Cyprus University of Technology

Faculty of Geotechnical Sciences and Environmental Management

Master's Thesis

Diversity of Bacteriophages in Cyprus Agricultural Soil

ANDREAS ROSSIS

Limassol, May 2025

CYPRUS UNIVERSITY OF TECHNOLOGY

Faculty of Geotechnical Sciences and Environmental Management

Department of Agricultural Sciences, Biotechnology and Food Science

Master's Thesis

Diversity of Bacteriophages in Cyprus Agricultural Soil

ANDREAS ROSSIS

Dr. George Botsaris

Dean of Faculty of Geotechnical Sciences and Environmental Management

Limassol, May 2025

Approval Form

Master's Thesis

Diversity of Bacteriophages in Cyprus Agricultural Soil

Presented by

Andreas Rossis

Supervisor: Dr. George Botsaris, Dean

Member of the committee: Dr. Michalis Omirou, Research Officer

Member of the committee: Dr. Loukas Kanetis Associate Professor

Cyprus University of Technology

Limassol, May 2025

Copyrights

The approval of the thesis by the Department of Agricultural Sciences, Biotechnology and Food Science does not necessarily imply the approval by the Department of the writer's views.

Acknowledgments:

First and foremost, I would like to express my heartfelt gratitude to my parents, whose continuous support and encouragement have been the foundation of my academic and personal growth. I am deeply grateful to my supervisors, Dr. Michalis Omirou and Dr. George Botsaris, firstly for introducing me to the extremely interesting and important subject. Their insightful feedback and support throughout the research process helped me structure my study and improve my research potential. Their expertise played a crucial role in shaping this work. I would also like to thank the staff of the Department of Agrobiotechnology at the Agricultural Research Institute for creating an environment that promotes curiosity and learning through daily scientific conversations and encouragement. Special thanks to my colleagues, Dr. Johanna Rhodosthenus, for her contribution, troubleshooting, and guidance through Next Generation Sequencing, Dr. Panayota Pissaridou, for the introduction to Linux-based systems, and Dr. Christiana Hadjimichael, whose company, help with sampling and scientific guidance were invaluable both for this project and my academic growth. Furthermore, I would like to thank Minas Trattou from the Cyprus Institute. His expertise and assistance with any computational problems that occurred significantly helped us use the HPC servers more efficiently.

Lastly, to everyone involved in this project and everyone who supported me. No matter the size of your contribution, this accomplishment wouldn't be possible without you.

ABSTRACT

Bacteriophages are prokaryotic viruses that infect bacteria and archaea. They exist in all ecosystems and are considered the most abundant biological entities on earth. Phages, with their ability to infect and lyse bacteria, are vital to their respective habitats as they regulate bacterial communities, mediate horizontal gene transfer, and cycle nutrients. Biodiversity, abundance, functions, and interactions with their host have been extensively studied in marine ecosystems, but their importance in terrestrial ecosystems has recently started to draw attention. Soil ecosystems are considered one of the most biologically complex environments on Earth, and the role of phages in it is vital both ecologically and agriculturally. Additionally, the improvement of high throughput sequencing and metagenomics has helped scientists and phage hunters worldwide to shed light on the subject. As no research has been conducted on soil phage diversity in Cyprus, we have selected multiple agricultural fields across the island. Soil geochemical properties were measured, and eDNA was extracted for high throughput sequencing (shotgun metagenomics). Bioinformatics analysis was then performed with various viral prediction tools to detect phages in soil metagenomes accurately. Furthermore, α-diversity and β-diversity were calculated and analysed to assess the possible effects of soil chemistry and spatial metrics on bacteriophage communities. In total, 2064 unique phage taxa were observed across all samples. Additionally, α- diversity indices where significantly correlated with Nitrogen, and β- diversity along with variance partitioning revealed nitrogen and elevation as significant factors explaining between sample variance. Furthermore, indicator species analysis revealed 35 phage taxa as significant indicators across the 3 nitrogen groups.

Keywords: Soil, Bacteriophages, Metagenome, Contigs, Diversity

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. Introduction

1.1 Background and Significance

Bacteriophages (phages) are viruses that infect and replicate within bacteria and archaea. They are the most abundant and diverse biological entities on the planet, with an estimated global population of 10^{31} particles(Bergh et al., 1989; Breitbart & Rohwer, 2005; Harper, 2011). These viruses are ubiquitous across all ecosystems and play a fundamental role in bacterial population control, gene transfer, and nutrient cycling(Rohwer & Thurber, 2009; Suttle, 2007). Their ecological and evolutionary importance is increasingly evident not only in marine systems, where their role in the viral shunt is well documented(Wilhelm & Suttle, 1999), but also in terrestrial environments, where phage-bacterial dynamics shape microbiomes in both natural and agricultural contexts(Abedon, 2008; X. Wang et al., 2024).

1.1.1. Historical Perspective and Foundations of Phage Research

Phages were first discovered independently by Frederick Twort in 1915 and Félix d'Hérelle in 1917. While Twort observed a "glassy transformation" of Staphylococcus cultures, it was d'Hérelle who convincingly demonstrated a transmissible, filterable agent capable of lysing Shigella and established its viral nature(D'Herelle, 2007; Twort, 1915). This discovery laid the groundwork for the use of phages as antibacterial agents in the pre-antibiotic era. Early clinical trials and field applications in the 1920s and 1930s demonstrated promising results on the concept of phage therapy, particularly in Eastern Europe and the Soviet Union(Chanishvili, 2012; Harper, 2011).

Despite their potential, inconsistent outcomes, the emergence of antibiotics, and a limited understanding of phage biology led to a decline in phage therapy research post-World War II(Summers, 2001). However, in the USSR and Georgia, institutions such as the Eliava Institute continued to develop phage-based treatments, preserving clinical interest through the 20th century(Sulakvelidze et al., 2001).

Phages have also been essential in shaping molecular biology. Key discoveries in gene regulation, recombination, lysogeny, and genome packaging were derived

from studies using bacteriophage λ and T4, including the Hershey-Chase experiment that proved DNA as the hereditary material(BLAKELY, 2004; Hershey & Chase, 1952; Luria & Human, 1952).

1.1.2. Phage Structure, Genomic Architecture, and Evolutionary Plasticity

Phages exhibit remarkable structural and genetic diversity. Morphologically, they are categorized into several groups based on tail structure (tailed phages), presence of envelopes, capsid symmetry, and genome type (dsDNA, ssDNA, dsRNA, ssRNA). The most studied group, the tailed dsDNA phages previously classified under *Caudovirales*, includes families such as *Myoviridae* (contractile tails), *Siphoviridae* (long, non-contractile tails), and *Podoviridae* (short tails), though these groupings have been restructured due to polyphyly(Ackermann, 2005; Harper, 2011).

Phage genomes vary widely, from <5 kb in ssDNA phages to >500 kb in jumbo phages. Their genomes are highly modular, displaying a mosaic architecture shaped by extensive horizontal gene transfer and recombination(Al-Shayeb et al., 2020; CHEN, 2025; Hendrix et al., 1999; Lawrence, 2002). Many phages carry accessory genes that can enhance host survival or viral fitness, including auxiliary metabolic genes (AMGs) that participate in nutrient cycles (e.g., photosynthesis, sulfur oxidation) or modulate stress responses(Roux et al., 2016; Thompson et al., 2011).

The genetic plasticity of phages is a key factor in their adaptability and ecological success. Genome exchange occurs through recombination, modular swapping, or transduction, contributing to the dynamic evolution of both phages and their bacterial hosts. This fluidity has made traditional taxonomy difficult, as few conserved genes are shared across all phages(Hatfull, 2008).

1.1.3. Taxonomy and Classification of Bacteriophages

Bacteriophage taxonomy has undergone major revisions as viral genomics and environmental virology have expanded our understanding of phage diversity. The classification is maintained by the International Committee on Taxonomy of Viruses (ICTV), which has shifted from morphology-based groupings toward a genome-informed hierarchical system. This change reflects the realization that

traditional morphology-based taxa, such as the families *Myoviridae*, *Siphoviridae*, and *Podoviridae* are polyphyletic and do not adequately represent evolutionary relationships(Ackermann, 2005; Rohwer & Edwards, 2002; Turner et al., 2021).

Traditionally, tailed phages were grouped within the now-obsolete order *Caudovirales*, which included over 95% of isolated phages. This order was divided based on tail morphology: contractile (*Myoviridae*), long non-contractile (*Siphoviridae*), and short tails (*Podoviridae*)(Ackermann, 2005; Harper, 2011). These groups typically possess double-stranded DNA (dsDNA) genomes and infect a wide range of bacterial hosts. However, whole-genome sequencing revealed extensive diversity in replication strategies, gene content, and structural modules even within these morphological groups(Hatfull, 2008; Hendrix et al., 1999).

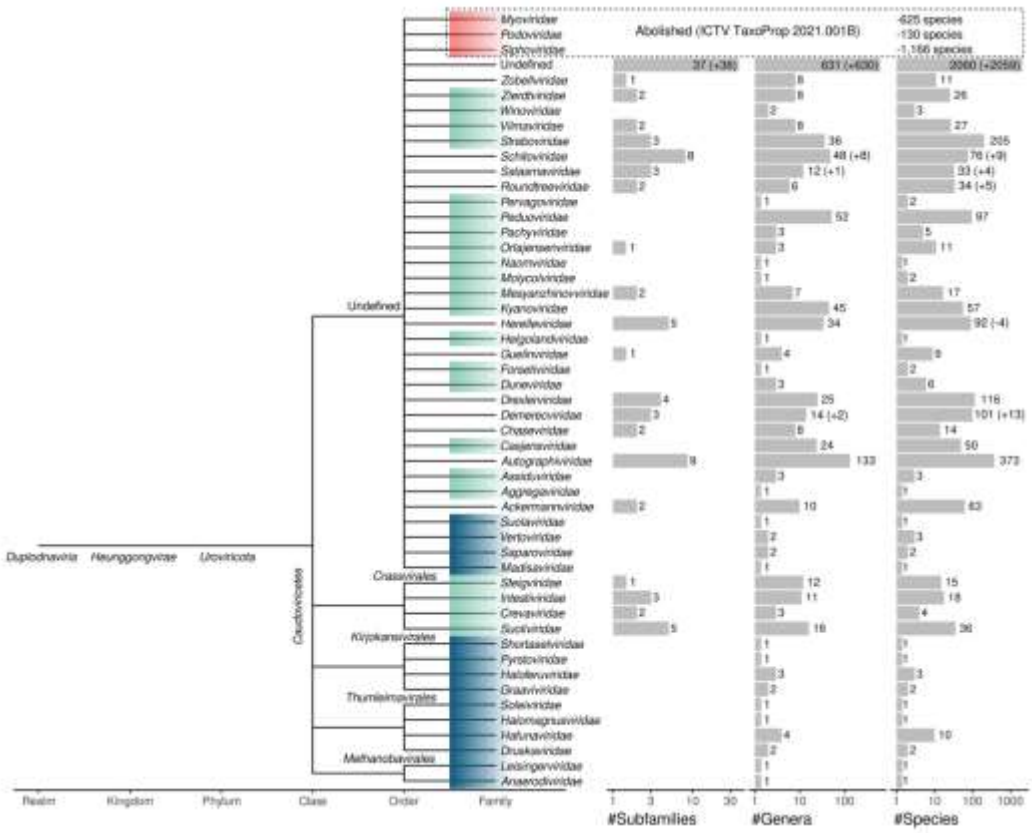

Figure 1: Revised phage taxonomy under the *Caudoviricetes* Class. The three previously main families were abolished due to polyphyly.(Turner et al., 2023)

Under the revised system, these families have been reorganized under the class ***Caudoviricetes***, which belongs to the viral realm ***Duplodnaviria***, encompassing all dsDNA viruses with head-tail structures, including herpesviruses(Adriaenssens &

Brister, 2017; Turner et al., 2021, 2023). Within *Caudoviricetes*, newer families have been established based on phylogenomic coherence. For example, *Herelleviridae* includes virulent phages related to the *Bacillus*-infecting SPO1 lineage and is characterized by large genomes (~130–150 kb), contractile tails, and headful DNA packaging mechanisms(Barylski et al., 2020).

Other prominent families include:

Autographiviridae: Characterized by short tails and the presence of a single-subunit RNA polymerase. This family includes the classic T7-like phages that infect *Enterobacteriaceae*, and their genomes are generally compact (~40 kb) with tightly organized early-late transcription modules (Harper, 2011).

Ackermannviridae: A recently defined family with distinct tail morphologies and broad host ranges, including phages infecting *Pseudomonas* and *Citrobacter*. These phages have medium-sized genomes and often exhibit depolymerase activity linked to host surface structure recognition (Turner et al., 2021).

Beyond tailed phages, additional morphotypes are recognized, including:

Inoviridae: Filamentous, non-lytic phages with circular ssDNA genomes. They chronically infect Gram-negative hosts and are involved in biofilm formation and virulence modulation (Rakonjac, 2012).

Microviridae: Small icosahedral phages with circular ssDNA genomes (~5–6 kb), which replicate through rolling-circle mechanisms and infect *Enterobacteria* and *Bacteroidetes* (Harper, 2011).

Cystoviridae: The only known dsRNA phage family, with members like Φ6 that infect *Pseudomonas syringae* and possess segmented genomes and lipid envelopes(Poranen et al., 2017).

Phage genomes can range from minimalist (e.g., ~5 kb in *Microviridae*) to highly complex (e.g., >200 kb in jumbo phages), with large DNA viruses often encoding their own DNA or RNA polymerases, tRNAs, and even DNA repair machinery (Yuan & Gao, 2017). These "jumbo" and "megaphages" blur the boundary between viruses and cellular life and are frequently found in environmental samples including soils (Roux et al., 2015; X. Wang et al., 2024).

In environmental metagenomes, especially soils, the majority of recovered viral contigs remain unclassified due to the absence of universal marker genes and underrepresentation in curated databases. These unassigned viruses are often referred to as “viral dark matter” and may represent new lineages or families with unique genomic signatures and host ranges (Nayfach et al., 2021; Roux et al., 2015). This taxonomic obscurity underscores the value of exploratory metagenomic studies in previously uncharacterized ecosystems such as agricultural soils in the Mediterranean region.

The evolving classification of phages reflects their extraordinary diversity and modular evolution, highlighting the need to incorporate structural, and genomic criteria into taxonomic frameworks. As new isolates and viral genomes are characterized, the phage taxonomy will continue to expand, offering insights not only into evolutionary biology but also into practical applications such as host range prediction and biocontrol agent selection.

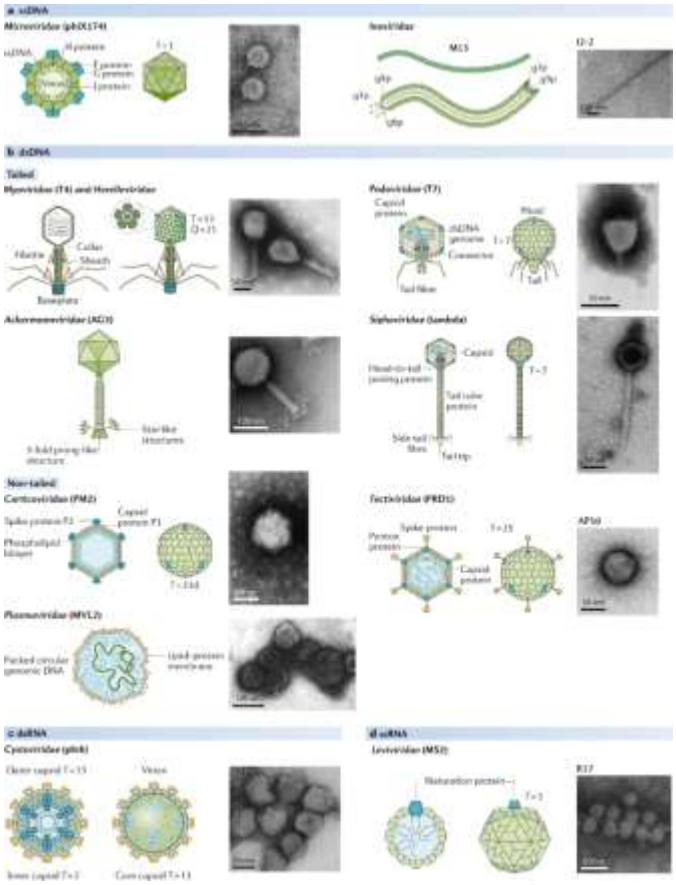

Figure 2: Schematic representation and transmission electron microscopy images visualizing diverse morphological structures of phages

1.1.4. Life Cycles and Strategies

Bacteriophages affect their environment depending on their lifecycles, which are mostly characterized by the lytic and lysogenic cycles and how strictly they alternate between those lifestyles. By the terms **lytic** (virulent) or **lysogenic** (temperate), we differentiate phages on their preference to either immediately lyse the host to multiply or remain integrated into the host's genome, therefore multiplying in parallel with cell division. Additionally, temperate phages can lyse their host in the event of host cell stress, like nutrient depletion and radiation(Mäntynen et al., 2021).

However, the simplicity of this categorization between phages overshadows the different strategies of phage replication and phage-host interactions. Other phage bacteriophage lifestyles include:

- **Chronic Infection** lifestyle describes the ability of certain phages to release progeny without the need for bacterial cell lysis to occur(Hoffmann-Berling & Mazé, 1964).
- **Pseudolysogeny** describes a state in which a bacteriophage's development stalls inside a host cell. The viral genome neither replicates independently (as during a lytic infection) nor integrates and is maintained in sync with host cell division (as in true lysogeny). Despite remaining intact, the phage genome is dormant awaiting improved adverse conditions, e.g., starvation or other stresses that trigger this response to subside, whereupon the virus can resume either the lytic cycle or establish a genuine lysogenic relationship(Łoś & Węgrzyn, 2012).

1.1.5. Phages as Ecological Agents in Microbial Ecosystems

Phages are integral to ecosystem function through their influence on microbial diversity, nutrient cycling, and genetic flux. In marine systems, they contribute to the "viral shunt," lysing host cells and redirecting organic carbon and nutrients away from higher trophic levels and into microbial loops (Suttle, 2005; Wilhelm & Suttle, 1999). Similar roles are hypothesized in soils, where phages regulate bacterial population size, affect community assembly, and mediate horizontal gene transfer (X. Wang et al., 2024).

In agricultural systems, particularly in the rhizosphere, phages impact both pathogenic and beneficial microbial populations. They can suppress soil-borne pathogens such as Ralstonia, Xylella, and Pseudomonas syringae, or inadvertently disrupt mutualistic bacteria like Rhizobium and Azospirillum, affecting nutrient uptake and plant growth (Jones et al., 2007; Le & Kurtböke, 2019; Xiong et al., 2020). Phages may also influence plant health indirectly through their modulation of microbial antagonism, quorum sensing, and metabolite production (Howard-Varona et al., 2017; Trivedi et al., 2020).

Environmental factors such as soil pH, texture, and moisture influence phage persistence and lifecycle dynamics. For instance, lysogeny is more common in drier, nutrient-poor soils, while lytic replication dominates in more moist and nutrient-rich environments (Jansson & Hofmockel, 2020; R. Wu et al., 2023).

1.1.6. Biotechnological and Therapeutic Applications

Bacteriophages are being explored for a wide range of applications beyond their natural ecological roles. The global rise in antimicrobial resistance has renewed interest in phage therapy as a targeted, self-amplifying alternative to antibiotics. Phages have been successfully used against multidrug-resistant infections in both clinical and agricultural settings (Czaplewski et al., 2016; Hagens & Loessner, 2007).

In agrobiotechnology and plant protection, phages are applied as biocontrol agents to reduce the load of bacterial phytopathogens while minimizing chemical pesticide use(Cooper, 2016). Products targeting Erwinia amylovora, Xanthomonas, and Pseudomonas are already approved for use in some countries (Balogh et al., 2008; Vu & Oh, 2020). Additionally, phages are used in diagnostics, biosensor development, vaccine delivery, and as tools in synthetic biology due to their specificity, modifiability, and minimal biosafety risks (Harper, 2011; Imran et al., 2023).

1.2 Bacteriophages in Soil Ecosystems

Soil ecosystems represent one of the most biologically complex environments on Earth, containing diverse microbial communities that are essential for nutrient cycling, plant health, and ecosystem stability. Within these microbiomes, bacteriophages are ubiquitous but remain significantly understudied in comparison to their aquatic and host-associated counterparts (Jansson, 2023; X. Wang et al., 2024). Recent studies have emphasized the ecological importance of soil phages, not only as agents of bacterial mortality but also as drivers of microbial evolution, nutrient flux, and ecosystem functioning.

1.2.1. Diversity and Abundance in Soil

Estimates suggest that soil contains up to 10^8-10^{10} virus-like particles (VLPs) per gram of dry soil(Jansson, 2023; Williamson et al., 2017), making phages among the most abundant biological entities in terrestrial ecosystems. However, most of these viruses remain uncharacterized, forming a component of the so-called "viral dark matter", genomic sequences with no known taxonomic assignment (Roux & Emerson, 2022). Advances in metagenomics and metatranscriptomics have enabled the detection of vast novel viral lineages in soil, particularly through the identification of viral operational taxonomic units (vOTUs) from bulk and rhizosphere soils (X. Wang et al., 2024).

The diversity of phages varies significantly between soil compartments. Rhizosphere soils, enriched with root exudates, exhibit higher phage abundance and taxonomic diversity than bulk soils, which are more nutrient poor (Bi et al., 2021; Swanson et al., 2009; Y. Wang et al., 2022). These differences are attributed to the heterogeneous distribution of microbial hosts, as well as environmental variables such as moisture, temperature, and plant development (Santos-Medellín et al., 2023; Yang et al., 2023).

1.2.2. Ecological Roles of Soil Phages

Soil bacteriophages impact microbial community dynamics primarily through top-down control mechanisms, including host lysis and the regulation of dominant bacterial populations (Brum et al., 2015; X. Wang et al., 2024). This is exemplified

by the “Kill-the-Winner” (KtW) hypothesis, where dominant bacterial taxa are preferentially lysed, thereby maintaining microbial diversity and preventing competitive exclusion (Våge et al., 2014). Conversely, the “Piggyback-the-Winner” (PtW) hypothesis suggests that under conditions of high bacterial density, temperate phages favor lysogeny, integrating into host genomes and coexisting with their hosts (Knowles et al., 2016). In addition to KtW and PtW, soil phages may employ a “Piggyback-the-Loser” (PtL) strategy, switching to lysogeny when bacterial densities are low to avoid unproductive lysis and ensure their own persistence via host replication (Hu et al., 2025; Voigt et al., 2021). Beyond these host abundance-dependent switches, coevolutionary theory describes two contrasting dynamics in phage-host arms races:

- Arms-Race Dynamics, hosts and phages engage in unidirectional escalation. Hosts continually evolve stronger resistance and phages evolve broader infectivity, often driving rapid genomic turnover and the fixation of resistance or infectivity alleles(Hampton et al., 2020).
- Fluctuating Selection Dynamics involve cyclical swings in allele frequencies, where different host resistance variants and phage infectivity types rise and fall in dominance over time, maintaining high polymorphism and stabilizing diversity in the community(Jdeed et al., 2025a).

Through these regulatory mechanisms, phages contribute to ecosystem stability and resilience, influencing not only bacterial abundance but also functional guild structure (Jansson, 2023). Virulent phages can reduce pathogen loads in the rhizosphere, suppressing disease while indirectly promoting beneficial microbes (Wei et al., 2015; S. Wu et al., 2025; Xiong et al., 2020). On the other hand, lysogenic conversion may introduce new traits into bacterial genomes, including virulence factors and stress tolerance mechanisms (Howard-Varona et al., 2017).

1.2.3. Nutrient Cycling

Bacteriophages play a vital role in biogeochemical cycles by facilitating the release of nutrients upon bacterial lysis, a phenomenon referred to as the "viral shunt" (Jansson, 2023; Suttle, 2005; X. Wang et al., 2024). This process liberates organic carbon, nitrogen, phosphorus, and other micronutrients, making them available to

other soil organisms. Studies in composting systems have shown that phages track the abundance of mesophilic and thermophilic bacteria, shaping microbial succession and nutrient turnover (Liao et al., 2023).

Beyond lysis, phages influence host metabolism through the expression of auxiliary metabolic genes (AMGs), viral genes that enhance host fitness under specific conditions. These genes encode enzymes such as chitosanases and endomannanases involved in the degradation of complex polysaccharides, thus supporting nutrient mineralization (Emerson et al., 2018; R. Wu et al., 2022). The expression of AMGs by both lytic and temperate phages modulates microbial metabolism, contributing to nutrient flux and microbial food web restructuring (Breitbart, 2012; Trubl et al., 2018).

1.2.4. Evolutionary Roles and Coevolution

In addition to ecological regulation, phages are pivotal in bacterial evolution through horizontal gene transfer (HGT). Lysogenic phages integrate into host genomes, carrying mobile genetic elements, resistance genes, and AMGs that alter host fitness and ecological roles (Howard-Varona et al., 2017; X. Wang et al., 2024). These interactions are especially complex in the rhizosphere, where spatial structure, nutrient gradients, and host density influence the costs and benefits of resistance (Brockhurst et al., 2003; Gómez & Buckling, 2011).

Experimental evidence suggests that phage and bacteria coevolution in soils is shaped by fluctuating selection pressures, including abiotic factors like temperature and moisture, as well as biotic factors such as plant species and cropping systems (Florent et al., 2022; Jdeed et al., 2025b; Yang et al., 2023). Resistance mutations often incur trade-offs, reducing competitiveness or symbiotic capacity, which can feedback into ecosystem-level processes such as nitrogen fixation or disease suppression (Brown & Rant, 2013; Jacott et al., 2017; Kolan et al., 2024).

1.2.5. Implications for Agricultural and Soil Health

Despite their importance, the roles of soil phages in terrestrial ecosystems, especially agriculture, are poorly understood compared to marine or human microbiomes (Braga et al., 2020; Breitbart & Rohwer, 2005; Jansson, 2023). Their

activity in the rhizosphere can modulate plant health outcomes, either by suppressing pathogens or altering the abundance of plant-growth-promoting bacteria (PGPRs). Moreover, phages may help maintain microbial diversity, which is associated with disease resistance, nutrient availability, and crop productivity (Fierer, 2017; Trivedi et al., 2020).

The One Health framework, which integrates human, animal, plant, and environmental health, increasingly recognizes the soil microbiome as a critical component (Banerjee & van der Heijden, 2023). In this context, understanding soil viral ecology is essential for managing microbial communities to support sustainable agriculture, mitigate pathogen outbreaks, and enhance ecosystem resilience under climate change and intensification pressures (Destoumieux-Garzón et al., 2018; Montgomery et al., 2024).

1.3 Agricultural Relevance of Soil Bacteriophages

The integration of bacteriophages into agricultural systems offers promising potential for targeted microbial management in both conventional and organic farming. Phages provide a species-specific and environmentally friendly alternative to chemical bacteriocides, with applications ranging from pathogen suppression to supporting soil microbial balance and nutrient cycling. Their unique biology makes them suitable for tailored interventions that do not compromise beneficial microbiota, and their self-replicating nature allows for sustained activity in appropriate host-rich environments (Vu & Oh, 2020).

1.3.1. Field and Greenhouse Trials of Phage Applications

Phage-based biocontrol has progressed from conceptual work to practical implementation through greenhouse and field trials. For example, phage therapy against *Pseudomonas syringae pv. Porri*, the causal agent of bacterial blight in leek, was tested using a six-phage cocktail across multiple field sites. One trial showed reduced symptom severity compared to controls, though results varied depending on environmental conditions (Rombouts et al., 2016).

In rice, foliar application of phages targeting *Xanthomonas oryzae pv. oryzae* resulted in over 50% reduction of bacterial blight symptoms under field conditions,

particularly when formulated with skim milk for UV protection (Chae et al., 2014). Similar outcomes were achieved in tomato, where a combination of phage cocktails and systemic acquired resistance inducers led to a 66% decrease in bacterial spot incidence (Balogh et al., 2003). Greenhouse trials using phages M5 and M8 to combat *Ralstonia solanacearum* in banana reported complete disease suppression (Ramírez et al., 2020), while postharvest applications in potato and cabbage reduced spoilage caused by *Pectobacterium* and *Erwinia* species (Schnabel et al., 1999; Yu et al., 2016). Even early 20th-century studies successfully applied phage suspensions to seeds, report showing a drop in Pantoea stewartii infection from 18% to 1.4% in corn (Thomas, 1935).

These trials highlight that while phages can be highly effective under controlled conditions, field efficacy is often influenced by environmental variability, formulation stability, and application timing (Flaherty et al., 2000; Jones et al., 2007).

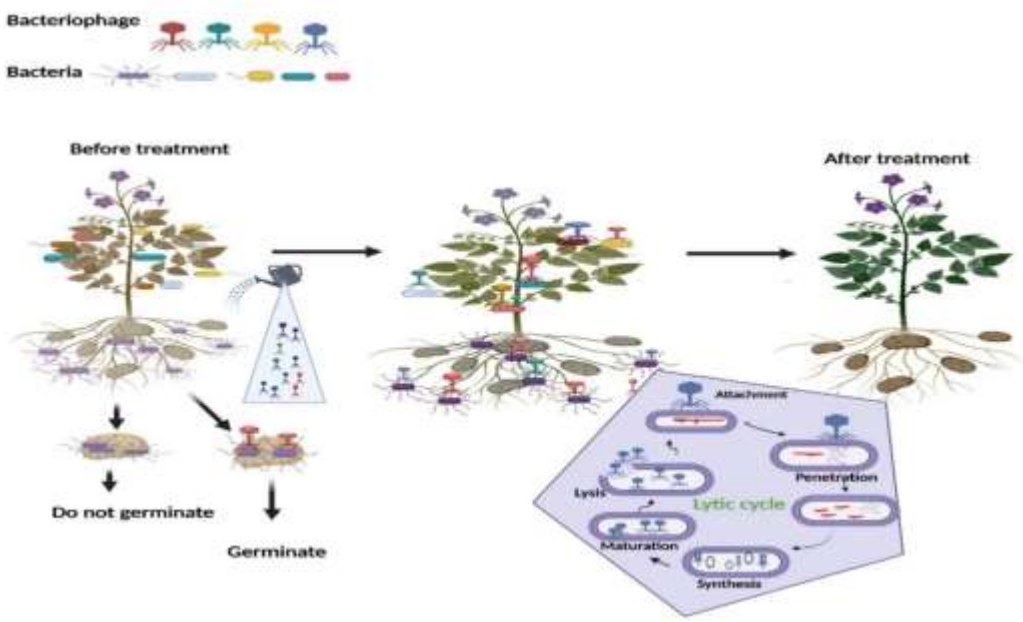

Figure 3: Visual representation of hypothetical bio control of pathogens with the use of phage therapy in agriculture(Jaglan et al., 2024)

1.3.2. Soil and Rhizosphere Functions Supporting Plant Health

Bacteriophages in the rhizosphere do more than control pathogens—they regulate microbial networks essential to soil health and plant productivity. By lysing

dominant bacterial taxa, phages promote diversity and functional redundancy, which are central to resilient soil microbiomes (X. Wang et al., 2024). These top-down dynamics, referred to as "Kill-the-Winner," maintain microbial equilibrium and suppress overproliferation of any single species (Gómez & Buckling, 2011).

Phages also contribute to nutrient turnover by lysing host cells, releasing carbon, nitrogen, and phosphorus into the surrounding soil. These nutrients can then be assimilated by other microbes or taken up by plants, accelerating nutrient cycling and root-associated productivity (Jansson, 2023). Crucially, temperate and even lytic phages may encode auxiliary metabolic genes (AMGs) that enhance microbial function. AMGs involved in phosphate solubilization, nitrogen metabolism, and oxidative stress resistance have been identified in viromes from compost-treated and organically managed soils (Emerson et al., 2018; R. Wu et al., 2022). Phages have also been shown to carry genes for succinoglycan biosynthesis—critical for root nodulation and beneficial symbiosis in legumes (Howard-Varona et al., 2017; X. Wang et al., 2024). These findings indicate that bacteriophages may serve as indirect biofertilizers by promoting microbial processes that support plant health and nutrient availability, particularly under environmental stress.

1.3.3. Commercialization, Formulation, and Application Strategies

Several phage-based biopesticide products have reached the market, most notably AgriPhage®, which is approved for use on tomatoes, peppers, and citrus in the United States (Buttimer et al., 2017). These products target pathogens such as Xanthomonas campestris and Pseudomonas syringae and are often incorporated into integrated pest management (IPM) frameworks (Jones et al., 2007).

To improve stability and persistence in field conditions, phages are often formulated with protective agents such as skim milk, sucrose, or corn flour. These additives shield viral particles from UV radiation and desiccation, particularly for foliar applications (Balogh et al., 2010; Flaherty et al., 2001). In citrus trials, combinations of phages with SAR inducers and stabilizers achieved over 80% disease control in field conditions (Iriarte et al., 2012).

From a deployment perspective, phages can be applied via foliar sprays, soil drenches, seed coatings, or in combination with live bacterial carriers that amplify

phage replication in situ (Gill & Abedon, 2003). Several delivery models emphasize application during early disease onset or at high host density to maximize infection success.

Although regulatory pathways remain inconsistent globally, frameworks are being developed to standardize phage product evaluation. A stepwise commercialization model includes laboratory screening, host range definition, fermentation scaling, environmental risk assessment, and field validation (León et al., 2024; Mohammadi et al., 2025).

1.3.4. Alignment with Sustainable and Organic Agriculture

Phage therapy aligns well with sustainability goals, offering a residue-free, host-specific solution compatible with organic farming practices. Because they do not harm beneficial microbes or pollinators, phages reduce non-target effects and preserve essential soil functions (Trivedi et al., 2020). In low-input and organic systems, where chemical options are limited, phages represent a promising alternative for disease control and microbiome management. Compost-amended soils show higher phage diversity and AMG content, suggesting that phage applications could be integrated into broader ecological farming frameworks (Liao et al., 2023; X. Wang et al., 2024). By supporting plant resilience, microbial diversity, nutrient mobilization, and suppressing pathogens, phages can contribute to productive and ecologically stable agroecosystems.

1.4 Approaches to Studying Bacteriophages

1.4.1. Traditional Culture-Based Techniques

Historically, the study of bacteriophages has relied on culture-based methods such as plaque assays, spot tests, and turbidity assays, which involve infecting a known bacterial host and visually observing lysis or growth inhibition (Flaherty et al., 2000; Schneider et al., 2018; Thomas, 1935). These assays enable direct isolation and characterization of phages and remain valuable in confirming infectivity and host specificity. However, they are inherently limited to culturable bacterial hosts, and many soil bacteria, and their associated phages are unculturable under standard laboratory conditions, leading to biased detection (Dennehy & Abedon, 2021;

Labrie & Moineau, 2000). Additionally, phage isolation protocols can introduce significant selection bias. Enrichment steps, while increasing yield, may select for fast-replicating or lytic phages at the expense of others. The choice of host strain, medium, and growth conditions (e.g., oxygen, pH, temperature) can all impact the recovery and diversity of phages isolated from a sample (Van Charante et al., 2021).

1.4.2. Molecular Methods

Advances in molecular biology have introduced PCR-based detection, allowing for specific and rapid identification of phage DNA directly from environmental samples. Quantitative PCR (qPCR) extends this further by enabling the quantification of phage abundance and is particularly useful in monitoring phage dynamics or contamination (Martín et al., 2008; Verreault et al., 2011). Yet, both traditional PCR and qPCR require prior knowledge of phage sequences, which is often lacking in environmental viromes (Binetti et al., 2005; del Rio et al., 2007).

1.4.3. Genome-by-Genome Strategy

An alternative to metaviromics is the genome-by-genome approach, where individual phages are isolated, sequenced, and annotated. This allows for controlled comparative genomic analyses and retains viable phage stocks for future phenotypic assays (Hanauer et al., 2006; Harper, 2011). However, this strategy still depends on a culturable host and thus cannot address the majority of uncultured soil phages.

1.4.4. Metagenomics and Metaviromics

To overcome these limitations, metagenomics, specifically shotgun metagenomic sequencing has become the dominant method for studying phages in complex environments such as soil, water, and host-associated microbiomes (Breitbart et al., 2002; Sullivan, 2015; Sullivan et al., 2017). This approach allows for the recovery of phage genomes without requiring culturing or prior sequence knowledge. DNA is extracted from filtered and concentrated samples, often treated with nucleases to remove free-floating DNA, and sequenced using platforms like Nextseq or NovaSeq.

However, identifying viral contigs from such data is non-trivial. Phages lack universal marker genes (e.g., 16S rRNA in bacteria), and many sequences recovered from soil represent "viral dark matter", unclassified viruses with no homology to known genomes (Mokili et al., 2012; Paez-Espino et al., 2016; Tringe & Hugenholtz, 2008). Biases also exist: for example, metagenomic protocols often favor dsDNA viruses, while RNA phages or ssDNA phages are underrepresented due to library preparation methods (Greninger, 2018; Kim & Bae, 2018).

1.4.5. Challenges and Integration

Despite the power of metagenomics, issues remain especially in soil, where high microbial diversity, genome fragmentation, and insufficient reference databases reduce resolution (Dennehy & Abedon, 2021; X. Wang et al., 2024). Consequently, combining both culture-based and sequence-based approaches, when feasible, offers a more comprehensive view of phage diversity.

This integrative strategy also supports host-phage linkage studies (e.g., Hi-C metagenomics), lifestyle prediction, and the exploration of auxiliary metabolic genes (AMGs) involved in nutrient cycling and phage-host interaction areas increasingly important for ecological and agricultural applications(R. Wu et al., 2023).

1.5 Aim

With the use of High-Throughput Next Generation Sequencing, bioinformatic analysis, and statistical analysis of diversity the main aim of this thesis is to initiate the documentation of phage taxa present in agricultural soils of Cyprus. To characterize the taxonomic and functional diversity of soil bacteriophage communities across multiple agricultural fields in Cyprus, thus filling a gap in Mediterranean virome research. Additionally, we aim to assess the potential trends in phage communities and ecological drivers shaping them. To analyze the influence of soil geochemical properties (e.g., pH, organic matter, nutrient content) and spatial metrics on phage community structure, α-diversity, and β-diversity, with an emphasis on barley fields across Cyprus. Finaly to establish a baseline dataset for future soil virome research in Cyprus, facilitating the development of phage-based indicators of soil health.

Research Questions

What is the composition and diversity of soil bacteriophage communities in Cypriot agricultural fields?

How do soil chemistry and spatial parameters shape phage α- and β-diversity in barley fields?

What baseline can be established for future monitoring and application of phages in Cyprus' agricultural soils?

Significance

This thesis represents the first systematic metagenomic investigation of soil phages in Cyprus, a region of pronounced agricultural and environmental diversity. By integrating high-throughput sequencing, rigorous bioinformatics, and advanced diversity analysis, this work advances our understanding of how phages influence soil microbial ecology and vise versa. The findings are expected to inform phage-based biocontrol, microbiome engineering, and soil health monitoring strategies for sustainable agriculture.

2. Materials and Methods

2.1 Materials

Consumables

- Falcon tube 50ml x52
- 15 Eppendorfs
- 96-well 0.8 ml Polypropylene Deepwell Storage (MIDI plate) x2
- Hard-Shell 96-well PCR plates
- Nextseq 2000 reagent cartridge (Illumina)
- Nextseq 2000 flow cell (Illumina)
- Nuclease free water
- 1.7 ml microcentrifuge tubes x34
- 8-tube strips
- Microseal 'B' adhesive seal x3
- 10 µl pipette tips
- 20 µl pipette tips
- 1000 µl pipette tips
- 80% ethanol (EtOH)
- Microseal 'F' foil seal
- Qubit Assay Tubes
- Kjeldahl digestion tubes x17
- Filter paper x17
- Whitman No. 5 filter papers x17
- 20ml Serological pipettes x2
- 10ml Serological pipettes x2
- Pasteur pipettes x17

Kits

- Illumina DNA Prep (24Samples) kit
- IDT for Illumina Nextera DNA UD Indexes set A
- Qubit dsDNAHS AssayKit

Equipment

- Sampling Auger
- 2mm and 4mm Soil sieves
- DNeasy PowerSoil Pro Kit (Qiagen)
- Retsch tissue lyser (Qiagen)
- Corning® LSE™ High Speed Microcentrifuge
- QIAcube (Qiagen)
- Qubit® Fluorometer 3.0
- 200µl - 1000µl micropipette
- 20 - 200µl micropipette
- 20µl micropipette
- 1 - 10µl micropipette
- 20µl multichannel pipettes
- 200µl multichannel pipettes
- Illumina Nextseq 2000
- Vortexer
- Microplate centrifuge
- Agilent 4150 TapeStation System
- Biometra Tone thermal cycler
- Tube spinner
- Precision scale
- Laminar air flow cabinet
- Heater
- UDK 169 Kjeldahl Nitrogen Analayzer (Velp Scientifica)
- AutoKjel Autosampler (Velp Scientifica)
- 100ml volumetric flasks x17
- 5ml pipette
- Shaker (P SELECTA)
- Refrigerated floor-standing centrifuge Sigma 8KS
- Model 420 Flame photometer (Sherwood)
- Funnels x17

- SAN++® automated wet chemistry analyzer (Skalar)
- SA1100 sampler (Skalar)
- Volumetric flask 20ml
- 500ml Conic vials x17
- Burette
- Glass beaker 500ml
- HI5522 Laboratory Research Grade Benchtop meter (HANNA)

2.2 Methods

3.2.1. Sampling

Soil was sampled from seventeen agricultural fields across South Cyprus using a sampling auger at 10-15cm depth from 10 points in each location and homogenised. Fifteen of the samples originated from barley fields in the regions of Ora, Anarita, Athienou, Agia Anna, Dromolaxia, Avdimou, Kourdaka, Goudi and Chrysoxous. Furthermore, soil was sampled from a Citrus grove in the region of Akrotiri and one from a vineyard in the region of Kyperounta. The coordinates of each sampling field were noted captured using smartphone gps.

Table 1: Location spatial information of samples and crops present

Sample_ID	Region	Lat	Long	crops
S1	Kyperounta	34.94239	32.94944	Vineyard
S4	Ora	34.85799	33.19973	Barley
S14	Anarita	34.75668	32.53519	Barley
S15	Athienou	35.02087	33.50051	Barley
S16	Agia Anna	34.93843	33.48557	Barley
S17	Dromolaxia	34.89243	33.59582	Barley
S18	Avdimou	34.67853	32.78168	Barley

Sample_ID	Region	Lat	Long	crops
S20	Akrotiri	34.64070	32.94584	Citrus
S925	Kourdaka	34.85250	32.53845	Barley
S935	Goudi	34.99420	32.44637	Barley
S937	Goudi	34.99644	32.44899	Barley
S938	Goudi	34.99356	32.44971	Barley
S943	Chrysoxous	35.00755	32.43095	Barley
S946	Chrysoxous	35.00803	32.43636	Barley
S950	Chrysoxous	35.00236	32.44542	Barley
S987	Latsi	35.03306	32.38035	Barley
S989	Latsi	35.03540	32.39362	Barley

Figure 4: Visual representation of the distribution of sampling sites across Cyprus

3.2.2. Sample processing

Samples were stored at -20ºC until the time of further processing. Soil was defrosted at 4ºC for 24h and then sieved at 2mm. For downstream analysis 120gr of moist soil was dried in the oven at 100ºC(Webb & Adeloju, 2013) for the purpose of proceeding to soil health assessment and 0.3gr of moist soil per sample was used for eDNA extraction.

Soil Analysis:

For Acidity (pH) and electrical conductivity (EC) dried soil was mixed with dH_2O at 1:1 ration at 200rpm for 1h, then pH and EC were measured with HANNA HI5522 Laboratory Research Grade Benchtop meter. Organic matter was measured by the Walkey Black Method(WALKLEY & BLACK, 1934) on 1gr of dried soil per sample. For Total Kjeldahl Nitrogen(Bremner, 1960) 1gr of dried soil was measured on UDK 169(VELP SCIENTIFICA). Furthermore, for the determination of Phosphorus(Webb & Adeloju, 2013), 5gr of dried soil were used for Phosphorus for extraction. Standards and samples were prepared according to the SKALAR METHODS instructions, and the measurements were executed using a SAN++® Series Automated Wet Chemistry Analyzer. Additionally, Potassium and Sodium content was measured on 5gr of dried soil by briefly mixing with 1N CH3COONH4.

3.2.3. DNA extraction and sequencing:

Soil DNA was extracted using DNeasy PowerSoil Pro QIAcube Kit (Qiagen, Hilden, Germany) with 0.3 g of soil(Braga et al., 2020) from each sample four times. DNA concentration was then determined using Qubit™ 1X dsDNA High Sensitivity (HS) and Broad Range (BR) Assay Kits on Qubit Flex Fluorometer (ThermoFisher SCIENTIFIC).

Libraries were prepared according to the Illumina DNA prep guide using the Illumina DNA Prep kit. Libraries were visualized on Tapestation 4150(Agilent) and then loaded in and sequenced on a 300 cycle (paired end) using Nextseq 2000.

3.2.4. Bioinformatic Analysis

Quality of raw sequencing reads was performed using FastQC (v0.12.1), then fastp(Chen et al., 2018) (v0.23.4) was used for adapter trimming, low-quality reads and bases removal. De-novo assembly for each sample was carried out with megahit(D. Li et al., 2015) (v1.2.9) using the --presets-meta-sensitive.

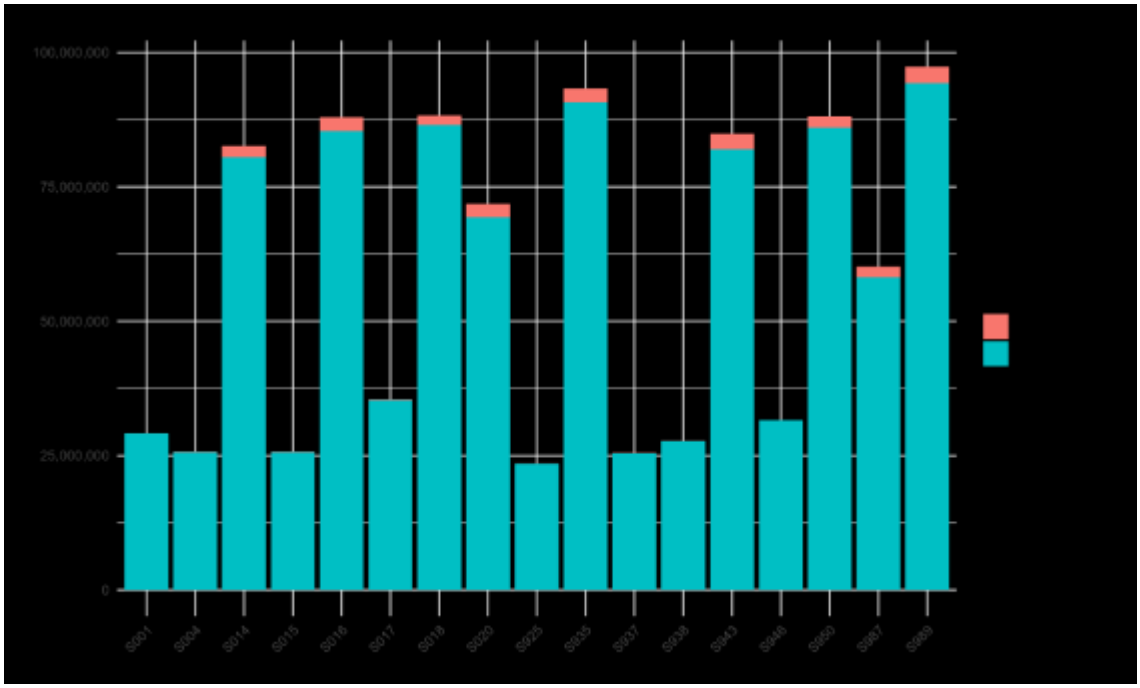

Figure 5: Barplot visualizing the absolute total sequencing depth of each sample depicting low-quality reads that were removed using fastp and retained reads for downstream analysis.

Sample	Retained_reads	Removed_reads
S001	99.644%	0.356%
S004	99.653%	0.347%
S014	97.432%	2.568%
S015	99.679%	0.321%
S016	97.112%	2.888%
S017	99.711%	0.289%
S018	97.970%	2.030%
S020	96.541%	3.459%
S925	99.685%	0.315%
S935	97.335%	2.665%
S937	99.658%	0.342%
S938	99.700%	0.300%
S943	96.573%	3.427%
S946	99.684%	0.316%
S950	97.499%	2.501%
S987	96.628%	3.372%
S989	96.834%	3.166%

Figure 6: Percentage of removed and retained reads per sample.

Assembled contigs were mapped with bwa-mem2(Vasimuddin et al., 2019) (v2.2.1) and alignment data were processed using SAMtools(H. Li et al., 2009) (v1.21) for the extraction of contig coverage and depth. Contigs of at least 1000bp were retained(Ru et al., 2023; R. Wu et al., 2023) for downstream analysis. Binning was not performed as it can introduce false positives(Ru et al., 2023). Retained contigs were then subjected to prediction of viral sequences using the tools VIBRANT(Kieft et al., 2020) (v1.2.1), Virsorter2(Ren et al., 2020) (v2.2.4), DeepVirFinder(Ren et al., 2020) and PPR-Meta(Fang et al., 2019) (v1.1). Cutoffs and cross validation according to literature were performed to reduce false positives in predicted contigs. Virsorter2 predicted contigs were retained with the suggested cutoff of 0.5 and DeepVirfinder with score > 0.9 and $p < 0.05$ cutoff(Ru et al., 2023; R. Wu et al., 2023). Viral quality, completeness and quality were then assessed using CheckV(Nayfach et al., 2021) (v1.0.1). All viral contigs were separated into two categories. 1) High-Confidence: Contigs identified as viral by at least two of the tools and contigs identified by one of the tools and assessed by CheckV as high or medium quality. 2) Low-Confidence: Contigs identified as viral by only one tool and low or not determined quality by CheckV. Only the sequences assigned to the High-Confidence category were considered as viruses. Additionally, coverage from SAMtools was examined to excise any contigs with values of less than 50%. Taxonomy was then assigned by MMseqs2(Gauthier et al., n.d.) (Release 17-b804f) taxonomy module using the NCBI viral RefSeq as references(Ru et al., 2023). Prediction for lifestyle of each contig was performed using Deephage with the suggested cutoff of lifestyle score <0.25 as temperate and >0.75 as virulent, with the values 0.25-0.75 as uncertain(Tange, 2025; S. Wu et al., 2021).

3.2.5. Computational analysis

Metadata files were then created containing information such as contig name, length, taxonomic name, taxid, coverage, depth, mapped reads and lineage. Csv Files were then filtered to isolate identified taxa under the calss of Caudoviricetes from other viruses (Moraru et al., n.d.). Abundance table was created by merging all sample files and adding reads for each unique taxid per sample.

To assess patterns in phage community composition and their relationship with environmental variables, we conducted a series of multivariate and univariate

analyses in R (v4.3.1) using the vegan, tidyverse, and indicspecies packages. Viral abundance tables were generated from metagenomic reads mapped to contigs taxonomically classified within the class *Caudoviricetes*, and only samples from barley cultivation fields (n = 15) were included in the analysis (excludes samples S1 and S20).

To evaluate the influence of soil chemical properties and spatial variables on phage alpha diversity in barley field soils, we quantified three diversity indices: Shannon, Simpson, and taxonomic richness. Linear regression models were constructed for each alpha diversity metric using soil variables (Nitrogen, pH, Organic Matter, Sodium, Electrical Conductivity, Phosphorus, Potassium) and spatial coordinates (Elevation, Latitude, Longitude) as predictors. An initial full model was defined for each metric, and stepwise model selection was performed using the Akaike Information Criterion (AIC) via the stepAIC() function to identify the most parsimonious model. To ensure model stability and interpretability, we assessed multicollinearity among predictors using variance inflation factors (VIFs). Predictors surpassing the threshold (VIF > 5), were iteratively removed and the model re-fitted until all remaining variables met acceptable collinearity thresholds. Final models were summarized using adjusted R^2 and significance values ($p < 0.05$), and the most influential predictors were retained for interpretation. All analyses were conducted in R (v4.4.3), using packages *car*, MASS, broom, and dplyr.

To visualize the influence of significant environmental gradients on phage β-diversity, we applied the ordisurf() function from the vegan package to an NMDS ordination based on Hellinger-transformed community data. Hellinger transformation was first applied using decostand(method = "hellinger") to standardize phage abundances and reduce the impact of zero inflation while preserving ecological distances. We performed PERMANOVA (adonis2) with 999 permutations using Euclidean distance. Predictors included soil chemistry (e.g., nitrogen, OM, pH, EC, etc.) and geographic variables (latitude, longitude, Elevation). Based on these results, we selected variables with a significant contribution to β-diversity ($p \leq 0.05$). NMDS was then conducted on the transformed matrix using Euclidean distance (metaMDS, k = 2, trymax = 100). To explore the spatial distribution of environmental drivers identified as significant by

PERMANOVA, we fitted smoothed response surfaces for Nitrogen and Elevation onto the NMDS ordination using ordisurf(). This function fits a generalized additive model (GAM) of the environmental variable against the NMDS axes, enabling contour visualization of the gradient's alignment with phage community structure. Each variable was plotted separately to facilitate interpretation.

Variance partition of phage β-diversity explained by environmental and spatial factors, we used the varpart() function in the vegan R package. Hellinger-transformed abundance data were used as the response matrix, and predictors were grouped into three explanatory matrices: (i) Nitrogen, (ii) Elevation, and (iii) Geography, the latter derived as the first principal component (PC1) from latitude and longitude. Partial redundancy analysis (partial RDA) was applied to estimate the unique and shared effects of each component. Significance of individual fractions was assessed using permutation-based RDA with 999 permutations. Only positive variance fractions are shown in the Venn diagram.

Aiming to identify phage taxa characteristic of different soil-nitrogen groups, we divided our samples into three equal sized groups (tertiles) based on total soil N concentration, then applied a multipatt-based Indicator Species Analysis. We computed the 0th, 33.3rd, 66.7th, and 100th percentiles of the Nitrogen measurements. Samples were assigned to one of three groups "Low N," "Medium N," or "High N" using the base R cut() function, ensuring roughly equal sample counts per group. The relative-abundance matrix and environmental metadata were aligned by sample ID. A sanity check (stopifnot()) confirmed that row names (samples) matched exactly between the two objects. We applied the multipatt() function from indicspecies, specifying the "r.g" association function and 999 permutations to assess significance (control = how(nperm = 999)).

This method computes, for each phage OTU, an indicator value measuring the fidelity and specificity of its occurrence within each nitrogen group and conducts a permutation test to derive a p-value. Phage OTUs with indicator-test p-values ≤ 0.05 were retained as significant indicators of Low, Medium, or High N soils. For each significant phage, we pulled its raw abundances across all samples, transposed to a sample-by-phage format, and merged with the N-group assignments. We then aggregated abundances by N group and phage, computed total abundance per phage

within each group, and converted to relative abundance (each phage's count divided by the group total). This workflow allowed us to pinpoint phage taxa whose distributions are non-randomly associated with soil-nitrogen levels, providing insights into how nutrient gradients structure viral communities in barley rhizosphere soils.

3. Results

3.1 Viral Contig Prediction

The tools Vibrant, Virsorter2, DeepVirFinder and PPR-Meta were used to predict viral contigs from metagenomically assembled contigs with a minimum length of 1000bp. In total, the tools predicted 2512, 29658, 70978, and 249668 contigs as viral across all samples, and after the application of the cross-validation between them 1546, 6858, 40283, and 44551 contigs were categorized as high confidence respectively. In total, 720, 355, 3852, 1412, 4444, 2158, 2169, 3994, 1088, 4083, 1619, 930, 3694, 1860, 2959, 3321 and 6520 across samples S1, S4, S14, S14, S15, S16, S17, S18, S20, S925, S935, S937, S938, S943, S946, S950, S987 and S989 respectively.

Table 2: Enumeration of total contigs identified by VIBRANT and their confidence level after cross-validation with other tools.

Sample	Total	High_confidence	Low_confidence
S1	455	210	245
S4	17	11	6
S14	191	112	79
S15	28	18	10
S16	172	118	54
S17	58	34	24
S18	136	106	30
S20	317	213	104
S925	46	36	10
S935	166	101	65

Sample	Total	High_confidence	Low_confidence
S937	32	27	5
S938	28	21	7
S943	155	94	61
S946	31	16	15
S950	81	55	26
S987	213	134	79
S989	386	240	146
Total	2,512	1,546	966

Table 3: Enumeration of total contigs identified by VirSorter v2 and their confidence level after cross-validation with other tools.

Sample	Total	High_confidence	Low_confidence
S1	1,793	290	1,503
S4	263	55	208
S14	2,703	597	2,106
S15	560	159	401
S16	1,965	561	1,404
S17	1,140	245	895
S18	2,156	440	1,716
S20	2,935	656	2,279

Sample	Total	High_confidence	Low_confidence
S925	671	155	516
S935	2,426	582	1,844
S937	768	246	522
S938	455	102	353
S943	2,773	589	2,184
S946	840	218	622
S950	1,820	374	1,446
S987	2,015	566	1,449
S989	4,375	1,023	3,352
Total	29,658	6,858	22,800

Table 4: Enumeration of total contigs identified by PPR-Meta and their confidence level after cross-validation with other tools.

Sample	Total	High_confidence	Low_confidence
S1	11,099	673	10,426
S4	2,093	353	1,740
S14	19,006	3,787	15,219
S15	6,190	1,398	4,792
S16	19,682	4,380	15,302
S17	10,875	2,113	8,762

Sample	Total	High_confidence	Low_confidence
S18	15,671	2,132	13,539
S20	26,924	3,917	23,007
S925	5,565	1,065	4,500
S935	21,336	4,020	17,316
S937	7,976	1,755	6,221
S938	4,904	920	3,984
S943	21,523	3,637	17,886
S946	9,188	1,838	7,350
S950	15,459	2,910	12,549
S987	16,802	3,261	13,541
S989	35,375	6,392	28,983
Total	249,668	44,551	205,117

Table 5: Enumeration of total contigs identified by DeepVirFinder and their confidence level after cross-validation with other tools.

Sample	Total	High_confidence	Low_confidence
S1	521	306	215
S4	668	320	348
S14	6,018	3,394	2,624
S15	2,300	1,329	971

Sample	Total	High_confidence	Low_confidence
S16	7,013	4,064	2,949
S17	3,472	2,000	1,472
S18	3,456	1,893	1,563
S20	6,058	3,455	2,603
S925	1,722	980	742
S935	6,336	3,669	2,667
S937	2,735	1,618	1,117
S938	1,546	864	682
S943	5,897	3,264	2,633
S946	3,028	1,708	1,320
S950	4,696	2,701	1,995
S987	5,077	2,922	2,155
S989	10,435	5,796	4,639
Total	70,978	40,283	30,695

3.2 Read Composition

Relative abundance was calculated considering the reads mapped on high-confidence contigs. From the sum of trimmed sequencing reads across all samples, viral reads represented 0.275% Figure 7. From all viral reads, 23.128% were successfully assigned taxonomy under the class of *Caudoviricetes* (Caudoviricetes), 8.183% were eukaryotic viruses (Viral) and 68.689% consisted of reads assigned to unclassified contigs (Unclassified) Figure 8.

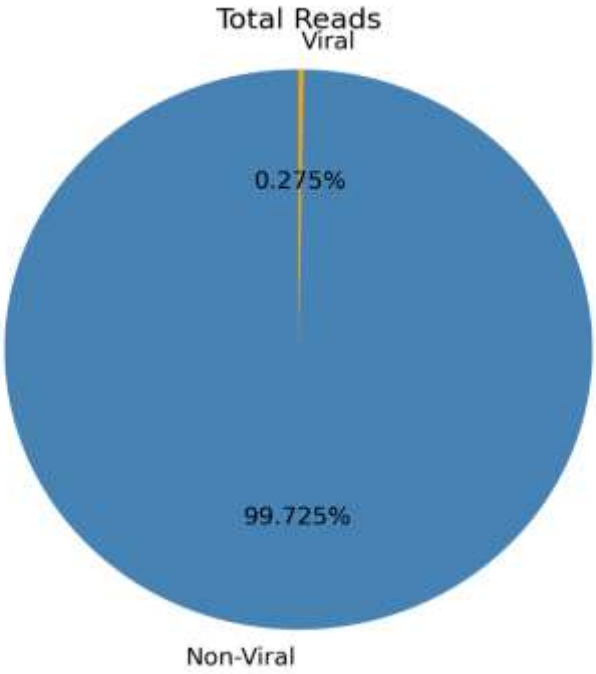

Figure 7: Summarized read count of all samples separated in high-confidence viral and non-viral reads.

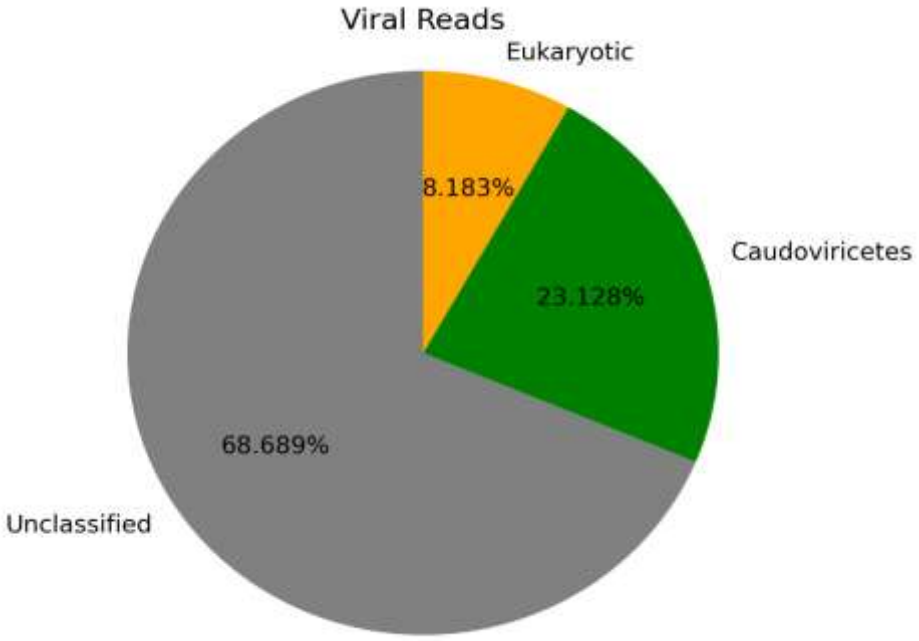

Figure 8: Summarized high-confidence viral read counts of all sequenced samples separating prokaryotic viruses (*Caudoviricetes*), eukaryotic viruses, and viral dark matter(unclassified).

All viral reads were separated for each sample as Eukariotic viral reads for eukaryotic viruses, *Caudooviricetes* reads for reads that were mapped on contigs with assigned taxonomy under the class of *Caudoviricetes* and Unclassified viral reads for contigs that were not assigned taxonomy by MMSeqs2 Figure 9. The percentage of mapped reads to eukaryotic viruses ranged from 2.847% (S4) to 11.654% (S1), excluding sample S20 that showcased 20.997%. *Caudoviricees* reads ranged from 12.764% (S15) to 30.453% (S4) besides sample S1 that 75.939% of its reads were included in this category. Unclassified viral reads represented the majority of viral reads across all samples with percentages of 61.047% to 80.272%, excluding sample S1 with 12.405% of its reads mapping to unclassified contigs. The complete list of percentages for each category per sample is available in **Error! Reference source not found.**.

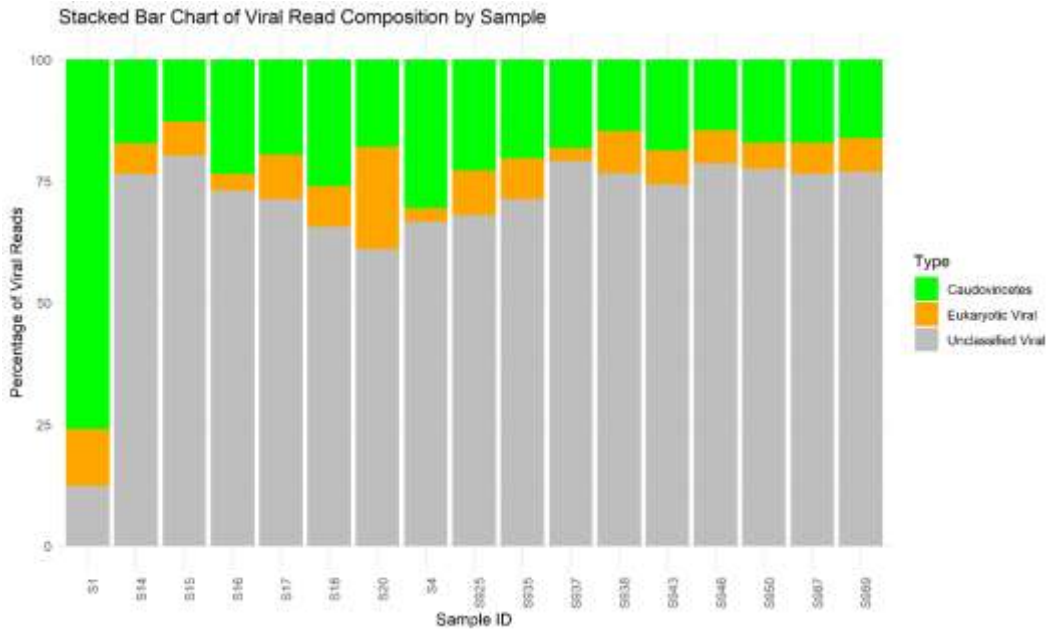

Figure 9: Normalized read composition of viral reads for each sample. Eukariotic viral: reads mapped to contigs identified as eukaryotic viruses. Caudoviricetes: reads mapped to contigs with successful taxonomy assignment to the class of *Caudoviricetes*. Unclassified_reads: reads mapped to contigs with no taxonomy assigned.

Sample_ID	Unclassified_viral_reads	Caudoviricetes_reads	Euk_viral_reads
S1	12.405429792466114	75.93975270448483	11.654817503049053
S4	66.69877970456005	30.45386426889317	2.847356026546778
S14	76.52964095795312	17.171576155242217	6.298782886804666
S15	80.27220077220078	12.764478764478765	6.963320463320463
S16	73.00748005744333	23.47393307778602	3.5185868647706466
S17	71.26191657604967	19.525765141837606	9.212318282112731
S18	65.61278783141947	26.062646546022965	8.324565622557566
S20	61.04750388075574	17.954820284560167	20.99767583468409
S925	68.1256575245272	22.794999143688987	9.079343331783818
S935	71.37567715322301	20.316667494988653	8.307655351788346
S937	79.13154691280029	18.231967579395043	2.636485507804675
S938	76.68491276086212	14.622533926331393	8.692553312806478
S943	74.42487409294975	18.721735849469304	6.853390057580944
S946	78.64790286975717	14.519867549668874	6.832229580573951
S950	77.49739826971796	17.1304228994313	5.372178830850739
S987	76.5520139337174	17.129702934119337	6.318283132163257
S989	76.94878846640664	16.126090920932658	6.925120612660704

Figure 10: Percentages of each category of viral reads per sample **Figure 9**

3.3 Diversity

We observed 233, 91, 541, 218, 565, 339, 445, 738, 206, 613, 253, 152, 544, 292, 444, 512 and 826 viral taxa in samples S1, S4, S14, S15, S16, S17, S18, S20, S925, S935, S937, S938, S943, S946, S950, S987, S989 respectively. Among the viral taxa 190, 81, 427, 149, 461, 244, 346, 548, 149, 474, 213, 106, 436, 219, 343, 401, and 652 taxa were specified as phages under the class *Caudoviricetes* (Richness by Sample Figure 11). In total, 2701 unique viral taxa were identified acrros all samples among them 2064 belonged to the class of *Caudoviricetes.* Furthermore, Alpha Diversity Shannon's index and Simpson's Index ranged between 3.156-5.925 and 0.890-0.993, respectively.

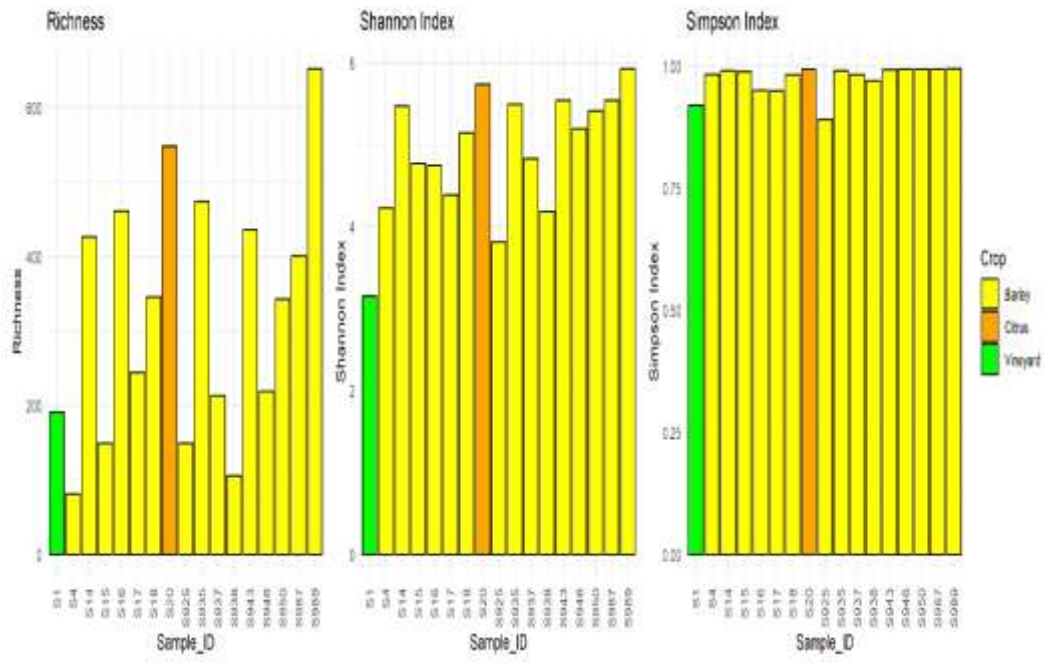

Figure 11: Alpha diversity Sannon index, Simpson index and richness derived from bacteriophage taxa relative abundance.

In sample S1, collected from a vineyard, *Sidiousvirus* accounted for the highest relative abundance at 18.42 %, while in sample S20 from a citrus grove, *Woodruffvirus* was most abundant at 7.03 %. All other samples originated from barley fields, and each exhibited a distinct dominant phage family: *Dolichocephalovirinae* was most abundant in S4 (6.30 %) and S14 (6.73 %), Stanwilliamsviridae peaked in S15 at 5.03 %, *Vividuovirus* reached 21.21 % in S16, and *Vequintavirinae* comprised 18.49 % of S17. *Foxquatrovirus* dominated S18 at 8.51 %, *Straboviridae* led S925 at 4.21 %, and *Stanwilliamsviridae* again dominated S935 at 15.29 %. In S937, *Richievirus* accounted for 10.46 %, *Rimavirus* 13.31 % of S938, and *Weiservirinae* 7.53 % of S946. The remaining barley-field samples showed *Dolichocephalovirinae* as the top family in S943 (5.80 %), S950 (5.09 %), and S987 (4.74 %), and *Bclasvirinae* in S989 (3.90 %).

As the dominant families and bacteriophages across all 17 samples, *Dolichocephalovirinae* emerged as the most frequently observed top family, leading in five samples. *Stanwilliamsviridae* was dominant in two samples and *Sidiousvirus*, *Woodruffvirus*, *Vividuovirus*, *Vequintavirinae*, *Foxquatrovirus*, *Straboviridae*, *Richievirus*, *Rimavirus*, *Weiservirinae*, and *Bclasvirinae* each

topped a single sample. Restricting the analysis to the 15 barley-field samples yields the same pattern: Dolichocephalovirinae dominated five fields, Stanwilliamsviridae two fields, and each of the other eight families dominated one field apiece. This distribution of highest-abundance families provides a clear overview of phage community structure in the vineyard, citrus grove, and barley-field soils, with *Dolichocephalovirinae* most consistently prevailing among the barley-field samples.

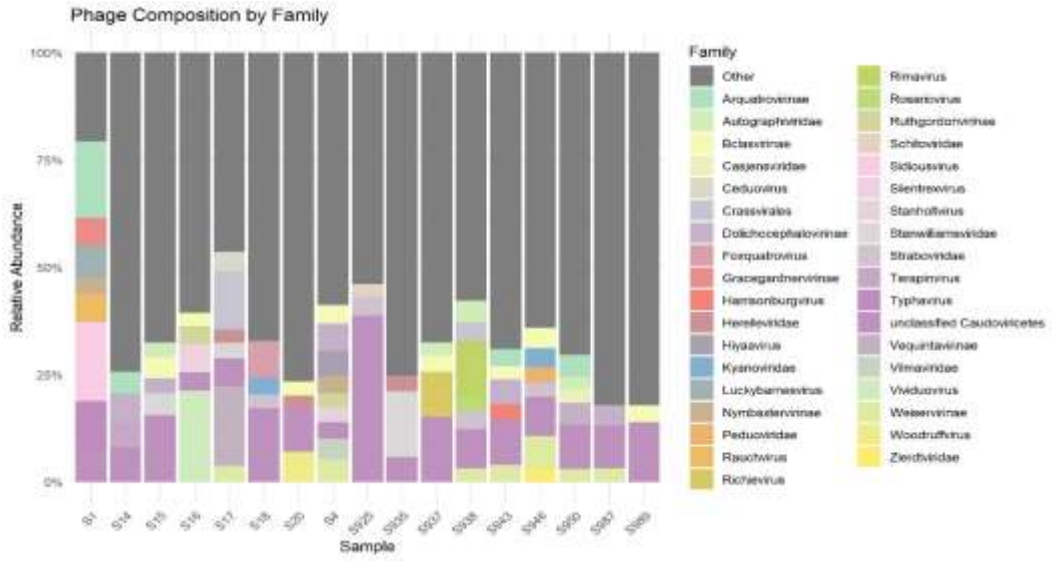

Figure 12: Relative abundance of bacteriophage families and uncategorized phages across soil samples. Stacked bar plots show the phage community composition at the family level, with low-abundance families (<3%) collapsed into the "Other" category (grey). Each bar represents a single sample normalized to total bacteriophage read counts.

3.4 Predicted Lifecycles:

Life cycles of contigs were predicted using DeePhage. The reads mapped to each contig identified as a phage with the same life cycle were added to calculate the abundance of phages with the respective lifestyle. Life cycles were categorized as temperate (lifestyle_score < 0.25), virulent (lifestyle_score > 0.75), and uncertain (lifestyle_score = 0.25 – 0.75). The percentages of mapped reads to temperate phages ranged from 21.558 (S938) to 87.79 (S1), for virulent phages fluctuated between 0.281 (S1) to 26.110 (S15), and reads mapped to phages with less confidence (uncertain) in their lifestyle prediction ranged between 11.929 and 65.127.

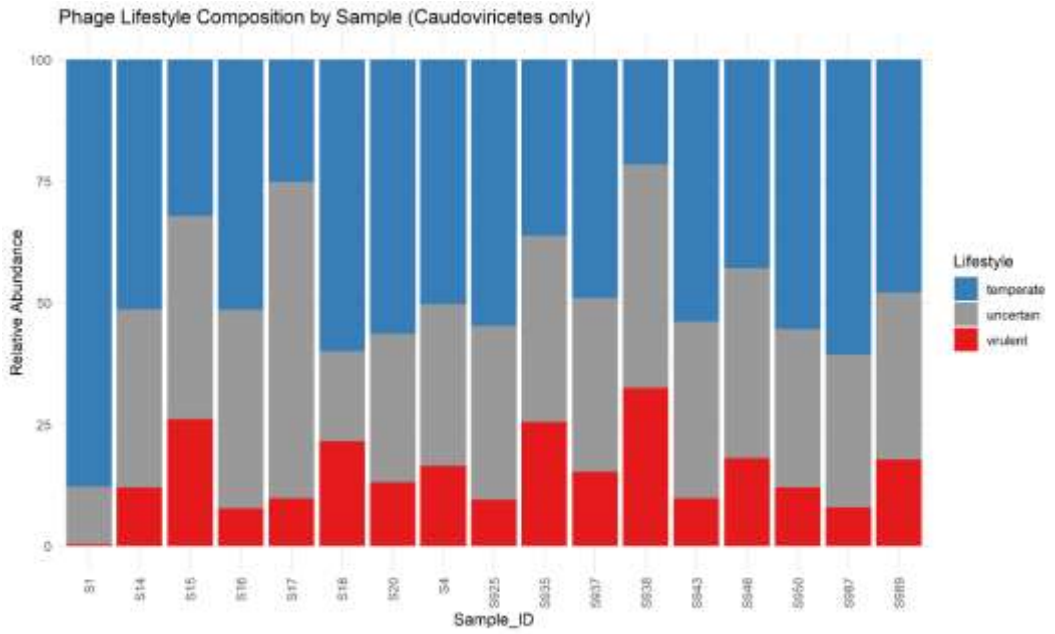

Figure 13: Relative abundance of phage lifestyles per sample. Temperate: reads mapped to phage contigs with lifestyle_score < 0.25. Virulent: reads mapped to phage contigs with lifestyle_score > 0.75. Uncertain: reads mapped to phage contigs with lifestyle_score 0.25-0.75.

Table 6: Percentages of relative abundance from lifecycle based read mapping (**Figure 13**)

Sample_ID	Temperate	Virulent	Uncertain
S1	87.790	0.281	11.929
S14	51.360	12.030	36.610
S15	32.186	26.110	41.704
S16	51.405	7.703	40.891
S17	25.090	9.783	65.127
S18	59.911	21.534	18.555
S20	56.350	13.130	30.520
S4	50.264	16.450	33.286

Sample_ID	Temperate	Virulent	Uncertain
S925	54.696	9.574	35.730
S935	36.206	25.412	38.382
S937	49.009	15.185	35.805
S938	21.558	32.468	45.974
S943	53.814	9.768	36.419
S946	42.880	18.023	39.097
S950	55.415	12.005	32.581
S987	60.611	7.985	31.404
S989	47.884	17.745	34.371

3.5 Statistical Analysis

4.5.1. Correlation of Alpha Diversity Metrics with Soil Variables

Spearman correlation analyses between each α-diversity metrics and seven soil parameters (Nitrogen, pH, Organic Matter, Electrical Conductivity, Elevation, Phosphorus, and Potassium) revealed multiple significant associations ($p \leq 0.05$). Simpson correlated negatively with Nitrogen (r= -0.7012534, p=0.003579736). Shannon also correlated negatively with Nitrogen (r= -0.6261191, p=0.012521605) and Elevation ($\rho = 0.24$, $p = 0.002$).

Table 7: Spearman correlation

Variable1	Variable2	Spearman_r	p_value
Elevation	Shannon	-0.5285714	0.045425102

Variable1	Variable2	Spearman_r	p_value
Nitrogen	Shannon	-0.6261191	0.012521605
Nitrogen	Simpson	-0.7012534	0.003579736

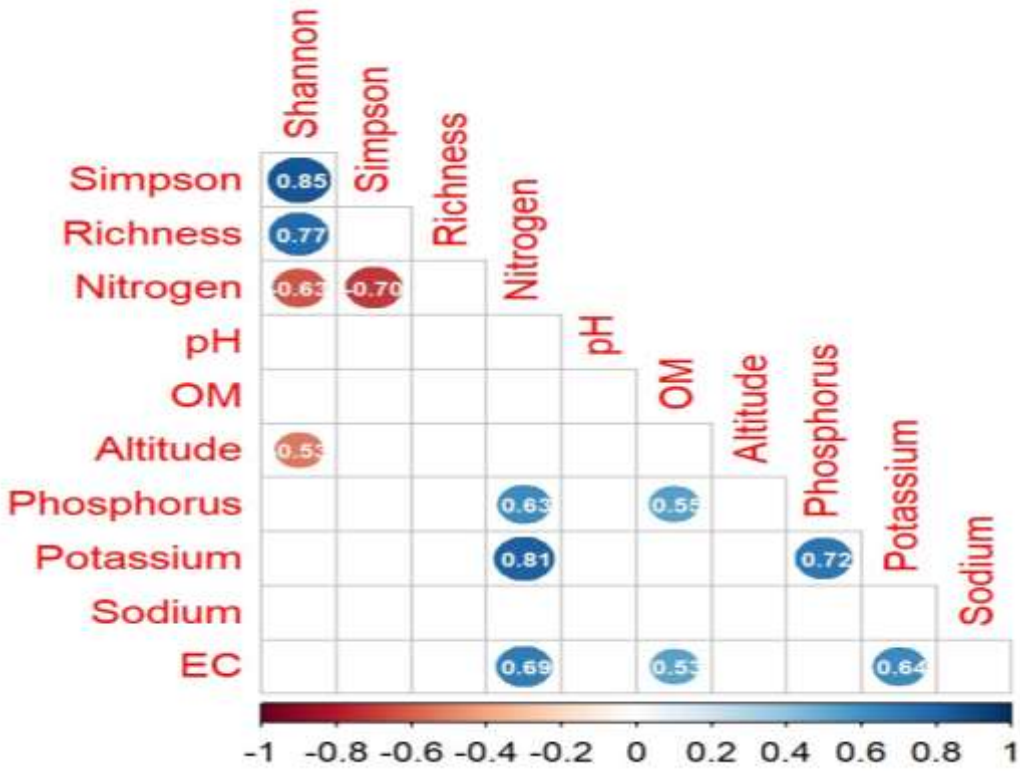

Figure 14: Spearman correlogram visualizing the correlations between α-diversity metrics, spatial gradients, and soil chemical properties.

Linear models were constructed to assess how soil chemical properties and spatial gradients predict alpha diversity metrics (Shannon, Simpson, and Richness). The models selected based on AIC indicated that Shannon Diversity Index was significantly predicted by Nitrogen (β = = -5.21, p = 0.00324) and Elevation (β = -0.00211, p = 0.0406). Simpson Diversity Index was influenced significantly by Nitrogen (β= -0.288, p = 7e-07), Elevation (β= 0.000127, p = 0.000125), Potassium (β = 1.9e-05, p = 0.00175), Sodium (β = -0.000193, p = 0.000714), EC (β = 4.14e-05, p = 0.000126), Latitude (β = 0.143, p = 0.000276), Longitude (β = -0.0281, p = 0.000666). Richness was associated with Elevation (β = -0.666, p = 0.032).

With the use of iterative pruning of Variance Inflation Factors all predictors in the cleaned models exhibited low multicollinearity with all VIF values below 3 (Table 8):

- Shannon model: Nitrogen (VIF = 1.64), Elevation (VIF= 1.27), Phosphorus (VIF= 1.76), Longitude (VIF= 1.19)
- Simpson model: Nitrogen (VIF = 1.38)
- Richness model: Elevation (VIF = 1.00)

After multicollinearity was filtered out by using iterative removal of Variance Inflation Factors (VIF), Shannon Diversity retained Nitrogen (β = -3.76, p = 0.00547) and Elevation (β = -0.00262, p = 0.0186) with R^2 = 0.774 and adjusted R^2 = 0.648. Simpson Diversity retained Nitrogen (β = -0.222, p = 0.00177) with R^2 = 0.774 and adjusted R^2 = 0.605. Richness retained Elevation (β = -0.666, p = 0.032) with R^2 = 0.429 and adjusted R^2 = 0.334.

Table 8: Variance Inflation Factors for Final Alpha-Diversity Models.

Metric	Variable	VIF
Shannon	Nitrogen	1.444976
Shannon	OM	2.270246
Shannon	Elevation	1.470807
Shannon	Sodium	1.664160
Shannon	Moisture	1.200692
Simpson	Nitrogen	1.376923
Simpson	pH	1.398623
Simpson	Elevation	1.352632
Simpson	Sodium	2.714388
Simpson	Latitude	2.650557

Metric	Variable	VIF
Simpson	Longitude	1.314872
Richness	Nitrogen	1.000001
Richness	Elevation	1.000001

Note. All VIFs are below the threshold of 5.

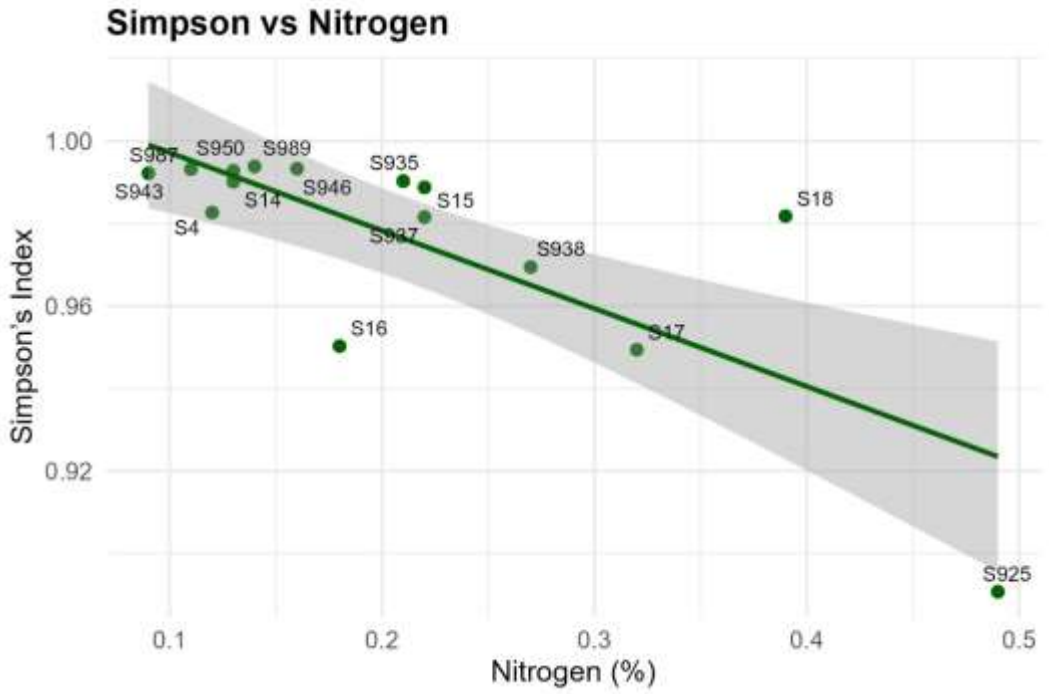

Figure 15: Scatterplot with linear regression line visualizing the inverse correlation of Simpson Diversity Index with soil nitrogen.

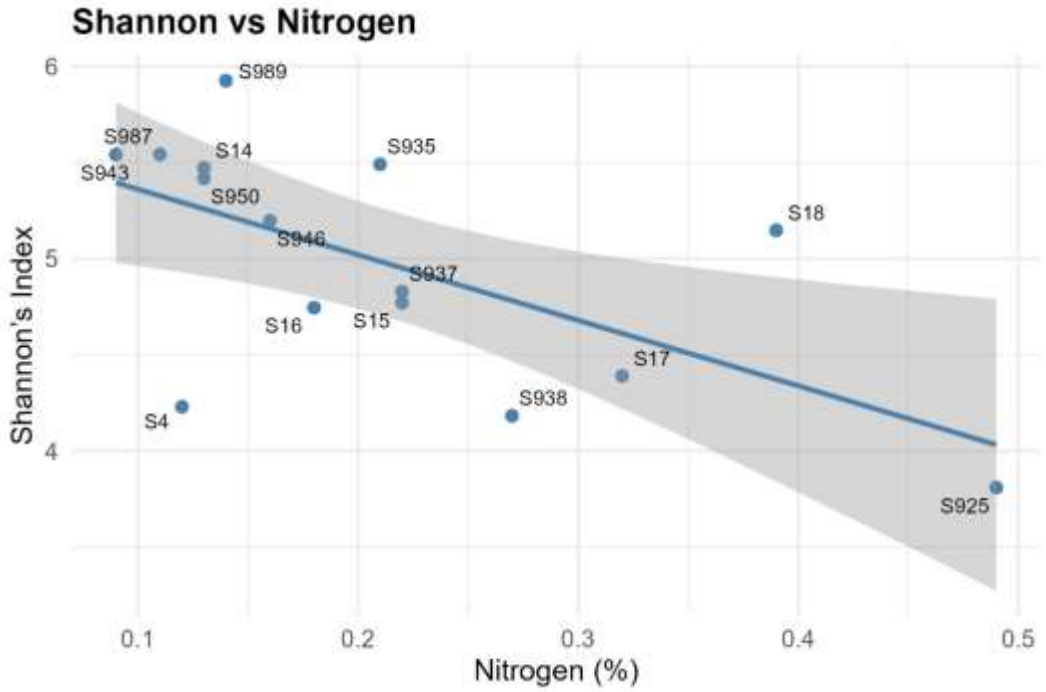

Figure 16: Scatterplot with linear regression line visualizing the inverse correlation of Shannon Diversity Index with soil nitrogen.

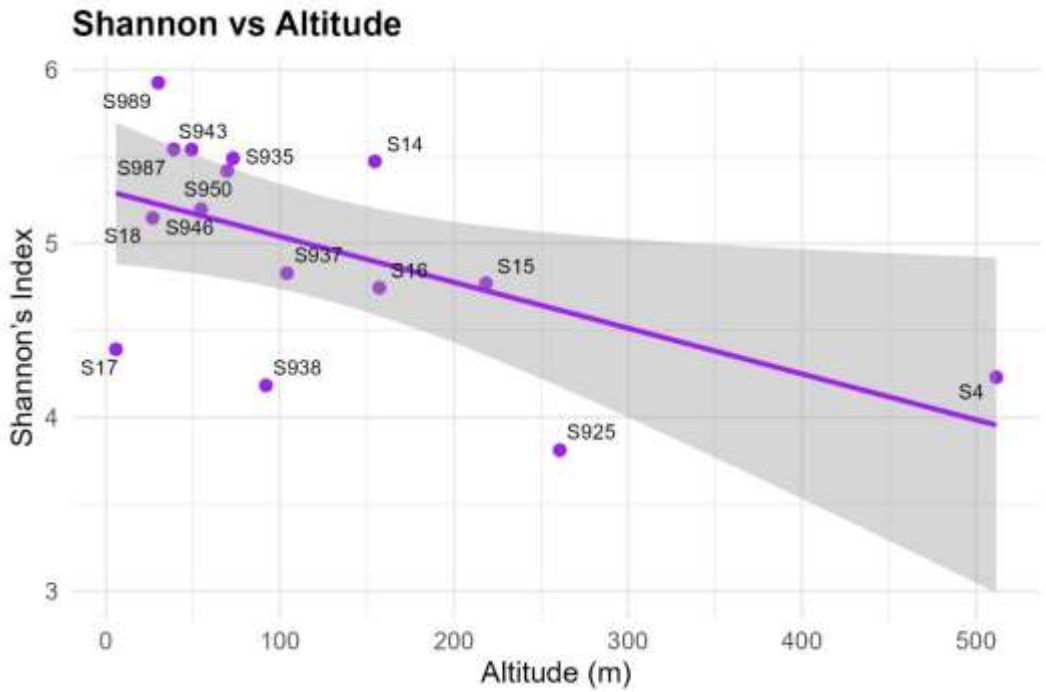

Figure 17: Scatterplot with linear regression line visualizing the inverse correlation of elevation with soil nitrogen.

4.5.2. PERMANOVA

Permutational multivariate analysis of variance (PERMANOVA) on Hellinger-transformed phage community data revealed that Nitrogen ($p = 0.019$) and Elevation ($p = 0.031$) significantly influenced β-diversity across barley field soils. The overall model explained approximately 59% of the variance in phage community composition, with Nitrogen accounting for the most significant proportion individually ($R^2 = 0.081$). Visualization of community composition was performed by overlaying smooth contour surfaces from the ordisurf function onto a two dimensional NMDS ordination of Hellinger-transformed phage-abundance data (stress = 0.161). The nitrogen gradient (blue contours) consistently separated samples along the first NMDS axis, indicating a clear nutrient-driven shift in viral assemblages. Elevation (red contours) also aligned with an orthogonal axis, revealing a secondary spatial structuring of phage communities (Figure 18).

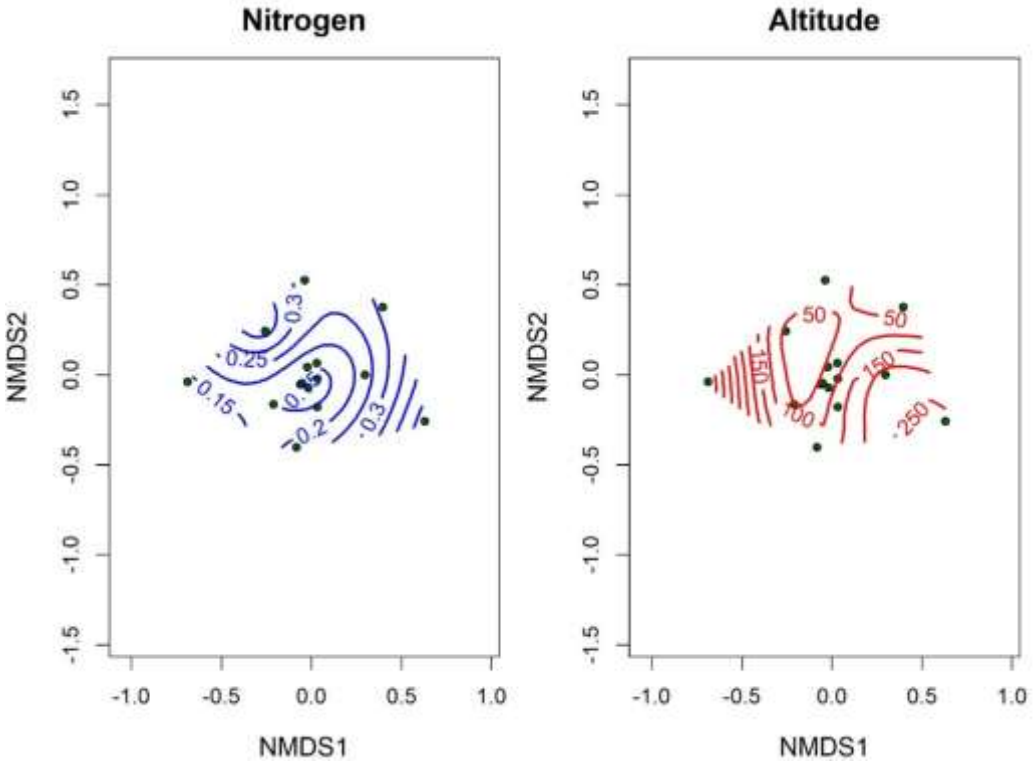

Figure 18: NMDS ordination of Hellinger-transformed phage community data (stress = 0.161) with environmental gradients overlaid using ordisurf. Left) Soil Nitrogen contours (blue = % N) reveal a gradient from low (0.15) to high (0.30) nitrogen concentrations across ordination space. Right) Elevation contours (red = meters) span from low (50 m) to high (250 m).

Table 9: Permutational ANOVA terms and p values for geochemical and spatial factors

Term	Df	SumOfSqs	R2	F	p.value
Nitrogen	1	0.8547356	0.08163207	1.1841052	0.019
pH	1	0.6864777	0.06556249	0.9510096	0.802
OM	1	0.7340657	0.07010741	1.0169355	0.357
EC	1	0.7451550	0.07116650	1.0322980	0.272
Elevation	1	0.8368344	0.07992241	1.1593059	0.031
Latitude	1	0.7624932	0.07282241	1.0563176	0.207
Longitude	1	0.7350780	0.07020410	1.0183379	0.373
Moisture	1	0.6975833	0.06662314	0.9663948	0.695
Residual	6	4.3310455	0.41363928		
Total	14	10.4705856	1.00000000		

4.5.3. Variance Partitioning

Variation partitioning analysis revealed that the unique effect of soil nitrogen accounted for the most significant proportion of explained variance in phage β-diversity (18.23%), followed by pure geography (8.93%) and pure Elevation (8.71%). An additional 4% of variance was explained jointly by all three predictors, while 77.34% of the variance remained unexplained. possibly reflecting unmeasured environmental or biological factors. The results indicate that nutrient status, topographic variation, and spatial configuration each independently shape phage assemblages, with nitrogen as the dominant ecological driver.

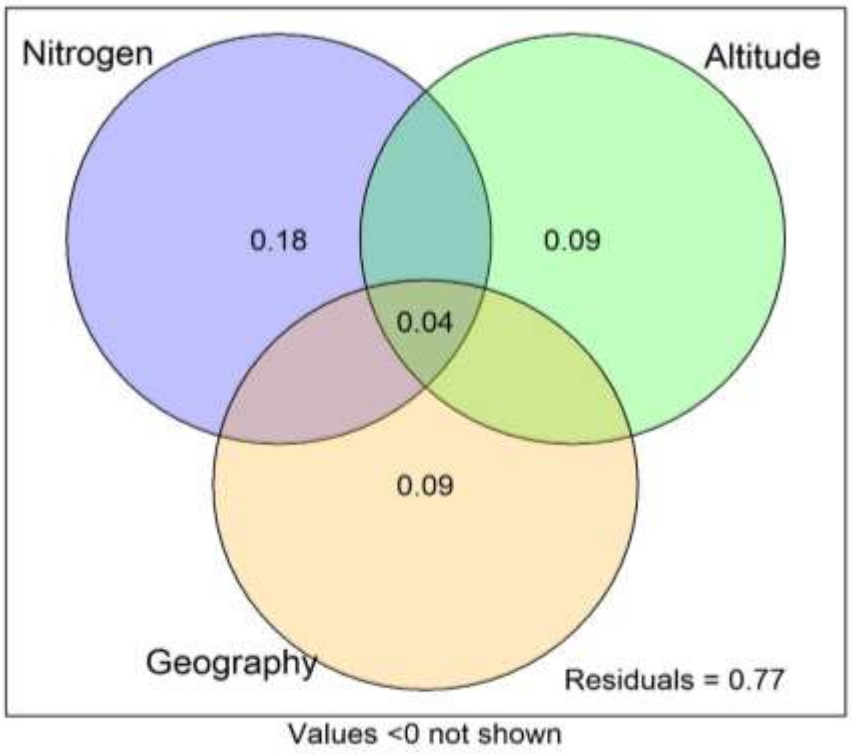

Figure 19: Venn diagram for variance partitioning of Nitrogen, Elevation and Geography (Latitude + Longitude)

4.5.4. Indicator Species Analysis

Indicator species analysis identified phages associated with low, medium, or high nitrogen groups. Samples were divided into three equal-sized groups based on soil nitrogen content: “Low N” (0.09 - 0.13%), “Medium N” (0.14 - 0.22%), and “High N” (0.27–0.49%). Each group contained 5, 6, and 4 samples, respectively. The mean nitrogen concentration for the Low, Medium, and High groups was 0.12%, 0.19%, and 0.37%, respectively. Indicator Species Analysis revealed 35 phage taxa significantly associated with each nitrogen grouping ($p \leq 0.05$). Fifteen significant indicator phages were identified in Low-N soils, 11 in Medium-N soils, and 8 in High-N soils. In Low-N samples, the most abundant indicator was Burkholderia phage phiE125 (MaxRel = 0.19). In Medium-N samples, Mycobacterium phage Serendipitous was the most abundant (MaxRel = 0.211). In High-N soils, 8 indicator phages were detected the most abundant was “uncultured phage cr107_1” (MaxRel = 0.176).

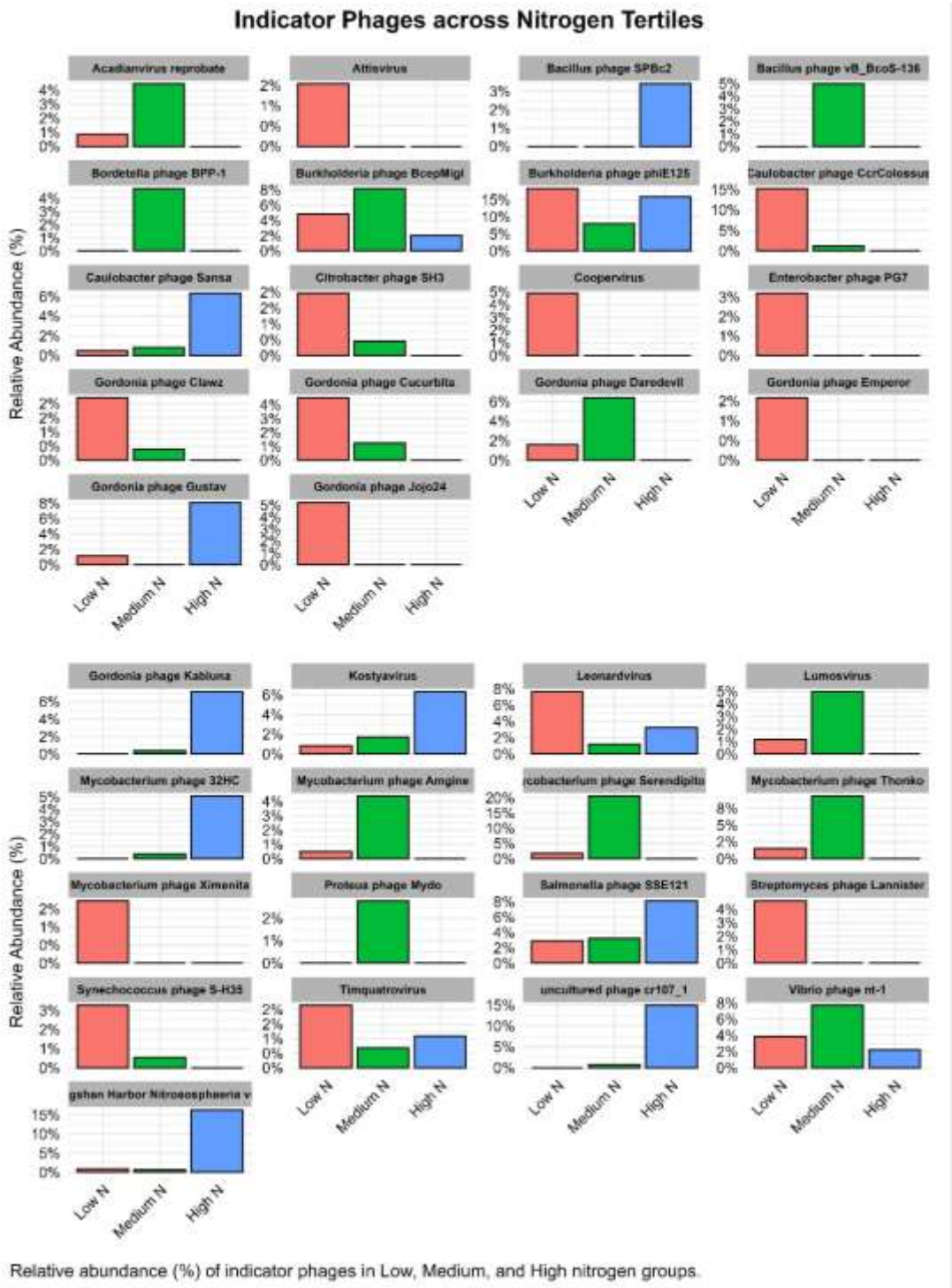

Figure 20: Bacteriophage indicators relative abundance across Nitrogen groups

Table 10: Significant phage indicator taxa across Nitrogen groups.

Phage	Low.N	Medium.N	High.N	index	p.value
Acadianvirus reprobate	0	1	0	2	0.016
Arthrobacter phage Yang	0	1	0	2	0.048
Attisvirus	1	0	0	1	0.035
Bacillus phage SPBc2	0	0	1	3	0.009
Bacillus phage vB_BcoS-136	0	1	0	2	0.008
Bordetella phage BPP-1	0	1	0	2	0.024
Burkholderia phage BcepMigl	0	1	0	2	0.041
Burkholderia phage phiE125	1	0	0	1	0.027
Caulobacter phage CcrColossus	1	0	0	1	0.005
Caulobacter phage Sansa	0	0	1	3	0.009
Citrobacter phage SH3	1	0	0	1	0.047
Coopervirus	1	0	0	1	0.031
Enterobacter phage PG7	1	0	0	1	0.039
Gordonia phage Clawz	1	0	0	1	0.037
Gordonia phage Cucurbita	1	0	0	1	0.020
Gordonia phage Daredevil	0	1	0	2	0.019
Gordonia phage Emperor	1	0	0	1	0.035
Gordonia phage Gustav	0	0	1	3	0.008

Phage	Low.N	Medium.N	High.N	index	p.value
Gordonia phage Jojo24	1	0	0	1	0.035
Gordonia phage Kabluna	0	0	1	3	0.032
Kostyavirus	0	0	1	3	0.010
Leonardvirus	1	0	0	1	0.027
Lumosvirus	0	1	0	2	0.042
Mycobacterium phage 32HC	0	0	1	3	0.029
Mycobacterium phage Amgine	0	1	0	2	0.017
Mycobacterium phage Serendipitous	0	1	0	2	0.021
Mycobacterium phage Ximenita	1	0	0	1	0.035
Proteus phage Mydo	0	1	0	2	0.043
Salmonella phage SSE121	0	0	1	3	0.037
Streptomyces phage Lannister	1	0	0	1	0.027
Synechococcus phage S-H35	1	0	0	1	0.045
Timquatrovirus	1	0	0	1	0.045
Vibrio phage nt-1	0	1	0	2	0.014
uncultured phage cr107_1	0	0	1	3	0.035

4. Discussion

Research on bacteriophages in soil ecosystems has recently gained attention compared to the well-studied marine ecosystem. In the present experiment, we performed shotgun sequencing on 17 individual libraries prepared from eDNA extracted from agricultural field soils across Cyprus. Then, with bioinformatic analysis, viral contigs were predicted and assigned taxonomy. Additionally, using statistical community ecological analysis, we investigated possible effects of soil chemical properties and spatial gradients on phage communities. Although additional investigation is needed to confidently assess phage diversity and the factors driving it, due to limitations in our research, this study represents the first comprehensive attempt to explore bacteriophage diversity in Cyprus agricultural soils using shotgun metagenomics and bioinformatics tools.

Figure 21: Phage Isolation Locations by GPS Coordinates(PhagesDB, n.d.).

4.1 Observed taxa

With phage diversity as the primary purpose of this thesis, we report a total of 2064 unique, high-confidence phage taxa identified across sampled locations, underscoring the rich and largely unexplored viral diversity present in terrestrial ecosystems of the Mediterranean region. Included in the observed taxa, 129

belonged to phages infecting bacteria of the genus *Pseudomonas*, 22 belonged to phages infecting the genus *Rhizobium*, 5 infecting the genus *Listeria,* and 6 infecting the genus *Xylella*. These genera represent both beneficial and pathogenic roles within their respective ecosystems. The *Pseudomonas* genus includes opportunistic pathogens like *P. aeruginosa*, known for its resistance to antibiotics and role in hospital-acquired infections(Qin et al., 2022), while some species showcase the ability to suppress phytopathogens and promote plant growth through various mechanisms(Alattas et al., 2024), such as P. fluorescens(David et al., 2018). *Rhizobium* species are vital in both agricultural and ecological contexts due to their nitrogen-fixing ability(Lindström & Mousavi, 2020). *Listeria* is best known for *L. monocytogenes*, a foodborne pathogen causing listeriosis, particularly in vulnerable populations(Vázquez-Boland et al., 2001). *Xylella fastidiosa* is a quarantine plant pathogen responsible for diseases in grapevines, olive trees, and citrus, significantly impacting agriculture(Loureiro et al., 2023). These genera are ecologically and economically relevant, making their associated phages valuable for biocontrol, therapeutic applications and for the study of their effect on beneficial bacteria populations in soil. Additionally, the potential use of these soil-derived phages is not limited just to agriculture and ecology, as the application of environmental phages in phage therapy has already been documented. In one study, phages isolated from sewage were formulated into a cocktail that successfully treated drug-resistant wound infections in a mammalian model(Regeimbal et al., 2016). Similarly, soil phages have been isolated and characterized for their lytic activity against foodborne pathogens, demonstrating strong specificity and effectiveness, and underscoring their potential for broader antimicrobial applications beyond food safety(Artawinata et al., 2023). The results of viral contig completeness via the CheckV tool presented two contigs as complete. Contig k141_895495 was assigned taxonomy to the *Caudoviricetes* class, and contig k141_1526914 was identified as *Xylella* phage Xfas53, which was the first lysogenic bacteriophage observed for *Xylella fastidiosa(Summer et al., 2010)*. As phages coexist with their hosts, the presence of *Xylella* phages in our samples and the fact that Xfas53 was assessed as complete should raise the question of whether, to our knowledge, this is the first report of *X. fastidiosa* in Cyprus. However, it is important to clarify that a designation of completeness by CheckV does not imply 100% nucleotide identity

to a known reference. Instead, completeness is inferred based on structural features such as direct terminal repeats (DTRs), flanking host DNA, or through gene content and genome length estimates derived from related viral genomes. Although Xfas53 was identified and assessed as complete, the sequence identity to reference phages was moderate, with a score of 75 percent identity, suggesting that it may represent a novel or divergent variant. Consequently, while the presence of a *Xylella* phage could indicate the potential presence of *X. fastidiosa*, this alone does not confirm the bacterium's occurrence in Cyprus. Additional targeted analyses, such as PCR or metagenomic detection of *X. fastidiosa* genomic sequences at a very high confidence, are required to verify its presence. Furthermore, significant abundance of unclassified high-confidence viral reads in our samples suggests the potential existence of novel bacteriophages that were not assigned taxonomy, either due to limited information in our assembly or reference bias.

4.2 Predicted Lifecycles

The predicted lifecycle profiles of phages across samples revealed substantial variation in the relative abundance of temperate, virulent, and uncertain lifestyle phages. Notably, temperate phages dominated most samples, with abundances reaching up to 87.79% in sample S1, while virulent phages were comparatively less abundant, peaking at 32.47% in sample S938. A considerable amount of reads were also mapped to phage sequences with uncertain lifecycle predictions (Deephage score range 0.25-0.75). This inability to predict lifecycles could be associated with the limitations of individual prediction tools, such as reliance on short contig lengths or specific training datasets, which may hinder confident classification. The inclusion of complementary tools that utilize different methodologies such as BACPHLIP(Hockenberry & Wilke, 2021), which predicts lifecycle based on conserved protein domains Replidec(Peng et al., 2022), which combines gene marker detection with a Naive Bayes classifier and includes chronic cycle prediction(Peng et al., 2022), PhaTYP, which applies a transformer-based model to phage protein sequences(Shang et al., 2023), and DeepPL, which uses DNA language modeling via DNABERT(Zhang et al., 2024) could enhance the resolution and confidence of phage lifecycle predictions. Furthermore, this uncertainty likely highlights the impact of reference bias in viral databases and the

inconsistent length/limited information in metagenomicaly assembled contigs through megahit. Reference bias refers to the fact that viral databases are poor compared to bacterial databases, and many of the included phages might not be characterized to a level that is satisfactory for the tools to perform predictions optimally. In addition, challenges in DNA extraction directly from soil, fragmentation, and low coverage of soil-derived metagenomes have already been researched in previous studies and can lead to contigs with less than the required information such as hallmark genes(Anthony et al., 2024; Wydro, 2022). These challenges underscore the need for integrative approaches and the current limitations in viral metagenomic studies.

The prevalence of temperate phages, especially in samples such as S1, S18, and S987, reflects a higher rate of lysogeny, particularly under environmental stress conditions such as drought or nutrient fluctuations. This pattern is consistent with the Piggyback-the-Loser and Piggyback-the-Winner models that underscore the tendency of phages to opt for a temperate strategy to persist within their hosts rather than lyse them under different biotic and abiotic conditions. Samples S938, S15 and S935 showed a relative abundance of lytic phages at 26.110%, 25.412% and 32.468%, respectively with only S938 presenting increased abundance of lytic phages compared to temperate ones. Overall, the variation in phage lifestyles across samples highlights potential shifts in phage-host dynamics. The dominance of temperate phages in many samples suggests that lysogeny may be the prevailing phage-host association strategy, possibly influenced by stress factors such as drought or nutrient deficiency. This trend is consistent with literature as previous studies have also revealed low soil moisture and precipitation to be associated with increased lysogeny(R. Wu et al., 2021, 2023). These findings support that phage behavior strategies reflect responses to local ecological conditions. Further investigation, incorporating host community profiles and environmental metadata, would help clarify the drivers of these lifecycle patterns.

4.3 Alpha diversity

The analysis of alpha diversity metrics in relation to soil environmental variables revealed that nitrogen concentration is a key predictor of phage community

structure in the sampled soils. Both Shannon and Simpson diversity indices exhibited strong and statistically significant negative correlations with nitrogen levels, indicating that increasing soil nitrogen is associated with a decline in phage diversity. Linear regression modeling further supported this trend, with nitrogen emerging again as a significant predictor for both Shannon ($\beta = -3.76$, $p = 0.00547$) and Simpson ($\beta = -0.222$, $p = 0.00177$) models. The consistency between correlation and model-based approaches reinforces the robustness of this relationship. Shannon and Simpson indices are commonly used measures of alpha diversity, with Shannon reflecting both richness and evenness of taxa in a community, and Simpson emphasizing the dominance or evenness by measuring the likelihood that two randomly selected individuals belong to different taxa. As Shannon and Simpson indices decrease with the increase of nitrogen concentration, this could be an indication that evenness of phages communities is decreasing while dominant phage taxa are becoming more prevalent.

From an ecological perspective, the decrease in diversity suggests a shift toward community dominance, where a few phage taxa prevail over others under elevated nitrogen conditions. Nitrogen enrichment has been documented to stimulate the growth of fast-growing bacterial taxa, potentially reducing host diversity and altering the availability of hosts for phages. This selective environment may favor specific host-phage pairs, leading to reduced evenness and richness in the phage community. Under increasing nitrogen content in soil, Kill-the-Winner would seem the dominant phage strategy. However, that would insinuate an increase in Bacteriophage communities' evenness, which does not align with the inverse correlations observed in our samples. Interestingly, the year and season that we collected our samples showcased extreme drought conditions in Cyprus. Previous research has shown that drought can “simplify” microbial diversities as specific drought-resistant taxa can dominate over less resistant ones. In combination with increasing soil nitrogen, proliferation might take place (Garg et al., 2025; Liu et al., 2025; Metze et al., 2023). Low soil moisture levels at the sampling sites introduce an important ecological constraint that favors lysogeny as a persistence strategy. This is also supported by the increased lysogeny trend in our predicted lifecycles data, aligning more closely with Piggyback-the-Winner and Piggyback-the-Loser

models. Both models are associated with increased lysogeny as under stress, virulent phages switch to Pseudolysogeny, aiming to maintain their integrity so they "Piggyback" their declining host until more favorable conditions occur. In contrast, true lysogenic phages in richer nitrogen can "behave" according to the Piggyback-the-Winner (PtW) model, utilizing drought-resistant taxa that can thrive comparatively in such conditions. Furthermore, lysogenic phages can be more abundant under drought by "induction," where they lyse hosts under stress in search of more suitable ones.

Combining the results of alpha diversity association with nitrogen, low soil moisture, and the confirmed increased reads mapped to temperate phages, it is possible to narrow down the possibility of which model is prevalent in barley fields under drought conditions. However, it is worth mentioning that for the prediction of lifecycles, Deephage was originally trained using deep learning algorithms on curated databases, which makes it unsuitable for identifying the state of a phage or prophage in real time. Therefore, we can not accurately assess if virulent phages are in a lytic state or pseudolysogeny and vice versa. Considering these multiple variables, we assume that based on Bacteriophage biology, the scenario with the highest possibility is that PtL and PtW might apply both to this environment under drought conditions. However, due to the design of Deephage increased lysogeny can only be attributed to induced lysogenic phages that exist free in the soil matrix, complemented by the PtW model.

In this study, a negative correlation was also observed between elevation and phage Shannon diversity, suggesting that phage communities become less diverse with increasing Elevation. This trend could reflect ecological simplification, whereby environmental filtering or stress limits the diversity of bacterial hosts, which in turn constrains phage diversity. This aligns with the broader ecological principle that elevation can act as a group of multiple factors, influencing temperature, moisture, nutrient composition, and plant–microbe interactions as soil type and environmental conditions change with the increase in Elevation. However, literature paints a complex and contrasting picture. A metatranscriptomic study in 2023(Merges et al., 2023) revealed that while bacterial activity declined with elevation, phage activity did not, suggesting a possible dissociation between host

and viral dynamics in high-Elevation soils. Additionally, another research by Siles J. et al. reported increased microbial abundance and respiration with Elevation in nutrient-rich forest soils, indicating that microbial responses to elevation are strongly mediated by site-specific conditions such as organic matter availability(Siles et al., 2016). Considering those results and the multiple variables checked by these publications, it seems that further study is needed to assess the true effect of elevation on phage communities. Additionally, multiple factors should be tested for significant effect such as temperature and solar UV rradiation. Although not in alignment with previous literature, the result of this study could provide a meaningful step in understanding the change in phage communities along different elevational gradients, as the conditions and soil types studied in this case are not consistent with previous literature. The study of Elevation as a predictor of phage communities α-diversity along different elevational gradients has not been assessed previously in agricultural fields under drought conditions and literature on this subject is very limited.

4.4 Beta diversity

PERMANOVA results revealed both soil nitrogen concentration and Elevation significantly influenced phage β-diversity across the sampled barley fields. Nitrogen was the strongest individual predictor, explaining approximately 8.2% of the variance in phage community composition ($p = 0.019$), while Elevation accounted for a similar fraction (7.9%, $p = 0.031$). These results suggest that multiple independent environmental gradients shape phage community turnover.

The NMDS ordination, overlaid with smooth ordisurf contours, visually reinforced these findings. Nitrogen structured samples along the first NMDS axis, indicating a consistent shift in community composition across a gradient of nutrient enrichment. As nitrogen availability influences bacterial community composition and growth rates, the pool of potential phage hosts and infection dynamics could be altered. Higher soil nitrogen content may favor specific bacteria, promoting the dominance of phages adapted to these fast-growing hosts or altering the balance between temperate and lytic lifestyles. Elevation, meanwhile, aligned more with the second NMDS axis, suggesting it independently structures phage communities,

potentially through associated shifts in soil moisture, temperature, or vegetation. Previous studies have shown that microbial communities at higher elevations may exhibit reduced host availability or functional diversity, and similar patterns may influence phage turnover. The overall stress value of 0.161 indicates a reasonable ordination fit, lending confidence to the visualization and the interpretation of phage community structuring. Recent studies investigating soil viral ecology against environmental gradients reinforce the relevance of this study's focus on elucidating phage community composition drivers. For example, Muscatt et al. examined viral communities across a soil depth profile and reported clear vertical stratification of viral assemblies, suggesting that abiotic gradients like oxygen availability and nutrient stratification strongly influence viral diversity and host-virus interactions(Muscatt et al., 2023). While this study focused on spatial depth rather than elevation or nutrients, the underlying theme that viral community turnover is tightly linked to environmental gradients resonates strongly with our findings.

Similarly, Coclet et al. demonstrated that seasonal snowmelt events significantly reshaped both DNA and RNA viral communities in a high-Elevation watershed. Their findings revealed significant changes in viral activity and composition correlated to environmental variability, further supporting that viral communities respond to ecological shifts(Coclet et al., 2023). Unlike our study, which found declining diversity and compositional shifts along stable spatial gradients (nitrogen and elevation), their work highlights temporal drivers of viral community structure. These studies highlight the importance of incorporating environmental context into soil virome research and provide external support for the ecological relevance of the observed patterns.

4.5 Variance partitioning

Variance partitioning analysis provides deeper insight into the relative and shared contributions of nitrogen, Elevation, and geographic location to phage β-diversity. Among these, soil nitrogen emerged as the most influential variable, explaining 18.2% of the variation independently, reinforcing the strong structuring role of nutrient availability identified earlier in the PERMANOVA and NMDS ordination.

This contribution suggests that nitrogen acts as a powerful ecological filter, potentially by selecting specific bacterial hosts or phage lifestyles better adapted to given conditions during the sampling period. In contrast, geography (latitude and longitude combined) and Elevation each explained a smaller but comparable portion of variance (approximately 9% each) when considered independently. These findings support the idea that spatial structuring and topographic variation contribute to phage community changes but are secondary to nitrogen in explanatory power. The geographic signal likely captures spatial heterogeneity in unmeasured environmental gradients such as plant composition, land use history, or microclimate differences that influence microbial and viral communities beyond the direct effects of elevation. Furthermore, 4% of the variation in phage composition was explained jointly by all three variables, highlighting overlap and interaction among ecological factors. However, 77.3% of the variation remained unexplained. The remaining fraction may reflect the need for a multivariate analysis considering unmeasured biological and environmental factors. These results reinforce that phage communities are structured by a complex interaction of environmental and spatial variables, with nutrient status as the dominant force. Yet, the large unexplained component also underscores the limitations of snapshot-based environmental surveys and highlights the need for future studies to incorporate functional, host-linked, and temporal data such as metatranscriptomics, prophage activation patterns, or fine-scale biotic interactions to capture a more complete picture of soil phage ecology. Delving deeper into the association of phages with soil nutrients and especially nitrogen a study in 2020 assessed the effect of phage communities on soil nitrogen. They performed a reciprocal transplant experiment which showed significant effect of phages on soil nitrogen, therefore suggesting that phages are not just driven by the interactions of soil nitrogen with their host but under specific conditions phage activity may regulate nitrogen concentrations(Braga et al., 2020).

4.6 Indicator species analysis

Bacteriophages have previously been proposed as bioindicators in monitoring context, such as assessing fecal contamination or sewage pollution(Rogovski et al., 2021). While such applications focus on their persistence and detectability as

proxies for contamination, our use of indicator species analysis extends this concept to an ecological frame. Identification of 35 significant phage indicators across nitrogen tertiles provides strong evidence that soil nitrogen levels selectively shape the composition of phage communities. By dividing samples into three nitrogen-based groups, the indicator species analysis revealed distinct viral taxa preferentially associated with low, medium, or high nitrogen environments, indicating specialization and potential host-nutrient interactions.

In low-nitrogen soils, 15 indicator phages were identified, including several infecting genera such as Burkholderia, Citrobacter, and Gordonia. Notably, Burkholderia phage phiE125 showcased the highest relative abundance (42.1%) among all indicators, indicating that the possible host may be more competitive under nitrogen-limited conditions. This could reflect a broader ecological strategy where oligotrophic bacterial hosts (and their associated phages) dominate in nutrient-poor soils, potentially through efficient nutrient acquisition or stable lysogenic associations. In contrast, medium-nitrogen soils were characterized by 11 unique phage indicators, including Mycobacterium phage Serendipitous (22.2% relative abundance) and Gordonia phage Daredevil. The intermediate nutrient context may support a mix of oligotrophic and copiotrophic microbial hosts, reflected in the co-occurrence of phages linked to both lifestyles. This intermediate zone could thus act as a transition point in phage community structure, supporting both lytic and temperate dynamics depending on local host availability and competitive interactions. High-nitrogen soils yielded 8 phage indicators, the most abundant being the uncultured phage cr107_1 (15.6%). Several of these phages were associated with genera like Mycobacterium, Gordonia, and Salmonella, which include fast-growing or opportunistic bacteria. The presence of these taxa in high-nutrient environments may indicate a shift toward copiotrophic hosts and associated viral taxa adapted to high-biomass, high-turnover systems. These patterns align with observations from other soil microbiome studies, where nutrient enrichment has been shown to reduce diversity while favoring specific hosts, optimized for nutrient-rich conditions(Bachtsevani et al., 2021; Duan et al., 2022; Shen et al., 2010). Indicator phages across nitrogen levels underscore the significant effect of soil nutrient status, with distinct viral populations emerging in both phages and host

respond to environmental conditions. These results complement the β-diversity findings from PERMANOVA and NMDS and highlight the value of phage-level taxonomic resolution in uncovering fine-scale ecological structuring. Future studies that link these phage indicators to their host taxa using metagenome-assembled genomes (MAGs) or CRISPR spacer matching would help confirm whether the observed patterns reflect host distribution shifts, lifestyle transitions (e.g., lytic to lysogenic), or viral adaptations to nutrient gradients.

5. Conclusions

Bacteriophages regulate bacterial community abundances and drive the evolutionary process through lytic cycles and multiple forms of lysogeny, their unique functions have intrigued the attention of researchers worldwide. As phages in terrestrial ecosystems still lack research compared to aquatic ones, our research provides the first report on bacterial communities in Cyprus' agricultural soils and their possible association with environmental factors. With the use of NGS (shotgun metagenomics) and multiple bioinformatic tools, we identified 2064 unique phage taxa across our samples. Lifecycle predictions painted a trent of increased lysogeny across most of our samples. Alpha diversity modeling and Spearman's correlation revealed soil nitrogen concentration as a significant predictor of Shannon's and Simpson's diversity indices, and elevation as a significant predictor of Shannon's index. Further analysis of β-diversity also showcased nitrogen and elevation as significant factors in explaining the variance between barley field soil samples with variance partitioning analisys showcasing approximately 18% of variance explained by nitrogen 9% was explained by elevation and another 9% from geographical coordinates. Those factors explained a further 4% fraction jointly, while 77.3% of variance remained unexplained. In addition Indicator Species Analysis identified 35 taxa as indicators of nitrogen content in soil. Fifteen indicator phages were identified for low nitrogen group, 11 unique phage indicators for the medium nitrogen group, and eight phages for the high nitrogen group. Nitrogen inverse correlation with α-diversity indices, significance in explaining between sample variance, along with increased lysogeny observation and drought conditions, consistently paint the picture of Piggyback-the-Winner model being the dominant strategy in our samples and possibly induction of lysogenic phages might have occurred. Though consistent with possible phage-host dynamics under given conditions, host prediction, bacteria identification in our samples and host contamination in viral sequences are needed to increase confidence in our conclusion. Furthermore, the inverse correlation between elevation and Shannon's index need further investigation. Only one publication has evaluated elevational grdients agains phage so far and with results contradicting the ones observed in this study. However the as soil sample type and environmental condition compared to

ours indicate that under drought condition different dynamics might take place. Certain conditions have hindered the optimal discovery of phage diversity and richness. As previously mentioned during the sampling period extreme drought conditions were at hand. Low levels of soil moisture were expected to drastically reduce phage recovery from soil, as low moisture is associated with low bacterial activity and therefore low phage activity. Additionally, viral sequence recovery could benefit from deeper sampling. Moving on to sample preparation, elution of VLPs from soil using optimized buffers, Tangential Flow Filtration systems, and concentration of VLPs in pellets using an ultracentrifuge prior to eDNA extraction could further increase phage recovery, though this methodology would probably require soil samples with more optimal moisture levels that could favor Kill-the-Winner model. Handling a metavirome instead of bulk soil metagenome in downstream bioinformatics analysis could be more time-efficient, with decreased effort in cross-validation. At the step of de-novo assembly, even though megahit is a tool widely used to metagenomicaly assemble contigs, SPADES with the argument --meta is documented to be more sensitive to identifying small genomes like viruses. The use of SPADES was infeasible at the time due to the extreme computational power required by the tool. Further steps that could improve this study would be increasing samples with diverse conditions among them (soil chemical properties, structure, spatial intervals, etc.). An increase in sample size and replicates could provide multiple ways of improving this study, like increasing the confidence in variance and diversity driving factors. Additionally, replicates would provide the opportunity to co-assemble multiple samples together for better identification and analysis of phage communities for better and more complete genome retrieval. Being the first report of phage diversity analysis in this country, and with unique environmental settings, this thesis provides significant prospects for researching terrestrial phages. Following this study the next logical step would be to carry out a replication of his research in conditions with optimal precipitation and soil moisture to assess the differences in phage-host dynamics under different conditions. Additionally, a time-series experiment would provide important insights into phage community changes. Most importantly, the focus should be on reassessing the significance of Elevation in phage community shaping, as literature is very limited in this setting.

BIBLIOGRAPHY

Abedon, S. T. (2008). Ecology of Viruses Infecting Bacteria. In *Encyclopedia of Virology* (pp. 71–77). Elsevier. https://doi.org/10.1016/B978-012374410-4.00745-7

Ackermann, H.-W. (2005). Classification of Bacteriophages. In *The Bacteriophages* (pp. 8–16). Oxford University PressNew York, NY. https://doi.org/10.1093/oso/9780195168778.003.0002

Adriaenssens, E., & Brister, J. R. (2017). How to Name and Classify Your Phage: An Informal Guide. *Viruses*, *9*(4), 70. https://doi.org/10.3390/v9040070

Alattas, H., Glick, B. R., Murphy, D. V., & Scott, C. (2024). Harnessing Pseudomonas spp. for sustainable plant crop protection. *Frontiers in Microbiology*, *15*. https://doi.org/10.3389/fmicb.2024.1485197

Al-Shayeb, B., Sachdeva, R., Chen, L.-X., Ward, F., Munk, P., Devoto, A., Castelle, C. J., Olm, M. R., Bouma-Gregson, K., Amano, Y., He, C., Méheust, R., Brooks, B., Thomas, A., Lavy, A., Matheus-Carnevali, P., Sun, C., Goltsman, D. S. A., Borton, M. A., … Banfield, J. F. (2020). Clades of huge phages from across Earth's ecosystems. *Nature*, *578*(7795), 425–431. https://doi.org/10.1038/s41586-020-2007-4

Anthony, W. E., Allison, S. D., Broderick, C. M., Chavez Rodriguez, L., Clum, A., Cross, H., Eloe-Fadrosh, E., Evans, S., Fairbanks, D., Gallery, R., Gontijo, J. B., Jones, J., McDermott, J., Pett-Ridge, J., Record, S., Rodrigues, J. L. M., Rodriguez-Reillo, W., Shek, K. L., Takacs-Vesbach, T., & Blanchard, J. L. (2024). From soil to sequence: filling the critical gap in genome-resolved metagenomics is essential to the future of soil microbial ecology. *Environmental Microbiome*, *19*(1), 56. https://doi.org/10.1186/s40793-024-00599-w

Artawinata, P. C., Lorraine, S., & Waturangi, D. E. (2023). Isolation and characterization of bacteriophages from soil against food spoilage and foodborne pathogenic bacteria. *Scientific Reports*, *13*(1), 9282. https://doi.org/10.1038/s41598-023-36591-6

Bachtsevani, E., Papazlatani, C. V., Rousidou, C., Lampronikou, E., Menkissoglu-Spiroudi, U., Nicol, G. W., Karpouzas, D. G., & Papadopoulou, E. S. (2021). Effects of the nitrification inhibitor 3,4-dimethylpyrazole phosphate (DMPP) on the activity and diversity of the soil microbial community under contrasting soil pH. *Biology and Fertility of Soils*, *57*(8), 1117–1135. https://doi.org/10.1007/s00374-021-01602-z

Balogh, B., Canteros, B. I., Stall, R. E., & Jones, J. B. (2008). Control of Citrus Canker and Citrus Bacterial Spot with Bacteriophages. *Plant Disease*, *92*(7), 1048–1052. https://doi.org/10.1094/PDIS-92-7-1048

Balogh, B., Jones, J. B., Momol, M. T., Olson, S. M., Obradovic, A., King, P., & Jackson, L. E. (2003). Improved Efficacy of Newly Formulated Bacteriophages for Management of Bacterial Spot on Tomato. *Plant Disease*, *87*(8), 949–954. https://doi.org/10.1094/PDIS.2003.87.8.949

Balogh, B., Jones, J., Iriarte, F., & Momol, M. (2010). Phage Therapy for Plant Disease Control. *Current Pharmaceutical Biotechnology*, *11*(1), 48–57. https://doi.org/10.2174/138920110790725302

Banerjee, S., & van der Heijden, M. G. A. (2023). Soil microbiomes and one health. *Nature Reviews Microbiology*, *21*(1), 6–20. https://doi.org/10.1038/s41579-022-00779-w

Barylski, J., Kropinski, A. M., Alikhan, N.-F., Adriaenssens, E. M., & Ictv Report Consortium. (2020). ICTV Virus Taxonomy Profile: Herelleviridae. *The Journal of General Virology*, *101*(4), 362–363. https://doi.org/10.1099/jgv.0.001392

Bergh, Ø., BØrsheim, K. Y., Bratbak, G., & Heldal, M. (1989). High abundance of viruses found in aquatic environments. *Nature*, *340*(6233), 467–468. https://doi.org/10.1038/340467a0

Bi, L., Yu, D., Du, S., Zhang, L., Zhang, L., Wu, C., Xiong, C., Han, L., & He, J. (2021). Diversity and potential biogeochemical impacts of viruses in bulk and rhizosphere soils. *Environmental Microbiology*, *23*(2), 588–599. https://doi.org/10.1111/1462-2920.15010

Binetti, A. G., Del Río, B., Martín, M. C., & Alvarez, M. A. (2005). Detection and characterization of Streptococcus thermophilus bacteriophages by use of the antireceptor gene sequence. *Applied and Environmental Microbiology*, *71*(10), 6096–6103. https://doi.org/10.1128/AEM.71.10.6096-6103.2005

BLAKELY, G. W. (2004). *A Genetic Switch, Third Edition, Phage Lambda Revisited* . M. Ptashne. Cold Spring Harbor Laboratory Press. 2004. 154 pages. ISBN 0 87969 716 4. Price $39 (paperback). *Genetical Research*, *84*(3), 193–194. https://doi.org/10.1017/S0016672304227276

Braga, L. P. P., Spor, A., Kot, W., Breuil, M. C., Hansen, L. H., Setubal, J. C., & Philippot, L. (2020). Impact of phages on soil bacterial communities and nitrogen availability under different assembly scenarios. *Microbiome*, *8*(1). https://doi.org/10.1186/s40168-020-00822-z

Breitbart, M. (2012). Marine Viruses: Truth or Dare. *Annual Review of Marine Science*, *4*(1), 425–448. https://doi.org/10.1146/annurev-marine-120709-142805

Breitbart, M., & Rohwer, F. (2005). Here a virus, there a virus, everywhere the same virus? *Trends in Microbiology*, *13*(6), 278–284. https://doi.org/10.1016/j.tim.2005.04.003

Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J. M., Segall, A. M., Mead, D., Azam, F., & Rohwer, F. (2002). Genomic analysis of uncultured marine viral communities. *Proceedings of the National Academy of Sciences*, *99*(22), 14250–14255. https://doi.org/10.1073/pnas.202488399

Bremner, J. M. (1960). Determination of nitrogen in soil by the Kjeldahl method. *The Journal of Agricultural Science*, *55*(1), 11–33. https://doi.org/10.1017/S0021859600021572

Brockhurst, M. A., Morgan, A. D., Rainey, P. B., & Buckling, A. (2003). Population mixing accelerates coevolution. *Ecology Letters*, *6*(11), 975–979. https://doi.org/10.1046/j.1461-0248.2003.00531.x

Brown, J. K. M., & Rant, J. C. (2013). Fitness costs and trade-offs of disease resistance and their consequences for breeding arable crops. *Plant Pathology*, *62*(S1), 83–95. https://doi.org/10.1111/ppa.12163

Brum, J. R., Ignacio-Espinoza, J. C., Roux, S., Doulcier, G., Acinas, S. G., Alberti, A., Chaffron, S., Cruaud, C., de Vargas, C., Gasol, J. M., Gorsky, G., Gregory, A. C., Guidi, L., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant, S., Poulos, B. T., … Sullivan, M. B. (2015). Patterns and ecological drivers of ocean viral communities. *Science*, *348*(6237). https://doi.org/10.1126/science.1261498

Buttimer, C., McAuliffe, O., Ross, R. P., Hill, C., O'Mahony, J., & Coffey, A. (2017). Bacteriophages and Bacterial Plant Diseases. *Frontiers in Microbiology*, *8*. https://doi.org/10.3389/fmicb.2017.00034

Chae, J.-C., Hung, N. B., Yu, S.-M., Lee, H. K., & Lee, Y. H. (2014). Diversity of Bacteriophages Infecting Xanthomonas oryzae pv. oryzae in Paddy Fields and Its Potential to Control Bacterial Leaf Blight of Rice. *Journal of Microbiology and Biotechnology*, *24*(6), 740–747. https://doi.org/10.4014/jmb.1402.02013

Chanishvili, N. (2012). *Phage Therapy—History from Twort and d'Herelle Through Soviet Experience to Current Approaches* (pp. 3–40). https://doi.org/10.1016/B978-0-12-394438-2.00001-3

CHEN, L. (2025). *Discovery and analysis of an 841 kbp phage genome: the largest known to date*. https://doi.org/10.1101/2025.01.14.633092

Coclet, C., Sorensen, P. O., Karaoz, U., Wang, S., Brodie, E. L., Eloe-Fadrosh, E. A., & Roux, S. (2023). Virus diversity and activity is driven by snowmelt and host dynamics in a high-Elevation watershed soil ecosystem. *Microbiome*, *11*(1), 237. https://doi.org/10.1186/s40168-023-01666-z

Cooper, I. R. (2016). A review of current methods using bacteriophages in live animals, food and animal products intended for human consumption. *Journal of Microbiological Methods*, *130*, 38–47. https://doi.org/10.1016/j.mimet.2016.07.027

Czaplewski, L., Bax, R., Clokie, M., Dawson, M., Fairhead, H., Fischetti, V. A., Foster, S., Gilmore, B. F., Hancock, R. E. W., Harper, D., Henderson, I. R., Hilpert, K., Jones, B. V, Kadioglu, A., Knowles, D., Ólafsdóttir, S., Payne, D., Projan, S., Shaunak, S., … Rex, J. H. (2016). Alternatives to antibiotics—a pipeline portfolio review. *The Lancet Infectious Diseases*, *16*(2), 239–251. https://doi.org/10.1016/S1473-3099(15)00466-1

David, B. V., Chandrasehar, G., & Selvam, P. N. (2018). Pseudomonas fluorescens : A Plant-Growth-Promoting Rhizobacterium (PGPR) With Potential Role in Biocontrol of Pests of Crops. In *Crop Improvement Through Microbial Biotechnology* (pp. 221–243). Elsevier. https://doi.org/10.1016/B978-0-444-63987-5.00010-4

del Rio, B., Binetti, A. G., Martín, M. C., Fernández, M., Magadán, A. H., & Alvarez, M. A. (2007). Multiplex PCR for the detection and identification of dairy bacteriophages in milk. *Food Microbiology*, *24*(1), 75–81. https://doi.org/10.1016/j.fm.2006.03.001

Dennehy, J. J., & Abedon, S. T. (2021). Adsorption: Phage Acquisition of Bacteria. In *Bacteriophages* (pp. 93–117). Springer International Publishing. https://doi.org/10.1007/978-3-319-41986-2_2

Destoumieux-Garzón, D., Mavingui, P., Boetsch, G., Boissier, J., Darriet, F., Duboz, P., Fritsch, C., Giraudoux, P., Le Roux, F., Morand, S., Paillard, C., Pontier, D., Sueur, C., & Voituron, Y. (2018). The One Health Concept: 10 Years Old and a Long Road Ahead. *Frontiers in Veterinary Science*, *5*. https://doi.org/10.3389/fvets.2018.00014

D'Herelle, F. (2007). On an invisible microbe antagonistic toward dysenteric bacilli: brief note by Mr. F. D'Herelle, presented by Mr. Roux. 1917. *Research in Microbiology*, *158*(7), 553–554. https://doi.org/10.1016/j.resmic.2007.07.005

Duan, N., Radosevich, M., Zhuang, J., DeBruyn, J. M., Staton, M., & Schaeffer, S. M. (2022). Identification of Novel Viruses and Their Microbial Hosts from Soils with Long-Term Nitrogen Fertilization and Cover Cropping Management. *MSystems*, *7*(6). https://doi.org/10.1128/msystems.00571-22

Emerson, J. B., Roux, S., Brum, J. R., Bolduc, B., Woodcroft, B. J., Jang, H. Bin, Singleton, C. M., Solden, L. M., Naas, A. E., Boyd, J. A., Hodgkins, S. B., Wilson, R. M., Trubl, G., Li, C., Frolking, S., Pope, P. B., Wrighton, K. C., Crill, P. M., Chanton, J. P., … Sullivan, M. B. (2018). Host-linked soil viral ecology along a permafrost thaw gradient. *Nature Microbiology*, *3*(8), 870–880. https://doi.org/10.1038/s41564-018-0190-y

Fang, Z., Tan, J., Wu, S., Li, M., Xu, C., Xie, Z., & Zhu, H. (2019). PPR-Meta: A tool for identifying phages and plasmids from metagenomic fragments using deep learning. *GigaScience*, *8*(6). https://doi.org/10.1093/gigascience/giz066

Fierer, N. (2017). Embracing the unknown: disentangling the complexities of the soil microbiome. *Nature Reviews Microbiology*, *15*(10), 579–590. https://doi.org/10.1038/nrmicro.2017.87

Flaherty, J. E., Harbaugh, B. K., Jones, J. B., Somodi, G. C., & Jackson, L. E. (2001). H-mutant Bacteriophages as a Potential Biocontrol of Bacterial Blight of Geranium. *HortScience*, *36*(1), 98–100. https://doi.org/10.21273/HORTSCI.36.1.98

Flaherty, J. E., Somodi, G. C., Jones, J. B., Harbaugh, B. K., & Jackson, L. E. (2000). Control of Bacterial Spot on Tomato in the Greenhouse and Field with H-mutant Bacteriophages. *HortScience*, *35*(5), 882–884. https://doi.org/10.21273/HORTSCI.35.5.882

Florent, P., Cauchie, H.-M., Herold, M., Jacquet, S., & Ogorzaly, L. (2022). Soil pH, Calcium Content and Bacteria as Major Factors Responsible for the Distribution of the Known Fraction of the DNA Bacteriophage Populations in Soils of Luxembourg. *Microorganisms*, *10*(7). https://doi.org/10.3390/microorganisms10071458

Garg, N., Lather, S., Rani, R., & Bisht, A. (2025). The impact of drought stress on soil microbial community, enzyme activities, and microbial diversity. In *Sustainable Agriculture under Drought Stress* (pp. 255–263). Elsevier. https://doi.org/10.1016/B978-0-443-23956-4.00017-X

Gauthier, C. H., Cresawn, S. G., & Hatfull, G. F. (n.d.). *2 3 PhaMMseqs: A new pipeline for constructing phage gene phamilies using MMseqs2 4 5 6 7 8*. https://doi.org/10.1093/g3journal/jkac233/6717792

Gill, J., & Abedon, S. T. (2003). Bacteriophage Ecology and Plants. *APSnet Feature Articles*. https://doi.org/10.1094/APSnetFeature-2003-1103

Gómez, P., & Buckling, A. (2011). Bacteria-Phage Antagonistic Coevolution in Soil. *Science*, *332*(6025), 106–109. https://doi.org/10.1126/science.1198767

Greninger, A. L. (2018). A decade of RNA virus metagenomics is (not) enough. *Virus Research*, *244*, 218–229. https://doi.org/10.1016/j.virusres.2017.10.014

Guo, J., Bolduc, B., Zayed, A. A., Varsani, A., Dominguez-Huerta, G., Delmont, T. O., Pratama, A. A., Gazitúa, M. C., Vik, D., Sullivan, M. B., & Roux, S. (2021). VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome*, *9*(1). https://doi.org/10.1186/s40168-020-00990-y

Hagens, S., & Loessner, M. J. (2007). Application of bacteriophages for detection and control of foodborne pathogens. *Applied Microbiology and Biotechnology*, *76*(3), 513–519. https://doi.org/10.1007/s00253-007-1031-8

Hampton, H. G., Watson, B. N. J., & Fineran, P. C. (2020). The arms race between bacteria and their phage foes. In *Nature* (Vol. 577, Issue 7790, pp. 327–336). Nature Research. https://doi.org/10.1038/s41586-019-1894-8

Harper, D. (2011). *Viruses*. Garland Science. https://doi.org/10.1201/9780429258466

Hatfull, G. F. (2008). Bacteriophage genomics. *Current Opinion in Microbiology*, *11*(5), 447–453. https://doi.org/10.1016/j.mib.2008.09.004

Hendrix, R. W., Smith, M. C. M., Burns, R. N., Ford, M. E., & Hatfull, G. F. (1999). Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. *Proceedings of the National Academy of Sciences*, *96*(5), 2192–2197. https://doi.org/10.1073/pnas.96.5.2192

Hershey, A. D., & Chase, M. (1952). INDEPENDENT FUNCTIONS OF VIRAL PROTEIN AND NUCLEIC ACID IN GROWTH OF BACTERIOPHAGE.

Journal of General Physiology, *36*(1), 39–56. https://doi.org/10.1085/jgp.36.1.39

Hockenberry, A. J., & Wilke, C. O. (2021). BACPHLIP: predicting bacteriophage lifestyle from conserved protein domains. *PeerJ*, *9*, e11396. https://doi.org/10.7717/peerj.11396

Hoffmann-Berling, H., & Mazé, R. (1964). Release of male-specific bacteriophages from surviving host bacteria. *Virology*, *22*(3), 305–313. https://doi.org/10.1016/0042-6822(64)90021-2

Howard-Varona, C., Hargreaves, K. R., Abedon, S. T., & Sullivan, M. B. (2017). Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *The ISME Journal*, *11*(7), 1511–1520. https://doi.org/10.1038/ismej.2017.16

Hu, C., Chen, X., Wei, W., Wallace, D., Liu, J., Zhang, Y., Zhang, L., Xu, D., Batt, J., Xiao, X., Shi, Q., Zheng, Q., Ma, R., Luo, T., Jiao, N., & Zhang, R. (2025). To kill or to piggyback: Switching of viral lysis-lysogeny strategies depending on host dynamics. *Science of the Total Environment*, *959*. https://doi.org/10.1016/j.scitotenv.2024.178233

Imran, A., Shehzadi, U., Islam, F., Afzaal, M., Ali, R., Ali, Y. A., Chauhan, A., Biswas, S., Khurshid, S., Usman, I., Hussain, G., Zahra, S. M., Shah, M. A., & Rasool, A. (2023). Bacteriophages and food safety: An updated overview. *Food Science & Nutrition*, *11*(7), 3621–3630. https://doi.org/10.1002/fsn3.3360

Iriarte, F. B., Obradović, A., Wernsing, M. H., Jackson, L. E., Balogh, B., Hong, J. A., Momol, M. T., Jones, J. B., & Vallad, G. E. (2012). Soil-based systemic delivery and phyllosphere in vivo propagation of bacteriophages. *Bacteriophage*, *2*(4), e23530. https://doi.org/10.4161/bact.23530

Jacott, C., Murray, J., & Ridout, C. (2017). Trade-Offs in Arbuscular Mycorrhizal Symbiosis: Disease Resistance, Growth Responses and Perspectives for Crop Breeding. *Agronomy*, *7*(4), 75. https://doi.org/10.3390/agronomy7040075

Jaglan, A. B., Vashisth, M., Sharma, P., Verma, R., Virmani, N., Bera, B. C., Vaid, R. K., Singh, R. K., & Anand, T. (2024). Phage Mediated Biocontrol:

A Promising Green Solution for Sustainable Agriculture. *Indian Journal of Microbiology*, *64*(2), 318–327. https://doi.org/10.1007/s12088-024-01204-x

Jansson, J. K. (2023). Soil viruses: Understudied agents of soil ecology. *Environmental Microbiology*, *25*(1), 143–146. https://doi.org/10.1111/1462-2920.16258

Jansson, J. K., & Hofmockel, K. S. (2020). Soil microbiomes and climate change. *Nature Reviews Microbiology*, *18*(1), 35–46. https://doi.org/10.1038/s41579-019-0265-7

Jdeed, G., Kravchuk, B., & Tikunova, N. V. (2025a). Factors Affecting Phage–Bacteria Coevolution Dynamics. In *Viruses* (Vol. 17, Issue 2). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/v17020235

Jdeed, G., Kravchuk, B., & Tikunova, N. V. (2025b). Factors Affecting Phage–Bacteria Coevolution Dynamics. *Viruses*, *17*(2), 235. https://doi.org/10.3390/v17020235

Jones, J. B., Jackson, L. E., Balogh, B., Obradovic, A., Iriarte, F. B., & Momol, M. T. (2007). Bacteriophages for Plant Disease Control. *Annual Review of Phytopathology*, *45*(1), 245–262. https://doi.org/10.1146/annurev.phyto.45.062806.094411

Kieft, K., Zhou, Z., & Anantharaman, K. (2020). VIBRANT: Automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome*, *8*(1). https://doi.org/10.1186/s40168-020-00867-0

Kim, M.-S., & Bae, J.-W. (2018). Lysogeny is prevalent and widely distributed in the murine gut microbiota. *The ISME Journal*, *12*(4), 1127–1141. https://doi.org/10.1038/s41396-018-0061-9

Knowles, B., Silveira, C. B., Bailey, B. A., Barott, K., Cantu, V. A., Cobian-Guëmes, A. G., Coutinho, F. H., Dinsdale, E. A., Felts, B., Furby, K. A., George, E. E., Green, K. T., Gregoracci, G. B., Haas, A. F., Haggerty, J. M., Hester, E. R., Hisakawa, N., Kelly, L. W., Lim, Y. W., … Rohwer, F. (2016).

Lytic to temperate switching of viral communities. *Nature*, *531*(7595), 466–470. https://doi.org/10.1038/nature17193

Kolan, D., Cattan-Tsaushu, E., Enav, H., Freiman, Z., Malinsky-Rushansky, N., Ninio, S., & Avrani, S. (2024). Tradeoffs between phage resistance and nitrogen fixation drive the evolution of genes essential for cyanobacterial heterocyst functionality. *The ISME Journal*, *18*(1). https://doi.org/10.1093/ismejo/wrad008

Labrie, S., & Moineau, S. (2000). Multiplex PCR for detection and identification of lactococcal bacteriophages. *Applied and Environmental Microbiology*, *66*(3), 987–994. https://doi.org/10.1128/AEM.66.3.987-994.2000

Lawrence, J. G. (2002). Shared Strategies in Gene Organization among Prokaryotes and Eukaryotes. *Cell*, *110*(4), 407–413. https://doi.org/10.1016/S0092-8674(02)00900-5

Le, S. T., & Kurtböke, İ. (2019). Bacteriophages as biocontrol agents in aquaculture. *Microbiology Australia*, *40*(1), 37. https://doi.org/10.1071/MA19003

León, M., Araya, J., Nuñez, M., Arce, M., Guzmán, F., Yáñez, C., Besoain, X., & Bastías, R. (2024). Evaluation of Different Formulations on the Viability of Phages for Use in Agriculture. *Viruses*, *16*(9), 1430. https://doi.org/10.3390/v16091430

Li, D., Liu, C.-M., Luo, R., Sadakane, K., & Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics (Oxford, England)*, *31*(10), 1674–1676. https://doi.org/10.1093/bioinformatics/btv033

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*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Liao, H., Liu, C., Ai, C., Gao, T., Yang, Q.-E., Yu, Z., Gao, S., Zhou, S., & Friman, V.-P. (2023). Mesophilic and thermophilic viruses are associated

with nutrient cycling during hyperthermophilic composting. *The ISME Journal*, *17*(6), 916–930. https://doi.org/10.1038/s41396-023-01404-1

Lindström, K., & Mousavi, S. A. (2020). Effectiveness of nitrogen fixation in rhizobia. *Microbial Biotechnology*, *13*(5), 1314–1335. https://doi.org/10.1111/1751-7915.13517

Liu, C., Chen, Z., Wang, X., Deng, Y., Tao, L., Zhou, X., & Deng, J. (2025). Response of Soil Phage Communities and Prokaryote–Phage Interactions to Long-Term Drought. *Environmental Science & Technology*, *59*(6), 3054–3066. https://doi.org/10.1021/acs.est.4c08448

Łoś, M., & Węgrzyn, G. (2012). *Pseudolysogeny* (pp. 339–349). https://doi.org/10.1016/B978-0-12-394621-8.00019-4

Loureiro, T., Mesquita, M. M., Dapkevicius, M. de L. E., Serra, L., Martins, Â., Cortez, I., & Poeta, P. (2023). *Xylella fastidiosa: An Updated Review*. https://doi.org/10.20944/preprints202308.2109.v1

Luria, S. E., & Human, M. L. (1952). A NONHEREDITARY, HOST-INDUCED VARIATION OF BACTERIAL VIRUSES. *Journal of Bacteriology*, *64*(4), 557–569. https://doi.org/10.1128/jb.64.4.557-569.1952

Mäntynen, S., Laanto, E., Oksanen, H. M., Poranen, M. M., & Díaz-Muñoz, S. L. (2021). Black box of phage-bacterium interactions: Exploring alternative phage infection strategies. In *Open Biology* (Vol. 11, Issue 9). Royal Society Publishing. https://doi.org/10.1098/rsob.210188

Martín, M. C., del Rio, B., Martínez, N., Magadán, A. H., & Alvarez, M. A. (2008). Fast real-time polymerase chain reaction for quantitative detection of Lactobacillus delbrueckii bacteriophages in milk. *Food Microbiology*, *25*(8), 978–982. https://doi.org/10.1016/j.fm.2008.07.004

Merges, D., Schmidt, A., Schmitt, I., Neuschulz, E. L., Dal Grande, F., & Bálint, M. (2023). Metatranscriptomics reveals contrasting effects of elevation on the activity of bacteria and bacterial viruses in soil. *Molecular Ecology*, *32*(23), 6552–6563. https://doi.org/10.1111/mec.16756

Metze, D., Schnecker, J., Canarini, A., Fuchslueger, L., Koch, B. J., Stone, B. W., Hungate, B. A., Hausmann, B., Schmidt, H., Schaumberger, A., Bahn, M., Kaiser, C., & Richter, A. (2023). Microbial growth under drought is confined to distinct taxa and modified by potential future climate conditions. *Nature Communications*, *14*(1), 5895. https://doi.org/10.1038/s41467-023-41524-y

Mohammadi, E., Rahimian, M., & Panahi, B. (2025). Bridging the gap: Phage manufacturing processes from laboratory to agri-food industry. *Virus Research*, *353*, 199537. https://doi.org/10.1016/j.virusres.2025.199537

Mokili, J. L., Rohwer, F., & Dutilh, B. E. (2012). Metagenomics and future perspectives in virus discovery. *Current Opinion in Virology*, *2*(1), 63–77. https://doi.org/10.1016/j.coviro.2011.12.004

Montgomery, D. R., Rabinowitz, P., Sipos, Y., & Wheat, E. E. (2024). Soil health: A common focus for one health and planetary health interventions. *One Health*, *18*, 100673. https://doi.org/10.1016/j.onehlt.2023.100673

Moraru, C., Nayfach, S., Lab, B., & States Vicente Pérez-Brocal, U. (n.d.). *Phage family classification under Caudoviricetes: A review of current tools using the latest ICTV classification framework*.

Muscatt, G., Cook, R., Millard, A., Bending, G. D., & Jameson, E. (2023). Viral metagenomics reveals diverse virus-host interactions throughout the soil depth profile. *MBio*, *14*(6). https://doi.org/10.1128/mbio.02246-23

Nayfach, S., Camargo, A. P., Schulz, F., Eloe-Fadrosh, E., Roux, S., & Kyrpides, N. C. (2021). CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nature Biotechnology*, *39*(5), 578–585. https://doi.org/10.1038/s41587-020-00774-7

Paez-Espino, D., Eloe-Fadrosh, E. A., Pavlopoulos, G. A., Thomas, A. D., Huntemann, M., Mikhailova, N., Rubin, E., Ivanova, N. N., & Kyrpides, N. C. (2016). Uncovering Earth's virome. *Nature*, *536*(7617), 425–430. https://doi.org/10.1038/nature19094

Peng, X., Ru, J., Mirzaei, M. K., & Deng, L. (2022). *Replidec - Use naive Bayes classifier to identify virus lifecycle from metagenomics data*. https://doi.org/10.1101/2022.07.18.500415

PhagesDB. (n.d.). *Phage Isolation Locations by GPS Coordinates*. Https://Phagesdb.Org/GPSmap/.

Poranen, M. M., Mäntynen, S., & Ictv Report Consortium. (2017). ICTV Virus Taxonomy Profile: Cystoviridae. *The Journal of General Virology*, *98*(10), 2423–2424. https://doi.org/10.1099/jgv.0.000928

Qin, S., Xiao, W., Zhou, C., Pu, Q., Deng, X., Lan, L., Liang, H., Song, X., & Wu, M. (2022). Pseudomonas aeruginosa: pathogenesis, virulence factors, antibiotic resistance, interaction with host, technology advances and emerging therapeutics. *Signal Transduction and Targeted Therapy*, *7*(1), 199. https://doi.org/10.1038/s41392-022-01056-1

Rakonjac, J. (2012). Filamentous Bacteriophages: Biology and Applications. In *Encyclopedia of Life Sciences*. Wiley. https://doi.org/10.1002/9780470015902.a0000777

Ramírez, M., Neuman, B. W., & Ramírez, C. A. (2020). Bacteriophages as promising agents for the biological control of Moko disease (Ralstonia solanacearum) of banana. *Biological Control*, *149*, 104238. https://doi.org/10.1016/j.biocontrol.2020.104238

Regeimbal, J. M., Jacobs, A. C., Corey, B. W., Henry, M. S., Thompson, M. G., Pavlicek, R. L., Quinones, J., Hannah, R. M., Ghebremedhin, M., Crane, N. J., Zurawski, D. V, Teneza-Mora, N. C., Biswas, B., & Hall, E. R. (2016). Personalized Therapeutic Cocktail of Wild Environmental Phages Rescues Mice from Acinetobacter baumannii Wound Infections. *Antimicrobial Agents and Chemotherapy*, *60*(10), 5806–5816. https://doi.org/10.1128/AAC.02877-15

Ren, J., Song, K., Deng, C., Ahlgren, N. A., Fuhrman, J. A., Li, Y., Xie, X., Poplin, R., & Sun, F. (2020). Identifying viruses from metagenomic data using deep learning. *Quantitative Biology*, *8*(1), 64–77. https://doi.org/10.1007/s40484-019-0187-4

Rogovski, P., Cadamuro, R. D., da Silva, R., de Souza, E. B., Bonatto, C., Viancelli, A., Michelon, W., Elmahdy, E. M., Treichel, H., Rodríguez-Lázaro, D., & Fongaro, G. (2021). Uses of Bacteriophages as Bacterial Control Tools and Environmental Safety Indicators. *Frontiers in Microbiology*, *12*. https://doi.org/10.3389/fmicb.2021.793135

Rohwer, F., & Edwards, R. (2002). The Phage Proteomic Tree: a Genome-Based Taxonomy for Phage. *Journal of Bacteriology*, *184*(16), 4529–4535. https://doi.org/10.1128/JB.184.16.4529-4535.2002

Rohwer, F., & Thurber, R. V. (2009). Viruses manipulate the marine environment. *Nature*, *459*(7244), 207–212. https://doi.org/10.1038/nature08060

Rombouts, S., Volckaert, A., Venneman, S., Declercq, B., Vandenheuvel, D., Allonsius, C. N., Van Malderghem, C., Jang, H. B., Briers, Y., Noben, J. P., Klumpp, J., Van Vaerenbergh, J., Maes, M., & Lavigne, R. (2016). Characterization of Novel Bacteriophages for Biocontrol of Bacterial Blight in Leek Caused by Pseudomonas syringae pv. porri. *Frontiers in Microbiology*, *7*. https://doi.org/10.3389/fmicb.2016.00279

Roux, S., & Emerson, J. B. (2022). Diversity in the soil virosphere: to infinity and beyond? *Trends in Microbiology*, *30*(11), 1025–1035. https://doi.org/10.1016/j.tim.2022.05.003

Roux, S., Hallam, S. J., Woyke, T., & Sullivan, M. B. (2015). Viral dark matter and virus–host interactions resolved from publicly available microbial genomes. *ELife*, *4*. https://doi.org/10.7554/eLife.08490

Roux, S., Solonenko, N. E., Dang, V. T., Poulos, B. T., Schwenck, S. M., Goldsmith, D. B., Coleman, M. L., Breitbart, M., & Sullivan, M. B. (2016). Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. *PeerJ*, *4*, e2777. https://doi.org/10.7717/peerj.2777

Ru, J., Khan Mirzaei, M., Xue, J., Peng, X., & Deng, L. (2023). ViroProfiler: a containerized bioinformatics pipeline for viral metagenomic data analysis. *Gut Microbes*, *15*(1). https://doi.org/10.1080/19490976.2023.2192522

Santos-Medellín, C., Blazewicz, S. J., Pett-Ridge, J., Firestone, M. K., & Emerson, J. B. (2023). Viral but not bacterial community successional patterns reflect extreme turnover shortly after rewetting dry soils. *Nature Ecology & Evolution*, *7*(11), 1809–1822. https://doi.org/10.1038/s41559-023-02207-5

Schnabel, E. L., Fernando, W. G. D., Meyer, M. P., Jones, A. L., & Jackson, L. E. (1999). BACTERIOPHAGE OF ERWINIA AMYLOVORA AND THEIR POTENTIAL FOR BIOCONTROL. *Acta Horticulturae*, *489*, 649–654. https://doi.org/10.17660/ActaHortic.1999.489.116

Schneider, G., Szentes, N., Horváth, M., Dorn, Á., Cox, A., Nagy, G., Doffkay, Z., Maróti, G., Rákhely, G., & Kovács, T. (2018). Kinetics of Targeted Phage Rescue in a Mouse Model of Systemic *Escherichia coli* K1. *BioMed Research International*, *2018*, 1–8. https://doi.org/10.1155/2018/7569645

Shang, J., Tang, X., & Sun, Y. (2023). PhaTYP: predicting the lifestyle for bacteriophages using BERT. *Briefings in Bioinformatics*, *24*(1). https://doi.org/10.1093/bib/bbac487

Shen, W., Lin, X., Shi, W., Min, J., Gao, N., Zhang, H., Yin, R., & He, X. (2010). Higher rates of nitrogen fertilization decrease soil enzyme activities, microbial functional diversity and nitrification capacity in a Chinese polytunnel greenhouse vegetable land. *Plant and Soil*, *337*(1–2), 137–150. https://doi.org/10.1007/s11104-010-0511-2

Siles, J. A., Cajthaml, T., Minerbi, S., & Margesin, R. (2016). Effect of Elevation and season on microbial activity, abundance and community structure in Alpine forest soils. *FEMS Microbiology Ecology*, *92*(3), fiw008. https://doi.org/10.1093/femsec/fiw008

Sulakvelidze, A., Alavidze, Z., & Morris, J. G. (2001). Bacteriophage Therapy. *Antimicrobial Agents and Chemotherapy*, *45*(3), 649–659. https://doi.org/10.1128/AAC.45.3.649-659.2001

Sullivan, M. B. (2015). Viromes, Not Gene Markers, for Studying Double-Stranded DNA Virus Communities. *Journal of Virology*, *89*(5), 2459–2461. https://doi.org/10.1128/JVI.03289-14

Sullivan, M. B., Weitz, J. S., & Wilhelm, S. (2017). Viral ecology comes of age. *Environmental Microbiology Reports*, *9*(1), 33–35. https://doi.org/10.1111/1758-2229.12504

Summer, E. J., Enderle, C. J., Ahern, S. J., Gill, J. J., Torres, C. P., Appel, D. N., Black, M. C., Young, R., & Gonzalez, C. F. (2010). Genomic and Biological Analysis of Phage Xfas53 and Related Prophages of *Xylella fastidiosa*. *Journal of Bacteriology*, *192*(1), 179–190. https://doi.org/10.1128/JB.01174-09

Summers, W. C. (2001). Bacteriophage Therapy. *Annual Review of Microbiology*, *55*(1), 437–451. https://doi.org/10.1146/annurev.micro.55.1.437

Suttle, C. A. (2005). Viruses in the sea. *Nature*, *437*(7057), 356–361. https://doi.org/10.1038/nature04160

Suttle, C. A. (2007). Marine viruses — major players in the global ecosystem. *Nature Reviews Microbiology*, *5*(10), 801–812. https://doi.org/10.1038/nrmicro1750

Swanson, M. M., Fraser, G., Daniell, T. J., Torrance, L., Gregory, P. J., & Taliansky, M. (2009). Viruses in soils: morphological diversity and abundance in the rhizosphere. *Annals of Applied Biology*, *155*(1), 51–60. https://doi.org/10.1111/j.1744-7348.2009.00319.x

Tange, O. (2025). *GNU Parallel 20250222 ('Grete Tange')*. Zenodo. https://doi.org/10.5281/zenodo.14911163

Thomas, R. C. (1935). *A bacteriophage in relation to Stewards disease of Corn.*

Thompson, L. R., Zeng, Q., Kelly, L., Huang, K. H., Singer, A. U., Stubbe, J., & Chisholm, S. W. (2011). Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proceedings of the National Academy of Sciences*, *108*(39). https://doi.org/10.1073/pnas.1102164108

Tringe, S. G., & Hugenholtz, P. (2008). A renaissance for the pioneering 16S rRNA gene. *Current Opinion in Microbiology*, *11*(5), 442–446. https://doi.org/10.1016/j.mib.2008.09.011

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*(11), 607–621. https://doi.org/10.1038/s41579-020-0412-1

Trubl, G., Jang, H. Bin, Roux, S., Emerson, J. B., Solonenko, N., Vik, D. R., Solden, L., Ellenbogen, J., Runyon, A. T., Bolduc, B., Woodcroft, B. J., Saleska, S. R., Tyson, G. W., Wrighton, K. C., Sullivan, M. B., & Rich, V. I. (2018). Soil Viruses Are Underexplored Players in Ecosystem Carbon Processing. *MSystems*, *3*(5). https://doi.org/10.1128/msystems.00076-18

Turner, D., Kropinski, A. M., & Adriaenssens, E. M. (2021). A roadmap for genome-based phage taxonomy. *Viruses*, *13*(3). https://doi.org/10.3390/v13030506

Turner, D., Shkoporov, A. N., Lood, C., Millard, A. D., Dutilh, B. E., Alfenas-Zerbini, P., van Zyl, L. J., Aziz, R. K., Oksanen, H. M., Poranen, M. M., Kropinski, A. M., Barylski, J., Brister, J. R., Chanisvili, N., Edwards, R. A., Enault, F., Gillis, A., Knezevic, P., Krupovic, M., … Adriaenssens, E. M. (2023). Abolishment of morphology-based taxa and change to binomial species names: 2022 taxonomy update of the ICTV bacterial viruses subcommittee. *Archives of Virology*, *168*(2). https://doi.org/10.1007/s00705-022-05694-2

Twort, F. W. (1915). AN INVESTIGATION ON THE NATURE OF ULTRA-MICROSCOPIC VIRUSES. *The Lancet*, *186*(4814), 1241–1243. https://doi.org/10.1016/S0140-6736(01)20383-3

Våge, S., Storesund, J. E., Giske, J., & Thingstad, T. F. (2014). Optimal defense strategies in an idealized microbial food web under trade-off between competition and defense. *PLoS ONE*, *9*(7). https://doi.org/10.1371/journal.pone.0101415

Van Charante, F., Holtappels, D., Blasdel, B., & Burrowes, B. H. (2021). Isolation of bacteriophages. *Bacteriophages: Biology, Technology, Therapy*, 433–464.

Vasimuddin, Md., Misra, S., Li, H., & Aluru, S. (2019). Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. *2019 IEEE*

International Parallel and Distributed Processing Symposium (IPDPS), 314–324. https://doi.org/10.1109/IPDPS.2019.00041

Vázquez-Boland, J. A., Kuhn, M., Berche, P., Chakraborty, T., Domínguez-Bernal, G., Goebel, W., González-Zorn, B., Wehland, J., & Kreft, J. (2001). Listeria pathogenesis and molecular virulence determinants. *Clinical Microbiology Reviews*, *14*(3), 584–640. https://doi.org/10.1128/CMR.14.3.584-640.2001

Verreault, D., Gendron, L., Rousseau, G. M., Veillette, M., Massé, D., Lindsley, W. G., Moineau, S., & Duchaine, C. (2011). Detection of Airborne Lactococcal Bacteriophages in Cheese Manufacturing Plants. *Applied and Environmental Microbiology*, *77*(2), 491–497. https://doi.org/10.1128/AEM.01391-10

Voigt, E., Rall, B. C., Chatzinotas, A., Brose, U., & Rosenbaum, B. (2021). Phage strategies facilitate bacterial coexistence under environmental variability. *PeerJ*, *9*. https://doi.org/10.7717/peerj.12194

Vu, N. T., & Oh, C.-S. (2020). Bacteriophage Usage for Bacterial Disease Management and Diagnosis in Plants. *The Plant Pathology Journal*, *36*(3), 204–217. https://doi.org/10.5423/PPJ.RW.04.2020.0074

WALKLEY, A., & BLACK, I. A. (1934). AN EXAMINATION OF THE DEGTJAREFF METHOD FOR DETERMINING SOIL ORGANIC MATTER, AND A PROPOSED MODIFICATION OF THE CHROMIC ACID TITRATION METHOD. *Soil Science*, *37*(1), 29–38. https://doi.org/10.1097/00010694-193401000-00003

Wang, X., Tang, Y., Yue, X., Wang, S., Yang, K., Xu, Y., Shen, Q., Friman, V.-P., & Wei, Z. (2024). The role of rhizosphere phages in soil health. *FEMS Microbiology Ecology*, *100*(5). https://doi.org/10.1093/femsec/fiae052

Wang, Y., Liu, Y., Wu, Y., Wu, N., Liu, W., & Wang, X. (2022). Heterogeneity of soil bacterial and bacteriophage communities in three rice agroecosystems and potential impacts of bacteriophage on nutrient cycling. *Environmental Microbiome*, *17*(1), 17. https://doi.org/10.1186/s40793-022-00410-8

Webb, B., & Adeloju, S. B. (2013). Evaluation of some wet digestions methods for reliable determination of total phosphorus in Australian soils. *Microchemical Journal*, *111*, 47–52. https://doi.org/10.1016/j.microc.2013.02.001

Wei, Z., Yang, T., Friman, V.-P., Xu, Y., Shen, Q., & Jousset, A. (2015). Trophic network architecture of root-associated bacterial communities determines pathogen invasion and plant health. *Nature Communications*, *6*(1), 8413. https://doi.org/10.1038/ncomms9413

Wilhelm, S. W., & Suttle, C. A. (1999). Viruses and Nutrient Cycles in the Sea. *BioScience*, *49*(10), 781–788. https://doi.org/10.2307/1313569

Williamson, K. E., Fuhrmann, J. J., Wommack, K. E., & Radosevich, M. (2017). Viruses in Soil Ecosystems: An Unknown Quantity Within an Unexplored Territory. *Annual Review of Virology*, *4*(1), 201–219. https://doi.org/10.1146/annurev-virology-101416-041639

Wu, R., Davison, M. R., Nelson, W. C., Graham, E. B., Fansler, S. J., Farris, Y., Bell, S. L., Godinez, I., Mcdermott, J. E., Hofmockel, K. S., & Jansson, J. K. (2021). DNA Viral Diversity, Abundance, and Functional Potential Vary across Grassland Soils with a Range of Historical Moisture Regimes. *MBio*, *12*(6), e0259521. https://doi.org/10.1128/mBio.02595-21

Wu, R., Davison, M. R., Nelson, W. C., Smith, M. L., Lipton, M. S., Jansson, J. K., McClure, R. S., McDermott, J. E., & Hofmockel, K. S. (2023). Hi-C metagenome sequencing reveals soil phage–host interactions. *Nature Communications*, *14*(1). https://doi.org/10.1038/s41467-023-42967-z

Wu, R., Smith, C. A., Buchko, G. W., Blaby, I. K., Paez-Espino, D., Kyrpides, N. C., Yoshikuni, Y., McDermott, J. E., Hofmockel, K. S., Cort, J. R., & Jansson, J. K. (2022). Structural characterization of a soil viral auxiliary metabolic gene product – a functional chitosanase. *Nature Communications*, *13*(1), 5485. https://doi.org/10.1038/s41467-022-32993-8

Wu, S., Fang, Z., Tan, J., Li, M., Wang, C., Guo, Q., Xu, C., Jiang, X., & Zhu, H. (2021). DeePhage: Distinguishing virulent and temperate phage-derived

sequences in metavirome data with a deep learning approach. *GigaScience*, *10*(9). https://doi.org/10.1093/gigascience/giab056

Wu, S., Zhang, W., Wang, D., Balcazar, J. L., Wang, G., Ye, M., Chao, H., Sun, M., & Hu, F. (2025). Bacteriophage–Bacteria Interactions Promote Ecological Multifunctionality in Compost-Applied Soils. *Environmental Microbiology*, *27*(3). https://doi.org/10.1111/1462-2920.70074

Wydro, U. (2022). Soil Microbiome Study Based on DNA Extraction: A Review. *Water*, *14*(24), 3999. https://doi.org/10.3390/w14243999

Xiong, W., Song, Y., Yang, K., Gu, Y., Wei, Z., Kowalchuk, G. A., Xu, Y., Jousset, A., Shen, Q., & Geisen, S. (2020). Rhizosphere protists are key determinants of plant health. *Microbiome*, *8*(1), 27. https://doi.org/10.1186/s40168-020-00799-9

Yang, K., Wang, X., Hou, R., Lu, C., Fan, Z., Li, J., Wang, S., Xu, Y., Shen, Q., Friman, V.-P., & Wei, Z. (2023). Rhizosphere phage communities drive soil suppressiveness to bacterial wilt disease. *Microbiome*, *11*(1), 16. https://doi.org/10.1186/s40168-023-01463-8

Yu, J.-G., Lim, J.-A., Song, Y.-R., Heu, S., Kim, G. H., Koh, Y. J., & Oh, C.-S. (2016). Isolation and Characterization of Bacteriophages Against Pseudomonas syringae pv. actinidiae Causing Bacterial Canker Disease in Kiwifruit. *Journal of Microbiology and Biotechnology*, *26*(2), 385–393. https://doi.org/10.4014/jmb.1509.09012

Yuan, Y., & Gao, M. (2017). Jumbo Bacteriophages: An Overview. *Frontiers in Microbiology*, *8*. https://doi.org/10.3389/fmicb.2017.00403

Zhang, Y., Mao, M., Zhang, R., Liao, Y.-T., & Wu, V. C. H. (2024). DeepPL: A deep-learning-based tool for the prediction of bacteriophage lifecycle. *PLOS Computational Biology*, *20*(10), e1012525. https://doi.org/10.1371/journal.pcbi.1012525

APPENDIX I

All supplementary data are included in the compressed folder that is provided along with this thesis document.

Includded files:

- Bacteriophages reads per sample
- Soil Metadata
- Statistical analysis script