Exploring the metagenome of *Geosiphon pyriformis* – the only known fungus with a photosynthetic endosymbiont

Essam Sorwar

Thesis submitted to the University of Ottawa
in partial Fulfillment of the requirements for the
Master of Science degree in Biology Specialization
in Bioinformatics

Department of Biology
Faculty of Science
University of Ottawa

ABSTRACT

Arbuscular mycorrhizal fungi (AMF), belonging to the Glomeromycotina sub-phylum, are vital symbionts for roughly 80% of terrestrial plants, providing essential services like water, nutrients, and pathogen protection in exchange for photosynthetic products. These natural "biofertilizers" play a crucial role in enhancing plant biodiversity and sustaining terrestrial ecosystems. Interestingly, AMF not only engage in endosymbiosis with plants but also harbor their own unique endobacterial microbiota. A prominent example of this is *Geosiphon pyriformis*, distinct within the AMF group for its ability to host a cyanobacterium within its fungal cells - a unique symbiotic relationship in the Fungal Kingdom. This specialized interaction, occurring in bladder-like structures, involves *G. pyriformis* nuclei, mollicute-like bacterial endosymbionts (MRE), and photosynthetically active *Nostoc punctiforme* cells; a cyanobacteria. This setup makes *G. pyriformis* an ideal model for studying the complex dynamics of microbial interactions in AMF systems. Despite advancements in genomic research shedding light on *G. pyriformis*, there remains a gap in fully understanding the genomics of its endosymbionts. To bridge this gap, I have assembled and analyzed genome data from *N. punctiforme* and MRE using metagenomic datasets from *G. pyriformis* bladders. This analysis, juxtaposed with genomes of related cyanobacteria and MRE, unveiled notable gene expansions in the *Nostoc* endosymbiont, particularly in mobile genetic elements. I also found that the *G. pyriformis* MRE is part of one of two distinct monophyletic MRE groups, all of which characterized by streamlined genomes that differ significantly in structure and composition from other MRE relatives. Overall, the current work provided new insights into the evolutionary and functional complexity of these symbiotic relationships.

RÉSUMÉ

Les champignons mycorhiziens arbusculaires (AMF), appartenant au sous-embranchement des Glomeromycotina, sont des symbiotes racinaires essentiels pour environ 80% des plantes terrestres, fournissant des services essentiels tels que l'eau, les nutriments et la protection contre les pathogènes en échange des produits de la photosynthèse des plantes. Ces "biofertilisants" naturels jouent un rôle crucial dans l'amélioration de la biodiversité végétale et dans le soutien des écosystèmes terrestres. Intéressant, les AMF ne s'engagent pas seulement dans l'endosymbiose avec les plantes, mais hébergent également leur propre microbiote endobactérien unique. Un exemple remarquable de ceci est *Geosiphon pyriformis*, distinct au sein du groupe AMF pour sa capacité exclusive à héberger une cyanobactérie dans ses cellules fongiques - une relation symbiotique unique. Cette interaction spécialisée, se produisant dans des structures en forme de vessie, implique des noyaux de *G. pyriformis*, des endosymbiotes bactériens de type mollicute (MRE) et des cellules de *Nostoc* punctiforme actives en photosynthèse. Cette configuration fait de *G. pyriformis* un modèle idéal pour étudier la dynamique complexe des interactions microbiennes dans les systèmes AMF. Malgré les avancées de la recherche génomique éclairant sur la biologie de *G. pyriformis*, il reste un écart à combler pour comprendre pleinement la génomique de ses endosymbiotes. Pour combler cet écart, les chercheurs ont assemblé des données génomiques de N. punctiforme et de MRE en utilisant des ensembles de données métagénomiques provenant des vessies de *G. pyriformis*. Cette analyse, comparée aux génomes des cyanobactéries et des MRE apparentés, a révélé des expansions de gènes notables chez l'endosymbiote *Nostoc*, en particulier dans les éléments génétiques mobiles. Elle établit également que les MRE de Glomeromycotina forment un groupe monophylétique distinct, caractérisé par des génomes rationalisés qui diffèrent significativement en structure et en composition des autres parents MRE, fournissant de nouvelles perspectives sur la complexité évolutive et fonctionnelle de ces relations symbiotiques.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

First and foremost, I extend my heartfelt gratitude to my thesis supervisor, Dr. Nicolas Corradi, for welcoming me into his team as an honors student four years ago and subsequently inviting me to pursue my master's degree under his guidance. His mentorship has enriched my understanding in the field of bioinformatics and provided me with opportunities to collaborate with other highly qualified professionals. His consistent support, encouragement in my research endeavors, and appreciation for my contributions have been invaluable. Additionally, I am thankful to my committee members, Dr. Marina Cvetkovska and Dr. Julian Starr, for their insightful feedback and advice throughout my project.

I extend my gratitude to all members, past and present, of the Corradi Lab for the invaluable skills and knowledge imparted throughout my project. My academic journey has been profoundly influenced and shaped by every individual in the lab. I would like to offer special thanks to the past and present Postdoctoral Fellows, Dr. Yildirir, Dr. Oliveira, and Dr. Chandrababunaidu, for their expertise in bioinformatics. I will cherish the camaraderie that I had with other students in the lab, namely Matt, Robbie, and Ken. Thanks also to my colleagues, Dr. Alp Oran and Cindy Lesage-Pelletier, for the rewarding experience of guiding young scientists. A very special acknowledgment to Eric Ellison – our journey from being fellow students to co-teaching assistants and mentoring the next generation has been extraordinary. I am profoundly grateful for your friendship and support, and I wish you all the best in your future endeavors.

I am deeply thankful to my parents for instilling in me the courage and strength needed to navigate and surmount the challenges encountered throughout my academic path. To my friends in Alberta, I am indebted for your relentless encouragement to pursue my ambitions and for inspiring my personal growth. Additionally, my heartfelt appreciation goes out to my closest friends, Mabolaji Fawole, Jamiul and Jahidul Alam, and Richard Donghoon Kim. Your wisdom and unwavering support have been essential pillars during my master's studies, and I am profoundly grateful for your presence in my life.

LIST OF FIGURES

LIST OF TABLES

ABREVIATIONS

AMF – Arbuscular mycorrhizal fungi

Mbp – Mega base pairs

Kbp – Kilo base pairs

Bp – Base pairs

IE – Insertion/Excision

rRNA – Ribosomal RNA

MRE – Mollicutes/Mycoplasma Related Endosymbionts

KEGG – Kyoto Encyclopedia of Genes and Genomes

KO – KEGG Orthology

MGE – Mobile Genetic Element

TE – Transposable Elements

BLAST – Basic Alignment Search Tool

ROC – Root Organ Culture

MAG – Metagenome Assembled Genome

BUSCO – Benchmarking Universal Single-Copy Orthologs

MA- Mycorrhizal Association

CHAPTER ONE

Introduction

1.1 Arbuscular Mycorrhizal Fungal Symbiosis

1.1.1 Unique Symbiosis

The arbuscular mycorrhizal symbiosis (AMS) is an ancient and ubiquitous mutualism between many terrestrial plants and a group of soil fungi of the sub-phylum *Glomeromycotina* called arbuscular mycorrhizal fungi (AMF). The AMS dates back to more than 400 million years ago (Remy et al., 1994), and is believed to have played a key role in the colonization of land by plants (Paszkowski, 2006; Pirozynski & Malloch, 1975). It is also essential for the productivity and sustainability of many agronomically important plant species (Bagyaraj et al., 2015; Gianinazzi-pearson et al., 1996). During the AMS, the fungus colonizes the root cortex of the plant and forms specialized structures called arbuscules, which are intracellular haustoria that interface with the plant cytoplasm. Figure 1.1 details the mechanism by which they form arbuscules for nutrient exchange with the roots of their plant host. These arbuscules are the main site of nutrient exchange between the symbionts (Kameoka et al., 2019; W. Wang et al., 2017). The fungus produces new arbuscules as it grows, but they also senesce rapidly in a non-synchronized manner throughout the root system (Gianinazzi-pearson et al., 1996). The fungus confers multiple benefits to the plant, such as improved water use efficiency (Al-Karaki, 1998; Bi et al., 2018), enhanced uptake of phosphorous, nitrogen (Amaya-Carpio et al., 2009; Beslemes et al., 2023), and increased resistance to biotic and abiotic stresses (Begum et al., 2019; Dowarah et al., 2022), while the plant supplies the fungus with carbon (mainly in the form of lipids and carbohydrates; (Keymer et al., 2017) required growth and development. AMF are obligate biotrophs, which means that they depend on a living plant host for their carbon source (Bago & Bécard, 2002). The plant can regulate the degree of fungal colonization and the extent of symbiotic benefits, depending on its defense responses (Liao et al., 2018) and

the expression of genes involved in plant-microbe interactions (Gianinazzi-Pearson, 1996).

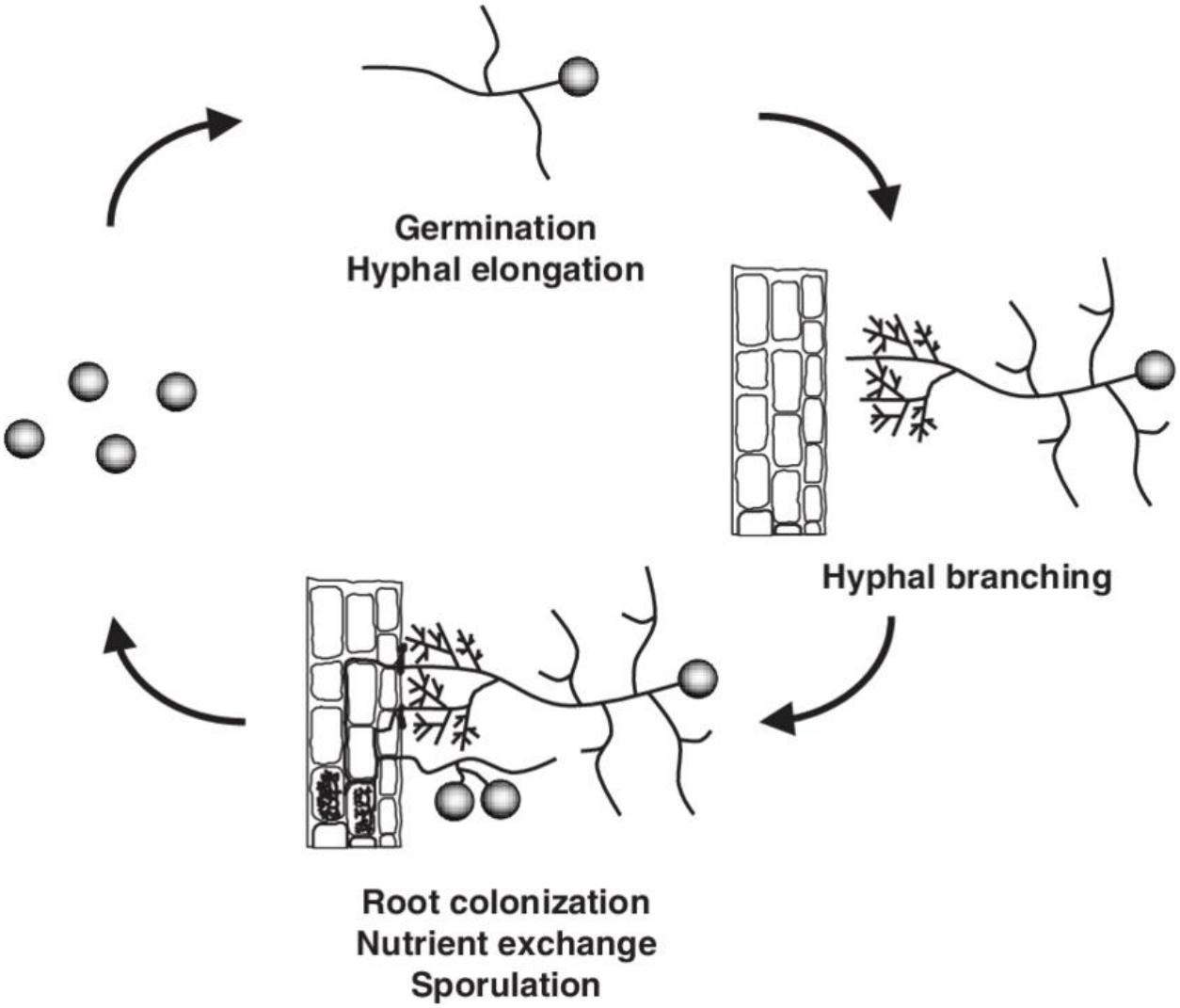

Figure 1.1 A**rbuscular-Mycorrhizal Spore Symbiosis with a plant root.** AMF produces spores in the soil, which can start growing on their own. With the assistance of chemicals coming from plant roots, they grow more hyphae. They penetrate the root and make special structures called arbuscules, where they exchange nutrients with the plant. They need the plant to survive and reproduce, so they cannot make new spores without being in a symbiotic relationship with a plant.

1.1.2 Potential applications of AMF

The rising global population and food demand will exhaust the fertilizer reserves in the

foreseeable future, necessitating the exploration of alternative methods to meet the market needs (Roy-Bolduc & Hijri, 2010). AMF are a long-established and unique mutualist of plants allowing plants to reach nutrients via hyphal networks that could not be reached by their roots alone. As such, AMF can enhance the plant's ability to utilize the existing soil nutrients applications, in turn acting as a "biofertilizer" for agricultural purposes. This offers a sustainable alternative to the harmful chemical fertilizers that are widely used worldwide and cause environmental problems such as eutrophication (Savci, 2012; Tilman, 1999) and allow the use of lower phosphate inputs, which are constantly depleted from phosphate rock reserves(Cooper et al., 2011). In fact, organic farming practices with AMF have been used in many agricultural sectors for decades (Oehl et al., 2004). However, the effectiveness of AMF inoculation is variable and often inconsistent (Guo et al., 2023) ranging from positive (Kaur et al., 2020) to neutral effects on crop yield (Kokkoris et al., 2019). Moreover, the plant responses to AMF, specifically the widely available commercial strain *R. irregularis,* inoculation depend on the genetic diversity of both the fungus and the host (Koch et al., 2006), suggesting that the fitness outcomes of this symbiosis are influenced by the interplay of different genotypes and phenotypes. Therefore, it is evident that more research is required to optimize the use of AMF as a viable solution for the agricultural challenges.

Mining, desertification, and overgrazing are all human activities that have adversely impacted the environment, as there has been observable loss of function of ecosystems and reduction of biodiversity and soil productivity. AMF are beneficial microorganisms that form symbiotic associations with most plants and have various positive effects on plant and soil health. Therefore, they are considered potential candidates for ecological restoration efforts (Asmelash et al., 2016; F. Wang, 2017). AMF can enhance the survival, growth, and

establishment of tree and shrub seedlings on degraded soils, creating favorable habitats for other organisms. They can also increase the diversity and stability of plant communities and prevent the invasion of alien species. Moreover, they can improve soil quality by promoting soil aggregation, organic matter decomposition, and nutrient cycling (Wei et al., 2019; Xu et al., 2018), which are all factors that reduce soil erosion and compaction. AMF can be applied to restore damaged ecosystems by inoculating seeds or seedlings with AMF spores, or by adding AMF-enriched substrates or composts to the soil (Jia et al., 2023; Zhou et al., 2022).

1.2 AMF Nuclear Biology and Evolution

A distinctive characteristic of AMF is the presence of hundreds of thousands of identical haploid nuclei in their cytoplasm. Moreover, there is no indication that formal sex takes place in these fungi. Consequently, the genetic structure and evolutionary mode of AMF have been contentious for years until a groundbreaking study (Ropars et al., 2016) discovered the existence of homo-dikaryotic-like phases in the model AMF species, *Rhizophagus irregularis*.

Specifically, a genome analysis of five strains of *Rhizophagus irregularis*, a model AMF, revealed that some isolates of this species harbor two genetically distinct nuclear populations, a cellular state that is indicative of sexual reproduction in multinucleate fungi. In particular, the detection of strains carrying two distinct genotypes implies that AMF might possess a sexual life cycle, which entails the hyphal fusion and nuclear exchange among compatible strains, followed by the recombination of compatible nuclei at different mating-type loci. In fact, the presence of two putative and markedly different mating-type *MAT*-loci, a genomic region that determines sexual identity in fungi, has been verified by single nucleus sequencing and sequencing of whole genomes (E. C. Chen et al., 2019; Sperschneider et al., 2023).

In summary, the genetic organization of AMF conforms to that of other multinucleate fungi. In particular, AMF isolates can be either homokaryotic with identical nuclei coexisting in the same cytoplasm (AMF homokaryons), or dikaryotic with two distinct nuclear populations partitioned among thousands of nuclei in a shared cytoplasm (Sperschneider et al., 2023).

1.3 AMF Genomics

To conduct genome analyses of Arbuscular Mycorrhizal Fungi (AMF), substantial amounts of DNA free of contaminants are required, which are usually obtained by employing Root Organ Culture (ROC) systems. However, recent investigations have revealed that homokaryotic strains constitute more than 95% of ROC, from the *Rhizophagus* genus. Consequently, the current knowledge of AMF genomics relies on AMF homokaryons of the *Rhizophagus* genus.

R. irregularis DAOM-197198 (DOAM stands for the Department of Agriculture, Ottawa, Mycology), whose descendants are established in Quebec, Canada, was the first arbuscular mycorrhizal fungus (AMF) whose genome was sequenced. This sequencing effort has made it a pivotal research model in the field of AMF molecular biology. Its genome analysis uncovered features that are typical of most *Glomeromycotina* genomes. For example, the size of the DAOM-197198 genome assembly was consistent with previous flow cytometry measurements (Bianciotto & Bonfante, 1992; Sędzielewska et al., 2011) and showed that most *Glomeromycotina* genomes are considerably larger than those of their fungal relatives.

The genome size of AMF varies widely, from 50 Mb in some early diverging taxa to a maximum of 784 Mb in *Gigaspora margarita*. The genome size is positively correlated with the DNA repeats content of the symbiont. As a case in point, transposable elements (TEs) and other DNA repeats may account for 86% of the large genome of *G. margarita*. The gene content also

correlates with the genome size, ranging from about 10,000 genes in *Paraglomus occultum* to an average of 24,000 genes in more derived taxa with bigger genomes.

Moreover, genome analyses have revealed the reason for the high host dependency of these organisms. Specifically, all *Glomeromycotina* with sequenced genomes lack some essential enzymes for fatty acid biosynthesis and organic matter degradation. This suggests that AMF have lost these functions in their evolutionary history and why they cannot grow in axenic culture (Dallaire & Paszkowski, 2023). Genomic studies have played a crucial role in revealing common features in AMF, especially the expansion of gene families related to symbiosis and stress response, which indicate their adaptations to various environments and plant hosts (Dallaire & Paszkowski, 2023; Jing et al., 2022). For instance, AMF has enlarged gene families for ABC transporters and carbohydrate-active enzymes, which are presumably involved in nutrient uptake and plant cell wall modification, respectively (E. C. H. Chen et al., 2018). Moreover, these studies have detected the expansion of gene families for heat shock proteins and glutathione S-transferases, which are involved in protein folding and oxidative stress response, respectively (Hildebrandt et al., 2007). These genomic features and functions reflect the ecological and evolutionary success of AMF as obligate plant symbionts.

1.4 AMF Bacterial Microbiota

1.4.1 Mollicutes/Mycoplasma-related Bacteria

AMF participate in complex microbial networks, where they not only establish symbiosis with plants, but also harbor endobacteria, such as Mollicutes/Mycoplasma-related (MRE) and Burkholderia- related endobacteria (BRE) (Bonfante & Desirò, 2017). MRE are widespread in many AMF families, while BRE has only been detected in the AMF family Gigasporaceae

(Bonfante & Desirò, 2017). Bacterial endosymbionts of fungi are usually obligate dependents of their host for survival, which can influence fungal physiology and fitness. The roles and interactions (i.e., mutualistic, or parasitic) of these endobacteria remain unclear, as no attempts to eliminate MRE from AMF have succeeded. AMF bacterial endosymbionts have experienced reductive evolution, resulting in the loss of many core genes and a reduced metabolic capacity due to host adaptation (Uehling et al., 2023). A study in 2015 showed that MRE cannot synthesize purine and pyrimidine de novo and depend on their fungal host for their nucleotides (Naito et al., 2015). In other fungal subphyla under the *Mucoromycota* phylum, BRE and MRE endobacteria are also present in many species of *Mucoromycotina* and *Mortierellomycotina*. They also affect the ecological functions of their hosts as seen in *Rhizophus microsporus* of the *Mucoromycota* subphylum, where the endobacteria produce toxins, which enhance metabolic capabilities and the pathogenic nature of their hosts (Partida-Martinez & Hertweck, 2005). The endobacteria in the *Mortierellomycotina* also resemble the AMF endobacteria, in that the MREs in *Lobosporangium* impact their host's functioning in selective pressures (Desirò et al., 2018).

1.4.2 Geosiphon pyriformis

Geosiphon pyriformis is a unique fungus that forms an endocytobiotic association with a cyanobacterium, specifically *Nostoc punctiforme* (A. Schüßler & Kluge, 2001). This symbiotic relationship, known as endocyanosis, is characterized by a "siphonal bladder" made of a swollen fungal hyphal that now contains the photosynthetically active *Nostoc* filaments and heterocysts, as well *as G. pyriformis* nuclei and MRE. This raises the question of whether *G. pyriformis* is truly an AMF, as it does not establish a symbiosis with plants.

The *Geosiphon-Nostoc* symbiosis is the only known fungal endosymbiosis with cyanobacteria, and despite its uniqueness, it has now been demonstrated that this species is a

member of the *Glomeromycotina* (Figure 1.2) (Malar et al., 2021). Specifically, genome analyses of this species have shown that this taxon is a member of the AMF order *Archesporales*, and remarkably, genome investigations have revealed that this species is homokaryotic and carries all *Glomeromycotina* hallmarks, including the absence of genes to produce fatty acids and a reduced sets of plant cell wall degrading enzymes.

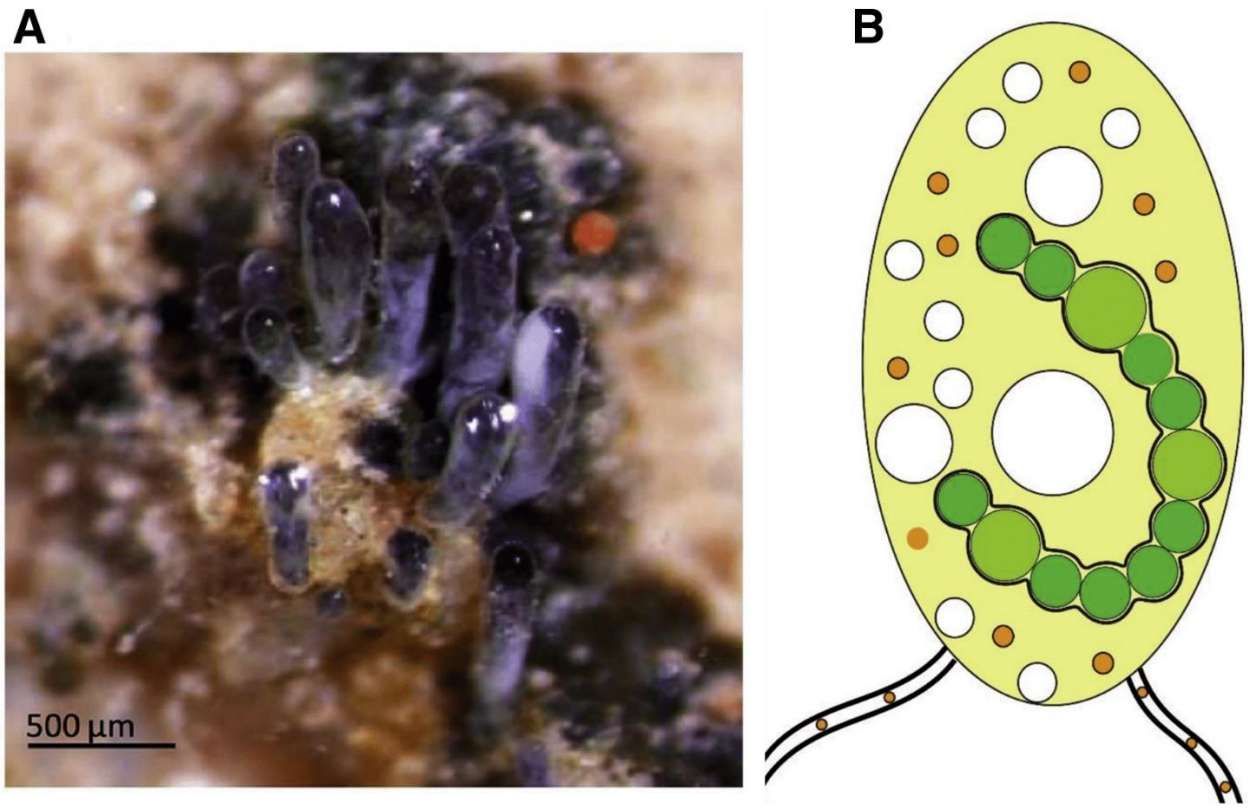

Figure 1.2 The shape and structure of the round cells made by *Geosiphon pyriformis* and the cyanobacteria living inside them. A) A picture of the soil where *G. pyriformis* lives in the wild, located in the Spessart mountains located of Germany. The soil has small, round structures called bladders, which are made by *G. pyriformis*. **B)** A drawing of what the bladders look like inside, based on Kluge, (2002). The bladders have tiny green cells called *Nostoc*, which can fix nitrogen from the air. Some of the *Nostoc* cells are lighter green and have a special shape. These are called heterocysts. The bladders also have orange dots, which are the nuclei of G. pyriformis. The bladders have empty spaces called vacuoles. Thin, thread-like structures called hyphae come out of the bladders and connect them to each other. (Malar et al., 2021)

In addition to providing insights into the evolution of *Glomeromycotina, G. pyriformis* offers a chance to reveal the functions and interactions of between these prominent symbionts

and their endobacteria. Specifically, the data used to acquire the *G. pyriformis* genome also contains sequence data from its endobacteria, and these can be extracted and analyzed using appropriate metagenomics approaches.

1.4.3 Nostoc: Cyanobacterium involved in Symbiotic Networks

Nostoc encompasses a genus of cyanobacteria renowned for their versatility and ability to form symbiotic relationships with a wide array of hosts, including fungi, plants, and lichens (Sand-Jensen, 2014). Therefore, these cyanobacteria have been research topics, where studies are conducted in further understanding their ability to form symbiotic networks within microbial communities (Gagunashvili & Andrésson, 2018). For centuries, the *Nostoc* genus has been utilized as nutritious food source and in traditional medicine across Asia. They can provide immense nutritional value as they have high content in proteins, vitamins and carbohydrates, making them a staple in the diets of people in China, India and Indonesia (Gao, 1998; Johnson et al., 2008; Li & Guo, 2018; Roney et al., 2009). *Nostoc* species are more widely known for their ability to fix atmospheric nitrogen (J. S. Singh et al., 2016), contributing significantly to the nutrient dynamics of ecosystem they reside in.

Just like AMF, *Nostoc* is also heavily researched as a key to resolving the climate crisis as it plays a role in carbon fixation. They can sequester carbon, thus reducing atmospheric CO2 levels. Therefore, in-depth research into this genus could offer insights into natural strategies for combating climate change (Llamas et al., 2023). *Nostoc* is recognized for its other various biotechnological applications, especially due to its capability to synthesize compounds that hold significant value in the pharmaceutical industry. These include peptides and peptide-like structures, including cytotoxic cryptophycins, antiviral cyanovirin-N, and the antitoxic nostocyclopeptides (Fidor et al., 2020, 2019; Niveshika et al., 2016). These compounds,

alongside others such as hepatotoxic microcystins and potent protease inhibitors like cyanopeptolins, anabaenopeptins, and microginins (Fidor et al., 2019), exhibit a range of biological activities, including antifungal, antibacterial, and antitumor properties (Fidor et al., 2019; Nowruzi et al., 2018). They have also been a topic of interest when regarding biofertilizers, due to their chemical composition and photosynthetic efficiency (Zahra et al., 2020), which allows them to produce fatty acids which can potentially be utilized for biofuel production (Nosheen et al., 2021).

In the context of AMF symbiosis, currently the only known complete symbiosis with cyanobacteria is *Geosiphon pyriformis*, where *Nostoc* plays a pivotal role in its host survival (Malar et al., 2021). *G. pyriformis*, a distinct AMF species, hosts cyanobacteria specific strain of *Nostoc punctiforme* (SAG 69.79) within a bladder like symbiosome. Although previous studies have begun to shed light on the complex dynamics of microbial symbioses and the evolutionary pathways of symbiotic organisms, many aspects remain underexplored. Metabolically, the *N. punctiforme* acts as a photosymbiont; it engages in CO2 fixation both in light and darkness, with different products being synthesized under these conditions (Kluge, 2002a). This photosynthetic activity supplies organic carbon to the fungus such as phosphate esters, polyglucans, free sugars, some amino acids, and organic acids. Conversely, *Nostoc* benefits from a presumably improved access to nutrients facilitated by the host.

1.5 Read Binning in Metagenomics

Metagenomics is a technique that enables the study of microbial communities by directly recovering and sequencing genetic material from environmental samples (Zhang et al., 2021). This technique has several advantages over traditional culture-based methods, such as the ability

to capture the genetic diversity and function of microbial communities without the need to isolate individual species, and the potential to identify novel genes and pathways that reveal the ecological and evolutionary relationships of microbes and their roles in the community they inhabit (Hilton et al., 2016). Metagenomics is especially useful in bioinformatics, where it can overcome the challenges of isolating, culturing, and sequencing microorganisms that are difficult or impossible to grow in the laboratory. Early studies have demonstrated that metagenomics can provide a more comprehensive and accurate picture of the microbial diversity present in various environments than conventional culturing methods (Stein et al., 1996).

One of the first approaches to move away from culture-based methods was 16S ribosomal RNA (rRNA) gene sequencing, which involves amplifying and sequencing specific regions of the 16S rRNA gene, a universal marker for bacterial identification (Eckburg et al., 2005; Ley et al., 2006; Woese, 1987). However, 16S rRNA gene sequencing has some limitations, such as the inability to provide information on the metabolic capabilities and the genetic variations among different strains of the same species. To address this issue and to obtain a more detailed view of the species composition and function of a microbial community, it is necessary to extract groups of metagenomic contigs and assemble them into metagenome-assembled genomes (MAGs). Current technology allows for the preparation of Illumina libraries, where metagenomic DNA is fragmented, attached to adapters, and amplified by Illumina sequencing platforms (Hu et al., 2021).

In the era of genome-resolved metagenomics, a new challenge emerges on how to extract the genomes of interest efficiently and accurately from the massive amount of sequenced data. This requires a process of organizing the metagenomic reads or contigs based on patterns or features found in the sequences. One such process is binning (Kunin Victor et al., 2008), where

sequences from a next-generation sequencer or assembled contigs are assigned to a specific organism or taxonomic group. Binning can be based on similarity, where algorithms use organism-specific characteristics of the DNA to separate sequences. One such character is k-mer frequency, which refers to the occurrence rate of sequence of length 'k' in the DNA, providing a signature that can indicate genomic similarity. The other characteristic used to separate sequences is depth coverage. An example of a similarity-based binning algorithm is CONCOCT (Alneberg et al., 2014), which clusters contigs by using a Gaussian mixture model (a statistical machine learning model that assumes there are normally distributed subpopulations within an overall population) and a variational Bayesian inference framework (Figure 1.3).

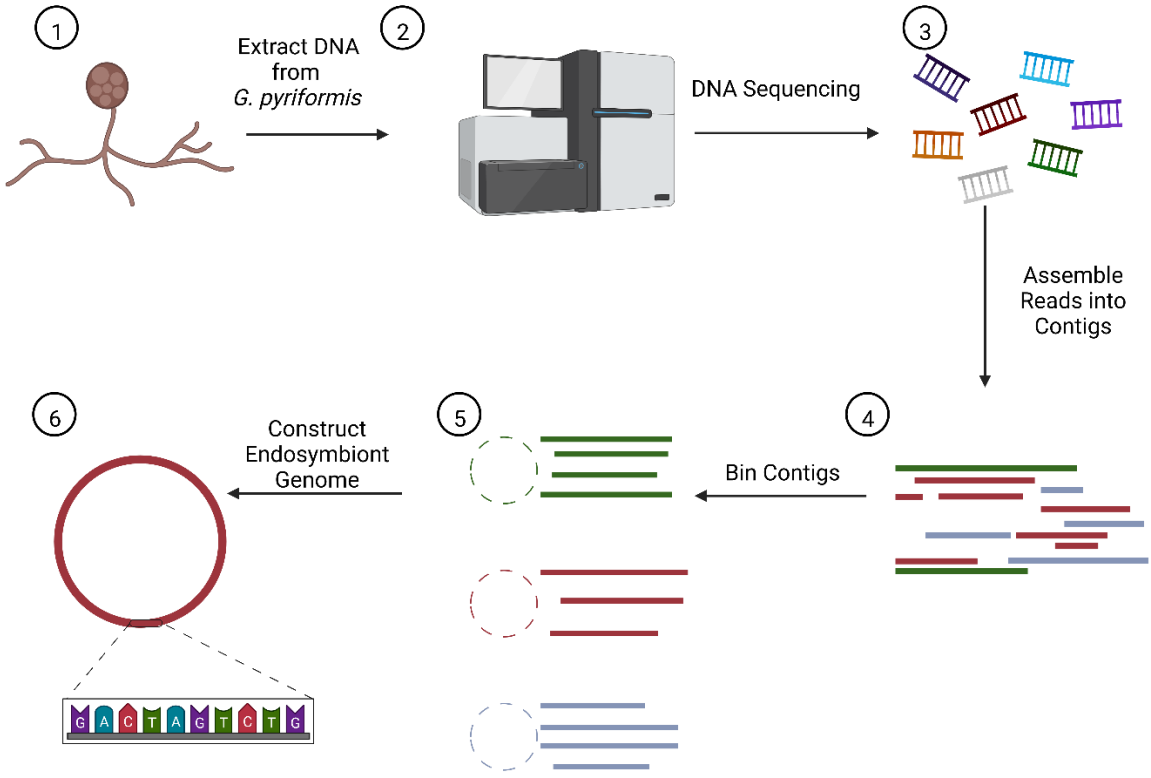

Figure 1.3 Metagenomic workflow to construct the endobacterial genomes present in the *G. pyriformis* bladder. Figure created with BioRender.

1.6 Research Justification and Goals

Geosiphon pyriformis is a unique member of the Glomeromycotina sub-phylum, which are all obligate symbionts that depend on their hosts for survival. Unlike most Glomeromycotina, which form mutualistic associations with plant roots and exchange fatty acids, *G. pyriformis* does not associate with plants, but instead forms a symbiotic relationship with endobacteria that reside within its bladder-like structures. This remarkable feature makes *G. pyriformis* an ideal model organism to investigate the ecological and evolutionary implications of endobacterial symbiosis in AMF. However, one of the major challenges in studying this system is the difficulty of isolating and culturing the endobacteria from the fungal host in the laboratory. Therefore, to address this issue, we employ metagenomics as a powerful tool to explore the genomic diversity and composition of the endobacterial community in *G. pyriformis*. The main objective of this study is to characterize the endobacterial genomes in *G. pyriformis* and compare them with those of other AMF-related taxa, using the available metagenomic data from the fungal bladders. To achieve this goal, we analyze the sequence data of the bladder samples with a metagenomic binning pipeline that assigns sequence reads to different genomic bins based on their k-mer frequency. By applying a comparative genomic approach, we aim to identify common and unique genomic features in the *Nostoc* and putative MRE species that co-exist within the *G. pyriformis* bladders, and to elucidate their functional roles and evolutionary histories in relation to their AMF hosts.

CHAPTER TWO

Assembly and comparative analyses of the *Geosiphon pyriformis* metagenome

2.1 Abstract

Geosiphon pyriformis is a representative of the fungal sub-phylum Glomeromycotina, and the only known example of an endosymbiosis of cyanobacteria within a fungal cell. This symbiosis takes place in bladders that contain *G. pyriformis* nuclei, mollicute-like bacterial endosymbionts (MRE), as well as photosynthetically active and dividing *Nostoc punctiforme* cells. While recent genome analyses provided key insights into the biology of *G. pyriformis*, information on the genome content and biology of its endosymbionts is currently unavailable. To address this, we collected and analyzed metagenomic data from the bladders of *G. pyriformis* where *N. punctiforme* and MRE resides, to ensure analyses focused on the organs directly involved in the symbiosis. By comparing this data with the genetic information of related cyanobacteria and MREs from other AMF species, we aimed to uncover the genetic content of these organisms and how these interact at a genetic level and establish a symbiotic relationship. Our analyses reveal significant gene expansions affecting the *Nostoc* endosymbiont, particularly in mobile elements, but also in genes potentially involved in the degradation of xenobiotics. We also confirm that the *Glomeromycotina* MRE are monophyletic and carry a highly streamlined genome. These genomes also differ dramatically in both structure and content, including in the presence of enzymes involved in environmental sensing and stress response.

2.2 Keywords

Arbuscular mycorrhizal fungi, Metagenomics, *Nostoc punctiforme*, Mollicutes-Related Endobacteria, Genome Assembly, Endosymbionts, Mobile Genetic Elements

2.3 Introduction

Arbuscular mycorrhizal fungi (AMF, sub-phylum Glomeromycotina (Spatafora et al., 2016)) are key components of the underground fungal network as they form a mutualistic and widespread relationship with terrestrial plants called the mycorrhizal association (MA) symbiosis (F. M. Martin et al., 2017). In the MA symbiosis, the hyphae of the fungal partner colonize the root cortex of the plant to produce tree-like structures called "arbuscules" (Corradi & Bonfante, 2012). In these, nutrients are exchanged between the partners; on one hand, the plant provides photosynthetically fixed carbon to the plant, on the other hand, the fungus improves the transfer of soil nutrients (i.e., phosphorus, nitrogen, potassium) and water uptake for the plant (Smith & Read, 2010). Generally, this symbiosis improves plant fitness and ecosystem productivity (der Heijden et al., 1998), and for these reasons, AMF are considered keystone mutualists in terrestrial ecosystems (Smith & Read, 2010).

Observations from fossil records and genome analyses of extant plants show that these fungi are highly successful symbionts that colonized and coevolved with the roots of up to 80% of terrestrial plant species (Redecker et al., 2000; Rich et al., 2021). As such, AMF might have impacted the evolution of land plants and the biodiversity of ecosystems. These fungi also carry peculiar cellular and evolutionary features, which include the presence of thousands of nuclei with one or two genotypes within a large syncytium (Kokkoris et al., 2020) and a lack of observable sexual reproduction despite the presence of mating-related gene, evidence of recombination, and life stages like those of distant sexual fungi (Halary et al., 2011; Riley & Corradi, 2013; Ropars et al., 2016; Sperschneider et al., 2023).

AMF can also harbour Mollicutes/Mycoplasma-related (MRE) and *Burkholderia*-related endobacteria (BRE). MRE are abundant in many Glomeromycotina, whereas BRE has only been

seen in the AMF family Gigasporaceae (Bonfante & Desiro, 2017; Naumann et al., 2010). MRE seem to be obligately dependent on their host for survival, and can affect fungal physiology and fitness (Salvioli et al., 2016), as well as the gene transcription of both mycorrhizal partners (Venice et al., 2021). The functions and lifestyle (i.e., mutualistic, or parasitic) of MRE are still unclear, and like most intracellular bacteria these endosymbionts have undergone reductive evolution leading to a limited metabolic capacity (Amses et al., 2023; McCutcheon, 2021). For example, MRE lack genes to synthesize purine and pyrimidine *de novo*, and may thus rely on their fungal hosts for their nucleotides (Naito et al., 2015).

BRE and MRE endobacteria can also be present in closely related fungi of the phylum Mucoromycota (Longley et al., 2023) and, as seen in AMF, these can impact the ecological functions of their hosts. For example, in *Rhizopus microsporus* (sub-phylum Mucoromycotina) the endobacteria can produce toxins, increasing metabolic capabilities and the pathogenic nature of the fungal host (Partida-Martinez & Hertweck, 2005), or can influence the sexuality of specific taxa (Mondo et al., 2017). The endobacteria present in the sub-phylum Mortierellomycotina also show similarity with the AMF MRE in that they can affect their hosts functioning (Desirò et al., 2018).

Arguably, the mutual *Geosiphon-Nostoc* association is one of the most intriguing terrestrial endosymbiosis, as it presents the only known example of a cyanobacterium growing inside fungal cells (Kluge, 2002b). Unlike other Glomeromycotina, *G. pyriformis* is not known to produce mycorrhizae (Malar C et al., 2021a; Arthur Schüßler & Wolf, 2005), and relies instead on the endosymbiont *Nostoc punctiforme* (strain SAG 69.79) to obtain essential metabolites such as fatty acids. Within bladders, several nuclei from *G. pyriformis* co-exist with

photosynthetically active, nitrogen-fixing *N. punctiforme* and undefined MRE (Gehrig et al., 1996; Kluge, 2002b).

While genome analyses of *G. pyriformis* revealed unique biological features of this species, including expansions in gene families involved in processing lipids and genes horizontally transferred from bacteria (Malar C et al., 2021a), basic information on the endosymbionts of this fungus is still lacking. In particular, the genome content, diversity, and putative function of the symbiotic *N. punctiforme* (SAG 69.79) and MRE endosymbionts are unknown. To begin addressing these questions, we identified and paired-end/mate Illumina reads belonging to these endosymbionts from available *G. pyriformis* metagenome datasets using binning techniques (Malar C et al., 2021a). We then assembled these reads and genome assemblies were investigated to determine their uniqueness and similarities to closely related organisms' relatives through comparative genomics.

2.4 Materials and Methods

2.4.1 Endosymbiont genome assembly and annotation

Extracted DNA of symbiotic bladders of the *Geosiphon-Nostoc* association (Malar C et al., 2021b), from the Spessart Mountains in Germany, was implemented into a read binning approach. This DNA was subjected to a 125 bp paired-end and 5 kb mate-pair Illumina Hiseq 4000 sequencing. Read duplication/overrepresentation and contaminants were removed using FastQC (Andrews, 2023) and adapter removal using Cutadapt (M. Martin, 2011). The metagenome reads were then assembled into contigs using SPAdes V3.12.0 (Prjibelski et al., 2020).

The assembled contigs were binned with CONCOCT (Alneberg et al., 2014), which clusters the contigs based on their sequence composition and coverage. The clusters were annotated using BLAST 2.6.0+ (Altschul et al., 1990; Buchfink et al., 2021) and the two bins that primarily aligned with *Nostoc* or MRE were kept for downstream analysis. The paired-end and mate-pairs Illumina reads that mapped onto contigs obtained with SPAdes V3.12.0 (Prjibelski et al., 2020) were scaffolded using MaSuRCA 3.3.0 (Zimin et al., 2013). All genome alignments were conducted using MUMmer3 (*marbl/MUMmer3*, 2023) using default parameters, and incorrectly assembled segments were removed using Seqkit (Shen et al., 2016).

Bacterial protein-coding genes were predicted for all the genomes using Bakta v1.8.2 (Schwengers et al., 2021). This tool was applied to all AMF MRE genomes and the other closely related *Nostoc* species analysed in this study to maintain consistency in the quality of annotation. Proteins from the newly assembled *N. punctiforme* (SAG 69.79) endosymbiont genome were clustered into orthologous groups alongside to four *Nostoc* species using FastOrtho (*GitHub - olsonanl/FastOrtho*, n.d.). The other *Nostoc* species analyzed include the reference *Nostoc punctiforme* ATCC 29133/PCC 73102 (GCF_000020025.1), the free-living *Nostoc sphaeroides* Kutzing En (PRJNA488142), and *Nostoc edaphicum* CCNP1411 (GCA_014023275) (Huo et al., 2021) and *Nostoc* sp. 'Peltigera membranacea cyanobiont' N6 (GCA_002949735), a lichen symbiont. These species were chosen as representing lifestyles that significantly differ from SAG 69.79 growing in *G. pyriformis* bladders. The same procedure was conducted on all available MRE genomes.

In all cases, the putative gene functions were investigated using the COG (Galperin et al., 2021) (Clusters of Orthologous Groups) database and with DIAMOND (Buchfink et al., 2021). In parallel, the metabolic functionality of these orthologues was examined with the Kyoto

Encyclopedia of Genes and Genomes (Kanehisa et al., 2016) (KEGG). Mobile genetic elements (MGE) abundance and expansions among *Nostoc* species analysed in this study were investigated using the mobileOG-db database (Brown et al., 2022; *GitHub - clb21565/mobileOG-db: code repo for mobileOG-db*, n.d.) and Proksee (Grant et al., 2023), with 80% for minimum sequence identity threshold. The annotation of putative secreted proteins was performed using SignalP-6.0 (Petersen et al., 2011), with the mode parameter set to *slow_sequential* and the organism set to *other options*. To enhance the accuracy of our findings, proteins concurrently identified as transmembrane proteins by the DeepTMHMM (Hallgren et al., 2022) online tool were excluded from the final list of putative secreted proteins.

2.4.2 Phylogenetic analysis

MRE phylogenomic analysis was constructed by a set of 14 single-copy proteins from all nine *Mollicutes* species, including *Mycoplasmoides genitalium* as the outgroup species. To construct the *N. punctiforme* tree, partial 16S rRNA gene sequences were extracted from the newly assembled genome with barrnap (Seemann, 2023). The phylogenetic tree was reconstructed using the maximum-likelihood approach with the best-fit substitution model option for individual partitions defined by each gene marker implemented in IQ-TREE v.1.6 (Nguyen et al., 2015).

2.4.3 Inspecting cases of Horizontal Gene transfer

Evidence of horizontal gene transfers (HGT) from the Glomeromycotina fungal host to its bacterial endosymbionts was investigated using previously established procedures (Sun et al., 2019). Specifically, the predicted bacterial proteins from the two *G. pyriformis* endosymbionts were compared against the NCBI nr protein database with DIAMOND using the BLASTp (query

and subject coverage to greater than or equal to 50% and e-values below $1e^{-5}$ (Sun et al., 2019)). Proteins that did not exhibit at least twice as many fungal hits as bacterial taxa were filtered out, while the others were subjected to phylogenetic analysis to confirm their foreign origin (HGT candidates). The location of HGT candidates within the genome was also manually inspected with Integrative Genomics Viewer (IGV) (Robinson et al., 2023) to ensure that these were surrounded by bacterial genes with high coverage, and did not originate from chimeric assemblies errors/contaminants.

2.5 Results

2.5.1 Endosymbiont genome assembly and annotation

Publicly available 125 bp paired-end and 5 kb mate-pair Illumina reads, derived from *G. pyriformis* bladders (Malar C et al., 2021a), were used to identify *Nostoc* and MRE genome data using binning procedures successfully implemented to obtain quality genome data from multiple Glomeromycotina (Malar C et al., 2021a, 2022; Sun et al., 2019) (Figure 2.1). These procedures resulted in the identification of 148 clusters, based on the read sequence composition and coverage. Clusters containing reads that only aligned with *Nostoc* species or Glomeromycotina MRE were kept, inspected manually, and used for genome assembly with MaSuRCA (Zimin et al., 2013).

Following assembly, a total of 18 genome scaffolds showed high sequence similarity and aligned well against the *N. punctiforme* ATCC 29133/PCC 73102 reference sequence (GCF_000020025.1). Substantial differences in structure between that genome is also found (Figure S1). These contigs were deemed to represent *bona-fide* genome information from the *Geosiphon* cyanobiont (referred to hereafter as *Np*-SAG 69.79). The total genome size of *Np*-SAG 69.79 is 9.9 Mb, with a GC content of 40.76% (Table 1). This compares with a total size of

9.1 Mb for the *N. punctiforme* (*Np*-PCC 73102) reference sequence (GCF_000020025.1), and with sizes of 8.9 Mb, 8.3 Mb and 6.7 Mb and average GC content of 41.5% in closely related *Nostoc* species with distinct lifestyles, including the free-living *N. sphaeroides* (*Ns*-Kutzing En) and *N. edaphicum* (*Ne*-CCNP1411) and the lichen symbiont *Nostoc* sp. '*Peltigera membranacea* cyanobiont' N6 (*Nsp.* PMC N6).

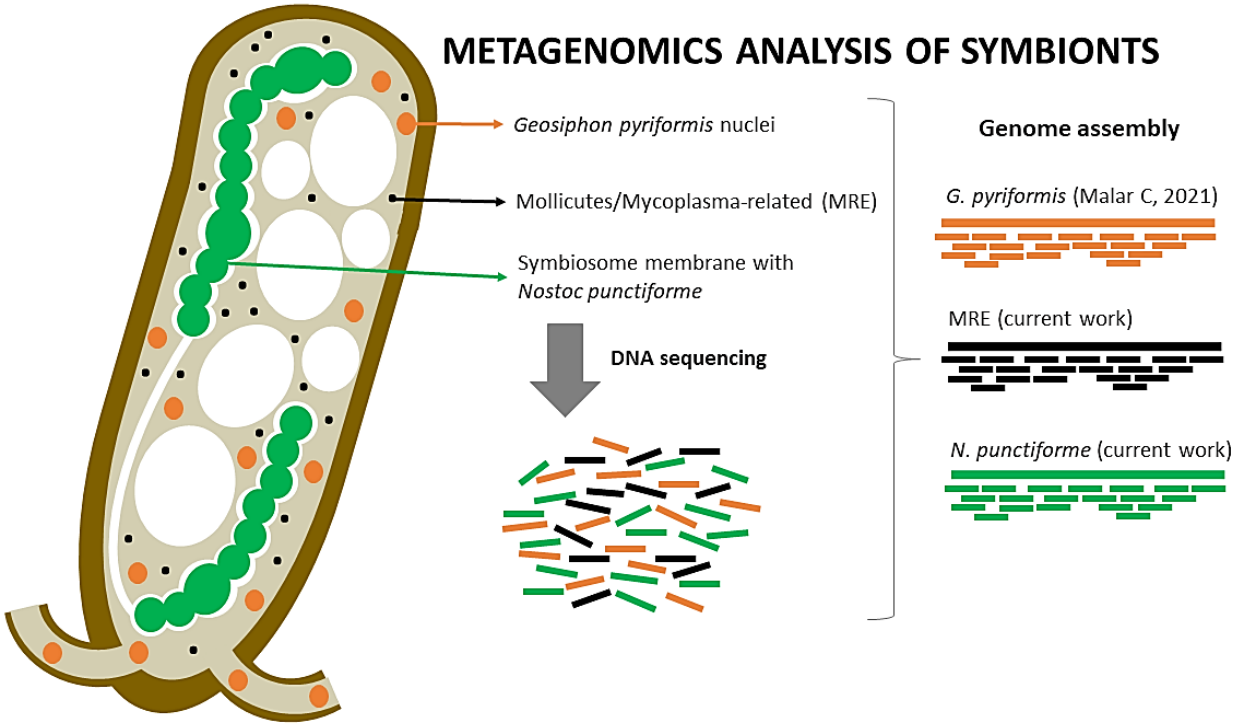

Figure 2.1. Metagenomics analysis of endosymbionts present in the bladder of *Geosiphon-Nostoc* symbiosis. This diagram illustrates the sequences are binned based on their K-mer frequencies and coverage to construct the genomic profiles of endobacteria located within the bladder of *Geosiphon-Nostoc* symbiosis.

In the case of MRE, we were unable to identify well aligned genome regions among any publicly available MRE genomes, as well as between our putative scaffolds from *Geosiphon* MRE and publicly available MRE genomes (Naito et al., 2015; Sun et al., 2019; Torres-Cortés et al., 2015). The inability to fully align these genomes underpins the presence of rapid changes in the

sequence and structure of MRE genomes of Glomeromycotina (Figure S2). To identify *bona-fide G. pyriformis* MRE scaffolds (defined hereafter as *Gp*-MRE) in the absence of high sequence/genome similarities, putative MRE scaffolds were identified based similarity in GC content. gene functions and gene sequence homology with other Glomeromycota MRE (Table 2). Altogether, this procedure led to the identification of 3 putative *Gp*-MRE scaffolds, totaling a genome size of 712,504 bp (Table 2), with an average GC content of 32.63%. All scaffolds show higher sequence similarity to Glomeromycotina MRE than to any other organism. No evidence of multiple distinct MRE, as seen for example in *Diversispora epigaea* (*De*MRE-I and II (Sun et al., 2019)), was found in *G. pyriformis.*

Table 1 Genome statistics of *Np*-SAG 69.79 and other closely related *Nostoc* species.

	***Np*-SAG 69.79**	***Np*-PCC 73102**	***Ne*-CCNP1411**	***Ns*-Kutzing En**	***Nsp.* PMC N6**
Genome size (bp)	9 904 587	9 059 191	8 316 316	6 684 490	8 898 590
% BUSCO Genome Compl.	98.4	99.4	99.4	98.3	99.3
No. scaffolds	18	6	6	5	11
%GC	41.5	41.4	41.6	41.6	41.4
No. rRNAs	8	12	9	12	12
No. tRNAs	105	90	76	74	133
No. of CDS	8689	7733	7054	5955	7574
% Coding density	84.2	81.9	81.4	77.7	82.4
No. pseudogenes	183	64	70	47	31
Accession no.	This Study	GCA_000020025	GCA_014023275	GCA_003443655	GCA_002949735

CDS, coding DNA sequence; GC, guanine-cytosine; rRNA ribosomal RNA; tRNA, transfer RNA *Np*-PCC 73102, *N. punctiforme*; *Ne*-CCNP1411, *N. edaphicum*; *Ns*-Kutzing En, *N. sphaeroides; Nsp*. PMC N6, *Nostoc* sp. '*Peltigera membranacea* cyanobiont'. The comparison of *Np*-SAG 69.79 with other *Nostoc* species and an overview of the newly assembled genome, which is used in this study for further analysis.

Table 2 Genome statistics of *G. pyriformis*' Mollicutes/Mycoplasma-related endobacteria (MRE) and other MRE.

	***Gp*MRE**	***De*MREI-1**	***De*MREI-2**	***De*MREII**	***Dh*MRE**	***Ce*MRE**	***Rc*MRE**	***Rv*MRE**
Genome size (bp)	712 504	629 893	607 619	571 166	648 879	662 952	739 936	1 227 948
% BUSCO Genome Compl.	49.6	35.1	36.5	42.4	41.7	31.1	33.8	39
No. scaffolds	3	1	1	1	1	67	34	99
%GC	33	32	29	34	34	32	34	34
No. rRNAs	3	3	3	3	3	0	3	1
No. tRNAs	34	27	30	32	35	37	32	31
No. of CDS	1239	1085	1055	913	1302	1200	1241	2294
% Coding density	56.8	53.9	56	53.8	63.1	56.2	55.2	56.3
No. pseudogenes	27	229	205	159	213	61	8	491
Accession no.	This study	NWVA00000000	NWVA00000000	NWVA00000000	LN828718	JPXH00000000	JPXG00000000	JQIB00000000

CDS, coding DNA sequence; GC, guanine-cytosine; rRNA ribosomal RNA; tRNA, transfer RNA.
This table presents genome data for *G. pyriformis* MRE and various Mollicutes/Mycoplasma-related endobacteria (MRE) in other AMF species. MREs are a group of endobacteria often found with streamlined genome possibly due to reductive evolution.

Gene annotations were performed on *Np*-SAG 69.79 and *Gp*-MRE scaffolds using Bakta (Schwengers et al., 2021). Identical procedures were used on publicly available *Nostoc* spp. relatives and MRE genomes to allow direct genome comparisons (Table 1, Figure 2.2a). For *Np*-SAG 69.79, 8,691 genes were identified, and a BUSCO gene repertoire completeness of 98.4% (*Index of /v4/data/lineages/*, n.d.). In comparison, the reference genome *Nostoc punctiforme* ATCC 29133/PCC 73102 showed 7,118 genes with a BUSCO completeness of 99.4%. Other *Nostoc* species, such as *Nostoc sp. 'Peltigera membranacea cyanobiont'* N6, *N. edaphicum* CCNP1411, and *N. sphaeroides* Kutzing En, presented 7,578, 7,056, and 5,957 genes, respectively. The *Gp*-MRE scaffold have a complete BUSCO score of 49.6% and encodes 1,239 predicted proteins and 27 pseudogenes, compares with the average BUSCO genome assessment of 37.09%, and 1,299 predicted proteins and 192 predicted pseudogenes for other sequenced AMF MRE (Table 2, Figure 2.2b).

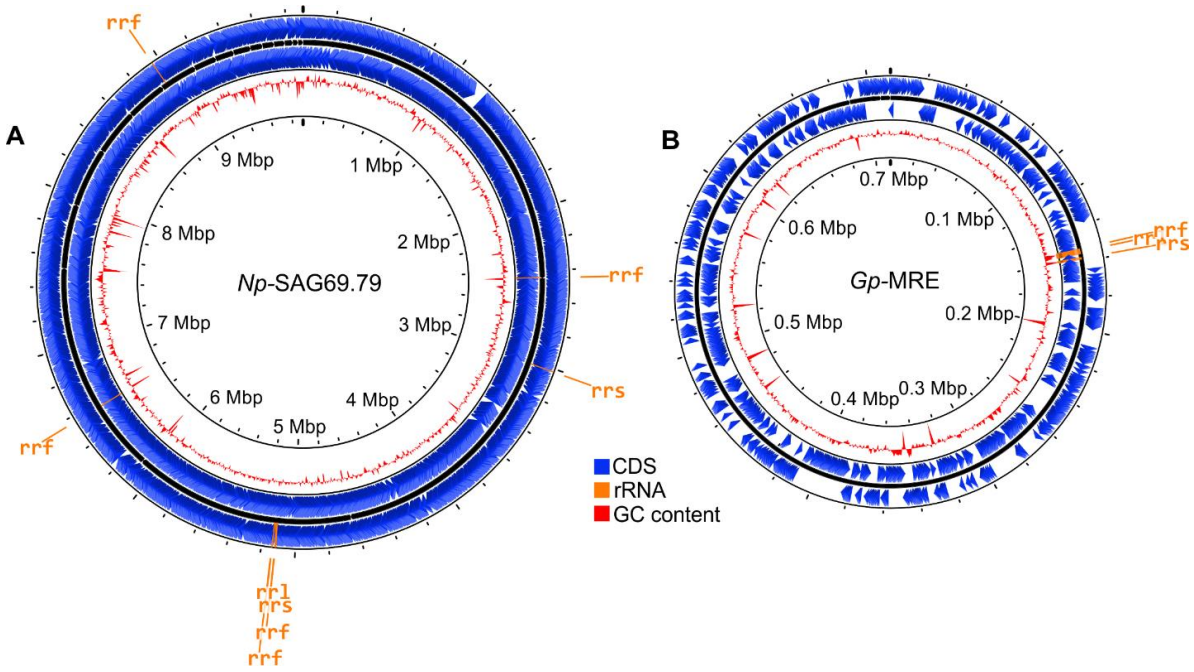

Figure 2.2. Endosymbiont genomes present in *G. pyriformis* microbiota. (**A**) Depicts the genome of *Np*-SAG 69.79, a cyanobacterium residing uniquely in the cytoplasm *Gp*-MRE of the obligate symbiont *G. pyriformis*. (**B**) The genome of Gp-MRE. Black circular arrows depict the scaffolds of the assemblies. Coding sequences (CDS) are represented by blue arrows, while rRNA genes are highlighted in orange. The 16S rRNA gene regions are labeled as rrs. The 23S/5S rRNA gene regions are labeled as rrl/rrf. The innermost ring indicates the GC content of each assembly.

2.5.2 Phylogenetic analyses of G. pyriformis endosymbionts

To determine the relationship of *Np*-SAG 69.79, the 16S ribosomal RNA (rRNA) gene was extracted from its genome and used with homologous sequences from *N. punctiforme* reference species and *Nostoc* species available in NCBI to generate a maximum likelihood phylogenetic tree (Figure S3). This analysis showed that *Np*-SAG 69.79 clusters in a single clade that contains multiple strains from culture collections sharing identical 16S rRNA sequences – e.g., *N. punctiforme* Jbr02, PCC-73102, SAG 69.79.

To investigate the origin of *Gp*-MRE, the amino-acid sequences of 14 single-copy orthologs were extracted from all available MRE genomes and used to produce a maximum likelihood phylogenetic tree (Figure 2.3). This analysis supports previous findings, based on based on other protein encoding genes (Sun et al., 2019), showcasing that AMF MRE separate into two main clusters/lineages, namely MRE-I and II. Among these, *Gp*-MRE clustered within the lineage that contains MRE-II.

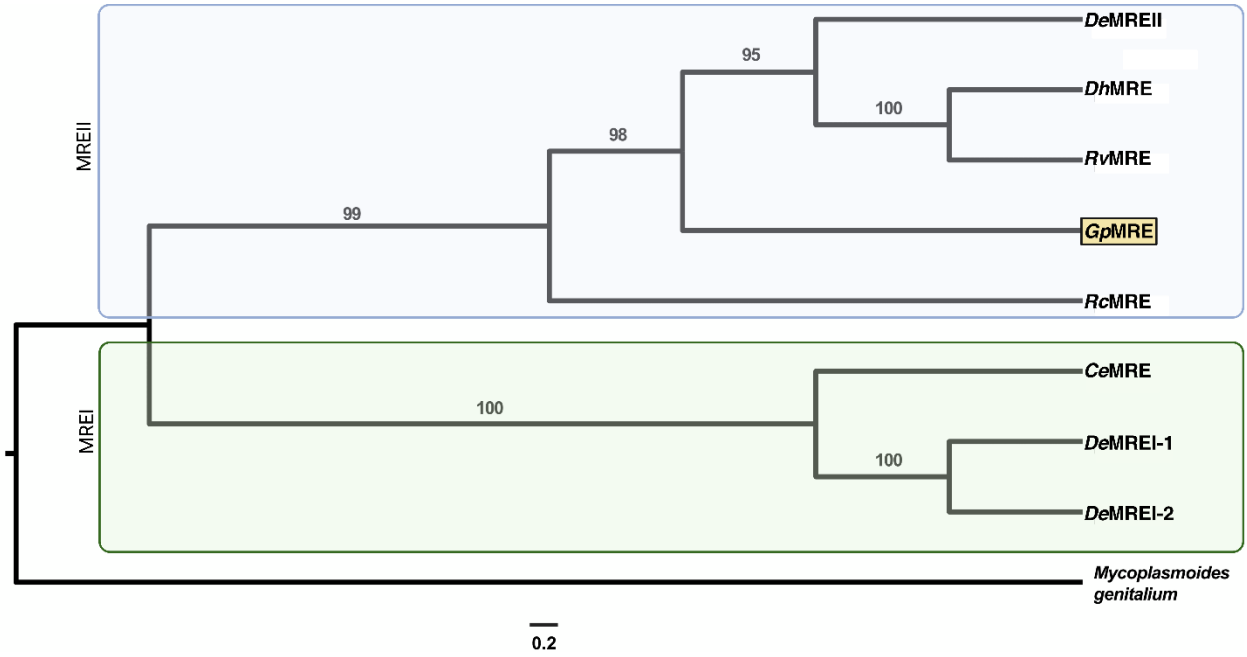

Figure 2.3 Phylogenetic tree that delineates the evolutionary relationships among sequenced MREs within *G. pyriformis*, compared with MREs from diverse AMF species. It underscores the diversification and shared ancestry of endosymbiotic relationships across fungal lineages. The analysis includes complete genome assemblies of MREs from multiple species and incorporates three metagenome assemblies, providing insights into their ancient associations with AMF. The scale represents 0.2 amino acid substitutions per site. Bootstrap support values over 80% are shown above branches.

2.5.3 Comparative analyses of Nostoc and MRE Genomes

To identify the shared functions and biological novelties present in *Np*-SAG 69.79, we clustered the genes present in the genome of *Np*-SAG 69.79 and those of free-living and lichen associated

symbiotic *Nostoc* strains with sequenced genomes into shared orthologues and unique genes using FastOrtho (*GitHub - olsonanl/FastOrtho*, n.d.) (Figure 2.4a).

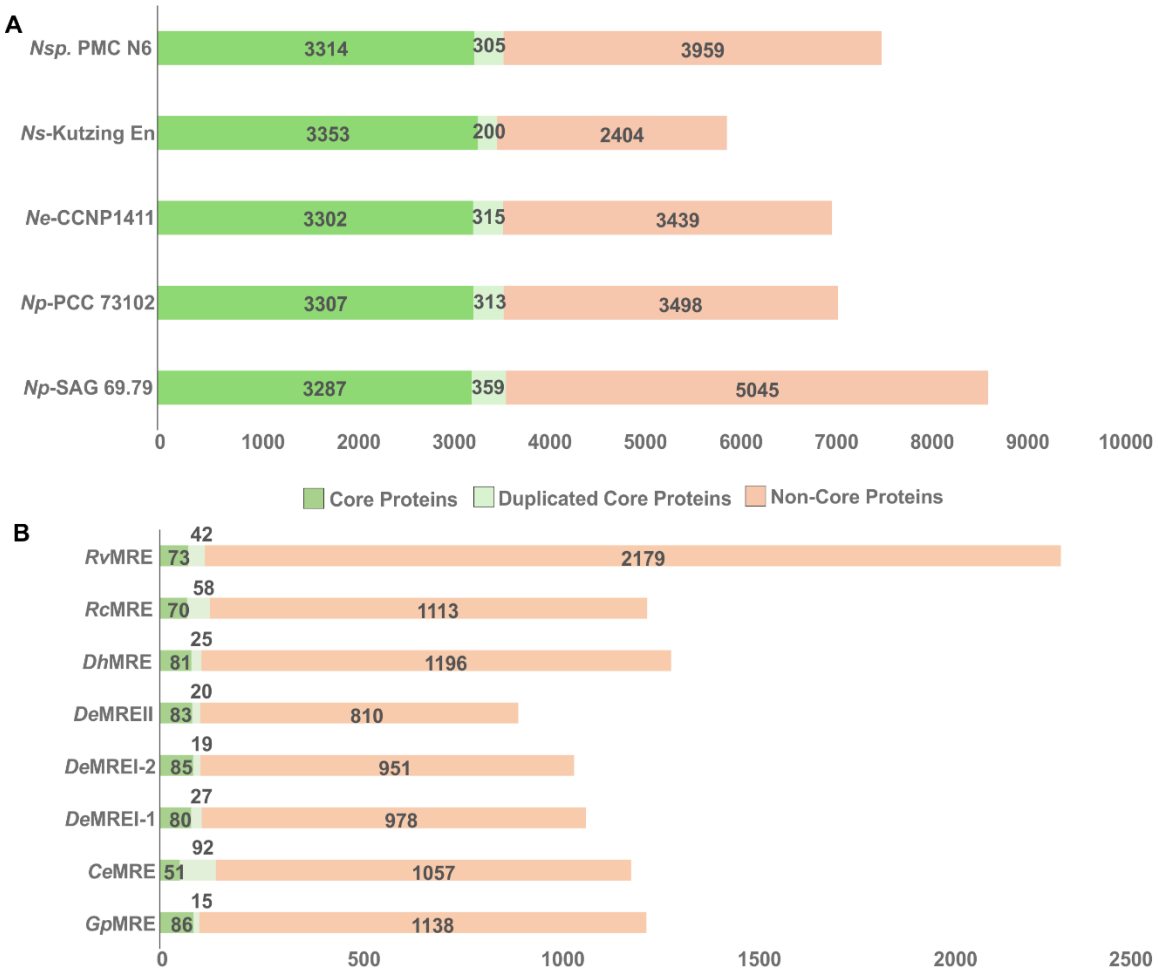

Figure 2.4 Ortholog Identification and Comparison in Endobacterial Genomes of *G. pyriformis*. Orthologous genes are indicated in green, distinguishing between single-copy core genes and duplicated core genes. The non-core genes are a combination of unique and differentially present coded proteins and are also highlighted in orange. **(A)** Shows the *Np*-SAG 69.79 genome in comparison with other closely related *Nostoc* species, emphasizing the core proteins shared among these five species. **(B)** Illustrates the core and non-core protein distribution in *Np*-SAG 69.79 compared with other AMF MREs, underscoring the significant genetic divergence among these species.

This analysis revealed that 41.95% of the gene repertoire across all analyzed *Nostoc* species

consists of core genes, with a notable enrichment in the COG categories for “Energy Production & Conversion” and “Signal Transduction Mechanisms”. Besides shared genes, *Np*-SAG 69.79 exhibits a unique gene enrichment, particularly in the category called “Mobilome” which is sixfold more prevalent compared to the reference strain *N. punctiforme* PCC 73102 (Table S1). *N. punctiforme* genomes also differ in putative gene functions. For example, the *Np*-SAG 69.79 genome carries a specific gene that encodes for a putative catechol 1,2-dioxygenase (K03381) known to involved in xenobiotic degradation and carries more candidate secreted proteins (SP) than their relatives – i.e. 560 SP *vs* an average of 427 SP in close relatives.

For MRE genomes, only 3.3% of their gene repertoire is composed of core genes shared by all the AMF MREs (Figure 2.4b, Table S2). Shared genes include an exonuclease ABC subunit B, which plays a putative role in nucleotide excision repair mechanisms, as well as a thioredoxin reductase. Annotated *Gp*-MRE specific genes using COG and KEGG databases (Table S3 & S4), reveal that *Gp*-MRE carries putative DNA and RNA helicases, as well as homologs of DNA (cytosine-5)-methyltransferase 1 and an ATP-dependent Lon Protease not found in close relatives. Other Glomeromycotina MRE carry specific functions, for example, the *Claroideoglomus etunicatum* and *Racocetra verrucosa* MRE encode for a protein superoxide dismutase, which is known to help defense against reactive oxygen species. Other MRE-specific genes include nucleoside-diphosphate kinases and a protein known to play a role in plant-pathogen interactions, the elongation factor Tu (Table S5).

2.5.4 Diversity and expansions of mobile genetic elements expansion in Np-SAG 67.79

To investigate the significant enrichment in the “mobilome” category in *Np*-SAG 69.79, the diversity of mobile genetic elements (MGE) was identified among all *N. punctiforme* strains genome analysed in this study. The mobile genetic elements (MGE) in procaryotes are

categorized into five classes: Phage (P), Replication/Recombination/Repair (RRR), Stability/transfer/defense (STD), Transfer (T), and Integration/excision (IE) (Brown et al., 2022). All these elements, which represent on average 4% of the *Nostoc* genomes analysed, promote bacteria defense, transformation, and variability (Brown et al., 2022) and vary in number among all *Nostoc* relatives in this study (Table 3).

Table 3. Mobile Genetic Elements (MGE) identified in *NP*-SAG 69.79 and other closely related *Nostoc* species.

	Np-SAG 69.79	***Np-PCC 73102***	***Ne-CCNP1411***	***Ns-Kutzing En***	***Nsp. PMC N6***
Transfer	25	13	17	7	23
Stability/transfer/defense	22	16	25	21	40
Replication/recombination/repair	29	15	18	14	23
Phage	12	16	13	9	17
Integration/excision	368	283	238	54	165

Np-PCC 73102; *N. punctiforme PCC 73102*, *Ne*-CCNP1411; *N. edaphicum, Ns*-Kutzing En; *N. sphaeroides*, *Nsp*. PMC N6; *Nostoc* sp. '*Peltigera membranacea* cyanobiont'.
The analysis conducted reveals a notable expansion in various categories of MGEs across these genomes. Particularly striking is the significant increase in the Insertion and Excision category observed in the genome of *Np*-SAG 69.79.

Among these, the IE class, which includes DNA transposons, show the highest variability among the *Nostoc* strains investigated —ranging from 54 copies in *N. sphaeroides* to 368 copies in *Np*-SAG 69.79 (Figure 2.5). The majority IE class sequences are either present in a single copy, or present up to 10 copies, indicating reduced transposition activity (Durrant et al., 2020). *Np*-SAG 69.79 also differs by notable unique expansions of three IE domains (InsB, ISPlu6L, and Tnpa) from insertion IS1, IS982, and IS200 families, respectively.

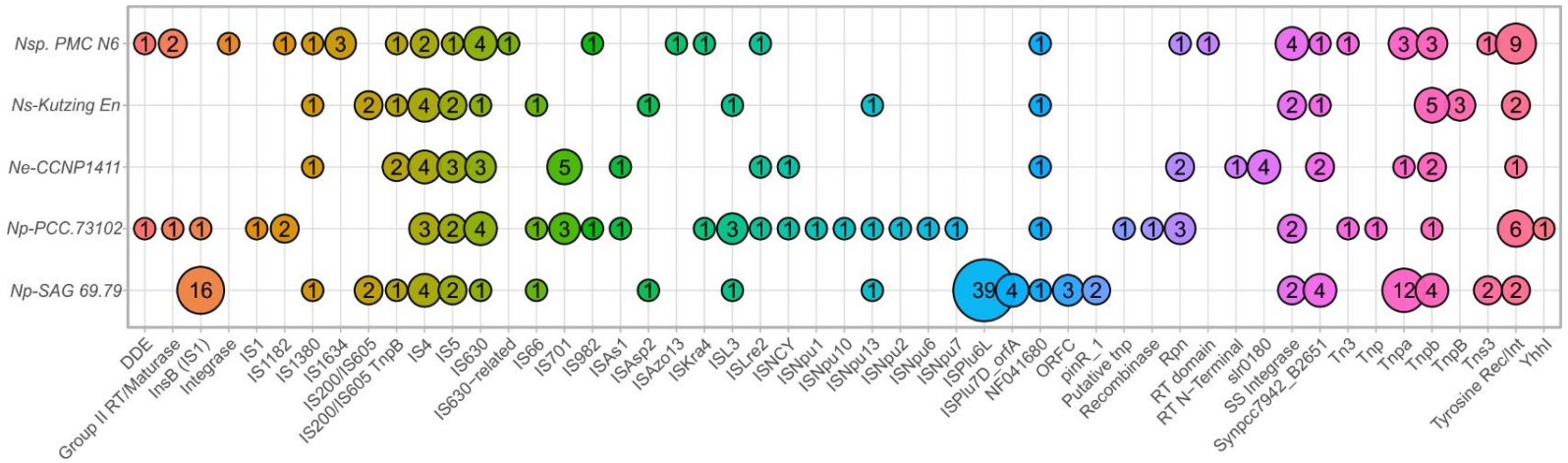

Figure 2.5. **Insertion/Excision domains in *Np*-SAG 69.79 and related *Nostoc* species.** On the x-axis are the domains found in each Nostoc species (y-axis). The circles represent, proportionally, the number of certain domains in a species.

2.5.3 A putative horizontal gene transfer between AMF and Geosiphon MRE

Past analyses of MRE have revealed the presence of potential horizontal gene transfers (HGT) between these genomes and other bacteria, as well as between those and the fungal host (Torres-Cortés et al., 2015). To identify if similar events also occurred in *Np*-SAG 69.79 and *Gp*-MRE, their genes were compared against available gene sequence databases to identify genes of possible eukaryotic origin and candidate HGT events were further investigated using phylogenetics. Overall, these analyses found no convincing HGT candidate in *Np*-SAG 69.79, however, a potential event was identified in *Gp*-MRE.

The putative *Gp*-MRE HGT includes a gene with highest sequence similarity to AMF proteins containing P-loop NTPase domains (Figure S4). Close homologue of this gene is not found outside Glomeromycotina/MRE, and contig inspections showed that candidate is located in between genes of bacterial origin (MRE), altogether indicating that homology mirrors a

transfer the host rather than representing a miss-assembled or contaminant contig (Torres-Cortés et al., 2015).

2.6 Discussion

Using k-mer based binning of Illumina reads, we were able to assemble, annotate, and compare the genomic content of two *Geosiphon* endosymbionts, namely *Np*-SAG 69.79 and *Gp*-MRE. Comparative analyses revealed that *Np*-SAG 69.79 and *Gp*-MRE harbor striking, unique genomic features that functionally distinguish them from free-living relatives.

In *Np*-SAG 69.79, the load of insertions sequences (IS) domains outweighs that of any known species of this genus with a sequenced genome available. It was proposed that gene loss resulting from the nutritionally rich environment of the host would relax control against potential deleterious proliferation of selfish elements in the symbiont (Siguier et al., 2015), and this could result in the large MGE expansions we observed in *Np*-SAG 69.79. It is intriguing to speculate that the gene disruptions promoted by the expansion of transposons would, over time, increase the dependency of the host and reduce the virulence (Fayad et al., 2020; Vandecraen et al., 2017), thus contributing to the establishment of long-term coevolution between the *Np*-SAG 69.79 and *G. pyriformis*.

Np-SAG 69.79 also carries a gene potentially involved in the degradation of xenobiotics. Catechol 1,2-dioxygenase is an enzyme that plays a crucial role in microbes, as it is involved in the degradation of biologically relevant aromatic compounds such as Benzoate, Fluorobenzoate and Toluene (D. P. Singh et al., 2018; Strunk & Engesser, 2013; C.-L. Wang et al., 2006). The presence of catechol 1,2-dioxygenase in this cyanobacterium suggests it has the metabolic capability to degrade benzoate, particularly in the presence of oxygen. This enzyme is part of a pathway that could potentially allow the cyanobacterium to utilize benzoate as a supplemental

source of carbon and energy under specific conditions. However, further research is needed to confirm if benzoate can serve as a significant carbon and energy source for this organism. An alternative explanation for the presence of this enzyme, along with other xenobiotic degrading enzymes found in the genome, is that the cyanobacterium acts as a detoxifier to protect its fungal host from environmental stress. Indeed, Benzoate and Fluorobenzoate are known antifungal compounds (Krebs et al., 1983; Valderrama et al., 2012) and bacterial degradation of these compounds may reflect additional benefit for maintaining *Np*-SAG 69.79 within *G. pyriformis* bladders.

For *Gp*-MRE, our analyses showed that sequence divergence in these highly adapted symbionts is extreme, leading not only to a streamlining of gene and metabolic contents but also in the extensive rearrangements and divergence that preclude direct sequence comparisons between MRE relatives. This divergence is also linked to the presence of unique genes/functions in MRE, which in *Gp*-MRE include a putative DNA (cytosine-5)-methyl transferase 1 and an ATP-dependent Lon Protease. The functional advantages conferred by specific proteins in *Gp*-MRE are challenging to assess. However, drawing on insights from model organisms, we can infer that the genes encoding DNA (cytosine-5)-methyltransferase 1 and ATP-dependent Lon Protease in *Gp*-MRE may be involved in epigenetic regulation, robust protein quality control, and most plausibly responding to stress (Ponger & Li, 2005; Van Melderen & Aertsen, 2009). In this case, these capabilities would ensure long-term survival and effective functioning of the symbiotic relationship between MRE and AMF.

Strain-specific superoxide dismutase, an enzyme crucial for protecting endosymbionts (and potentially their hosts) from reactive oxygen species produced in the cytoplasm by mitochondria, has been identified across various MREs, including *Gp*-MRE. This ensemble of

enzymes, each playing a pivotal role in environmental sensing and stress response, may underscore the potential of MRE to adapt and respond to dynamic environmental conditions; a trait that is beneficial in the context of its symbiotic existence.

Acknowledgments

The research is funded by the Discovery program of the Natural Sciences and Engineering Research Council (RGPIN2020-05643) and a Discovery Accelerator Supplements Program (RGPAS-2020-00033) to NC. NC is a University of Ottawa Research Chair in Microbial Genomics. JINO and MM were founded by Mitacs Accelerate Program (IT16902) and a Discovery Accelerator Supplements Program (RGPAS-2020-00033).

2.7 Supplementary Data

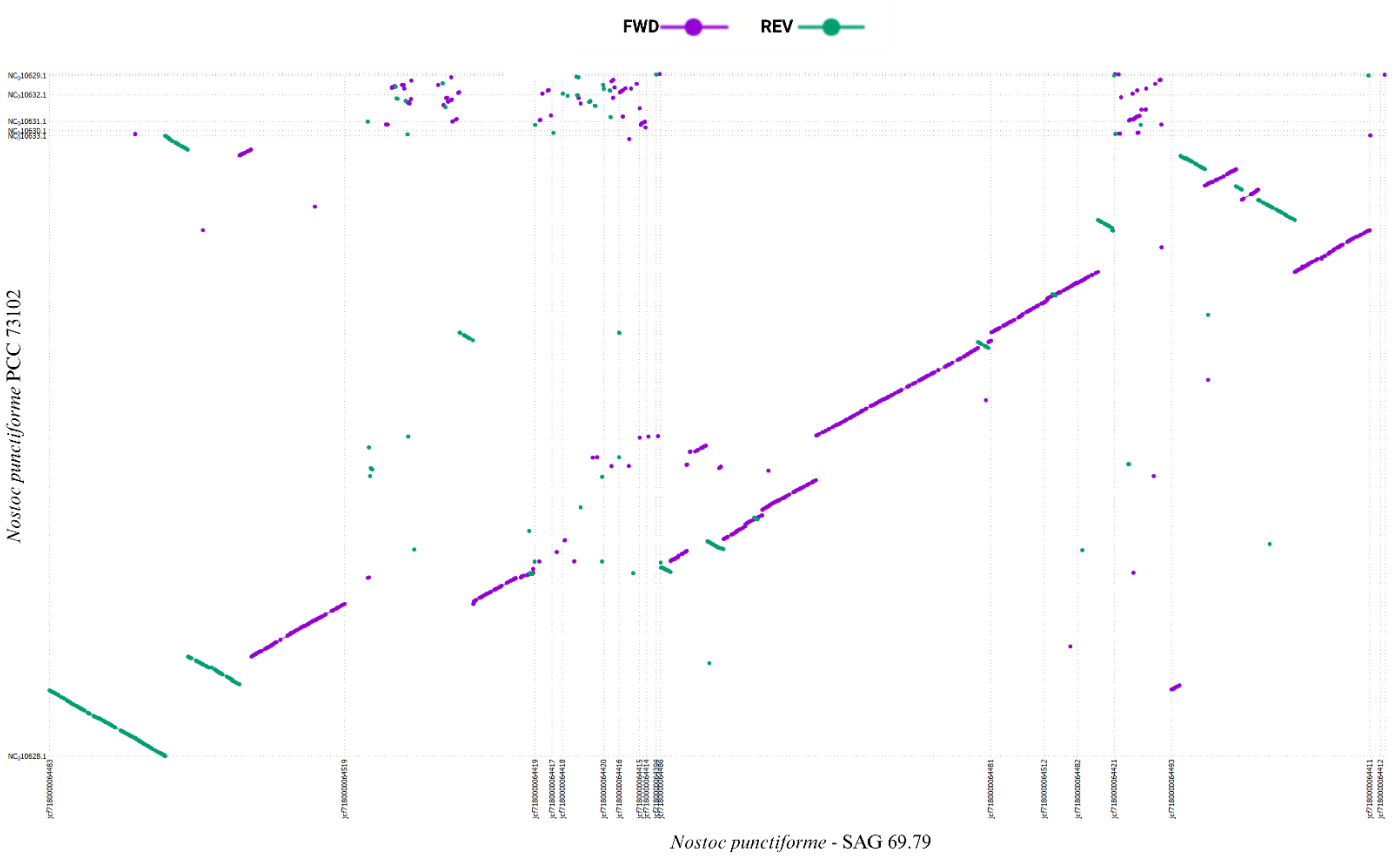

Figure S1. Alignment of *Nostoc* Scaffolds to the Reference *Nostoc* Punctiforme PCC 73102 Genome. Genomic scaffolds obtained from the metagenomic binning approach are aligned to the reference N. punctiforme genome assembly. Scaffolds that successfully mapped are deemed reliable and suitable for further analysis. The x-axis contains scaffolds from the *Nostoc* made in this study, and the reference *N. punctiforme* genome is on the y-axis.

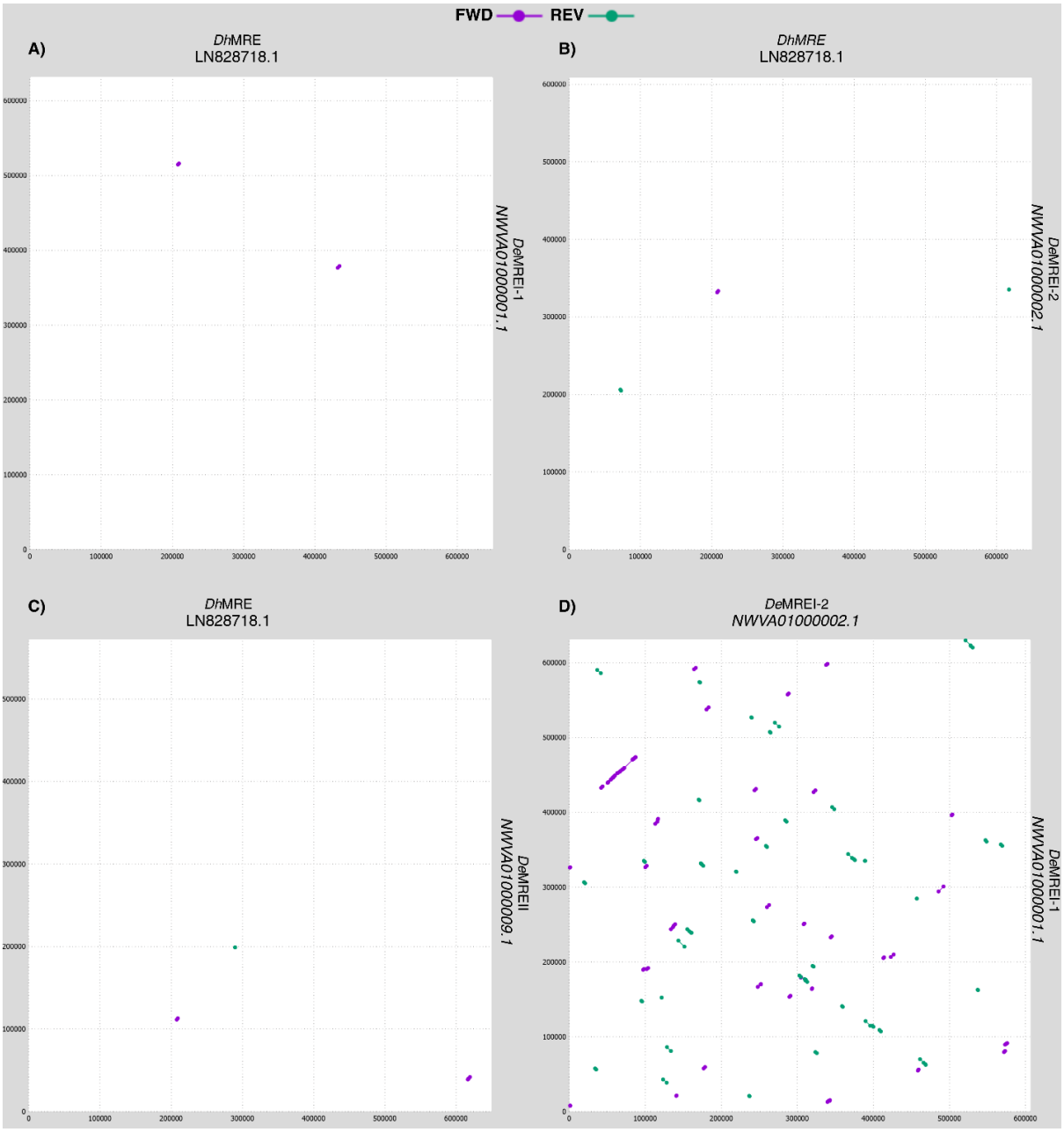

Figure S2. Comparative Alignment of Two Complete Mollicutes-Related Endobacteria (MRE) Assembles. This figure illustrated the alignment of two fully assembled MRE genomes, highlighting their divergence. The comparison showcases distinct genomic variations and structural differences between the two assemblies, emphasizing the evolutionary divergence and uniqueness of each MRE genome. A) Alignment of *Dentiscutata heterogama* MRE assembly in the x -axis and *Diversispora Epigaea* MREI-1 assembly in y-axis B) Alignment of *Dentiscutata heterogama* MRE assembly in the x -axis and *Diversispora Epigaea* MREI-2 assembly in y-axis C) Alignment of *Dentiscutata heterogama* MRE assembly in the x -axis and *Diversispora Epigaea* MREII assembly in y-axis D) Alignment of *Diversispora Epigaea* MREI-1 assembly in x-axis and *Diversispora Epigaea* MREI-2 assembly in y-axis.

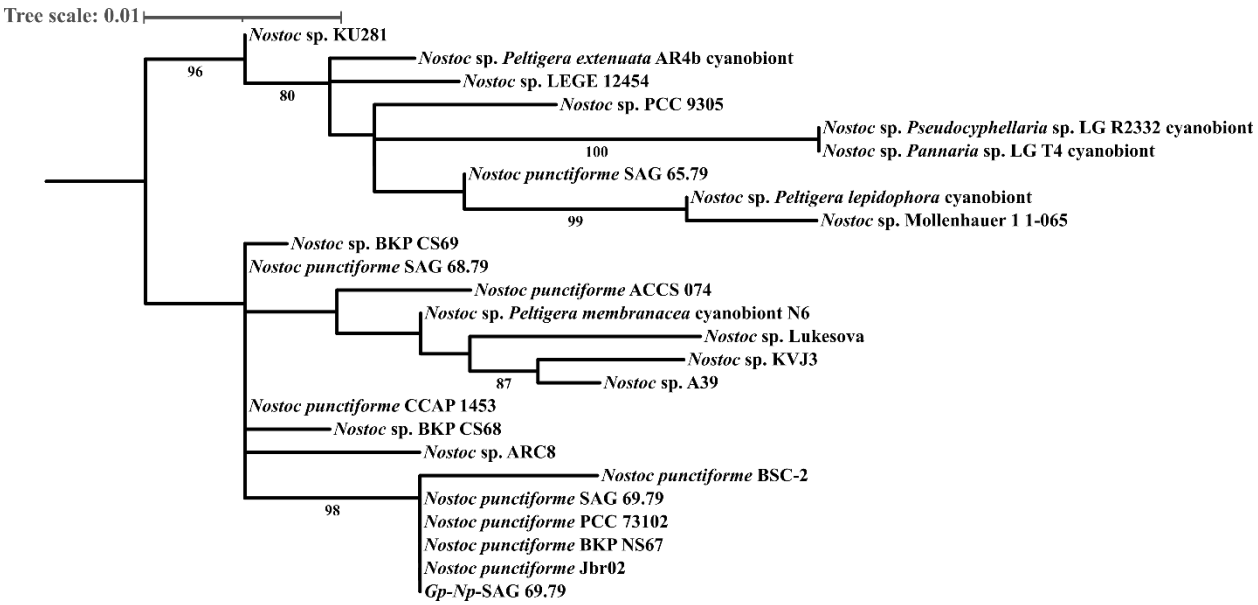

Figure S3. Maximum Likelihood Phylogenetic Tree of *Np*-SAG 69.79 and Related *Nostoc* Strains. The tree is based on the 16S rRNA gene sequences, showing high conservation among *N. punctiforme* strains and distinctly clustering *Np*-SAG 69.79 in a separate clade with multiple similar strains from culture collections. The scale represents 0.01 nucleotide substitutions per site. Bootstrap support values above 80% are shown below branches.

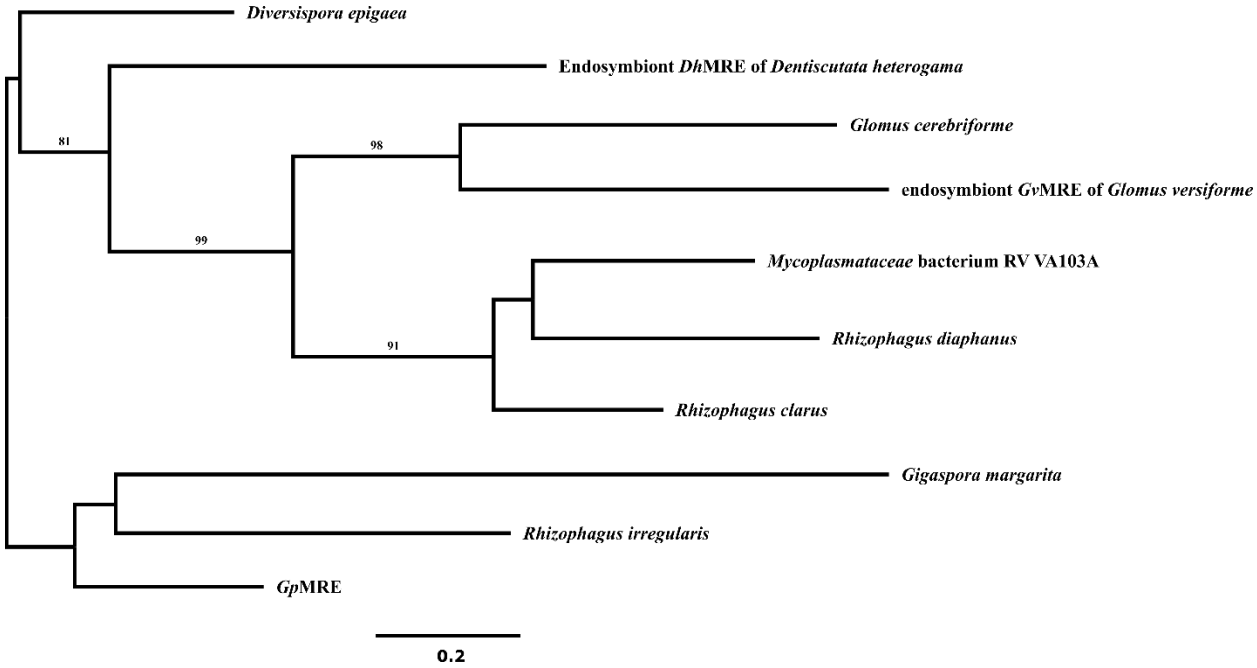

Figure S4. Putative Horizontal Gene Transfer (HGT) in Gp-MRE. The figure depicts a gene in *Gp*-MRE with high sequence similarity to AMF proteins, specifically those with P-loop NTPase domains, and partial homology to MRE genes. The gene's location suggests a potential HGT from the host. The scale represents 0.2 amino acid substitution per site. Bootstrap values over 80% are shown above branches.

Table S1. Functional Enrichment Heatmap of unique protiens found in each *Nostoc* species via COG Analysis. *Np*-PCC 73102; *Nostoc punctiforme* PCC 73102, *Ne*-CCNP1411; *Nostoc edaphicum*, Ns-Kutzing En; *Nostoc sphaeroides*, *Nsp*. PMC N6; *Nostoc* sp. '*Peltigera membranacea* cyanobiont'. The functional enrichment of strain-specific unique proteins identified by FastOrtho analysis based on the COG (Clusters of Orthologous Groups) database.

COG Categories	Np-SAG 69.79	Np-PCC 73102	Ne-CCNP1411	Ns-Kutzing En	Nsp. PMC N6
(X) Mobilome: Prophages, Transposons	395	62	108	52	63
(U) Intracellular trafficking, secretion, and vesicular transport	8	4	3	2	3
(FV) Nucleotide transport and metabolism + Defense mechanism	3	0	1	0	0
(G) Carbohydrate transport and metabolism	26	11	11	10	16
(S) Function Unknown	51	30	30	22	32
(UW) Intracellular trafficking, secretion, and vesicular transport + Extracellular structures	5	0	0	1	2
(CR) Energy production and conversion + General Function	3	1	0	0	1
(HR) Coenzyme Transport and metabolism + General function	11	4	7	4	4
(DR) Cell cycle control, cell division, chromosome partitioning + General function	2	0	0	1	0
(T) Signal transduction mechanisms	66	49	29	37	40
(MDT) Cell Wall/Envelope + Cell cycle control + Signal transduction mechanisms	2	0	1	0	1
(O) Posttranslational modification, protein turnover, chaperones	14	22	20	19	19

Table S2. Functional Categorization of Core Proteins in the Glomeromycotina *Mollicutes/Mycoplasma-related endobacteria* (MRE) using COG Database. Each protein is assigned a COG ID and falls into specific functional categories as indicated. This table forms part of a larger analysis aimed at understanding the genomic landscape of MREs by examining the frequency of COG categories present in core proteins extracted from the FastOrtho analysis.

COG ID	COG category	Percent Abundance	Raw Hits
COG0124	J	3.614457831	3
COG0013	J	3.614457831	3
COG0520	E	3.614457831	3
COG1185	J	3.614457831	3
COG0322	L	3.614457831	3
COG0423	J	3.614457831	3
COG0085	K	2.409638554	2
COG0532	J	2.409638554	2
COG0466	O	2.409638554	2
COG0595	J	2.409638554	2
COG0094	J	2.409638554	2
COG0018	J	2.409638554	2
COG1692	T	2.409638554	2
COG0012	J	2.409638554	2
COG0210	L	2.409638554	2
COG0481	J	2.409638554	2
COG0495	J	2.409638554	2
COG0272	L	2.409638554	2
COG0533	J	2.409638554	2
COG0081	J	1.204819277	1
COG0222	J	1.204819277	1
COG0086	K	1.204819277	1
COG1253	R	1.204819277	1
COG0195	K	1.204819277	1
COG1160	J	1.204819277	1
COG0064	J	1.204819277	1
COG0125	F	1.204819277	1
COG0088	J	1.204819277	1
COG0258	L	1.204819277	1
COG0556	L	1.204819277	1
COG0037	J	1.204819277	1
COG0522	J	1.204819277	1
COG0052	J	1.204819277	1
COG0492	O	1.204819277	1
COG0184	J	1.204819277	1
COG0203	J	1.204819277	1

COG0802	J	1.204819277	1
COG0186	J	1.204819277	1
COG0093	J	1.204819277	1
COG0198	J	1.204819277	1
COG0564	J	1.204819277	1
COG0216	J	1.204819277	1
COG0215	J	1.204819277	1
COG0335	J	1.204819277	1
COG0112	E	1.204819277	1
COG2094	L	1.204819277	1
COG1214	J	1.204819277	1
COG0445	J	1.204819277	1
COG0275	J	1.204819277	1
COG0267	J	1.204819277	1
COG0193	J	1.204819277	1
COG0484	O	1.204819277	1
COG1131	V	1.204819277	1
COG1132	V	1.204819277	1
COG0359	J	1.204819277	1
COG0561	HR	1.204819277	1
COG0482	J	1.204819277	1
COG0172	J	1.204819277	1

Table S3. Functional Categorization of Unique Proteins in the Glomeromycotina Mollicutes/Mycoplasma-related endobacteria (MRE) using COG Database. The analysis specifically explores the prevalence of various COG categories within unique *Gp*MRE specific proteins identified through the FastOrtho analysis. This table is a component of a broader analysis focused on understanding the genomic composition of *Gp*MRE.

COG ID	COG category	Percent Abundance	Raw Hits
COG0286	V	15.38461538	6
COG2189	L	12.82051282	5
COG0270	L	10.25641026	4
COG1061	KL	7.692307692	3
COG0209	F	7.692307692	3
COG0732	V	5.128205128	2
COG3335	X	5.128205128	2
COG0050	J	2.564102564	1
COG0513	L	2.564102564	1
COG0847	L	2.564102564	1
COG0112	E	2.564102564	1
COG3677	X	2.564102564	1
COG3440	V	2.564102564	1
COG0466	O	2.564102564	1
COG1541	H	2.564102564	1
COG1192	DN	2.564102564	1
COG0507	L	2.564102564	1
COG1328	F	2.564102564	1
COG0666	T	2.564102564	1
COG3415	V	2.564102564	1
COG0468	L	2.564102564	1

Table S4. Correlating unique *GpMRE* specific proteins with KEGG Orthology and Definitions. Unique *Gp*MRE specific proteins extracted from the FastOrtho were analyzed with KEGG Database to better obtain a holistic view of metabolic pathways present.

Query	Kegg Orthology	Definition
DIPOKK_04140		
DIPOKK_01025		
DIPOKK_05425		
DIPOKK_05420		
DIPOKK_00715		
DIPOKK_00010		
DIPOKK_01170		
DIPOKK_00720		
DIPOKK_00015		
DIPOKK_01175		
DIPOKK_04300		
DIPOKK_04270		
DIPOKK_00115		
DIPOKK_03020		
DIPOKK_03115		
DIPOKK_03435		
DIPOKK_03110		
DIPOKK_03430		
DIPOKK_00145		
DIPOKK_00330		
DIPOKK_00470		
DIPOKK_00920		
DIPOKK_00990		
DIPOKK_00995		
DIPOKK_01055		
DIPOKK_01150		
DIPOKK_01235		
DIPOKK_01365		
DIPOKK_01510		
DIPOKK_01645		
DIPOKK_01755		
DIPOKK_01830		
DIPOKK_01915		
DIPOKK_02105		
DIPOKK_02110		
DIPOKK_02215		
DIPOKK_02280		

DIPOKK_02360		
DIPOKK_02370		
DIPOKK_02570		
DIPOKK_02595		
DIPOKK_02650		
DIPOKK_02665		
DIPOKK_02725		
DIPOKK_02740		
DIPOKK_02810		
DIPOKK_02890		
DIPOKK_02985		
DIPOKK_03025		
DIPOKK_03095		
DIPOKK_03150		
DIPOKK_03345		
DIPOKK_03550		
DIPOKK_03705		
DIPOKK_03900	K01338	lon; ATP-dependent Lon protease [EC:3.4.21.53]
DIPOKK_03985		
DIPOKK_03995		
DIPOKK_04025		
DIPOKK_04030		
DIPOKK_04230		
DIPOKK_04275		
DIPOKK_04285		
DIPOKK_04315		
DIPOKK_04345		
DIPOKK_04400		
DIPOKK_04410	K00558	DNMT1, dcm; DNA (cytosine-5)-methyltransferase 1 [EC:2.1.1.37]
DIPOKK_04525		
DIPOKK_04620		
DIPOKK_04670		
DIPOKK_04695		
DIPOKK_04700		
DIPOKK_04790		
DIPOKK_04915		
DIPOKK_05000		
DIPOKK_05010		
DIPOKK_05365		
DIPOKK_05485		
DIPOKK_05490		
DIPOKK_05505		
DIPOKK_05510		
DIPOKK_05535		

DIPOKK_05655		
DIPOKK_05665		
DIPOKK_05695		
DIPOKK_05700		
DIPOKK_05750		
DIPOKK_05755		
DIPOKK_06040	K07494	K07494; putative transposase
DIPOKK_06080		
DIPOKK_06090		
DIPOKK_06265		
DIPOKK_06330		

Table S5. Detailed KEGG Orthology and Gene Function Analysis of differentially present genes in *Gp*MRE. This table presents a comprehensive summary of KEGG Orthology (KO) terms associated with genes that are differentially present in the *Gp*MRE dataset. It has been utilized in conjunction with FastOrtho analysis to identify which genes are uniquely present in certain Glomeromycota species, while absent in others.

Query	Kegg Orthology	Definition
DIPOKK_00050		
DIPOKK_00075		
DIPOKK_00080		
DIPOKK_00085		
DIPOKK_00090	K03046	rpoC; DNA-directed RNA polymerase subunit beta' [EC:2.7.7.6]
DIPOKK_00100		
DIPOKK_00105		
DIPOKK_00110		
DIPOKK_00135		
DIPOKK_00140		
DIPOKK_00150		
DIPOKK_00160	K02834	rbfA; ribosome-binding factor A
DIPOKK_00200		
DIPOKK_00230		
DIPOKK_00235		
DIPOKK_00240		
DIPOKK_00245		
DIPOKK_00265		
DIPOKK_00270		
DIPOKK_00305		
DIPOKK_00335	K02838	frr, MRRF, RRF; ribosome recycling factor
DIPOKK_00360		
DIPOKK_00365		
DIPOKK_00405		
DIPOKK_00425		
DIPOKK_00445		
DIPOKK_00480		
DIPOKK_00495		
DIPOKK_00520		
DIPOKK_00535		
DIPOKK_00540		
DIPOKK_00545		
DIPOKK_00575		
DIPOKK_00580		
DIPOKK_00595		
DIPOKK_00605		
DIPOKK_00610		
DIPOKK_00615	K02890	RP-L22, MRPL22, rplV; large subunit ribosomal protein L22

DIPOKK_00620	K02965	RP-S19, rpsS; small subunit ribosomal protein S19
DIPOKK_00625	K02886	RP-L2, MRPL2, rplB; large subunit ribosomal protein L2
DIPOKK_00630	K02892	RP-L23, MRPL23, rplW; large subunit ribosomal protein L23
DIPOKK_00655	K02906	RP-L3, MRPL3, rplC; large subunit ribosomal protein L3
DIPOKK_00675		
DIPOKK_00680		
DIPOKK_00685		
DIPOKK_00730	K03702	uvrB; excinuclease ABC subunit B
DIPOKK_00735	K02968	RP-S20, rpsT; small subunit ribosomal protein S20
DIPOKK_00740	K01507	ppa; inorganic pyrophosphatase [EC:3.6.1.1]
DIPOKK_00745		
DIPOKK_00760		
DIPOKK_00765		
DIPOKK_00770		
DIPOKK_00775		
DIPOKK_00780		
DIPOKK_00790		
DIPOKK_00795		
DIPOKK_00800		
DIPOKK_00810		
DIPOKK_00815	K03685	rnc, DROSHA, RNT1; ribonuclease III [EC:3.1.26.3]
DIPOKK_00835	K03798	ftsH, hflB; cell division protease FtsH [EC:3.4.24.-]
DIPOKK_00845		
DIPOKK_00860		
DIPOKK_00865		
DIPOKK_00875	K03536	rnpA; ribonuclease P protein component [EC:3.1.26.5]
DIPOKK_00890	K04564	SOD2; superoxide dismutase, Fe-Mn family [EC:1.15.1.1]
DIPOKK_00930		
DIPOKK_00965		
DIPOKK_00980		
DIPOKK_01060	K03427	hsdM; type I restriction enzyme M protein [EC:2.1.1.72]
DIPOKK_01065		
DIPOKK_01075		
DIPOKK_01105	K02867	RP-L11, MRPL11, rplK; large subunit ribosomal protein L11
DIPOKK_01245		
DIPOKK_01250	K01151	nfo; deoxyribonuclease IV [EC:3.1.21.2]
DIPOKK_01260		
DIPOKK_01290		
DIPOKK_01295		
DIPOKK_01300		
DIPOKK_01370		
DIPOKK_01385		
DIPOKK_01390	K02948	RP-S11, MRPS11, rpsK; small subunit ribosomal protein S11
DIPOKK_01395	K02952	RP-S13, rpsM; small subunit ribosomal protein S13
DIPOKK_01400	K02518	infA; translation initiation factor IF-1
DIPOKK_01405	K01265	map; methionyl aminopeptidase [EC:3.4.11.18]

DIPOKK_01415		
DIPOKK_01435	K02876	RP-L15, MRPL15, rplO; large subunit ribosomal protein L15
DIPOKK_01440	K02988	RP-S5, MRPS5, rpsE; small subunit ribosomal protein S5
DIPOKK_01445	K02881	RP-L18, MRPL18, rplR; large subunit ribosomal protein L18
DIPOKK_01450	K02933	RP-L6, MRPL6, rplF; large subunit ribosomal protein L6
DIPOKK_01460		
DIPOKK_01465		
DIPOKK_01470		
DIPOKK_01485	K02950	RP-S12, MRPS12, rpsL; small subunit ribosomal protein S12
DIPOKK_01490	K02992	RP-S7, MRPS7, rpsG; small subunit ribosomal protein S7
DIPOKK_01495		
DIPOKK_01500	K02355	fusA, GFM, EFG; elongation factor G
DIPOKK_01505	K02358	tuf, TUFM; elongation factor Tu
DIPOKK_01520	K03664	smpB; SsrA-binding protein
DIPOKK_01550	K00567	ogt, MGMT; methylated-DNA-[protein]-cysteine S-methyltransferase [EC:2.1.1.63]
DIPOKK_01560		
DIPOKK_01580		
DIPOKK_01600	K13993	HSP20; HSP20 family protein
DIPOKK_01610		
DIPOKK_01615		
DIPOKK_01650		
DIPOKK_01675	K02982	RP-S3, rpsC; small subunit ribosomal protein S3
DIPOKK_01680		
DIPOKK_01685	K02904	RP-L29, rpmC; large subunit ribosomal protein L29
DIPOKK_01700		
DIPOKK_01740		
DIPOKK_01745		
DIPOKK_01765		
DIPOKK_01775		
DIPOKK_01800	K07494	K07494; putative transposase
DIPOKK_01815		
DIPOKK_01820		
DIPOKK_01825		
DIPOKK_01835		
DIPOKK_01870		
DIPOKK_01950		
DIPOKK_02075		
DIPOKK_02100		
DIPOKK_02145		
DIPOKK_02150		
DIPOKK_02155		
DIPOKK_02160		
DIPOKK_02165		
DIPOKK_02170		
DIPOKK_02180		

DIPOKK_02200		
DIPOKK_02210		
DIPOKK_02220		
DIPOKK_02225		
DIPOKK_02235	K03530	hupB; DNA-binding protein HU-beta
DIPOKK_02250	K03671	TXN, trxA; thioredoxin
DIPOKK_02285	K02909	RP-L31, rpmE; large subunit ribosomal protein L31
DIPOKK_02290	K00567	ogt, MGMT; methylated-DNA-[protein]-cysteine S-methyltransferase [EC:2.1.1.63]
DIPOKK_02325		
DIPOKK_02330	K01867	WARS, trpS; tryptophanyl-tRNA synthetase [EC:6.1.1.2]
DIPOKK_02340		
DIPOKK_02345		
DIPOKK_02350		
DIPOKK_02355		
DIPOKK_02395		
DIPOKK_02405		
DIPOKK_02410		
DIPOKK_02440	K07171	mazF, ndoA, chpA; mRNA interferase MazF [EC:3.1.-.-]
DIPOKK_02445		
DIPOKK_02455		
DIPOKK_02460		
DIPOKK_02465		
DIPOKK_02470		
DIPOKK_02475		
DIPOKK_02480		
DIPOKK_02485		
DIPOKK_02510		
DIPOKK_02540		
DIPOKK_02545		
DIPOKK_02685		
DIPOKK_02690	K03218	rlmB; 23S rRNA (guanosine2251-2'-O)-methyltransferase [EC:2.1.1.185]
DIPOKK_02715		
DIPOKK_02720		
DIPOKK_02815		
DIPOKK_02820	K00554	trmD; tRNA (guanine37-N1)-methyltransferase [EC:2.1.1.228]
DIPOKK_02830	K01889	FARSA, pheS; phenylalanyl-tRNA synthetase alpha chain [EC:6.1.1.20]
DIPOKK_02835		
DIPOKK_02855	K02888	RP-L21, MRPL21, rplU; large subunit ribosomal protein L21
DIPOKK_02920		
DIPOKK_02925		
DIPOKK_02930		
DIPOKK_02935		
DIPOKK_02940		
DIPOKK_02945		
DIPOKK_02955		

DIPOKK_02960		
DIPOKK_02980		
DIPOKK_03030		
DIPOKK_03040	K00940	ndk, NME; nucleoside-diphosphate kinase [EC:2.7.4.6]
DIPOKK_03145		
DIPOKK_03220	K03652	MPG; DNA-3-methyladenine glycosylase [EC:3.2.2.21]
DIPOKK_03235		
DIPOKK_03250		
DIPOKK_03265		
DIPOKK_03280		
DIPOKK_03310		
DIPOKK_03355		
DIPOKK_03360	K21498	higA-1; antitoxin HigA-1
DIPOKK_03375		
DIPOKK_03380		
DIPOKK_03420		
DIPOKK_03425		
DIPOKK_03440		
DIPOKK_03530		
DIPOKK_03540		
DIPOKK_03575		
DIPOKK_03795		
DIPOKK_03820		
DIPOKK_03825		
DIPOKK_03835	K09976	K09976; uncharacterized protein
DIPOKK_03845	K02887	RP-L20, MRPL20, rplT; large subunit ribosomal protein L20
DIPOKK_03850		
DIPOKK_03855		
DIPOKK_03860		
DIPOKK_03885	K02503	HINT1_2, hinT, hit; histidine triad (HIT) family protein [EC:3.9.1.-]
DIPOKK_03890		
DIPOKK_03905		
DIPOKK_03910		
DIPOKK_03970	K02899	RP-L27, MRPL27, rpmA; large subunit ribosomal protein L27
DIPOKK_03980		
DIPOKK_04035		
DIPOKK_04080		
DIPOKK_04090		
DIPOKK_04115		
DIPOKK_04120	K21498	higA-1; antitoxin HigA-1
DIPOKK_04175		
DIPOKK_04190		
DIPOKK_04210		
DIPOKK_04225		
DIPOKK_04245		
DIPOKK_04250		

DIPOKK_04295		
DIPOKK_04320		
DIPOKK_04335	K07171	mazF, ndoA, chpA; mRNA interferase MazF [EC:3.1.-.-]
DIPOKK_04340	K04763	xerD; integrase/recombinase XerD
DIPOKK_04355		
DIPOKK_04360		
DIPOKK_04370		
DIPOKK_04375		
DIPOKK_04420		
DIPOKK_04445	K10563	mutM, fpg; formamidopyrimidine-DNA glycosylase [EC:3.2.2.23 4.2.99.18]
DIPOKK_04450		
DIPOKK_04460	K02520	infC, MTIF3; translation initiation factor IF-3
DIPOKK_04475		
DIPOKK_04480		
DIPOKK_04510	K04043	dnaK, HSPA9; molecular chaperone DnaK
DIPOKK_04515		
DIPOKK_04520		
DIPOKK_04530	K03695	clpB; ATP-dependent Clp protease ATP-binding subunit ClpB
DIPOKK_04555		
DIPOKK_04565	K06218	relE, stbE; mRNA interferase RelE/StbE
DIPOKK_04570	K18918	relB; RHH-type transcriptional regulator, rel operon repressor / antitoxin RelB
DIPOKK_04605		
DIPOKK_04640		
DIPOKK_04650	K03496	parA, soj; chromosome partitioning protein
DIPOKK_04655		
DIPOKK_04690		
DIPOKK_04705		
DIPOKK_04720		
DIPOKK_04725		
DIPOKK_04750	K02963	RP-S18, MRPS18, rpsR; small subunit ribosomal protein S18
DIPOKK_04755		
DIPOKK_04820		
DIPOKK_04835		
DIPOKK_04880		
DIPOKK_04890		
DIPOKK_04900	K21498	higA-1; antitoxin HigA-1
DIPOKK_04925		
DIPOKK_04970		
DIPOKK_04985		
DIPOKK_04990		
DIPOKK_05005		
DIPOKK_05015		
DIPOKK_05070		
DIPOKK_05080		
DIPOKK_05120		
DIPOKK_05135		

DIPOKK_05195		
DIPOKK_05280		
DIPOKK_05285		
DIPOKK_05300		
DIPOKK_05315		
DIPOKK_05325		
DIPOKK_05330		
DIPOKK_05335		
DIPOKK_05340	K02946	RP-S10, MRPS10, rpsJ; small subunit ribosomal protein S10
DIPOKK_05345		
DIPOKK_05350	K18682	rny; ribonucrease Y [EC:3.1.-.-]
DIPOKK_05410		
DIPOKK_05495		
DIPOKK_05500		
DIPOKK_05520	K03650	mnmE, trmE, MSS1; tRNA modification GTPase [EC:3.6.-.-]
DIPOKK_05600		
DIPOKK_05620		
DIPOKK_05635	K02871	RP-L13, MRPL13, rplM; large subunit ribosomal protein L13
DIPOKK_05640	K02996	RP-S9, MRPS9, rpsI; small subunit ribosomal protein S9
DIPOKK_05645		
DIPOKK_05685		
DIPOKK_05710	K03427	hsdM; type I restriction enzyme M protein [EC:2.1.1.72]
DIPOKK_05715		
DIPOKK_05830		
DIPOKK_05835		
DIPOKK_05840		
DIPOKK_05855		
DIPOKK_05890	K04567	KARS, lysS; lysyl-tRNA synthetase, class II [EC:6.1.1.6]
DIPOKK_05895		
DIPOKK_05920		
DIPOKK_05985		
DIPOKK_05995		
DIPOKK_06035		
DIPOKK_06055		
DIPOKK_06060		
DIPOKK_06075		
DIPOKK_06085	K03553	recA; recombination protein RecA
DIPOKK_06185	K03496	parA, soj; chromosome partitioning protein
DIPOKK_06190		
DIPOKK_06195		
DIPOKK_06205		
DIPOKK_06210		
DIPOKK_06260		
DIPOKK_06335		
DIPOKK_06345		

CHAPTER THREE

Discussion, Future Directions, and Concluding Remarks

3.1 Importance of Research

This study employed a metagenomic approach to enhance understanding of the microbiota associated with AMF, specifically focusing on *G. pyriformis*. Unlike typical AMF that engages in mutualistic exchanges with plant roots, *G. pyriformis* forms unique symbiotic associations with endobacteria within its bladder-like structures (Malar et al., 2021). This distinction made *G. pyriformis* an invaluable model organism for exploring the ecological and evolutionary aspects of endobacterial symbiosis in AMF. A major challenge in this research area was the complexity involved in isolating and culturing these endobacteria from their fungal host in laboratory settings. To navigate this obstacle, this study harnessed the power of metagenomics to examine the genomic diversity and structure of the endobacterial community within *G. pyriformis*. The primary objective was to analyze the genomic composition of these endobacteria and conduct comparative genomic assessments with other AMF-associated taxa. By employing a metagenomic binning strategy, this research identified unique and shared genomic characteristics among the *Nostoc* and potential MRE species cohabiting the *G. pyriformis* bladders, thereby shedding light on their functional roles and evolutionary histories in symbiotic relationships with AMF hosts.

3.2 Summary of Novel Findings

3.2.1 Genomic Assembly and Annotation of Endosymbionts

In their previous research, Malar et al. (2021) employed a metagenomic binning approach to successfully assemble the genome of *G. pyriformis*. Building upon this foundation, the current study extends their methodology which identified 18 genome scaffolds aligned with *N. punctiforme,* surpassing its size and GC content with 9.9 Mb and 40.76%, respectively. The study also found three *G. pyriformis* MRE scaffolds totaling 712,504bp and performed gene

annotations revealing 8,691 genes in *Np*-SAG 69.79. Significantly, the analysis achieved a 98.4% BUSCO completeness score for *Np*-SAG 69.79, compared to a 49.6% BUSCO score for the *Gp*-MRE genome. This discrepancy highlights significant genomic diversity and structure variations among endosymbionts. Benchmarking a newly assembled genome with BUSCO scores is an important step, as it offers an objective measure of genomic completeness by checking for the presence of conserved genes that expected to be found in all members of a specific lineage. Thus, the relatively high BUSCO score for *Np*-SAG 69.79 and *Gp*-MRE displays thoroughness and integrity of the endosymbiont genomic assemblies and annotations, providing a solid foundation for further analysis.

3.2.2 Phylogenetic Investigation of Endosymbionts in G. pyriformis

Analyzing the 16S rRNA gene from the Np-SAG 69.79 genome and comparing it with sequences from other *Nostoc* strains revealed a unique phylogenetic placement for Np-SAG 69.79 within a distinct clade, including strains like *N. punctiforme Jbr02.* Additionally, a separate phylogenetic analysis on 14 single-copy orthologs across eight MRE genomes showed *Gp*-MRE closely aligning with the MRE-II cluster, supporting a shared evolutionary path within the *Diversisporales* order.

3.2.3 Genomic features of Endosymbionts in G. pyriformis

Gene analysis of Np-SAG 69.79 revealed 41.95% core genes shared with other *Nostoc* species, highlighting significant functional similarities, especially in “Energy Production & Conversion” and “Signal Transduction Mechanisms,” and unique expansions in genes related to “Mobilome” and “Replication & Repair.” Conversely, MRE genomes show a divergent gene set, with only 3.3% core genes shared across relatives, emphasizing their evolutionary divergence and unique genetic features, including distinct enzymes and absence of secreted proteins in *Gp*MRE.

3.3 Further Implications and Future Directions

3.3.1 Implications of AMF's reliance on endobacteria

Through the k-mer based binning of Illumina reads, efforts have been successful in assembling, annotating, and comparing the genomes of two *Geosiphon* endosymbionts; *Np*-SAG 69.79 and *Gp*-MRE. This process led to the uncovering of unique genomic features that distinguish these endosymbionts from their free-living counterparts. *Np*-SAG 69.79, despite sharing genome size and GC content with close relatives, displayed an unusually high load of insertion sequences (IS) domains, surpassing any known strain in its genus with sequenced genomes. This abundance of mobile genetic elements (MGEs) is thought to be a consequence of gene losses in the nutrient-rich environment of its host, leading to less control over the proliferation of these elements. Such expansions in MGEs might increase host dependency and reduce virulence over time, fostering long-term coevolution with *G. pyriformis*. Additionally, *Np*-SAG 69.79 possesses a unique gene for catechol 1,2-dioxygenase, suggesting its capability to degrade xenobiotics like benzoate, a potential supplemental carbon and energy source, or act as a detoxifier protecting its host from environmental stressors. This aspect requires further research to confirm its significance in the organism's metabolism.

Gp-MRE shows extreme sequence divergence, leading to streamlined gene and metabolic content. This divergence has resulted in extensive rearrangements, making direct comparisons with MRE relatives challenging. Unique genes in *Gp*-MRE, like DNA (cytosine-5)-methyltransferase 1 and ATP-dependent Lon Protease—a key enzyme involved in protein quality control that degrades damaged or misfolded proteins using energy from ATP, hinting at roles in epigenetic regulation, protein quality control, and stress response. These functions are essential for the symbiont's survival and symbiosis efficacy. The Lon Protease, specifically, is

vital for maintaining cellular protein homeostasis and responding to stress, and it is possible that its presence supports the symbiont's adaptation and resilience in the symbiotic environment. Furthermore, the identification of strain-specific superoxide dismutase in other MREs suggests a protective role against reactive oxygen species, highlighting MREs' adaptability and response capabilities in dynamic environments. These findings not only elucidate the complex genomic landscape of these endosymbionts but also underscore their critical roles in their symbiotic relationship with an AMF species.

3.3.2 Future Directions

Exploring the environmental impact of *Geosiphon* endosymbionts, for example through RNAseq analysis, offers a valuable approach to further identify their ecological roles. RNAseq can reveal gene expression patterns under varying environmental conditions, providing insights into how these symbionts and their specific genes respond to and influence their surroundings. This could unveil their adaptation mechanisms, interaction dynamics with the host and the environment, and contributions to ecosystem functioning, ultimately enhancing or confirming our knowledge of dynamics at play in this symbiotic relationship.

The functional analysis of unique genes in *Np*-SAG 69.79 and *Gp*-MRE presents fertile ground for future research. By employing techniques like gene knockout or overexpression, scientists can now investigate the specific roles and regulatory mechanisms of these unique genes. For example, the study of catechol 1,2-dioxygenase in Np-SAG 69.79 could elucidate its potential role in xenobiotic degradation and environmental detoxification. Similarly, analyzing the function of the DNA (cytosine-5)-methyltransferase 1 in *Gp*-MRE can provide insights into epigenetic regulation within these symbionts. These analyses not only contribute to our understanding of symbiont biology but also reveal how these unique genetic traits enhance

symbiont survival and symbiotic efficiency. Such studies could also pave the way for biotechnological applications, leveraging these genes for environmental or agricultural purposes.

The future adoption of metagenomic binning processes in studying endosymbionts within other AMF species will be crucial for expanding our understanding of these complex symbiotic relationships. By applying metagenomic binning to a wider range of AMF species, researchers can now continue this work to conduct comparative genomic studies to understand landscape of endosymbionts. In fact, this binning approach has demonstrated its effectiveness in extracting reliable genomic information from multiple species across various taxa. This knowledge is invaluable in elucidating the specific roles and contributions of different endosymbionts in their symbiotic partnerships. Furthermore, understanding these relationships and functions at a genomic level could pave the way for leveraging these endosymbionts in agricultural applications, such as enhancing crop growth, improving soil health, and increasing resistance to environmental stresses. This approach not only has the potential to revolutionize our understanding of AMF-endosymbiont interactions but also to contribute significantly to sustainable agricultural practices.

3.4 Concluding Remarks

In this study, genome content and origin of two endosymbionts in *G. pyriformis* bladders were uncovered, illustrating the effectiveness of binning approaches in acquiring genomic data from both AMF species and their bacterial endosymbionts. This research paves the way for future studies aimed at extracting similar information from other non-axenically cultivable organisms and their symbionts. A significant discovery is the expansion of mobile elements in the

cyanobacterium *Np*-SAG 69.79, mirroring patterns seen in other symbionts. This expansion suggests that endosymbionts follow similar mechanisms to establish and maintaining a stable association within the host's bladders. Additionally, our findings support the idea that MRE are monophyletic, yet they have undergone substantial genome shuffling. This has led to a scarcity of shared genes and a completely altered genome structure among these endosymbionts. Nonetheless, these MRE possess specific proteins, which may be crucial for mitigating stress within the shared cytoplasm. Our study also further demonstrates the value of k-mer binning approaches in retrieving genome and biological data, not just from challenging subjects like pot-cultured and *Nostoc*-carrying Glomeromycotina, but also from their endobacterial symbionts. This methodology offers a robust tool for unraveling complex biological relationships and genomic structures in symbiotic organisms.

References

Al-Karaki, G. N. (1998). Benefit, cost and water-use efficiency of arbuscular mycorrhizal durum wheat grown under drought stress. *Mycorrhiza*, *8*(1), 41–45.

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

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410.

Amaya-Carpio, L., Davies, F. T., Fox, T., & He, C. (2009). Arbuscular mycorrhizal fungi and organic fertilizer influence photosynthesis, root phosphatase activity, nutrition, and growth of Ipomoea carnea ssp. fistulosa. *Photosynthetica*, *47*(1), 1–10.

Amses, K., Desiró, A., Bryson, A., Grigoriev, I., Mondo, S., Lipzen, A., LaButti, K., Riley, R., Singan, V., Salazar-Hamm, P., King, J., Ballou, E., Pawlowska, T., Adeleke, R., Bonito, G., & Uehling, J. (2023). Convergent reductive evolution and host adaptation in Mycoavidus bacterial endosymbionts of Mortierellaceae fungi. *Fungal Genetics and Biology: FG & B*, *169*(103838), 103838.

Andrews, S. (2023). *FastQC* [Java]. https://github.com/s-andrews/FastQC

Asmelash, F., Bekele, T., & Birhane, E. (2016). The Potential Role of Arbuscular Mycorrhizal Fungi in the Restoration of Degraded Lands. *Frontiers in Microbiology*, *7*, 1095.

Bago, B., & Bécard, G. (2002). Bases of the obligate biotrophy of arbuscular mycorrhizal fungi. In S. Gianinazzi, H. Schüepp, J. M. Barea, & K. Haselwandter (Eds.), *Mycorrhizal Technology in Agriculture: From Genes to Bioproducts* (pp. 33–48). Birkhäuser Basel.

Bagyaraj, D. J., Sharma, M. P., & Maiti, D. (2015). Phosphorus nutrition of crops through arbuscular mycorrhizal fungi. *Current Science*, *108*(7), 1288–1293.

Begum, N., Qin, C., Ahanger, M. A., Raza, S., Khan, M. I., Ashraf, M., Ahmed, N., & Zhang, L. (2019). Role of Arbuscular Mycorrhizal Fungi in Plant Growth Regulation: Implications in Abiotic Stress Tolerance. *Frontiers in Plant Science*, *10*, 1068.

Beslemes, D., Tigka, E., Roussis, I., Kakabouki, I., Mavroeidis, A., & Vlachostergios, D. (2023). Effect of Arbuscular Mycorrhizal Fungi on Nitrogen and Phosphorus Uptake Efficiency and Crop Productivity of Two-Rowed Barley under Different Crop Production Systems. *Plants*, *12*(9). https://doi.org/10.3390/plants12091908

Bi, Y., Qiu, L., Zhakypbek, Y., Jiang, B., Cai, Y., & Sun, H. (2018). Combination of plastic film mulching and AMF inoculation promotes maize growth, yield and water use efficiency in the semiarid region of Northwest China. *Agricultural Water Management*, *201*, 278–286.

Bianciotto, V., & Bonfante, P. (1992). Quantification of the nuclear DNA content of two arbuscular mycorrhizal fungi. *Mycological Research*, *96*(12), 1071–1076.

Bonfante, P., & Desiro, A. (2017). Who lives in a fungus[quest] The diversity, origins and functions of fungal endobacteria living in Mucoromycota. *The ISME Journal*. https://doi.org/10.1038/ismej.2017.21

Bonfante, P., & Desirò, A. (2017). Who lives in a fungus? The diversity, origins and functions of fungal endobacteria living in Mucoromycota. *The ISME Journal*, *11*(8), 1727–1735.

Brown, C. L., Mullet, J., Hindi, F., Stoll, J. E., Gupta, S., Choi, M., Keenum, I., Vikesland, P., Pruden, A., & Zhang, L. (2022). mobileOG-db: a Manually Curated Database of Protein Families Mediating the Life Cycle of Bacterial Mobile Genetic Elements. *Applied and Environmental Microbiology*, *88*(18), e0099122.

Buchfink, B., Reuter, K., & Drost, H.-G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods*, *18*(4), 366–368.

Chen, E. C. H., Morin, E., Beaudet, D., Noel, J., Yildirir, G., Ndikumana, S., Charron, P., St-Onge, C., Giorgi, J., Krüger, M., Marton, T., Ropars, J., Grigoriev, I. V., Hainaut, M., Henrissat, B., Roux, C., Martin, F., & Corradi, N. (2018). High intraspecific genome diversity in the model arbuscular mycorrhizal symbiont Rhizophagus irregularis. *The New Phytologist*, *220*(4), 1161–1171.

Chen, E. C., Mathieu, S., Sedzielewska-Toro, K., Peart, M., Pelin, A., Ndikumana, S., Ropars, J., Dreissig, S., Fuchs, J., Brachmann, A., & Corradi, N. (2019). Correction: Single nucleus sequencing reveals evidence of inter-nucleus recombination in arbuscular mycorrhizal fungi. *ELife*, *8*. https://doi.org/10.7554/eLife.46860

Cooper, J., Lombardi, R., Boardman, D., & Carliell-Marquet, C. (2011). The future distribution and production of global phosphate rock reserves. *Resources, Conservation and Recycling*, *57*, 78–86.

Corradi, N., & Bonfante, P. (2012). The Arbuscular Mycorrhizal Symbiosis: Origin and Evolution of a Beneficial Plant Infection. *PLoS Pathogens*, *8*(4), e1002600.

Dallaire, A., & Paszkowski, U. (2023). Genomes of Arbuscular Mycorrhizal Fungi. In B. Scott & C. Mesarich (Eds.), *Plant Relationships: Fungal-Plant Interactions* (pp. 67–81). Springer International Publishing.

der Heijden, M. G. A., Klironomos, J. N., Ursic, M., Moutoglis, P., Streitwolf-Engel, R., Boller, T., Wiemken, A., & Sanders, I. R. (1998). Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. *Nature*, *396*(6706), 69–72.

Desirò, A., Hao, Z., Liber, J. A., Benucci, G. M. N., Lowry, D., Roberson, R., & Bonito, G. (2018). Mycoplasma-related endobacteria within Mortierellomycotina fungi: diversity, distribution and functional insights into their lifestyle. *The ISME Journal*, *12*(7), 1743–1757.

Dowarah, B., Gill, S. S., & Agarwala, N. (2022). Arbuscular Mycorrhizal Fungi in Conferring Tolerance to Biotic Stresses in Plants. *Journal of Plant Growth Regulation*, *41*(4), 1429–1444.

Durrant, M. G., Li, M. M., Siranosian, B. A., Montgomery, S. B., & Bhatt, A. S. (2020). A Bioinformatic Analysis of Integrative Mobile Genetic Elements Highlights Their Role in Bacterial Adaptation. *Cell Host & Microbe*, *28*(5), 767.

Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E., & Relman, D. A. (2005). Diversity of the human intestinal microbial flora. *Science (New York, N.Y.)*, *308*(5728), 1635–1638.

Fayad, N., Kallassy Awad, M., & Mahillon, J. (2020). IS982 and kin: new insights into an old IS family. *Mobile DNA*, *11*, 24.

Fidor, A., Grabski, M., Gawor, J., Gromadka, R., Węgrzyn, G., & Mazur-Marzec, H. (2020). Nostoc edaphicum CCNP1411 from the Baltic Sea-A new producer of nostocyclopeptides. *Marine Drugs*, *18*(9), 442.

Fidor, A., Konkel, R., & Mazur-Marzec, H. (2019). Bioactive peptides produced by Cyanobacteria of the genus Nostoc: A review. *Marine Drugs*, *17*(10), 561.

Gagunashvili, A. N., & Andrésson, Ó. S. (2018). Distinctive characters of Nostoc genomes in cyanolichens. *BMC Genomics*, *19*(1). https://doi.org/10.1186/s12864-018-4743-5

Galperin, M. Y., Wolf, Y. I., Makarova, K. S., Vera Alvarez, R., Landsman, D., & Koonin, E. V. (2021). COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Research*, *49*(D1), D274–D281.

Gao, K. (1998). Chinese studies on the edible blue-green alga, Nostoc flagelliforme: A review. *J. Appl. Phycol*, *10*, 37–49.

Gehrig, H., Schüßler, A., & Kluge, M. (1996). Geosiphon pyriforme, a fungus forming endocytobiosis with Nostoc (cyanobacteria), is an ancestral member of the glomales: Evidence by SSU rRNA analysis. *Journal of Molecular Evolution*. https://doi.org/10.1007/BF02352301

Gianinazzi-Pearson, V. (1996). Plant Cell Responses to Arbuscular Mycorrhizal Fungi: Getting to the Roots of the Symbiosis. *The Plant Cell*, *8*(10), 1871–1883.

Gianinazzi-pearson, V., Dumas-gaudot, E., Gollotte, A., Alaoui, A. T.-, & Gianinazzi, S. (1996). Cellular and molecular defence-related root responses to invasion by arbuscular mycorrhizal fungi. *The New Phytologist*, *133*(1), 45–57.

GitHub - clb21565/mobileOG-db: code repo for mobileOG-db. (n.d.). Retrieved June 27, 2023, from https://github.com/clb21565/mobileOG-db/tree/main

GitHub - olsonanl/FastOrtho. (n.d.). Retrieved June 27, 2023, from https://github.com/olsonanl/FastOrtho

Grant, J. R., Enns, E., Marinier, E., Mandal, A., Herman, E. K., Chen, C.-Y., Graham, M., Van Domselaar, G., & Stothard, P. (2023). Proksee: in-depth characterization and visualization of bacterial genomes. *Nucleic Acids Research*, *51*(W1), W484–W492.

Guo, X., Li, M., Jiang, S., Yang, L., Guo, S., Xing, L., & Wang, T. (2023). Arbuscular mycorrhizal fungi inoculation exerts weak effects on species- and community-level growth traits for invading or native plants under nitrogen deposition. *Frontiers in Ecology and Evolution*, *11*. https://doi.org/10.3389/fevo.2023.1152213

Halary, S., Malik, S.-B., Lildhar, L., Slamovits, C. H., Hijri, M., & Corradi, N. (2011). Conserved Meiotic Machinery in Glomus spp., a Putatively Ancient Asexual Fungal Lineage. *Genome Biology and Evolution*, *3*(0), 950–958.

Hallgren, J., Tsirigos, K. D., Pedersen, M. D., Armenteros, J. J. A., Marcatili, P., Nielsen, H., Krogh, A., & Winther, O. (2022). DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. In *bioRxiv* (p. 2022.04.08.487609). https://doi.org/10.1101/2022.04.08.487609

Hildebrandt, U., Regvar, M., & Bothe, H. (2007). Arbuscular mycorrhiza and heavy metal tolerance. *Phytochemistry*, *68*(1), 139–146.

Hilton, S. K., Castro-Nallar, E., Pérez-Losada, M., Toma, I., McCaffrey, T. A., Hoffman, E. P., Siegel, M. O., Simon, G. L., Johnson, W. E., & Crandall, K. A. (2016). Metataxonomic and Metagenomic Approaches vs. Culture-Based Techniques for Clinical Pathology. *Frontiers in Microbiology*, *7*, 484.

Hu, T., Chitnis, N., Monos, D., & Dinh, A. (2021). Next-generation sequencing technologies: An overview. *Human Immunology*, *82*(11), 801–811.

Huo, D., Li, H., Cai, F., Guo, X., Qiao, Z., Wang, W., Yu, G., & Li, R. (2021). Genome evolution of filamentous Cyanobacterium Nostoc species: From facultative symbiosis to free living. *Microorganisms*, *9*(10), 2015.

Index of /v4/data/lineages/. (n.d.). Retrieved October 1, 2023, from https://busco-data.ezlab.org/v4/data/lineages/

Jia, T., Zhang, Y., Yao, Y., Wang, Y., Liang, X., Zheng, M., Zhao, L., & Chai, B. (2023). Effects of AMF inoculation on the eco-physiological characteristics of Imperata cylindrica under differing soil nitrogen conditions. *Frontiers in Plant Science*, *14*, 1134995.

Jing, S., Li, Y., Zhu, L., Su, J., Yang, T., Liu, B., Ma, B., Ma, F., Li, M., & Zhang, M. (2022). Transcriptomics and metabolomics reveal effect of arbuscular mycorrhizal fungi on growth and development of apple plants. *Frontiers in Plant Science*, *13*, 1052464.

Johnson, H., King, S., Banack, S., Webster, C., Callanaupa, W., & Cox, P. (2008). Nostoc commune) used as a dietary item in the Peruvian highlands produce the neurotoxic amino acid BMAA. *J Ethnopharmacol*, *118*, 159–165.

Kameoka, H., Maeda, T., Okuma, N., & Kawaguchi, M. (2019). Structure-Specific Regulation of Nutrient Transport and Metabolism in Arbuscular Mycorrhizal Fungi. *Plant & Cell Physiology*, *60*(10), 2272–2281.

Kanehisa, M., Sato, Y., & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology*, *428*(4), 726–731.

Kaur, J., Chavana, J., Soti, P., Racelis, A., & Kariyat, R. (2020). Arbuscular mycorrhizal fungi (AMF) influences growth and insect community dynamics in Sorghum-sudangrass (Sorghum x drummondii). *Arthropod-Plant Interactions*, *14*(3), 301–315.

Keymer, A., Pimprikar, P., Wewer, V., Huber, C., Brands, M., Bucerius, S. L., Delaux, P.-M., Klingl, V., Röpenack-Lahaye, E. von, Wang, T. L., Eisenreich, W., Dörmann, P., Parniske, M., & Gutjahr, C. (2017). Lipid transfer from plants to arbuscular mycorrhiza fungi. *ELife*, *6*. https://doi.org/10.7554/eLife.29107

Kluge, M. (2002a). A fungus eats a Cyanobacterium: The story of the geosiphon pyriformis endocyanosis. *Biology and Environment. Proceedings of the Royal Irish Academy. Section B. Dublin*, *102B*(1), 11–14.

Kluge, M. (2002b). A fungus eats a cyanobacterium: The story of the Geosiphon pyriformis endocyanosis. *Biology and Environment*. https://doi.org/10.3318/BIOE.2002.102.1.11

Koch, A. M., Croll, D., & Sanders, I. R. (2006). Genetic variability in a population of arbuscular mycorrhizal fungi causes variation in plant growth. *Ecology Letters*, *9*(2), 103–110.

Kokkoris, V., Li, Y., Hamel, C., Hanson, K., & Hart, M. (2019). Site specificity in establishment of a commercial arbuscular mycorrhizal fungal inoculant. *The Science of the Total Environment*, *660*, 1135–1143.

Kokkoris, V., Stefani, F., Dalpé, Y., Dettman, J., & Corradi, N. (2020). Nuclear Dynamics in the Arbuscular Mycorrhizal Fungi. *Trends in Plant Science*, *25*(8), 765–778.

Krebs, H. A., Wiggins, D., Stubbs, M., Sols, A., & Bedoya, F. (1983). Studies on the mechanism of the antifungal action of benzoate. *Biochemical Journal*, *214*(3), 657–663.

Kunin Victor, Copeland Alex, Lapidus Alla, Mavromatis Konstantinos, & Hugenholtz Philip. (2008). A Bioinformatician's Guide to Metagenomics. *Microbiology and Molecular Biology Reviews: MMBR*, *72*(4), 557–578.

Ley, R. E., Peterson, D. A., & Gordon, J. I. (2006). Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell*, *124*(4), 837–848.

Li, Z., & Guo, M. (2018). Healthy efficacy of Nostoc commune Vaucher. *Oncotarget*, *9*(18), 14669–14679.

Liao, D., Wang, S., Cui, M., Liu, J., Chen, A., & Xu, G. (2018). Phytohormones Regulate the

Development of Arbuscular Mycorrhizal Symbiosis. *International Journal of Molecular Sciences*, *19*(10). https://doi.org/10.3390/ijms19103146

Llamas, A., Leon-Miranda, E., & Tejada-Jimenez, M. (2023). Microalgal and nitrogen-fixing bacterial consortia: From interaction to biotechnological potential. *Plants*, *12*(13). https://doi.org/10.3390/plants12132476

Longley, R., Robinson, A., Liber, J. A., Bryson, A. E., Morales, D. P., LaButti, K., Riley, R., Mondo, S. J., Kuo, A., Yoshinaga, Y., Daum, C., Barry, K., Grigoriev, I. V., Desirò, A., Chain, P. S. G., & Bonito, G. (2023). Comparative genomics of Mollicutes-related endobacteria supports a late invasion into Mucoromycota fungi. *Communications Biology*, *6*(1), 948.

Malar C, M., Krüger, M., Krüger, C., Wang, Y., Stajich, J. E., Keller, J., Chen, E. C. H., Yildirir, G., Villeneuve-Laroche, M., Roux, C., Delaux, P.-M., & Corradi, N. (2021a). The genome of Geosiphon pyriformis reveals ancestral traits linked to the emergence of the arbuscular mycorrhizal symbiosis. *Current Biology: CB*, *31*(7), 1578–1580.

Malar C, M., Krüger, M., Krüger, C., Wang, Y., Stajich, J. E., Keller, J., Chen, E. C. H., Yildirir, G., Villeneuve-Laroche, M., Roux, C., Delaux, P.-M., & Corradi, N. (2021b). The genome of Geosiphon pyriformis reveals ancestral traits linked to the emergence of the arbuscular mycorrhizal symbiosis. *Current Biology: CB*, *31*(7), 1570-1577.e4.

Malar, C. M., Krüger, M., Krüger, C., Wang, Y., Stajich, J. E., Keller, J., Chen, E., Yildirir, G., Villeneuve-Laroche, M., & Roux, C. (2021). The genome of Geosiphon pyriformis reveals ancestral traits linked to the emergence of the arbuscular mycorrhizal symbiosis. *Current Biology*, *31*, 1570–1577.

Malar C, M., Wang, Y., Stajich, J. E., Kokkoris, V., Villeneuve-Laroche, M., Yildirir, G., & Corradi, N. (2022). Early branching arbuscular mycorrhizal fungus Paraglomus occultum carries a small and repeat-poor genome compared to relatives in the Glomeromycotina. *Microbial Genomics*, *8*(4). https://doi.org/10.1099/mgen.0.000810

marbl/MUMmer3. (2023). [C++]. MarBL. https://github.com/marbl/MUMmer3

Martin, F. M., Uroz, S., & Barker, D. G. (2017). Ancestral alliances: Plant mutualistic symbioses with fungi and bacteria. *Science*, *356*(6340), eaad4501.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12.

McCutcheon, J. P. (2021). The genomics and cell biology of host-beneficial intracellular infections. *Annual Review of Cell and Developmental Biology*, *37*(1), 115–142.

Mondo, S. J., Lastovetsky, O. A., Gaspar, M. L., Schwardt, N. H., Barber, C. C., Riley, R., Sun, H., Grigoriev, I. V., & Pawlowska, T. E. (2017). Bacterial endosymbionts influence host sexuality and reveal reproductive genes of early divergent fungi. *Nature Communications*, *8*(1), 1843.

Naito, M., Morton, J. B., & Pawlowska, T. E. (2015). Minimal genomes of mycoplasma-related endobacteria are plastic and contain host-derived genes for sustained life within Glomeromycota. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(25), 7791–7796.

Naumann, M., Schüssler, A., & Bonfante, P. (2010). The obligate endobacteria of arbuscular mycorrhizal fungi are ancient heritable components related to the Mollicutes. *The ISME Journal*, *4*(7), 862–871.

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msu300

Niveshika, Verma, E., Mishra, A. K., Singh, A. K., & Singh, V. K. (2016). Structural elucidation and molecular docking of a novel antibiotic compound from Cyanobacterium Nostoc sp. MGL001. *Frontiers in Microbiology*, *7*, 1899.

Nosheen, S., Ajmal, I., & Song, Y. (2021). Microbes as biofertilizers, a potential approach for sustainable crop production. *Sustainability*, *13*(4), 1868.

Nowruzi, B., Haghighat, S., Fahimi, H., & Mohammadi, E. (2018). Nostoc cyanobacteria species: a new and rich source of novel bioactive compounds with pharmaceutical potential. *Journal of Pharmaceutical Health Services Research: An Official Journal of the Royal Pharmaceutical Society of Great Britain*, *9*(1), 5–12.

Oehl, F., Sieverding, E., Mäder, P., Dubois, D., Ineichen, K., Boller, T., & Wiemken, A. (2004). Impact of long-term conventional and organic farming on the diversity of arbuscular mycorrhizal fungi. *Oecologia*, *138*(4), 574–583.

Partida-Martinez, L. P., & Hertweck, C. (2005). Pathogenic fungus harbours endosymbiotic bacteria for toxin production. *Nature*, *437*(7060), 884–888.

Paszkowski, U. (2006). A journey through signaling in arbuscular mycorrhizal symbioses 2006. *The New Phytologist*, *172*(1), 35–46.

Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, *8*(10), 785–786.

Pirozynski, K. A., & Malloch, D. W. (1975). The origin of land plants: a matter of mycotrophism. *Bio Systems*, *6*(3), 153–164.

Ponger, L., & Li, W.-H. (2005). Evolutionary diversification of DNA methyltransferases in eukaryotic genomes. *Molecular Biology and Evolution*, *22*(4), 1119–1128.

Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A., & Korobeynikov, A. (2020). Using SPAdes De Novo Assembler. *Current Protocols in Bioinformatics*, *70*(1), e102.

Redecker, D., Kodner, R., & Graham, L. E. (2000). Glomalean fungi from the Ordovician.

Science, *289*(5486), 1920–1921.

Remy, W., Taylor, T. N., Hass, H., & Kerp, H. (1994). Four hundred-million-year-old vesicular arbuscular mycorrhizae. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(25), 11841–11843.

Rich, M. K., Vigneron, N., Libourel, C., Keller, J., Xue, L., Hajheidari, M., Radhakrishnan, G. V., Le Ru, A., Diop, S. I., Potente, G., Conti, E., Duijsings, D., Batut, A., Le Faouder, P., Kodama, K., Kyozuka, J., Sallet, E., Bécard, G., Rodriguez-Franco, M., … Delaux, P.-M. (2021). Lipid exchanges drove the evolution of mutualism during plant terrestrialization. *Science (New York, N.Y.)*, *372*(6544), 864–868.

Riley, R., & Corradi, N. (2013). Searching for clues of sexual reproduction in the genomes of arbuscular mycorrhizal fungi. *Fungal Ecology*, *6*(1), 44–49.

Robinson, J. T., Thorvaldsdottir, H., Turner, D., & Mesirov, J. P. (2023). igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). *Bioinformatics*, *39*(1), btac830.

Roney, B. R., Renhui, L., Banack, S. A., Murch, S., Honegger, R., & Cox, P. A. (2009). Consumption of fa cai Nostoc soup: a potential for BMAA exposure from Nostoc cyanobacteria in China? *Amyotrophic Lateral Sclerosis: Official Publication of the World Federation of Neurology Research Group on Motor Neuron Diseases*, *10 Suppl 2*(sup2), 44–49.

Ropars, J., Toro, K. S., Noel, J., Pelin, A., Charron, P., Farinelli, L., Marton, T., Krüger, M., Fuchs, J., Brachmann, A., & Corradi, N. (2016). Evidence for the sexual origin of heterokaryosis in arbuscular mycorrhizal fungi. *Nature Microbiology*, *1*(6), 16033.

Roy-Bolduc, A., & Hijri, M. (2010). The Use of Mycorrhizae to Enhance Phosphorus Uptake: A Way Out The Phosphorus Crisis. *Roy-Bolduc and Hijri J Biofertil Biopestici*, *2*, 1.

Salvioli, A., Ghignone, S., Novero, M., Navazio, L., Venice, F., Bagnaresi, P., & Bonfante, P. (2016). Symbiosis with an endobacterium increases the fitness of a mycorrhizal fungus, raising its bioenergetic potential. *The ISME Journal*, *10*(1), 130–144.

Sand-Jensen, K. (2014). Ecophysiology of gelatinous Nostoc colonies: unprecedented slow growth and survival in resource-poor and harsh environments. *Annals of Botany*, *114*(1), 17–33.

Savci, S. (2012). Investigation of Effect of Chemical Fertilizers on Environment. *APCBEE Procedia*, *1*, 287–292.

Schüßler, A., & Kluge, M. (2001). Geosiphon pyriforme, an endocytosymbiosis between fungus and Cyanobacteria, and its meaning as a model system for arbuscular mycorrhizal research. In *Fungal Associations* (pp. 151–161). Springer Berlin Heidelberg.

Schüßler, Arthur, & Wolf, E. (2005). Geosiphon pyriformis—a Glomeromycotan Soil Fungus Forming Endosymbiosis with Cyanobacteria. In *In Vitro Culture of Mycorrhizas*.

Schwengers, O., Jelonek, L., Dieckmann, M. A., Beyvers, S., Blom, J., & Goesmann, A. (2021). Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microbial Genomics*, *7*(11). https://doi.org/10.1099/mgen.0.000685

Sędzielewska, K. A., Fuchs, J., Temsch, E. M., Baronian, K., Watzke, R., & Kunze, G. (2011). Estimation of the Glomus intraradices nuclear DNA content. *The New Phytologist*, *192*(4), 794–797.

Seemann, T. (2023). *Barrnap* [Perl]. https://github.com/tseemann/barrnap

Shen, W., Le, S., Li, Y., & Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLOS ONE*, *11*(10), e0163962.

Siguier, P., Gourbeyre, E., Varani, A., Ton-Hoang, B., & Chandler, M. (2015). Everyman's Guide to Bacterial Insertion Sequences. *Microbiology Spectrum*, *3*(2), MDNA3-0030–2014.

Singh, D. P., Prabha, R., Gupta, V. K., & Verma, M. K. (2018). Metatranscriptome Analysis Deciphers Multifunctional Genes and Enzymes Linked With the Degradation of Aromatic Compounds and Pesticides in the Wheat Rhizosphere. *Frontiers in Microbiology*, *9*, 1331.

Singh, J. S., Kumar, A., Rai, A. N., & Singh, D. P. (2016). Cyanobacteria: A precious bio-resource in agriculture, ecosystem, and environmental sustainability. *Frontiers in Microbiology*, *7*, 529.

Smith, S. E., & Read, D. J. (2010). *Mycorrhizal Symbiosis* (3rd ed., pp. 1–88). Academic Press.

Spatafora, J. W., Chang, Y., Benny, G. L., Lazarus, K., Smith, M. E., Berbee, M. L., Bonito, G., Corradi, N., Grigoriev, I., Gryganskyi, A., & Others. (2016). A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia*, *108*(5), 1028–1046.

Sperschneider, J., Yildirir, G., Rizzi, Y. S., Malar C, M., Mayrand Nicol, A., Sorwar, E., Villeneuve-Laroche, M., Chen, E. C. H., Iwasaki, W., Brauer, E. K., Bosnich, W., Gutjahr, C., & Corradi, N. (2023). Arbuscular mycorrhizal fungi heterokaryons have two nuclear populations with distinct roles in host-plant interactions. *Nature Microbiology*. https://doi.org/10.1038/s41564-023-01495-8

Stein, J. L., Marsh, T. L., Wu, K. Y., Shizuya, H., & DeLong, E. F. (1996). Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *Journal of Bacteriology*, *178*(3), 591–599.

Strunk, N., & Engesser, K.-H. (2013). Degradation of fluorobenzene and its central metabolites 3-fluorocatechol and 2-fluoromuconate by Burkholderia fungorum FLU100. *Applied Microbiology and Biotechnology*, *97*(12), 5605–5614.

Sun, X., Chen, W., Ivanov, S., MacLean, A. M., Wight, H., Ramaraj, T., Mudge, J., Harrison, M. J., & Fei, Z. (2019). Genome and evolution of the arbuscular mycorrhizal fungus Diversispora epigaea (formerly Glomus versiforme) and its bacterial endosymbionts. *The New Phytologist*. https://doi.org/10.1111/nph.15472

Tilman, D. (1999). Global environmental impacts of agricultural expansion: the need for sustainable and efficient practices. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(11), 5995–6000.

Torres-Cortés, G., Ghignone, S., Bonfante, P., & Schüßler, A. (2015). Mosaic genome of endobacteria in arbuscular mycorrhizal fungi: Transkingdom gene transfer in an ancient mycoplasma-fungus association. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1501540112

Uehling, J. K., Salvioli, A., Amses, K. R., Partida-Martínez, L. P., Bonito, G., & Bonfante, P. (2023). *Bacterial Endosymbionts of Mucoromycota Fungi: Diversity and Function of their Interactions* (S. Pöggeler & T. James, Eds.; pp. 177–205). Springer International Publishing.

Valderrama, J. A., Durante-Rodríguez, G., Blázquez, B., García, J. L., Carmona, M., & Díaz, E. (2012). Bacterial degradation of benzoate. *The Journal of Biological Chemistry*, *287*(13), 10494–10508.

Van Melderen, L., & Aertsen, A. (2009). Regulation and quality control by Lon-dependent proteolysis. *Research in Microbiology*, *160*(9), 645–651.

Vandecraen, J., Chandler, M., Aertsen, A., & Van Houdt, R. (2017). The impact of insertion sequences on bacterial genome plasticity and adaptability. *Critical Reviews in Microbiology*, *43*(6), 709–730.

Venice, F., Chialva, M., Domingo, G., Novero, M., Carpentieri, A., Salvioli di Fossalunga, A., Ghignone, S., Amoresano, A., Vannini, C., Lanfranco, L., & Bonfante, P. (2021). Symbiotic responses of *Lotus japonicus* to two isogenic lines of a mycorrhizal fungus differing in the presence/absence of an endobacterium. *The Plant Journal: For Cell and Molecular Biology*, *108*(6), 1547–1564.

Wang, C.-L., You, S.-L., & Wang, S.-L. (2006). Purification and characterization of a novel catechol 1,2-dioxygenase from Pseudomonas aeruginosa with benzoic acid as a carbon source. *Process Biochemistry*, *41*(7), 1594–1601.

Wang, F. (2017). Arbuscular Mycorrhizas and Ecosystem Restoration. In *Arbuscular Mycorrhizas and Stress Tolerance of Plants* (pp. 245–292). Springer Singapore.

Wang, W., Shi, J., Xie, Q., Jiang, Y., Yu, N., & Wang, E. (2017). Nutrient Exchange and Regulation in Arbuscular Mycorrhizal Symbiosis. *Molecular Plant*, *10*(9), 1147–1158.

Wei, L., Vosátka, M., Cai, B., Ding, J., Lu, C., Xu, J., Yan, W., Li, Y., & Liu, C. (2019). The role of arbuscular mycorrhiza fungi in the decomposition of fresh residue and soil organic carbon: A mini-review. *Soil Science Society of America Journal. Soil Science Society of America*, *83*(3), 511–517.

Woese, C. R. (1987). Bacterial evolution. *Microbiological Reviews*, *51*(2), 221–271.

Xu, J., Liu, S., Song, S., Guo, H., Tang, J., Yong, J. W. H., Ma, Y., & Chen, X. (2018).

Arbuscular mycorrhizal fungi influence decomposition and the associated soil microbial community under different soil phosphorus availability. *Soil Biology & Biochemistry*, *120*, 181–190.

Zahra, Z., Choo, D. H., Lee, H., & Parveen, A. (2020). Cyanobacteria: Review of current potentials and applications. *Environments*, *7*(2), 13.

Zhang, L., Chen, F., Zeng, Z., Xu, M., Sun, F., Yang, L., Bi, X., Lin, Y., Gao, Y., Hao, H., Yi, W., Li, M., & Xie, Y. (2021). Advances in Metagenomics and Its Application in Environmental Microorganisms. *Frontiers in Microbiology*, *12*, 766364.

Zhou, W., Zhang, M., Tao, K., & Zhu, X. (2022). Effects of arbuscular mycorrhizal fungi and plant growth-promoting rhizobacteria on growth and reactive oxygen metabolism of tomato fruits under low saline conditions. *Biocell: Official Journal of the Sociedades Latinoamericanas de Microscopia Electronica ... et. Al*, *46*(12), 2575–2582.

Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L., & Yorke, J. A. (2013). The MaSuRCA genome assembler. *Bioinformatics* . https://doi.org/10.1093/bioinformatics/btt476