells (Gorjanović *et al.*, 2004). Moreover, during the maturation of grains and the malting process, microbial stress induces a defensive immunological response in barley, leading to the accumulation of plant defensins, including antimicrobial peptides (Kaur *et al.*, 2012). These peptides have been implicated in premature yeast flocculation (PYF) by disrupting yeast cell metabolism, respiration, and cell membrane integrity, causing irreversible cellular damage (van Nierop *et al.*, 2004; Panteloglou *et al.*, 2012; Qin *et al.*, 2023). Studies have also shown that fungi, rather than bacteria, are more closely associated with PYF throughout the brewing process (Kaur *et al.*, 2012; MacIntosh *et al.*, 2014). Fungal contamination in the field barley, specifically by *Fusarium graminearum* and *Cochliobolus sativus*, are known to cause PYF behaviour (MacIntosh *et al.*, 2014).

Furthermore, it has been reported that treating malt husk with extracellular fungal enzymes before mashing can cause PYF (Kaur *et al.*, 2012).

## 1.4 Barley as a key ingredient in beer production

### 1.4.1 Introduction and adaptability

Barley (*Hordeum vulgare* L.) is an important cereal crop globally and is used for both human and animal consumption (Newman, 2006). Globally, it occupies the fourth position in terms of both quantity produced and area of cultivation, following maize, rice and wheat (Bera *et al.*, 2018). Its importance in early agricultural cultures is demonstrated by the fact that it was one of the earliest plants domesticated and played a key part in the transition from nomadic to sedentary lifestyles (Hornsey, 2003).

Barley is regarded as the most adaptable cereal because of its ability to grow in a wide range of climates and geographical regions (Quan *et al.*, 2016; Kumar *et al.*, 2013). It thrives in both temperate and tropical climates and can withstand harsher circumstances than many other cereals (Arendt and Zannini, 2013; Kumar *et al.*, 2013). Barley can grow well on poor soils and tolerate droughts, making it a dependable crop in areas where other cereals may fail (Hao *et al.*, 2022). It is widely grown in countries such as Canada, Germany and Russia, but also thrives in high-altitude countries such as Nepal and Tibet, where it is grown on mountain slopes at higher elevations than any other cereal crops (Kumar *et al.*, 2013; Bianco *et al.*, 2018; Hao *et al.*, 2022). Barley also has a short growing season which allows it to evade seasonal challenges such as late spring frosts or summer droughts (Hao *et al.*, 2022). This adaptability makes barley an essential crop in many parts of the world, contributing to food security and agricultural sustainability (Kumar *et al.*, 2013).

### 1.4.2 The benefits of using barley in beer production

As mentioned previously, beer can be made from a number of grains, such as barley, rice, millet, corn, and sorghum (Hornsey, 2003). However, the barley seed has unique enzymatic and physical characteristics, making it the preferred grain for the brewing industry (Newman, 2006; Schwarz and Li, 2010). The multilayer seed structure of barley is important, notably the aleurone layer, which is rich in malting enzymes (Rani and Bhardwaj, 2021). Furthermore, the

endosperm of barley seeds contains a substantial starch reserve that is required after germination (Rani and Bhardwaj, 2021).

Although other cereals have an aleurone layer, barley is the most effective at breaking down the endosperm and converting starch to simpler sugars (Arendt and Zannini, 2013). Barley enzymes, particularly α-amylase, are extremely thermally stable, which is important for mashing as temperature changes can impede enzyme activity (Laus *et al.*, 2022). This stability ensures consistent performance across brewing temperatures, allowing for a uniform conversion of starch to sugar (Parés Viader *et al.*, 2021; Laus *et al.*, 2022). Barley enzyme effectiveness is also influenced by temperature and pH; α-amylase and β-amylase perform best at specific temperatures and pH ranges, allowing for a more successful and regulated conversion process (Parés Viader *et al.*, 2021).

The composition and structure of barley provides further evidence of its suitability for beer production. With starch accounting for more than 70% of its dry mass, barley has a high starch-to-husk ratio which is a benefit in the mashing process (Yu *et al.*, 2017). The barley husk not only protects the kernel during malting, but it is also detachable, producing a natural filter bed during the lautering stage of brewing, which aids in the filtration process (Yangcheng *et al.*, 2016). This attribute is not found in other grain used for brewing. Furthermore, barley has a higher diastatic power (DP) than other grains (Rani *et al.*, 2022). This suggests that it has high starch-converting efficiency, which adds to its suitability as the grain of choice for beer production.

## 1.5 Protein and nitrogen content of barley and their impact on beer production

### 1.5.1 The impact of grain protein content on beer quality

Barley has a protein content ranging from 9 to 12%; this provides an ideal balance for yeast nutrition, effective fermentation, and the development of beer body and foam stability (Zhu, 2017). This balance is essential since too much protein will generate haze and undesirable flavours, whilst too little can limit mouthfeel and foam retention (Xie *et al.*, 2023). In contrast, grains such as wheat, maize, and rice have a wide range of protein levels (Arendt and Zannini, 2013). This can impact the brewing process and beer quality in a variety of ways. Wheat, for example, may improve foam stability while increasing the probability of haze formation,

whereas maize and rice, with lower protein content, may be preferred for brewing lighter beer styles due to their impact on body and foam stability (Briggs *et al.*, 2004). Considering the importance of grain protein content in defining the taste and quality of the finished beer, it is essential to select barley grains with adequate protein content (Briggs *et al.*, 2004). An effective approach to assess this is through a measurement of the nitrogen content of the grain.

### 1.5.2 Nitrogen as an indicator of protein levels

#### 1.5.2.1 Introduction and sources of nitrogen in barley

Nitrogen in barley grains originates from two primary sources, the soil and the atmosphere (Pohlman and McColl, 1982). The initial source is soil nitrogen (Toews and Soper, 1978). Through their roots, barley plants assimilate nitrogen available in the soil, which exists in several forms, such as nitrate ($NO_3^-$) and ammonium ($NH_4^+$) (Söderberg and Bååth, 2004). Alternatively, barley benefits from the conversion of atmospheric nitrogen ($N_2$) into a usable form for plants (either ammonium or nitrate) by certain bacteria (Pohlman and McColl, 1982).

Nitrogen is a key component of plant growth, and has a considerable impact on germination rate (Finkel, 2011). The correlation between the nitrogen content of seeds and their protein levels is based on the premise that proteins are composed of amino acids (Kibite and Evans, 1984; Beauregard and Hefford, 2006). All amino acids incorporate nitrogen, which is an essential component of their structure and accounts for approximately 16% of their weight (Yu and Fukagawa, 2020). Given this composition, nitrogen in seeds can serve as a reliable indicator for calculating protein content.

#### 1.5.2.2 Techniques for nitrogen content determination in barley grains

Assessing the nitrogen content in barley grains can be achieved using several methods such near-infrared spectroscopy (NIRS) (Angelino and Convention, 1996), the Kjeldahl method (Buckee, 1994) and the Dumans method (Ebeling, 1968). NIRS is a non-invasive technique that illuminates a sample of barley with near-infrared light and measures the amount of light absorbed by the sample. Different compounds in the seed, including protein, absorb different amounts of light, making it easier to determine protein and, by extension, nitrogen content (Sileoni *et al.*, 2015). The NIRS process is swift, bypasses the need for sample preparation and allows for the evaluation of numerous barley quality aspects beyond protein content, such as moisture, starch, and oil content (Sileoni *et al.*, 2015). However, it requires calibration using

known samples, and its accuracy can be impacted by factors such as the temperature of the grain and the equipment employed (Sileoni *et al.*, 2015).

In parallel, the Kjeldahl method is a standard technique for estimating the total nitrogen (and by proxy the protein content) of a sample (Buckee, 1994). The sample is digested in sulfuric acid to convert organic nitrogen to ammonium, after which it is converted to ammonia and distilled. The resultant ammonia is then titrated to determine its nitrogen content (Marcó *et al.*, 2002). Another laboratory analysis to measure nitrogen content encompasses the Dumas method, which involves the combustion of a barley grain sample to release nitrogen gas (Ebeling, 1968). The nitrogen content is then quantified based on the volume of nitrogen gas produced (Ebeling, 1968).

It is important to note that while these methodologies can provide an approximation of protein content, they do not reveal information about the specific proteins present or their function, which impacts malting quality. Thus, additional tests, such as those for enzymatic activity, GC and grain rigidity, must be conducted to determine the suitability of barley grains for malting (Kumar *et al.*, 2013).

#### 1.5.2.3 Nitrogen grain content impact on brewing.

In brewing, the preferred nitrogen content in barley grains is typically between 1.4 and 1.8% on a dry weight basis (Kassie and Tesfaye, 2019). Barley with a nitrogen content below 1.4% might lead to poor yeast nutrition and inadequate beer head formation, whereas nitrogen levels above 1.8% can jeopardise beer clarity and stability (Kassie and Tesfaye, 2019). Low nitrogen levels can stress yeast cells, resulting in off-flavours, whereas high nitrogen levels might cause excessive fermentation and undesirable byproducts (Kassie and Tesfaye, 2019). Brewers often tailor nitrogen levels to suit the specific beer style, striving for the desired balance of flavour, aroma and physical characteristics (Stewart *et al.*, 2017). For example, ales often require a higher nitrogen content to support their robust flavour and fuller body (Stewart *et al.*, 2017). Accurate measurement and management of nitrogen levels is therefore essential for ensuring the suitability of barley grains for high-quality malt and beer production, with the required nitrogen content varying according to brewing specifications (Agu and Palmer, 2001). Maltsters and brewers meticulously manage nitrogen levels by carefully selecting barley varieties and modifying the conditions of malting and brewing. This precise control maximises protein utilisation, thereby minimising its impact on the quality of beer (Schwarz and Li, 2010).

### 1.6 Processing barley for brewing from harvest to storage

Barley is generally ready for harvest approximately four months after sowing, though some varieties may mature in about two months (Henry, 2016). It is advisable to harvest barley when the grain moisture level is between 16-18% (w/w) (Henry, 2016). This moisture range is optimal for direct combining, a process that integrates harvesting and threshing into a single step (Khokonova, 2011). Following harvest, it is important to clean the grain to eliminate foreign materials and then dry it to a desired moisture content of 12% (Henry, 2016). Subsequently, the barley is stored in silos, which offer effective protection against external air and pest contamination (Henry, 2016; Khokonova, 2011). Silos are also designed to avert moisture and mould growth through their high-quality sealing, yet they provide the option to ventilate to maintain air quality (Ghali, 2014). With proper temperature and moisture management, barley can be stored for extended periods (Mallett, 2014). The duration of storage can, however, vary based on the moisture content of the grain and its intended use (Khokonova, 2011). For instance, barley intended for brewing, with a moisture content of 10.5% or lower, can be stored for up to 18 months at temperatures ranging from 10 to 20°C (Khokonova *et al.*, 2015; Henry, 2016).

Monitoring the internal temperature of the silos is important; the optimal temperature ranges are recommended to be 20–23°C during summer and below 15°C in winter (Ghali, 2014). If these temperature guidelines are not adhered to, barley may experience accelerated ageing or a decline in viability and GC (Angelino and Convention, 1996). Managing humidity levels is equally important; changes in humidity can affect the moisture content of barley, leading to microbial growth or spoilage (Henry, 2016; Khokonova *et al.*, 2015). The duration of storage distinctly impacts the quality of barley (Flannigan, 1978; Henry, 2016). As storage time increases, barley grains may gradually lose their vitality, leading to slower germination rates and, in some cases, seed decay (Mwazha *et al.*, 2022). Prolonged storage can also negatively affect enzyme activity within the barley (Felšöciová *et al.*, 2021; Mwazha *et al.*, 2022).

Several microorganisms are found in different parts of the seed (Bokulich and Bamforth, 2013). Changes in storage conditions, such as moisture, temperature, oxygen levels and storage duration, can alter the microbial dynamics of the grain or promote the proliferation of microorganisms that may have detrimental effects, such as the growth of pathogenic microorganisms (Bokulich and Bamforth, 2013). (Khokonova *et al.*, 2015).

## 1.7 Introduction to the seed microbiome

The term microbiome refers to a community of microorganisms, including archaea, bacteria, fungi and viruses, that reside within a specific environment (Marchesi and Ravel, 2015). This environment may vary, extending from the human body to seed, soil, or a body of water (Berg *et al.*, 2020). Each microbiome is unique, with its composition influenced by the environment it occupies and other external factors (Berg *et al.*, 2020). The seed microbiome refers to the community of microorganisms that live in, on and around seeds (Wassermann *et al.*, 2021). These microorganisms can transfer to subsequent plant generations, thereby playing a critical role in plant health (Abdelfattah *et al.*, 2023). The seed microbiome comprises both pathogenic and beneficial microorganisms and includes microorganisms that colonise seeds through two primary modes: endophytic and epiphytic colonisation (Abdelfattah *et al.*, 2023). Endophytes are microorganisms, typically bacteria or fungi, that reside in the internal tissues of seeds, where they vertically transmit to their progeny seedlings (Díaz Herrera *et al.*, 2016; Hodgson *et al.*, 2014). In contrast, epiphytes are microorganisms that live on the surface of seeds (Hodgson *et al.*, 2014). These epiphytes may penetrate the seed tissues and be passed from parents to offspring through vertical transmission, or they may spread between distinct seeds or plants via horizontal transmission (Hodgson *et al.*, 2014; War *et al.*, 2023).

#### 1.7.1.1 Beneficial role of the seed microbiome

Seed-associated microorganisms contribute to the improvement of seedling establishment as well as growth and development, especially in soils lacking nutrients and under harsh conditions (Compant *et al.*, 2019; War *et al.*, 2023). Firstly, they improve nutrient uptake and assimilation through various processes such as phosphate solubilisation, siderophore production, 1-aminocyclopropane-1-carboxylate (ACC) deaminase activity and nitrogen fixation (Bashir *et al.*, 2022, Pohlman and McColl, 1982; Chimwamurombe *et al.*, 2016; Truyens *et al.*, 2015). Secondly, they play a role in biological control, offering protection against pathogens and competing against rival plant species (Díaz Herrera *et al.*, 2016; Saeed *et al.*, 2021). Thirdly, they boost plant metabolic processes such as modulating reactive oxygen species (ROS) levels and regulating phytohormones such as auxin, cytokinin and gibberellin (Pitzschke, 2016; Shahzad *et al.*, 2016; Scott *et al.*, 2007). Furthermore, they confer resistance to abiotic stresses including drought, salinity and heavy metal contamination (Oukala *et al.*, 2021; Trivedi *et al.*, 2020).

#### 1.7.1.2 The detrimental effects of the seed microbiome

Not all microorganisms associated with the seed microbiome are beneficial. The susceptibility of seeds to microbial pathogens can be attributed to their high nutrient content and protective nature, which provides pathogens an attractive habitat and a food source (Turner *et al.*, 2020). Microbial pathogens may produce spores and toxins known to impair seed quality and GC (Noots *et al.*, 1999). Seed germination and growth can be further hindered by the depletion of nutrients, which can occur because of the enzymatic breakdown of complex compounds by the microbial pathogens. Microbial enzymes work together to degrade plant defences, making plant tissues more susceptible to pathogen invasion (War *et al.*, 2023). Endoglucanases and exoglucanases, for example, are enzymes that target cellulose, a key component of seed cell walls (Kubicek *et al.*, 2014; de Vries and Visser, 2001). Pectinases are enzymes that break down pectin, which is required for the plant cell wall to remain intact (Walton, 1994; Bonnin *et al.*, 2014). Lipases damage the lipid components of the plant cell membrane, compromising the structural integrity of the seed (Walton, 1994).

### 1.7.2 The barley seed microbiome

The barley microbiome refers to the community of microorganisms that live in association with roots, stems, leaves and seeds of barley plants (Noots *et al.*, 1999). Recent studies employing advanced sequencing technologies, such as metagenomics and amplicon sequencing, have revealed the vast microbial diversity associated with barley seeds (Bziuk *et al.*, 2021; Saeed *et al.*, 2021). These studies have identified a wide range of bacterial taxa, including Proteobacteria, Actinobacteria, Firmicutes, and Bacteroidetes, as well as diverse fungal communities dominated by Ascomycota and Basidiomycota (Bziuk *et al.*, 2021; Felšöciová *et al.*, 2020; Felšöciová *et al.*, 2021).

Among the bacteria, *Pseudomonas* is a predominant genus within the barley seed microbiome (Bziuk *et al.*, 2021; Tshisekedi *et al.*, 2024). Members of this genus are known for their roles in suppressing soil-borne diseases, enhancing stress tolerance, and promoting plant growth by producing phytohormones (Pham *et al.*, 2017; Fröhlich *et al.*, 2012). *Bacillus* species, also found abundantly, offer protection against various pathogens through the production of antibiotics and antifungal compounds and are involved in nutrient mobilisation, which is crucial for seedling development (Smith *et al.*, 1992; Lowe *et al.*, 2005). Additionally, genera such as *Enterobacter* and *Pantoea* are often found in barley seeds, where they promote plant

growth by fixing nitrogen, solubilizing phosphate, and producing growth-promoting substances (Lv *et al.*, 2022; Singh *et al.*, 2021).

The fungal taxa associated with barley seeds include species such as *Fusarium, Alternaria,* and *Cladosporium* (Felšöciová *et al.*, 2021). While some *Fusarium* species can cause diseases in barley, others may be neutral or beneficial, competing with pathogenic microbes and contributing to the microbial balance (Díaz Herrera *et al.*, 2016). *Alternaria,* commonly associated with the phyllosphere, also inhabits barley seeds, with roles varying from pathogenic to symbiotic depending on environmental conditions and the presence of other microorganisms (Felšöciová *et al.*, 2021). *Cladosporium,* known for its ubiquitous presence, can be part of the normal seed microbiota, aiding in biotic and abiotic stress tolerance (Felšöciová *et al.*, 2021).

### 1.7.3 Bacterial communities in barley

Bacteria are the most abundant microorganisms in soil, playing pivotal roles in nutrient cycling and disease suppression (Felšöciová *et al.*, 2020; Noots *et al.*, 1999; Shala *et al.*, 2023). Commonly found genera include *Actinobacteria*, *Bacillus*, *Paenibacillus, Pantoea*, and *Pseudomonas* (Felšöciová *et al.*, 2020; Noots *et al.*, 1999; Shala *et al.*, 2023). These bacteria are fundamental in promoting seed growth by facilitating processes improving mineral nutrition, nitrogen fixation and phosphate solubilization (Yulan Chen *et al.*, 2022; Wassermann *et al.*, 2021; Bziuk *et al.*, 2021).

Several strains within the barley seed microbiome are particularly beneficial, enhancing plant growth and providing resistance against pathogens (Felšöciová *et al.*, 2020; Noots *et al.*, 1999; Shala *et al.*, 2023). Notably, certain strains of *Bacillus* and *Pseudomonas* produce plant growth-promoting substances such as indole acetic acid (IAA) and siderophores, which not only enhance plant growth but also suppress disease-causing pathogens (Bziuk *et al.*, 2021). For example, strains of *Paenibacillus* produce biofilms that effectively control *Fusarium graminearum* (Taheri *et al.*, 2022), while strains of *Pseudomonas* offer resistance against barley leaf disease caused by *Rhynchosporium secalis* and suppress the growth of the root pathogen *Gaeumannomyces graminis* (Dutilloy *et al.*, 2022; Fröhlich† *et al.*, 2012). *Pantoea agglomerans* inhibits the proliferation of *Pseudomonas syringae pv. syringae*, which causes basal kernel blight (Braun-Kiewnick *et al.*, 2000). Additionally, *Duffyella gerundensis* has been shown to inhibit the growth of *Fusarium tricinctum* and reduce toxin levels associated with this fungus (Gnonlonfoun *et al.*, 2023). However, not all microorganisms in barley seeds

are beneficial; *Erwinia persicina*, for instance, causes a decline in seed quality due to the pink staining it produces (Kawaguchi *et al.*, 2021).

The bacterial communities in stored barley play a role in determining the quality and safety of the grains (Felšöciová *et al.*, 2020). At the onset of storage, *Pantoea agglomerans* is typically predominant, but over time, a diverse array of bacteria including *Alcaligenes spp.*, *Arthrobacter globiformis*, *Clavibacter iranicum*, *Kosakonia cowanii*, *Lactobacillus*, *Pseudomonas fluorescens*, *Pseudomonas libanensis*, *Pseudomonas poae*, *Pseudomonas synxantha*, and *Staphylococcus haemolyticus* can also be found in barley seeds, though the functional roles of these bacteria during storage remains largely unexplored (Felšöciová *et al.*, 2020; Noots *et al.*, 1999). Research indicates that while the abundance of coliform bacteria tends to increase, LAB, which are vital for the malting process, tend to decrease as storage progresses (Felšöciová *et al.*, 2020). Additionally, spore-forming bacteria such as *Bacillus altitudinis* may degrade seed quality and complicate the management of microbial populations in stored grains (Dar *et al.*, 2021).

### 1.7.4 Fungal communities in barley

Fungi are crucial for organic matter decomposition and nutrient cycling, with common genera including *Alternaria*, *Cladosporium*, *Fusarium*, *Penicillium*, and *Rhizoctonia*, involved in interactions affecting seed health and storage quality (Wen Chen *et al.*, 2022; Felšöciová *et al.*, 2020; Flannigan, 1978; Flannigan, 1969; Laitila *et al.*, 2007). For instance, *Fusarium* species are known for producing mycotoxins that can lead to diseases such as *Fusarium* head blight (Alisaac and Mahlein, 2023). *Alternaria* spp. are responsible for leaf spot diseases that diminish crop yields and degrade barley seed quality (Agostinetto *et al.*, 2020). Additionally, the conditions required for the proliferation of fungi such as *Penicillium* and *Cladosporium*, which thrive with a water activity (*aw*) above 0.90 corresponding to moisture levels of 20 to 25%, are common in fields and storage areas, highlighting the need for effective environmental management to maintain barley quality (Felšöciová *et al.*, 2021).

Fungi proliferate post-harvest and throughout the storage of barley, often associated with the occurrence of pathogens that can deteriorate grain quality (Wen Chen *et al.*, 2022). The mycobiome of stored grains predominantly includes mould genera such as *Aspergillus*, *Alternaria*, *Cladosporium*, *Fusarium*, *Mucor*, *Rhizopus*, and *Penicillium*, which are capable of surviving long periods under varying storage conditions (Noots *et al.*, 1999; Flannigan, 2003;

Blackburn, 2006; Felšöciová *et al.*, 2021). These fungi begin to germinate and expand once the moisture content in barley surpasses 15%, with certain species such as *Aspergillus* and *Fusarium* being particularly detrimental as they invade seeds and degrade essential nutrients, thereby depleting the energy reserves necessary for seed germination and early plant development (Blackburn, 2006; Boeira *et al.*, 1999; Birr *et al.*, 2020).

Moreover, the diversity of fungal species detected in barley changes throughout the storage period, adapting to lower water activity levels, thus showing resilience even in drier conditions (Felšöciová *et al.*, 2021; Laitila *et al.*, 2007; Noots *et al.*, 1999). Yeasts such as *Candida*, *Pichia*, *Rhodotorula*, *Saccharomyces*, *Sporobolomyces*, and *Torulopsis* also colonise stored barley grains, adding a layer of microbial interaction (Flannigan, 2003; Noots *et al.*, 1999). While some yeasts, like *Saccharomyces*, play a protective role through biocontrol activities that suppress pathogenic filamentous fungi, their presence can also lead to spoilage by increasing moisture content and the accumulation of unwanted metabolites (Bokulich and Bamforth, 2013).

### 1.7.5 The unexplored aspect of barley microbiomes for brewing

Our current knowledge of the barley seed microbiome hints at a complex web of interactions that are crucial to the storage life and quality of barley seeds, yet the specifics of these relationships need to be elaborated (Bziuk *et al.*, 2021; Felšöciová *et al.*, 2020; Felšöciová *et al.*, 2021). Short-term nitrogen fertilisation has been proven in studies to have an impact on agricultural soil by influencing microbial community composition and nitrogen mineralization processes (Ouyang and Norton, 2020). In addition, increasing nitrogen intake has been shown to increase the complexity of microbial networks in soil (Ouyang and Norton, 2020). These findings demonstrate the impact of nitrogen on the microbial community composition in agricultural settings.

The impact of nitrogen content in barley grain on seed quality is widely understood (Agu and Palmer, 2001). However, the intricacies of how nitrogen content impacts the microbial population in barley grain, and thus seed quality, are largely unresolved. It has been reported that nitrogen levels in the barley rhizosphere influence the selective enrichment of microbial taxa, with low nitrogen levels improving the genetic control of the host over its microbiota (Alegria Terrazas *et al.*, 2022). Furthermore, taxonomic diversity plays a role in functional

specialisation, which promotes plant development in nitrogen-limited environments (Alegria Terrazas *et al.*, 2022).

One key unsolved aspect is how grain nitrogen content affects the microbiome of barley stored for beer production, compared to that of harvested seeds. Although the changes in the barley microbiome during storage are known, its functional capacities during this period remain unclear (Felšöciová *et al.*, 2020; Felšöciová *et al.*, 2021). To address this gap, a more specific strategy could include identifying microbial signatures or indicator species in the microbiome that are capable of predicting grain quality or signal potential shortcomings, allowing for more precise interventions. Furthermore, having a functional profile of what these taxa are capable of would be quite beneficial. This would be best performed through meta-omics analysis, which has potential for expanding our understanding (Fadiji and Babalola, 2020).

### 1.8 Meta-omics approaches to study temporal microbiome changes

Meta-omic techniques, particularly metagenomics and metaproteomics, have revolutionised our understanding of microbial communities by providing comprehensive insights into both the taxonomic composition and functional potential of these communities (Zhang *et al.*, 2019). In the context of barley seed storage, these techniques could offer an opportunity to delve into the dynamics of the microbiome over time, especially under varying nitrogen conditions, though the application of such approaches remains largely unexplored for barley (Bziuk *et al.*, 2021; Tshisekedi *et al.*, 2023).

Metagenomics allows researchers to capture a snapshot of the genetic potential of microbial communities by sequencing their collective DNA (Desai *et al.*, 2012). This approach provides information on the presence and relative abundance of different microbial taxa, offering insights into how microbial communities shift over time and in response to environmental changes, such as the nitrogen content of barley seeds (Fadiji and Babalola, 2020). For example, studies on other crops like wheat and maize have demonstrated how metagenomics can reveal microbial shifts during storage (Kaushal *et al.*, 2020; Akinola *et al.*, 2021; Aswini *et al.*, 2023). In these systems, it was observed that storage conditions, including temperature and moisture content, can significantly alter microbial community structure (Solanki *et al.*, 2019; Li *et al.*, 2023). Nitrogen fertilisation has similarly been shown to drive changes in microbial diversity and composition in maize rhizospheres (Miranda-Carrazco *et al.*, 2022; Liu *et al.*, 2024). These

findings suggest that metagenomics is an ideal tool for investigating microbial changes in stored barley seeds, particularly in relation to nitrogen content.

Metaproteomics, on the other hand, takes this analysis a step further by linking genetic potential to actual protein expression (Abraham *et al.*, 2014). While metagenomics provides information on what microorganisms could achieve on their genetic repertoire, metaproteomics reveals what microorganisms are doing by analysing the proteins they express (Heyer *et al.*, 2015; Abraham *et al.*, 2014). This technique is especially effective for understanding functional changes within microbial communities over time. For instance, studies on rice storage have used metaproteomics to identify microbial proteins involved in starch degradation and protein synthesis, functions crucial for the preservation of seed quality (Salvato *et al.*, 2021). In nitrogen-fertilised environments, metaproteomics has been used to track how microbial communities adjust their metabolic processes in response to nitrogen availability (Pan *et al.*, 2024). These examples highlight the potential of metaproteomics to provide functional insights into how microbial communities associated with barley seeds may respond to varying nitrogen levels over prolonged storage periods.

While these techniques have been applied to a range of crops, the use of meta-omics to explore temporal changes in microbial communities during the storage of barley seeds remains uncharted. The integration of metagenomics and metaproteomics can provide a comprehensive understanding of both the taxonomic shifts and functional adaptations of microbial communities over time, particularly in response to nitrogen modulation (Zhang *et al.*, 2019; Graw *et al.*, 2021). This study, therefore, aims to fill this gap by applying both approaches to investigate the dynamics of the barley seed microbiome under different nitrogen conditions and storage durations, providing novel insights into how these factors influence microbial composition and function.

## 1.9 Research Aims and Objectives

The role of nitrogen in shaping microbial communities, particularly in the context of seed storage, remains an underexplored area in microbial ecology. While several studies have examined the effects of nitrogen on soil microbiomes, there is a limited understanding of how nitrogen levels within seeds influence microbial taxa and function over time. This knowledge gap is particularly evident in malting barley, where the effects of varying nitrogen concentrations on the seed microbiome during long-term storage have yet to be fully characterised.

Addressing this gap is essential as nitrogen plays a key role in microbial metabolic activities, which directly affect seed health, preservation, and ultimately malting quality. This research aims to explore these dynamics by employing advanced metagenomic and metaproteomic techniques to analyse how nitrogen content and storage periods influence both the composition and functionality of microbial communities in barley seeds.

The study will achieve this aim through the following objectives:

**Objective 1: Assessment of microbial taxonomic abundance**

This objective aims to estimate microbial taxa abundance in the barley seed microbiome with varying nitrogen contents across different storage times. By doing so, this research will determine whether certain microbial taxa are favoured or suppressed under specific nitrogen conditions. This will provide insights into how nitrogen shapes the microbial ecosystem over time, allowing for the identification of taxa that may play beneficial or detrimental roles in seed health and storage.

**Objective 2: Reconstruction of metagenome-assembled genomes (MAGs)**

Through the reconstruction of MAGs, this study will explore the genomic composition of microbial communities in relation to seed nitrogen content and storage conditions. By examining the presence of distinct genomic features, the research will determine if nitrogen-modulated communities harbour specific genetic capabilities, such as nitrogen fixation or enhanced metabolic versatility. This analysis will help pinpoint the nitrogen-responsive microoganisms, potentially including nitrogen-utilising bacteria or fungi, that contribute to seed preservation or deterioration.

**Objective 3: Assembly and annotation of metagenomic reads**

The assembly and annotation of metagenomic reads will create a comprehensive protein profile of the microbiome, linking genetic potential to microbial metabolic capabilities. By doing so, this research will uncover how the microbial community's metabolic potential shifts in response to nitrogen content, providing a deeper understanding of which metabolic pathways or if functions are enriched or diminished under different nitrogen conditions. This is particularly relevant for understanding microbial contributions to seed health and nitrogen cycling during storage.

**Objective 4: Analysis of protein expression through metaproteomics**

Metaproteomic analysis will offer direct insights into the functional roles that microoganisms play under varying nitrogen levels and storage conditions by assessing actual protein expression. This objective is critical for identifying which microbial functions are activated or repressed in response to nitrogen modulation, offering a real-time view of microbial activity in the stored seeds. Such insights could reveal key proteins involved in processes like nutrient cycling, seed preservation, or even spoilage, linking microbial function to practical outcomes in barley storage and malting.

## 1.10 References


Abdelfattah, A., Tack, A.J.M., Lobato, C., Wassermann, B., Berg, G. (2023). From seed to seed: the role of microbial inheritance in the assembly of the plant microbiome. *Trends in Microbiology*, **31**: 346–355.

Abraham, P.E., Giannone, R.J., Xiong, W., Hettich, R.L. (2014). Metaproteomics: Extracting and Mining Proteome Information to Characterize Metabolic Activities in Microbial Communities. *Current Protocols in Bioinformatics*, **46**: 13.26.1-13.26.14.

Abubakar, S. (2019). Effect of Storage Conditions on wheat Quality. *International Journal of Food and Allied Sciences*, **4**: 20–23.

Agostinetto, L., Casa, R.T., Bogo, A., Alves Neto, L., Vieira Junior, J. de A.L., Fingstag, M.D., Rosa, J.M. da. (2020). Viabilidade de *Alternaria alternata*, *Bipolaris sorokiniana* and *Drechslera teres* transmitidos por sementes de cevada no sul do Brasil. *Summa Phytopathologica*, **46**: 26–30.

Agribook. (2021). Barley. *Agribook Digital* https://agribook.co.za/agronomy/barley/. Visited: 9th August 2021.

Agu, R.C., Palmer, G.H. (2001). The Effect of Nitrogen Level on the Performance of Malting Barley Varieties during Germination. *Journal of the Institute of Brewing*, **107**: 93–98.

Ahmed, A.K.A., Shi, X., Hua, L., Manzueta, L., Qing, W., Marhaba, T., Zhang, W. (2018). Influences of Air, Oxygen, Nitrogen, and Carbon Dioxide Nanobubbles on Seed Germination and Plant Growth. *Journal of Agricultural and Food Chemistry*, **66**: 5117–5124.

Akinola, S.A., Ayangbenro, A.S., Babalola, O.O. (2021). The diverse functional genes of maize rhizosphere microbiota assessed using shotgun metagenomics. *Journal of the Science of Food and Agriculture*, **101**: 3193–3201.

Alegria Terrazas, R., Robertson-Albertyn, S., Corral, A.M., Escudero-Martinez, C., Kapadia, R., Balbirnie-Cumming, K., Morris, J., Hedley, P.E., Barret, M., Torres-Cortes, G., Paterson, E., Baggs, E.M., Abbott, J., Bulgarelli, D. (2022). Defining Composition and Function of the Rhizosphere Microbiota of Barley Genotypes Exposed to Growth-Limiting Nitrogen Supplies. *mSystems*, **7**: e0093422.

Alisaac, E., Mahlein, A.-K. (2023). *Fusarium* Head Blight on Wheat: Biology, Modern Detection and Diagnosis and Integrated Disease Management. *Toxins*, **15**: 192.

Angelino, S.A.G.F., Convention, O.B.O.T.A.C.O.T.E.B. (1996). Determination of the Moisture and Nitrogen Contents of Barley and Malt by Near Infrared Spectroscopy (nirs). *Journal of the Institute of Brewing*, **102**: 73–74.

Aranguren, M., Castellón, A., Aizpurua, A. (2020). Crop Sensor Based Non-destructive Estimation of Nitrogen Nutritional Status, Yield, and Grain Protein Content in Wheat. *Agriculture*, **10**: 148.

Arendt, E.K., Zannini, E. (2013). *Cereal Grains for the Food and Beverage Industries*. Elsevier, 519 pp.

Armitage, D.M., Baxter, E.D., Knight, J., Wilkin, D.R., Woods, J.L. (2004). Malting Barley: Europe. *Crop Post-Harvest: Science and Technology*. John Wiley & Sons, Ltd, pp. 117–127.

Aswini, K., Suman, A., Sharma, P., Singh, P.K., Gond, S., Pathak, D. (2023). Seed endophytic bacterial profiling from wheat varieties of contrasting heat sensitivity. *Frontiers in Plant Science*, **14**.

Balet, S., Gous, P., Fox, G., Lloyd, J., Manley, M. (2020). Characterisation of starch quality from barley varieties grown in South Africa. *International Journal of Food Science & Technology*, **55**: 443–452.

Bamforth, C.W. (2017). Progress in Brewing Science and Beer Production. *Annual Review of Chemical and Biomolecular Engineering*, **8**: 161–176.

Beauregard, M., Hefford, M.A. (2006). Enhancement of essential amino acid contents in crops by genetic engineering and protein design. *Plant Biotechnology Journal*, **4**: 561–574.

Bera, S., Sabikhi, L., Singh, A.K. (2018). Assessment of malting characteristics of different Indian barley cultivars. *Journal of Food Science and Technology*, **55**: 704–711.

Berg, G., Rybakova, D., Fischer, D., Cernava, T., Vergès, M.-C.C., Charles, T., Chen, X., Cocolin, L., Eversole, K., Corral, G.H., Kazou, M., Kinkel, L., Lange, L., Lima, N., Loy, A., Macklin, J.A., Maguin, E., Mauchline, T., McClure, R., Mitter, B., Ryan, M., Sarand, I., Smidt, H., Schelkle, B., Roume, H., Kiran, G.S., Selvin, J., Souza, R.S.C. de, van Overbeek, L., Singh, B.K., Wagner, M., Walsh, A., Sessitsch, A., Schloter, M. (2020). Microbiome definition re-visited: old concepts and new challenges. *Microbiome*, **8**: 103.

Bianco, A., Fancello, F., Balmas, V., Zara, G., Dettori, M., Budroni, M. (2018). The microbiome of Sardinian barley and malt. *Journal of the Institute of Brewing*, **124**: 344–351.

Birr, T., Hasler, M., Verreet, J.-A., Klink, H. (2020). Composition and Predominance of *Fusarium* Species Causing *Fusarium* Head Blight in Winter Wheat Grain Depending on Cultivar Susceptibility and Meteorological Factors. *Microorganisms*, **8**: 617.

Blackburn, C. de W. (2006). *Food Spoilage Microorganisms*. Woodhead Publishing, 737 pp.

Boeira, L. s., Bryce, J. h., Stewart, G. g., Flannigan, B. (1999). Inhibitory Effect of *Fusarium* Mycotoxins on Growth of Brewing Yeasts. 1 Zearalenone and Fumonisin B1*. *Journal of the Institute of Brewing*, **105**: 366–375.

Bokulich, N.A., Bamforth, C.W. (2013). The microbiology of malting and brewing. *Microbiology and molecular biology reviews: MMBR*, **77**: 157–172.

Bonnin, E., Garnier, C., Ralet, M.-C. (2014). Pectin-modifying enzymes and pectin-derived materials: applications and impacts. *Applied Microbiology and Biotechnology*, **98**: 519–532.

Boonchoo, S., Fukai, S., Hetherington, S.E. (1998). Barley yield and grain protein concentration as affected by assimilate and nitrogen availability. *Australian Journal of Agricultural Research*, **49**: 695–706.

Booysen, C., Dicks, L.M.T., Meijering, I., Ackermann, A. (2002). Isolation, identification and changes in the composition of lactic acid bacteria during the malting of two different barley cultivars. *International Journal of Food Microbiology*, **76**: 63–73.

Boulton, C. (2006). 11 - Fermentation of beer. In: Bamforth, C.W. (ed) *Brewing*. Woodhead Publishing Series in Food Science, Technology and Nutrition. Woodhead Publishing, pp. 228–253.

Braun-Kiewnick, A., Jacobsen, B.J., Sands, D.C. (2000). Biological Control of *Pseudomonas syringae* pv. *syringae*, the Causal Agent of Basal Kernel Blight of Barley, by Antagonistic *Pantoea agglomerans*. *Phytopathology*, **90**: 368–375.

Briggs, D.E. (1998). *Malts and Malting*. Springer Science & Business Media, 832 pp.

Briggs, D.E., Brookes, P.A., Stevens, R., Boulton, C.A. (2004). *Brewing: Science and Practice*. Elsevier, 900 pp.

Briggs, D.E., Hough, J.S., Stevens, R., Young, T.W. (1981). *Malting and Brewing Science: Malt and Sweet Wort*. Springer Science & Business Media, 404 pp.

Brookes, P.A., Lovett, D.A., MacWilliam, I.C. (1976). The Steeping of Barley. a Review of the Metabolic Consequences of Water Uptake, and Their Practical Implications. *Journal of the Institute of Brewing*, **82**: 14–26.

Buckee, G.K. (1994). Determination of Total Nitrogen in Barley, Malt and Beer by Kjeldahl Procedures and the Dumas Combustion Methodcollaborative Trial. *Journal of the Institute of Brewing*, **100**: 57–64.

Burger, W.C., LaBerge, D.E. (1985). Malting and Brewing Quality. *Barley*. John Wiley & Sons, Ltd, pp. 367–401.

Bziuk, N., Maccario, L., Straube, B., Wehner, G., Sørensen, S.J., Schikora, A., Smalla, K. (2021). The treasure inside barley seeds: microbial diversity and plant beneficial bacteria. *Environmental Microbiome*, **16**: 20.

Calderbank, J., Hammond, J.R.M. (1989). Influence of Nitrate and Bacterial Contamination on the Formation of Apparent Total N-Nitroso Compounds (atnc) During Fermentation. *Journal of the Institute of Brewing*, **95**: 277–281.

Chandel, A., Mann, R., Kaur, J., Norton, S., Edwards, J., Spangenberg, G., Sawbridge, T. (2021). Implications of Seed Vault Storage Strategies for Conservation of Seed Bacterial Microbiomes. *Frontiers in Microbiology*, **12**.

Chen, Wen, Cheung, H.Y.K., McMillan, M., Turkington, T.K., Izydorczyk, M.S., Gräfenhan, T. (2022). The dynamics of indigenous epiphytic bacterial and fungal communities of barley grains through the commercial malting process in Western Canada. *Current Research in Food Science*, **5**: 1352–1364.

Chen, Yulan, Liang, J., Zia, A., Gao, X., Wang, Y., Zhang, L., Xiang, Q., Zhao, K., Yu, X., Chen, Q., Penttinen, P., Nyima, T., Gu, Y. (2022). Culture dependent and independent characterization of endophytic bacteria in the seeds of highland barley. *Frontiers in Microbiology*, **13**.

Chimwamurombe, P.M., Grönemeyer, J.L., Reinhold-Hurek, B. (2016). Isolation and characterization of culturable seed-associated bacterial endophytes from gnotobiotically grown Marama bean seedlings. *FEMS Microbiology Ecology*, **92**: fiw083.

Collins, K.J.E., Rogerson, C.M., Rogerson, J. (2023). The Evolution of the Craft Beer Industry in the Global South: The Experience of South Africa. *Studia Periegetica*, **44**: 49–72.

Compant, S., Samad, A., Faist, H., Sessitsch, A. (2019). A review on the plant microbiome: Ecology, functions, and emerging trends in microbial application. *Journal of Advanced Research*, **19**: 29–37.

Dar, M.A., Dhole, N.P., Xie, R., Pawar, K.D., Ullah, K., Rahi, P., Pandit, R.S., Sun, J. (2021). Valorization Potential of a Novel Bacterial Strain, *Bacillus altitudinis* RSP75, towards Lignocellulose Bioconversion: An Assessment of Symbiotic Bacteria from the Stored Grain Pest, *Tribolium castaneum*. *Microorganisms*, **9**: 1952.

Desai, N., Antonopoulos, D., Gilbert, J.A., Glass, E.M., Meyer, F. (2012). From genomics to metagenomics. *Current Opinion in Biotechnology*, **23**: 72–76.

Díaz Herrera, S., Grossi, C., Zawoznik, M., Groppa, M.D. (2016). Wheat seeds harbour bacterial endophytes with potential as plant growth promoters and biocontrol agents of Fusarium graminearum. *Microbiological Research*, **186–187**: 37–43.

Domin, M., Kluza, F., Góral, D., Nazarewicz, S., Kozłowicz, K., Szmigielski, M., Ślaska-Grzywna, B. (2020). Germination Energy and Capacity of Maize Seeds Following Low-Temperature Short Storage. *Sustainability*, **12**: 46.

Doran, P.J., Briggs, D.E. (1993). Microbes and Grain Germination. *Journal of the Institute of Brewing*, **99**: 165–170.

Dutilloy, E., Oni, F.E., Esmaeel, Q., Clément, C., Barka, E.A. (2022). Plant Beneficial Bacteria as Bioprotectants against Wheat and Barley Diseases. *Journal of Fungi*, **8**: 632.

Ebeling, M.E. (1968). The Dumas Method for Nitrogen in Feeds. *Journal of Association of Official Analytical Chemists*, **51**: 766–770.

Edney, M.J., O'Donovan, J.T., Turkington, T.K., Clayton, G.W., McKenzie, R., Juskiw, P., Lafond, G.P., Brandt, S., Grant, C.A., Harker, K.N., Johnson, E., May, W. (2012). Effects of seeding rate, nitrogen rate and cultivar on barley malt quality. *Journal of the Science of Food and Agriculture*, **92**: 2672–2678.

Evans, D. e., Sheehan, M. c., Stewart, D. c. (1999). The Impact of Malt Derived Proteins on Beer Foam Quality. Part II: The Influence of Malt Foam-positive Proteins and Non-starch Polysaccharides on Beer Foam Quality. *Journal of the Institute of Brewing*, **105**: 171–178.

Ezzeldin, S., El-Wazir, A., Enany, S., Muhammad, A., Johar, D., Osama, A., Ahmed, E., Shikshaky, H., Magdeldin, S. (2019). Current Understanding of Human Metaproteome Association and Modulation. *Journal of Proteome Research*, **18**: 3539–3554.

Fadiji, A.E., Babalola, O.O. (2020). Metagenomics methods for the study of plant-associated microbial communities: A review. *Journal of Microbiological Methods*, **170**: 105860.

Felšöciová, S., Kowalczewski, P.Ł., Krajčovič, T., Dráb, Š., Kačániová, M. (2021). Effect of Long-Term Storage on Mycobiota of Barley Grain and Malt. *Plants (Basel, Switzerland)*, **10**: 1655.

Felšöciová, S., Kowalczewski, P.Ł., Krajčovič, T., Dráb, Š., Kačániová, M. (2020). Quantitative and Qualitative Composition of Bacterial Communities of Malting Barley Grain and Malt during Long-Term Storage. *Agronomy*, **10**: 1301.

Finkel, T. (2011). Signal transduction by reactive oxygen species. *Journal of Cell Biology*, **194**: 7–15.

Flannigan, B. (1969). Microflora of dried barley grain. *Transactions of the British Mycological Society*, **53**: 371–379.

Flannigan, B. (1978). Primary contamination of barley and wheat grain storage fungi. *Transactions of the British Mycological Society*, **71**: 37–42.

Flannigan, B. (2003). The microbiota of barley and malt. In: Priest, F.G. & Campbell, I. (eds) *Brewing Microbiology*. Springer US, Boston, MA, pp. 113–180.

Fox, G.P. (2010). Chemical Composition in Barley Grains and Malt Quality. In: Zhang, G. & Li, C. (eds) *Genetics and Improvement of Barley Malt Quality*. Advanced Topics in Science and Technology in China. Springer, Berlin, Heidelberg, pp. 63–98.

Fröhlich, A., Buddrus-Schiemann, K., Durner, J., Hartmann, A., von Rad, U. (2012). Response of barley to root colonization by *Pseudomonas* sp. DSMZ 13134 under laboratory, greenhouse, and field conditions. *Journal of Plant Interactions*, **7**: 1–9.

Ghali, A. (2014). *Circular Storage Tanks and Silos, Third Edition*. CRC Press, 392 pp.

Gnonlonfoun, E., Fotin, G., Schwebel, S., Colin, C., Salles, M., Parant, V., Alonso, P., Moreau, F., Borges, F., Revol-Junelles, A.-M., Schmitt, M., Framboisier, X., Fick, M., Rondags, E. (2023). Impact of *Erwinia gerundensis* as a Biocontrol Agent on the Sanitary and Technological Quality of Barley Malt. *Journal of the American Society of Brewing Chemists*, **0**: 1–14.

Gorjanović, S., Sužnjević, D., Beljanski, M., Ostojić, S., Gorjanović, R., Vrvić, M., Hranisavljević, J. (2004). Effects of Lipid-Transfer Protein from Malting Barley Grain on Brewers Yeast Fermentation. *Journal of the Institute of Brewing*, **110**: 297–302.

Graw, S., Chappell, K., Washam, C.L., Gies, A., Bird, J., Robeson, M.S., Byrum, S.D. (2021). Multi-omics data integration considerations and study design for biological systems and disease. *Molecular Omics*, **17**: 170–185.

Guine, R. de P.F., Correia, P.M. dos R. (2013). *Engineering Aspects of Cereal and Cereal-Based Products*. CRC Press, 370 pp.

Guo, B., Li, D., Lin, S., Li, Y., Wang, S., Lv, C., Xu, R. (2019). Regulation of nitrogen availability results in changes in grain protein content and grain storage subproteomes in barley (*Hordeum vulgare* L.). *PloS One*, **14**: e0223831.

Gyllang, H., Saetmark, L., Martinson, E. (1977). The influence of some fungi on malt quality.

Haikara, A., Home, S. (1991). Mash filtration difficulties caused by split barley kernels, a microbiological problem. *Mallas ja Olut*, 133–141.

Hao, Z., Wang, Y., Guo, X., De, J. (2022). Deciphering the core seed endo-bacteriome of the highland barley in Tibet plateau. *Frontiers in Plant Science*, **13**.

Hawthorne, D.B., Shaw, R.D., Davine, D.F., Kavanagh, T.E., Clarke, B.J. (1991). Butyric Acid Off-Flavors in Beer: Origins and Control1. *Journal of the American Society of Brewing Chemists*, **49**: 4–8.

Henry, R.J. (2016). Barley: Harvesting, Storage, and Transport. *Reference Module in Food Science*. Elsevier.

Heyer, R., Kohrs, F., Reichl, U., Benndorf, D. (2015). Metaproteomics of complex microbial communities in biogas plants. *Microbial Biotechnology*, **8**: 749–763.

Hodgson, S., de Cates, C., Hodgson, J., Morley, N.J., Sutton, B.C., Gange, A.C. (2014). Vertical transmission of fungal endophytes is widespread in forbs. *Ecology and Evolution*, **4**: 1199–1208.

Hornsey, I.S. (2003). *A History of Beer and Brewing*. Royal Society of Chemistry, 764 pp.

Hudson, O.P. (1986). Malting Technology. *Journal of the Institute of Brewing*, **92**: 115–122.

Kassie, M., Tesfaye, K. (2019). Malting Barley Grain Quality and Yield Response to Nitrogen Fertilization in the Arsi Highlands of Ethiopia. *Journal of Crop Science and Biotechnology*, **22**: 225–234.

Kaur, M., Bowman, J.P., Stewart, D.C., Sheehy, M., Janusz, A., Speers, R.A., Koutoulis, A., Evans, D.E. (2012). TRFLP analysis reveals that fungi rather than bacteria are associated with premature yeast flocculation in brewing. *Journal of Industrial Microbiology and Biotechnology*, **39**: 1821–1832.

Kaushal, M., Mahuku, G., Swennen, R. (2020). Metagenomic Insights of the Root Colonizing Microbiome Associated with Symptomatic and Non-Symptomatic Bananas in *Fusarium* Wilt Infected Fields. *Plants*, **9**: 263.

Kawaguchi, A., Abe, D., Saito, T., Nogata, Y., Nomiyama, K., Kohyama, N., Takahashi, A., Yoshioka, T., Ishikawa, N., Tomioka, K. (2021). Pink seed of barley caused by *Erwinia persicina*. *Journal of General Plant Pathology*, **87**: 106–109.

Kelly, L., Briggs, D.E. (1992). The Influence of the Grain Microflora on the Germinative Physiology of Barley. *Journal of the Institute of Brewing*, **98**: 395–400.

Khokonova, M.B. (2011). Effect of seed planting depth on brewing qualities of barley grain and malt. *Russian Agricultural Sciences*, **37**: 436–438.

Khokonova, M.B., Karashaeva, A.S., Zavalin, A.A. (2015). Quality of brewing malt depending on the storage conditions of barley. *Russian Agricultural Sciences*, **41**: 508–511.

Kibite, S., Evans, L.E. (1984). Causes of negative correlations between grain yield and grain protein concentration in common wheat. *Euphytica*, **33**: 801–810.

Kubicek, C.P., Starr, T.L., Glass, N.L. (2014). Plant Cell Wall–Degrading Enzymes and Their Secretion in Plant-Pathogenic Fungi. *Annual Review of Phytopathology*, **52**: 427–451.

Kumar, D., Kumar, V., Verma, R.P.S., Kharub, A.S., Sharma, I. (2013). Quality parameter requirement and standards for malt barley - a review. *Agricultural Reviews*, **34**: 313–317.

Laitila, A., Kotaviita, E., Peltola, P., Home, S., Wilhelmson, A. (2007). Indigenous Microbial Community of Barley Greatly Influences Grain Germination and Malt Quality. *Journal of the Institute of Brewing*, **113**: 9–20.

Laitila, A., Manninen, J., Priha, O., Smart, K., Tsitko, I., James, S. (2018). Characterisation of barley-associated bacteria and their impact on wort separation performance. *Journal of the Institute of Brewing*, **124**: 314–324.

Laus, A., Endres, F., Hutzler, M., Zarnkow, M., Jacob, F. (2022). Isothermal Mashing of Barley Malt: New Insights into Wort Composition and Enzyme Temperature Ranges. *Food and Bioprocess Technology*, **15**: 2294–2312.

Li, Y., Wang, C., Chang, H., Zhang, Y., Liu, S., He, W. (2023). Metagenomics reveals the effect of long-term fertilization on carbon cycle in the maize rhizosphere. *Frontiers in Microbiology*, **14**.

Liu, C., Wang, J., Wang, Y., Li, L., Feng, Z., Xian, Y., Jiang, Y., Yu, J., Tong, T., Li, X., Yao, M. (2024). Crop rotation and fertilization shape the microbiomes of maize rhizosphere soil with distinct mechanisms. *Plant and Soil*.

Lowe, D.P., Ulmer, H.M., Barta, R.C., Goode, D.L., Arendt, E.K. (2005). Biological Acidification of a Mash Containing 20% Barley Using *Lactobacillus amylovorus* FST 1.1: Its Effects on Wort and Beer Quality. *Journal of the American Society of Brewing Chemists*, **63**: 96–106.

Lv, L., Luo, J., Ahmed, T., Zaki, H.E.M., Tian, Y., Shahid, M.S., Chen, J., Li, B. (2022). Beneficial Effect and Potential Risk of *Pantoea* on Rice Production. *Plants*, **11**: 2608.

MacIntosh, A.J., MacLeod, A., Beattie, A.D., Eck, E., Edney, M., Rossnagel, B., Speers, R.A. (2014). Assessing the Effect of Fungal Infection of Barley and Malt on Premature Yeast Flocculation1. *Journal of the American Society of Brewing Chemists*, **72**: 66–72.

Magliano, P.N., Prystupa, P., Gutiérrez-Boem, F.H. (2014). Protein content of grains of different size fractions in malting barley. *Journal of the Institute of Brewing*, **120**: 347–352.

Mallett, J. (2014). *Malt: A Practical Guide from Field to Brewhouse*. Brewers Publications, 298 pp.

Marchesi, J.R., Ravel, J. (2015). The vocabulary of microbiome research: a proposal. *Microbiome*, **3**: 31.

Marcó, A., Rubio, R., Compañó, R., Casals, I. (2002). Comparison of the Kjeldahl method and a combustion method for total nitrogen determination in animal feed. *Talanta*, **57**: 1019–1026.

Miranda-Carrazco, A., Navarro-Noya, Y.E., Govaerts, B., Verhulst, N., Dendooven, L. (2022). Nitrogen Fertilizer Application Alters the Root Endophyte Bacterial Microbiome in Maize Plants, but Not in the Stem or Rhizosphere Soil. *Microbiology Spectrum*, **10**: e01785-22.

Mwazha, M., Mugadza, D.T., Manhokwe, S., Njini, M., Jombo, T.Z. (2022). Comparison of the Changes in Seed Germination Vigour with Prolonged Postharvest Storage for *Hordeum vulgare* Varieties, Hope and Sierra Malting Barley Varieties, in Zimbabwe. *Journal of Food Quality*, **2022**: e9612855.

Newman. (2006). A Brief History of Barley Foods. *Cereal Foods World*.

van Nierop, S.N.E., Cameron-Clarke, A., Axcell, B.C. (2004). Enzymatic Generation of Factors from Malt Responsible for Premature Yeast Flocculation. *Journal of the American Society of Brewing Chemists*, **62**: 108–116.

Noots, I., Delcour, J.A., Michiels, C.W. (1999). From field barley to malt: detection and specification of microbial activity for quality aspects. *Critical Reviews in Microbiology*, **25**: 121–153.

O'Sullivan, T. f., Walsh, Y., O'Mahony, A., Fitzgerald, G. f., van Sinderen, D. (1999). A Comparative Study of Malthouse and Brewhouse Microflora. *Journal of the Institute of Brewing*, **105**: 55–61.

Oukala, N., Aissat, K., Pastor, V. (2021). Bacterial Endophytes: The Hidden Actor in Plant Immune Responses against Biotic Stress. *Plants*, **10**: 1012.

Ouyang, Y., Norton, J.M. (2020). Short-Term Nitrogen Fertilization Affects Microbial Community Composition and Nitrogen Mineralization Functions in an Agricultural Soil. *Applied and Environmental Microbiology*, **86**: e02278-19.

Pan, H., Wattiez, R., Gillan, D. (2024). Soil Metaproteomics for Microbial Community Profiling: Methodologies and Challenges. *Current Microbiology*, **81**: 257.

Panteloglou, A.G., Smart, K.A., Cook, D.J. (2012). Malt-induced premature yeast flocculation: current perspectives. *Journal of Industrial Microbiology & Biotechnology*, **39**: 813–822.

Parés Viader, R., Yde, M.S.H., Hartvig, J.W., Pagenstecher, M., Carlsen, J.B., Christensen, T.B., Andersen, M.L. (2021). Optimization of Beer Brewing by Monitoring α-Amylase and β-Amylase Activities during Mashing. *Beverages*, **7**: 13.

Pham, V.T.K., Rediers, H., Ghequire, M.G.K., Nguyen, H.H., De Mot, R., Vanderleyden, J., Spaepen, S. (2017). The plant growth-promoting effect of the nitrogen-fixing endophyte *Pseudomonas stutzeri* A15. *Archives of Microbiology*, **199**: 513–517.

Pitzschke, A. (2016). Developmental Peculiarities and Seed-Borne Endophytes in Quinoa: Omnipresent, Robust Bacilli Contribute to Plant Fitness. *Frontiers in Microbiology*, **7**.

Pohlman, A.A., McColl, J.G. (1982). Nitrogen fixation in the rhizosphere and rhizoplane of barley. *Plant and Soil*, **69**: 341–352.

Poutanen, K.S. (2020). 5 - Cereal raw material pretreatment. In: Perdon, A.A., Schonauer, S.L., & Poutanen, K.S. (eds) *Breakfast Cereals and How They Are Made (Third Edition)*. AACC International Press, Oxford, pp. 97–107.

Priest, F.G., Cowbourne, M.A., Hough, J.S. (1974). Wort Enterobacteria—a Review. *Journal of the Institute of Brewing*, **80**: 342–356.

Qin, Q., Liu, J., Hu, S., Yu, J., Zhang, L., Huang, S., Yang, M., Wang, L. (2023). Exploring Fungal Species Diversity in the Premature Yeast Flocculation (PYF) of Barley Malt Using Deep Sequencing. *Journal of the American Society of Brewing Chemists*, **81**: 333–339.

Quan, X., Zeng, J., Ye, L., Chen, G., Han, Z., Shah, J.M., Zhang, G. (2016). Transcriptome profiling analysis for two Tibetan wild barley genotypes in responses to low nitrogen. *BMC Plant Biology*, **16**: 30.

Rani, H., Bhardwaj, R.D., Kaur, S., Grewal, S.K. (2022). Investigating diverse barley (*Hordeum vulgare* L.) germplasm for thermostability of β-amylase: A key player of diastatic power. *Cereal Chemistry*, **99**: 1339–1348.

Saeed, Q., Xiukang, W., Haider, F.U., Kučerik, J., Mumtaz, M.Z., Holatko, J., Naseem, M., Kintl, A., Ejaz, M., Naveed, M., Brtnicky, M., Mustafa, A. (2021). Rhizosphere Bacteria in Plant Growth Promotion, Biocontrol, and Bioremediation of Contaminated Sites: A Comprehensive Review of Effects and Mechanisms. *International Journal of Molecular Sciences*, **22**: 10529.

Salvato, F., Hettich, R.L., Kleiner, M. (2021). Five key aspects of metaproteomics as a tool to understand functional interactions in host-associated microbiomes. *PLoS pathogens*, **17**: e1009245.

Schwarz, P., Li, Y. (2010). Malting and Brewing Uses of Barley. *Barley*. John Wiley & Sons, Ltd, pp. 478–521.

Schwarz, Z.J., Paul. (2020). Advances in postharvest storage and handling of barley: methods to prevent or reduce mycotoxin contamination. *Achieving Sustainable Cultivation of Barley*. Burleigh Dodds Science Publishing.

Scott, B., Takemoto, D., Tanaka, A. (2007). Fungal Endophyte Production of Reactive Oxygen Species is Critical for Maintaining the Mutualistic Symbiotic Interaction Between Epichloë festucae and Perennial Ryegrass. *Plant Signaling & Behavior*, **2**: 171–173.

Shahzad, R., Waqas, M., Khan, A.L., Asaf, S., Khan, M.A., Kang, S.-M., Yun, B.-W., Lee, I.-J. (2016). Seed-borne endophytic *Bacillus amyloliquefaciens* RWL-1 produces gibberellins and regulates endogenous phytohormones of *Oryza sativa*. *Plant Physiology and Biochemistry*, **106**: 236–243.

Shala, N., Dreshaj, A., Hoxha, I., Sejfijaj, O., Millaku, B. (2023). The Impact of Climate Conditions on the Yield of Some Fall Barley Cultivars for the Production of Beer in the Kosovo Plain. *Ecological Engineering & Environmental Technology*, **24**: 261–268.

Shu, K., Meng, Y.J., Shuai, H.W., Liu, W.G., Du, J.B., Liu, J., Yang, W.Y. (2015). Dormancy and germination: How does the crop seed decide? *Plant Biology (Stuttgart, Germany)*, **17**: 1104–1112.

Sileoni, V., Marconi, O., Perretti, G. (2015). Near-infrared Spectroscopy in the Brewing Industry. *Critical Reviews in Food Science and Nutrition*, **55**: 1771–1791.

Singh, P., Singh, R.K., Li, H.-B., Guo, D.-J., Sharma, A., Lakshmanan, P., Malviya, M.K., Song, X.-P., Solanki, M.K., Verma, K.K., Yang, L.-T., Li, Y.-R. (2021). Diazotrophic Bacteria *Pantoea dispersa* and *Enterobacter asburiae* Promote Sugarcane Growth by Inducing Nitrogen Uptake and Defense-Related Gene Expression. *Frontiers in Microbiology*, **11**.

Smith, N.A., Smith, P., Woodruff, C.A. (1992). The role of *Bacillus* spp. In n-nitrosamine formation during wort production. *Journal of the Institute of Brewing*, **98**: 409–414.

Söderberg, K.H., Bååth, E. (2004). The influence of nitrogen fertilisation on bacterial activity in the rhizosphere of barley. *Soil Biology and Biochemistry*, **36**: 195–198.

Solanki, M.K., Abdelfattah, A., Britzi, M., Zakin, V., Wisniewski, M., Droby, S., Sionov, E. (2019). Shifts in the Composition of the Microbiota of Stored Wheat Grains in Response to Fumigation. *Frontiers in Microbiology*, **10**.

Stewart, G.G., Russell, I., Anstruther, A. (2017). *Handbook of Brewing*. CRC Press, 799 pp.

Taheri, E., Tarighi, S., Taheri, P. (2022). Characterization of root endophytic *Paenibacillus polymyxa* isolates with biocontrol activity against *Xanthomonas translucens* and *Fusarium graminearum*. *Biological Control*, **174**: 105031.

Toews, W.H., Soper, R.J. (1978). Effects of nitrogen source, method of placement and soil type on seedling emergence and barley crop yields. *Canadian Journal of Soil Science*, **58**: 311–320.

Trivedi, P., Leach, J.E., Tringe, S.G., Sa, T., Singh, B.K. (2020). Plant–microbiome interactions: from community assembly to plant health. *Nature Reviews Microbiology*, **18**: 607–621.

Truyens, S., Weyens, N., Cuypers, A., Vangronsveld, J. (2015). Bacterial seed endophytes: genera, vertical transmission and interaction with plants. *Environmental Microbiology Reports*, **7**: 40–50.

Tshisekedi, K.A., De Maayer, P., Botes, A. (2024). Metagenomic sequencing and reconstruction of 82 microbial genomes from barley seed communities. *Scientific Data*, **11**: 484.

Tuomi, T., Laakso, S., Rosenqvist, H. (1995). Plant Hormones in Fungi and Bacteria from Malting Barley. *Journal of the Institute of Brewing*, **101**: 351–357.

Turner, R.E., Ebelhar, M.W., Wilkerson, T., Bellaloui, N., Golden, B.R., Irby, J.T., Martin, S. (2020). Effects of Purple Seed Stain on Seed Quality and Composition in Soybean. *Plants*, **9**: 993.

Van Nierop, S.N.E., Rautenbach, M., Axcell, B.C., Cantrell, I.C. (2006). The Impact of Microorganisms on Barley and Malt Quality—A Review. *Journal of the American Society of Brewing Chemists*, **64**: 69–78.

de Vries, R.P., Visser, J. (2001). Aspergillus Enzymes Involved in Degradation of Plant Cell Wall Polysaccharides. *Microbiology and Molecular Biology Reviews*, **65**: 497–522.

Walton, J.D. (1994). Deconstructing the Cell Wall. *Plant Physiology*, **104**: 1113–1118.

War, A.F., Bashir, I., Reshi, Z.A., Kardol, P., Rashid, I. (2023). Insights into the seed microbiome and its ecological significance in plant life. *Microbiological Research*, **269**: 127318.

Wassermann, B., Rybakova, D., Adam, E., Zachow, C., Bernhard, M., Müller, M., Mancinelli, R., Berg, G. (2021). Studying Seed Microbiomes. In: Carvalhais, L.C. & Dennis, P.G. (eds) *The Plant Microbiome: Methods and Protocols*. Methods in Molecular Biology. Springer US, New York, NY, pp. 1–21.

Wunderlich, S., Back, W. (2009). 1 - Overview of Manufacturing Beer: Ingredients, Processes, and Quality Criteria. In: Preedy, V.R. (ed) *Beer in Health and Disease Prevention*. Academic Press, San Diego, pp. 3–16.

Xie, J., Hong, Y., Gu, Z., Cheng, L., Li, Z., Li, C., Ban, X. (2023). Highland Barley Starch: Structures, Properties, and Applications. *Foods*, **12**: 387.

Yangcheng, H., Gong, L., Zhang, Y., Jane, J. (2016). Pysicochemical properties of Tibetan hull-less barley starch. *Carbohydrate Polymers*, **137**: 525–531.

Yu, W., Tan, X., Zou, W., Hu, Z., Fox, G.P., Gidley, M.J., Gilbert, R.G. (2017). Relationships between protein content, starch molecular structure and grain size in barley. *Carbohydrate Polymers*, **155**: 271–279.

Yu, Y.-M., Fukagawa, N.K. (2020). Chapter 2 - Protein and amino acids. In: Marriott, B.P., Birt, D.F., Stallings, V.A., & Yates, A.A. (eds) *Present Knowledge in Nutrition (Eleventh Edition)*. Academic Press, pp. 15–35.

Zhang, X., Li, L., Butcher, J., Stintzi, A., Figeys, D. (2019). Advancing functional and translational microbiome research using meta-omics approaches. *Microbiome*, **7**: 154.

Zhu, F. (2017). Barley Starch: Composition, Structure, Properties, and Modifications. *Comprehensive Reviews in Food Science and Food Safety*, **16**: 558–579.

# Chapter 2 Microbial community dynamics of stored barley seeds with high and low nitrogen content

## 2.1 Abstract

The impact of storage conditions on barley quality is pivotal for beer production, with the microbial composition during storage being a critical factor. This study aimed to elucidate the shifts in microbial communities associated with barley seeds, with high and low nitrogen grain contents, stored for varying periods. To achieve this, whole metagenomes of barley seeds were sequenced under different nitrogen and storage time conditions, and the resulting reads were taxonomically classified using the Kraken2 and Bracken pipelines. A notable bacterial predominance was observed, particularly with taxa from the genera *Erwinia*, *Pantoea*, *Pseudomonas* and *Stenotrophomonas*. Distinct microbial genera were more prevalent in seeds with high nitrogen content compared to those with low nitrogen content. Furthermore, notable changes in the microbial community structure correlated with the duration of storage. The detection of lytic phages suggests possible impacts on bacterial populations and overall community structures. Fungi were absent from seeds at harvest and those with high nitrogen content. These findings could provide brewers with insights for optimising storage conditions and time to potentially enhance malt quality.

## 2.2 Introduction

The quality of barley grain plays a pivotal role in determining the flavour profile of both malt and beer (Bamforth, 2017). Compromised barley quality, due to suboptimal cultivation and storage, directly affects the brewing process, yielding beer with poor flavour (Briggs *et al.*, 1981; Briggs *et al.*, 2004; Bamforth, 2017). This poses major financial risks to the brewing industry (Langridge, 2018). Following harvest, barley seeds may be stored for a period of time before malting to break seed dormancy (Felšöciová *et al.*, 2021). Storage conditions such as aeration level, integumentary grain tissue integrity, moisture content and temperature, are strictly monitored and controlled (Felšöciová *et al.*, 2021; Abubakar, 2019).

Seed spoilage during storage can be caused by a variety of microorganisms, including bacteria, yeasts, and filamentous fungi (Suzuki, 2020; Pitt and Hocking, 1997). These microorganisms

can be introduced to the seed at various stages, including during cultivation, harvesting and postharvest handling (Tyc *et al.*, 2020). The presence of fungi is especially problematic as they not only causes spoiling but can also introduce mycotoxins, which pose a health risk to both humans and animals (Bingham *et al.*, 2012; Felšöciová *et al.*, 2021). Previous studies have identified a wide range of fungi associated with barley at various stages, including storage and malting (Laitila, 2015; Suzuki, 2020). These fungi include *Alternaria*, *Aspergillus*, *Fusarium*, *Geotrichum* and *Penicillium* (Laitila, 2015; Laitila *et al.*, 2007; Felšöciová *et al.*, 2021). The presence of these toxigenic fungi emphasises the importance of strict storage conditions and times to preserve the quality of barley grain (Laitila, 2015).

Bacteria play an important role in seed deterioration during storage with certain species having a particularly strong influence. Among these, *Pseudomonas syringae* is well known for impairing seeds through the production of phytotoxins and extracellular polysaccharides that promote infection and colonisation (Afkhamifar *et al.*, 2023). Its pathogenicity is further enhanced by ice nucleation activity, predisposing seeds to frost damage under cold storage conditions (Lindow *et al.*, 1982). Similarly, *Xanthomonas campestris*, with its diverse pathovars, is implicated in causing black rot in seeds (Batista *et al.*, 2021). This bacterium facilitates disease through the secretion of extracellular enzymes and xanthan gum, which collectively contribute to the disintegration of seed tissues and obstruction of the plant vascular systems (Nejadmansouri *et al.*, 2020). *Pectobacterium carotovorum*, though primarily associated with soft rot in vegetables, also poses a risk to seeds stored in humid environments by releasing a suite of plant cell wall-degrading enzymes, such as cellulases, pectinases, and proteases, leading to the deterioration of seed integrity (Wasendorf *et al.*, 2022; Ranjan and Singh, 2020).

Studies have shown that bacterial communities associated with barley seed undergo substantial changes from pre-harvest to post-storage conditions (Angelino and Convention, 1996; Bziuk *et al.*, 2021). In pre-harvest barley, *Pantoea agglomerans* is the most abundant and commonly reported species within the culturable microbial community (Flannigan, 2003). Several studies have also tracked changes within the bacterial community of stored barley and have noted increases in coliform bacteria and decreases in lactic acid bacteria (LAB) (Felšöciová *et al.*, 2020; Abubakar, 2019). These shifts, however, are largely influenced by factors such as the initial microflora composition, storage conditions and duration (Abubakar, 2019).

The nitrogen content of barley grains is a critical factor in determining the quality of beer (Adeboye *et al.*, 2011). A grain with lower nitrogen content is preferable as grains with a higher nitrogen content may have higher protein levels (Agu and Palmer, 2001) which can result in undesirable foam stability and off-flavours in the beer (Ahmed *et al.*, 2018). Microorganisms play an important part in the nitrogen cycle, turning atmospheric nitrogen into usable forms and recycling it through ecosystems, thus impacting the nitrogen content of crops such as barley (Agu and Palmer, 2001). Their activities ensure the availability of appropriate nitrogen sources, which is essential for optimising barley grain quality for beer production (Baslam *et al.*, 2021). While the impact of nitrogen content on beer quality is well-documented, the relationship between barley nitrogen content and microbial dynamics during storage is less understood.

This knowledge gap presents an opportunity for further research, which could provide valuable insights into improving barley storage and the beer production process. Here, using metagenomics approaches, the microbial community composition associated with low and high nitrogen content barley seeds from harvest through to nine months of storage was determined.

## 2.3 Methodology

### 2.3.1 Sample collection and processing

Malting barley (*Hordeum vulgare* L.) samples of a single cultivar (Kadie) were sourced from Anheuser-Busch InBev (AB InBev) storage facilities in the Western Cape Province, South Africa. The collected samples were derived from four distinct time points, including barley seeds at harvest as well as those stored in silos for three, six and nine months, respectively (Table S1). At each time point, barley seeds were selected based on their nitrogen content; those with nitrogen contents < 1.5% and > 1.5% were classified as having low and high nitrogen contents, respectively. Nitrogen content classification was determined by AB InBev. The low and high nitrogen samples collected at different time points originated from a single harvest but were stored in different locations. Despite this, storage conditions were standardised across facilities to minimise variability and maintain comparability. All samples were aseptically collected and stored at -20°C to inhibit microbial growth.

### 2.3.2 DNA isolation, library preparation and sequencing

A sterile mortar and pestle were used to crush ~10 g of each barley sample. The residue was suspended in 40 ml of phosphate buffered saline (PBS; pH7.4, 0.137M NaCl, 0.01M $Na_2HPO_4$ and 0.0018M $KH_2PO_4$). The samples were then sonicated for 7 min at 18 W amplitude in a 30 second on-off pulsating schedule in a Microson Ultrasonic Cell Disruptor (Wazobia Enterprise, USA). Following sonication, the mixtures were centrifuged for 1 min at 4000 x g to separate the supernatant from the solid residue. The supernatants were transferred to a filtration system comprised of an autoclaved polycarbonate filter holder and a 0.45 µm pore Sartorius-Stedim Biotech filter membrane. The microbial cells collected on the filter paper were transferred to sterile Eppendorf tubes. Metagenomic DNA (mDNA) extraction was performed on these samples using the ZymoBIOMICS DNA/RNA Miniprep Kit (Zymo Research, USA) according to the manufacturer's protocol.

Library preparation and sequencing were carried out at Molecular Research LP (MR DNA, Texas, USA). Briefly, the mDNA concentration was evaluated using the Qubit® dsDNA HS Assay Kit (Life Technologies, USA). A total of 50 ng mDNA per sample was used to prepare the libraries, where samples underwent simultaneous fragmentation and addition of adapter sequences (Nextera, Illumina, USA). These adapters were used during a limited-cycle polymerase chain reaction (PCR) in which unique indices were added to the sample. Following the library preparation, the final concentration of the libraries was measured using the Qubit® dsDNA HS Assay Kit, and the average library size was determined using the Agilent 2100 Bioanalyzer (Agilent Technologies, USA). The libraries were then pooled in equimolar ratios of 0.6nM and sequenced using the Illumina NovaSeq 6000 system with paired end configuration (2 × 250 bp) for 500 cycles.

### 2.3.3 Quality trimming and host DNA contamination assessment

Raw sequence reads were evaluated for quality using FastQC (v0.12.1)(Andrews, 2010) and MultiQC (v1.15) (Ewels *et al.*, 2016). Trimmomatic v. 0.36 (Bolger *et al.*, 2014) was used to filter out reads shorter than 36 bp or with an average quality score lower than 15. Host DNA contamination in metagenomic sequences from barley seed was assessed by alignment against the barley reference genome (*H. vulgare* L., NCBI assembly accession number: GCF_904849725.1). The latter reference genome was indexed and the paired-end metagenomic reads were aligned against this index using Bowtie2 (v2.5.1) and SAMtools

(v1.19) (Langmead and Salzberg, 2012; Li *et al.*, 2009). Reads that did not align were classified as microbial, whereas aligned reads were considered as host contamination.

### 2.3.4 Taxonomic classification with Kraken2-Bracken pipeline

Taxonomic classification of the trimmed paired-end metagenomic reads was conducted using the Kraken2-Bracken pipeline (Lu *et al.*, 2022). Kraken2 is a pipeline for classifying genetic material from microbial communities, employing a k-mer-based approach to assign taxonomic labels to short DNA sequences, which are then matched against a comprehensive database (Wood *et al.*, 2019). This method is well-suited for handling the diverse range of sequences in metagenomic datasets. Kraken2 (v2.1.3) was used with default parameters (k = 31) with NCBI RefSeq (Release 202) reference database (https://www.ncbi.nlm.nih.gov/refseq/), which includes complete genomes of archaea, bacteria, fungi and viruses.

Bracken (Bayesian Reestimation of Abundance with Kraken) (v2.7.0), complementing Kraken2, was employed for more accurate taxonomic abundance estimation at different taxonomic ranks, especially at the species level. It functions by applying a statistical model to distribute reads with precision (Lu *et al.*, 2017). Post-classification, Bracken reports were converted to the MetaPhlAn format using the kreport2mpa.py script of KrakenTools (Lu *et al.*, 2022). This process involves aggregating reports for a comprehensive overview, extracting specific taxonomic ranks for clarity and optimizing file headers to remove redundant extensions. The output from Bracken, in MetaPhlAn Profile Analysis (MPA) format, was used for representing and analysing microbial community composition.

### 2.3.5 Diversity and statistical analysis

Alpha diversity refers to the variety and abundance of microbial species within individual barley grain samples, measuring the internal complexity of each sample (Xia and Sun, 2023). Beta diversity, on the other hand, compares microbial communities from different samples, emphasising the differences in microbial makeup (Maziarz *et al.*, 2018). Microbial diversity was analysed using R version 4.2.2 (RStudio Team, 2020). The MPA-formatted data were combined with sample metadata after being transformed into a long format. The Shannon Diversity Index was employed to quantify alpha diversity, reflecting species richness and evenness for each individual sample (Shannon, 1948). Calculations were performed using the diversity function in the *vegan* package (Dixon, 2003), with data categorized by sample ID, nitrogen content and storage duration. The impact of barley seed nitrogen content and storage

duration on diversity was assessed using the Kruskal-Wallis test (Kruskal and Wallis, 1952), a non-parametric method comparing diversity indices across groups to identify statistically significant differences. These results were visualized as boxplots, highlighting diversity trends in relation to barley nitrogen content and storage duration.

To investigate beta diversity, the Bray-Curtis dissimilarity matrix was used, followed by Non-metric Multidimensional Scaling (NMDS) for graphical representation (Bray and Curtis, 1957). The Kruskal-Wallis test and ANalysis Of SIMilarities (ANOSIM) were employed to evaluate the effects of barley nitrogen content and storage duration on beta diversity (Clarke, 1993). Results were presented through boxplots and ordination plots, offering insights into how these environmental factors influenced microbial diversity.

### 2.3.6 Taxonomic data analysis

The MPA data underwent filtration at both the phylum and genus levels, with the goal of calculating the relative abundance of each microorganism (Truong *et al.*, 2015). Given the negligible number of reads attributed to the domain Archaea, the study primarily concentrated on studying bacteria, fungi and viruses. The bacterial MPA data were divided into two groups. The first group, **core**, includes taxa found in all samples. The second group, **non-core**, includes taxa not found in all samples. We then identified the top ten core and non-core genera for bacteria, fungi and viruses. To better understand how different factors affect the samples, we first sorted the samples by their nitrogen content, either high or low, and created plots. Then, the same samples were grouped by storage duration, from harvest up to nine months. All data analyses and visualizations were conducted in RStudio version 4.2.2 (RStudio Team, 2020). Essential packages used included *dplyr* for data manipulation, *tidyr* for data tidying, *ggplot2* for plot creation, and *RColorBrewer* for enhancing colour schemes in visual representations.

## 2.4 Results and discussion

### 2.4.1 Metagenomic diversity and read distribution across different storage duration

A total of eight metagenomes were sequenced (Table S2.1), with a total yield of ~365.3 million reads, averaging around 45.7 million reads for each sample. The sample stored for three months with high nitrogen content (S3H) produced the highest number of reads (~76 million), while

the sample stored for nine months with low nitrogen content (S9L) yielded the lowest number (~20 million) of reads (Table S2.1). It should be noted that the average number of reads decreased by nearly 50% over the nine month storage period.

Kraken2 was used to classify the metagenomic reads for each sample (Wood *et al.*, 2019). Sequences that had a match in the database were categorised as classified, while sequences without a match were referred to as unclassified (Fig. 2.1). In most samples > 70% of the reads were classified. Sample S9L (nine months storage, low nitrogen), had the highest percentage of classified reads at 81.41%. In contrast, Sample S3L (3 months storage, low nitrogen) had the lowest percentage at 68.19%. The relatively high level of unclassified sequences may be inherent of the current microbial databases available for classification, which may not be representative of the unculturable majority (Xu *et al.*, 2023).

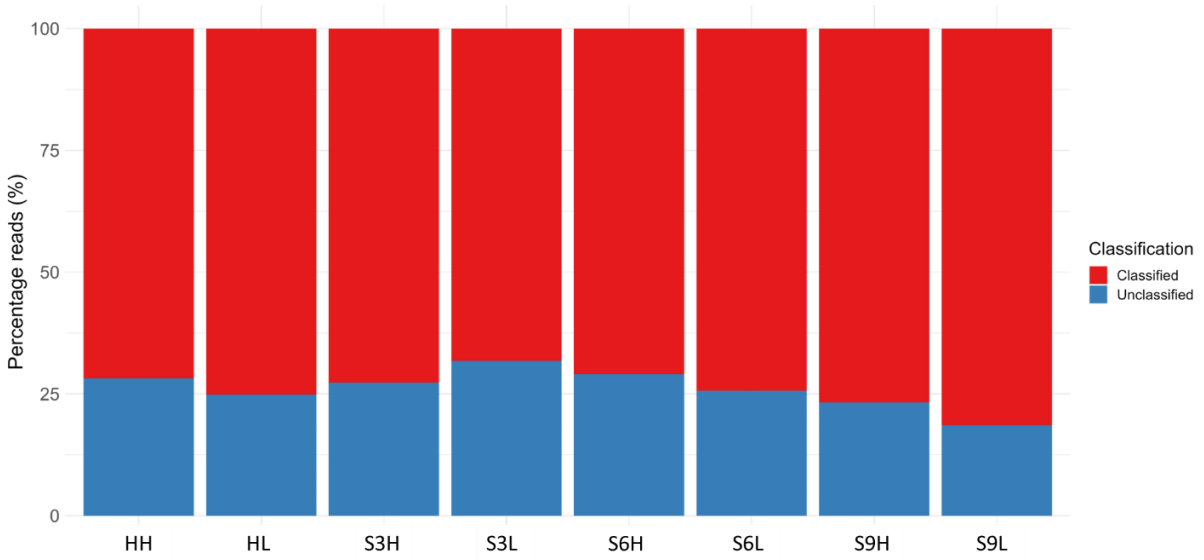


Figure 2.1 Proportional taxonomic classification of reads using the Kraken2-Bracken suite (Lu *et al.*, 2022). Each bar visually contrasts the share of classified reads (in red) against those that remained unclassified (in blue). HH/HL = harvest high/low nitrogen; S3L/S3H, S6L/S6H, S9L/S9H = three, six, nine months storage low/high nitrogen.

### 2.4.2 Microbial diversity of barley seeds is influenced by storage duration

The Shannon Diversity Index (Shannon, 1948) was used to evaluate the microbial diversity of barley grain samples with high and low nitrogen content (Fig. 2.2A). Minimal difference in mean diversity between high and low nitrogen content was observed, suggesting that nitrogen

content may not substantively impact overall microbial diversity. The Kruskal-Wallis rank sum test (Kruskal and Wallis, 1952) further confirmed that the differences in microbial diversity between high and low nitrogen content barley seeds were not statistically significant (chi-squared: 0.0833, p = 0.7728), indicating that the observed variations are likely not attributable to nitrogen content alone. However, intra-sample variability is greater among the low nitrogen samples compared to the high nitrogen samples (Fig. 2.2A; Table S2.2), pointing to greater variability in microbial composition on low nitrogen content barley seeds.

The Bray-Curtis dissimilarity metric was further applied to assess the composition of microbial communities of the barley grain samples (Bray and Curtis, 1957) (Fig. 2.2B). This showed that the microbial compositions of seeds collected at harvest were comparatively similar. Over time however, this similarity diminished, suggesting a shift in the microbial community structure with increased storage duration in seeds with different nitrogen contents. The Bray-Curtis PCoA highlighted that after six months of storage, microbial communities were significantly distinct, primarily due to a dysbiosis observed in sample S6L (six months storage and low nitrogen) (Fig. 2.2B; Table S2.3). The ANOSIM test for Bray-Curtis dissimilarity by storage indicated an R statistic of 0.5417, with a significance level of 0.034.

The diversity analyses demonstrate that storage duration, rather than nitrogen content, significantly influences the microbial diversity of barley seeds, as evidenced by minimal differences between high and low nitrogen content but notable shifts in microbial communities over time. The Bray-Curtis dissimilarity and ANOSIM tests indicate that microbial compositions become significantly distinct with extended storage, pointing to changes in community structure (Cao *et al.*, 2022; Felšöciová *et al.*, 2021). Importantly, the S6L sample, which exhibited dysbiosis, was included in all downstream analyses to maintain consistency in the study design. This sample was not discarded, as it provided valuable insight into the potential variability within the microbial communities over time. However, it is acknowledged that dysbiosis in S6L could have introduced some bias into the results, and further replication would strengthen the conclusions regarding community shifts over time. This underlines the importance of conducting microbial ecology studies with adequate replication to account for such variability (Felšöciová *et al.*, 2021). Nevertheless, the observed shifts in microbial composition suggest that factors associated with storage duration play a critical role in shaping the microbial ecosystem within barley seeds, overshadowing the impact of nitrogen content.

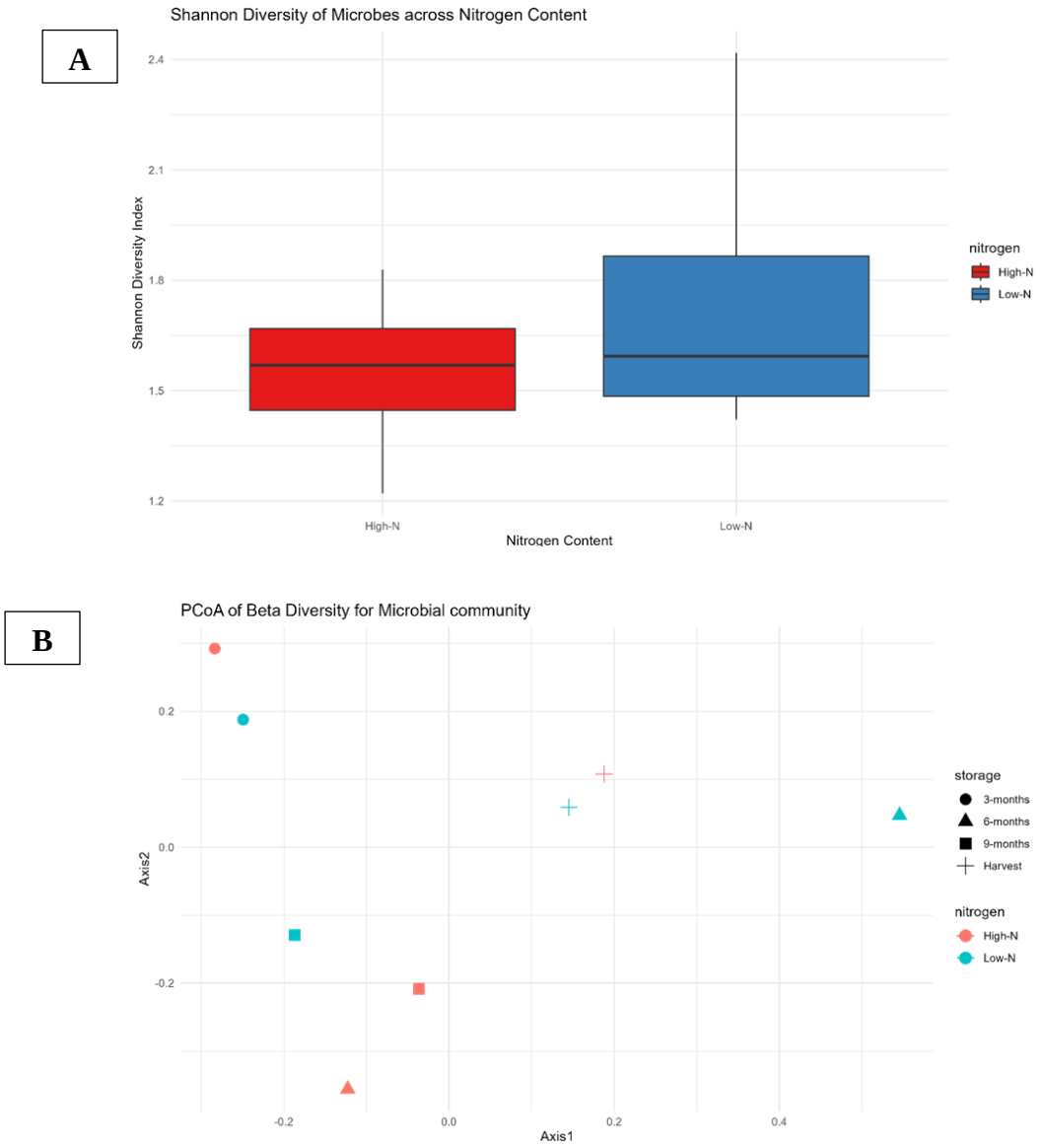


Figure 2.2 Microbial diversity and community dissimilarity in barley grains with high and low nitrogen content and at different storage durations. (A) Box plot on the basis of the Shannon diversity index reflecting microbial diversity in barley grains with high (red box) and low (blue box) nitrogen content. The central line in each box denotes the median diversity index, and the whiskers show the full range of the data. (B) Principle Component of Analysis (PCoA) plot of beta diversity measured by Bray-Curtis dissimilarity (Bray and Curtis, 1957) across the different storage sampling time points.

### 2.4.3 The seed microbiome is dominated by bacterial communities

Bacteria represented the dominant microorganisms associated with the barley seeds, contributing 99.7% of all classified reads (~111 million reads) (Fig. 2.3A, Table S2.4). A total of 381 distinct bacterial genera were observed across the eight barley metagenomic samples, encompassing 103,266,174 reads (Table S2.5). Fungi and phages were less prevalent, comprising 0.0707 and 0.1903% of the reads, respectively. Archaea, represented mainly by the Euryarchaeota, were scantily detected, contributing 0.0001% (112 reads) of the total classified reads. While this might suggest a higher bacterial abundance, it is crucial to consider the nuances of sequencing technology and experimental design. The use of short-read Illumina sequencing, particularly with 250 bp paired-end reads, inherently favours the recovery of bacterial genomes over eukaryotic ones (Jo *et al.*, 2020). This is due to the smaller, often less complex genomes of bacteria, which are more likely to be efficiently sequenced and assembled with short-read technologies. Eukaryotic genomes, being larger and more complex due to introns and repetitive sequences, might not be as thoroughly sampled or accurately assembled, leading to underrepresentation in the dataset (Borodovsky and Lomsadze, 2011; Richard, 2020). Additionally, the overall coverage of the metagenome may play a critical role; insufficient metagenomic coverage can lead to incomplete sampling of the community, disproportionately affecting organisms with larger genomes or those present in lower abundance (Knight *et al.*, 2018).

A total of twelve bacterial phyla were identified. The phyla Pseudomonadota, Bacteroidota, Actinomycetota, and Bacillota have the most widespread distribution across the barley seed samples. Pseudomonadota dominates, contributing more than 80% of the total reads in all samples. The least represented phyla are Deinococcus-Thermus, Fusobacteria, Tenericutes and Verrucomicrobia, (Fig. 2.3B, Table S2.4). The dominant phyla are not unique to barley but have previously been identified in various other seed microbiomes, including *Brassicaceae*, pumpkin, rice and wild cabbage seeds (Tyc *et al.*, 2020; Raj *et al.*, 2019; Adam *et al.*, 2018; Matsumoto *et al.*, 2021). Microorganisms associated with these phyla have been detected on the surface of seeds from a wide range of plants, including barley (Truyens *et al.*, 2015). Similar to our findings, the abundance of these phyla in barley seed was reported in a study that examined the microbiome of different barley genotypes (Bziuk *et al.*, 2021).

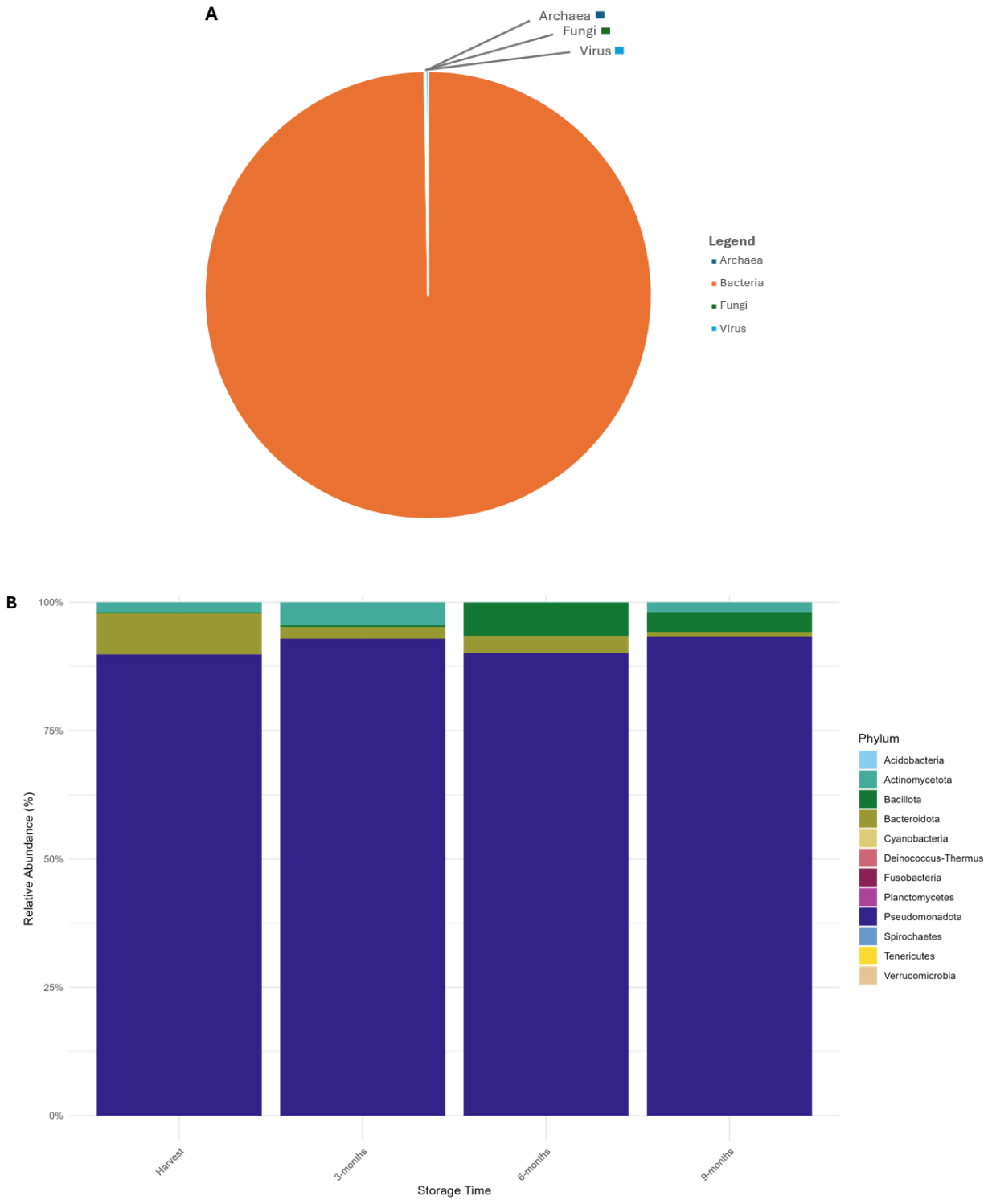


Figure 2.3 (A) Taxonomic composition of classified reads using Kraken2-Bracken suite (Lu *et al.*, 2022). (B) Relative abundance of bacterial phyla within the barley samples. Samples are grouped by storage duration (Harvest, three, six, and nine months).

### 2.4.4 The core microbial community of barley seed

This study delineates the core microbiome of barley seed samples, highlighting the top ten bacteria, which vary based on storage duration and grain nitrogen content (Fig. 2.4A and B; Table S2.5). Members of the genera *Erwinia*, *Pantoea*, *Pseudomonas*, and *Stenotrophomonas* comprised over 50% of the taxa (Fig. 2.4A and B). Additionally, other genera identified among the top ten include *Chryseobacterium*, *Klebsiella*, *Massilia*, *Citrobacter*, *Leclercia*, and *Xanthomonas* (Fig. 2.4A and B). The distribution of these genera varies with both storage duration and grain nitrogen content. For example, genera such as *Erwinia*, *Pantoea*, *Pseudomonas*, and *Stenotrophomonas* are consistently present across all storage times but show variations in their relative abundances (Fig. 2.4A). In contrast, other genera such as *Citrobacter*, *Klebsiella*, and *Leclercia* only increase in abundance at later stages of storage (Fig. 2.4A). This suggests the core microbial community may respond to changing conditions within the storage environment (Felšöciová *et al.*, 2020). Conversely, the distribution of the top ten core genera relative to grain nitrogen content exhibits a degree of stability (Fig. 2.4B), suggesting that variations in nitrogen content has minimal impact on the core bacterial communities.

An ideal plant core microbiome is made up of microbial communities consistently found in different plants grown across various locations and soil types (Shade and Handelsman, 2012). The microbial genera identified in this chapter are also found in the rhizosphere, roots, bulk soil, and across different barley genotypes (Yang *et al.*, 2017; Bziuk *et al.*, 2021; Bulgarelli *et al.*, 2015; Rahman *et al.*, 2018). This suggests they play an important role as stable components of the core barley microbiome. Members of the core genera, such as *Erwinia*, *Pantoea*, and *Pseudomonas* are known to exhibit a range of interactions with barley seed (Yang *et al.*, 2017; Bziuk *et al.*, 2021; Rahman *et al.*, 2018). For instance, *Erwinia gerundensis* acts as a biological control agent in barley malt (Gnonlonfoun *et al.*, 2023). Conversely, *Erwinia persicina* is associated with pink seed disease in barley (Kawaguchi *et al.*, 2021). Members of the genus *Pantoea* are ubiquitous and known for their adaptability and ecological versatility (De Maayer *et al.*, 2017; Sabra and Ahmed, 2022). For example, *Pantoea agglomerans* has been reported to inhibit basal kernel blight, typically caused by *Pseudomonas syringae pv. Syringae* (Braun-Kiewnick *et al.*, 2000). Additionally, members of the genus *Pseudomonas* are involved in several beneficial roles in plants, including nitrogen fixation and phosphorus solubilization (Yan *et al.*, 2008; Pham *et al.*, 2017). However, it is important to acknowledge that genus-level

identification may encompass species with both beneficial and detrimental effects, highlighting the importance of species-specific analyses to fully grasp their potential impacts (Chen *et al.*, 2020).

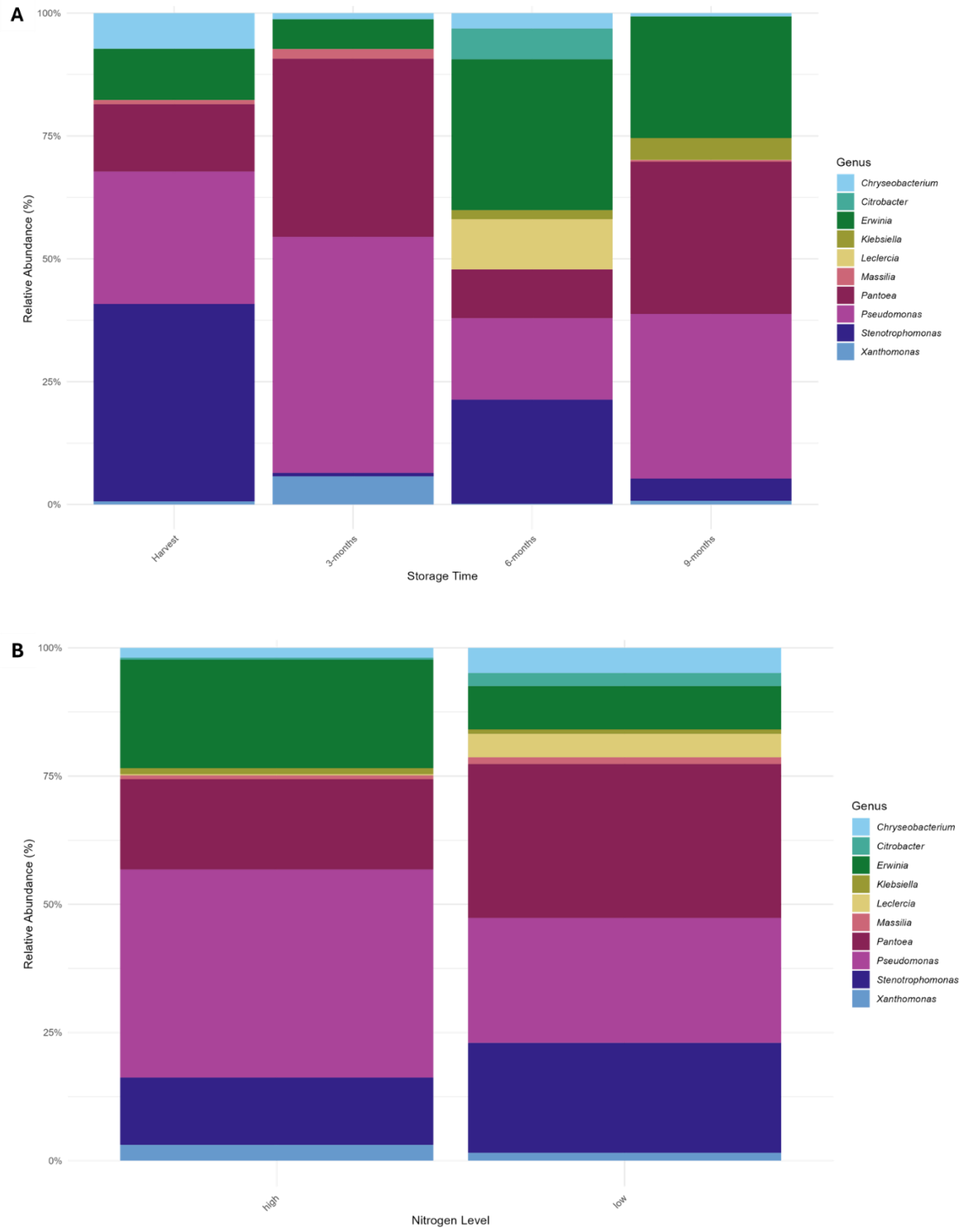


Figure 2.4 Relative abundance of core bacterial community within barley samples, categorised by, (A) Storage duration (harvest, three, six and nine months). (B) grain nitrogen content (high and low).

### 2.4.5 The non-core microbial community of barley seed

Non-core taxa contribute to the ecological flexibility and functional diversity of microbial communities, enabling them to adapt to a range of environmental conditions (Shade and Handelsman, 2012). The composition of non-core genera changes with storage duration and grain nitrogen content (Fig. 2.5 A; Table S2.5). Initially, at harvest and three months, genera such as *Pedobacter*, *Plantibacter*, and *Pseudarthrobacter* were most abundant, while *Enterococcus*, *Lactococcus*, and *Saccharibacillus* were more abundant in the later stages of storage (six and nine months) (Fig. 2.5A). A distinct pattern emerges when comparing microbial composition in environments with high versus low grain nitrogen content (Fig. 2.5 B). In samples with high nitrogen, *Plantibacter*, *Pseudarthrobacter*, and *Saccharibacillus* collectively represent over 50% of the microbial population (Fig. 2.5B; Table S2.5). In contrast, in low nitrogen samples, *Enterococcus* and *Lactococcus* are the predominant genera, also exceeding 50% combined (Fig. 2.5B). This pattern suggests that changing conditions selectively favour these distinct microbial communities.

The genera identified in the early stages of storage and high nitrogen grain content are known to colonise the endosphere, phyllosphere and rhizosphere (Liao *et al.*, 2024; Park *et al.*, 2020; Woo *et al.*, 2024). In contrast, the genera prevalent in later storage phases and low grain nitrogen content, such as *Enterococcus* and *Lactococcus*, are lactic acid bacteria commonly found in barley malt (Doi, 2018; Song *et al.*, 2017). Although they have benefits as inoculants for enhancing malt quality and safety, their uncontrolled proliferation during germination can lead to the synthesis of organic acids, which can compromise beer quality by modifying its flavour and stability (Laitila *et al.*, 2018; Mokoena, 2017).

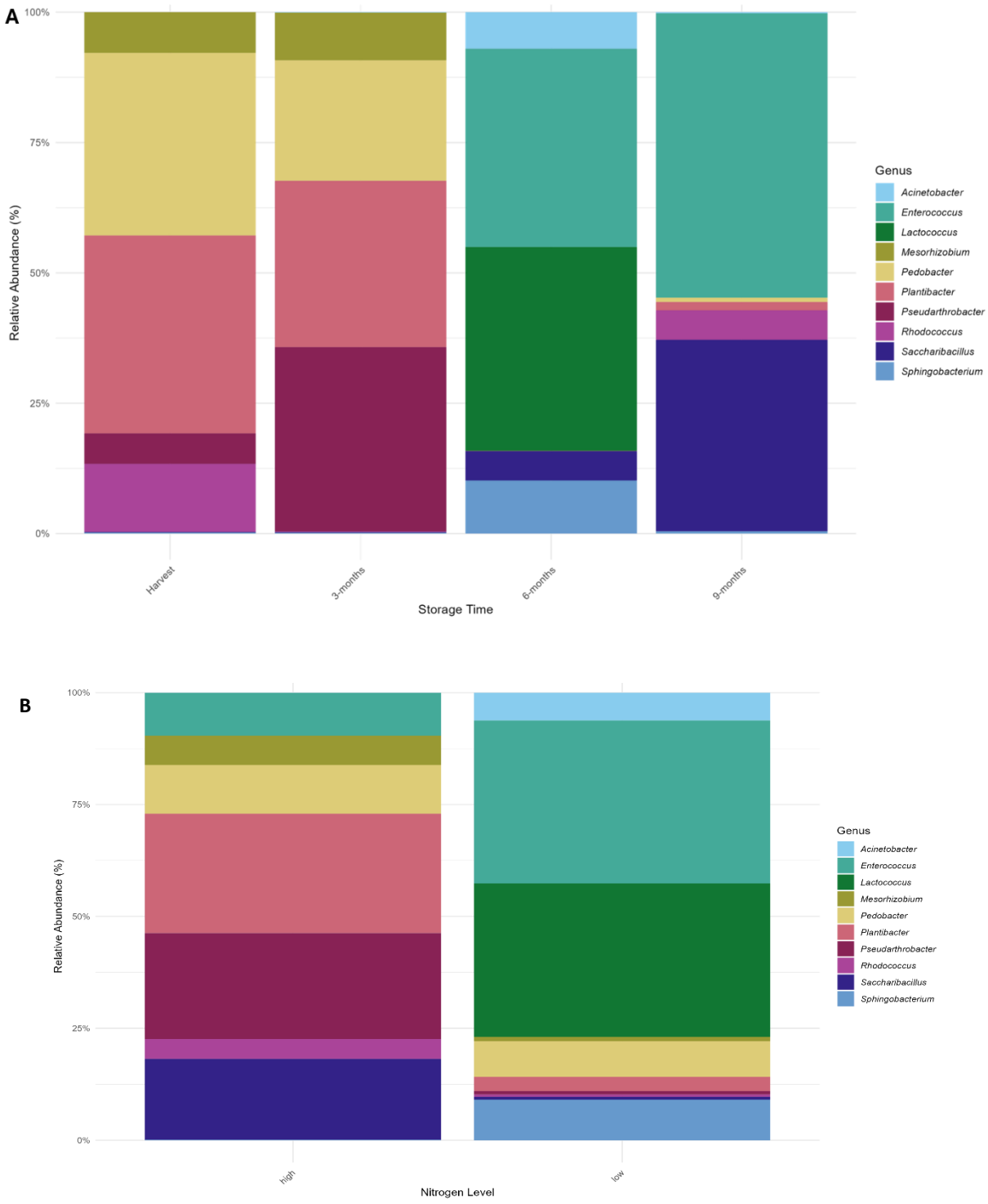


Figure 2.5 Relative abundance of the non-core bacterial community within barley samples, categorised by. (A) Storage duration (harvest, three, six and nine months). (B) Grain nitrogen content (high and low).

### 2.4.6 Microbial richness variations in barley seeds across storage duration

A total of 192 distinct bacterial genera were identified but not universally, showing a varied prevalence depending on storage duration. Also, a total of 140 genera were unique to each

storage time point. In total, 17, 80, 41 and two genera were unique to the harvest, three month, six month and nine-month storage duration, respectively (Fig. 2.6). The variation in genus richness at different storage time points underlines a complex interaction between the storage environment and microbial dynamics. Particularly, the increase in species richness up to the six-month storage suggests that the early and mid-storage periods are critical for microbial diversity.

The decline in diversity observed between the six- and nine-month intervals points towards a shift in the microbial ecosystem, potentially driven by competitive interactions among microbial species or changes in the nutritional landscape of the barley seeds as they undergo physiological changes in preparation for germination. This suggests that the storage environment, while controlled, does not remain static in its influence on microbial communities (Abubakar, 2019). Although the storage system is highly regulated, interactions within the microbial community, such as competition for nutrients or the production of antimicrobial compounds, likely play a role, alongside abiotic factors like moisture and temperature fluctuations (Chandel *et al.*, 2021). However, it is important to recognise that this controlled environment may not entirely reflect a natural ecological setting. These findings underscore the importance of considering microbial interactions in managing microbial dynamics in stored grain, although further research is needed to fully understand the ecological relevance of these dynamics in artificial storage environments.

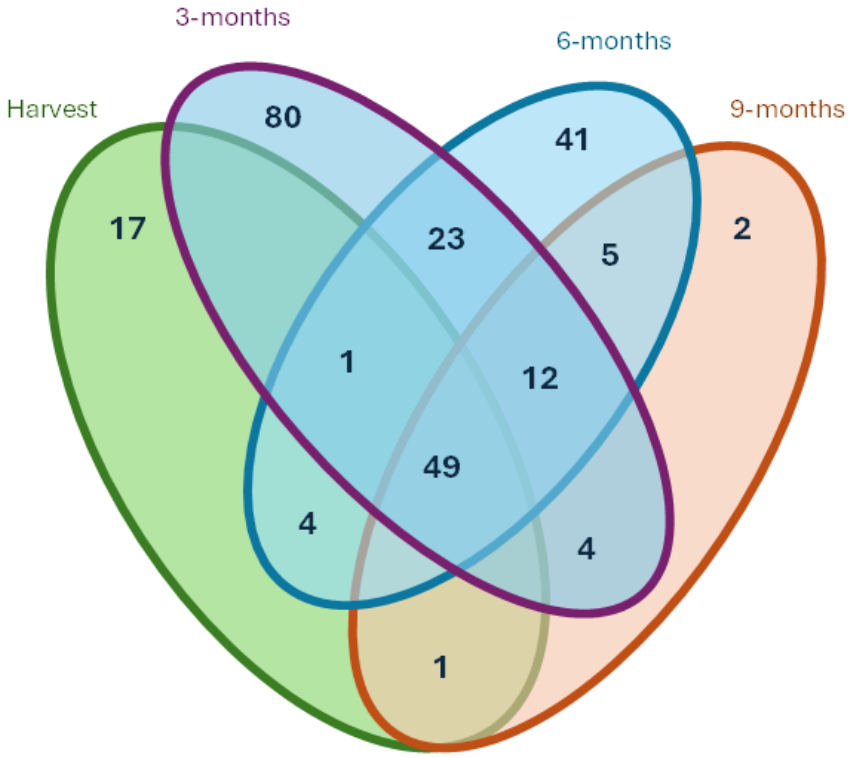


Figure 2.6 The number of shared and unique bacterial genera associated with barley seeds over storage duration.

### 2.4.7 Differential distribution of phage genera linked to nitrogen content and storage duration in barley seed

A total of 211,526 viral reads were identified, representative of 59 distinct bacteriophage genera. The abundance of the top ten phage genera was calculated by combining samples first by storage duration (Fig. 2.7A) then by grain nitrogen content (Fig. 2.7B). As storage duration increases, a greater diversity of genera becomes evident (Fig. 2.7A). At harvest, the phage population was predominantly comprised of members of the genus *Mimasvirus*, accounting for 91%, with *Petsuvirus* making up the remaining 9%. *Mimasvirus* is noteworthy for including phage GAP32, which targets the opportunistic pathogen *Cronobacter sakazakii* (Abbasifar *et al.*, 2014). The genus *Cronobacter* was observed in our samples (Table S2.6). Three months into storage, there was a shift towards a dominance of *Feofaniavirus*, which constituted 99% of the population. By six months, diversity began to increase, with *Skunavirus* at 60% and *Cornellvirus* at 20%, alongside the emergence of four additional genera (Fig. 2.7B). Phages belonging to the genus *Skunavirus* have a known association with species of *Lactococcus*, especially *Lactococcus lactis* (Lemay *et al.*, 2020; Doré *et al.*, 2021). The highest number of

sequences reads identified as *Lactococcus* (772,822) was found in the sample (S6L) coinciding with the detection of *Skunavirus* (Fig. 2.8A; Table S2.6). This phage is strictly lytic and has been shown to infect and kill *L. lactis* strains (Lemay *et al.*, 2020; Doré *et al.*, 2021). The genus *Cornellvirus* encompasses the phage *Cornellvirus* SLAM_phiST1N3, known for its capacity to infect multidrug-resistant strains of *Salmonella enterica serovar Typhimurium* (Choi *et al.*, 2024). Consistent with the presence of this phage, the highest number of *Salmonella* sequence reads, totalling 29,787, was recorded in sample (S6L) where *Cornellvirus* was detected (Table S2.6). At nine months of storage, five genera were observed with *Mosigvirus* (32%), *Kanagawavirus* (29%), and *Feofaniavirus* (28%) being the most abundant (Fig 2.7A). The evolving diversity over time could reflect changes in the microbiome associated with the seeds, host availability, or environmental conditions during storage facilitating the proliferation of different viral genera.

In samples with high nitrogen content, the distribution among phage genera was notably uniform, with *Sauletekiovirus*, *Feofaniavirus*, and *Ravinvirus* each contributing approximately 17-18% to the viral community (Fig. 2.7B). *Sauletekiovirus*, identified as a genus of lytic phages, are associated with *P. agglomerans* (Žukauskienė *et al.*, 2021), aligning with the finding that *Pantoea*, one of the top ten core genera in this study (Table S2.6;Fig. 2.8B). *Feofaniaviruses*, also lytic in nature, targets the genus *Erwinia* (Kim *et al.*, 2021), present across all samples. *Ravinvirus*, known for infecting *Escherichia coli* (Ravin, 2011), corresponds with the detection of the genus *Escherichia* in our samples (Table S2.6). In contrast, the phages associated with barley seed samples with low nitrogen contents were distinct, with the genera *Skunavirus* and *Cornellvirus* (64 and 21% of viral reads, respectively) dominating (Fig. 2.7B).

The presence of lytic phages in this study supports phage-bacterial host interaction. The presence of lytic phages can lead to the lysis of specific bacterial populations, thereby potentially altering community compositions and nutrient cycling processes (Hobbs and Abedon, 2016; Bertani, 1953). Bacterial genera such as *Cronobacter*, *Escherichia*, and *Salmonella*, showed a relatively low number of reads (Table S2.6). It is possible that their associations with lytic phages in the study could be responsible for their low read number. Further investigation into the interactions between these lytic phages and their bacterial hosts will be crucial in understanding the extent of their impact on microbial communities.

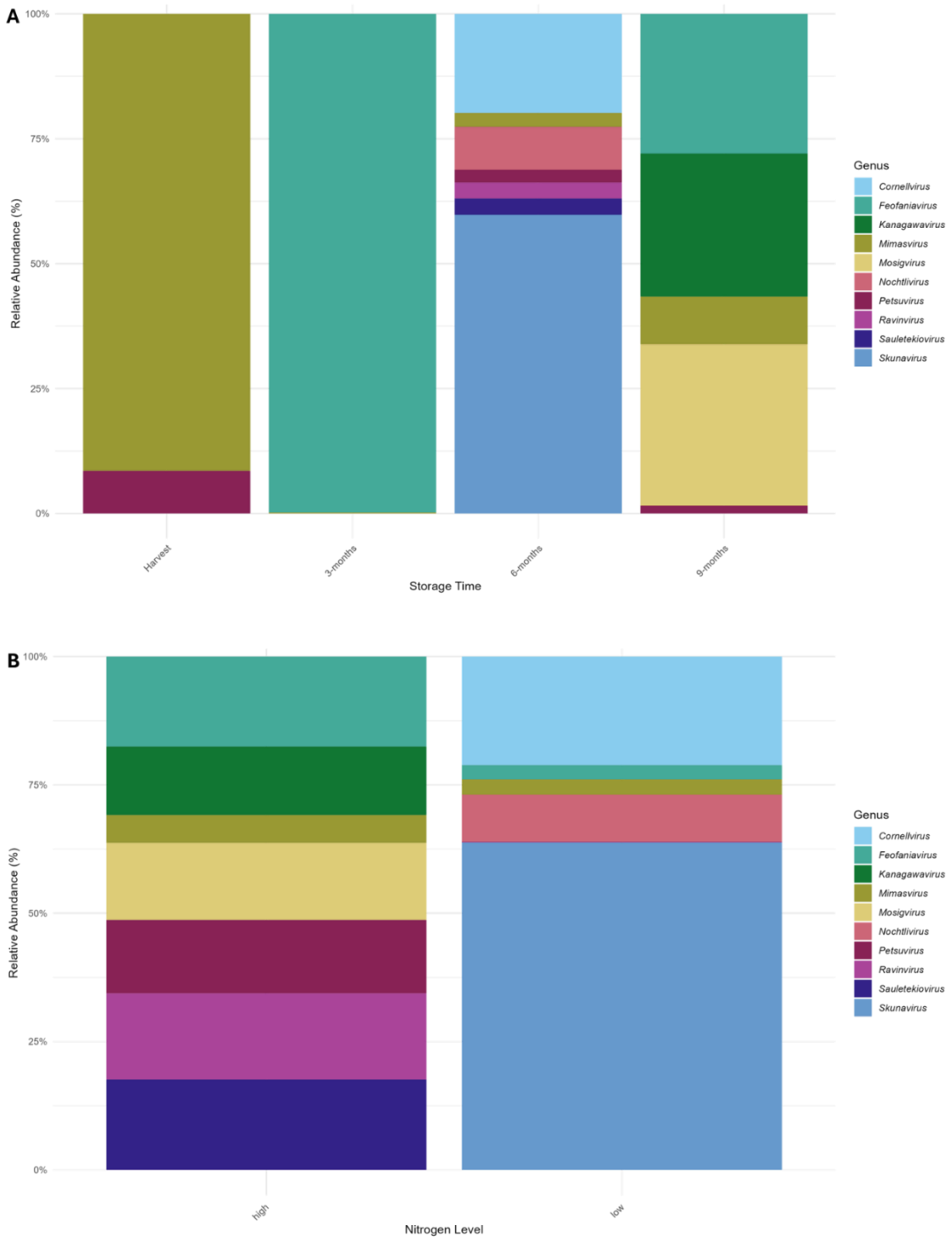


Figure 2.7 Relative abundance of phages within the barley metagenome. (A) Storage duration samples (B) Nitrogen content samples.

### 2.4.8 The barley seed microbiome exhibits no fungal presence at harvest or under conditions of high nitrogen content

A total of 46,979 fungal reads were identified, with an absence of such reads in samples from the initial harvest and those from barley seeds with high nitrogen content (Table S2.7). Among the low nitrogen content barley seed samples, a total of 19,052 fungal reads were observed at three months, 7,030 reads at six months and 20,897 reads at nine months of storage (Table S2.5). At three months of storage, the genera *Fusarium* (41%), *Candida* (21%), *Aspergillus* (17%) and *Thermothelomyces* (13%) were observed (Fig. 2.9). There was a shift at six months of storage, with *Eremothecium* accounting for all the detected fungal reads. At nine months in storage, the fungal distribution shifted again, including *Cryptococcus* (36%), *Fusarium* (20%), *Botrytis* (16%) and *Thermothelomyces* (11%) (Fig. 2.8).

Fungi exploit the storage phase to establish and proliferate (Laitila, 2015). This pattern suggests that the environmental conditions and microbial interactions within these specific contexts may inherently deter fungal colonisation and growth (Wang and Kuzyakov, 2024). High nitrogen content appears to create conditions less conducive to fungal presence, potentially through the competitive exclusion of nitrogen-using bacteria (Masuda *et al.*, 2024). This notion is supported by the exclusive detection of fungal species in low-nitrogen samples, where the reduced bacterial competition may allow fungi to utilise the available nutrients more freely.

The observed fungal communities were dominated by the genus *Fusarium*, which is recognised for its role in spoilage and mycotoxin production, with profound implications for agriculture. *Fusarium* is also associated with *Fusarium* Head Blight (FHB), a disease that significantly affects wheat and barley crops (Alisaac and Mahlein, 2023). The exclusive colonisation by *Eremothecium* at the six-month storage time point, may be indicative of an environment that possibly favours its rapid growth, potentially due to unique substrate characteristics or a state of dysbiosis. This shift correlates with the observed dysbiosis in bacterial communities at the same storage duration (Fig 2.6B), suggesting a broader ecological disruption. The presence of *Eremothecium* in storage environments is important, as this genus is known for its pathogenic potential, including causing yeast-spot disease in soybean seeds, often in association with the stink bug *Riptortus clavatus* (Kimura *et al.*, 2009; Kimura *et al.*, 2008; Zečević *et al.*, 2023).

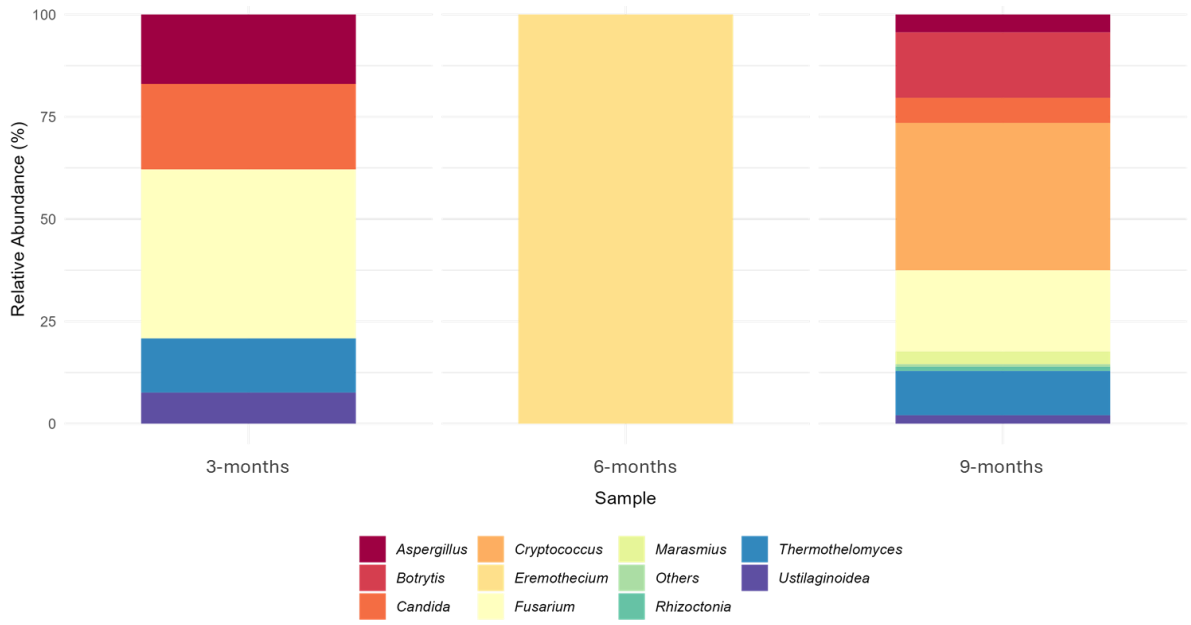


Figure 2.8 Relative abundance of fungal genera within the barley metagenome over a nine month storage period.

## 2.5 Conclusions

Almost 400 million metagenomic reads were generated from eight barley seed samples with high and low nitrogen content, stored for four different periods of time. More than 70% of the reads per sample were classified, leaving a substantial fraction unclassified due to the existing limitations within various metagenomic databases. The Shannon diversity metrics have revealed that seed nitrogen content does not significantly affect overall diversity. This observation was further supported by Bray-Curtis dissimilarity analyses, which also highlighted a shift in community structure over time, particularly noting a dysbiosis at the six-month storage duration. The microbial composition was predominantly bacterial, with both core and non-core bacteria shedding light on how different genera respond to nitrogen content and storage duration. Phages were closely associated with bacterial genera, with a noted decline in bacterial diversity corresponding with an increase in phage populations. Of note, fungal reads were absent from harvest barley seed samples and those characterised by high nitrogen content, suggesting conditions that may inhibit fungal growth. These findings contribute to the existing body of knowledge of the barley seed microbiome and could provide insight for optimal storage practices and microbial dynamics.

While these findings highlight important microbial shifts during storage, it is important to acknowledge that the storage environment represents an artificial system, and thus, the microbial dynamics observed here may not fully reflect those found in natural ecosystems. Nonetheless, competitive interactions, resource availability, and abiotic factors such as moisture and temperature are still relevant within this controlled system. Further research could help clarify how these factors influence microbial communities under storage conditions and how these dynamics may differ from those in natural settings. Future chapters will delve into genome reconstruction from these metagenomic data and further explore the functional profiles at the community level.

## 2.6 References


Abbasifar, R., Griffiths, M.W., Sabour, P.M., Ackermann, H.-W., Vandersteegen, K., Lavigne, R., Noben, J.-P., Alanis Villa, A., Abbasifar, A., Nash, J.H.E., Kropinski, A.M. (2014). Supersize me: *Cronobacter sakazakii* phage GAP32. *Virology*, **460–461**: 138–146.

Abubakar, S. (2019). Effect of Storage Conditions on wheat Quality. *International Journal of Food and Allied Sciences*, **4**: 20–23.

Adam, E., Bernhart, M., Müller, H., Winkler, J., Berg, G. (2018). The *Cucurbita pepo* seed microbiome: genotype-specific composition and implications for breeding. *Plant and Soil*, **422**: 35–49.

Adeboye, M.K.A., Bala, A., Osunde, A.O., Uzoma, A.O., Odofin, A.J., Lawal, B.A. (2011). Assessment of soil quality using soil organic carbon and total nitrogen and microbial properties in tropical agroecosystems. *Agricultural Sciences*, **2**: 34–40.

Afkhamifar, A., Moslemkhani, C., Hasanzadeh, N., Razmi, J. (2023). Interactions of seed-borne bacterial pathogens *Xanthomonas translucens* and *Pseudomonas syringae* pv *syringae* on wheat. *Journal of Plant Pathology*, **105**: 859–867.

Agu, R.C., Palmer, G.H. (2001). The Effect of Nitrogen Level on the Performance of Malting Barley Varieties during Germination. *Journal of the Institute of Brewing*, **107**: 93–98.

Ahmed, A.K.A., Shi, X., Hua, L., Manzueta, L., Qing, W., Marhaba, T., Zhang, W. (2018). Influences of Air, Oxygen, Nitrogen, and Carbon Dioxide Nanobubbles on Seed Germination and Plant Growth. *Journal of Agricultural and Food Chemistry*, **66**: 5117–5124.

Alisaac, E., Mahlein, A.-K. (2023). *Fusarium* Head Blight on Wheat: Biology, Modern Detection and Diagnosis and Integrated Disease Management. *Toxins*, **15**: 192.

Andrews, S. (2010). Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Visited: 5th September 2019.

Angelino, S.A.G.F., Convention, O.B.O.T.A.C.O.T.E.B. (1996). Determination of the Moisture and Nitrogen Contents of Barley and Malt by Near Infrared Spectroscopy (nirs). *Journal of the Institute of Brewing*, **102**: 73–74.

Bamforth, C.W. (2017). Progress in Brewing Science and Beer Production. *Annual Review of Chemical and Biomolecular Engineering*, **8**: 161–176.

Baslam, M., Mitsui, T., Sueyoshi, K., Ohyama, T. (2021). Recent Advances in Carbon and Nitrogen Metabolism in C3 Plants. *International Journal of Molecular Sciences*, **22**: 318.

Batista, J.N.G., Ferreira, M.A. da S.V., Quezado-Duval, A.M. (2021). Molecular and phenotypic characterization of *Xanthomonas campestris pv. campestris* causing black rot in *Brassica* crops in Brazil. *Tropical Plant Pathology*, **46**: 684–701.

Bertani, G. (1953). Lysogenic Versus Lytic Cycle of Phage Multiplication. *Cold Spring Harbor Symposia on Quantitative Biology*, **18**: 65–70.

Bingham, I.J., Hoad, S.P., Thomas, W.T.B., Newton, A.C. (2012). Yield response to fungicide of spring barley genotypes differing in disease susceptibility and canopy structure. *Field Crops Research*, **139**: 9–19.

Bolger, A.M., Lohse, M., Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**: 2114–2120.

Borodovsky, M., Lomsadze, A. (2011). Eukaryotic gene prediction using GeneMark.hmm-E and GeneMark-ES. *Current Protocols in Bioinformatics*, **Chapter 4**: Unit 4.6.1-10.

Braun-Kiewnick, A., Jacobsen, B.J., Sands, D.C. (2000). Biological Control of *Pseudomonas syringae pv. syringae*, the Causal Agent of Basal Kernel Blight of Barley, by Antagonistic Pantoea agglomerans. *Phytopathology*, **90**: 368–375.

Bray, J.R., Curtis, J.T. (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, **27**: 326–349.

Briggs, D.E., Brookes, P.A., Stevens, R., Boulton, C.A. (2004). *Brewing: Science and Practice*. Elsevier, 900 pp.

Briggs, D.E., Hough, J.S., Stevens, R., Young, T.W. (1981). *Malting and Brewing Science: Malt and Sweet Wort*. Springer Science & Business Media, 404 pp.

Bulgarelli, D., Garrido-Oter, R., Münch, P.C., Weiman, A., Dröge, J., Pan, Y., McHardy, A.C., Schulze-Lefert, P. (2015). Structure and Function of the Bacterial Root Microbiota in Wild and Domesticated Barley. *Cell Host & Microbe*, **17**: 392–403.

Bziuk, N., Maccario, L., Straube, B., Wehner, G., Sørensen, S.J., Schikora, A., Smalla, K. (2021). The treasure inside barley seeds: microbial diversity and plant beneficial bacteria. *Environmental Microbiome*, **16**: 20.

Chandel, A., Mann, R., Kaur, J., Norton, S., Edwards, J., Spangenberg, G., Sawbridge, T. (2021). Implications of Seed Vault Storage Strategies for Conservation of Seed Bacterial Microbiomes. *Frontiers in Microbiology*, **12**: 784796.

Chen, L.-X., Anantharaman, K., Shaiber, A., Eren, A.M., Banfield, J.F. (2020). Accurate and complete genomes from metagenomes. *Genome Research*, **30**: 315–333.

Choi, Y., Kwak, M.-J., Kang, M.-G., Kang, A.N., Lee, W., Mun, D., Choi, H., Park, J., Eor, J.Y., Song, M., Kim, J.N., Oh, S., Kim, Y. (2024). Molecular characterization and environmental impact of newly isolated lytic phage SLAM_phiST1N3 in the *Cornellvirus* genus for biocontrol of a multidrug-resistant *Salmonella* Typhimurium in the swine industry chain. *Science of The Total Environment*, **922**: 171208.

Clarke, K.R. (1993). Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology*, **18**: 117–143.

De Maayer, P., Aliyu, H., Vikram, S., Blom, J., Duffy, B., Cowan, D.A., Smits, T.H.M., Venter, S.N., Coutinho, T.A. (2017). Phylogenomic, Pan-genomic, Pathogenomic and Evolutionary Genomic Insights into the Agronomically Relevant Enterobacteria Pantoea ananatis and Pantoea stewartii. *Frontiers in Microbiology*, **8**.

Dixon, P. (2003). VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, **14**: 927–930.

Doi, Y. (2018). Lactic acid fermentation is the main aerobic metabolic pathway in *Enterococcus faecalis* metabolizing a high concentration of glycerol. *Applied Microbiology and Biotechnology*, **102**: 10183–10192.

Doré, L., Pageau, G., Bourque-Leblanc, F., Dupuis, M.-È., Lessard-Hurtubise, R., Lacasse, G., Plante, P.-L., Labrie, S., P. Jolicoeur, A., Rousseau, G.M., Tremblay, D.M., Moineau, S. (2021). Complete Genome Sequences of 10 *Lactococcal Skunavirus* Phages Isolated from Cheddar Cheese Whey Samples in Canada. *Microbiology Resource Announcements*, **10**: 10.1128/mra.00098-21.

Ewels, P., Magnusson, M., Lundin, S., Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, **32**: 3047–3048.

Felšöciová, S., Kowalczewski, P.Ł., Krajčovič, T., Dráb, Š., Kačániová, M. (2021). Effect of Long-Term Storage on Mycobiota of Barley Grain and Malt. *Plants (Basel, Switzerland)*, **10**: 1655.

Felšöciová, S., Kowalczewski, P.Ł., Krajčovič, T., Dráb, Š., Kačániová, M. (2020). Quantitative and Qualitative Composition of Bacterial Communities of Malting Barley Grain and Malt during Long-Term Storage. *Agronomy*, **10**: 1301.

Flannigan, B. (2003). The microbiota of barley and malt. In: Priest, F.G. & Campbell, I. (eds) *Brewing Microbiology*. Springer US, Boston, MA, pp. 113–180.

Gnonlonfoun, E., Fotin, G., Schwebel, S., Colin, C., Salles, M., Parant, V., Alonso, P., Moreau, F., Borges, F., Revol-Junelles, A.-M., Schmitt, M., Framboisier, X., Fick, M., Rondags, E. (2023). Impact of *Erwinia gerundensis* as a Biocontrol Agent on the Sanitary and Technological Quality of Barley Malt. *Journal of the American Society of Brewing Chemists*, **0**: 1–14.

Hobbs, Z., Abedon, S.T. (2016). Diversity of phage infection types and associated terminology: the problem with 'Lytic or lysogenic'. *FEMS Microbiology Letters*, **363**: fnw047.

Jo, J., Oh, J., Park, C. (2020). Microbial community analysis using high-throughput sequencing technology: a beginner's guide for microbiologists. *Journal of Microbiology*, **58**: 176–192.

Kawaguchi, A., Abe, D., Saito, T., Nogata, Y., Nomiyama, K., Kohyama, N., Takahashi, A., Yoshioka, T., Ishikawa, N., Tomioka, K. (2021). Pink seed of barley caused by *Erwinia persicina*. *Journal of General Plant Pathology*, **87**: 106–109.

Kim, S.G., Roh, E., Park, J., Giri, S.S., Kwon, J., Kim, S.W., Kang, J.W., Lee, S.B., Jung, W.J., Lee, Y.M., Cho, K., Park, S.C. (2021). The Bacteriophage pEp_SNUABM_08 Is a Novel Singleton Siphovirus with High Host Specificity for *Erwinia pyrifoliae*. *Viruses*, **13**: 1231.

Kimura, S., Tokumaru, S., Kuge, K. (2008). *Eremothecium ashbyi* causes soybean yeast-spot and is associated with stink bug, *Riptortus clavatus*. *Journal of General Plant Pathology*, **74**: 275–280.

Kimura, S., Tokumaru, S., Kuge, K. (2009). *Eremothecium coryl*i and *E.ashbyi* cause yeast spot of azuki bean. *Journal of General Plant Pathology*, **75**: 322–324.

Knight, R., Vrbanac, A., Taylor, B.C., Aksenov, A., Callewaert, C., Debelius, J., Gonzalez, A., Kosciolek, T., McCall, L.-I., McDonald, D., Melnik, A.V., Morton, J.T., Navas, J., Quinn, R.A., Sanders, J.G., Swafford, A.D., Thompson, L.R., Tripathi, A., Xu, Z.Z., Zaneveld, J.R., Zhu, Q., Caporaso, J.G., Dorrestein, P.C. (2018). Best practices for analysing microbiomes. *Nature Reviews Microbiology*, **16**: 410–422.

Kruskal, W.H., Wallis, W.A. (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, **47**: 583–621.

Laitila, A. (2015). 6 - Toxigenic fungi and mycotoxins in the barley-to-beer chain. In: Hill, A.E. (ed) *Brewing Microbiology*. Woodhead Publishing Series in Food Science, Technology and Nutrition. Woodhead Publishing, Oxford, pp. 107–139.

Laitila, A., Kotaviita, E., Peltola, P., Home, S., Wilhelmson, A. (2007). Indigenous Microbial Community of Barley Greatly Influences Grain Germination and Malt Quality. *Journal of the Institute of Brewing*, **113**: 9–20.

Laitila, A., Manninen, J., Priha, O., Smart, K., Tsitko, I., James, S. (2018). Characterisation of barley-associated bacteria and their impact on wort separation performance. *Journal of the Institute of Brewing*, **124**: 314–324.

Langmead, B., Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, **9**: 357–359.

Langridge, P. (2018). Economic and Academic Importance of Barley. In: Stein, N. & Muehlbauer, G.J. (eds) *The Barley Genome*. Compendium of Plant Genomes. Springer International Publishing, Cham, pp. 1–10.

Lemay, M.-L., Maaß, S., Otto, A., Hamel, J., Plante, P.-L., Rousseau, G.M., Tremblay, D.M., Shi, R., Corbeil, J., Gagné, S.M., Becher, D., Moineau, S. (2020). A Lactococcal Phage Protein Promotes Viral Propagation and Alters the Host Proteomic Response During Infection. *Viruses*, **12**: 797.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**: 2078–2079.

Liao, L., Qin, Q., Yi, D., Lai, Q., Cong, B., Zhang, H., Shao, Z., Zhang, J., Chen, B. (2024). Evolution and adaptation of terrestrial plant-associated *Plantibacter* species into remote marine environments. *Molecular Ecology*, **33**: e17385.

Lindow, S.E., Arny, D.C., Upper, C.D. (1982). Bacterial Ice Nucleation: A Factor in Frost Injury to Plants 1. *Plant Physiology*, **70**: 1084–1089.

Lu, J., Breitwieser, F.P., Thielen, P., Salzberg, S.L. (2017). Bracken: estimating species abundance in metagenomics data. *PeerJ Computer Science*, **3**: e104.

Lu, J., Rincon, N., Wood, D.E., Breitwieser, F.P., Pockrandt, C., Langmead, B., Salzberg, S.L., Steinegger, M. (2022). Metagenome analysis using the Kraken software suite. *Nature Protocols*, **17**: 2815–2839.

Masuda, Y., Mise, K., Xu, Z., Zhang, Z., Shiratori, Y., Senoo, K., Itoh, H. (2024). Global soil metagenomics reveals distribution and predominance of Deltaproteobacteria in nitrogen-fixing microbiome. *Microbiome*, **12**: 95.

Matsumoto, H., Fan, X., Wang, Yue, Kusstatscher, P., Duan, J., Wu, S., Chen, S., Qiao, K., Wang, Yiling, Ma, B., Zhu, G., Hashidoko, Y., Berg, G., Cernava, T., Wang, M. (2021). Bacterial seed endophyte shapes disease resistance in rice. *Nature Plants*, **7**: 60–72.

Maziarz, M., Pfeiffer, R.M., Wan, Y., Gail, M.H. (2018). Using standard microbiome reference groups to simplify beta-diversity analyses and facilitate independent validation. *Bioinformatics*, **34**: 3249–3257.

Mokoena, M.P. (2017). Lactic Acid Bacteria and Their Bacteriocins: Classification, Biosynthesis and Applications against Uropathogens: A Mini-Review. *Molecules (Basel, Switzerland)*, **22**: 1255.

Nejadmansouri, M., Shad, E., Razmjooei, M., Safdarianghomsheh, R., Delvigne, F., Khalesi, M. (2020). Production of xanthan gum using immobilized *Xanthomonas campestris* cells: Effects of support type. *Biochemical Engineering Journal*, **157**: 107554.

Park, M.-K., Park, Y.-J., Kim, M., Kim, M.-C., Ibal, J.C., Kang, G.-U., Lee, G.-D., Tagele, S.B., Kwon, H.-J., Kang, M.-S., Kim, M.-H., Kim, S.-Y., Shin, J.-H. (2020). Complete genome sequence of a plant growth-promoting bacterium *Pseudarthrobacter* sp. NIBRBAC000502772, isolated from shooting range soil in the Republic of Korea. *The Microbiological Society of Korea*, **56**: 390–393.

Pham, V.T.K., Rediers, H., Ghequire, M.G.K., Nguyen, H.H., De Mot, R., Vanderleyden, J., Spaepen, S. (2017). The plant growth-promoting effect of the nitrogen-fixing endophyte *Pseudomonas stutzeri* A15. *Archives of Microbiology*, **199**: 513–517.

Pitt, J.I., Hocking, A.D. (1997). *Fungi and Food Spoilage*. Blackie Academic & Professional, 618 pp.

Rahman, M.M., Flory, E., Koyro, H.-W., Abideen, Z., Schikora, A., Suarez, C., Schnell, S., Cardinale, M. (2018). Consistent associations with beneficial bacteria in the seed endosphere of barley (*Hordeum vulgare* L.). *Systematic and Applied Microbiology*, **41**: 386–398.

Raj, G., Shadab, M., Deka, S., Das, M., Baruah, J., Bharali, R., Talukdar, N.C. (2019). Seed interior microbiome of rice genotypes indigenous to three agroecosystems of Indo-Burma biodiversity hotspot. *BMC Genomics*, **20**: 924.

Ranjan, D.R.K., Singh, D. (2020). Effect of temperature and inoculums level on development of soft rot of potato caused by *Erwinia carotovora subsp. carotovora* and their molecular detection through poymearse chain reaction. *Journal of Pharmacognosy and Phytochemistry*, **9**: 1414–1419.

Ravin, N.V. (2011). N15: The linear phage–plasmid. *Plasmid*, **65**: 102–109.

Richard, G.-F. (2020). Eukaryotic Pangenomes. In: Tettelin, H. & Medini, D. (eds) *The Pangenome: Diversity, Dynamics and Evolution of Genomes*. Springer International Publishing, Cham, pp. 253–291.

Sabra, M.A., Ahmed, A.E. (2022). The Influence of The Isolated *Enterobacter* spp and *Pantoea* sp., on Barley's Phosphorus Uptake Grown in Calcareous Soil. *Egyptian Academic Journal of Biological Sciences, G. Microbiology*, **14**: 47–63.

Shade, A., Handelsman, J. (2012). Beyond the Venn diagram: the hunt for a core microbiome. *Environmental Microbiology*, **14**: 4–12.

Shannon, C.E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, **27**: 379–423.

Song, A.A.-L., In, L.L.A., Lim, S.H.E., Rahim, R.A. (2017). A review on *Lactococcus lactis*: from food to factory. *Microbial Cell Factories*, **16**: 55.

Suzuki, K. (2020). Emergence of New Spoilage Microorganisms in the Brewing Industry and Development of Microbiological Quality Control Methods to Cope with This Phenomenon: A Review. *Journal of the American Society of Brewing Chemists*, **78**: 245–259.

Truong, D.T., Franzosa, E.A., Tickle, T.L., Scholz, M., Weingart, G., Pasolli, E., Tett, A., Huttenhower, C., Segata, N. (2015). MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nature Methods*, **12**: 902–903.

Truyens, S., Weyens, N., Cuypers, A., Vangronsveld, J. (2015). Bacterial seed endophytes: genera, vertical transmission and interaction with plants. *Environmental Microbiology Reports*, **7**: 40–50.

Tyc, O., Putra, R., Gols, R., Harvey, J.A., Garbeva, P. (2020). The ecological role of bacterial seed endophytes associated with wild cabbage in the United Kingdom. *MicrobiologyOpen*, **9**: e00954.

Wang, C., Kuzyakov, Y. (2024). Mechanisms and implications of bacterial-fungal competition for soil resources. *The ISME journal*, **18**: wrae073.

Wasendorf, C., Schmitz-Esser, S., Eischeid, C.J., Leyhe, M.J., Nelson, E.N., Rahic-Seggerman, F.M., Sullivan, K.E., Peters, N.T. (2022). Genome analysis of *Erwinia persicina* reveals implications for soft rot pathogenicity in plants. *Frontiers in Microbiology*, **13**.

Woo, H., Chhetri, G., Kim, I., So, Y., Park, S., Jung, Y., Seo, T. (2024). *Pedobacter rhodius* sp. nov. and *Pedobacter punctiformis* sp. nov., isolated from soil. *Antonie van Leeuwenhoek*, **117**: 72.

Wood, D.E., Lu, J., Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology*, **20**: 257.

Xia, Y., Sun, J. (2023). Alpha Diversity. In: Xia, Y. & Sun, J. (eds) *Bioinformatic and Statistical Analysis of Microbiome Data: From Raw Sequences to Advanced Modeling with QIIME 2 and R*. Springer International Publishing, Cham, pp. 289–333.

Xu, R., Rajeev, S., Salvador, L.C.M. (2023). The selection of software and database for metagenomics sequence analysis impacts the outcome of microbial profiling and pathogen detection. *PloS One*, **18**: e0284031.

Yan, Y., Yang, J., Dou, Y., Chen, M., Ping, S., Peng, J., Lu, W., Zhang, W., Yao, Z., Li, H., Liu, W., He, S., Geng, L., Zhang, X., Yang, F., Yu, H., Zhan, Y., Li, D., Lin, Z., Wang, Y., Elmerich, C., Lin, M., Jin, Q. (2008). Nitrogen fixation island and rhizosphere competence traits in the genome of root-associated *Pseudomonas stutzeri* A1501. *Proceedings of the National Academy of Sciences of the United States of America*, **105**: 7564–7569.

Yang, L., Danzberger, J., Schöler, A., Schröder, P., Schloter, M., Radl, V. (2017). Dominant Groups of Potentially Active Bacteria Shared by Barley Seeds become Less Abundant in Root Associated Microbiome. *Frontiers in Plant Science*, **8**: 1005.

Zečević, K., Sudimac, M., Majstorović, H., Stanković, I., Petrović, B., Delibašić, G., Krstić, B. (2023). First Report of Yeast-Spot Disease of Soybean Seeds Caused by *Eremothecium coryli* in Serbia. *Plant Disease*, **107**: 232.

Žukauskienė, E., Šimoliūnienė, M., Truncaitė, L., Skapas, M., Kaupinis, A., Valius, M., Meškys, R., Šimoliūnas, E. (2021). *Pantoea* Bacteriophage vB_PagS_AAS23: A Singleton of the Genus *Sauletekiovirus*. *Microorganisms*, **9**: 668.

# Chapter 3 Reconstruction of metagenome assembled genomes of microorganisms associated with stored barley seeds with varied nitrogen contents

## 3.1 Abstract

Seed microbiomes play important roles in the physiological processes required for effective crop development. The barley seed microbiome influences not only crop yield but also the quality of barley-derived products. The diversity, organisation and functional capacities of the barley seed microbial community are largely unexplored. In this study metagenome-assembled genomes (MAGs) were derived from metagenomes obtained from barley seeds of varied nitrogen grain content sampled across different storage times. The majority (91%) of the MAGs had a completeness score above 75%. The MAGs were classified into 26 bacterial genera belonging to four distinct phyla: Pseudomonadota, Bacteroidota, Actinomycetota, and Bacillota, revealing a diverse microbial community associated with stored barley seeds. The MAGs were structurally and functionally annotated using a range of *in silico* tools. OrthoFinder analysis revealed a limited number of single-copy orthologues and only 66 core orthogroups across the 82 MAGs analysed in this study, reflecting the substantial taxonomic diversity observed within the sample set. Additionally, Cluster of Orthologous Groups (COG) analysis via eggNOG revealed that these COGs play could be involved in critical cellular functions, particularly in translation, ribosomal structure, and biogenesis (14%), and replication, recombination, and repair (14%), highlighting their importance in adapting to environmental shifts in barley seed microbiomes. Plant pathogenicity-related proteins were also identified in MAGs from the genera *Duffyela*, *Erwinia*, and *Klebsiella*. This study enriches our understanding of the genomic content and potential functions of the predominant microbial species in barley seeds, offering insights into their roles within the barley seed microbiome.

## 3.2 Introduction

The barley seed microbiome comprises a diverse assembly of bacteria, fungi and other microorganisms located on and within the seeds (Noots *et al.*, 1999; Bianco *et al.*, 2018). While many of these microorganisms are beneficial, some contribute to pathogenicity (Noots *et al.*,

1999). In the context of beer production, unwanted microbial activity can extend seed dormancy, impair germination and decrease malt yield (Bokulich and Bamforth, 2013). These microbial influences may degrade wort quality, induce yeast flocculation and, ultimately, affect beer quality (Bamforth, 2017; Guo *et al.*, 2021).

In recent years, the reduced costs of next-generation sequencing (NGS) has catalysed a surge in microbiome sequencing projects across various cereal crops (Goodwin *et al.*, 2016; Grant, 2022). Research has primarily focused on the wheat microbiome, with maize and rice also receiving substantial attention (Michl *et al.*, 2023). By contrast, the barley microbiome has not been studied as extensively. Our current understanding is based on culture-dependent and culture-independent studies that mainly report on the taxonomy of microorganisms associated with the barley seed (Bziuk *et al.*, 2021; Felšöciová *et al.*, 2021; Noots *et al.*, 1999). These studies have successfully identified key microbial taxa within the core barley microbiome, including beneficial microorganisms and pathogens (Bziuk *et al.*, 2021; Felšöciová *et al.*, 2021; Noots *et al.*, 1999). However, the primary limitation of purely taxonomic studies lies in their limited ability to evaluate the functional capacities of these microorganisms (Muhamad Rizal *et al.*, 2020). A more effective strategy would involve the reconstruction of metagenome-assembled genomes (Setubal, 2021).

Metagenome-assembled genomes (MAGs) represent an advance in the reconstruction of microbial genomes directly from metagenome data (Setubal, 2021). In recent years, the scientific community has reported several thousand MAGs across diverse environments and in association with a range of different hosts, markedly enhancing our understanding of microbial populations and their interactions with their respective habitats (He *et al.*, 2021; Kayani *et al.*, 2021; Chen *et al.*, 2020). Notably, a considerable proportion of these MAGs are attributed to novel species, thereby contributing to the reduction of so-called microbial “dark matter” (Jiao *et al.*, 2021; Lok, 2015). The construction of metagenome-assembled genomes (MAGs) begins with metagenomic sequencing of mixed-species samples to generate a vast array of sequence data (Wu *et al.*, 2022). Subsequently, this data is assembled into longer fragments known as contigs, a challenge compounded by the mixture of sequences from various organisms (Nurk *et al.*, 2017). Following assembly, the contigs are sorted into bins, with each bin approximating the genome of an individual microorganism (Alneberg *et al.*, 2014). After binning, the MAGs undergo a rigorous quality assessment, typically involving the identification of single-copy marker genes to evaluate their completeness and accuracy (Parks *et al.*, 2015). Finally, the

MAGs are taxonomically classified and their genes functionally annotated using gene and protein databases, to uncover their potential roles (Chaumeil *et al.*, 2020; Zhou *et al.*, 2022; Stewart *et al.*, 2019).

Our understanding of plant microbiomes has substantially improved through reconstruction of MAGs (Fadiji and Babalola, 2020). For instance, MAGs have shed light on the microbial networks that underpin plant health, revealing how microbiomes build networks to preserve functionality across many plant-associated variables (Singh *et al.*, 2020). Additionally, MAGs have shed light on techniques for improving plant health and stress tolerance, providing insights into the biochemical, physiological and molecular aspects of plant-microbe interactions under stress (Gagliano *et al.*, 2022; Dindhoria *et al.*, 2024). While several studies have documented the taxonomic composition of microbial communities associated with barley seeds (Felšöciová *et al.*, 2020; Felšöciová *et al.*, 2021; Bziuk *et al.*, 2021), comprehensive knowledge of the functionality of the barley seed microbiome remains limited. Building upon the metagenome sequence of barley samples in Chapter 2, here we constructed MAGs from these metagenomic datasets and functionally annotated the resultant genomes. As such, the genomes of prevalent bacterial taxa from the barley seed communities could be reconstructed. Furthermore, these MAGs give insight into the functional potential of the taxa associated with barley seeds.

## 3.3 Methods

### 3.3.1 Metagenomic assembly, binning, and refinement

The metagenomic read datasets processed in Chapter 2 (section 2.3.2) were used for MAG reconstruction. A *de novo* assembly of the reads was performed using metaSPAdes v3.15.3 (Nurk *et al.*, 2017) with default parameters. The integrity and quality of the final assemblies were evaluated using QUAST v5.2.0 (Gurevich *et al.*, 2013). Subsequently, three binning tools, employing distinct algorithms, namely MaxBin v2.0 (Wu *et al.*, 2016), MetaBAT2 (Kang *et al.*, 2019) and CONCOCT v1.0.0 (Alneberg *et al.*, 2014), were used to assign the resultant contigs to bins. MaxBin groups contigs together using an Expectation-Maximization algorithm and Markov Chain Monte Carlo methods, focusing on tetranucleotide frequency and differential coverage (Wu *et al.*, 2016). It also improves the accuracy of bins by searching for highly conserved marker genes within the assembled contigs, serving as reliable genetic

markers to differentiate microbial species or strains (Wu *et al.*, 2016). MetaBAT2 adopts an adaptive binning algorithm based on tetranucleotide frequency and differential coverage, with a unique approach for contig distance measurement to optimise accuracy and minimise misclassification (Kang *et al.*, 2019). CONCOCT employs a Gaussian mixture model to cluster contigs, considering nucleotide composition and coverage across samples, effectively partitioning data into genetically related groups suitable for short-read sequencing data analysis (Alneberg *et al.*, 2014). The completeness and contamination levels of the generated bins were evaluated using CheckM v1.2.2 (Parks *et al.*, 2015).

Following the preliminary binning, the resultant bins were processed using the *MetaWRAP-Bin_refinement* module of the MetaWRAP v1.3 pipeline (Uritskiy *et al.*, 2018) for further refinement. The parameters for this module were set to -c 70 and -x 10 (completeness > 70 % and contamination < 10%). The module enhances the bin quality by integrating the outputs from the initial binning algorithms, thereby improving genome completeness and reducing contamination (Uritskiy *et al.*, 2018). The refinement process involves a detailed assessment of the quality of each bin, reallocating contigs between bins and selecting bins with completeness and contamination criteria outlined above (Uritskiy *et al.*, 2018). Subsequently, the bins were reassembled using the *MetaWRAP-reassemble_bins* module aiming to rectify misassemblies and bolster genomic integrity.

After refinement and reassembly, the bins were dereplicated to reduce redundancy and improve the uniqueness of the MAGs. Using dRep v2.6.2 (Olm *et al.*, 2017), with a 95% average nucleotide identity (ANI) threshold, closely related bins were identified and consolidated, resulting in a collection of 82 non-redundant MAGs.

### 3.3.2 Phylogenetic analysis and classification of MAGs

For taxonomic assignment of the MAGs, the *classify_wf* workflow from GTDB-Tk v2.3.2 was executed using default settings, (Parks *et al.*, 2022) with the reference database GTDB release 207v2 (Parks *et al.*, 2022). A comprehensive phylogenetic tree encompassing 82 species-level bacterial MAGs was derived from 120 bacterial marker genes using the *gtdbtk_infer* module in GTDB-Tk (Parks *et al.*, 2022). To improve interpretation and visualisation, the tree was annotated using iTOL v5 (Letunic and Bork, 2021). While GTDB is among the most robust databases available for taxonomic classification, it is important to recognise that reliance on a single database may limit the detection of novel or less-represented taxa. Future work could

incorporate additional databases or complementary methods to address these potential gaps and enhance taxonomic discovery in underexplored microbial environments.

### 3.3.3 Structural and functional annotation

Structural annotation of MAGs was performed using Prokka v1.14.5 (Seemann, 2014), a pipeline that integrates multiple algorithms and tools to accurately predict genomic features. Prodigal was utilised for coding sequence identification (Hyatt *et al.*, 2010), while Aragorn facilitated the detection of tRNA and tmRNA genes (Laslett and Canback, 2004).

### 3.3.4 Analysis of orthologous gene clusters in the MAGs

The protein sequences from structurally annotated MAGs were analysed using OrthoFinder v2.5.5 (Emms and Kelly, 2015). OrthoFinder groups proteins into orthogroups, which are sets of proteins descended from a single ancestral gene shared by all included species (Emms and Kelly, 2015). OrthoFinder starts by performing an all-against-all protein sequence similarity search. By default, it uses DIAMOND (Buchfink *et al.*, 2015), a sequence aligner that is considerably quicker than BLAST for large datasets. Based on the results from DIAMOND, OrthoFinder employs a graph-based approach to infer orthogroups. Essentially, these are groups of orthologues and closely related paralogues. OrthoFinder constructs a graph where each node represents a protein, and an edge between any two genes indicates a similarity score from DIAMOND. It then uses various graph clustering algorithms to delineate orthogroups, grouping together genes that are more similar to each other than to genes in other groups (Emms and Kelly, 2015).

The results from OrthoFinder identified which orthogroups were core (shared by all taxa), unique to a single taxon, or shared by at least two species within the MAGs. This differentiation was based on specific criteria such as nitrogen grain content (high or low) and storage time (harvest, three, six and nine months) of the samples. The output from OrthoFinder was then fractioned into different segments. Proteins shared by all MAGs were classified as the “core orthogroups”. Next the orthogroups were grouped by nitrogen grain content and storage time to identify orthogroups that are shared and unique to these specific groups.

### 3.3.5 Functional annotation of Orthogroups

A single representative from each identified orthogroup was selected and analysed using a local implementation of eggNOG mapper v2.1.12 (Cantalapiedra *et al.*, 2021). Functional annotation was carried out using the HMMER sequence similarity search tool for protein family profile searches (Eddy, 2011), targeting the eggNOG database v5.0 (Huerta-Cepas *et al.*, 2019). The sequences that show similarity are mapped to their corresponding orthologous groups in the eggNOG database. These groups are based on evolutionary relationships and are assumed to have retained similar functions across species (Koonin, 2005). Once the sequences are mapped to specific orthologous groups, the tool assigns functional annotations to the input sequences. These annotations are derived from the known functions of the genes or proteins within each orthologous group. The eggNOG annotations also categorised the proteins based on their Clusters of Orthologous Groups (COGs) functional categories (Yin *et al.*, 2012). This approach enabled the exploration of the potential biological roles of these proteins across different MAGs, providing insights into the functional capabilities of each group.

### 3.3.6 Pathogenicity proteins in the reconstructed MAGs

To assess the presence of genes related to plant pathogenicity in the MAGs, the protein sequences from all MAGs were aligned against the Pathogen-Host Interactions Database (PHI-base version 5.0) (Urban *et al.*, 2020). This database is a collection of well-studied genes and proteins that have been experimentally confirmed to be involved in pathogen-host interactions (Urban *et al.*, 2015; Urban *et al.*, 2020). Proteins related to pathogenicity were sourced from the PHI-base (available at http://www.phi-database.org) and subjected to a BLASTP analysis (Shiryev *et al.*, 2007) against the annotated MAGs. Proteins displaying >70% amino acid identity and an alignment coverage of > 70% were identified as potential pathogenicity protein orthologues in the MAGs. Additionally, the proportion of proteins matching PHI-base homologs that exceeded 70% was calculated for each MAG. This was calculated by dividing the number of PHI-base hits per MAG by the total number of proteins in that MAG with greater than 70% amino acid identity. These ratios were depicted using bar plots and incorporated into a phylogenetic tree through iTOL to facilitate data interpretation and visualisation (Letunic and Bork, 2021).

## 3.4 Results and discussion

### 3.4.1 Genome reconstruction of the barley microbial community

In this study, a total of 82 metagenome-assembled genomes (MAGs) were recovered from the metagenome data generated from eight barley seed samples stored for different lengths of time (harvest, three, six and nine months) and with different nitrogen contents (high or low) (Table 3.1). The highest number of MAGs were derived from the metagenomes of barley seeds stored for three months (29 MAGs), followed by those sampled at harvest (27 MAGs). The lowest number of MAGs could be assembled form the barley stored for six months (10 MAGs). Typically, a greater number of MAGs were derived from the low-nitrogen content barley seeds (46 MAGs) than high-nitrogen content barley seeds (36 MAGs), with the exception of the harvest samples, where similar numbers of MAGs could be assembled (Table 3.1).

Table 3.1 Number of metagenome-assembled genomes (MAGs) recovered per barley metagenome samples.

| Sample | Storage duration | Grain nitrogen content | Number of MAGs |
|---|---|---|---|
| HHN | Harvest | High | 14 |
| HLN | Harvest | Low | 13 |
| T1HN | 3 months | High | 11 |
| T1LN | 3 months | Low | 18 |
| T2HN | 6 months | High | 4 |
| T2LN | 6 months | Low | 6 |
| T3HN | 9 months | High | 7 |
| T3LN | 9 months | Low | 9 |

### 3.4.2 Quality metrics of the barley seed MAGs

Genome completeness determines the accuracy and reliability of the genomic data acquired (Albanese and Donati, 2021), allowing for better functional inference and taxonomy classification (Eisenhofer *et al.*, 2023). Conversely, lower completeness levels may result in gaps in the genome, potentially leading to the omission of essential genes and pathways (Chen *et al.*, 2020). Studies have shown that accurate genomic analysis requires high-quality MAGs, which are defined by completeness levels greater than 90% and contamination levels less than 10% (He *et al.*, 2021; Liu *et al.*, 2021). All MAGs demonstrated completeness > 75%, with

50/82 being >90% complete (Fig. 3.1A and B; Table S3.2). These completeness values are in alignment with the high-quality draft criterion of the Minimum Information about a Metagenome-Assembled Genome (MIMAG) standards for Bacteria and Archaea (Bowers *et al.*, 2017). Further, most of the MAGs (~91%; 75/82) registered contamination levels < 5%, whereas the remaining seven MAGs exhibited contaminant levels between 5 and 10%, indicating the reliability and integrity of our dataset (Fig. 3.2A-D, Table S3.2). We identified a notable negative correlation between genome completeness and contamination ($r = -0.498$, $p < 0.00001$; Fig. 3.2A). In parallel, our data demonstrated a positive relationship between genome size and N50 metric ($r = 0.251$, $p = 0.023$; Fig. 3.2 B), indicating that larger genomes are often associated with superior assembly contiguity.

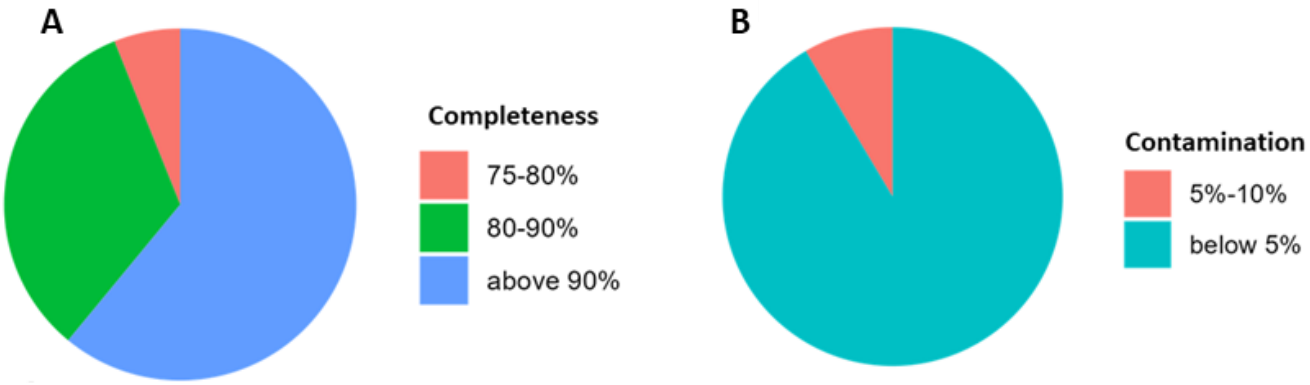


Figure 3.1 Comparative analysis of phylum distribution, MAGs completeness, and contamination. (A) Depicts the completeness of MAGs reconstructed from the metagenomic data, categorised into three ranges: 75-80%, 80-90%, and above 90%, which represent the percentage of the expected genome content present. (B) Illustrates the estimated contamination in the reconstructed MAGs, with categories indicating the percentage of sequences identified as foreign or contaminant, split into two ranges: 5-10% and below 5%.

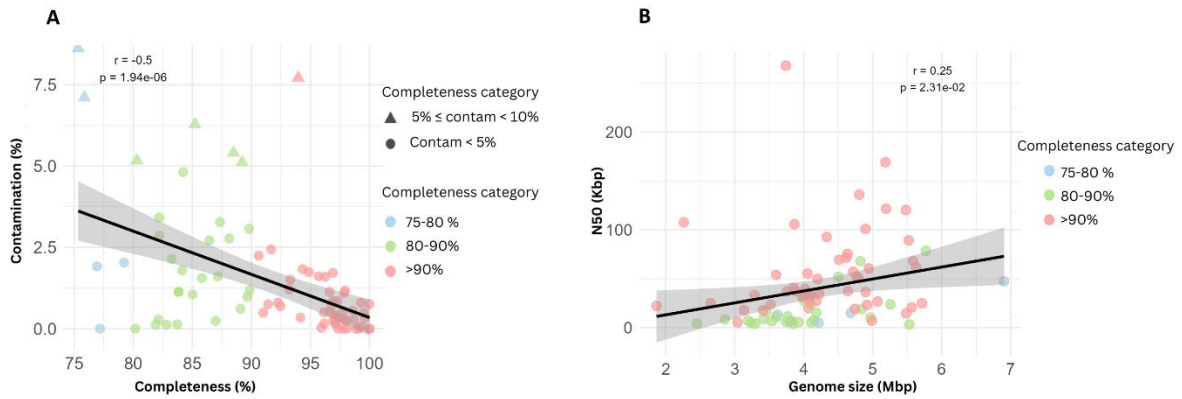


Figure 3.2 Correlations in metagenome-assembled genomes (MAGs). (A) relationship between completeness and contamination of MAGs (B) Correlation between genome size (Mbp) and N50 values (Kbp).

### 3.4.3 Taxonomic composition the MAGs

Taxonomic evaluation using the Genome Taxonomy Database Toolkit (GTDB-Tk) (Chaumeil *et al.*, 2020) revealed that the barley-associated MAG dataset was dominated by members of the phylum Pseudomonadota, comprising 53.7% (44/82) of the total MAGs (Fig. 3.3).The remainder of the MAGs belong to the phyla Bacteroidota 19.5% (16/82), Bacillota 17% (14/82) and Actinomycetota 9.7% (8/82) (Fig. 3.3, Table S3.2). These findings are consistent with our observations in Chapter 2 (Fig. 2.4, and Table S2.3), in which these four phyla were the most abundant, with Pseudomonadota accounting for the greatest number of sequences reads. Furthermore, these phyla have been observed to dominate the seed microbial communities of various plants, including barley (Bziuk *et al.*, 2021; Tyc *et al.*, 2020; Raj *et al.*, 2019; Adam *et al.*, 2018; Matsumoto *et al.*, 2021).

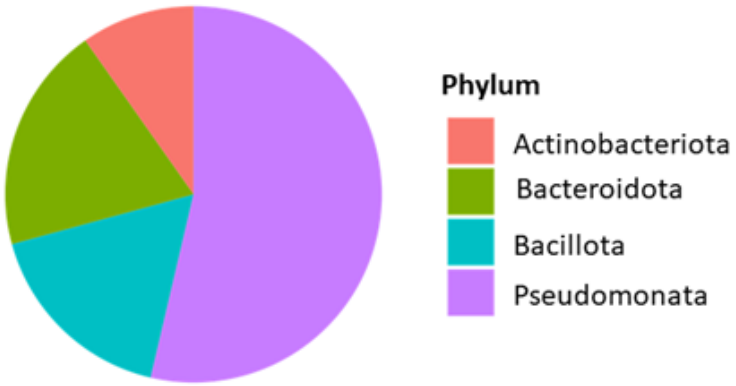


Figure 3.3 Relative abundance of microbial phyla identified in the metagenomic dataset.

MAGs from the phylum Pseudomonadota were found in abundance in all samples, except for the low-nitrogen content barley seeds stored for six months, where the Bacillota dominated. MAGs belonging to the phylum Bacillota were more common after six and nine months, with four and six MAGs, respectively, whereas only one MAG was detected from barley at harvest and barley seeds stored for three months (Table S3.2). This is consistent with the findings in Chapter 2 (Table S2.3) which show that the six month samples (S6L and S6H) included the highest number of Bacillota sequence reads.

The barley-seed-derived MAGs were further classified at the genus and species levels where possible. A total of 26 bacterial genera (Fig. 3.3, Table S3.2) were observed across the barley

metagenomes. The most abundant genera included *Pseudomonas*_E (9 MAGs), *Chryseobacterium* (7 MAGs), *Duffyella* (7 MAGs) and *Erwinia* (6 MAGs) (Fig. 3.4). It is noteworthy that sixteen MAGs could not be classified at the species level, highlighting the underexplored microbial diversity associated with barley seeds (Table S3.2). Most of the MAGs not assigned to a species belong to the phylum Bacteroidota (Table S3.2).

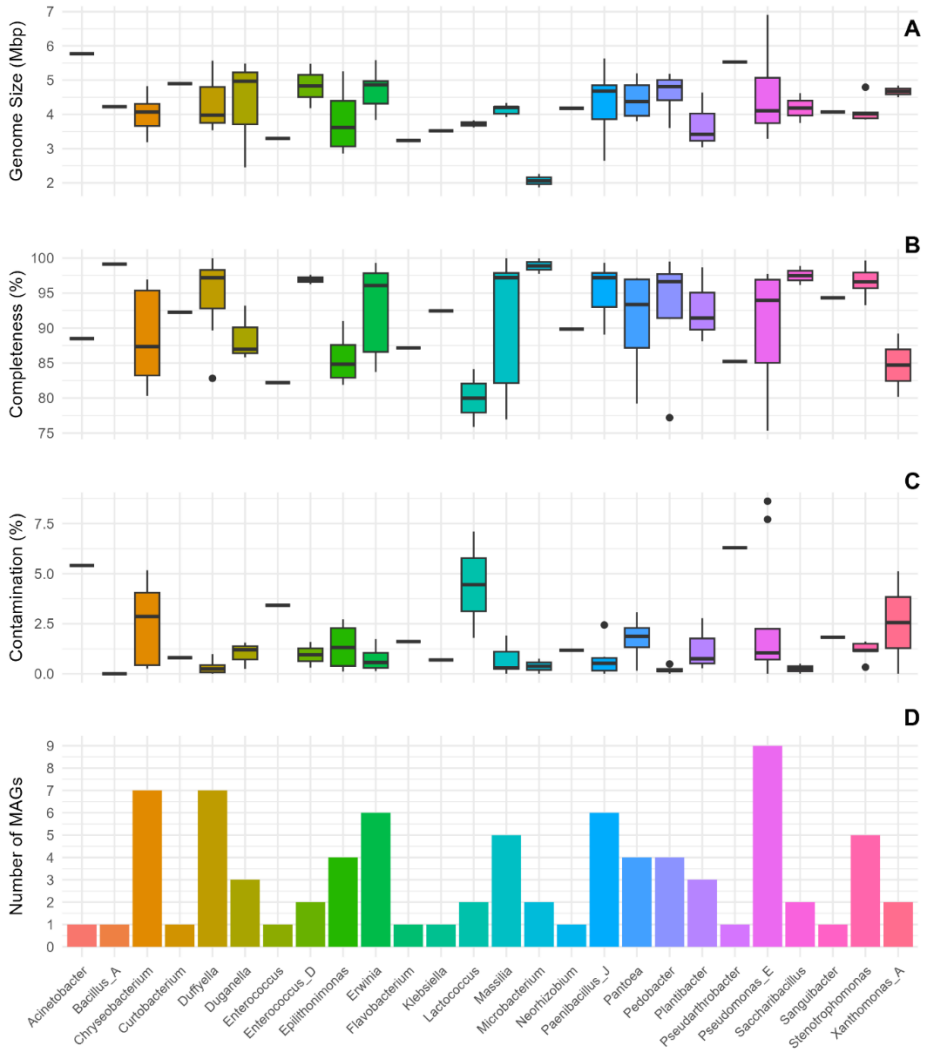


Figure 3.4 Genomic metrics of the identified bacterial genera**.** (A) Genome size distributions across genera. (B) Representation of genome completeness (C) contamination levels in assembled genomes (D) Number of Metagenome-assembled genomes (MAGs) identified per genus. The letter following the genus names indicates a novel genus as predicted by the GTDB (Genome Taxonomy Database) (Parks *et al.*, 2022) classification system.

### 3.4.3.1 Distribution of MAG genera per nitrogen content and storage duration

The distribution of MAGs by genus, as well as their abundance patterns grouped based on barley seed nitrogen grain content, exhibit trends closely aligning with the metagenomic read-based taxonomy previously reported in Chapter 2 (Table 2.1). For instance, in Chapter 2 (Figure 2.4), reads assigned to the genus *Pseudomonas* were abundant in barley seeds from both high and low nitrogen environments. Similarly, reads associated with the genus *Pantoea* and *Stenotrophomonas* were more abundant in barley samples with low nitrogen, a trend that is consistent with the higher number of MAGs observed in this study (Fig. 3.4A).

The abundance of MAGs at the different storage time point also mirrors the patterns seen in Chapter 2 (Figure 2.5). For example, at harvest in reads assigned to the genus *Chryseobacterium* and *Stenotrophomonas* were abundant (Chapter 2 Fig. 2.4); the same trend is also observed in this chapter (Fig.3.5A). *Pseudomonas*_E MAGs were found in all storage samples, while *Chryseobacterium* was most abundant in the harvest samples. Similarly, relative abundance profiles of the classified metagenomic reads showed that *Pseudomonas* (*Pseudomonas*_E) is not included in the kraken2 database (Lu *et al.*, 2022) recorded the highest read count across all samples and *Chryseobacterium* was predominantly found at harvest (Chapter 2; Fig. 2.4A). *Lactococcus*, which had the highest number of reads in Chapter 2 (Fig. 2.6B) at six months, mirrored this with *Lactococcus* MAGs only recovered from metagenomic DNA at the six month storage time point (Fig. 3.5B). Additionally, *Pedobacter*, *Plantibacter* and *Pseudarthrobacter*, primarily observed at harvest and six months of storage based on relative metagenomic read abundance (Fig. 3.5B), showed a similar presence in MAGs at these storage time points (Fig. 3.5B).

The congruence between taxonomic classifications derived from sequence reads and those from MAGs implies that taxa with higher read abundance are more likely to be assembled into bins, whereas those with lower abundance often fail to meet completeness criteria (Alneberg *et al.*, 2014). While the MAGs discussed in this chapter do not provide extensive insights into the overall microbial community composition, further exploration through annotation and comparative genomics will offer at least a preliminary understanding of the genomic content of dominant taxa within the barley samples.

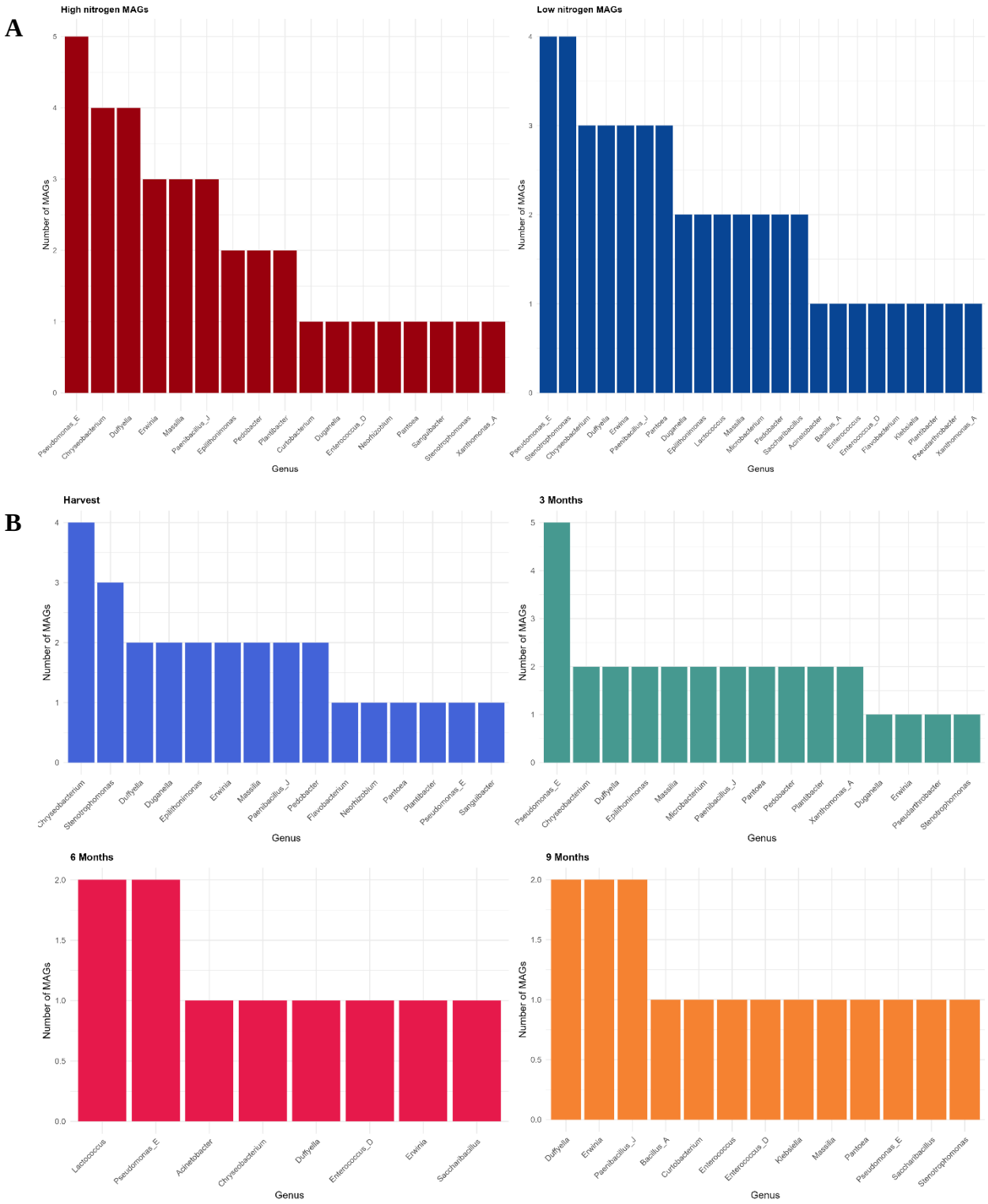


Figure 3.5 Distribution of metagenome-assembled genomes (MAGs) genera. (A) MAGs count per grain nitrogen content (high or low) and (B) storage duration (harvest, three, six and nine months).

### 3.4.4 The number of MAGs identified decreased over storage time

A decrease in the number of MAGs throughout the storage periods was observed (Fig. 3.6). This is consistent with what was observed in Chapter 2 (Section 2.4.1, Table S2.1), where the number of sequence reads obtained per storage period also decreased (Fig. 3.6). The binning process of MAGs involves assembling microbial genomes from raw sequencing data, which requires that the number of reads for a particular microorganism exceed a certain threshold to ensure accurate and reliable genome assembly (Alneberg *et al.*, 2014). This suggests that the microorganisms were sufficiently prevalent to be detected through their genetic material. The MAG assembly process improves our understanding of microbial communities by identifying which taxa or strains predominate at specific nitrogen grain content and storage time. The reduction in MAG counts over storage duration indicates that fewer microbial species maintain the abundance necessary for detection and assembly into MAGs.

This observation suggests a decline in microbial diversity and abundance, which could be attributed to several factors. One possible explanation is Gause's law of competitive exclusion, which states that two species competing for the same limited resources cannot coexist at constant population levels (Gause, 1934; Hening and Nguyen, 2020). Additionally, even under controlled silo storage conditions, subtle changes in nutrient availability are inevitable over time (Abubakar, 2019). As nutrients deplete, competition among microorganisms intensifies (Wieder *et al.*, 2015; Zhu *et al.*, 2017). This competition may impact previously abundant species, making it difficult for them to thrive. Consequently, their reduced abundance hinders the assembly of their reads into MAGs.

The reduction in microbial diversity with prolonged storage could have both positive and negative implications for the health and resilience of the barley seeds. A diverse microbial community is often linked to improved resistance against pathogens, enhanced nutrient availability and overall plant health (Harman *et al.*, 2021; Trivedi *et al.*, 2020). A decline in microbial diversity might expose the seeds to a higher risk of colonisation by opportunistic pathogens or lead to a decrease in beneficial microbial interactions (Hardoim, 2019; Rosenblueth and Martínez-Romero, 2006). Empirical validation of these hypotheses through *in vitro* assays is essential to elucidate the specific impacts and guide effective seed storage practices.

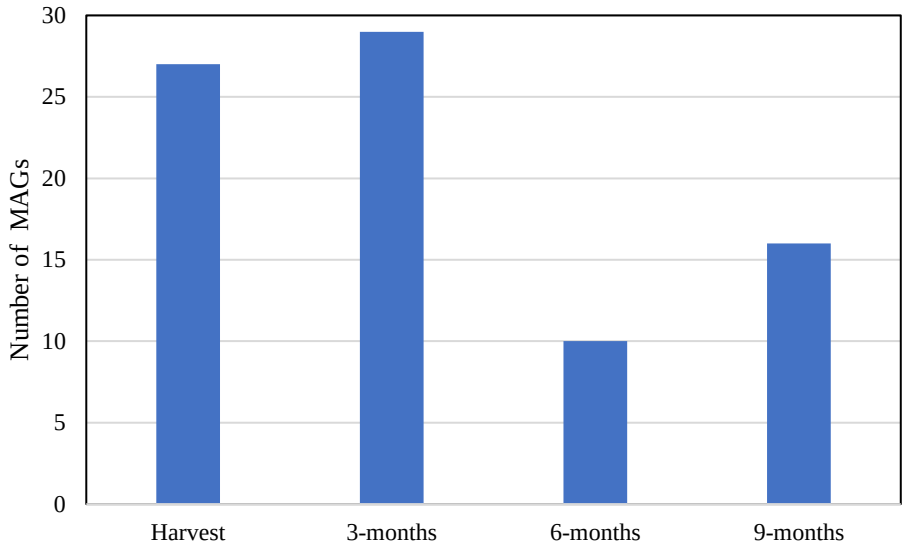


Figure 3.6 The abundance of metagenome-assembled genomes (MAGs) recovered per storage time.

### 3.4.5 Functional analysis of the MAGs hints at taxonomic diversity

Structural annotation of the 82 MAGs yielded a total of 317,760 proteins. Of these, 95.5% (317,760 proteins) were assigned to orthogroups, while the remaining 4.5% (14,847 proteins) were singletons, occurring in a single MAG (Table 3.3). The proteins were distributed across 34,233 orthogroups. Notably, the analysis identified 707 orthogroups that are distinct to each MAG, comprising 1,682 proteins, which represent merely 0.5% of the total (Table 3.3). The presence of these MAG-specific orthogroups, indicates both functional conservation and a degree of specialisation (Table 3.3) (Plett *et al.*, 2023). Such distinct adaptations improve the generalist functions seen in shared and core orthogroups, establishing a collaborative community in which diverse microorganisms play different roles, ultimately benefiting the group as a whole (Yajima *et al.*, 2023). The low incidence of single-copy orthogroups (SCOs) observed in the dataset is primarily a consequence of taxonomic diversity, reflecting a population-wide orthology analysis. The level of taxonomic diversity among the taxa analysed is a critical factor; as it increases, there is a natural decline in both SCOs and shared orthogroups (Steenwyk *et al.*, 2022; Emms and Kelly, 2015). This pattern illustrates the genetic distance, which significantly escalates among microorganisms from different phyla.

Table 3.2 OrthoFinder general statistics (Emms and Kelly, 2015).

| | |
|---|---|
| Number of species (MAGs) | 82 |
| Number of proteins | 332607 |
| Number of proteins in orthogroups | 317760 |
| Number of unassigned proteins | 14847 |
| Percentage of proteins in orthogroups | 95,5 |
| Percentage of unassigned proteins | 4,5 |
| Number of orthogroups | 34233 |
| Number of MAG-specific orthogroups | 707 |
| Number of proteins in MAG-specific orthogroups | 1682 |
| Percentage of proteins in MAG-specific orthogroups | 0,5 |
| Mean orthogroup size | 16,4 |
| Median orthogroup size | 5 |
| G50 (assigned proteins) | 56 |
| G50 (all proteins) | 51 |
| O50 (assigned proteins) | 1281 |
| O50 (all proteins) | 1420 |
| Number of orthogroups with all species present | 66 |
| Number of single-copy orthogroups | 1 |

### 3.4.6 Diversity of orthogroups in MAGS from barley seeds with different nitrogen grain contents and stored for different durations

The classification of orthogroups in MAGs based on nitrogen grain content (high or low) and over storage time (harvest, three, six and nine months) revealed differences in orthogroup diversity. In low nitrogen samples, 11,128 distinct orthogroups were identified, with 1,331 being shared by at least two species and 9,797 singletons (Figure 3.7A). In contrast, high-nitrogen samples exhibited a lower diversity, with only 6,182 orthogroups, including 425 shared and 5,757 as singletons (Figure 3.7A). The greater diversity in low-nitrogen MAGs can be attributed to a larger sample size (46 MAGs in low nitrogen versus 36 in high nitrogen), which provides a broader genetic base, enhancing the likelihood of identifying unique orthogroups and taxonomic diversity. Therefore, the observed differences in orthogroup diversity reflects underlying taxonomic diversity, pointing to variations in the microbial community that are yet to be fully explored.

Further analysis over different storage durations also demonstrated a dynamic pattern in the diversity of orthogroups (Figure 3.7B). Initially, at harvest, 5,281 orthogroups were identified, including 526 shared and 4,755 singletons. After three months of storage, the number modestly increased to 5,578 distinct orthogroups, comprising 718 shared and 4,860 singletons (Fig. 3.7B). However, a noticeable decrease was observed by six months, with only 2,334 unique orthogroups remaining, including 91 shared and 2,243 singletons. Similar results were observed in nine months storage including 85 shared and 3,360 (Fig 3.7). This trend aligns with the observed decrease in the number of MAGs and raw sequence reads over time, reflecting a similar pattern to the high-singleton prevalence seen under different nitrogen conditions.

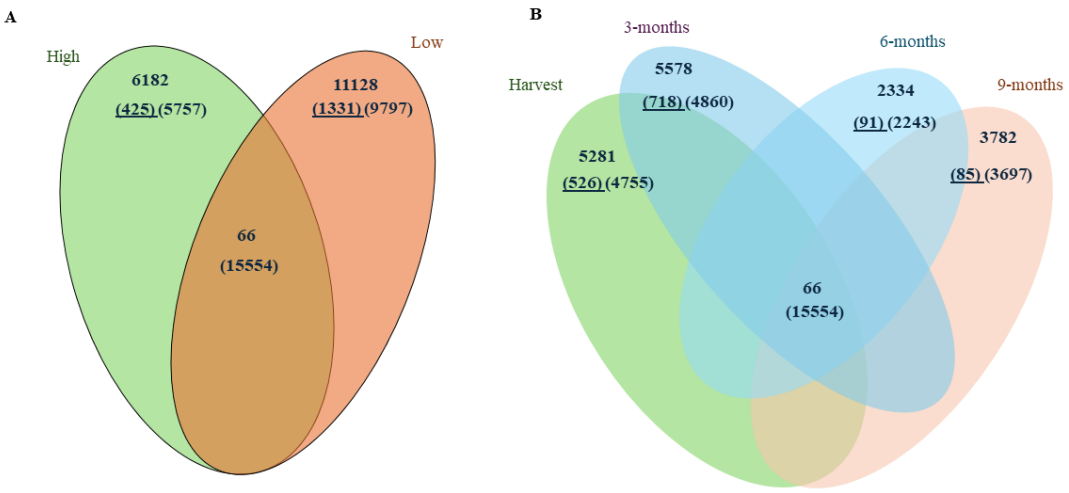


Figure 3.7 Shared and unique orthogroups in metagenome-assembled genomes (MAGs). (A) Shared and unique orthogroups of MAGs based on grain nitrogen content (high or low) (B) Shared and unique orthogroups of MAGs based on storage duration (harvest, three, six and nine months). The underlined number represent orthogroups shared by at least two species.

### 3.4.7 Functional insights into core and unique orthogroups in barley seed microbiome through COG analysis

Approximatively 35% of proteins shared by at least two MAG taxa from the OrthoFinder output did not have orthologues in the eggNOG database and were classified in the COG category ‘function unknown’ (S) (Fig. 3.8). Analysis of the core proteins COG data revealed that the highest proportion of proteins belong to the COG functional categories ‘translation,

ribosomal structure, and biogenesis' (J; 14%) and 'DNA replication, recombination, and repair' (L;14%). The high representation of proteins involved in these categories in the barley seed microbiome underscores the fundamental cellular functions necessary for growth and adaptation to environmental conditions (Truyens *et al.*, 2015) (Fig. 3.8).

The COG functional categorisation of unique proteins from MAGs derived from barley seeds with differing nitrogen content reveals a distinct functional adaptation within the seed microbiome. In MAGs from low nitrogen conditions, the dominant COG category (aside from category S) is 'transcription' (K; 10%). In contrast, MAGs from high nitrogen conditions show a prevalence in the category of 'cell wall/membrane/envelope biogenesis' (M; 11%) (Fig. 3.8)

The analysis of COG categories across different storage times for MAGs derived from barley seeds reveals an intriguing shift in the dominant metabolic functions of the microbial communities (Fig. 3.8). At harvest and continuing through three months of storage, the predominant functional category is 'carbohydrate transport' and 'metabolism' (G; 11% and 10% respectively). This trend changes by nine months, where 'amino acid transport and metabolism' (E; 22%) becomes the dominant function. Initially, the microbial communities exploit the abundant carbohydrate resources, but as these decrease in availability, they may increasingly turn to amino acids derived from the breakdown of seed proteins or from nitrogenous compounds in the seed environment (Pan *et al.*, 2022). This shift could impact the nutritional quality and shelf life of the grain, potentially affecting its viability and economic value (Ziegler *et al.*, 2021).

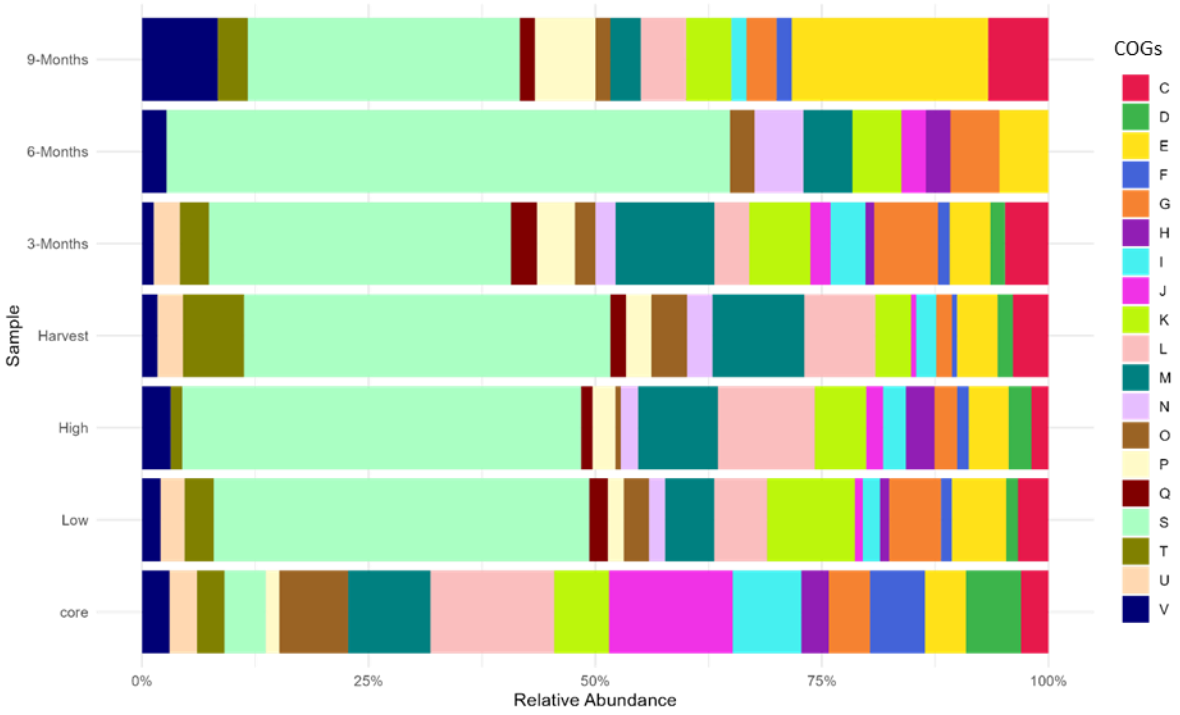


Figure 3.8 Functional annotation of unique proteins in metagenome assembled genomes (MAGs). Cluster of orthologous function (COGs) (Huerta-Cepas *et al.*, 2019). of shared and unique orthogroups of MAGs based on nitrogen grain content and storage time. **A:** RNA processing and modification; **B:** Chromatin structure and dynamics; **C:** Energy production and conversion; **D:** Cell cycle control, cell division, chromosome partitioning; **E:** Amino acid transport and metabolism, **F:** Nucleotide transport and metabolism; **G:** Carbohydrate transport and metabolism; **H:** Coenzyme transport and metabolism; **I:** Lipid transport and metabolism; **J:** Translation, ribosomal structure and biogenesis; **K:** Transcription; **L:** Replication, recombination and repair; **M:** Cell wall/membrane/envelope biogenesis; **N:** Cell motility; **O:** Posttranslational modification, protein turnover, chaperones; **P:** Inorganic ion transport and metabolism; **Q:** Secondary metabolites biosynthesis, transport and catabolism; **S:** Unknown; **T:** Signal transduction mechanisms; **U:** Intracellular trafficking, secretion, and vesicular transport; **V:** Defence mechanisms; **W:** Extracellular structures; **Y:** Nuclear structure; **Z:** Cytoskeleton.

### 3.4.8 Survey of proteins involved in pathogenicity in the MAGs

The barley seed microbiome is populated by a varied range of microorganisms, some of which are known to play a role in pathogenicity (Bziuk *et al.*, 2021; Noots *et al.*, 1999). The PHI database (Urban *et al.*, 2020) was searched to identify proteins in each MAG exhibiting over 70% homology with known pathogenic proteins. Of the 82 MAGs analysed, 20 exhibited a pathogenicity ratio exceeding 50%. Notably, MAGs from the genera *Erwinia*, *Duffyella*, *Klebsiella* and *Pantoea*, showed particularly high pathogenicity ratios Fig 3.9. For instance, *Erwinia persicina* MAGs (6 MAGs) exhibit pathogenicity ratios ranging from 0.67 to 0.74. Members of the genus *Erwinia* are implicated in causing plant diseases, with *E. persicina* specifically known for causing pink seed disease in barley (Kawaguchi *et al.*, 2021; Wasendorf *et al.*, 2022). Similarly, *Duffyella gerundensis* MAGs (6 MAGs) displayed ratios above 0.73

(Fig. 3.9). The species *D. gerundensis*, previously classified as *E. gerundensis*, was reclassified following taxonomic analyses that determined it did not belong to the *Erwinia* species group (Soutar and Stavrinides, 2022).

*Pantoea agglomerans* MAGs (2 MAGs) demonstrated a pathogenicity ratio of 0.86, reflecting its dual functionality within plant ecosystems as both a pathogen and a plant growth promoter (Lv *et al.*, 2022). Specifically, *P. agglomerans* has been implicated in reducing seed germination, seedling growth and overall seed quality in oats (Wang *et al.*, 2022). Additionally, *Klebsiella grimontii* MAG70 displayed the highest ratio of 0.87 (Fig. 3.9). This species is known for its antifungal activity against *Fusarium oxysporum*, a pathogen that impacts soybean plants (Wang *et al.*, 2023). In contrast, genera such as *Chryseobacterium* (7 MAGs), *Microbacterium* (2 MAGs) and *Pedobacter* (4 MAGs) exhibited much lower pathogenicity ratios (Fig. 3.9). Member of these genera are mainly associated with plant growth promotion roles in plants (Hayat *et al.*, 2010).

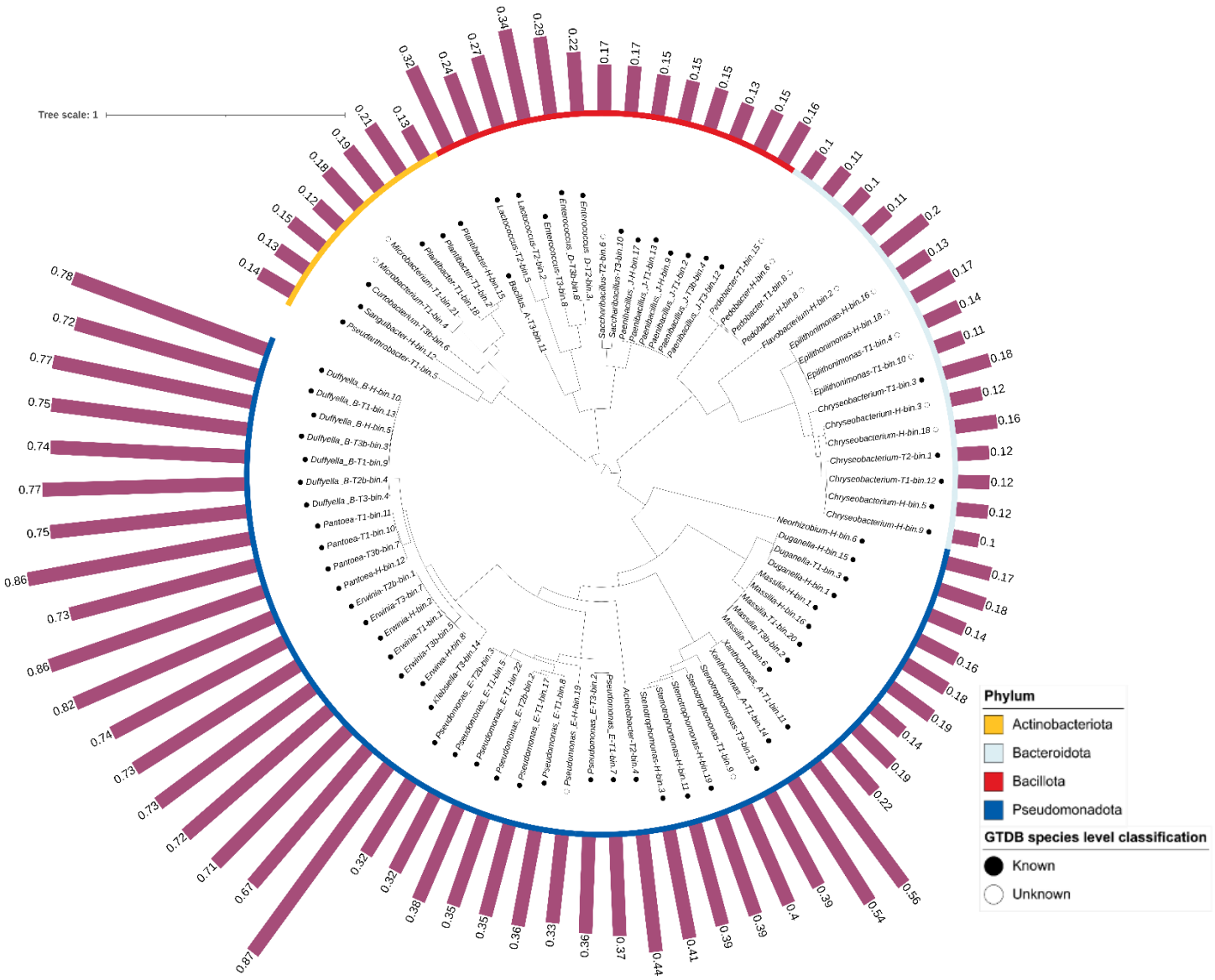


Figure 3.9 Phylogenetic relationships of bacterial metagenome-assembled genomes (MAGs). Phylogenetic tree of 82 species-level MAGs based on 120 conserved bacterial marker genes. Bar plot displaying the ratio of pathogenicity-related proteins, each with homology greater than 70% to entries in the Pathogen Host Interaction (PHI) database (Urban *et al.*, 2020), identified in each MAG.

## 3.5 Conclusions

This study has reconstructed 82 high-quality MAGs from barley seeds of varying nitrogen content and storage durations, providing detailed insights into the genomic content of the seed-associated microbiome. The analysis revealed substantial taxonomic diversity, with MAGs classified into 26 bacterial genera from four distinct phyla. Despite this diversity, functional conservation was evident in the presence of core orthogroups related to essential cellular functions, including translation, ribosomal structure, and DNA repair. These functions suggest that the barley seed microbiome plays a critical role in maintaining microbial viability and activity during extended storage, potentially contributing to the overall stability of the seed environment.

Notably, the identification of proteins associated with plant pathogenicity from genera such as *Erwinia* and *Klebsiella* underscores the complex role of these microorganisms. While these findings do not directly indicate a pathogenic impact on barley seeds during storage, they suggest that these microorganisms may possess latent pathogenic capabilities, which could become relevant under specific conditions. This highlights the need for further investigation into the ecological roles of these microbes within the stored seed microbiome.

While this study offers valuable genomic insights into the barley seed microbiome, it is important to acknowledge its limitations. The reconstructed MAGs represent only a portion of the microbial community, focusing primarily on higher-abundance taxa. Future studies should incorporate deeper sequencing and complementary approaches, such as single-cell genomics, to capture the full complexity of the microbial community, including low-abundance species that may play crucial roles in barley seed health and preservation.

## 3.6 References


Abubakar, S. (2019). Effect of Storage Conditions on wheat Quality. *International Journal of Food and Allied Sciences*, **4**: 20–23.

Adam, E., Bernhart, M., Müller, H., Winkler, J., Berg, G. (2018). The *Cucurbita pepo* seed microbiome: genotype-specific composition and implications for breeding. *Plant and Soil*, **422**: 35–49.

Albanese, D., Donati, C. (2021). Genome Recovery, Functional Profiling, and Taxonomic Classification from Metagenomes. *Methods in Molecular Biology (Clifton, N.J.)*, **2242**: 153–172.

Alneberg, J., Bjarnason, B.S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., Lahti, L., Loman, N.J., Andersson, A.F., Quince, C. (2014). Binning metagenomic contigs by coverage and composition. *Nature Methods*, **11**: 1144–1146.

Bamforth, C.W. (2017). Progress in Brewing Science and Beer Production. *Annual Review of Chemical and Biomolecular Engineering*, **8**: 161–176.

Bianco, A., Fancello, F., Balmas, V., Zara, G., Dettori, M., Budroni, M. (2018). The microbiome of Sardinian barley and malt. *Journal of the Institute of Brewing*, **124**: 344–351.

Bokulich, N.A., Bamforth, C.W. (2013). The microbiology of malting and brewing. *Microbiology and molecular biology reviews: MMBR*, **77**: 157–172.

Bowers, R.M., Kyrpides, N.C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T.B.K., Schulz, F., Jarett, J., Rivers, A.R., Eloe-Fadrosh, E.A., Tringe, S.G., Ivanova, N.N., Copeland, A., Clum, A., Becraft, E.D., Malmstrom, R.R., Birren, B., Podar, M., Bork, P., Weinstock, G.M., Garrity, G.M., Dodsworth, J.A., Yooseph, S., Sutton, G., Glöckner, F.O., Gilbert, J.A., Nelson, W.C., Hallam, S.J., Jungbluth, S.P., Ettema, T.J.G., Tighe, S., Konstantinidis, K.T., Liu, W.-T., Baker, B.J., Rattei, T., Eisen, J.A., Hedlund, B., McMahon, K.D., Fierer, N., Knight, R., Finn, R., Cochrane, G., Karsch-Mizrachi, I., Tyson, G.W., Rinke, C., Genome Standards Consortium, Lapidus, A., Meyer, F., Yilmaz, P., Parks, D.H., Eren, A.M., Schriml, L., Banfield, J.F., Hugenholtz, P., Woyke, T. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature Biotechnology*, **35**: 725–731.

Buchfink, B., Xie, C., Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, **12**: 59–60.

Bziuk, N., Maccario, L., Straube, B., Wehner, G., Sørensen, S.J., Schikora, A., Smalla, K. (2021). The treasure inside barley seeds: microbial diversity and plant beneficial bacteria. *Environmental Microbiome*, **16**: 20.

Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, **38**: 5825–5829.

Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., Parks, D.H. (2020). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*, **36**: 1925–1927.

Chen, L.-X., Anantharaman, K., Shaiber, A., Eren, A.M., Banfield, J.F. (2020). Accurate and complete genomes from metagenomes. *Genome Research*, **30**: 315–333.

Dindhoria, K., Kumar, Raghawendra, Bhargava, B., Kumar, Rakshak. (2024). Metagenomic assembled genomes indicated the potential application of hypersaline microbiome for plant growth promotion and stress alleviation in salinized soils. *mSystems*, **0**: e01050-23.

Eddy, S.R. (2011). Accelerated Profile HMM Searches. *PLOS Computational Biology*, **7**: e1002195.

Eisenhofer, R., Odriozola, I., Alberdi, A. (2023). Impact of microbial genome completeness on metagenomic functional inference. *ISME Communications*, **3**: 12.

Emms, D.M., Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, **16**: 157.

Fadiji, A.E., Babalola, O.O. (2020). Metagenomics methods for the study of plant-associated microbial communities: A review. *Journal of Microbiological Methods*, **170**: 105860.

Felšöciová, S., Kowalczewski, P.Ł., Krajčovič, T., Dráb, Š., Kačániová, M. (2021). Effect of Long-Term Storage on Mycobiota of Barley Grain and Malt. *Plants (Basel, Switzerland)*, **10**: 1655.

Felšöciová, S., Kowalczewski, P.Ł., Krajčovič, T., Dráb, Š., Kačániová, M. (2020). Quantitative and Qualitative Composition of Bacterial Communities of Malting Barley Grain and Malt during Long-Term Storage. *Agronomy*, **10**: 1301.

Gagliano, M.C., Sampara, P., Plugge, C.M., Temmink, H., Sudmalis, D., Ziels, R.M. (2022). Functional Insights of Salinity Stress-Related Pathways in Metagenome-Resolved Methanothrix Genomes. *Applied and Environmental Microbiology*, **88**: e02449-21.

Gause, G.F. (1934). Experimental Analysis of Vito Volterra's Mathematical Theory of the Struggle for Existence. *Science*, **79**: 16–17.

Goodwin, S., McPherson, J.D., McCombie, W.R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, **17**: 333–351.

Grant, K.R. (2022). Next-Generation Amplicon Sequencing: A Cost-Effective Method for Exploring Microbial Biodiversity. In: Kumar, A., Choudhury, B., Dayanandan, S., & Khan, M.L. (eds) *Molecular Genetics and Genomics Tools in Biodiversity Conservation*. Springer Nature, Singapore, pp. 203–236.

Guo, L.-Y., Xie, X., Zhao, N., Wang, M., Lin, Z.-P., Song, Y.-M., Jia, F.-C. (2021). Investigation of potential strategies to control malt-induced premature yeast flocculation. *Systems Microbiology and Biomanufacturing*, **1**: 193–199.

Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, **29**: 1072–1075.

Hardoim, P. (2019). The Ecology of Seed Microbiota. In: Verma, S.K. & White, J., James Francis (eds) *Seed Endophytes: Biology and Biotechnology*. Springer International Publishing, Cham, pp. 103–125.

Harman, G., Khadka, R., Doni, F., Uphoff, N. (2021). Benefits to Plant Health and Productivity From Enhancing Plant Microbial Symbionts. *Frontiers in Plant Science*, **11**.

Hayat, R., Ali, S., Amara, U., Khalid, R., Ahmed, I. (2010). Soil beneficial bacteria and their role in plant growth promotion: a review. *Annals of Microbiology*, **60**: 579–598.

He, C., Keren, R., Whittaker, M.L., Farag, I.F., Doudna, J.A., Cate, J.H.D., Banfield, J.F. (2021). Genome-resolved metagenomics reveals site-specific diversity of episymbiotic CPR bacteria and DPANN archaea in groundwater ecosystems. *Nature Microbiology*, **6**: 354–365.

Hening, A., Nguyen, D.H. (2020). The competitive exclusion principle in stochastic environments. *Journal of Mathematical Biology*, **80**: 1323–1351.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., von Mering, C., Bork, P. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, **47**: D309–D314.

Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, **11**: 119.

Jiao, J.-Y., Liu, L., Hua, Z.-S., Fang, B.-Z., Zhou, E.-M., Salam, N., Hedlund, B.P., Li, W.-J. (2021). Microbial dark matter coming to light: challenges and opportunities. *National Science Review*, **8**: nwaa280.

Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., Wang, Z. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*, **7**: e7359.

Kawaguchi, A., Abe, D., Saito, T., Nogata, Y., Nomiyama, K., Kohyama, N., Takahashi, A., Yoshioka, T., Ishikawa, N., Tomioka, K. (2021). Pink seed of barley caused by *Erwinia persicina*. *Journal of General Plant Pathology*, **87**: 106–109.

Kayani, M. ur R., Huang, W., Feng, R., Chen, L. (2021). Genome-resolved metagenomics using environmental and clinical samples. *Briefings in Bioinformatics*, **22**: bbab030.

Koonin, E.V. (2005). Orthologs, paralogs, and evolutionary genomics. *Annual Review of Genetics*, **39**: 309–338.

Laslett, D., Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Research*, **32**: 11–16.

Letunic, I., Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, **49**: W293–W296.

Liu, L., Wang, Y., Yang, Y., Wang, D., Cheng, S.H., Zheng, C., Zhang, T. (2021). Charting the complexity of the activated sludge microbiome through a hybrid sequencing strategy. *Microbiome*, **9**: 205.

Lok, C. (2015). Mining the microbial dark matter. *Nature*, **522**: 270–273.

Lu, J., Rincon, N., Wood, D.E., Breitwieser, F.P., Pockrandt, C., Langmead, B., Salzberg, S.L., Steinegger, M. (2022). Metagenome analysis using the Kraken software suite. *Nature Protocols*, **17**: 2815–2839.

Lv, L., Luo, J., Ahmed, T., Zaki, H.E.M., Tian, Y., Shahid, M.S., Chen, J., Li, B. (2022). Beneficial Effect and Potential Risk of Pantoea on Rice Production. *Plants*, **11**: 2608.

Matsumoto, H., Fan, X., Wang, Yue, Kusstatscher, P., Duan, J., Wu, S., Chen, S., Qiao, K., Wang, Yiling, Ma, B., Zhu, G., Hashidoko, Y., Berg, G., Cernava, T., Wang, M. (2021). Bacterial seed endophyte shapes disease resistance in rice. *Nature Plants*, **7**: 60–72.

Michl, K., Berg, G., Cernava, T. (2023). The microbiome of cereal plants: The current state of knowledge and the potential for future applications. *Environmental Microbiome*, **18**: 28.

Muhamad Rizal, N.S., Neoh, H., Ramli, R., A/L K Periyasamy, P.@ R., Hanafiah, A., Abdul Samat, M.N., Tan, T.L., Wong, K.K., Nathan, S., Chieng, S., Saw, S.H., Khor, B.Y. (2020). Advantages and Limitations of 16S rRNA Next-Generation Sequencing for Pathogen Identification in the Diagnostic Microbiology Laboratory: Perspectives from a Middle-Income Country. *Diagnostics*, **10**: 816.

Noots, I., Delcour, J.A., Michiels, C.W. (1999). From field barley to malt: detection and specification of microbial activity for quality aspects. *Critical Reviews in Microbiology*, **25**: 121–153.

Nurk, S., Meleshko, D., Korobeynikov, A., Pevzner, P.A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Research*, **27**: 824–834.

Olm, M.R., Brown, C.T., Brooks, B., Banfield, J.F. (2017). dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME journal*, **11**: 2864–2868.

Pan, W., Tang, S., Zhou, J., Liu, M., Xu, M., Kuzyakov, Y., Ma, Q., Wu, L. (2022). Plant–microbial competition for amino acids depends on soil acidity and the microbial community. *Plant and Soil*, **475**: 457–471.

Parks, D.H., Chuvochina, M., Rinke, C., Mussig, A.J., Chaumeil, P.-A., Hugenholtz, P. (2022). GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Research*, **50**: D785–D794.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, **25**: 1043–1055.

Plett, J.M., Miyauchi, S., Morin, E., Plett, K., Wong-Bajracharya, J., de Freitas Pereira, M., Kuo, A., Henrissat, B., Drula, E., Wojtalewicz, D., Riley, R., Pangilinan, J., Andreopoulos, W., LaButti, K., Daum, C., Yoshinaga, Y., Fauchery, L., Ng, V., Lipzen, A., Barry, K., Singan, V., Guo, J., Lebel, T., Costa, M.D., Grigoriev, I.V., Martin, F., Anderson, I.C., Kohler, A. (2023). Speciation Underpinned by Unexpected Molecular Diversity in the Mycorrhizal Fungal Genus *Pisolithus*. *Molecular Biology and Evolution*, **40**: msad045.

Raj, G., Shadab, M., Deka, S., Das, M., Baruah, J., Bharali, R., Talukdar, N.C. (2019). Seed interior microbiome of rice genotypes indigenous to three agroecosystems of Indo-Burma biodiversity hotspot. *BMC Genomics*, **20**: 924.

Rosenblueth, M., Martínez-Romero, E. (2006). Bacterial Endophytes and Their Interactions with Hosts. *Molecular Plant-Microbe Interactions®*, **19**: 827–837.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, **30**: 2068–2069.

Setubal, J.C. (2021). Metagenome-assembled genomes: concepts, analogies, and challenges. *Biophysical Reviews*, **13**: 905–909.

Shiryev, S.A., Papadopoulos, J.S., Schäffer, A.A., Agarwala, R. (2007). Improved BLAST searches using longer words for protein seeding. *Bioinformatics (Oxford, England)*, **23**: 2949–2951.

Singh, M.P., Singh, P., Singh, R.K., Solanki, M.K., Bazzer, S.K. (2020). Plant Microbiomes: Understanding the Aboveground Benefits. In: Solanki, M.K., Kashyap, P.L., & Kumari, B. (eds) *Phytobiomes: Current Insights and Future Vistas*. Springer, Singapore, pp. 51–80.

Soutar, C.D., Stavrinides, J. (2022). Phylogenomic analysis of the Erwiniaceae supports reclassification of *Kalamiella piersonii* to *Pantoea piersonii* comb. nov. and *Erwinia gerundensis* to the new genus *Duffyella* gen. nov. as *Duffyella gerundensis* comb. nov. *Molecular Genetics and Genomics*, **297**: 213–225.

Steenwyk, J.L., Goltz, D.C., Iii, T.J.B., Li, Y., Shen, X.-X., Rokas, A. (2022). OrthoSNAP: A tree splitting and pruning algorithm for retrieving single-copy orthologs from gene family trees. *PLOS Biology*, **20**: e3001827.

Stewart, R.D., Auffret, M.D., Snelling, T.J., Roehe, R., Watson, M. (2019). MAGpy: a reproducible pipeline for the downstream analysis of metagenome-assembled genomes (MAGs). *Bioinformatics*, **35**: 2150–2152.

Trivedi, P., Leach, J.E., Tringe, S.G., Sa, T., Singh, B.K. (2020). Plant–microbiome interactions: from community assembly to plant health. *Nature Reviews Microbiology*, **18**: 607–621.

Truyens, S., Weyens, N., Cuypers, A., Vangronsveld, J. (2015). Bacterial seed endophytes: genera, vertical transmission and interaction with plants. *Environmental Microbiology Reports*, **7**: 40–50.

Tyc, O., Putra, R., Gols, R., Harvey, J.A., Garbeva, P. (2020). The ecological role of bacterial seed endophytes associated with wild cabbage in the United Kingdom. *MicrobiologyOpen*, **9**: e00954.

Urban, M., Cuzick, A., Seager, J., Wood, V., Rutherford, K., Venkatesh, S.Y., De Silva, N., Martinez, M.C., Pedro, H., Yates, A.D., Hassani-Pak, K., Hammond-Kosack, K.E. (2020). PHI-base: the pathogen–host interactions database. *Nucleic Acids Research*, **48**: D613–D620.

Urban, M., Irvine, A.G., Cuzick, A., Hammond-Kosack, K.E. (2015). Using the pathogen-host interactions database (PHI-base) to investigate plant pathogen genomes and genes implicated in virulence. *Frontiers in Plant Science*, **6**.

Uritskiy, G.V., DiRuggiero, J., Taylor, J. (2018). MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome*, **6**: 158.

Wang, J., Chen, T., White, J.F., Wei, X., Li, X., Li, C. (2022). *Pantoea agglomerans*, a seed-borne plant pathogenic bacterium, decreased seed germination, seedling growth and seed quality of oat. *European Journal of Plant Pathology*, **162**: 667–679.

Wang, S., Zheng, L., Gao, A., Xiao, Y., Han, Z., Pan, H., Zhang, H. (2023). Antifungal activity of *Klebsiella grimontii* DR11 against *Fusarium oxysporum* causing soybean root rot. *Journal of Applied Microbiology*, **134**: lxad245.

Wasendorf, C., Schmitz-Esser, S., Eischeid, C.J., Leyhe, M.J., Nelson, E.N., Rahic-Seggerman, F.M., Sullivan, K.E., Peters, N.T. (2022). Genome analysis of *Erwinia persicina* reveals implications for soft rot pathogenicity in plants. *Frontiers in Microbiology*, **13**.

Wieder, W.R., Cleveland, C.C., Smith, W.K., Todd-Brown, K. (2015). Future productivity and carbon storage limited by terrestrial nutrient availability. *Nature Geoscience*, **8**: 441–444.

Wu, Y.-W., Simmons, B.A., Singer, S.W. (2016). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics*, **32**: 605–607.

Wu, Z., Wang, Y., Zeng, J., Zhou, Y. (2022). Constructing metagenome-assembled genomes for almost all components in a real bacterial consortium for binning benchmarking. *BMC Genomics*, **23**: 746.

Yajima, D., Fujita, H., Hayashi, I., Shima, G., Suzuki, K., Toju, H. (2023). Core species and interactions prominent in fish-associated microbiome dynamics. *Microbiome*, **11**: 53.

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research*, **40**: W445–W451.

Zhou, Z., Tran, P.Q., Breister, A.M., Liu, Y., Kieft, K., Cowley, E.S., Karaoz, U., Anantharaman, K. (2022). METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. *Microbiome*, **10**: 33.

Zhu, Q., Riley, W.J., Tang, J. (2017). A new theory of plant–microbe nutrient competition resolves inconsistencies between observations and model predictions. *Ecological Applications*, **27**: 875–886.

Ziegler, V., Paraginski, R.T., Ferreira, C.D. (2021). Grain storage systems and effects of moisture, temperature and time on grain quality - A review. *Journal of Stored Products Research*, **91**: 101770.

# Chapter 4 Functional insights into the barley seed microbiome using a meta-omics approach

## 4.1 Abstract

The barley seed microbiome plays a key role in agricultural sustainability and quality of this crop. While the taxonomic landscape of the barley microbiome is well documented, particularly its implications for brewing quality influenced by storage practices and nitrogen content, the functional dynamics of its microbial constituents are not fully understood. This study employed a comprehensive meta-omics strategy, integrating metagenomics and metaproteomics, to delineate the functional capabilities and active biological processes of the barley seed microbiome during storage. A total of eight samples were investigated in this study. The metagenome assemblies ranged in size from 78 to 180 megabases, comprising a total of 274,778 contigs, averaging 34,347 contigs per sample. Structural annotation identified 952,951 protein-coding sequences (CDS), while functional annotation mapped these sequences to a broad array of metabolic pathways. Clusters of Orthologous Groups (COGs) analysis suggested functions such as transcription, amino acid metabolism, and ion transport that are essential for microbial adaptation to the storage environment. Additionally, the detection of carbohydrate-active enzyme (CAZyme) families, including glycosyl transferases and glycoside hydrolases, indicates a microbial proficiency in modifying plant-derived substrates. The nutrient cycling pathways that were reconstructed from the metagenomic data hinted at potential anaerobic activities that increased with storage duration. Complementarily, metaproteomic analysis showed that a broad range of outer membrane proteins, porins and catalases are expressed by the microbial community, suggesting active roles in nutrient transport, cellular defence and oxidative stress mitigation. To our knowledge, this is the first study that integrates metagenomics with metaproteomics to investigate the barley seed microbiome during storage.

## 4.2 Introduction

Analysis of the microbial population associated with barley seeds offers insights into seed storage longevity, contamination hazards and the potential effects of these microorganisms on beer production (Cao *et al.*, 2022; Felšöciová *et al.*, 2021; Felšöciová *et al.*, 2020). Despite

substantial advancements, there is still a limited understanding of the barley seed microbiome in comparison to other cereals (Michl *et al.*, 2023). This is due in part to a shortage of comprehensive metagenomic datasets focusing on these microbial communities, as well as a previous reliance on culture-dependent methods (Rahman *et al.*, 2018; Tshisekedi *et al.*, 2024). While these strategies provide useful insights, their applicability is limited (Yulan Chen *et al.*, 2022). Currently, only a number of studies have adopted culture-independent approaches, with a primary focus on taxonomic classification (Bziuk *et al.*, 2021). Crucially, the functional capacity of microorganisms associated with the barley seed remains elusive.

The assembly of metagenomic sequencing reads into longer contiguous sequences (contigs) makes it possible to acquire a thorough picture of the functional profile of microbial communities (Fadiji and Babalola, 2020). This is supplemented by *in silico* functional analysis, which involves gene prediction to identify open reading frames (ORFs) in the assembled sequences (van der Walt *et al.*, 2017). These predicted genes are then aligned against those maintained in databases such as the Kyoto Encyclopaedia of Genes and Genomes (KEGG), Clusters of Orthologous Groups (COGs) or CAZy (Carbohydrate-Active enZymes) databases to determine functional roles based on gene similarity (Ogata *et al.*, 1999; Park *et al.*, 2018; Tatusov *et al.*, 2003). While metagenomics identifies the genetic material in environmental samples, shedding light on the potential roles of the microbiome (Desai *et al.*, 2012), it does not verify active gene expression or protein production (Fadiji and Babalola, 2020). Metaproteomics fills this gap by analysing proteins extracted from environmental samples (Abraham *et al.*, 2014). In metaproteomics, it is common practice to match mass spectrometry spectra against annotated metagenomic databases, enhancing specificity and accuracy in protein identification. This approach allows for a direct link between protein function and genetic context (Erickson *et al.*, 2012; Heyer *et al.*, 2019).

In earlier chapters (Chapters 2 and 3), the taxonomic diversity of the barley microbiome at different storage time points in silos (at harvest and after three, six and nine months of storage) and reconstructed metagenome-assembled genomes (MAGs) were explored. Building on this taxonomic foundation, the current chapter adopted a meta-omics approach to investigate the functional profiles of the barley seed microbiome. The annotated metagenome sequences were studied and then used as a database for metaproteomic protein identification by comparing mass spectral data to proteins (Fig. 4.1). This combination of metagenomic and metaproteomic

data attempts to close the gap between genetic potential and actual biological activity, providing additional insight into the barley seed microbiome.

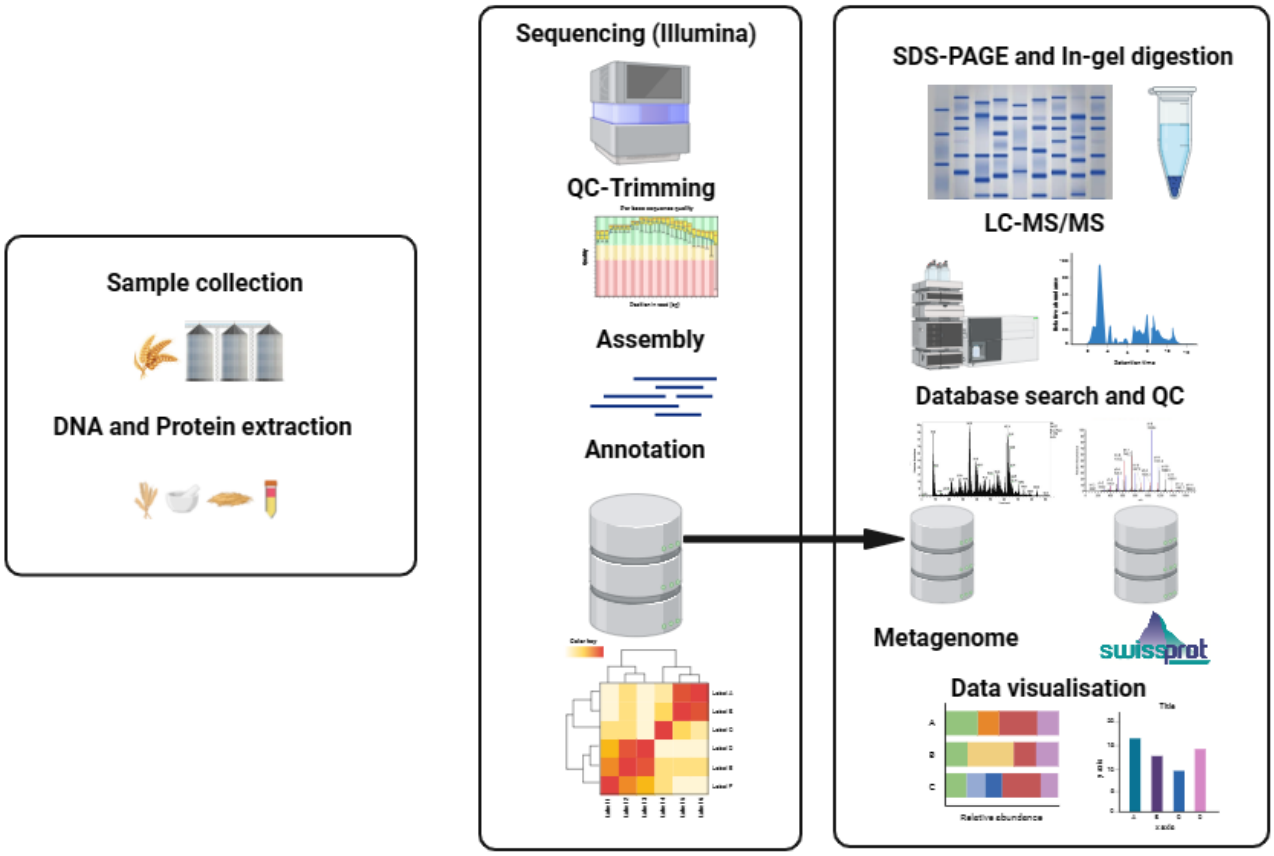


Figure 4.1 Methodology workflow of the meta-omics approach used in the study.

## 4.3 Methodology

### 4.3.1 Metagenomic analysis

#### 4.3.1.1 Sample collection, DNA extraction and metagenome sequencing

Barley seeds (Kadie cultivar) were collected at harvest and after three, six and nine months of storage at Anheuser-Busch InBev (AB InBev) storage facilities in the Western Cape Province, South Africa (Chapter 2; Section 2.3.1). At each time point barley seeds with high and low nitrogen content were collected (Chapter 2, Table S2.1). Subsequently, metagenomic DNA (mDNA) was extracted (Chapter 2; Section 2.3.2) and sequenced using the Illumina NovaSeq 6000 system with paired end configuration (2 × 150 bp) for 500 cycles (Chapter 2; Section

2.3.2). The resultant metagenomic sequence data were quality trimmed and curated (Chapter 2; Section 2.3.3).

#### 4.3.1.2 Metagenome assembly

The metagenomic data were assembled using MetaSPAdes version 3.15.3 (Nurk *et al.*, 2017). This tool has been developed specifically for the complex nature of metagenomes, which comprise genetic sequences from a diverse range of microorganisms found in environmental samples (Nurk *et al.*, 2017). MetaSPAdes utilises a de Bruijn graph algorithm as its foundation, starting with the rectification of errors in reads to improve accuracy (Compeau *et al.*, 2011). Subsequently, it merges overlapping reads into elongated sequences and then constructs de Bruijn graphs using various k-mer sizes (21, 33, and 55). This multi-k-mer strategy facilitates the assembly of genomic regions of varying complexity by accommodating different structures and coverage depths inherent in metagenomic datasets (Nurk *et al.*, 2017). The integrity and quality of the final assemblies was evaluated using QUAST v5.2.0 (Gurevich *et al.*, 2013).

#### 4.3.1.3 Structural and functional annotation

Structural annotation of the metagenomic assemblies was conducted using the Prokka v1.14.5 pipeline (Seemann, 2014). Prodigal identifies protein-coding sequences (Hyatt *et al.*, 2010). Concurrently, Aragorn is used for the prediction of tRNA genes and tmRNA, while Barrnap is employed for ribosomal RNA (rRNA) gene prediction (Laslett and Canback, 2004). phaseFor functional annotation, BLAST+ was used to match the predicted protein sequences against known proteins in the UniProtKB (SwissProt) database (UniProt Consortium, 2018). Additionally, HMMER was used to identify protein domains via hidden Markov models against the Pfam and TIGRFAM databases (Prakash *et al.*, 2017). Subsequently, the protein sequences generated by Prokka were processed using eggNOG-mapper v5.0, which employs the DIAMOND algorithm for rapid sequence alignment (Huerta-Cepas *et al.*, 2019) against the EggNOG database v.5.0 (Huerta-Cepas *et al.*, 2019).

While eggNOG-mapper was a core tool for functional annotation, it draws upon multiple underlying databases, ensuring comprehensive functional insights. These include the Clusters of Orthologous Groups (COGs) database (Tatusov *et al.*, 2003), KEGG Orthology (KO) terms (Ogata *et al.*, 1999), Pfam domains (Finn *et al.*, 2014), Gene Ontology (GO) terms (Harris *et al.*, 2004) to functionally annotate each protein. Eggnog-mapper also includes the dbCAN tool (Yin *et al.*, 2012) for the identification of carbohydrate-active enzymes (CAZymes). This

multi-database approach captures a broad array of functional attributes and avoids the limitations of relying on a single database, providing a robust and well-rounded annotation of the metagenomic data.

#### 4.3.1.4 Predicted coding sequences (CDS) coverage analysis

Paired-end reads were mapped to coding DNA sequences (CDS) within .ffn files, previously annotated using Prokka, by employing Bowtie2 v2.3.5.1 (Langmead and Salzberg, 2012). Coverage for each CDS was determined by computing the average depth of coverage, which involves dividing the accumulated per-base read depth by the total length of the CDS. This calculation was performed using the bedcov function of SAMtools v1.9 (Li *et al.*, 2009). The coverage data were then integrated with functional annotations obtained from eggnog-mapper (Huerta-Cepas *et al.*, 2019).

#### 4.3.1.5 Metabolic pathway reconstruction and metagenome mining

Genes known to enhance plant growth in barley and related seeds have been documented in previous studies (Akinola *et al.*, 2021; Robertson-Albertyn *et al.*, 2022). The presence of similar CDS was determined by manually searching within the functional annotation files. To reconstruct important metabolic pathways, the KO identifiers associated with key metabolic functions, such as amino acid utilisation, carbon degradation, fermentation, and nitrogen metabolism, were searched. These KO identifiers served as markers to locate and verify the corresponding genes in the functional annotation files, thus facilitating the reconstruction of relevant metabolic pathways based on the metagenomic data. The refined pathways were then redrawn to incorporate the additional information obtained from the genome mining process. The final visualisations of the nutrient cycling pathways were created using Microsoft PowerPoint.

### 4.3.2 Metaproteomic analyses

#### 4.3.2.1 Microbial cells separation

Approximately 10 g of barley seeds per sample were ground using a sterilized pestle and mortar. The powdered material was suspended in 40 ml of phosphate-buffered saline (PBS, pH 7.4). This solution was mixed by brief vortexing to ensure uniform consistency.

Subsequently, the solution was sonicated (Microson XL 200, United States of America) at 18 W, with a pattern of 30 s active and 30 s inactive, for a total of 7 min. After sonication, the suspension was centrifuged at 4,000 x g for 1 min. The clear supernatant was filtered through a 0.45 µm pore filter (Sartorius-Stedim Biotech).

#### 4.3.2.2 Protein Extraction

Microbial cells obtained from the filter paper were subjected to protein extraction using the sodium dodecyl sulphate (SDS) and acetone method, adhering to the protocol detailed by Hall and coworkers (2012). Initially, cells were resuspended in an extraction buffer containing 50 mM Tris (pH 7.5, adjusted with NaOH) and 1% (w/v) SDS. The mixture was subjected to vigorous vortexing for 2 min, sonication for 5 min on ice at 20% amplitude (Q125 Sonicator, Qsonica, Connecticut, USA), boiling for 15 min at 95 °C, and then gentle agitation on an orbital shaker (Reflecta Laboratories, South Africa) at 20 rpm for 1 h at 4 °C. Following centrifugation (1,000 x g for 5 min at 4°C), the supernatant, which contained the extracted proteins, was transferred into a sterile tube. Proteins in the supernatant were concentrated and purified through acetone precipitation, maintaining a 5:1 ratio of acetone to sample, using 1 ml of sample. The mixture was left to precipitate at −20°C overnight. The samples were centrifuged (1,000 x g for 1 min at 4°C), the supernatant was discarded, and the pellet was allowed to dry on ice for 10 min. The resulting pellets were redissolved in 50 mM Tris buffer (pH7.4), and a standard 30 µl volume of the protein solution was subjected to SDS-polyacrylamide gel electrophoresis (PAGE) as described by Laemmli (1970), using 12% (w/v) polyacrylamide gels stained with Coomassie Brilliant Blue (G-250, Bio-rad). Subsequently, the gels were sent to the Council for Scientific and Industrial Research (CSIR) in Pretoria, South Africa, for in-gel digestion and liquid chromatography-mass spectrometry (LC-MS) analysis.

#### 4.3.2.3 In-gel digestion

Protein lanes were divided into five separate sections based on band density and molecular weight range. From each sample, five gel bands were selected, excised and subjected to in-gel digestion, followed by analysis as described by Shevchenko and colleagues (2006). Proteins within the gel sections were reduced using 10 mM dithiothreitol (DTT) mixed with 25 mM of ammonium bicarbonate ($NH_4HCO_3$) and the mixture was maintained at 60°C for 1 h. After cooling to ambient temperature, gel sections were immersed in 100% acetonitrile for 10 min. After discarding the supernatant, the gel sections were exposed to 55 mM iodoacetamide

combined with 25 mM $NH_4HCO_3$. The reaction proceeded for 20 min under light-protected conditions. The supernatant was discarded, and the gel sections were dehydrated using a mixture of 25 mM $NH_4HCO_3$ and 50% (v/v) acetonitrile. Once completely dry, fresh trypsin solution (0.25% v/v) was introduced. Digestion was performed overnight at 37°C. The addition of 0.1% (v/v) formic acid halted digestion, after which the samples were dried under vacuum. The dried specimens were then prepared in a solution of 2% (v/v) acetonitrile and 0.2% (v/v) formic acid for liquid chromatography–mass spectrometry (LC-MS)

#### 4.3.2.4 Liquid Chromatography-Mass Spectrometry (LC-MS)

Peptide fragments derived from each gel section were analysed using a Dionex Ultimate 3000 RSLC system (ThermoFisher, USA) coupled with an AB Sciex 5600 TripleTOF mass spectrometer (Sciex, USA). The peptides were initially purified on an Acclaim PepMap C18 trap column (ThermoFisher, USA) and subsequently subjected to gradient elution on a Waters NanoEase MSH C18 column (ThermoFisher, USA). LC-MS analysis was conducted on the 5600 TripleTOF platform, employing a positive ion mode and data-dependent acquisition (DDA) mode. Precursor (MS) scans were collected from m/z 400-1500, followed by 40 fragment ion (MS/MS) scans, ranging from m/z 100-1800.

#### 4.3.2.5 Database generation

Database searching is a fundamental component of metaproteomic analysis (Saito *et al.*, 2019). It serves as the primary method for identifying proteins from the complex mixtures present in environmental samples. In a metaproteomic experiment, peptides and consequently the proteins from which they originate can be identified by matching the generated mass spectra against theoretical spectra derived from a protein sequence database (Heyer *et al.*, 2019). In this study, a customised database was generated for the metaproteomics analysis to enhance protein identification accuracy. The database incorporated protein sequences from the host organism, *Hordeum vulgare,* potential contaminants identified in-house (CSIR), and annotated sequences obtained earlier (Section 4.4.3). This ensured the database specificity to the samples, increasing the likelihood of accurate protein identification. A target-decoy approach was implemented to control the False Discovery Rate (FDR) (Aggarwal and Yadav, 2016). This is a statistical measure used to control the rate of Type I errors (false positives), ensuring the reliability of the results (Elias and Gygi, 2007). The total database consisted of 372,508 proteins (745,016 in the concatenated target-decoy protein database).

#### 4.3.2.6 Protein search

The raw LC-MS datasets (.wiff files) and generated database were analysed using the Sciex software Protein Pilot V5.010 (Seymour and Hunter, 2015). This software used the Paragon Database Search algorithm to identify peptides from the MS/MS spectra (Shilov *et al.*, 2007). Subsequently, the Pro Group algorithm was used to organise these identified peptides into a coherent lists of reliable peptide identifications. FDR analysis provided a confidence level for the results and a report was generated for each peptide identification result. The LC-MS datasets were segmented into ten parts for processing to enhance the sensitivity and coverage of the analysis. Each segment underwent two separate searches, culminating in a total of 20 sub-searches. A reversed search function was applied, which is a common strategy in proteomics used to estimate the FDR (Gupta *et al.*, 2011). The multi-precursor search function was enabled during the Fraglet and Taglet stages of mass spectrometry data analysis to accommodate different precursor masses, improving the identification of peptides, especially in complex mixtures (Shilov *et al.*, 2007). When the precursor charge state (z) was undetermined, the first three charge states were considered, a common practice in proteomics to account for the fact that peptides can carry multiple charges, increasing the likelihood of correctly identifying the peptide (Sharma *et al.*, 2010).

#### 4.3.2.7 Protein quantification and grouping

Protein grouping was conducted using Bayesian protein confidence calculations alongside peptide FDR analysis (Seymour and Hunter, 2015). The criteria for grouping proteins were based on Intersection ProtScore and Unused ProtScore values. In Protein Pilot, ProtScore is a scoring metric whereby the total ProtScore aggregates all peptide evidence associated with a specific protein and the Unused ProtScore represents the total unique peptide evidence for that protein (Seymour and Hunter, 2015). Protein-level ratios were derived from the median of peptide ratios, mitigating the influence of extreme values or outliers (Maia *et al.*, 2020). For peptide quantification, filtering based on signal-to-noise ratios and peptide confidence thresholds was used to enhance the accuracy of peptide identifications (Cox *et al.*, 2014). The signal-to-noise ratio contrasts the level of desired signal against background noise, with a higher ratio indicating minimal background noise (Maia *et al.*, 2020). The peptide confidence threshold serves as a cut-off value, beyond which peptides are deemed confidently identified. Additionally, specific considerations for discordant and shared peptides were applied to decide their inclusion or exclusion in the analysis. Discordant peptides, which deviate from the

majority, and shared peptides, common to several proteins, were evaluated through this logic to ensure the inclusion of only the most reliable peptide identifications in the final analysis.

#### 4.3.2.8 Protein data filtering

The range of spectral counts found in the eight samples varied from 17,702 to 29,299 (Table S4.2). To control false positive identifications, we used a protein FDR of 1 % (Seymour and Hunter, 2015). Protein hits obtained from the databases were filtered based on the criterion that their Unused ProtScore satisfied the 1 % FDR. This was followed by host and contaminant removal. In addition to filtering at 1 % FDR and host removal, only those data points that exhibited a confidence score of ≥ 95 %, coupled with a contribution score of ≥ 0.1 (10 %), were retained. Subsequently, the curated peptides belonging to the identified proteins were parsed through the Unipept web application (https://unipept.ugent.be/) for protein search and functional annotation (Verschaffelt *et al.*, 2021).

#### 4.3.2.9 Functional annotation and taxonomy of peptides

Functional annotation of peptides and their taxonomic classification involved several steps. Peptides from the proteins identified by Protein Pilot V5.010 (Seymour and Hunter, 2015) were further analysed using the Unipept application (Verschaffelt *et al.*, 2021). This platform receives peptide sequences, typically from metaproteomic experiments, and provides both taxonomic and functional annotations. For taxonomic classification, Unipept uses the Lowest Common Ancestor (LCA) algorithm, which identifies the deepest node, or the lowest common ancestor, in the taxonomic tree that includes all organisms sharing a specific peptide. This process draws on taxonomic information from the NCBI taxonomy database (http://www.ncbi.nlm.nih.gov/taxonomy). For functional annotation, Unipept uses data from multiple databases, including Enzyme Commission (E.C.) numbers, GO terms, and InterPro entries (Harris *et al.*, 2004; Jones *et al.*, 2014). This data helps elucidate the functions of the peptides and the roles of the organisms in a specific environmental context.

## 4.4 Results and discussion

### 4.4.1 Metagenome assembly metrics

The metagenomes of the eight barley seed samples in this study were assembled using MetaSpades (Nurk *et al.*, 2017). Assembly sizes of samples ranged from 78 to 180 megabases (Mb). The largest assembly was observed in the high nitrogen content barley seeds sampled at harvest (HH), whereas the smallest assembly was obtained from the low nitrogen content grain maintained in storage for nine months (S9L) (Table 4.1). The assembly sizes correlate with the number of reads obtained from each sample (Chapter 2; Table 2.1) and the number of MAGs recovered per sample (Chapter 3; Table 3.1). The observed trend of decreasing assembly size with increased storage time raises intriguing questions about the stability and longevity of microbial communities in stored grain (Wisnoski and Lennon, 2021). It is possible that microbial population is initially diverse but reduces as storage conditions become less favourable or as microbial competition and succession lead to the dominance of a few adapted species (Dutta *et al.*, 2022).

Metagenome assembly produced a total of 274,778 contigs across all samples, averaging 34,347 contigs per sample (Table 4.1). The highest count was observed in sample S6L with 36,111 contigs, while the lowest was in sample S9L with 17,032 contigs. The N50 values ranged from 137,89 bp in sample S3H to 468,6 bp in sample S6L (Table 4.1). The N50 value is a statistical metric that indicates the median length of the contigs, where the shortest contig equals half of the overall genome length (Parra *et al.*, 2009). Higher N50 values suggest a higher-quality assembly characterised by longer contigs (Parra *et al.*, 2009; Wang and Wang, 2023). All samples exhibited N50 values exceeding 1,000 bp, indicative of robust assembly quality, a result consistent with the superior outcomes reported using SPAdes for complex microbial communities (van der Walt *et al.*, 2017).

The number of contigs per sample does not necessarily correlate with the assembly sizes (Papudeshi *et al.*, 2017). While a high number of contigs can indicate extensive coverage of a metagenome, it can also imply fragmentation during the assembly process (Papudeshi *et al.*, 2017; van der Walt *et al.*, 2017). Notably, sample S6L, which had the highest number of contigs (36,111), also demonstrated the lowest N50 value (Give the value), suggesting a fragmented assembly, possibly due to the presence of many unique, short sequences that could not be fully assembled (van der Walt *et al.*, 2017). This fragmentation may also explain the lower number

of MAGs recovered from sample S6L (Chapter 3). This highlights the bioinformatic challenges faced in metagenomic sequencing, particularly in complex environmental samples where genomic overlap among different species can complicate the accurate assembly of complete genomes (Berg *et al.*, 2020).

Samples with a high nitrogen content generally exhibited larger assembly sizes, with the exception of the barley seeds stored for six months (S6H and S6L) (Table 4.1). A higher nitrogen content within the seeds likely enhances the availability of nitrogen for microorganisms residing in or colonising the seed (Liu *et al.*, 2023). As these microorganisms consume nitrogen, they proliferate, thereby increasing the overall microbial biomass (Arduini *et al.*, 2006). This increase may be directly reflected in the amount of microbial DNA present, contributing to larger metagenomic assembly sizes upon sequencing of the seed microbiome. The six months storage samples did, however, display signs of dysbiosis (Chapter 2; Section 2.4.1), which could explain the deviation observed in these samples.

Table 4.1 Metagenome assembly and structural annotation metrics.

| Sample | **HH** | **HL** | **S3H** | **S3L** | **S6H** | **S6L** | **S9H** | **S9L** |
|---|---|---|---|---|---|---|---|---|
| Grain nitrogen content | High | Low | High | Low | High | Low | High | Low |
| Storage duration | Harvest | Harvest | 3-months | 3-months | 6-months | 6-months | 9-months | 9-months |
| #contigs | 26,069 | 23,100 | 23,526 | 17,737 | 13,574 | 36,111 | 20,368 | 17,032 |
| Total_length (Mb) | 159.01 | 129.63 | 146.25 | 109.46 | 80.11 | 134.23 | 108.08 | 78.17 |
| GC (%) | 54.06 | 56.24 | 57.92 | 52.52 | 53.91 | 53.98 | 54.44 | 55.68 |
| N50 | 130,49 | 118,78 | 137,89 | 137,88 | 100,21 | 468,6 | 921,5 | 724,1 |
| CDS | 74,436 | 151,214 | 169,009 | 75,592 | 103,736 | 136,775 | 135,364 | 106,825 |
| rRNA | 36 | 28 | 42 | 15 | 17 | 27 | 19 | 21 |
| tRNA | 720 | 1,273 | 1,138 | 357 | 940 | 1,003 | 1,028 | 1,067 |
| mRNA | 17 | 21 | 12 | 13 | 16 | 17 | 11 | 6 |
| eggNOGs-query | 135,443 | 124,932 | 85,980 | 116,142 | 67,085 | 108,812 | 87,574 | 62,211 |
| COG_category | 120,401 | 111,592 | 76,654 | 103,974 | 61,441 | 97,379 | 79,196 | 56,316 |
| KEGG_ko | 71,134 | 66,609 | 46,162 | 61,747 | 41,171 | 61,105 | 50,710 | 35,874 |
| CAZy | 1726 | 1539 | 1084 | 1454 | 1015 | 1525 | 1123 | 820 |

### 4.4.2 Structural and functional annotation metrics

Structural annotation using Prokka identified a total of 952,951 CDS, averaging 119,119 CDS per sample (Table 4.1). The number of CDS varied among the samples, with the highest count of 169,009 CDS observed in sample HH and the lowest count of 75,592 CDS in sample S6H (Table 4.1). Notably, the distribution of CDS across samples did not consistently align with metagenome size or contig counts; for example, sample S9L, which had one of the smallest metagenome sizes and lowest contig counts, displayed one of the highest numbers of CDS (Table 4.1). The number of CDS might not correspond to the assembly size or the number of contigs because the assembly size contains both coding and non-coding regions and contigs may contain several genes or incomplete genes, resulting in multiple CDS (Papudeshi *et al.*, 2017).

The number of CDS also appeared to vary with nitrogen content and storage duration (Table 4.1). At harvest, the high nitrogen content samples contained a markedly lower number of CDS (74,436) compared to the other samples. However, after three months of storage, the CDS count in the high nitrogen content samples surged to 169,009, the highest across all samples, before slightly decreasing at six months and then increasing again at nine months. In contrast, the low nitrogen content samples exhibited a different trend, starting with a high CDS count at harvest (151,214), which then dramatically dropped to 75,592 after three months of storage, increased again at six months and remained relatively stable at nine months of storage. The variation in CDS numbers across different storage times and nitrogen content suggests potential adaptations of the microbial communities to the environmental conditions within the barley seeds (Agu and Palmer, 2001).

### 4.4.3 Functional annotation of the metagenome assemblies

#### 4.4.3.1 Clusters of Orthologous Groups (COG) functions

COGs are collections of homologous proteins that are used to study evolutionary relationships (Tatusov *et al.*, 2003). Each COG consists of individual proteins or groups of paralogues that are assumed to have evolved from an ancestral protein. This assumption allows for the functional annotation of uncharacterised proteins based on their evolutionary relationships, thus providing a robust framework for exploring the genetic potential and functional diversity of microbial communities in metagenomic studies (Tatusov *et al.*, 2003; Tatusov *et al.*, 1997). Functional annotation of CDS using the eggNOG mapper tool (Huerta-Cepas *et al.*, 2019)

identified 706,953 COGs across all samples (Table S4.2). The distribution of these COGs showed variation in the number of COGs based on storage time and grain nitrogen content (Table S4.2). For instance, the high nitrogen harvest sample (HH) exhibited the highest count at 120,401 COGs, while sample S9L recorded the lowest at 56,316 COGs.

The protein coding sequences belonged to a total of 23 distinct COG categories (Table S4.2). The COG category function unknown (S) was the most represented, comprising approximately 21 % (167,146 entries) of the proteins with a COG category assignment (Fig. 4.2; Table S4.2). Metagenomic studies often uncover genes/proteins of unknown function from uncultured microorganisms that are underrepresented in current databases (Lok, 2015).

In terms of COGs with known functions, the COG category with the largest number of CDS was transcription (K; 67,581 total entries), closely followed by the category for amino acid transport and metabolism (E; 60,147 total entries) (Fig. 4.2; Table S4.2). Transcription is a fundamental biological process essential for all living organisms, as it initiates the pathway of gene expression by converting genetic information from DNA into RNA (Mejía-Almonte *et al.*, 2020). Similarly, the category for amino acid transport and metabolism includes genes involved in the uptake, transport, synthesis, and breakdown of amino acids; these are the building blocks of proteins (Marin and Krämer, 2007).

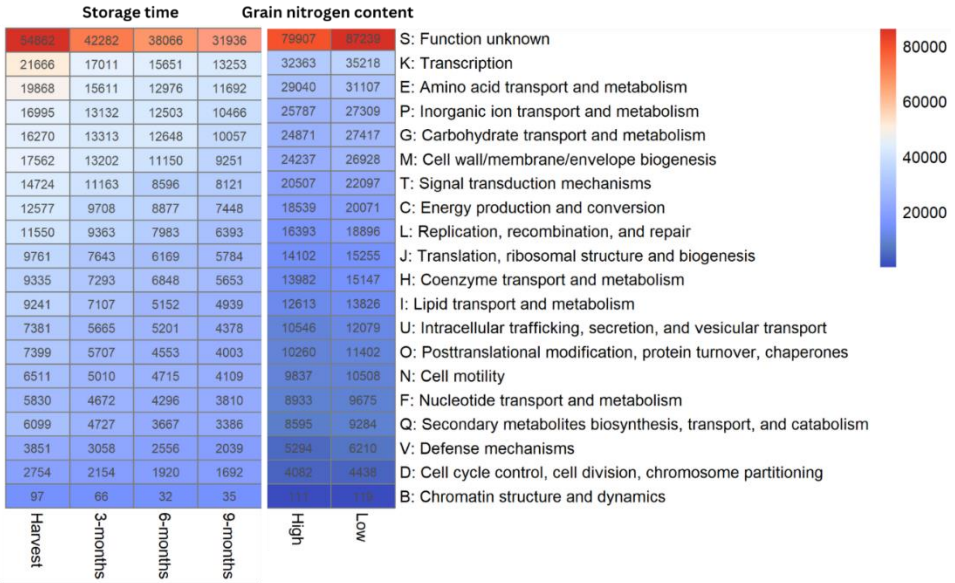


Figure 4.2 Heatmap illustrating the distribution of Clusters of Orthologous Groups (COGs) (Tatusov *et al.*, 2003) across different variables. (Left) Distribution by storage times: harvest, three months, six months, and nine months. (Right) Distribution by grain nitrogen content: high and low.

The third most populous COG category was inorganic ion transport and metabolism (P; 53,096 total entries) (Fig. 4.2). This category includes proteins that manage the uptake and utilisation of inorganic ions, which are vital for various cellular processes such as osmoregulation, electron transport and nutrient assimilation (Shabala and Shabala, 2011). Effective ion management can help microorganisms maintain cellular balance and energy production(Chandel *et al.*, 2021). This was followed by the COG category carbohydrate transport and metabolism (G; 52,288 total entries) (Fig. 4.2). This category encompasses enzymes and transport systems involved in the degradation and synthesis of carbohydrates, which are essential for energy production and storage (Biely, 2012). The presence of these proteins suggests that bacteria may impact the carbohydrate economy of the seed environment, potentially influencing seed decay or preservation during storage (Boraston *et al.*, 2004).

When grouped based on storage time and grain nitrogen content, it was observed that the number of COGs per category decreased with increasing storage duration (Fig. 4.2). The decline in COGs could suggest a simplification of the microbial community as the storage period extends (Chandel *et al.*, 2021). This trend may also be linked to the initially lower read counts obtained, making recovery less likely, particularly given that the highest number of CDS was observed at harvest and after three months of storage, indicating a potential decline as storage progresses.

#### 4.4.3.2 Distribution of CAZymes and carbon degradation proteins across the metagenomic samples

Carbohydrate-Active Enzymes (CAZymes) play a role in the formation and breakdown of complex carbohydrates and glycoconjugates (Cantarel *et al.*, 2009). These enzymes are fundamental in plant-microbial interactions, facilitating the degradation of complex plant biomass into simpler components (Deveau *et al.*, 2008). As such, analysis of the CAZyme composition can provide insight into the microbiome-barley seed interactions.

Across all samples, a total of 10,286 CAZymes were identified, spanning six enzyme classes and 90 families (Table S4.3). The highest number of CAZymes was observed at harvest in the high nitrogen content barley seeds (HH) and the lowest in sample S9L, with 1,726 and 820 representatives, respectively. Overall, a decrease in the total number of CAZymes correlates with extended storage times (Fig.4.3A and B). Additionally, samples with lower nitrogen content consistently exhibited higher counts of CAZymes (Fig.4.3A and B).

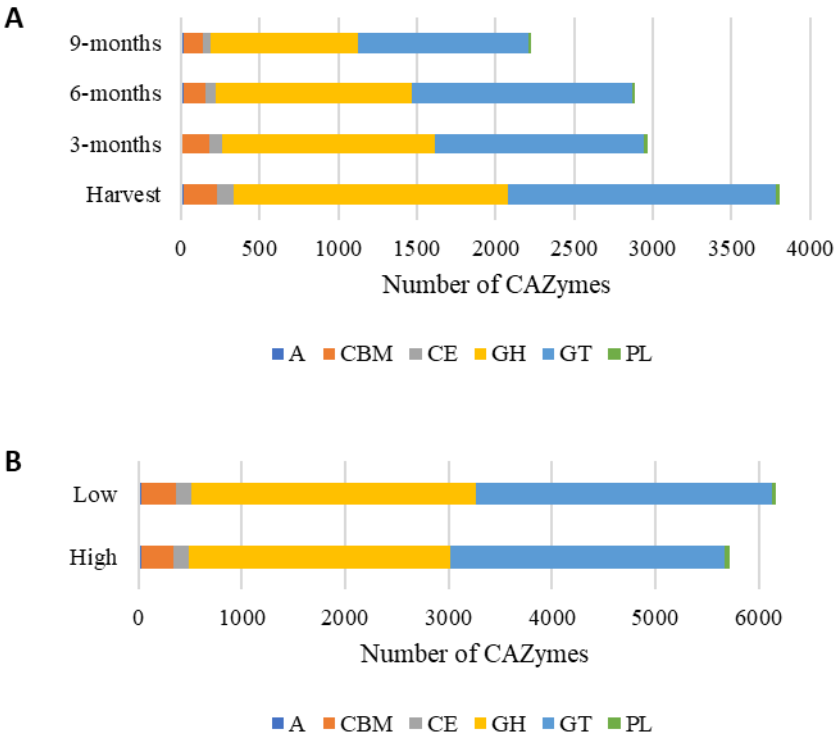


Figure 4.3 Distribution of CAZymes families (Park *et al.*, 2018) in barley seed microbiome metagenomes. The bar charts illustrate the quantitative representation of different CAZyme class: Auxiliary Activities (AA), Carbohydrate-Binding Modules (CBM), Carbohydrate Esterases (CE), Glycoside Hydrolases (GH), Glycosyl Transferases (GT), Polysaccharide Lyases (PL). (A) Distribution across storage times (harvest, three, six, and nine months). (B) Distribution by grain nitrogen content (high and low).

More than half (56.6%) of the identified CAZymes belong to the enzyme class glycosyl transferases (GTs). Within this class, the GT2 family was the most abundant (1,580 total entries), followed by the GT4 family (1,130 total entries) (Table S4.3). Glycosyl transferases are involved in the biosynthesis of glycosidic bonds (Zhu *et al.*, 2013). The GT2 family includes enzymes such as cellulose synthase, which synthesises beta-glucan chains integral to microbial biofilms and extracellular matrices (Schmid *et al.*, 2016). The GT4 family consists largely of glycogen synthases, required for microbial glycogen biosynthesis and energy management (Zhu *et al.*, 2013).

Glycoside hydrolases (GH) form the second-largest CAZyme class, encompassing 5,274 CAZymes, with the GH13 family accounting for 1,246 CAZymes (Fig. 4.4; Table S4.3). This family, often referred to as the alpha-amylase family, includes enzymes that break down starch into simpler sugars such as maltose and glucose (Stam *et al.*, 2006; Franceus and Desmet, 2020). This suggests the microbiome integrates strategies for utilising starch and other energy-

rich compounds within the seeds, which are needed for microbial energy acquisition, but could reduce the starch reserves essential for malting in beer production (Parés Viader *et al.*, 2021). Metagenome mining of glycosyl hydrolases also identified beta-glucosidases (BglX, BglB), which are pivotal for cellulose degradation, and beta-galactosidases (LacZ), which facilitate the conversion of beta-galactosides into monosaccharides, aiding in the metabolism of plant-derived sugars (Rosenberger *et al.*, 2012). Enzymes of the classes Auxiliary activities (AA - primarily involved in lignin degradation), carbohydrate esterases (CE - involved in modifying esterified plant cell components) and polysaccharide lyases (PL - degrade pectin) were found in lower abundance in the barley seed metagenomes. This may be due to the less complex barley seed cell wall, which further has negligible amounts of pectin and lignin (Shafiei *et al.*, 2023; Sidar *et al.*, 2020; Langenaeken *et al.*, 2020).

Overall, analysis of the CAZYmes highlights a robust system for carbohydrate manipulation. These enzymes could prepare the seeds for more efficient malting by breaking down complex carbohydrates into simpler sugars required for fermentation (Parés Viader *et al.*, 2021).

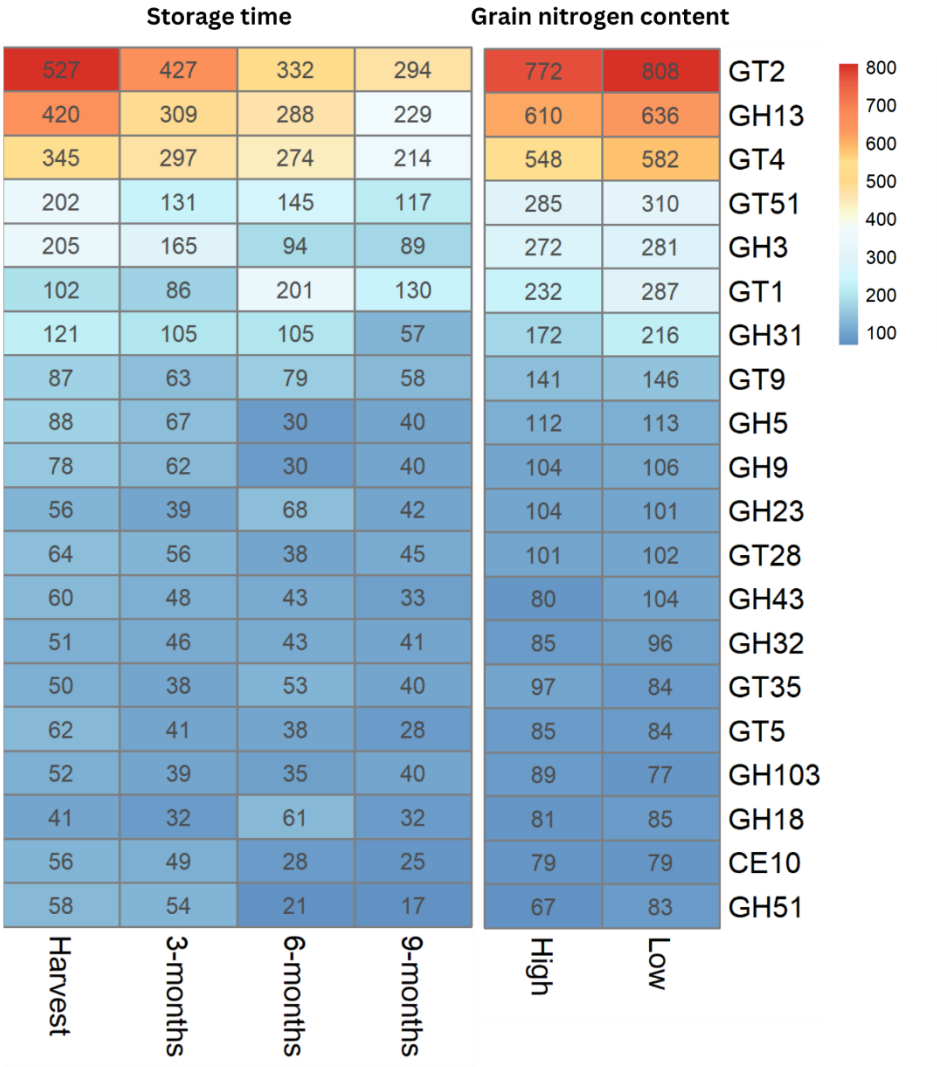


Figure 4.4 Comparative analysis of Carbohydrate-Active enZYmes (CAZymes) (Park *et al.*, 2018) family abundance in barley seed metagenomes. The heatmaps display the abundance counts of various CAZyme subfamilies. (Left) Distribution across storage times (harvest, 3, 6, and 9 months). (Right) Distribution by grain nitrogen content (high and low). Carbohydrate Esterases (CE), Glycoside Hydrolases (GH) and Glycosyl Transferases (GT).

### 4.4.4 Reconstruction of metabolic pathways from the barley microbiome

Reconstruction of the metabolic pathways of seed microbiomes facilitates an understanding of the complex interactions between plants and their associated microbial communities (Zhou *et al.*, 2022). This section details metabolic pathways related to amino acids, carbohydrate degradation, fermentation, and nitrogen (Table S4.4). Examining these pathways elucidates how seed storage conditions and nitrogen content may influence the metabolic activities of the seed microbiome. Additionally, it provides insights into how these metabolic processes contribute to the nutritional quality of barley seed (Bziuk *et al.*, 2021; Rahman *et al.*, 2018).

#### 4.4.4.1 Survey of proteins involved in fermentation

Fermentation processes in the barley seed microbiome may support microbial survival and activity under anaerobic or low-oxygen conditions (Bokulich and Mills, 2012). These conditions often occur in environments where seeds are densely packed, limiting the availability of oxygen (Chandel *et al.*, 2021). Under such circumstances, traditional aerobic metabolic pathways are not feasible, prompting microorganisms to rely on anaerobic respiration or fermentation for energy production (Chandel *et al.*, 2021; Hirko *et al.*, 2023). Additionally, the lactate fermentation pathway, facilitated by enzymes such as L-lactate dehydrogenase, becomes particularly advantageous in these anaerobic settings, helping microorganisms efficiently produce energy while contributing to the acidity and potential flavour profiles of the grains used in brewing(Gaspar *et al.*, 2007; Laitila *et al.*, 2018).

Metagenomic mining of the barley seed microbiome identified key proteins involved in fermentation, such as acetyl-CoA synthetase (AcS) and formate C-acetyltransferase (PflD) (Table S4.4). Acetyl-CoA synthetase converts acetate into acetyl-CoA, thus enhancing the utilisation of acetate and influencing the carbon cycle in the seed environment (Krivoruchko *et al.*, 2015). Formate C-acetyltransferase catalyses the reversible transformation of pyruvate and CoA into acetyl-CoA and formate (Bar-Even, 2016). The presence of these enzymes indicates a microbiome adapted to survive under anaerobic or low-oxygen conditions (Chandel *et al.*, 2021).

Additionally, orthologues of L-lactate dehydrogenase (Idh) were observed in all samples, with particularly high read coverage depth in samples S6H and S9H (Table S4.4). This enzyme catalyses the conversion of pyruvate to lactate in the lactate fermentation pathway (Gaspar *et al.*, 2007). Notably, samples S6H and S9H also showed a high abundance of reads assigned to

the genera *Lactococcus* and *Enterococcus* (Chapter 2; Section 2.4.5) that are known for their ability to produce lactate via lactate dehydrogenase (Song *et al.*, 2017; Cruz *et al.*, 2013).

#### 4.4.4.2 Survey of proteins involved in amino acid metabolism

Amino acids are an essential source of nitrogen, vital for growth of almost all living organisms (Reineke and Schlömann, 2023). Consequently, the metabolism of amino acids by the seed microbiome could impact the nitrogen cycle within the plant-microorganism ecosystem (Jetten, 2008). Metagenomic analysis identified several proteins involved in amino acid metabolism, primarily aminotransferases/transaminases (Table S4.4). These enzymes catalyse the transfer of an amino group from a donor to a recipient molecule, thereby transforming amino acids into other forms usable by the plant or its associated microorganisms (Givan, 1980; Dold *et al.*, 2016). These enzymes included histidinol-phosphate aminotransferase (HisC), involved in histidine biosynthesis; phosphoserine aminotransferase (SerC), which converts 3-phosphohydroxypyruvate to 3-phosphoserine; and branched-chain amino acid aminotransferase (IlvE), required for the synthesis and degradation of leucine, isoleucine and valine (Nelson, 2018; Egle *et al.*, 2015; Rudat *et al.*, 2012). These activities may facilitate the conversion of seed nitrogen into forms that are more accessible for use by both the microorganisms and the plant (Marin and Krämer, 2007).

#### 4.4.4.3 Survey of proteins involved in nitrogen metabolism

Metagenomic analysis of the barley seed microbiome identified a range of nitrogen cycling processes. The reconstruction of nitrogen cycling within the barley seed microbiome revealed that key processes were present in all metagenome samples. These processes included assimilatory and dissimilatory nitrate reduction to ammonium, denitrification, nitrite reduction to ammonium, while nitrogen fixation was restricted to only two barley metagenome samples (Table S4.4).

Nitrite ammonification transforms nitrite into ammonia, a less toxic form, and nitrate reduction converts nitrate into nitrite, both of which are essential for nitrogen assimilation in microbial metabolism (Egas *et al.*, 2024; Tiso and Schechter, 2015). The ubiquity of these pathways could suggest a fundamental role in enabling the microbial community to manage nitrogen compounds effectively. Denitrification is primarily an anaerobic process where microorganisms reduce nitrate to gaseous nitrogen forms, functioning as an alternative respiratory pathway when oxygen is scarce (Zumft, 1997). Silos and storage facilities often create low-oxygen environments due to dense packing and limited air circulation (Ghali, 2014).

The ubiquity of these pathways could suggest a fundamental role in enabling the microbial community to manage nitrogen compounds effectively.

Nitrogen fixation refers to the biological conversion of atmospheric nitrogen into a more reactive form that can be used biologically, typically ammonia (Table S4.4) (Burris and Roberts, 1993). Proteins related to nitrogen fixation (NifD, NifH, and NifK) were observed in only two samples, namely S6L and S9H (Table S4.4). The appearance of nitrogen fixation genes later in the storage period could result from the observed succession in the microbial community (Chapter 2; Section 2.4.7). As storage conditions change, the environment may become more favourable for nitrogen-fixing bacteria to proliferate (Jetten, 2008; Fani *et al.*, 2000).

### 4.4.5 Metaproteomic analysis

#### 4.4.5.1 General metrics

A total of 3,267 proteins were identified across all barley seed samples at a 1% FDR, with an average of 408 proteins per sample (Table S4.5). The number of proteins identified is lower than the CDS obtained from the metagenome of the same samples (Fig. 4.5; Table S4.5). This discrepancy can be attributed to the fact that not all genes are expressed simultaneously or at the same level (Desai *et al.*, 2012). Some genes may not be expressed under the specific conditions at the time of sampling, leading to no detectable proteins (Salvato *et al.*, 2021). Additionally, challenges in protein identification in metaproteomics contribute to underreporting (Saito *et al.*, 2019). As an emerging field, metaproteomics faces hurdles such as the complexity of the sample, the presence of low-abundance proteins and the limitations of analytical methods, all of which can reduce protein identification rates (Abraham *et al.*, 2014; Saito *et al.*, 2019). Secreted proteins were underrepresented in the data, likely due to losses incurred during protein isolation steps that may fail to effectively capture or preserve them. Given their less stable and more soluble characteristics, these proteins are prone to degradation or loss throughout the extraction and purification processes (Saito *et al.*, 2019).

The highest number of proteins (604 proteins) were identified in sample S9H (Fig. 4.5; Table S4.5). This sample also recorded a high number of CDS (135,364). Conversely, sample S6L had the lowest number of identified proteins (93), despite having one of the highest numbers of CDS (136,775). The number of proteins per storage period does not follow a clear ascending

or descending trend; instead, it fluctuates. The samples stored for nine months consistently show the highest counts of identified proteins in the dataset (Figure 4.5). This suggests that as storage time progresses, the microbial community likely undergoes changes in response to storage conditions (Mordant and Kleiner, 2021), which could lead to variations in the types and quantities of proteins produced.

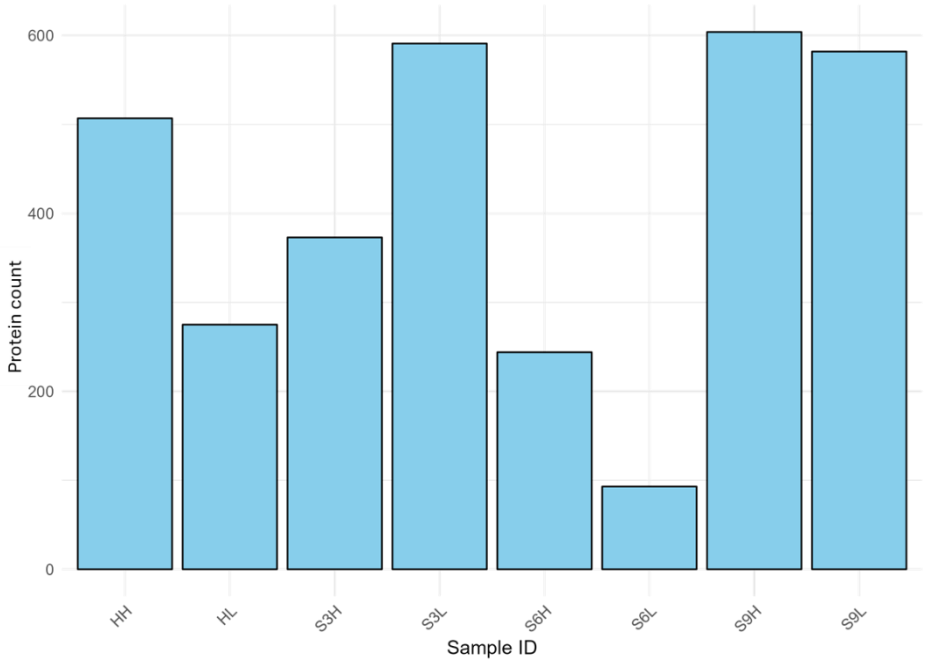


Figure 4.5 Protein count of metaproteome samples identified using Protein Pilot (Seymour and Hunter, 2015) and Unipept (Verschaffelt *et al.*, 2021). HH: Harvest, High grain nitrogen content; HL: Harvest, Low grain nitrogen content; S3H: 3 months storage, High grain nitrogen content; S3L: 3 months storage, Low grain nitrogen content; S6H: 6 months storage, High grain nitrogen content; S6L: 6 months storage, Low grain nitrogen content; S9H: 9 months storage, High grain nitrogen content; S9L: 9 months storage, Low grain nitrogen content.

#### 4.4.5.2 Taxonomic classification of peptides

Taxonomic classification of peptides against the NCBI database via Unipept revealed that the proteins belong to four bacterial phyla, namely the Pseudomonata, Bacteroidota, Bacillota, and Actinomycetota, as well as the viral phylum Uroviricota (Fig. 4.6). Seed-associated bacteria across a wide range of plant taxa predominantly belong to the bacterial phyla Actinobacteria, Bacteroidetes, Firmicutes and Proteobacteria, as documented in several studies (Bulgarelli *et al.*, 2015; Bziuk *et al.*, 2021; Rahman *et al.*, 2018). The dominance of these phyla in soil likely explains their frequent encounters with seeds during development and post-harvest periods (Fierer *et al.*, 2012).

The phylum Pseudomonata accounted for 92% of the total peptides identified (Fig 4.6), aligning with the taxonomy of metagenomic reads (Chapter 2; Section 2.4.1), which suggests not only their abundance but also the substantial activity linked to this phylum within the barley seed microbiome (Bziuk *et al.*, 2021). Archaea and fungi were detected in the metagenomic data in very low proportions but were not observed in the protein data. This absence might indicate that their protein concentrations are too low for detection, given the sensitivity of protein extraction methods (Abraham *et al.*, 2014).

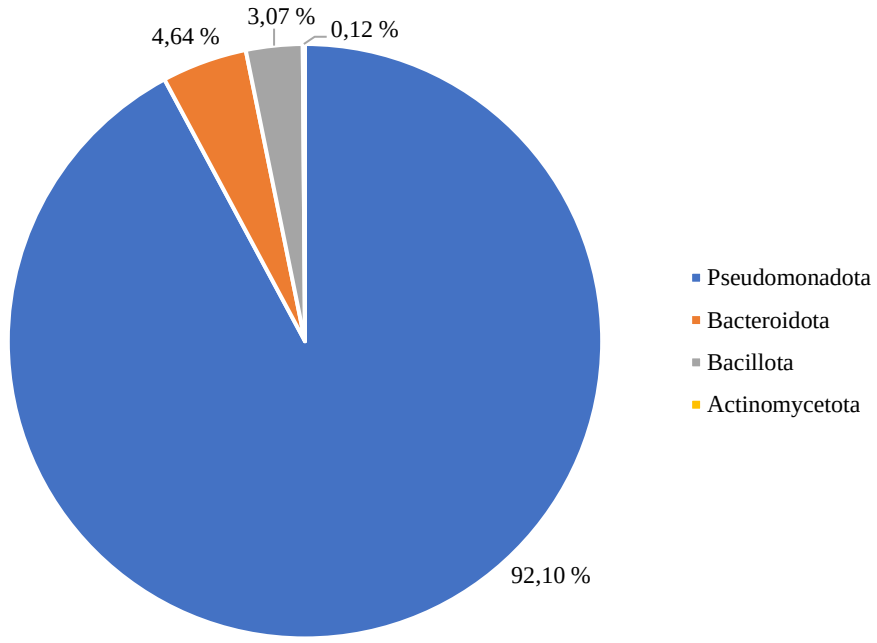


Figure 4.6 Proportional distribution of proteins, categorised by microbial phyla, based on identifications from the InterPro protein database (Jones *et al.*, 2014).

The metaproteomic analysis identified bacterial proteins associated with 22 genera (Fig. 4.7). Proteins from the genus *Pseudomonas* were the most prevalent, comprising 45% of the total protein count, followed by *Pantoea* (16%), *Xanthomonas* (14%) and *Duffyella* (11%). These genera were consistently identified in both the metagenomic reads and account for the majority of the assembled MAGS (chapter 2 and 3 respectively) Previous studies using RNA and DNA analyses have noted the frequent occurrence of *Paenibacillus*, *Pantoea*, *Pseudomonas* and *Stenotrophomonas* within the barley seed microbiome (Yang *et al.*, 2017; Rahman *et al.*, 2018; Bziuk *et al.*, 2021). The identification of proteins related to these bacterial genera cements their

status as core components of the barley seed microbiome. Additionally, given that the proteomic analysis reflects active protein expression, it suggests that they play a role in the functioning of the microbiome and its interaction with the barley seed (Erickson *et al.*, 2012). Notably, proteins from other prevalent genera identified from the metagenomic data (Chapter 2; Section 2.4.4), such as *Citrobacter*, *Klebsiella* and *Leclercia*, were not detected in the proteomic data, suggesting that these may be less active participants in the seed microbiome.

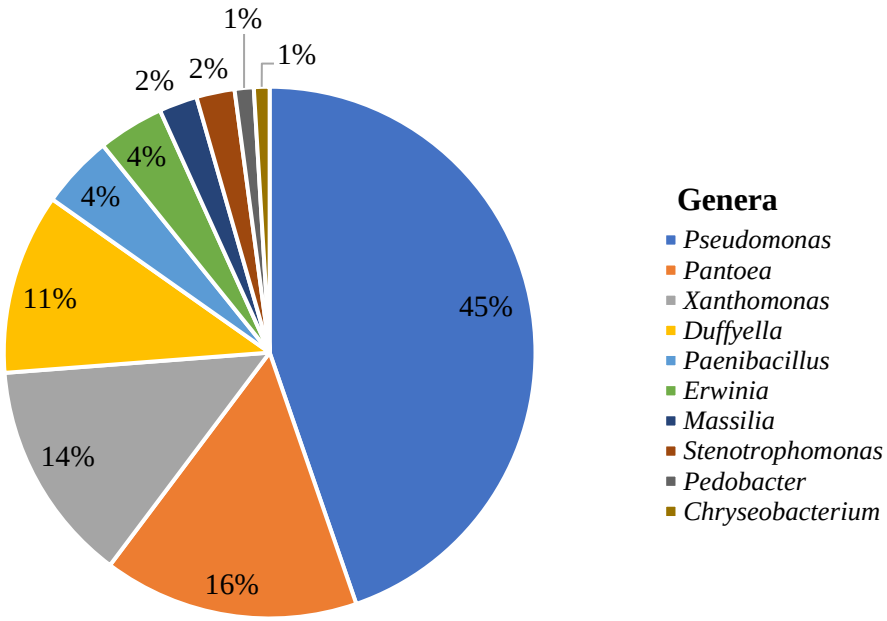


Figure 4.7 Distribution of the top ten microbial genera identified based on InterPro protein identifications (Jones *et al.*, 2014).

### 4.4.5.3 Distribution of proteins of the dominant genera based on storage time

The taxonomic profiling of proteins from barley seeds at different storage durations reveals deviations from the metagenomic data (Fig. 4.8). At harvest, proteins predominantly linked to *Pseudomonas* (33%), *Pantoea* (22%), and *Erwinia* (11%) were observed (Fig. 4.8). As storage progressed, a shift in the dominant genera was noted; after three months of storage, *Xanthomonas* proteins were most abundant (40%), surpassing *Pseudomonas* (25%) and *Pantoea* (15%) (Fig. 4.8). After six months of storage, *Pseudomonas* (48%) regained dominance, followed closely by *Pantoea* (28%) and *Paenibacillus* (14%). At nine months of storage, *Pseudomonas* proteins further increased in prevalence (69%), followed by *Duffyella* (13%) and *Pantoea* (9%) (Fig. 4.8).

These observations confirm that the barley seed microbiome undergoes changes over time (Felšöciová *et al.*, 2020). Additionally, it highlights disparities between the taxonomic abundance observed in metagenomic data and metaproteomic data (Fig. 4.8). For example, while metagenomic data at harvest highlighted that the genus *Stenotrophomonas* was prevalent, proteomic data indicated members of the genus *Pantoea* and *Pseudomonas* as the most active. Similarly, while *Xanthomonas* showed a higher protein count at the three month storage point, it had a lower read count in metagenomic data, suggesting varied microbial activities. This disparity continues at six and nine months, where the proteomic data show *Pseudomonas* (48, 69%, respectively), *Pantoea* (28, 9%, respectively), and *Paenibacillus* (14, 14%, respectively) as the more active genera, contrasting with the metagenomic prevalence of *Erwinia* and *Stenotrophomonas* (Chapter 2; Section 2.4.4). These findings underscore the differences between genomic and proteomic data in detecting taxa presence within a community. The observed differences suggest that proteomic data more directly reflects microbial activity, whereas metagenomic data primarily indicates the presence of taxa (Schiebenhoefer *et al.*, 2019). This distinction could result in varied conclusions regarding the functional impact of the microbiome.

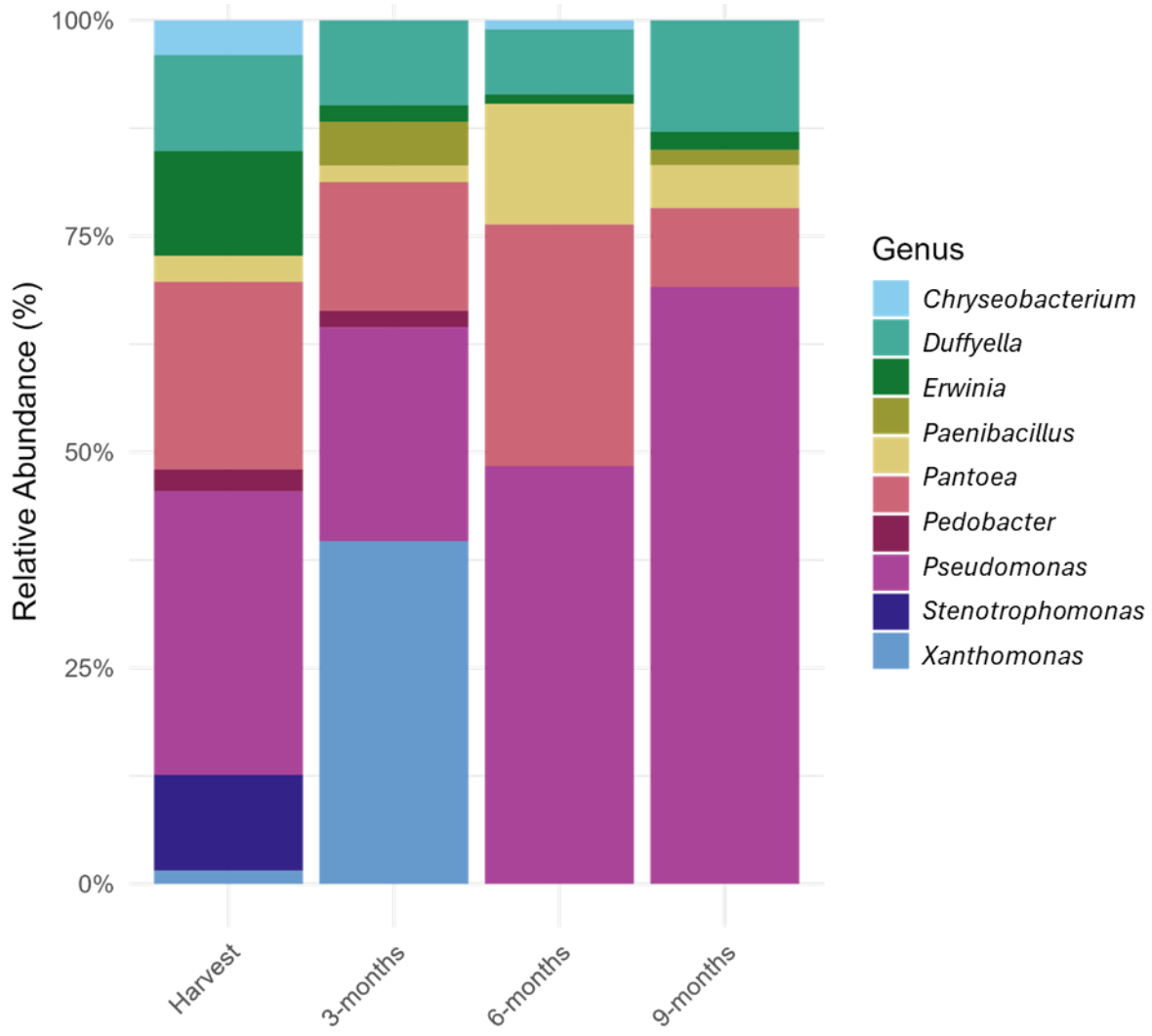


Figure 4.8 Bar plot of relative abundance showing the distribution of proteins of the top ten microbial genera in relation to storage time (harvest, three, six, and nine months).

#### 4.4.5.4 Distribution of proteins of the dominant genera based on grain nitrogen content

Variations in the taxonomic prevalence of peptides were observed between the high and low nitrogen content barley seeds (Fig. 4.9). The genus *Pseudomonas* exhibited the highest protein counts in both high (51%) and low (39.6%) nitrogen content samples, consistently dominating the protein profile (Fig. 4.9). In the high nitrogen samples, *Pantoea* was the second most abundant (15%) in terms of protein abundance, while in the low nitrogen samples, *Xanthomonas* proteins were more prevalent (22.5%) (Fig. 4.9). *Paenibacillus* showed a marked preference for high nitrogen conditions, making up 7% of proteins compared to only 2.1% in low nitrogen samples. *Stenotrophomonas* was found exclusively in high nitrogen samples (5%) and *Massilia* was only detected in low nitrogen samples (4.2%) (Fig. 4.9).

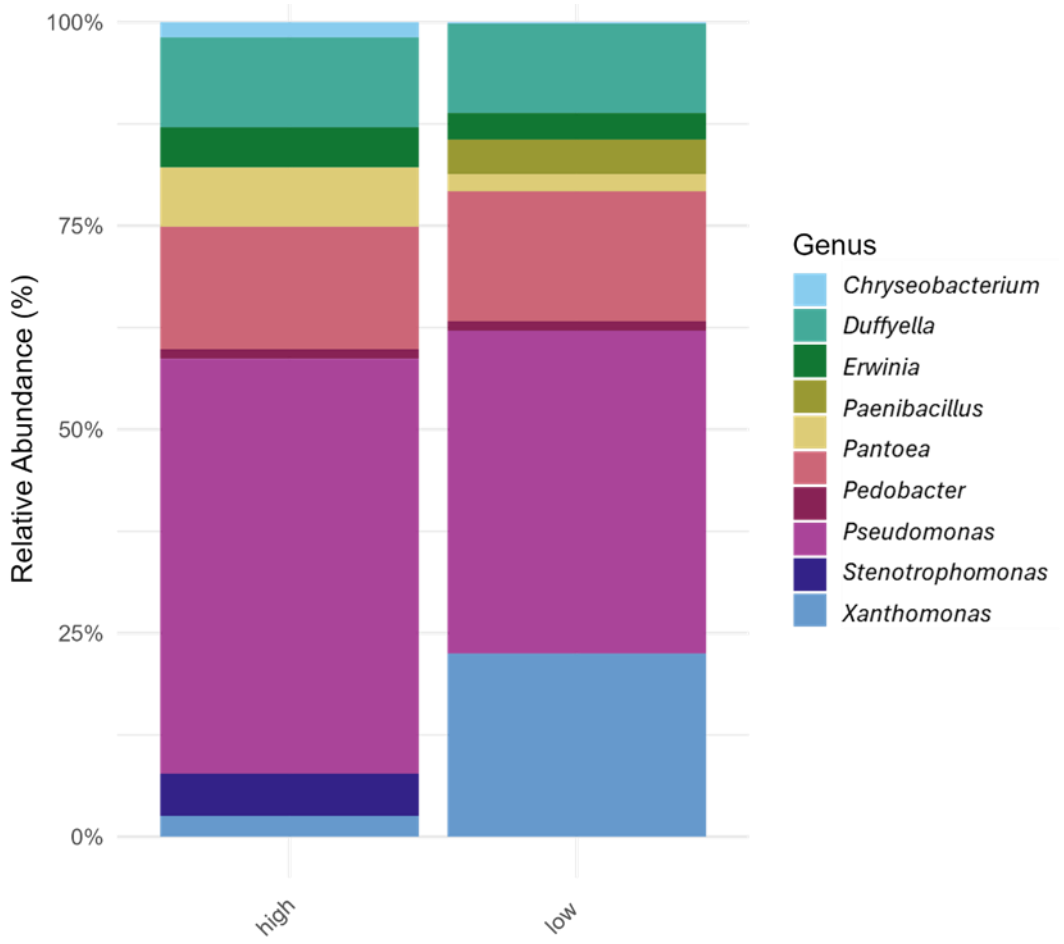


Figure 4.9 Bar plot of relative abundance showing the distribution of proteins of the top ten microbial genera in relation to grain nitrogen content (high vs low).

The metaproteomic data from barley seeds suggests that the nitrogen content may influence the activity of certain microbial genera. Genera such as *Paenibacillus*, *Pantoea* and *Pseudomonas* thrive regardless of nitrogen levels. *Stenotrophomonas*, however, exhibits a high protein count in high nitrogen conditions, suggesting that it may be more active when nitrogen is abundant, even though it appears in both high and low nitrogen metagenome samples. This indicates that its activity is likely enhanced in nitrogen rich environments (Reineke and Schlömann, 2023).

#### 4.4.5.5 Functional analysis of metaproteomic data

A total of 3,268 proteins were assigned to InterPro functions across 371 families (Fig. 4.10). The most prevalent bacterial protein families identified included outer membrane proteins (OMPs, 29%), catalases (18%), porins and TonB-dependent receptors (each at 13%), pyridine

nucleotide-disulfide oxidoreductases (7%), and solute-binding protein family (6%). Smaller proportions were observed for proteins involved in flagellar motility (flagellin), FAD/NAD(P)-binding domains (4%), OmpA-like domains (3%), and glucose metabolic processes (3%) (Fig. 4.10).

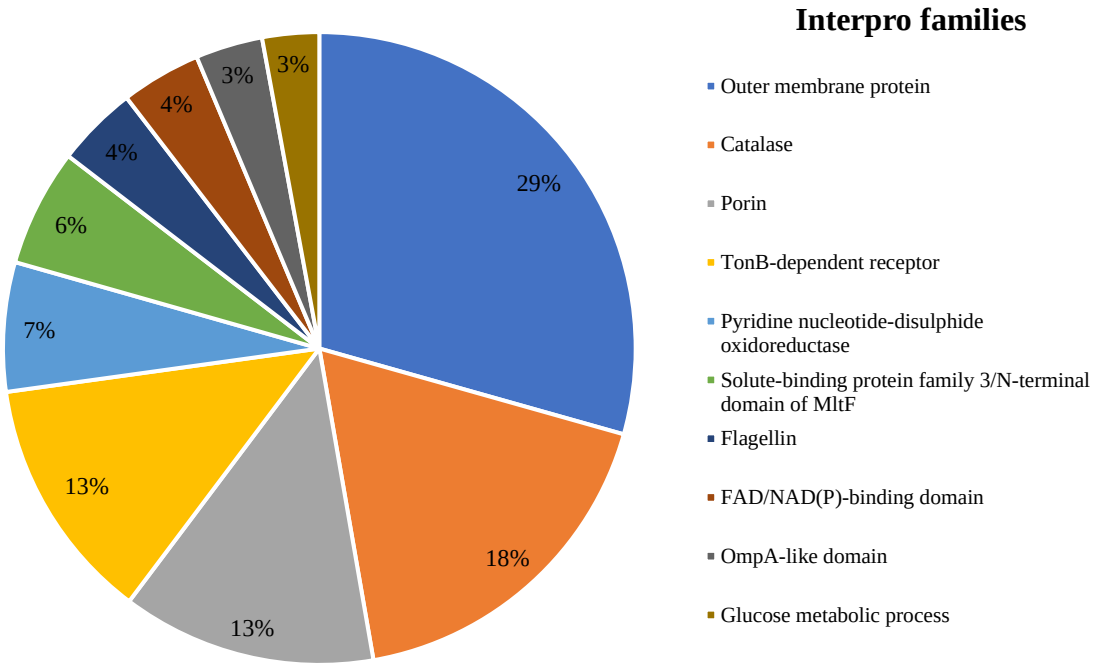


Figure 4.10 Abundance of the top ten bacterial proteins identified based on InterPro (Jones *et al.*, 2014).

The abundance of membrane-associated proteins underscores a primary focus on maintaining structural integrity. OMPs regulate the permeability of microbial cell envelopes, allowing selective transport of nutrients and metabolites and protecting against toxins, thereby shaping cellular structure and morphology (Nikaido, 2003). In Gram-negative bacteria, OMPs are particularly versatile, engaging in a range of functions, including signalling, host cell adhesion, catalysing key reactions, and facilitating both active and passive solute transport (Shearer *et al.*, 2019). Additionally, porins play a role in passively transporting hydrophilic molecules across membranes (Achouak *et al.*, 2001). TonB-dependent receptors, typically located in the outer membrane of Gram-negative bacteria, play a role in the uptake and transportation of substrates, including iron siderophore complexes, vitamin B12, nickel complexes and carbohydrates (Noinaj *et al.*, 2010).

Catalases are the second most prevalent function within the microbial community (Fig 4.10). They play a pivotal role in defending against oxidative stress by catalysing the decomposition of hydrogen peroxide into water and oxygen (Aebi, 1984). This crucial reaction protects microbial cells from the damaging effects of reactive oxygen species (ROS), which can originate from environmental stresses or metabolic activities (Glorieux and Calderon, 2017). By mitigating oxidative damage, catalases sustain the resilience and health of the microbial community (Aebi, 1984). They also support vital processes such as cell development, differentiation, and metabolite production (Glorieux and Calderon, 2017).

## 4.5 Conclusions

This study delved into the functional capabilities of the barley seed microbiome using metagenomic and metaproteomic approaches. Functional annotation of the proteins encoded on the metagenomes identified that the microbiome encompasses a broad array of key functions, including metabolic functions (e.g., carbohydrate and amino acid metabolism) and nutrient cycling pathways that could ensure their succession in the barley seed environment throughout long-term storage. This analysis also highlighted how variations in storage duration and nitrogen content influence the microbial dynamics within the barley seed microbiome. Variations in the microbiome taxonomy and protein profiles over different storage durations and grain nitrogen content can reveal how specific microbial communities and their metabolic activities are optimised for barley seed preservation and preparation for brewing.

These findings were complemented by metaproteomic investigations, which provided insights into protein expression by the barley seed microbiome. The presence of proteins linked to key microbial genera, as discussed in Chapters 2 and 3, affirmed their active roles within the microbiome. The proteins identified were predominantly involved in functions related to outer membrane proteins and catalases, crucial for bacterial survival and adaptation under various conditions.

While this integrated meta-omics approach provides insights into the microbial ecology of stored barley seeds, there are inherent limitations to these techniques, particularly in the context of complex microbial communities. For instance, functional annotation often relies on reference databases that may not comprehensively cover the microbial diversity present, potentially leading to incomplete or biased functional profiles. Additionally, metagenomic

sequencing can fail to capture the full scope of low-abundance microbial species, which may still play critical roles in the ecosystem. Furthermore, while metaproteomics can reveal expressed proteins, it does not capture transient interactions or dynamic functional changes within the community over time. These limitations must be considered when interpreting the results, particularly regarding the broader ecological interactions and their real-time impact on barley seed health and preservation.

Overall, the integration of metagenomic and metaproteomic analyses enhances our understanding of the microbial dynamics within the barley seed microbiomes, effectively bridging the gap between genomic potential and actual protein expression under real-time conditions. However, future studies incorporating more refined approaches, such as single-cell sequencing or real-time functional assays, may help overcome these limitations and provide deeper insights into the functional interactions within complex microbial communities.

## 4.6 References


Abraham, P.E., Giannone, R.J., Xiong, W., Hettich, R.L. (2014). Metaproteomics: Extracting and Mining Proteome Information to Characterize Metabolic Activities in Microbial Communities. *Current Protocols in Bioinformatics*, **46**: 13.26.1-13.26.14.

Achouak, W., Heulin, T., Pagès, J.-M. (2001). Multiple facets of bacterial porins. *FEMS Microbiology Letters*, **199**: 1–7.

Aebi, H. (1984). Catalase in vitro. *Methods in Enzymology*, **105**: 121–126.

Aggarwal, S., Yadav, A.K. (2016). False Discovery Rate Estimation in Proteomics. In: Jung, K. (ed) *Statistical Analysis in Proteomics*. Methods in Molecular Biology. Springer, New York, NY, pp. 119–128.

Agu, R.C., Palmer, G.H. (2001). The Effect of Nitrogen Level on the Performance of Malting Barley Varieties during Germination. *Journal of the Institute of Brewing*, **107**: 93–98.

Akinola, S.A., Ayangbenro, A.S., Babalola, O.O. (2021). The diverse functional genes of maize rhizosphere microbiota assessed using shotgun metagenomics. *Journal of the Science of Food and Agriculture*, **101**: 3193–3201.

Arduini, I., Masoni, A., Ercoli, L., Mariotti, M. (2006). Grain yield, and dry matter and nitrogen accumulation and remobilization in durum wheat as affected by variety and seeding rate. *European Journal of Agronomy*, **25**: 309–318.

Bar-Even, A. (2016). Formate Assimilation: The Metabolic Architecture of Natural and Synthetic Pathways. *Biochemistry*, **55**: 3851–3863.

Berg, G., Rybakova, D., Fischer, D., Cernava, T., Vergès, M.-C.C., Charles, T., Chen, X., Cocolin, L., Eversole, K., Corral, G.H., Kazou, M., Kinkel, L., Lange, L., Lima, N., Loy, A., Macklin, J.A., Maguin, E., Mauchline, T., McClure, R., Mitter, B., Ryan, M., Sarand, I., Smidt, H., Schelkle, B., Roume, H., Kiran, G.S., Selvin, J., Souza, R.S.C. de, van Overbeek, L., Singh, B.K., Wagner, M., Walsh, A., Sessitsch, A., Schloter, M. (2020). Microbiome definition re-visited: old concepts and new challenges. *Microbiome*, **8**: 103.

Biely, P. (2012). Microbial carbohydrate esterases deacetylating plant polysaccharides. *Biotechnology Advances*, **30**: 1575–1588.

Bokulich, N.A., Mills, D.A. (2012). Differentiation of mixed lactic acid bacteria communities in beverage fermentations using targeted terminal restriction fragment length polymorphism. *Food Microbiology*, **31**: 126–132.

Boraston, A.B., Bolam, D.N., Gilbert, H.J., Davies, G.J. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *The Biochemical Journal*, **382**: 769–781.

Bulgarelli, D., Garrido-Oter, R., Münch, P.C., Weiman, A., Dröge, J., Pan, Y., McHardy, A.C., Schulze-Lefert, P. (2015). Structure and Function of the Bacterial Root Microbiota in Wild and Domesticated Barley. *Cell Host & Microbe*, **17**: 392–403.

Burris, R.H., Roberts, G.P. (1993). Biological nitrogen fixation. *Annual Review of Nutrition*, **13**: 317–335.

Bziuk, N., Maccario, L., Straube, B., Wehner, G., Sørensen, S.J., Schikora, A., Smalla, K. (2021). The treasure inside barley seeds: microbial diversity and plant beneficial bacteria. *Environmental Microbiome*, **16**: 20.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research*, **37**: D233–D238.

Cao, D., Lou, Y., Jiang, X., Zhang, D., Liu, J. (2022). Fungal Diversity in Barley Under Different Storage Conditions. *Frontiers in Microbiology*, **13**.

Chandel, A., Mann, R., Kaur, J., Norton, S., Edwards, J., Spangenberg, G., Sawbridge, T. (2021). Implications of Seed Vault Storage Strategies for Conservation of Seed Bacterial Microbiomes. *Frontiers in Microbiology*, **12**: 784796.

Chen, Y., Liang, J., Zia, A., Gao, X., Wang, Y., Zhang, L., Xiang, Q., Zhao, K., Yu, X., Chen, Q., Penttinen, P., Nyima, T., Gu, Y. (2022). Culture dependent and independent characterization of endophytic bacteria in the seeds of highland barley. *Frontiers in Microbiology*, **13**.

Compeau, P.E.C., Pevzner, P.A., Tesler, G. (2011). How to apply de Bruijn graphs to genome assembly. *Nature Biotechnology*, **29**: 987–991.

Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., Mann, M. (2014). Accurate Proteome-wide Label-free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ. *Molecular & Cellular Proteomics : MCP*, **13**: 2513–2526.

Cruz, M.R., Graham, C.E., Gagliano, B.C., Lorenz, M.C., Garsin, D.A. (2013). *Enterococcus faecalis* Inhibits Hyphal Morphogenesis and Virulence of Candida albicans. *Infection and Immunity*, **81**: 189–200.

Desai, N., Antonopoulos, D., Gilbert, J.A., Glass, E.M., Meyer, F. (2012). From genomics to metagenomics. *Current Opinion in Biotechnology*, **23**: 72–76.

Deveau, A., Kohler, A., Frey-Klett, P., Martin, F. (2008). The Major Pathways of Carbohydrate Metabolism in the Ectomycorrhizal Basidiomycete *Laccaria bicolor* S238N. *The New Phytologist*, **180**: 379–390.

Dold, S.-M., Syldatk, C., Rudat, J. (2016). Transaminases and their Applications. *Green Biocatalysis*. John Wiley & Sons, Ltd, pp. 715–746.

Dutta, S., Choi, S.Y., Lee, Y.H. (2022). Temporal Dynamics of Endogenous Bacterial Composition in Rice Seeds During Maturation and Storage, and Spatial Dynamics of the Bacteria During Seedling Growth. *Frontiers in Microbiology*, **13**.

Egas, R.A., Kurth, J.M., Boeren, S., Sousa, D.Z., Welte, C.U., Sánchez-Andrea, I. (2024). A novel mechanism for dissimilatory nitrate reduction to ammonium in *Acididesulfobacillus acetoxydan*s. *mSystems*, **9**: e00967-23.

Egle, K., Beschow, H., Merbach, W. (2015). Nitrogen allocation in barley: Relationships between amino acid transport and storage protein synthesis during grain filling. *Canadian Journal of Plant Science*, **95**: 451–459.

Elias, J.E., Gygi, S.P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature Methods*, **4**: 207–214.

Erickson, A.R., Cantarel, B.L., Lamendella, R., Darzi, Y., Mongodin, E.F., Pan, C., Shah, M., Halfvarson, J., Tysk, C., Henrissat, B., Raes, J., Verberkmoes, N.C., Fraser, C.M., Hettich, R.L., Jansson, J.K. (2012). Integrated Metagenomics/Metaproteomics Reveals Human Host-Microbiota Signatures of Crohn's Disease. *PLOS ONE*, **7**: e49138.

Fadiji, A.E., Babalola, O.O. (2020). Metagenomics methods for the study of plant-associated microbial communities: A review. *Journal of Microbiological Methods*, **170**: 105860.

Fani, R., Gallo, R., Liò, P. (2000). Molecular Evolution of Nitrogen Fixation: The Evolutionary History of the nifD, nifK, nifE, and nifN Genes. *Journal of Molecular Evolution*, **51**: 1–11.

Felšöciová, S., Kowalczewski, P.Ł., Krajčovič, T., Dráb, Š., Kačániová, M. (2021). Effect of Long-Term Storage on Mycobiota of Barley Grain and Malt. *Plants (Basel, Switzerland)*, **10**: 1655.

Felšöciová, S., Kowalczewski, P.Ł., Krajčovič, T., Dráb, Š., Kačániová, M. (2020). Quantitative and Qualitative Composition of Bacterial Communities of Malting Barley Grain and Malt during Long-Term Storage. *Agronomy*, **10**: 1301.

Fierer, N., Leff, J.W., Adams, B.J., Nielsen, U.N., Bates, S.T., Lauber, C.L., Owens, S., Gilbert, J.A., Wall, D.H., Caporaso, J.G. (2012). Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proceedings of the National Academy of Sciences*, **109**: 21390–21395.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L.L., Tate, J., Punta, M. (2014). Pfam: the protein families database. *Nucleic Acids Research*, **42**: D222–D230.

Franceus, J., Desmet, T. (2020). Sucrose Phosphorylase and Related Enzymes in Glycoside Hydrolase Family 13: Discovery, Application and Engineering. *International Journal of Molecular Sciences*, **21**: 2526.

Gaspar, P., Neves, A.R., Shearman, C.A., Gasson, M.J., Baptista, A.M., Turner, D.L., Soares, C.M., Santos, H. (2007). The lactate dehydrogenases encoded by the ldh and ldhB genes in *Lactococcus lactis* exhibit distinct regulation and catalytic properties − comparative modeling to probe the molecular basis. *The FEBS Journal*, **274**: 5924–5936.

Ghali, A. (2014). *Circular Storage Tanks and Silos, Third Edition*. CRC Press, 392 pp.

Givan, C.V. (1980). 8 - Aminotransferases in Higher Plants. In: Miflin, B.J. (ed) *Amino Acids and Derivatives*. Academic Press, pp. 329–357.

Glorieux, C., Calderon, P.B. (2017). Catalase, a remarkable enzyme: targeting the oldest antioxidant enzyme to find a new cancer treatment approach. *Biological Chemistry*, **398**: 1095–1108.

Gupta, N., Bandeira, N., Keich, U., Pevzner, P.A. (2011). Target-Decoy Approach and False Discovery Rate: When Things May Go Wrong. *Journal of the American Society for Mass Spectrometry*, **22**: 1111–1120.

Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, **29**: 1072–1075.

Hall, E., Besemer, K., Kohl, L., Preiler, C., Riedel, K., Schneider, T., Wanek, W., Battin, T. (2012). Effects of Resource Chemistry on the Composition and Function of Stream Hyporheic Biofilms. *Frontiers in Microbiology*, **3**.

Harris, M.A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C., Richter, J., Rubin, G.M., Blake, J.A., Bult, C., Dolan, M., Drabkin, H., Eppig, J.T., Hill, D.P., Ni, L., Ringwald, M., Balakrishnan, R., Cherry, J.M., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S., Fisk, D.G., Hirschman, J.E., Hong, E.L., Nash, R.S., Sethuraman, A., Theesfeld, C.L., Botstein, D., Dolinski, K., Feierbach, B., Berardini, T., Mundodi, S., Rhee, S.Y., Apweiler, R., Barrell, D., Camon, E., Dimmer, E., Lee, V., Chisholm, R., Gaudet, P., Kibbe, W., Kishore, R., Schwarz, E.M., Sternberg, P., Gwinn, M., Hannick, L., Wortman, J., Berriman, M., Wood, V., de la Cruz, N., Tonellato, P., Jaiswal, P., Seigfried, T., White, R., Gene Ontology Consortium. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research*, **32**: D258-261.

Heyer, R., Schallert, K., Büdel, A., Zoun, R., Dorl, S., Behne, A., Kohrs, F., Püttker, S., Siewert, C., Muth, T., Saake, G., Reichl, U., Benndorf, D. (2019). A Robust and Universal Metaproteomics Workflow for Research Studies and Routine Diagnostics Within 24 h Using Phenol Extraction, FASP Digest, and the MetaProteomeAnalyzer. *Frontiers in Microbiology*, **10**: 1883.

Hirko, B., Mitiku, H., Getu, A. (2023). Role of fermentation and microbes in cacao fermentation and their impact on cacao quality. *Systems Microbiology and Biomanufacturing*, **3**: 509–520.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., von Mering, C., Bork, P. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, **47**: D309–D314.

Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, **11**: 119.

Jetten, M.S.M. (2008). The microbial nitrogen cycle. *Environmental Microbiology*, **10**: 2903–2909.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, **30**: 1236–1240.

Krivoruchko, A., Zhang, Y., Siewers, V., Chen, Y., Nielsen, J. (2015). Microbial acetyl-CoA metabolism and metabolic engineering. *Metabolic Engineering*, **28**: 28–42.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**: 680–685.

Laitila, A., Manninen, J., Priha, O., Smart, K., Tsitko, I., James, S. (2018). Characterisation of barley-associated bacteria and their impact on wort separation performance. *Journal of the Institute of Brewing*, **124**: 314–324.

Langenaeken, N.A., Ieven, P., Hedlund, E.G., Kyomugasho, C., van de Walle, D., Dewettinck, K., Van Loey, A.M., Roeffaers, M.B.J., Courtin, C.M. (2020). Arabinoxylan, β-glucan and pectin in barley and malt endosperm cell walls: a microstructure study using CLSM and cryo-SEM. *The Plant Journal: For Cell and Molecular Biology*, **103**: 1477–1489.

Langmead, B., Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature methods*, **9**: 357–359.

Laslett, D., Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Research*, **32**: 11–16.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**: 2078–2079.

Liu, Y., Chu, G., Stirling, E., Zhang, H., Chen, S., Xu, C., Zhang, X., Ge, T., Wang, D. (2023). Nitrogen fertilization modulates rice seed endophytic microbiomes and grain quality. *Science of The Total Environment*, **857**: 159181.

Lok, C. (2015). Mining the microbial dark matter. *Nature*, **522**: 270–273.

Maia, T.M., Staes, A., Plasman, K., Pauwels, J., Boucher, K., Argentini, A., Martens, L., Montoye, T., Gevaert, K., Impens, F. (2020). Simple Peptide Quantification Approach for MS-Based Proteomics Quality Control. *ACS Omega*, **5**: 6754–6762.

Marin, K., Krämer, R. (2007). Amino Acid Transport Systems in Biotechnologically Relevant Bacteria. In: Wendisch, V.F. (ed) *Amino Acid Biosynthesis ~ Pathways, Regulation and Metabolic Engineering*. Springer, Berlin, Heidelberg, pp. 289–325.

Mejía-Almonte, C., Busby, S.J.W., Wade, J.T., van Helden, J., Arkin, A.P., Stormo, G.D., Eilbeck, K., Palsson, B.O., Galagan, J.E., Collado-Vides, J. (2020). Redefining fundamental concepts of transcription initiation in bacteria. *Nature reviews. Genetics*, **21**: 699–714.

Michl, K., Berg, G., Cernava, T. (2023). The microbiome of cereal plants: The current state of knowledge and the potential for future applications. *Environmental Microbiome*, **18**: 28.

Mordant, A., Kleiner, M. (2021). Evaluation of Sample Preservation and Storage Methods for Metaproteomics Analysis of Intestinal Microbiomes. *Microbiology Spectrum*, **9**: e01877-21.

Nelson, E.B. (2018). The seed microbiome: Origins, interactions, and impacts. *Plant and Soil*, **422**: 7–34.

Nikaido, H. (2003). Molecular Basis of Bacterial Outer Membrane Permeability Revisited. *Microbiology and Molecular Biology Reviews*, **67**: 593–656.

Noinaj, N., Guillier, M., Barnard, T.J., Buchanan, S.K. (2010). TonB-dependent transporters: regulation, structure, and function. *Annual review of microbiology*, **64**: 43–60.

Nurk, S., Meleshko, D., Korobeynikov, A., Pevzner, P.A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Research*, **27**: 824–834.

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., Kanehisa, M. (1999). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, **27**: 29–34.

Papudeshi, B., Haggerty, J.M., Doane, M., Morris, M.M., Walsh, K., Beattie, D.T., Pande, D., Zaeri, P., Silva, G.G.Z., Thompson, F., Edwards, R.A., Dinsdale, E.A. (2017). Optimizing and evaluating the reconstruction of Metagenome-assembled microbial genomes. *BMC genomics*, **18**: 915.

Parés Viader, R., Yde, M.S.H., Hartvig, J.W., Pagenstecher, M., Carlsen, J.B., Christensen, T.B., Andersen, M.L. (2021). Optimization of Beer Brewing by Monitoring α-Amylase and β-Amylase Activities during Mashing. *Beverages*, **7**: 13.

Park, Y.-J., Jeong, Y.-U., Kong, W.-S. (2018). Genome Sequencing and Carbohydrate-Active Enzyme (CAZyme) Repertoire of the White Rot *Fungus Flammulina elastica*. *International Journal of Molecular Sciences*, **19**.

Parra, G., Bradnam, K., Ning, Z., Keane, T., Korf, I. (2009). Assessing the gene space in draft genomes. *Nucleic Acids Research*, **37**: 289–297.

Prakash, A., Jeffryes, M., Bateman, A., Finn, R.D. (2017). The HMMER Web Server for Protein Sequence Similarity Search. *Current Protocols in Bioinformatics*, **60**: 3.15.1-3.15.23.

Rahman, M.M., Flory, E., Koyro, H.-W., Abideen, Z., Schikora, A., Suarez, C., Schnell, S., Cardinale, M. (2018). Consistent associations with beneficial bacteria in the seed endosphere of barley (*Hordeum vulgare* L.). *Systematic and Applied Microbiology*, **41**: 386–398.

Reineke, W., Schlömann, M. (2023). The Microbial Nitrogen Cycle. In: Reineke, W. & Schlömann, M. (eds) *Environmental Microbiology*. Springer, Berlin, Heidelberg, pp. 291–305.

Robertson-Albertyn, S., Abbott, J.C., Concas, F., Brown, L.H., Orr, J.N., George, T.S., Bulgarelli, D. (2022). Genome-Annotated Bacterial Collection of the Barley Rhizosphere Microbiota. *Microbiology Resource Announcements*, **11**: e01064-21.

Rosenberger, A.F.N., Hangelmann, L., Hofinger, A., Wilson, I.B.H. (2012). UDP-xylose and UDP-galactose synthesis in *Trichomonas vaginalis*. *Molecular and Biochemical Parasitology*, **181**: 53–56.

Rudat, J., Brucher, B.R., Syldatk, C. (2012). Transaminases for the synthesis of enantiopure beta-amino acids. *AMB Express*, **2**: 11.

Saito, M.A., Bertrand, E.M., Duffy, M.E., Gaylord, D.A., Held, N.A., Hervey, W.J.I., Hettich, R.L., Jagtap, P.D., Janech, M.G., Kinkade, D.B., Leary, D.H., McIlvin, M.R., Moore, E.K., Morris, R.M., Neely, B.A., Nunn, B.L., Saunders, J.K., Shepherd, A.I., Symmonds, N.I., Walsh, D.A. (2019). Progress and Challenges in Ocean Metaproteomics and Proposed Best Practices for Data Sharing. *Journal of Proteome Research*, **18**: 1461–1476.

Salvato, F., Hettich, R.L., Kleiner, M. (2021). Five key aspects of metaproteomics as a tool to understand functional interactions in host-associated microbiomes. *PLoS pathogens*, **17**: e1009245.

Schiebenhoefer, H., Van Den Bossche, T., Fuchs, S., Renard, B.Y., Muth, T., Martens, L. (2019). Challenges and promise at the interface of metaproteomics and genomics: an overview of recent progress in metaproteogenomic data analysis. *Expert Review of Proteomics*, **16**: 375–390.

Schmid, J., Heider, D., Wendel, N.J., Sperl, N., Sieber, V. (2016). Bacterial Glycosyltransferases: Challenges and Opportunities of a Highly Diverse Enzyme Class Toward Tailoring Natural Products. *Frontiers in Microbiology*, **7**.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, **30**: 2068–2069.

Seymour, S., Hunter, C. (2015). ProteinPilot™ Software Overview High Quality, In-Depth Protein Identification and Protein Expression Analysis.

Shabala, S., Shabala, L. (2011). Ion transport and osmotic adjustment in plants and bacteria. **2**: 407–419.

Shafiei, R., Hooper, M., McClellan, C., Oakey, H., Stephens, J., Lapierre, C., Tsuji, Y., Goeminne, G., Vanholme, R., Boerjan, W., Ralph, J., Halpin, C. (2023). Downregulation of barley ferulate 5-hydroxylase dramatically alters straw lignin structure without impact on mechanical properties. *Frontiers in Plant Science*, **13**.

Sharma, V., Eng, J.K., Feldman, S., von Haller, P.D., MacCoss, M.J., Noble, W.S. (2010). Precursor charge state prediction for electron transfer dissociation tandem mass spectra. *Journal of Proteome Research*, **9**: 5438–5444.

Shearer, J., Jefferies, D., Khalid, S. (2019). Outer Membrane Proteins OmpA, FhuA, OmpF, EstA, BtuB, and OmpX Have Unique Lipopolysaccharide Fingerprints. *Journal of Chemical Theory and Computation*, **15**: 2608–2619.

Shevchenko, A., Tomas, H., Havli, J., Olsen, J.V., Mann, M. (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nature Protocols*, **1**: 2856–2860.

Shilov, I.V., Seymour, S.L., Patel, A.A., Loboda, A., Tang, W.H., Keating, S.P., Hunter, C.L., Nuwaysir, L.M., Schaeffer, D.A. (2007). The Paragon Algorithm, a Next Generation Search Engine That Uses Sequence Temperature Values and Feature Probabilities to Identify Peptides from Tandem Mass Spectra *. *Molecular & Cellular Proteomics*, **6**: 1638–1655.

Sidar, A., Albuquerque, E.D., Voshol, G.P., Ram, A.F.J., Vijgenboom, E., Punt, P.J. (2020). Carbohydrate Binding Modules: Diversity of Domain Architecture in Amylases and Cellulases From Filamentous Microorganisms. *Frontiers in Bioengineering and Biotechnology*, **8**.

Song, A.A.-L., In, L.L.A., Lim, S.H.E., Rahim, R.A. (2017). A review on *Lactococcus lactis*: from food to factory. *Microbial Cell Factories*, **16**: 55.

Stam, M.R., Danchin, E.G.J., Rancurel, C., Coutinho, P.M., Henrissat, B. (2006). Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of α-amylase-related proteins. *Protein Engineering, Design and Selection*, **19**: 555–562.

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., Rao, B.S., Smirnov, S., Sverdlov,

A.V., Vasudevan, S., Wolf, Y.I., Yin, J.J., Natale, D.A. (2003). The COG database: an updated version includes eukaryotes. *BMC bioinformatics*, **4**: 41.

Tatusov, R.L., Koonin, E.V., Lipman, D.J. (1997). A Genomic Perspective on Protein Families. *Science*, **278**: 631–637.

Tiso, M., Schechter, A.N. (2015). Nitrate Reduction to Nitrite, Nitric Oxide and Ammonia by Gut Bacteria under Physiological Conditions. *PLOS ONE*, **10**: e0119712.

Tshisekedi, K.A., De Maayer, P., Botes, A. (2024). Metagenomic sequencing and reconstruction of 82 microbial genomes from barley seed communities. *Scientific Data*, **11**: 484.

UniProt Consortium, T. (2018). UniProt: the universal protein knowledgebase. *Nucleic Acids Research*, **46**: 2699.

Verschaffelt, P., Van Den Bossche, T., Martens, L., Dawyndt, P., Mesuere, B. (2021). Unipept Desktop: A Faster, More Powerful Metaproteomics Results Analysis Tool. *Journal of Proteome Research*, **20**: 2005–2009.

van der Walt, A.J., van Goethem, M.W., Ramond, J.-B., Makhalanyane, T.P., Reva, O., Cowan, D.A. (2017). Assembling metagenomes, one community at a time. *BMC Genomics*, **18**: 521.

Wang, P., Wang, F. (2023). A proposed metric set for evaluation of genome assembly quality. *Trends in Genetics*, **39**: 175–186.

Wisnoski, N.I., Lennon, J.T. (2021). Stabilising role of seed banks and the maintenance of bacterial diversity. *Ecology Letters*, **24**: 2328–2338.

Yang, L., Danzberger, J., Schöler, A., Schröder, P., Schloter, M., Radl, V. (2017). Dominant Groups of Potentially Active Bacteria Shared by Barley Seeds become Less Abundant in Root Associated Microbiome. *Frontiers in Plant Science*, **8**: 1005.

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., Xu, Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research*, **40**: W445–W451.

Zhou, Z., Tran, P.Q., Breister, A.M., Liu, Y., Kieft, K., Cowley, E.S., Karaoz, U., Anantharaman, K. (2022). METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. *Microbiome*, **10**: 33.

Zhu, F., Wu, R., Zhang, H., Wu, H. (2013). Structural and biochemical analysis of a bacterial glycosyltransferase. *Methods in Molecular Biology (Clifton, N.J.)*, **1022**: 29–39.

Zumft, W.G. (1997). Cell biology and molecular basis of denitrification. *Microbiology and molecular biology reviews: MMBR*, **61**: 533–616.

# Chapter 5 Summary

The barley seed microbiome plays a critical role in determining the storage life and quality of seeds, however, the intricate interactions within this microbial community and their precise influence on seed health and longevity remain unknown. This research explored the microbiome of South African barley seeds, with either high or low nitrogen content, at harvest and after three, six and nine months of storage in silos. Metagenome sequencing and taxonomic classification of the resultant reads revealed a predominance of bacterial reads with minimal fungal presence. Lytic dsDNA phages were also detected within the sequence reads. An in-depth analysis of read distribution across different storage periods (harvest, three, six and nine months) and grain nitrogen contents (high and low) showed that each phase had a distinct microbial composition, with varying relative abundances. Diversity analysis indicated that the duration of storage, rather than the nitrogen content of the grain, impacted the microbial diversity associated with the barley seeds. These findings directly address the research question concerning the role of nitrogen in microbial dynamics by demonstrating that while nitrogen content did not significantly impact overall microbial diversity, it did influence specific community compositions. This advances the understanding of how nitrogen modulates the microbial ecosystem within barley seeds, particularly in relation to community functionality over time. Core microbial genera identified in this study align with previous research looking at the barley microbiome. Binning of the metagenomic reads facilitated the reconstruction of 82 high-quality metagenome-assembled genomes (MAGs) representative of the core microbial taxa. Comparative genomic analysis showed that these MAGs contained few shared single-copy orthologues, with an extensive accessory genome observed for each distinct MAG, suggesting they function as a consortium playing complementary roles within the barley seed environment. Additionally, certain MAGs harboured pathogenicity determinants that could be detrimental to the health of the barley seeds, and ultimately malt quality.

Metagenome assembly, followed by structural and functional annotation identified a broad array of metabolic pathways and carbohydrate active enzymes highlighting the key role of the barley seed microbiome in nutrient cycling and by extension the health, germination and growth of the barley seed. Metaproteomic analysis identified a broad array of expressed microbial proteins associated with the barley seed microbiome, with a large proportion of these

associated with membrane protection and oxidative stress, potentially triggered by changes in environmental conditions during prolonged storage.

The application of metagenomic and metaproteomic analyses revealed that the South African barley microbiome comprises a diverse array of microorganisms, which is anything but static. Instead, the microbiome shifts in response to prolonged storage and the different nitrogen content of the barley grains, both in terms of microbial community and their projected activities. This may be used to inform both on the storage practices and usage of barley grains with specific nitrogen contents to ensure barley grains of the highest nutritional and microbial quality for downstream applications, such as beer brewing. These findings not only contribute to our understanding of microbial dynamics in barley storage but also have broader implications for agricultural practices, particularly in managing nitrogen levels during barley storage. By optimising nitrogen content, it may be possible to promote beneficial microbial communities that enhance seed quality and preservation, further supporting sustainable storage practices in the brewing and grain industries.

One notable observation from the study was the distinct behaviour of the S6L sample, which consistently presented different quality and assembly metrics compared to other samples. S6L displayed signs of dysbiosis and had the highest number of contigs but the lowest N50 value, indicating a fragmented assembly. These unique characteristics suggest that S6L may represent an outlier, which poses challenges for interpreting the microbial community dynamics during storage. Replication could have confirmed the status of S6L as an outlier, and the inclusion of replicates in future studies could help ensure robust conclusions.

While the metagenomic and metaproteomic analyses shed light on who is associated with the barley seeds and what they are doing, further insights into how the bacterial community is interacting with the barley seed would be of benefit. Future research should build upon these findings by exploring microbial interactions and functional dynamics in greater detail. Expanding the scope to include metametabolomics and metatranscriptomics would provide more comprehensive insights into the metabolic exchanges and gene expression patterns within the barley seed microbiome during storage. In addition, studies that employ greater biological replication, including the use of multiple cultivars and environmental conditions, would help validate the current findings and account for variability such as the dysbiosis observed in the S6L sample. It would also be valuable to investigate how nitrogen levels influence microbial interactions, particularly focusing on nitrogen cycling, functional activity, and its broader

impact on barley seed preservation. Exploring the combined effects of nitrogen and environmental factors such as temperature and moisture would also provide a more comprehensive understanding of microbial dynamics in storage. Finally, studies that integrate machine learning or artificial intelligence (AI) driven models to predict microbial behaviour and optimise storage conditions would be highly beneficial for the brewing industry.

# Appendix

# Chapter 2 Appendix

Table S2.1. Overview of sample collection, sequencing metrics, and genomic data for barley seed microbiome study.

| Sample ID | Nitrogen grain content | Depot location | Storage-time | Reads | Contamination Ratio (%) | accession | Average quality of reads (PHRED score) |
|---|---|---|---|---|---|---|---|
| HH | High (>1.5%) | Klipdale | Harvest | 50840506 | 0.0128 | SRR27317622 | 36 |
| HL | Low (< 1.5%) | Caledon | Harvest | 44861190 | 0.0827 | SRR27317623 | 36 |
| S3H | High (>1.5%) | Caledon | 3-months | 76518028 | 0.0751 | SRR27317620 | 36 |
| S3L | Low (< 1.5%) | Napier | 3-months | 56163948 | 0.8122 | SRR27317621 | 36 |
| S6H | High (>1.5%) | Protem | 6-months | 37911140 | 0.0069 | SRR27317618 | 36 |
| S6L | Low (< 1.5%) | Krige | 6-months | 52681492 | 0.0059 | SRR27317619 | 36 |
| S9H | High (>1.5%) | Krige | 9-months | 25566560 | 0.2103 | SRR27317616 | 36 |
| S9L | Low (< 1.5%) | Protem | 9-months | 20722664 | 2.7368 | SRR27317617 | 36 |

Table S2.2. Summary of alpha diversity statistics for barley seed metagenomic samples.

| Sample ID | Nitrogen | Storage | Shannon Diversity |
|---|---|---|---|
| HHN | High-N | Harvest | 1.62 |
| HLN | Low-N | Harvest | 1.68 |
| T1HN | High-N | 3-months | 1.52 |
| T1LN | Low-N | 3-months | 1.42 |
| T2HN | High-N | 6-months | 1.22 |
| T2LN | Low-N | 6-months | 2.42 |
| T3HN | High-N | 9-months | 1.83 |
| T3LN | Low-N | 9-months | 1.51 |

Kruskal-Wallis test for Shannon diversity by storage

Kruskal-Wallis chi-squared = 1.1667, df = 3, p-value = 0.761

Kruskal-Wallis test for Shannon diversity by nitrogen

Table S2.3. Summary of beta diversity statistics for barley seed metagenomic samples.

| Sample ID | HLN | T1LN | T2LN | T3LN | HHN | T1HN | T2HN | T3HN |
|---|---|---|---|---|---|---|---|---|
| HLN | 0.0 | 0.46 | 0.49 | 0.42 | 0.18 | 0.57 | 0.57 | 0.45 |
| T1LN | 0.46 | 0.0 | 0.84 | 0.42 | 0.5 | 0.33 | 0.62 | 0.54 |
| T2LN | 0.49 | 0.84 | 0.0 | 0.8 | 0.51 | 0.88 | 0.81 | 0.72 |
| T3LN | 0.42 | 0.42 | 0.8 | 0.0 | 0.59 | 0.55 | 0.45 | 0.41 |
| HHN | 0.18 | 0.5 | 0.51 | 0.59 | 0.0 | 0.6 | 0.63 | 0.45 |
| T1HN | 0.57 | 0.33 | 0.88 | 0.55 | 0.6 | 0.0 | 0.7 | 0.64 |
| T2HN | 0.57 | 0.62 | 0.81 | 0.45 | 0.63 | 0.7 | 0.0 | 0.44 |
| T3HN | 0.45 | 0.54 | 0.72 | 0.41 | 0.45 | 0.64 | 0.44 | 0.0 |

ANOSIM Test for Bray-Curtis Dissimilarity by Nitrogen

ANOSIM statistic R: -0.07292, Significance: 0.64, Number of permutations: 999

ANOSIM Test for Bray-Curtis Dissimilarity by Storage

ANOSIM statistic R: 0.5417, Significance: 0.034, Number of permutations: 999

Table S2.4 Kingdom and phylum distribution of metagenomic reads.

| **Kingdom** | **Phylum** | **HL** | **HH** | **S3L** | **S3H** | **S6L** | **S6H** | **S9L** | **S9H** |
|---|---|---|---|---|---|---|---|---|---|
| Bacteria | Pseudomonadota | 12986078 | 13968330 | 14631301 | 21177324 | 13574863 | 11033650 | 6610882 | 7392268 |
| Bacteria | Bacteroidota | 1311673 | 1094279 | 582331 | 289454 | 918729 | 421 | 85698 | 41518 |
| Bacteria | Actinomycetota | 124141 | 502281 | 171710 | 1549330 | 14118 | 1294 | 256967 | 51588 |
| Bacteria | Bacillota | 9679 | 15098 | 50230 | 93903 | 1629501 | 142251 | 90762 | 466506 |
| Phages | Uroviricota (Virus) | 1338 | 1957 | 29562 | 153258 | 4024 | 14792 | 2195 | 4400 |
| Fungi | Ascomycota (Fungi) | 7352 | 8024 | 3535 | 7047 | 15066 | 15 | 6856 | 18373 |
| Fungi | Basidiomycota (Fungi) | 423 | 0 | 0 | 0 | 5949 | 0 | 5118 | 780 |
| Bacteria | Spirochaetes | 48 | 51 | 60 | 211 | 0 | 0 | 0 | 0 |
| Bacteria | Cyanobacteria | 19 | 0 | 57 | 46 | 24 | 0 | 20 | 0 |
| Bacteria | Planctomycetes | 0 | 13 | 20 | 18 | 21 | 10 | 0 | 0 |
| Archea | Euryarchaeota (Archea) | 0 | 10 | 0 | 11 | 0 | 0 | 20 | 17 |
| Bacteria | Verrucomicrobia | 0 | 0 | 13 | 0 | 0 | 0 | 0 | 0 |
| Bacteria | Fusobacteria | 0 | 0 | 0 | 12 | 0 | 0 | 0 | 0 |
| Bacteria | Acidobacteria | 0 | 0 | 0 | 0 | 12 | 0 | 0 | 0 |
| Bacteria | Deinococcus-Thermus | 0 | 0 | 10 | 0 | 0 | 0 | 0 | 0 |
| Bacteria | Tenericutes | 0 | 0 | 10 | 0 | 0 | 0 | 0 | 0 |

Table S2.5 Core and non-bacterial genera of barley seed microbiome.

| **Category** | **Genus** | **HL** | **HH** | **S9H** | **S6L** | **S9L** | **S6H** | **S3H** | **S3L** |
|---|---|---|---|---|---|---|---|---|---|
| Core | *Achromobacter* | 136 | 395 | 41 | 96159 | 27 | 12 | 304 | 222 |
| Core | *Aeromonas* | 24 | 33 | 46 | 647 | 96 | 32 | 96 | 69 |
| Core | *Azotobacter* | 14 | 19 | 32 | 21 | 10 | 11 | 94 | 43 |
| Core | *Bacillus* | 684 | 176 | 190235 | 174 | 989 | 11989 | 4157 | 167 |
| Core | *Brenneria* | 15 | 17 | 19 | 274 | 50 | 85 | 23 | 22 |
| Core | *Burkholderia* | 223 | 309 | 93 | 313 | 92 | 43 | 970 | 774 |
| Core | *Cedecea* | 57 | 293 | 151 | 1287 | 151 | 597 | 114 | 124 |
| Core | *Chryseobacterium* | 1120110 | 853586 | 32663 | 660035 | 56823 | 165 | 99934 | 323654 |
| Core | *Citrobacter* | 5634 | 1566 | 2769 | 1108262 | 3077 | 196765 | 2655 | 3668 |
| Core | *Clavibacter* | 20047 | 20001 | 6828 | 1280 | 88156 | 519 | 82865 | 13289 |
| Core | *Clostridium* | 11 | 11 | 468 | 878 | 75 | 11 | 35 | 84 |
| Core | *Cronobacter* | 129 | 63 | 115 | 3903 | 56 | 526 | 317 | 184 |
| Core | *Curtobacterium* | 40975 | 14270 | 15426 | 1545 | 108754 | 491 | 281302 | 30530 |
| Core | *Dickeya* | 384 | 89 | 195 | 865 | 220 | 265 | 202 | 232 |
| Core | *Edwardsiella* | 467 | 131 | 148 | 187 | 167 | 120 | 245 | 445 |
| Core | *Enterobacter* | 2671 | 1745 | 30145 | 601963 | 3330 | 175755 | 2874 | 2223 |
| Core | *Erwinia* | 1280489 | 1553131 | 1978325 | 109655 | 1352391 | 6303737 | 1073965 | 971993 |
| Core | *Escherichia* | 438 | 244 | 2215 | 14472 | 302 | 3595 | 499 | 3743 |
| Core | *Flavobacterium* | 6386 | 2001 | 89 | 742 | 192 | 61 | 2435 | 3019 |
| Core | *Frigoribacterium* | 1122 | 8451 | 948 | 66 | 14992 | 30 | 4104 | 4038 |
| Core | *Hafnia* | 107 | 59 | 110 | 197 | 18 | 523 | 53 | 73 |
| Core | *Klebsiella* | 1365 | 1802 | 596008 | 374996 | 2591 | 5548 | 2486 | 1625 |
| Core | *Kluyvera* | 16 | 13 | 38 | 1553 | 20 | 179 | 38 | 31 |
| Core | *Kosakonia* | 262 | 157 | 514 | 97892 | 316 | 1276 | 633 | 2037 |
| Core | *Leclercia* | 174 | 128 | 4205 | 2016197 | 543 | 126885 | 325 | 1618 |
| Core | *Lelliottia* | 259 | 102 | 61567 | 31765 | 3256 | 713 | 829 | 231 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Core | *Lonsdalea* | 32 | 10 | 64 | 62 | 10 | 25 | 157 | 26 |
| Core | *Massilia* | 138916 | 98933 | 3136 | 259 | 43941 | 615 | 278352 | 393871 |
| Core | *Microbacterium* | 5393 | 31750 | 333 | 2101 | 1339 | 13 | 505238 | 9228 |
| Core | *Mixta* | 368 | 105 | 316 | 852 | 238 | 1989 | 363 | 300 |
| Core | *Morganella* | 44 | 13 | 35 | 95 | 23 | 55 | 66 | 42 |
| Core | *Nissabacter* | 425 | 32 | 62 | 59 | 20 | 132 | 137 | 50 |
| Core | *Paenibacillus* | 5216 | 8541 | 27451 | 5178 | 4960 | 14915 | 45730 | 42449 |
| Core | *Pantoea* | 2635287 | 1095015 | 969409 | 271573 | 3199546 | 1800410 | 5218757 | 7091193 |
| Core | *Paraburkholderia* | 762 | 701 | 26 | 46 | 47 | 11 | 9169 | 1730 |
| Core | *Pectobacterium* | 1100 | 412 | 1680 | 1438 | 2046 | 2754 | 627 | 941 |
| Core | *Pluralibacter* | 87 | 380 | 101 | 929 | 40 | 68 | 107 | 24 |
| Core | *Pseudomonas* | 3032583 | 4309770 | 2885783 | 1417522 | 1633126 | 2047298 | 11685145 | 4639277 |
| Core | *Rahnella* | 826 | 341 | 1408 | 16445 | 1885 | 665 | 1095 | 660 |
| Core | *Ralstonia* | 242 | 310 | 16 | 279 | 13 | 46 | 1901 | 462 |
| Core | *Raoultella* | 415 | 215 | 1921 | 208824 | 1159 | 392 | 275 | 227 |
| Core | *Salmonella* | 141 | 364 | 1182 | 29787 | 263 | 2979 | 118 | 118 |
| Core | *Serratia* | 3241 | 2485 | 2545 | 9528 | 4383 | 5557 | 2483 | 1258 |
| Core | *Sphingomonas* | 30257 | 62832 | 2441 | 396 | 31051 | 608 | 128709 | 84458 |
| Core | *Stenotrophomonas* | 4841231 | 6102215 | 514716 | 4429884 | 92911 | 69 | 163857 | 71292 |
| Core | *Tatumella* | 12 | 12 | 38 | 89 | 88 | 42 | 28 | 19 |
| Core | *Vibrio* | 11 | 12 | 32 | 348 | 91 | 355 | 115 | 137 |
| Core | *Xanthomonas* | 99170 | 82002 | 68276 | 29748 | 33589 | 588 | 1435416 | 510985 |
| Core | *Yersinia* | 866 | 1305 | 1496 | 2314 | 645 | 1138 | 1015 | 266 |
| Non-Core | *Acetobacter* | 0 | 0 | 0 | 0 | 16 | 0 | 0 | 0 |
| Non-Core | *Acidithiobacillus* | 0 | 0 | 0 | 14 | 0 | 0 | 0 | 0 |
| Non-Core | *Acidovorax* | 124 | 376 | 0 | 12892 | 14 | 0 | 155 | 148 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Non-Core | *Acinetobacter* | 0 | 10 | 169 | 138942 | 556 | 12 | 466 | 264 |
| Non-Core | *Actinomyces* | 0 | 19 | 0 | 0 | 0 | 0 | 0 | 0 |
| Non-Core | *Actinoplanes* | 0 | 21 | 0 | 0 | 0 | 0 | 21 | 0 |
| Non-Core | *Actinotalea* | 0 | 20 | 0 | 0 | 0 | 0 | 0 | 0 |
| Non-Core | *Advenella* | 0 | 0 | 0 | 52 | 0 | 0 | 0 | 0 |
| Non-Core | *Aequorivita* | 0 | 10 | 0 | 0 | 0 | 0 | 12 | 0 |
| Non-Core | *Aeromicrobium* | 0 | 146 | 10 | 0 | 0 | 0 | 0 | 0 |
| Non-Core | *Afipia* | 0 | 0 | 25 | 0 | 55 | 0 | 12 | 0 |
| Non-Core | *Agreia* | 945 | 2121 | 50 | 41 | 76 | 0 | 1964 | 933 |
| Non-Core | *Agrobacterium* | 15502 | 10144 | 302 | 577 | 33 | 0 | 1923 | 1042 |
| Non-Core | *Agrococcus* | 0 | 47 | 0 | 0 | 22 | 0 | 61 | 0 |
| Non-Core | *Agromyces* | 22 | 314 | 25 | 12 | 31 | 0 | 363 | 134 |
| Non-Core | *Alcaligenes* | 10 | 11 | 37 | 724 | 0 | 0 | 13 | 0 |
| Non-Core | *Alcanivorax* | 0 | 0 | 0 | 18 | 0 | 0 | 25 | 0 |
| Non-Core | *Alicycliphilus* | 0 | 0 | 0 | 31 | 0 | 0 | 14 | 10 |
| Non-Core | *Alistipes* | 0 | 0 | 0 | 11 | 0 | 0 | 0 | 0 |
| Non-Core | *Allosphingosinicella* | 0 | 0 | 0 | 0 | 0 | 0 | 17 | 0 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Non-Core | *Altererythrobacter* | 0 | 0 | 0 | 0 | 0 | 0 | 15 | 0 |
| Non-Core | *Aminobacter* | 0 | 0 | 0 | 0 | 0 | 0 | 44 | 0 |
| Non-Core | *Amniculibacterium* | 36 | 49 | 0 | 0 | 0 | 0 | 36 | 31 |
| Non-Core | *Amycolatopsis* | 0 | 0 | 0 | 0 | 0 | 0 | 21 | 0 |
| Non-Core | *Amylolactobacillus* | 0 | 0 | 0 | 19 | 0 | 0 | 0 | 0 |
| Non-Core | *Anoxybacillus* | 0 | 0 | 0 | 0 | 0 | 0 | 15 | 0 |
| Non-Core | *Apibacter* | 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Non-Core | *Aquitalea* | 0 | 0 | 0 | 14 | 0 | 0 | 22 | 13 |
| Non-Core | *Arachidicoccus* | 0 | 0 | 0 | 29 | 0 | 0 | 0 | 17 |
| Non-Core | *Arenimonas* | 10 | 0 | 0 | 18 | 0 | 0 | 0 | 0 |
| Non-Core | *Aromatoleum* | 0 | 0 | 0 | 0 | 0 | 0 | 17 | 15 |
| Non-Core | *Arsenophonus* | 11 | 59 | 0 | 0 | 0 | 0 | 0 | 14 |
| Non-Core | *Arthrobacter* | 118 | 196 | 0 | 0 | 14 | 0 | 3401 | 119 |
| Non-Core | *Asticcacaulis* | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 0 |
| Non-Core | *Atlantibacter* | 15 | 0 | 46 | 8796 | 15 | 40 | 12 | 23 |
| Non-Core | *Aureimonas* | 0 | 44 | 0 | 0 | 0 | 0 | 37 | 29 |
| Non-Core | *Azoarcus* | 17 | 13 | 29 | 14 | 0 | 0 | 37 | 64 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Non-Core | *Azospira* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 |
| Non-Core | *Azospirillum* | 13 | 18 | 0 | 16 | 0 | 0 | 31 | 16 |
| Non-Core | *Bacteroides* | 11 | 13 | 0 | 18 | 0 | 0 | 14 | 31 |
| Non-Core | *Bergeyella* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| Non-Core | *Bifidobacterium* | 0 | 0 | 0 | 0 | 0 | 0 | 17 | 10 |
| Non-Core | *Blastochloris* | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 0 |
| Non-Core | *Blastomonas* | 0 | 0 | 0 | 0 | 0 | 0 | 16 | 0 |
| Non-Core | *Blattabacterium* | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 11 |
| Non-Core | *Blautia* | 0 | 0 | 0 | 23 | 0 | 0 | 0 | 0 |
| Non-Core | *Bordetella* | 72 | 148 | 0 | 474 | 13 | 0 | 110 | 148 |
| Non-Core | *Bosea* | 56 | 49 | 265 | 0 | 716 | 0 | 288 | 46 |
| Non-Core | *Brachybacterium* | 0 | 36 | 0 | 45 | 0 | 0 | 25 | 0 |
| Non-Core | *Bradyrhizobium* | 1454 | 1216 | 36 | 61 | 65 | 0 | 6207 | 1141 |
| Non-Core | *Brevibacillus* | 0 | 0 | 0 | 11 | 0 | 0 | 11 | 0 |
| Non-Core | *Brevibacterium* | 0 | 23 | 0 | 251 | 0 | 0 | 44 | 14 |
| Non-Core | *Brevundimonas* | 7922 | 12789 | 561 | 32172 | 115 | 0 | 1196 | 4863 |
| Non-Core | *Brochothrix* | 0 | 0 | 0 | 43 | 0 | 0 | 0 | 0 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Non-Core | *Brucella* | 0 | 35 | 0 | 31302 | 25 | 0 | 20 | 17 |
| Non-Core | *Buttiauxella* | 37 | 0 | 27 | 1446 | 14 | 1586 | 31 | 32 |
| Non-Core | *Caballeronia* | 0 | 0 | 0 | 0 | 0 | 0 | 33 | 18 |
| Non-Core | *Candidatus_Symbiopectobacterium* | 28 | 18 | 12 | 46 | 0 | 27 | 0 | 0 |
| Non-Core | *Capnocytophaga* | 11 | 16 | 0 | 0 | 0 | 0 | 20 | 12 |
| Non-Core | *Carnobacterium* | 0 | 0 | 14 | 105 | 9471 | 0 | 12 | 0 |
| Non-Core | *Caulobacter* | 16 | 25 | 10 | 80 | 37 | 0 | 85 | 49 |
| Non-Core | *Cellulomonas* | 0 | 532 | 53 | 2652 | 16 | 0 | 159 | 43 |
| Non-Core | *Cellulophaga* | 17 | 0 | 0 | 0 | 0 | 0 | 0 | 20 |
| Non-Core | *Cellulosimicrobium* | 0 | 133 | 0 | 0 | 0 | 0 | 42 | 0 |
| Non-Core | *Cereibacter* | 0 | 11 | 0 | 0 | 0 | 0 | 0 | 0 |
| Non-Core | *Chania* | 0 | 0 | 0 | 16 | 0 | 12 | 0 | 22 |
| Non-Core | *Chitinophaga* | 0 | 0 | 0 | 109 | 0 | 0 | 0 | 13 |
| Non-Core | *Chromobacterium* | 20 | 17 | 0 | 20 | 0 | 0 | 61 | 48 |
| Non-Core | *Cloacibacterium* | 125 | 309 | 0 | 52 | 0 | 0 | 154 | 352 |
| Non-Core | *Cnuibacter* | 0 | 10 | 0 | 0 | 0 | 0 | 13 | 13 |
| Non-Core | *Cobetia* | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 0 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Non-Core | *Cohnella* | 0 | 0 | 0 | 44 | 23 | 0 | 22 | 0 |
| Non-Core | *Collimonas* | 108 | 109 | 0 | 0 | 17 | 0 | 463 | 279 |
| Non-Core | *Comamonas* | 25 | 36 | 0 | 51686 | 14 | 0 | 79 | 118 |
| Non-Core | *Corynebacterium* | 0 | 48 | 0 | 39 | 0 | 0 | 37 | 23 |
| Non-Core | *Croceicoccus* | 0 | 0 | 0 | 0 | 0 | 0 | 15 | 14 |
| Non-Core | *Cruoricaptor* | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 10 |
| Non-Core | *Cryobacterium* | 21 | 172 | 13 | 0 | 13 | 0 | 116 | 61 |
| Non-Core | *Cupriavidus* | 292 | 447 | 27 | 304 | 83 | 0 | 2978 | 1395 |
| Non-Core | *Curvibacter* | 0 | 0 | 0 | 14 | 0 | 0 | 0 | 0 |
| Non-Core | *Cutibacterium* | 43 | 20 | 23 | 0 | 0 | 0 | 25 | 65 |
| Non-Core | *Dechloromonas* | 0 | 0 | 0 | 11 | 0 | 0 | 0 | 14 |
| Non-Core | *Delftia* | 48 | 38 | 19 | 32131 | 0 | 0 | 86 | 86 |
| Non-Core | *Denitratisoma* | 0 | 0 | 0 | 37 | 0 | 0 | 0 | 0 |
| Non-Core | *Dermacoccus* | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 0 |
| Non-Core | *Desulfohalobium* | 0 | 0 | 0 | 0 | 0 | 15 | 67 | 0 |
| Non-Core | *Devosia* | 0 | 20 | 0 | 0 | 0 | 0 | 17 | 0 |
| Non-Core | *Diaminobutyricimonas* | 0 | 16 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Non-Core | *Diaphorobacter* | 15 | 20 | 0 | 311 | 0 | 0 | 49 | 47 |
| Non-Core | *Dokdonella* | 33 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Non-Core | *Duganella* | 2235 | 2752 | 39 | 20 | 363 | 0 | 11722 | 2938 |
| Non-Core | *Dyella* | 18 | 0 | 0 | 76 | 0 | 0 | 27 | 14 |
| Non-Core | *Dysgonomonas* | 0 | 0 | 0 | 205 | 0 | 0 | 0 | 0 |
| Non-Core | *Echinicola* | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 0 |
| Non-Core | *Elizabethkingia* | 191 | 348 | 53 | 204 | 29 | 0 | 301 | 277 |
| Non-Core | *Empedobacter* | 0 | 28 | 17 | 26113 | 19511 | 0 | 32 | 16 |
| Non-Core | *Ensifer* | 32 | 88 | 0 | 15 | 0 | 0 | 47 | 37 |
| Non-Core | *Enterococcus* | 0 | 0 | 115967 | 751067 | 66701 | 579 | 53 | 13 |
| Non-Core | *Epilithonimonas* | 5918 | 9252 | 244 | 1117 | 541 | 0 | 7572 | 6461 |
| Non-Core | *Erythrobacter* | 0 | 10 | 0 | 0 | 0 | 0 | 27 | 13 |
| Non-Core | *Ewingella* | 57 | 22 | 19 | 30 | 0 | 16 | 15 | 129 |
| Non-Core | *Faecalibacter* | 0 | 0 | 0 | 218 | 111 | 0 | 0 | 0 |
| Non-Core | *Falsirhodobacter* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| Non-Core | *Ferribacterium* | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 0 |
| Non-Core | *Fictibacillus* | 0 | 0 | 0 | 0 | 0 | 12 | 0 | 0 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Non-Core | *Flammeovirga* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 18 |
| Non-Core | *Frateuria* | 15 | 0 | 0 | 44 | 0 | 0 | 15 | 0 |
| Non-Core | *Frondihabitans* | 10 | 118 | 10 | 0 | 127 | 0 | 98 | 35 |
| Non-Core | *Fusobacterium* | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 0 |
| Non-Core | *Georgenia* | 0 | 27 | 0 | 0 | 0 | 0 | 0 | 0 |
| Non-Core | *Gibbsiella* | 0 | 0 | 0 | 97 | 0 | 15 | 95 | 14 |
| Non-Core | *Glaciihabitans* | 0 | 16 | 0 | 0 | 0 | 0 | 19 | 0 |
| Non-Core | *Glaciimonas* | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 11 |
| Non-Core | *Glutamicibacter* | 0 | 0 | 0 | 22 | 0 | 0 | 0 | 0 |
| Non-Core | *Glycocaulis* | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 0 |
| Non-Core | *Gordonia* | 0 | 20 | 0 | 0 | 0 | 0 | 26 | 0 |
| Non-Core | *Gryllotalpicola* | 0 | 11 | 0 | 0 | 0 | 0 | 19 | 0 |
| Non-Core | *Haematobacter* | 0 | 0 | 0 | 12 | 0 | 0 | 0 | 0 |
| Non-Core | *Halanaerobium* | 12 | 10 | 0 | 0 | 0 | 0 | 0 | 0 |
| Non-Core | *Halomonas* | 25 | 33 | 392 | 80 | 28 | 0 | 107 | 122 |
| Non-Core | *Halopseudomonas* | 31 | 25 | 31 | 15 | 0 | 13 | 133 | 158 |
| Non-Core | *Halotalea* | 0 | 18 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Non-Core | *Herbaspirillum* | 206 | 242 | 0 | 14 | 86 | 0 | 989 | 728 |
| Non-Core | *Herbiconiux* | 14 | 101 | 0 | 0 | 0 | 0 | 56 | 21 |
| Non-Core | *Humibacter* | 0 | 30 | 0 | 0 | 0 | 0 | 26 | 0 |
| Non-Core | *Hydrogenophaga* | 58 | 47 | 0 | 129 | 37 | 0 | 84 | 80 |
| Non-Core | *Hymenobacter* | 223 | 302 | 110 | 0 | 435 | 16 | 2123 | 447 |
| Non-Core | *Isoptericola* | 0 | 38 | 0 | 0 | 0 | 0 | 10 | 0 |
| Non-Core | *Janthinobacterium* | 2324 | 4507 | 55 | 131 | 2588 | 0 | 9672 | 3704 |
| Non-Core | *Jejubacter* | 0 | 0 | 40 | 196 | 0 | 30 | 0 | 0 |
| Non-Core | *Jeongeupia* | 0 | 0 | 0 | 0 | 0 | 0 | 17 | 25 |
| Non-Core | *Jeotgalibaca* | 0 | 0 | 0 | 146 | 0 | 0 | 0 | 0 |
| Non-Core | *Kaistella* | 908 | 1697 | 56 | 109 | 80 | 0 | 1178 | 1389 |
| Non-Core | *Ketogulonicigenium* | 0 | 0 | 0 | 14 | 0 | 0 | 0 | 0 |
| Non-Core | *Kineococcus* | 0 | 10 | 0 | 0 | 0 | 0 | 0 | 0 |
| Non-Core | *Kingella* | 0 | 20 | 0 | 0 | 0 | 0 | 0 | 0 |
| Non-Core | *Kinneretia* | 13 | 15 | 0 | 0 | 0 | 0 | 32 | 20 |
| Non-Core | *Kocuria* | 0 | 17 | 0 | 0 | 15 | 0 | 25 | 0 |
| Non-Core | *Kordia* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Non-Core | *Lacinutrix* | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 0 |
| Non-Core | *Lacrimispora* | 0 | 0 | 3016 | 0 | 0 | 0 | 0 | 0 |
| Non-Core | *Lacticaseibacillus* | 0 | 0 | 0 | 84 | 0 | 0 | 0 | 0 |
| Non-Core | *Lactiplantibacillus* | 0 | 0 | 0 | 62 | 0 | 0 | 0 | 0 |
| Non-Core | *Lactobacillus* | 0 | 0 | 0 | 11 | 0 | 0 | 0 | 0 |
| Non-Core | *Lactococcus* | 0 | 0 | 0 | 772822 | 0 | 0 | 12 | 0 |
| Non-Core | *Laribacter* | 0 | 0 | 0 | 70 | 0 | 0 | 0 | 0 |
| Non-Core | *Latilactobacillus* | 0 | 0 | 10 | 1806 | 0 | 0 | 0 | 0 |
| Non-Core | *Lawsonella* | 0 | 17 | 0 | 0 | 0 | 0 | 0 | 0 |
| Non-Core | *Leeuwenhoekiella* | 0 | 0 | 0 | 0 | 0 | 0 | 20 | 0 |
| Non-Core | *Leifsonia* | 86 | 290 | 13 | 0 | 30 | 0 | 389 | 171 |
| Non-Core | *Lentilactobacillus* | 0 | 0 | 0 | 87 | 0 | 0 | 0 | 0 |
| Non-Core | *Leptospira* | 48 | 47 | 0 | 0 | 0 | 0 | 207 | 46 |
| Non-Core | *Leucobacter* | 13 | 84 | 24 | 733 | 0 | 0 | 115 | 32 |
| Non-Core | *Leuconostoc* | 0 | 0 | 139 | 12448 | 0 | 0 | 0 | 0 |
| Non-Core | *Levilactobacillus* | 0 | 0 | 0 | 137 | 0 | 0 | 0 | 0 |
| Non-Core | *Ligilactobacillus* | 0 | 0 | 0 | 65 | 0 | 0 | 0 | 0 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Non-Core | *Limnohabitans* | 0 | 0 | 0 | 40 | 0 | 0 | 0 | 0 |
| Non-Core | *Limosilactobacillus* | 0 | 0 | 0 | 95 | 0 | 0 | 0 | 0 |
| Non-Core | *Listeria* | 0 | 0 | 11 | 3197 | 0 | 11 | 10 | 0 |
| Non-Core | *Loigolactobacillus* | 0 | 0 | 0 | 12 | 0 | 0 | 0 | 0 |
| Non-Core | *Luteibacter* | 159 | 41 | 0 | 23 | 0 | 0 | 40 | 58 |
| Non-Core | *Luteimonas* | 540 | 1226 | 52 | 842 | 210 | 0 | 239 | 570 |
| Non-Core | *Lysinibacillus* | 0 | 0 | 0 | 0 | 0 | 0 | 19 | 0 |
| Non-Core | *Lysobacter* | 1400 | 1365 | 131 | 4615 | 105 | 0 | 641 | 209 |
| Non-Core | *Macrococcus* | 0 | 0 | 0 | 0 | 0 | 0 | 17 | 0 |
| Non-Core | *Maribacter* | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 0 |
| Non-Core | *Mariniflexile* | 0 | 0 | 0 | 10 | 0 | 0 | 0 | 0 |
| Non-Core | *Marinobacter* | 0 | 15 | 0 | 23 | 0 | 0 | 40 | 14 |
| Non-Core | *Marisediminicola* | 0 | 0 | 0 | 0 | 0 | 0 | 16 | 0 |
| Non-Core | *Martelella* | 0 | 34 | 0 | 15 | 0 | 0 | 13 | 0 |
| Non-Core | *Melaminivora* | 0 | 0 | 0 | 43 | 0 | 0 | 0 | 13 |
| Non-Core | *Mesorhizobium* | 14160 | 13100 | 17 | 37 | 18 | 0 | 66617 | 7181 |
| Non-Core | *Metakosakonia* | 11 | 0 | 0 | 1259 | 0 | 21 | 0 | 0 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Non-Core | *Methylobacterium* | 317 | 608 | 29 | 0 | 33 | 0 | 2722 | 257 |
| Non-Core | *Methylorubrum* | 731 | 670 | 712 | 0 | 1259 | 12 | 4213 | 441 |
| Non-Core | *Methylosinus* | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 0 |
| Non-Core | *Microbulbifer* | 15 | 0 | 0 | 0 | 0 | 0 | 17 | 0 |
| Non-Core | *Microcella* | 0 | 38 | 0 | 0 | 0 | 0 | 30 | 0 |
| Non-Core | *Micrococcus* | 0 | 18 | 0 | 0 | 0 | 0 | 40 | 0 |
| Non-Core | *Microlunatus* | 0 | 0 | 0 | 0 | 0 | 0 | 18 | 0 |
| Non-Core | *Micromonospora* | 0 | 0 | 0 | 0 | 0 | 0 | 13 | 0 |
| Non-Core | *Micropruina* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 |
| Non-Core | *Microterricola* | 0 | 64 | 0 | 0 | 0 | 0 | 47 | 23 |
| Non-Core | *Mongoliitalea* | 0 | 0 | 0 | 0 | 0 | 0 | 17 | 0 |
| Non-Core | *Moraxella* | 0 | 0 | 0 | 35 | 0 | 0 | 0 | 0 |
| Non-Core | *Mucilaginibacter* | 96 | 129 | 0 | 20 | 0 | 0 | 200 | 452 |
| Non-Core | *Musicola* | 0 | 0 | 0 | 15 | 0 | 0 | 0 | 16 |
| Non-Core | *Mycetocola* | 10 | 137 | 12 | 0 | 0 | 0 | 89 | 18 |
| Non-Core | *Mycobacterium* | 0 | 17 | 0 | 0 | 0 | 0 | 22 | 11 |
| Non-Core | *Mycobacteroides* | 0 | 0 | 16 | 0 | 84 | 0 | 57 | 0 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Non-Core | *Mycolicibacterium* | 27 | 61 | 0 | 10 | 0 | 0 | 157 | 43 |
| Non-Core | *Myroides* | 0 | 0 | 0 | 9893 | 79 | 0 | 15 | 13 |
| Non-Core | *Neisseria* | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 0 |
| Non-Core | *Neorhizobium* | 288 | 2138 | 24 | 344 | 10 | 0 | 522 | 544 |
| Non-Core | *Niallia* | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 0 |
| Non-Core | *Niastella* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 |
| Non-Core | *Nitratireductor* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| Non-Core | *Nitrobacter* | 0 | 0 | 0 | 0 | 0 | 0 | 17 | 0 |
| Non-Core | *Nitrosomonas* | 0 | 0 | 0 | 14 | 0 | 0 | 0 | 0 |
| Non-Core | *Niveispirillum* | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 0 |
| Non-Core | *Nocardia* | 0 | 23 | 0 | 0 | 0 | 0 | 10 | 0 |
| Non-Core | *Nocardioides* | 0 | 60 | 0 | 0 | 12 | 0 | 105 | 21 |
| Non-Core | *Nocardiopsis* | 0 | 17 | 0 | 0 | 0 | 0 | 0 | 0 |
| Non-Core | *Noviherbaspirillum* | 13 | 12 | 0 | 0 | 11 | 0 | 60 | 62 |
| Non-Core | *Novosphingobium* | 40 | 53 | 11 | 643 | 26 | 0 | 167 | 94 |
| Non-Core | *Obesumbacterium* | 0 | 0 | 0 | 17 | 10 | 0 | 26 | 0 |
| Non-Core | *Ochrobactrum* | 0 | 49 | 0 | 121 | 0 | 0 | 0 | 0 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Non-Core | *Oerskovia* | 0 | 53 | 0 | 57 | 0 | 0 | 12 | 0 |
| Non-Core | *Olivibacter* | 0 | 0 | 0 | 133 | 0 | 0 | 10 | 0 |
| Non-Core | *Orrella* | 11 | 0 | 0 | 15 | 0 | 0 | 31 | 19 |
| Non-Core | *Ottowia* | 0 | 0 | 0 | 27 | 0 | 0 | 19 | 10 |
| Non-Core | *Paenarthrobacter* | 13 | 19 | 0 | 0 | 0 | 0 | 344 | 0 |
| Non-Core | *Paludibacterium* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 |
| Non-Core | *Pandoraea* | 44 | 50 | 0 | 48 | 12 | 0 | 132 | 158 |
| Non-Core | *Pannonibacter* | 0 | 14 | 0 | 0 | 0 | 0 | 28 | 0 |
| Non-Core | *Parabacteroides* | 0 | 0 | 0 | 0 | 0 | 0 | 32 | 0 |
| Non-Core | *Paracoccus* | 0 | 28 | 0 | 38 | 12 | 25 | 34 | 219 |
| Non-Core | *Paraliobacillus* | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 0 |
| Non-Core | *Paraoerskovia* | 0 | 25 | 0 | 0 | 0 | 0 | 0 | 0 |
| Non-Core | *Pasteurella* | 0 | 0 | 0 | 0 | 53 | 0 | 21 | 25 |
| Non-Core | *Pediococcus* | 0 | 0 | 0 | 24 | 0 | 0 | 0 | 0 |
| Non-Core | *Pedobacter* | 56220 | 65836 | 1859 | 52 | 908 | 0 | 65229 | 121195 |
| Non-Core | *Peribacillus* | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 0 |
| Non-Core | *Peteryoungia* | 0 | 15 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Non-Core | *Photobacterium* | 0 | 0 | 0 | 24 | 0 | 0 | 0 | 0 |
| Non-Core | *Photorhabdus* | 0 | 0 | 0 | 10 | 0 | 0 | 16 | 0 |
| Non-Core | *Phyllobacterium* | 0 | 18 | 0 | 0 | 0 | 0 | 0 | 0 |
| Non-Core | *Phytobacter* | 0 | 21 | 26 | 3733 | 0 | 260 | 78 | 33 |
| Non-Core | *Pigmentiphaga* | 1295 | 5903 | 39 | 51 | 0 | 0 | 1445 | 483 |
| Non-Core | *Pistricoccus* | 0 | 0 | 0 | 10 | 0 | 0 | 0 | 0 |
| Non-Core | *Planococcus* | 0 | 0 | 0 | 11 | 0 | 0 | 32 | 0 |
| Non-Core | *Plantibacter* | 16112 | 116189 | 3386 | 1000 | 1900 | 0 | 205109 | 52319 |
| Non-Core | *Plesiomonas* | 0 | 0 | 0 | 29 | 0 | 0 | 0 | 0 |
| Non-Core | *Polaribacter* | 13 | 27 | 0 | 15 | 0 | 0 | 25 | 40 |
| Non-Core | *Polaromonas* | 0 | 15 | 0 | 24 | 0 | 0 | 41 | 43 |
| Non-Core | *Polymorphobacter* | 0 | 0 | 0 | 0 | 0 | 0 | 26 | 16 |
| Non-Core | *Polynucleobacter* | 0 | 0 | 0 | 17 | 0 | 0 | 0 | 13 |
| Non-Core | *Porphyrobacter* | 0 | 0 | 0 | 0 | 15 | 0 | 11 | 0 |
| Non-Core | *Prevotella* | 0 | 0 | 0 | 0 | 0 | 0 | 15 | 21 |
| Non-Core | *Priestia* | 0 | 0 | 0 | 0 | 0 | 0 | 24 | 0 |
| Non-Core | *Protaetiibacter* | 0 | 46 | 0 | 0 | 10 | 0 | 63 | 21 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Non-Core | *Proteus* | 0 | 0 | 30 | 1360 | 12 | 33 | 0 | 473 |
| Non-Core | *Providencia* | 0 | 27 | 16 | 75 | 38 | 85 | 24 | 30 |
| Non-Core | *Pseudarthrobacter* | 10917 | 9440 | 0 | 0 | 0 | 0 | 278952 | 6407 |
| Non-Core | *Pseudescherichia* | 0 | 0 | 0 | 3560 | 0 | 104 | 0 | 0 |
| Non-Core | *Pseudoalteromonas* | 0 | 0 | 0 | 16 | 0 | 234 | 10 | 327 |
| Non-Core | *Pseudochrobactrum* | 0 | 0 | 0 | 156 | 0 | 0 | 0 | 0 |
| Non-Core | *Pseudocitrobacter* | 31 | 0 | 69 | 867 | 35 | 73 | 23 | 10 |
| Non-Core | *Pseudogulbenkiania* | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 0 |
| Non-Core | *Pseudolabrys* | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 0 |
| Non-Core | *Pseudolysobacter* | 21 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Non-Core | *Pseudonocardia* | 0 | 0 | 0 | 0 | 0 | 0 | 22 | 0 |
| Non-Core | *Pseudopedobacter* | 10 | 0 | 0 | 116 | 0 | 0 | 10 | 14 |
| Non-Core | *Pseudorhizobium* | 12 | 103 | 0 | 0 | 0 | 0 | 12 | 16 |
| Non-Core | *Pseudoxanthomonas* | 575 | 157 | 56 | 2204 | 0 | 0 | 197 | 57 |
| Non-Core | *Psychrobacter* | 11 | 5407 | 5148 | 0 | 160 | 64 | 18 | 468 |
| Non-Core | *Pulveribacter* | 0 | 0 | 0 | 21 | 0 | 0 | 0 | 0 |
| Non-Core | *Pusillimonas* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Non-Core | *Qipengyuania* | 0 | 0 | 0 | 0 | 0 | 0 | 14 | 0 |
| Non-Core | *Ramlibacter* | 0 | 0 | 0 | 0 | 0 | 0 | 16 | 0 |
| Non-Core | *Rathayibacter* | 2146 | 21424 | 1079 | 51 | 2033 | 0 | 13398 | 2946 |
| Non-Core | *Rhizobium* | 869 | 4697 | 73 | 703 | 23 | 0 | 1016 | 1041 |
| Non-Core | *Rhizorhabdus* | 17 | 20 | 0 | 35 | 0 | 0 | 125 | 22 |
| Non-Core | *Rhodanobacter* | 43 | 17 | 0 | 183 | 0 | 0 | 17 | 0 |
| Non-Core | *Rhodococcus* | 1002 | 44552 | 8657 | 148 | 10326 | 0 | 690 | 224 |
| Non-Core | *Rhodoferax* | 21 | 30 | 0 | 131 | 10 | 0 | 88 | 82 |
| Non-Core | *Rhodoplanes* | 0 | 0 | 0 | 0 | 0 | 0 | 26 | 0 |
| Non-Core | *Rhodopseudomonas* | 39 | 50 | 0 | 0 | 0 | 0 | 258 | 53 |
| Non-Core | *Riemerella* | 0 | 19 | 0 | 0 | 0 | 0 | 27 | 13 |
| Non-Core | *Roseomonas* | 10 | 399 | 0 | 0 | 0 | 0 | 70 | 26 |
| Non-Core | *Rouxiella* | 12 | 0 | 0 | 79 | 22 | 11 | 34 | 22 |
| Non-Core | *Ruania* | 0 | 22 | 0 | 0 | 0 | 0 | 11 | 0 |
| Non-Core | *Runella* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 |
| Non-Core | *Saccharibacillus* | 0 | 452 | 120747 | 11621 | 2169 | 99307 | 219 | 1030 |
| Non-Core | *Salegentibacter* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Non-Core | *Salinibacterium* | 0 | 73 | 0 | 0 | 11 | 0 | 62 | 25 |
| Non-Core | *Sanguibacter* | 0 | 30900 | 2723 | 59 | 443 | 0 | 3928 | 404 |
| Non-Core | *Scandinavium* | 0 | 0 | 17 | 379 | 0 | 57 | 10 | 21 |
| Non-Core | *Serpentinimonas* | 0 | 140 | 0 | 63 | 0 | 0 | 0 | 0 |
| Non-Core | *Shewanella* | 0 | 10 | 31 | 33 | 18 | 1048 | 38 | 315 |
| Non-Core | *Shigella* | 0 | 0 | 0 | 33 | 0 | 0 | 39 | 0 |
| Non-Core | *Shimwellia* | 51 | 0 | 0 | 144 | 41 | 45 | 20 | 24 |
| Non-Core | *Shinella* | 19 | 60 | 0 | 43 | 0 | 0 | 70 | 31 |
| Non-Core | *Simplicispira* | 0 | 0 | 0 | 11 | 0 | 0 | 0 | 11 |
| Non-Core | *Sinomonas* | 0 | 0 | 0 | 0 | 0 | 0 | 22 | 0 |
| Non-Core | *Sinorhizobium* | 15 | 85 | 0 | 25 | 0 | 0 | 28 | 33 |
| Non-Core | *Sodalis* | 0 | 0 | 14 | 109 | 219 | 31 | 31 | 27 |
| Non-Core | *Solimonas* | 0 | 0 | 0 | 0 | 0 | 0 | 15 | 0 |
| Non-Core | *Solitalea* | 0 | 12 | 0 | 0 | 0 | 0 | 11 | 15 |
| Non-Core | *Sphingobacterium* | 429 | 234 | 0 | 201486 | 1533 | 0 | 832 | 348 |
| Non-Core | *Sphingobium* | 216 | 273 | 38 | 421 | 127 | 0 | 1245 | 336 |
| Non-Core | *Sphingopyxis* | 51 | 104 | 0 | 14 | 32 | 0 | 404 | 114 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Non-Core | *Sphingorhabdus* | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 0 |
| Non-Core | *Sphingosinicella* | 0 | 0 | 0 | 0 | 0 | 0 | 40 | 23 |
| Non-Core | *Sphingosinithalassobacter* | 14 | 18 | 0 | 0 | 13 | 0 | 144 | 43 |
| Non-Core | *Spirosoma* | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 0 |
| Non-Core | *Spongiibacter* | 0 | 0 | 0 | 0 | 0 | 0 | 66 | 0 |
| Non-Core | *Sporosarcina* | 0 | 0 | 0 | 0 | 0 | 0 | 24 | 0 |
| Non-Core | *Staphylococcus* | 13 | 0 | 12 | 240 | 26 | 18 | 491 | 27 |
| Non-Core | *Streptococcus* | 0 | 0 | 0 | 3919 | 386 | 15 | 39 | 10 |
| Non-Core | *Streptomyces* | 77 | 435 | 66 | 60 | 123 | 0 | 630 | 214 |
| Non-Core | *Subtercola* | 15 | 60 | 0 | 0 | 0 | 0 | 54 | 23 |
| Non-Core | *Sulfuriferula* | 0 | 0 | 0 | 11 | 0 | 0 | 28 | 14 |
| Non-Core | *Sulfuritortus* | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 0 |
| Non-Core | *Swingsia* | 0 | 0 | 0 | 0 | 0 | 12 | 0 | 0 |
| Non-Core | *Tardibacter* | 0 | 0 | 0 | 0 | 0 | 0 | 15 | 0 |
| Non-Core | *Tardiphaga* | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 0 |
| Non-Core | *Tenacibaculum* | 0 | 10 | 0 | 0 | 0 | 0 | 0 | 15 |
| Non-Core | *Tessaracoccus* | 0 | 0 | 0 | 0 | 0 | 0 | 25 | 0 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Non-Core | *Tetragenococcus* | 0 | 0 | 0 | 35 | 0 | 0 | 0 | 0 |
| Non-Core | *Thalassospira* | 0 | 0 | 0 | 33 | 0 | 0 | 0 | 0 |
| Non-Core | *Thauera* | 0 | 0 | 16 | 31 | 0 | 0 | 23 | 21 |
| Non-Core | *Thermodesulfobium* | 0 | 0 | 0 | 0 | 0 | 0 | 18 | 0 |
| Non-Core | *Thermomonas* | 279 | 85 | 18 | 818 | 0 | 0 | 70 | 20 |
| Non-Core | *Thiopseudomonas* | 0 | 0 | 0 | 13 | 0 | 0 | 0 | 0 |
| Non-Core | *Tsuneonella* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 |
| Non-Core | *Undibacterium* | 22 | 22 | 0 | 0 | 13 | 0 | 72 | 134 |
| Non-Core | *Usitatibacter* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 13 |
| Non-Core | *Vagococcus* | 0 | 0 | 10 | 138 | 0 | 0 | 0 | 0 |
| Non-Core | *Variovorax* | 1512 | 1513 | 18 | 142 | 37 | 0 | 6576 | 1695 |
| Non-Core | *Verminephrobacter* | 0 | 0 | 0 | 147 | 0 | 0 | 17 | 16 |
| Non-Core | *Vitreoscilla* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 |
| Non-Core | *Weeksella* | 0 | 0 | 0 | 21 | 0 | 0 | 0 | 0 |
| Non-Core | *Weissella* | 0 | 0 | 0 | 6560 | 0 | 0 | 0 | 0 |
| Non-Core | *Wenzhouxiangella* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| Non-Core | *Winogradskyella* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 13 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Non-Core | *Xenorhabdus* | 0 | 0 | 16 | 27 | 15 | 135 | 0 | 15 |
| Non-Core | *Xylella* | 0 | 0 | 0 | 52 | 0 | 0 | 0 | 0 |
| Non-Core | *Yokenella* | 10 | 0 | 0 | 424 | 0 | 49 | 19 | 0 |
| Non-Core | *Zhihengliuella* | 0 | 0 | 0 | 0 | 0 | 0 | 118 | 0 |
| Non-Core | *Zunongwangia* | 0 | 0 | 0 | 0 | 0 | 0 | 17 | 13 |
| Non-Core | *Zymobacter* | 0 | 0 | 0 | 0 | 0 | 0 | 51 | 0 |

Table S2.6 Distribution Phages genera.

| **Genus** | **HL** | **HH** | **S3H** | **S3L** | **S6H** | **S6L** | **S9H** | **S9L** |
|---|---|---|---|---|---|---|---|---|
| *Aarhusvirus* | 0 | 0 | 0 | 325 | 0 | 0 | 40 | 0 |
| *Agricanvirus* | 0 | 0 | 0 | 0 | 0 | 0 | 1174 | 0 |
| *Asteriusvirus* | 0 | 0 | 0 | 0 | 0 | 11 | 10 | 0 |
| *Cbunavirus* | 0 | 0 | 0 | 90 | 0 | 331 | 0 | 0 |
| *Ceduovirus* | 0 | 0 | 0 | 0 | 0 | 82 | 0 | 0 |
| *Certrevirus* | 1188 | 0 | 0 | 40 | 71 | 0 | 11 | 387 |
| *Cornellvirus* | 0 | 0 | 0 | 0 | 0 | 30199 | 0 | 0 |
| *Dhakavirus* | 0 | 0 | 0 | 0 | 0 | 0 | 80 | 0 |
| *Dibbivirus* | 0 | 0 | 0 | 0 | 0 | 0 | 23 | 0 |
| *Eneladusvirus* | 0 | 91 | 0 | 0 | 0 | 1442 | 598 | 0 |
| *Feofaniavirus* | 0 | 0 | 2073 | 3154 | 0 | 0 | 2874 | 800 |
| *Flaumdravirus* | 0 | 21 | 0 | 0 | 0 | 0 | 0 | 13 |
| *Gelderlandvirus* | 0 | 0 | 0 | 0 | 0 | 0 | 40 | 0 |
| *Ghunavirus* | 0 | 0 | 0 | 0 | 0 | 0 | 78 | 213 |
| *Henunavirus* | 0 | 0 | 0 | 0 | 77 | 0 | 244 | 0 |
| *Ithacavirus* | 0 | 0 | 0 | 0 | 0 | 212 | 0 | 0 |
| *Jedunavirus* | 0 | 0 | 0 | 0 | 0 | 1817 | 0 | 0 |
| *Jerseyvirus* | 0 | 0 | 0 | 0 | 0 | 176 | 0 | 0 |
| *Jiaodavirus* | 0 | 0 | 0 | 0 | 0 | 0 | 1887 | 0 |
| *Johnsonvirus* | 0 | 0 | 0 | 0 | 10 | 0 | 0 | 0 |
| *Kagunavirus* | 0 | 0 | 0 | 0 | 0 | 2094 | 0 | 0 |
| *Kanagawavirus* | 0 | 0 | 0 | 0 | 0 | 0 | 3762 | 0 |
| *Karamvirus* | 0 | 0 | 0 | 0 | 0 | 0 | 158 | 0 |
| *Kuttervirus* | 0 | 0 | 0 | 0 | 0 | 28 | 0 | 0 |
| *Littlefixvirus* | 0 | 0 | 0 | 0 | 0 | 0 | 21 | 0 |
| *Loessnervirus* | 0 | 0 | 0 | 0 | 0 | 587 | 2871 | 0 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *Marfavirus* | 0 | 0 | 0 | 0 | 0 | 0 | 50 | 0 |
| *Mimasvirus* | 0 | 279 | 0 | 10 | 0 | 4259 | 1251 | 0 |
| *Moabitevirus* | 0 | 0 | 0 | 0 | 0 | 23 | 0 | 0 |
| *Moonvirus* | 0 | 0 | 0 | 0 | 0 | 0 | 987 | 0 |
| *Mosigvirus* | 0 | 0 | 0 | 0 | 0 | 0 | 4244 | 0 |
| *Mydovirus* | 0 | 0 | 0 | 0 | 13 | 411 | 424 | 0 |
| *Naesvirus* | 0 | 0 | 30 | 0 | 0 | 140 | 0 | 0 |
| *Nochtlivirus* | 0 | 0 | 0 | 0 | 0 | 13077 | 0 | 0 |
| *Otagovirus* | 30 | 0 | 0 | 38 | 0 | 0 | 0 | 29 |
| *Peatvirus* | 0 | 0 | 0 | 0 | 0 | 2263 | 0 | 0 |
| *Petsuvirus* | 15 | 11 | 0 | 0 | 3891 | 16 | 125 | 80 |
| *Phutvirus* | 0 | 13 | 0 | 0 | 0 | 0 | 0 | 12 |
| *Pifdecavirus* | 0 | 106 | 0 | 0 | 0 | 23 | 0 | 48 |
| *Plaisancevirus* | 0 | 0 | 0 | 0 | 0 | 23 | 0 | 0 |
| *Pseudotevenvirus* | 0 | 0 | 0 | 0 | 0 | 0 | 386 | 0 |
| *Pylasvirus* | 0 | 0 | 0 | 0 | 0 | 19 | 0 | 0 |
| *Ravinvirus* | 0 | 0 | 0 | 0 | 4749 | 95 | 0 | 0 |
| *Risingsunvirus* | 0 | 0 | 0 | 0 | 29 | 0 | 0 | 0 |
| *Riverridervirus* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 21 |
| *Sauletekiovirus* | 0 | 0 | 0 | 0 | 4960 | 0 | 0 | 0 |
| *Seunavirus* | 0 | 0 | 0 | 0 | 726 | 554 | 1628 | 0 |
| *Shizishanvirus* | 0 | 0 | 0 | 0 | 0 | 0 | 59 | 0 |
| *Skunavirus* | 0 | 0 | 0 | 0 | 0 | 91068 | 0 | 0 |
| *Sugarlandvirus* | 0 | 0 | 0 | 0 | 0 | 0 | 21 | 0 |
| *Tegunavirus* | 0 | 0 | 0 | 0 | 0 | 0 | 56 | 0 |
| *Tequatrovirus* | 0 | 0 | 0 | 0 | 0 | 0 | 23 | 0 |
| *Uetakevirus* | 0 | 0 | 0 | 0 | 0 | 23 | 0 | 0 |
| *Uliginvirus* | 0 | 0 | 0 | 39 | 0 | 92 | 69 | 1490 |
| *Vedamuthuvirus* | 0 | 0 | 0 | 0 | 0 | 673 | 0 | 0 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *Waedenswilvirus* | 0 | 0 | 0 | 0 | 0 | 16 | 0 | 0 |
| *Winklervirus* | 0 | 0 | 0 | 0 | 0 | 0 | 123 | 0 |
| *Yonginvirus* | 0 | 0 | 0 | 0 | 31 | 0 | 0 | 0 |
| *Zurivirus* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 151 |

Table S2.7 Distribution of fungal genera.

| **Genus** | **S3L** | **S6L** | **S9L** |
|---|---|---|---|
| *Aspergillus* | 3232 | 0 | 912 |
| *Botrytis* | 0 | 0 | 3342 |
| *Candida* | 3981 | 0 | 1281 |
| *Cercospora* | 0 | 0 | 123 |
| *Cryptococcus* | 0 | 0 | 7532 |
| *Eremothecium* | 0 | 7030 | 0 |
| *Fusarium* | 7872 | 0 | 4150 |
| *Marasmius* | 0 | 0 | 647 |
| *Rhizoctonia* | 0 | 0 | 222 |
| *Thermothelomyces* | 2522 | 0 | 2263 |
| *Ustilaginoidea* | 1445 | 0 | 425 |

# Chapter 3 Appendix

Table S3.2 Detailed assembly and annotation metrics for barley seed microbiome metagenome-assembled genomes (MAGs).

| Sample id | Nitrogen-content | Bin-name | Completeness (%) | Contamination (%) | Total length | GC (%) | N50 | Phylum | Genus | Species |
|---|---|---|---|---|---|---|---|---|---|---|
| T2LN | Low | MAG58-bin2 | 97.77 | 0.0 | 1865236 | 38.18 | 22371 | Firmicutes | *Lactococcus* | *Lactococcus garvieae_A* |
| T2LN | Low | MAG61-bin5 | 100.0 | 0.757 | 2260998 | 34.98 | 118370 | Firmicutes | *Lactococcus* | *Lactococcus lactis* |
| T2LN | Low | MAG60-bin4 | 85.81 | 1.556 | 2451992 | 43.03 | 4356 | Pseudomonadota | *Acinetobacter* | *Acinetobacter variabilis* |
| T3LN | Low | MAG75-bin8 | 97.07 | 0.436 | 2648119 | 37.53 | 25727 | Firmicutes | *Enterococcus* | *Enterococcus faecalis* |
| HHN | High | MAG20-bin18 | 81.89 | 0.118 | 2856754 | 39.91 | 10588 | Bacteroidota | *Epilithonimonas* | |
| T3HN | High | MAG82-bin8 | 91.42 | 0.754 | 3039522 | 42.99 | 5647 | Firmicutes | *Enterococcus_D* | *Enterococcus_D casseliflavus* |
| HLN | Low | MAG4-bin16 | 91.01 | 0.490 | 3133833 | 39.55 | 25933 | Bacteroidota | *Epilithonimonas* | |
| T3HN | High | MAG80-bin6 | 84.23 | 4.812 | 3192441 | 71.33 | 8733 | Actinobacteriota | *Curtobacterium* | *Curtobacterium flaccumfaciens_A* |
| HLN | Low | MAG8-bin2 | 87.16 | 1.607 | 3239357 | 39.87 | 4923 | Bacteroidota | *Flavobacterium* | |
| T2LN | Low | MAG59-bin3 | 97.35 | 0.0 | 3287730 | 43.14 | 48287 | Firmicutes | *Enterococcus_D* | *Enterococcus_D casseliflavus* |
| T1HN | High | MAG49-bin2 | 82.21 | 3.420 | 3297731 | 69.80 | 4591 | Actinobacteriota | *Plantibacter* | *Plantibacter sp015351235* |
| T1LN | Low | MAG38-bin21 | 90.62 | 2.247 | 3411048 | 69.06 | 17946 | Actinobacteriota | *Microbacterium* | |
| HHN | High | MAG18-bin15 | 88.12 | 2.777 | 3417083 | 70.06 | 17891 | Actinobacteriota | *Plantibacter* | *Plantibacter sp015351235* |
| HHN | High | MAG23-bin3 | 82.20 | 2.859 | 3455050 | 34.81 | 22649 | Bacteroidota | *Chryseobacterium* | |
| T1LN | Low | MAG44-bin9 | 92.45 | 0.691 | 3522157 | 66.62 | 25013 | Pseudomonadota | *Stenotrophomonas* | |
| T1HN | High | MAG47-bin12 | 82.82 | 0.122 | 3538543 | 34.27 | 7567 | Bacteroidota | *Chryseobacterium* | *Chryseobacterium indoltheticum* |
| T1LN | Low | MAG29-bin10 | 99.50 | 0.490 | 3600078 | 39.11 | 54766 | Bacteroidota | *Epilithonimonas* | |
| T3HN | High | MAG81-bin7 | 89.08 | 0.613 | 3601193 | 56.28 | 13890 | Pseudomonadota | *Pantoea* | *Pantoea agglomerans* |
| T1LN | Low | MAG35-bin18 | 75.86 | 7.108 | 3617542 | 69.26 | 13948 | Actinobacteriota | *Plantibacter* | *Plantibacter cousiniae* |
| T1HN | High | MAG51-bin4 | 100.0 | 0.0 | 3740955 | 38.87 | 267888 | Bacteroidota | *Epilithonimonas* | |
| T1HN | High | MAG46-bin11 | 85.03 | 1.051 | 3743315 | 68.29 | 6415 | Pseudomonadota | *Xanthomonas_A* | *Xanthomonas_A translucens* |
| T1LN | Low | MAG41-bin4 | 96.14 | 0.505 | 3754105 | 67.87 | 37408 | Actinobacteriota | *Microbacterium* | |
| T1LN | Low | MAG30-bin11 | 89.66 | 0.983 | 3764179 | 56.22 | 13016 | Pseudomonadota | *Pantoea* | *Pantoea agglomerans* |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| T3LN | Low | MAG69-bin12 | 79.22 | 2.030 | 3798932 | 39.59 | 6488 | Firmicutes | *Paenibacillus_J* | *Paenibacillus_J nuruki* |
| T3HN | High | MAG76-bin2 | 83.82 | 1.133 | 3804786 | 64.81 | 10259 | Pseudomonadota | *Massilia* | *Massilia aurea_B* |
| T1LN | Low | MAG32-bin14 | 84.13 | 1.793 | 3821329 | 68.39 | 8412 | Pseudomonadota | *Xanthomonas_A* | *Xanthomonas_A translucens* |
| T1HN | High | MAG53-bin6 | 83.73 | 0.133 | 3838592 | 65.23 | 4645 | Pseudomonadota | *Massilia* | *Massilia aurea_B* |
| HLN | Low | MAG9-bin3 | 93.29 | 1.505 | 3852006 | 65.94 | 42222 | Pseudomonadota | *Stenotrophomonas* | *Stenotrophomonas sp014621775* |
| HHN | High | MAG24-bin5 | 94.15 | 0.343 | 3865160 | 34.12 | 108584 | Bacteroidota | *Chryseobacterium* | *Chryseobacterium indoltheticum* |
| HLN | Low | MAG7-bin19 | 95.69 | 1.613 | 3885098 | 66.41 | 43286 | Pseudomonadota | *Stenotrophomonas* | *Stenotrophomonas sp002836635* |
| HHN | High | MAG19-bin16 | 82.14 | 0.286 | 3924242 | 64.95 | 8808 | Pseudomonadota | *Massilia* | *Massilia aurea_B* |
| HHN | High | MAG15-bin10 | 98.03 | 0.245 | 3979189 | 55.75 | 51484 | Pseudomonadota | *Duffyella* | *Duffyella gerundensis* |
| HLN | Low | MAG2-bin12 | 89.81 | 3.073 | 4008689 | 55.44 | 30922 | Pseudomonadota | *Pantoea* | *Pantoea sp002419935* |
| T3HN | High | MAG77-bin3 | 96.61 | 1.136 | 4014262 | 56.04 | 37421 | Pseudomonadota | *Duffyella* | *Duffyella gerundensis* |
| T2HN | High | MAG66-bin4 | 97.18 | 0.245 | 4021096 | 55.43 | 51456 | Pseudomonadota | *Duffyella* | *Duffyella gerundensis* |
| HHN | High | MAG16-bin11 | 99.64 | 0.330 | 4054845 | 65.80 | 72028 | Pseudomonadota | *Stenotrophomonas* | *Stenotrophomonas sp014621775* |
| T1LN | Low | MAG31-bin13 | 96.93 | 0.260 | 4069491 | 55.76 | 35365 | Pseudomonadota | *Duffyella* | *Duffyella gerundensis* |
| HHN | High | MAG17-bin12 | 94.33 | 1.830 | 4070944 | 70.83 | 20846 | Actinobacteriota | *Sanguibacter* | *Sanguibacter inulinus* |
| HLN | Low | MAG10-bin5 | 97.17 | 0.327 | 4080115 | 55.66 | 62804 | Pseudomonadota | *Duffyella* | *Duffyella gerundensis* |
| HLN | Low | MAG13-bin9 | 87.34 | 3.284 | 4097961 | 34.00 | 25722 | Bacteroidota | *Chryseobacterium* | *Chryseobacterium indoltheticum* |
| T3LN | Low | MAG73-bin4 | 96.78 | 0.714 | 4105162 | 55.65 | 57133 | Pseudomonadota | *Duffyella* | *Duffyella gerundensis* |
| T3LN | Low | MAG68-bin11 | 83.25 | 2.143 | 4110094 | 34.99 | 5938 | Firmicutes | *Bacillus_A* | *Bacillus_A bombysepticus* |
| T1LN | Low | MAG37-bin20 | 83.84 | 1.133 | 4150027 | 64.71 | 14133 | Pseudomonadota | *Massilia* | *Massilia aurea_B* |
| HHN | High | MAG25-bin6 | 89.86 | 1.180 | 4180807 | 60.69 | 23553 | Pseudomonadota | *Neorhizobium* | *Neorhizobium soli* |
| T1HN | High | MAG56-bin9 | 97.60 | 0.309 | 4181855 | 55.70 | 35876 | Pseudomonadota | *Duffyella* | *Duffyella gerundensis* |
| T1HN | High | MAG50-bin3 | 97.87 | 1.102 | 4201248 | 39.58 | 49864 | Bacteroidota | *Chryseobacterium* | *Chryseobacterium sp001421435* |
| HLN | Low | MAG1-bin1 | 76.94 | 1.916 | 4216299 | 64.66 | 9399 | Pseudomonadota | *Massilia* | *Massilia aurea_B* |
| T1LN | Low | MAG42-bin5 | 99.12 | 0.0 | 4223639 | 66.34 | 38378 | Actinobacteriota | *Pseudarthrobacter* | *Pseudarthrobacter equi* |
| T2LN | Low | MAG57-bin1 | 100.0 | 0.0 | 4332316 | 33.55 | 117162 | Bacteroidota | *Chryseobacterium* | *Chryseobacterium piscium* |
| T3HN | High | MAG79-bin5 | 80.17 | 0.0 | 4502603 | 55.98 | 82812 | Pseudomonadota | *Erwinia* | *Erwinia persicina* |
| T2HN | High | MAG63-bin1 | 96.57 | 0.519 | 4512341 | 55.87 | 96926 | Pseudomonadota | *Erwinia* | *Erwinia persicina* |
| T1LN | Low | MAG36-bin2 | 98.86 | 0.0 | 4617697 | 39.64 | 80281 | Firmicutes | *Paenibacillus_J* | *Paenibacillus_J nuruki* |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| T3LN | Low | MAG74-bin7 | 98.68 | 0.273 | 4634138 | 55.65 | 91004 | Pseudomonadota | *Erwinia* | *Erwinia persicina* |
| HLN | Low | MAG5-bin17 | 98.08 | 0.0 | 4640035 | 39.66 | 37393 | Firmicutes | *Paenibacillus_J* | *Paenibacillus_J nuruki* |
| T3LN | Low | MAG72-bin2 | 77.19 | 0.0 | 4681908 | 63.75 | 26823 | Pseudomonadota | *Pseudomonas_E* | *Pseudomonas_E sp003253735* |
| HHN | High | MAG27-bin9 | 99.31 | 0.069 | 4717584 | 39.64 | 57215 | Firmicutes | *Paenibacillus_J* | *Paenibacillus_J nuruki* |
| T3HN | High | MAG78-bin4 | 96.90 | 1.716 | 4740412 | 39.44 | 19164 | Firmicutes | *Paenibacillus_J* | *Paenibacillus_J nuruki* |
| T1LN | Low | MAG28-bin1 | 96.90 | 0.860 | 4768914 | 55.97 | 64761 | Pseudomonadota | *Erwinia* | *Erwinia persicina* |
| T1LN | Low | MAG45-bin10 | 97.93 | 1.170 | 4792467 | 55.28 | 59038 | Pseudomonadota | *Pantoea* | *Pantoea agglomerans* |
| HHN | High | MAG22-bin2 | 97.30 | 0.338 | 4805805 | 55.63 | 152377 | Pseudomonadota | *Erwinia* | *Erwinia persicina* |
| HLN | Low | MAG6-bin18 | 80.32 | 5.172 | 4821760 | 34.43 | 73923 | Bacteroidota | *Chryseobacterium* | |
| T3LN | Low | MAG71-bin15 | 89.22 | 5.119 | 4838580 | 67.28 | 19361 | Pseudomonadota | *Stenotrophomonas* | *Stenotrophomonas sp003086775* |
| T1LN | Low | MAG43-bin8 | 91.67 | 2.435 | 4896167 | 59.56 | 101115 | Pseudomonadota | *Pseudomonas_E* | *Pseudomonas_E syringae* |
| T1HN | High | MAG54-bin7 | 92.26 | 0.803 | 4897798 | 63.55 | 37031 | Pseudomonadota | *Pseudomonas_E* | *Pseudomonas_E sp003253735* |
| T1HN | High | MAG48-bin13 | 98.00 | 0.275 | 4919275 | 39.48 | 27429 | Firmicutes | *Paenibacillus_J* | *Paenibacillus_J nuruki* |
| HHN | High | MAG26-bin8 | 97.10 | 0.159 | 4942342 | 39.06 | 74375 | Bacteroidota | *Pedobacter* | |
| HLN | Low | MAG3-bin15 | 86.98 | 0.237 | 4967183 | 64.14 | 11980 | Pseudomonadota | *Duganella* | *Duganella phyllosphaerae* |
| T3LN | Low | MAG67-bin10 | 94.86 | 1.739 | 4988107 | 59.53 | 8126 | Firmicutes | *Saccharibacillus* | *Saccharibacillus brassicae* |
| T2HN | High | MAG65-bin3 | 97.75 | 0.537 | 5070665 | 61.48 | 29080 | Pseudomonadota | *Pseudomonas_E* | *Pseudomonas_E poae* |
| HLN | Low | MAG11-bin6 | 96.17 | 0.159 | 5183976 | 38.97 | 193357 | Bacteroidota | *Pedobacter* | |
| T1HN | High | MAG55-bin8 | 97.12 | 0.159 | 5194491 | 38.99 | 148394 | Bacteroidota | *Pedobacter* | |
| T2HN | High | MAG64-bin2 | 86.44 | 2.718 | 5255736 | 59.62 | 25025 | Pseudomonadota | *Pseudomonas_E* | *Pseudomonas_E viridiflava* |
| T1LN | Low | MAG33-bin15 | 96.25 | 1.594 | 5480430 | 39.01 | 122493 | Bacteroidota | *Pedobacter* | |
| HHN | High | MAG14-bin1 | 93.20 | 1.201 | 5485102 | 64.32 | 17745 | Pseudomonadota | *Duganella* | *Duganella phyllosphaerae* |
| T3LN | Low | MAG70-bin14 | 98.58 | 0.553 | 5523418 | 56.06 | 113978 | Pseudomonadota | *Klebsiella* | *Klebsiella grimontii* |
| T2LN | Low | MAG62-bin6 | 85.22 | 6.291 | 5532076 | 56.67 | 3460 | Firmicutes | *Saccharibacillus* | |
| T1LN | Low | MAG40-bin3 | 95.93 | 0.026 | 5566168 | 64.22 | 31007 | Pseudomonadota | *Duganella* | *Duganella phyllosphaerae* |
| HLN | Low | MAG12-bin8 | 99.28 | 0.796 | 5586454 | 54.51 | 92296 | Pseudomonadota | *Erwinia* | *Erwinia persicina* |
| T1HN | High | MAG52-bin5 | 97.28 | 0.847 | 5629308 | 60.65 | 64684 | Pseudomonadota | *Pseudomonas_E* | *Pseudomonas_E rhodesiae* |
| T1LN | Low | MAG39-bin22 | 93.96 | 7.712 | 5714451 | 61.45 | 25000 | Pseudomonadota | *Pseudomonas_E* | *Pseudomonas_E poae* |
| T1LN | Low | MAG34-bin17 | 88.49 | 5.414 | 5770567 | 59.19 | 85504 | Pseudomonadota | *Pseudomonas_E* | *Pseudomonas_E viridiflava_C* |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| HHN | High | MAG21-bin19 | 75.34 | 8.620 | 6904257 | 61.22 | 51262 | Pseudomonadota | *Pseudomonas_E* | |

# Chapter 4 Appendix

Table S4.2 Distribution of clusters of orthologous groups (COG) across metagenomics sample.

| **COGs** | **HH** | **HL** | **S3H** | **S3L** | **S6H** | **S6L** | **S9H** | **S9L** |
|---|---|---|---|---|---|---|---|---|
| S: Function unknown | 28395 | 26467 | 18271 | 24011 | 14201 | 23865 | 19040 | 12896 |
| K: Transcription | 11522 | 10144 | 7021 | 9990 | 6052 | 9599 | 7768 | 5485 |
| E: Amino acid transport and metabolism | 10523 | 9345 | 6435 | 9176 | 5263 | 7713 | 6819 | 4873 |
| P: Inorganic ion transport and metabolism | 9002 | 7993 | 5590 | 7542 | 5089 | 7414 | 6106 | 4360 |
| G: Carbohydrate transport and metabolism | 8660 | 7610 | 5297 | 8016 | 5162 | 7486 | 5752 | 4305 |
| M: Cell wall/membrane/envelope biogenesis | 9000 | 8562 | 5744 | 7458 | 4180 | 6970 | 5313 | 3938 |
| T: Signal transduction mechanisms | 7619 | 7105 | 4773 | 6390 | 3429 | 5167 | 4686 | 3435 |
| C: Energy production and conversion | 6556 | 6021 | 4168 | 5540 | 3453 | 5424 | 4362 | 3086 |
| L: Replication, recombination, and repair | 5856 | 5694 | 3944 | 5419 | 2851 | 5132 | 3742 | 2651 |
| J: Translation, ribosomal structure and biogenesis | 5034 | 4727 | 3313 | 4330 | 2447 | 3722 | 3308 | 2476 |
| H: Coenzyme transport and metabolism | 4751 | 4584 | 3122 | 4171 | 2810 | 4038 | 3299 | 2354 |
| I: Lipid transport and metabolism | 4843 | 4398 | 2976 | 4131 | 1958 | 3194 | 2836 | 2103 |
| U: Intracellular trafficking, secretion, and vesicular transport | 3717 | 3664 | 2447 | 3218 | 1839 | 3362 | 2543 | 1835 |
| O: Posttranslational modification, protein turnover, chaperones | 3755 | 3644 | 2498 | 3209 | 1689 | 2864 | 2318 | 1685 |
| N: Cell motility | 3311 | 3200 | 2168 | 2842 | 1971 | 2744 | 2387 | 1722 |
| F: Nucleotide transport and metabolism | 2962 | 2868 | 1993 | 2679 | 1750 | 2546 | 2228 | 1582 |
| Q: Secondary metabolites biosynthesis, transport, and catabolism | 3246 | 2853 | 1913 | 2814 | 1436 | 2231 | 2000 | 1386 |
| V: Defense mechanisms | 1934 | 1917 | 1267 | 1791 | 876 | 1680 | 1217 | 822 |
| D: Cell cycle control, cell division, chromosome partitioning | 1459 | 1295 | 873 | 1281 | 757 | 1163 | 993 | 699 |
| B: Chromatin structure and dynamics | 54 | 43 | 26 | 40 | 9 | 23 | 22 | 13 |
| W: Extracellular structures | 54 | 46 | 27 | 22 | 4 | 37 | 14 | 10 |
| Z: Cytoskeleton | 42 | 39 | 15 | 20 | 8 | 44 | 25 | 12 |
| A: RNA processing and modification | 18 | 18 | 15 | 14 | 8 | 17 | 12 | 11 |

Table S4.3 Distribution of CAZymes and carbon degradation proteins across the metagenomic samples.

| CAZy | HH | HL | S3H | S3L | S6H | S6L | S9H | S9L |
|---|---|---|---|---|---|---|---|---|
| AA10 | 8 | 8 | 8 | 4 | 5 | 11 | 12 | 6 |
| CBM15 | 0 | 1 | 1 | 0 | 1 | 3 | 0 | 0 |
| CBM48 | 106 | 78 | 54 | 84 | 41 | 66 | 53 | 42 |
| CBM6 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 1 |
| CBM73 | 8 | 8 | 8 | 4 | 5 | 11 | 10 | 6 |
| CE1 | 23 | 24 | 16 | 18 | 14 | 24 | 18 | 10 |
| CE10 | 34 | 22 | 18 | 31 | 15 | 13 | 12 | 13 |
| GH102 | 18 | 15 | 8 | 12 | 13 | 15 | 12 | 11 |
| GH103 | 30 | 22 | 19 | 20 | 17 | 18 | 23 | 17 |
| GH105 | 12 | 17 | 10 | 10 | 14 | 17 | 14 | 7 |
| GH13 | 230 | 190 | 128 | 181 | 115 | 173 | 137 | 92 |
| GH130 | 5 | 5 | 1 | 3 | 3 | 4 | 2 | 0 |
| GH18 | 22 | 19 | 15 | 17 | 22 | 39 | 22 | 10 |
| GH19 | 26 | 25 | 9 | 20 | 12 | 18 | 15 | 9 |
| GH20 | 24 | 25 | 20 | 23 | 11 | 14 | 12 | 9 |
| GH23 | 30 | 26 | 19 | 20 | 29 | 39 | 26 | 16 |
| GH26 | 9 | 11 | 7 | 9 | 7 | 2 | 6 | 5 |
| GH29 | 18 | 21 | 21 | 20 | 3 | 12 | 4 | 5 |
| GH3 | 110 | 95 | 70 | 95 | 42 | 52 | 50 | 39 |
| GH30 | 15 | 11 | 10 | 8 | 2 | 6 | 3 | 3 |
| GH31 | 61 | 60 | 38 | 67 | 37 | 68 | 36 | 21 |
| GH32 | 30 | 21 | 15 | 31 | 21 | 22 | 19 | 22 |
| GH37 | 14 | 16 | 9 | 13 | 15 | 29 | 14 | 7 |
| GH38 | 6 | 5 | 2 | 7 | 10 | 16 | 7 | 6 |
| GH39 | 9 | 8 | 4 | 8 | 3 | 1 | 5 | 4 |
| GH4 | 1 | 0 | 0 | 0 | 10 | 15 | 9 | 3 |
| GH43 | 32 | 28 | 17 | 31 | 12 | 31 | 19 | 14 |
| GH5 | 47 | 41 | 30 | 37 | 11 | 19 | 24 | 16 |
| GH51 | 34 | 24 | 19 | 35 | 7 | 14 | 7 | 10 |
| GH52 | 1 | 1 | 1 | 1 | 2 | 0 | 1 | 0 |
| GH6 | 2 | 2 | 2 | 3 | 1 | 1 | 1 | 0 |
| GH63 | 0 | 0 | 0 | 0 | 1 | 5 | 2 | 0 |
| GH65 | 12 | 10 | 9 | 13 | 5 | 19 | 8 | 6 |
| GH68 | 8 | 10 | 7 | 9 | 6 | 1 | 4 | 6 |
| GH73 | 0 | 0 | 0 | 0 | 5 | 8 | 2 | 0 |
| GH77 | 27 | 24 | 17 | 18 | 18 | 16 | 17 | 10 |
| GH8 | 10 | 5 | 4 | 5 | 8 | 18 | 8 | 5 |
| GH88 | 2 | 3 | 2 | 2 | 6 | 17 | 8 | 4 |
| GH9 | 42 | 36 | 27 | 35 | 11 | 19 | 24 | 16 |
| GH94 | 6 | 6 | 2 | 5 | 3 | 8 | 3 | 5 |
| GH95 | 14 | 14 | 14 | 13 | 2 | 6 | 0 | 5 |
| GT1 | 49 | 53 | 39 | 47 | 71 | 130 | 73 | 57 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| GT17 | 3 | 2 | 3 | 3 | 6 | 2 | 3 | 3 |
| GT19 | 25 | 27 | 16 | 20 | 11 | 19 | 14 | 11 |
| GT2 | 288 | 239 | 176 | 251 | 142 | 190 | 166 | 128 |
| GT20 | 25 | 22 | 15 | 21 | 13 | 24 | 11 | 11 |
| GT21 | 1 | 0 | 0 | 0 | 1 | 4 | 1 | 0 |
| GT25 | 4 | 3 | 3 | 2 | 1 | 9 | 2 | 3 |
| GT26 | 10 | 13 | 8 | 11 | 9 | 15 | 11 | 8 |
| GT28 | 33 | 31 | 24 | 32 | 17 | 21 | 27 | 18 |
| GT30 | 28 | 26 | 14 | 15 | 13 | 21 | 14 | 11 |
| GT35 | 28 | 22 | 19 | 19 | 27 | 26 | 23 | 17 |
| GT36 | 3 | 5 | 4 | 6 | 8 | 9 | 6 | 6 |
| GT39 | 5 | 1 | 2 | 6 | 2 | 2 | 2 | 4 |
| GT4 | 184 | 161 | 128 | 169 | 116 | 158 | 120 | 94 |
| GT5 | 30 | 32 | 23 | 18 | 14 | 24 | 18 | 10 |
| GT51 | 106 | 96 | 58 | 73 | 49 | 96 | 72 | 45 |
| GT52 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| GT56 | 4 | 4 | 4 | 3 | 10 | 12 | 6 | 3 |
| GT73 | 4 | 4 | 1 | 3 | 2 | 5 | 5 | 2 |
| GT8 | 3 | 5 | 3 | 2 | 10 | 23 | 3 | 2 |
| GT80 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| GT81 | 5 | 5 | 2 | 5 | 3 | 1 | 3 | 4 |
| GT84 | 4 | 4 | 0 | 2 | 1 | 4 | 1 | 2 |
| GT9 | 44 | 43 | 28 | 35 | 36 | 43 | 33 | 25 |
| PL11 | 7 | 8 | 8 | 5 | 4 | 1 | 2 | 2 |
| PL4 | 5 | 6 | 3 | 7 | 4 | 1 | 3 | 2 |
| CBM20 | 2 | 5 | 3 | 2 | 0 | 2 | 2 | 0 |
| CBM50 | 1 | 2 | 2 | 3 | 0 | 8 | 3 | 2 |
| GH121 | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| GH15 | 5 | 4 | 3 | 4 | 0 | 1 | 1 | 1 |
| GH16 | 4 | 3 | 2 | 2 | 0 | 0 | 0 | 0 |
| GH28 | 1 | 1 | 1 | 1 | 0 | 3 | 2 | 1 |
| GH33 | 3 | 2 | 1 | 2 | 0 | 2 | 1 | 1 |
| GH97 | 4 | 9 | 6 | 3 | 0 | 3 | 0 | 1 |
| GT3 | 3 | 3 | 3 | 1 | 0 | 1 | 1 | 1 |
| GT70 | 2 | 2 | 2 | 2 | 0 | 0 | 0 | 0 |
| PL8 | 2 | 1 | 2 | 1 | 0 | 1 | 4 | 2 |
| GH101 | 1 | 1 | 0 | 1 | 0 | 2 | 1 | 0 |
| GT66 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 1 |
| GT87 | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| GT71 | 2 | 2 | 0 | 1 | 0 | 1 | 0 | 0 |
| GT14 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GT89 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| CBM5 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| GH116 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| GH47 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| GH66 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| GH82 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| GT11 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |

Table S4.4 Reconstruction of metabolic pathways from the barley microbiome.

| | KEGG_ko | Encoded protein (KEGG annotation) | *Protein* | Protein read coverage per sample | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | HL | HH | S3L | S3H | S6L | S6H | S9H | S9L |
| **Amino acid** | K00817 | hisC; histidinol-phosphate aminotransferase [EC:2.6.1.9] | *HisC* | 663 | 675 | 850 | 471 | 634 | 301 | 513 | 415 |
| | K00831 | serC, PSAT1; phosphoserine aminotransferase [EC:2.6.1.52] | *SerC, PSAT1* | 447 | 413 | 490 | 266 | 303 | 269 | 345 | 240 |
| | K00826 | E2.6.1.42, ilvE; branched-chain amino acid aminotransferase [EC:2.6.1.42] | *E2.6.1.42, ilvE* | 478 | 508 | 453 | 293 | 223 | 231 | 288 | 240 |
| | K00821 | argD; acetylornithine/N-succinyldiaminopimelate aminotransferase [EC:2.6.1.11 2.6.1.17] | *ArgD* | 572 | 555 | 485 | 353 | 187 | 104 | 220 | 223 |
| | K00832 | tyrB; aromatic-amino-acid transaminase [EC:2.6.1.57] | *TyrB* | 320 | 267 | 294 | 255 | 349 | 288 | 198 | 162 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | K00812 | aspB; aspartate aminotransferase [EC:2.6.1.1] | *AspB* | 364 | 504 | 344 | 261 | 172 | 20 | 176 | 52 |
| | K02619 | pabC; 4-amino-4-deoxychorismate lyase [EC:4.1.3.38] | *PabC* | 241 | 327 | 294 | 188 | 210 | 200 | 238 | 174 |
| | K07250 | gabT; 4-aminobutyrate aminotransferase / (S)-3-amino-2-methylpropionate transaminase / 5-aminovalerate transaminase [EC:2.6.1.19 2.6.1.22 2.6.1.48] | *GabT* | 124 | 227 | 237 | 124 | 14 | 153 | 65 | 81 |
| | K05825 | LYSN; 2-aminoadipate transaminase [EC:2.6.1.-] | *LYSN* | 111 | 143 | 152 | 81 | 240 | 113 | 61 | 88 |
| | K00819 | rocD, OAT; ornithine--oxo-acid transaminase [EC:2.6.1.13] | *RocD, OAT* | 156 | 85 | 112 | 42 | 21 | 0 | 40 | 57 |
| | K00823 | puuE; 4-aminobutyrate aminotransferase [EC:2.6.1.19] | *PuuE* | 89 | 112 | 47 | 66 | 24 | 47 | 71 | 47 |
| | K03918 | lat; L-lysine 6-transaminase [EC:2.6.1.36] | *LaT* | 112 | 122 | 72 | 113 | 23 | 0 | 0 | 19 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | K00830 | AGXT; alanine-glyoxylate transaminase / serine-glyoxylate transaminase / serine-pyruvate transaminase [EC:2.6.1.44 2.6.1.45 2.6.1.51] | *AgxT* | 20 | 41 | 0 | 37 | 0 | 20 | 4 | 47 |
| | | | | | | | | | | | |
| **Carbon degradation** | K05349 | bglX; beta-glucosidase [EC:3.2.1.21] | *BglX* | 2747 | 3125,97 | 2892,954 | 2114,827 | 1202,93 | 1029,927 | 1035,802 | 1156,903 |
| | K01190 | lacZ; beta-galactosidase [EC:3.2.1.23] | *LacZ* | 1110 | 827,2761 | 1343,214 | 997,5625 | 644,9148 | 766,4954 | 429,8287 | 388,9496 |
| | K01176 | AMY, amyA, malS; alpha-amylase [EC:3.2.1.1] | *AmY, AmyA, MalS* | 493,3 | 412,4599 | 300,8974 | 330,4118 | 570,5961 | 379,8491 | 196,6601 | 275,1116 |
| | K01207 | nagZ; beta-N-acetylhexosaminidase [EC:3.2.1.52] | *NagZ* | 328,1 | 424,8242 | 505,0109 | 285,4906 | 404,6925 | 285,4731 | 215,0501 | 300,3768 |
| | K01811 | xylS, yicI; alpha-D-xyloside xylohydrolase [EC:3.2.1.177] | *XylS, YicI* | 239,2 | 317,5511 | 544,5119 | 275,7944 | 396,9284 | 339,7735 | 182,0487 | 176,9164 |
| | K05989 | ramA; alpha-L-rhamnosidase [EC:3.2.1.40] | *RamA* | 234,9 | 371,1447 | 813,983 | 276,5019 | 314,2917 | 200,7502 | 13,5231 | 14,9667 |
| | K01192 | E3.2.1.25, MANBA, manB; beta-mannosidase [EC:3.2.1.25] | *E3.2.1.25, MAnbA, ManB* | 193,9 | 430,6751 | 343,4769 | 271,345 | 248 | 239,9155 | 139,696 | 305,9322 |
| | K01214 | ISA, treX; isoamylase [EC:3.2.1.68] | *IsA, TreX* | 319,4 | 328,7972 | 580,2116 | 180,5562 | 195,4794 | 69,7134 | 101,051 | 98,8971 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | K01209 | abfA; alpha-L-arabinofuranosidase [EC:3.2.1.55] | *AbfA* | 310 | 354,0085 | 337,8002 | 281,8126 | 117 | 94 | 179,9483 | 91 |
| | K01183 | E3.2.1.14; chitinase [EC:3.2.1.14] | *E3.2.1.14* | 134,1 | 265,5779 | 112 | 168,7829 | 440,8003 | 226,8115 | 21 | 198,9328 |
| | K01779 | racD; aspartate racemase [EC:5.1.1.13] | *RacD* | 163,1 | 149,3747 | 164,9651 | 109,3593 | 332,4521 | 147,1821 | 65,5148 | 156,9655 |
| | K01218 | gmuG; mannan endo-1,4-beta-mannosidase [EC:3.2.1.78] | *GmuG* | 207,6 | 188,0896 | 195,932 | 268,9463 | 19,9659 | 170,5474 | 112 | 113,7915 |
| | K01198 | xynB; xylan 1,4-beta-xylosidase [EC:3.2.1.37] | *XynB* | 158 | 145,6541 | 161,0879 | 105,7242 | 397,4209 | 27 | 88,52576 | 165,8603 |
| | K05343 | treS; maltose alpha-D-glucosyltransferase / alpha-amylase [EC:5.4.99.16 3.2.1.1] | *treS* | 221,2 | 251,3303 | 368,8706 | 71,5155 | 97,9485 | 28 | 71,7937 | 70,7452 |
| | K05350 | bglB; beta-glucosidase [EC:3.2.1.21] | *BglB* | 135,7 | 242,3799 | 420,8994 | 115,9748 | 115,0754 | 33,9654 | 39,90476 | 25 |
| | K01195 | uidA, GUSB; beta-glucuronidase [EC:3.2.1.31] | *uidA, GUSB* | 160,8 | 190,0221 | 237,7076 | 120 | 62,9658 | 58 | 58,9484 | 57,7152 |
| | K01200 | pulA; pullulanase [EC:3.2.1.41] | *PulA* | 48,8 | 49 | 49 | 48,9831 | 216,9328 | 48,9831 | 33,73832 | 84,00567 |
| | K01178 | SGA1; glucoamylase [EC:3.2.1.3] | *Sga1* | 46,63 | 127,5167 | 109 | 88,9831 | 22,0482 | | 32 | 21 |
| | | | | | | | | | | | |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Fermentation** | K01895 | ACSS1_2, acs; acetyl-CoA synthetase [EC:6.2.1.1] | ACSS1_2, acs | 1 070 | 1 076 | 940 | 779 | 625 | 470 | 407 | 825 |
| | K00656 | E2.3.1.54, pflD; formate C-acetyltransferase [EC:2.3.1.54] | E2.3.1.54, *PflD* | 194 | 193 | 194 | 116 | 958 | 471 | 155 | 461 |
| | K00001 | E1.1.1.1, adh; alcohol dehydrogenase [EC:1.1.1.1] | E1.1.1.1, *AdIH* | 291 | 374 | 381 | 233 | 408 | 300 | 171 | 357 |
| | K00925 | ackA; acetate kinase [EC:2.7.2.1] | *AckA* | 218 | 207 | 326 | 139 | 212 | 114 | 186 | 165 |
| | K00016 | LDH, ldh; L-lactate dehydrogenase [EC:1.1.1.27] | LDH, *LdH* | 31 | 33 | 33 | 33 | 283 | 82 | 49 | 176 |
| | K00625 | pta; phosphate acetyltransferase [EC:2.3.1.8] | *PtA* | 17 | 33 | 33 | 17 | - | - | 17 | - |
| | | | | | | | | | | | |
| **Nitrogen metabolism** | K00372 | nasC, nasA; assimilatory nitrate reductase catalytic subunit [EC:1.7.99.-] | *NasC, NasA* | 458 | 568,7169 | 494,5649 | 295,5658 | 380,9155 | 369,8825 | 270,7828 | 317,7218 |
| | K00367 | narB; ferredoxin-nitrate reductase [EC:1.7.7.2] | *NarB* | 59 | 59 | 116,9206 | 59 | 0 | 0 | 0 | 0 |
| | K00366 | nirA; ferredoxin-nitrite reductase [EC:1.7.7.1] | *NirA* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 |
| | K00360 | nasB; assimilatory nitrate reductase electron transfer subunit [EC:1.7.99.-] | *NasB* | 7 | 61,83193 | 45,4906 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | K00374 | narI, narV; nitrate reductase gamma subunit [EC:1.7.5.1 1.7.99.-] | *NarI, NarV* | 65,71 | 46 | 63,9823 | 69,00001 | 139,9824 | 153,6227 | 47,7168 | 155,965 |
| | K02568 | napB; nitrate reductase (cytochrome), electron transfer subunit | *NapB* | 4 | 9 | 8,98148 | 9 | 33,88958 | 14,57346 | 0 | 9 |
| | K02305 | norC; nitric oxide reductase subunit C | *NorC* | 14,49 | 0 | 0 | 8 | | 0 | 0 | 8 |
| | K04561 | norB; nitric oxide reductase subunit B [EC:1.7.2.5] | *NorB* | 82 | 0 | 0 | 24 | 39 | 0 | 0 | 24 |
| | K15864 | nirS; nitrite reductase (NO-forming) / hydroxylamine reductase [EC:1.7.2.1 1.7.99.1] | *NirS* | 28 | 0 | 0 | 28 | 0 | 0 | 16,4164 | 29 |
| | K19340 | nosF; Cu-processing system ATP-binding protein | *NosF* | 0 | 0 | 0 | 0 | 10,6387 | 0 | 0 | 16 |
| | K00376 | nosZ; nitrous-oxide reductase [EC:1.7.2.4] | *NosZ* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 33 |
| | K19341 | nosY; Cu-processing system permease protein | *NosY* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 14 |
| | K00368 | nirK; nitrite reductase (NO-forming) [EC:1.7.2.1] | *NirK* | 17,98 | 62,76539 | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | K00370 | narG, narZ, nxrA; nitrate reductase / nitrite oxidoreductase, alpha subunit [EC:1.7.5.1 1.7.99.-] | *NarG, NarZ, NxrA* | 255,9 | 184,05 | 286,9999 | 189 | 242,3205 | 434,7477 | 204 | 520,9665 |
| | K00371 | narH, narY, nxrB; nitrate reductase / nitrite oxidoreductase, beta subunit [EC:1.7.5.1 1.7.99.-] | *NarH, NarY, NxrB* | 112 | 78 | 90,9832 | 104,8713 | 299,9829 | 258,1374 | 119,9493 | 220,478 |
| | K02567 | napA; nitrate reductase (cytochrome) [EC:1.9.6.1] | *NapA* | 115,8 | 17,84226 | 83,3293 | 42 | 42 | 67,5264 | 0 | 8,99999 |
| | K03385 | nrfA; nitrite reductase (cytochrome c-552) [EC:1.7.2.2] | *NrfA* | 0 | 0 | 0 | 0 | 24 | 13 | 0 | 0 |
| | K00362 | nirB; nitrite reductase (NADH) large subunit [EC:1.7.1.15] | *NirB* | 799,4 | 807,8825 | 855,7621 | 659,3153 | 604,8901 | 891,7947 | 471,5986 | 662,9792 |
| | K00363 | nirD; nitrite reductase (NADH) small subunit [EC:1.7.1.15] | *NirD* | 106,5 | 131,2613 | 97,65868 | 73,48071 | 76,03105 | 98,88971 | 57,54516 | 86,598 |
| | K02586 | nifD; nitrogenase molybdenum-iron protein alpha chain [EC:1.18.6.1] | *NifD* | 0 | 0 | 0 | 0 | 22,983 | 0 | 0 | 25 |
| | K02591 | nifK; nitrogenase molybdenum-iron protein beta chain [EC:1.18.6.1] | *NifK* | 0 | 0 | 0 | 0 | 31,9827 | 0 | 0 | 27 |
| | K02588 | nifH; nitrogenase iron protein NifH | *NifH* | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 15 |
| | | | | | | | | | | | |

Table S4.5 Metaproteomic analysis general metrics.

| Sample Id | #Peptide | Phylum | Genus | Interpro |
|---|---|---|---|---|
| HH | 507 | 253 | 130 | 399 |
| HL | 275 | 160 | 79 | 229 |
| S3H | 372 | 174 | 100 | 265 |
| S3L | 590 | 358 | 220 | 488 |
| S6H | 244 | 130 | 81 | 188 |
| S6L | 93 | 33 | 13 | 78 |
| S9H | 604 | 196 | 124 | 541 |