Université de Montréal

The evolution of RNA interference system, blue light sensing mechanism and circadian clock in *Rhizophagus irregularis* give insight on Arbuscular mycorrhizal symbiosis

by Soon-Jae Lee

Département de sciences biologiques
Faculté des Arts et des sciences

Thèse présentée à la Faculté des Arts et des sciences
en vue de l'obtention du grade de docteur
en sciences biologiques

August, 2018

Résumé

La symbiose mycorhizienne arbusculaire (MA) est une association formée par les racines des plantes et les champignons mycorhiziens arbusculaires (CMA). Ces champignons sont les plus anciens symbiotes des plantes et ils sont apparus il y a au moins 460 millions d'années avec l'émergence et l'évolution des plantes terrestres. Les CMA sont également les partenaires symbiotiques des plantes les plus répandus dans les écosystèmes et ils peuvent s'associer avec plus de 80% des espèces de plantes vasculaires. Les CMA appartiennent à une lignée fongique primitive dont la position phylogénétique est encore en débat. Les CMA sont des microorganismes biotrophes obligatoires qui dépendent entièrement du carbone provenant de la photosynthèse des plantes hôtes, autrement dit, les CMA ne peuvent assimiler le carbone qu'on association avec les racines des plantes. En échange, les CMA aident les plantes à absorber divers nutriments essentiels du sol, tels que le phosphore. En effet, les CMA absorbent les nutriments et ils les transportent à travers leurs hyphes jusqu'aux cellules des racines dans lesquelles ils forment une structure appelée arbuscule.

L'allocation des nutriments et les voies métaboliques interconnectées entre le champignon et l'hôte ont subi une pression sélective en tant que partenaires symbiotiques. En plus, les hyphes de des CMA agissent comme une niche écologique pour divers microbes du sol tels que les bactéries et les champignons, formant ainsi le pivot de la rhizosphère des racines. La symbiose MA est un élément essentiel pour comprendre la physiologie des plantes ainsi que l'écosystème. Malgré les rôles cruciaux des CMA dans les écosystèmes, leur génétique et leur évolution demeure méconnues. Le système d'interférence de l'ARN (ARNi), le mécanisme de détection de la lumière bleue et l'horloge circadienne sont des mécanismes importants qui sont impliqués dans la régulation de l'expression des gènes chez les champignons. Bien que son rôle reconnu dans la régulation des gènes et la traduction des protéines, en particulier dans la symbiose telle que celle des nématodes, le système ARNi n'a jamais été étudié chez les CMA. Pareil pour le cas du mécanisme de détection de la lumière bleue. Seules quelques études ont montré que la lumière bleue peut affecter la germination des spores et la croissance des hyphes des CMA, cependant son mécanisme n'a pas été décrit. Dans le cas de l'horloge circadienne, même si le rythme circadien est omniprésent chez les champignons et que le rythme diurne de la croissance

des hyphes a été reporté dans les CMA dans une étude sur le terrain, le mécanisme demeure méconnu. Le génome et le transcriptome du CMA modèle *Rhizophagus irregularis* isolat DAOM 197198, étaient publiquement disponibles et ils ont été exploité dans mon projet. L'objectif de ma thèse de doctorat visait donc à étudier l'évolution du système ARNi, du mécanisme de détection de la lumière bleue et de l'horloge circadienne dans le génome de *R. irregularis* à l'aide d'approches biologiques et bioinformatiques. Les objectifs spécifiques étaient de: 1) déterminer si le système ARNi est conservé dans le génome de *R. irregularis* et d'analyser les traits évolutifs de ses protéines; 2) décrire le mécanisme de détection de la lumière bleue dans le génome *R. irregularis ;* 3) étudier le mécanisme circadien fongique dans le génome de *R. irregularis*.

J'ai analysé les données génomiques et transcriptomiques pour rechercher les mécanismes conservés du système ARNi de *R. irregularis* et de certaines espèces de CMA qui lui sont apparentées. Deux phases du cycle de vie de *R. irregularis* (la phase de la germination des spores et la phase symbiotique avec des racines) ont été utilisées pour déterminer les profils d'expression des gènes en utilisant la PCR quantitative par transcriptase inverse (qPCR). J'ai identifié des traits évolutifs particuliers dans le système ARNi de *R. irregularis*, tels que le transfert de gènes horizontal (HGT) d'un gène important codant la protéine ribonucléase III, d'origine des cyanobactéries qui n'a jamais été observé chez aucun eucaryote. J'ai également trouvé et identifié un ancien mécanisme de détection de la lumière bleue corrélé à l'horloge circadienne. J'ai trouvé que le gène *frequency* est conservé dans le génome *R. irregularis* et que son expression est influencée par l'exposition à la lumière bleue. Ce qui est intéressant est que la protéine la plus importante de l'horloge circadienne (FRQ) n'a jamais été retrouvée dans d'autres lignées primitives fongiques, y compris chez Mucoromycotina, un sous-embranchement fongique considéré comme le plus proche des CMA.

Les résultats de mon projet de doctorat a significativement contribuer à la progression de nos connaissances sur les mécanismes importants qui régulent l'expression des gènes chez les CMA qui sont des partenaires symbiotiques des racines des plantes et les plus anciens et les plus répandus dans les écosystèmes. Mes résultats apportent également de nouvelles informations sur le transfert des gènes entre les cyanobactéries et les CMA, et ils ont élargi les connaissances de l'évolution du gène *frq* chez les champignons. De plus, la présence de gène *frq* dans le génome

de *R. irregularis* ouvre la voie à l'étude de la chronologique de la symbiose MA, qui peut être le modèle intéressant d'holobiontes des plantes.

Mots-clés : champignons mycorrhiziens arbusculaires(CMA), holobionte, évolution, symbiose, ARN d'interférence (ARNi), transfert horizontal des gènes, détection de la lumière, horloge circadianne , chronobiologie

Abstract

Arbuscular mycorrhizal (AM) symbiosis is formed by plant roots and arbuscular mycorrhizal fungi (AMF) which are the oldest symbiotic partners of plants and have evolved at least 460 million years ago with the emergence and evolution of land plants. AMF are also the most ubiquitous symbiotic partner of plants as they can colonize more than 80% of vascular plant species. AMF are an early diverged fungal lineage whose phylogenetic position is still under debate. AMF are obligatory plant root symbionts which depend on a source of carbon from the photosynthesis of host plants. In exchange, AMF help plants to absorb various essential soil nutrients, such as phosphorus and transfer these nutrients through their hyphae which have grown into and colonized plant root cells in which they form a structure called an arbuscule. The nutrient allocation and interlocked metabolic pathways between the fungus and the host underwent selective pressure as symbiotic partners. Moreover, AMF hyphae act as an ecological niche for various soil bacteria and other fungi, thus forming the backbone of the rhizospheric part of the plant. AM symbiosis is an essential element to understand plant physiology and ecosystem. Despite the crucial roles of AMF in ecosystems, their genetics and evolution are far from being understood. The RNA interference (RNAi) system, the blue light sensing mechanism and the circadian clock are important mechanisms which regulate expression of various genes in fungi.

Although it has an acknowledged role in gene regulation, especially with symbiosis as well as reflected selective pressures on core proteins in the cases of other symbiotic organisms such as nematode worms, the RNAi system has never been considered in AMF. The same is true for the case of blue light sensing mechanisms. Only a few studies showed that blue light can affect spore germination and hyphae growth of AMF, but the mechanism was not addressed. In the case of the circadian clock, even though circadian rhythms are ubiquitous in fungi and a diurnal rhythm of hyphae growth was reported in AMF during a field level study, the mechanism was unknown. Currently, the genome and transcriptome of the model AM fungus *Rhizophagus irregularis* isolate DAOM 197198, are publicly available and they were used in my studies. The objective of my Ph.D. project was therefore to study the evolution of the RNAi system, blue light sensing mechanisms and the circadian clock in *R. irregularis* genome using both

bioinformatic and molecular biological approaches. The specific objectives of my Ph.D. project were: 1) to investigate whether the RNAi system is conserved in the genome of *R. irregularis* and explore the evolutionary traits in its core proteins; 2) to describe the blue light sensing mechanism in the genome of *R. irregularis*; and 3) to search for a fungal circadian mechanism in the genome of *R. irregularis*.

I surveyed genomic and transcriptomic data to search for conserved elements of the RNAi system of *R. irregularis* and its relatives. Two life stages of the *R. irregularis* lifecycle (germination of spores without roots and established mycorrhizal symbiosis) were investigated for gene expressional profiles using reverse-transcriptase quantitative PCR (qPCR). I identified particular evolutionary traits in *R. irregularis* RNAi system, such as horizontal gene transfer (HGT) of its core gene coding ribonuclease III protein from autotrophic cyanobacteria, which has never been reported in any eukaryotes so far. I also found and identified an ancient mechanism of blue light sensing which is related to circadian clock. It was intriguing to find a conserved core gene (*frequency*) that responds to light exposure in the genome of this underground plant-root symbiont. At the same time, the circadian clock core component (FRQ) was not found in other basal fungal lineages including *Mucoromycotina*, a fungal subphylum which is considered as the closest relative of AMF.

The outcome of my Ph.D. project advanced our knowledge on important mechanisms which regulate the expression of various genes in the oldest and most ubiquitous symbiotic partner of plants. My results also provide new insight on the intimacy between cyanobacteria and AMF which resulted in a unique HGT in the RNAi system. It also expands the knowledge of evolution of the circadian *frq* gene in fungi. Furthermore, the presence of circadian clock and output genes in *R. irregularis* opens the door to the chronological study of AM symbiosis, which can be used as a model for the plant holobiont.

Keywords : arbuscular mycorrhizal fungi (AMF), holobiont, evolution, symbiosis, RNA interference (RNAi), horizontal gene transfer, light sensing, circadian clock, chronobiology

Table of contents

List of tables

List of figures

List of abbreviations

~: approximately

a.a.: amino acid

AM: arbuscular mycorrhiza

AMF: arbuscular mycorrhizal fungi

Bp: base pair(s)

cDNA: complementary deoxyribonucleic acid

DAOM: agriculture and agri-Food Canada national mycological herbarium

DNA: deoxyribonucleic acid

dNTP: deoxyribonucleoside triphosphate

e.g.: Latin: exempli gratia (English: for example)

i.e.: Latin: id est (English: that is)

in vitro: Latin: in glass

in vivo: Latin: within the living organism

µl: micro liter

µM: micro molar

mM: mili molar

ML: maximum likelihood

MM: minimal medium

mRNA: messenger RNA

MUCL: mycothèque de l'universite catholique de Louvain

NADH: nicotinamide adenine dinucleotide

ng: nano gram

NCBI: the national center for biotechnology information

NSERC: natural sciences and engineering research council (Canada)

Orf: open reading frames

PCR: polymerase chain reaction

qPCR: quantitative polymerase chain reaction

rDNA: ribosomal deoxyribonucleic acid

Ri T-DNA: transfer DNA transmission from *Agrobacterium* root-inducing plasmid into the genome of *Dacus carota*

rRNA: ribosomal ribonucleic acid

RNA: ribonucleic acid

RT-PCR: real-time polymerase chain reaction

sRNA: small ribonucleic acid

SSU: small subunit

Spp.: species

Taq: Thermus aquaticus

Tm: melting temperature

tRNA: transfer ribonucleic acid

UV: ultraviolet

v/v: volume/volume

w/v: weight/volume

This thesis is dedicated to my mother and father who have always mentally supported and encouraged me to do what I want to, the science. Thank you and I love you

Remerciements

I want to express my deep appreciation for the people who made this research possible. The first and foremost is my supervisor Dr. Mohamed Hijri who always encouraged and supported me to continue the challenging studies initiated by the speculations from the weird graduate student. Whenever I knocked his office, he always responded me with gentle smile and patience. I have learned not only the knowledges, but also the important attitude as a scientist, the patience. I would like to my very special gratitude to Dr. Marc St-Arnaud who always kindly shared his valuable knowledge of the science especially with arbuscular mycorrhizal symbiosis. I also have to express my very special appreciation to Dr. David Morse. I have learned the scientific knowledge as well as the enthusiasm as a scientist from him. Without kind supporting and sometimes sharp advice from Dr. Mohamed Hijri, Dr. Marc St-Arnaud and Dr. David Morse, I would have not been able to continue my research. I need to express my appreciation for Dr. Paul Harrison who always kindly advised me with his knowledge of bioinformatics. My appreciation should also go to my beloved parents, Young-Suk Kwon and Geun-Ho Lee and my brother Soon-Hyouk Lee.

I wish to express my sincere appreciation to formal postdoctoral fellows Dr. Yves Terrat and Dr. Sebastien Halary as well as formal lab member Dr. Denis Beaudet, Dr. Alice Roy-Bolduc, Dr. Ivan de la Providencia and Laurence Daubois for their kind support and valuable advice. I would like to thank to Yerim Heo for kind support in the illustrations and figures as well as the sharing of hard moments during my study. I also thank to Mengxuan Kong who not only kindly supported the experiments, but also shared the hard moments during the study. I would acknowledge Dr. Jacques Brodeur for the beneficial advices in my comprehensive exam.

For the last, but not least, I want to express my wholehearted thanks to all of my lovely friends including Raymond Archambault and Kit Leung for their motivation, encouragement and support.

Chapter 1 – General introduction

1.1. Arbuscular mycorrhizal fungi

Mycorrhiza is a symbiotic association between fungi and plants and it is one of the most intimate and ubiquitous symbioses found on earth (Smith and Read 2008). In fact, according to the statement of the BEG (European bank of Glomeromycota) in 1993, the majority of plants, strictly speaking, do not have roots – they have mycorrhiza (Schüßler, et al. 2007). The intimacy of the mycorrhizal symbiosis is directly reflected in its name, mycorrhiza (greek "mukés", meaning fungus, and "rhiza", meaning root**s**). Up to date, seven major groups of mycorrhizae have been characterized and classified as; (1) arbuscular mycorrhiza, (2) ectomycorrhiza, (3) ectendomycorrhiza (4) ericoid mycorrhiza (5) arbutoid mycorrhiza, (6) monotropoid mycorrhiza, and (7) orchid mycorrhiza (Smith and Read 2008). The different groups of mycorrhiza show diverged forms of symbiosis such as an intracellular structure (arbuscule in arbuscular mycorrhiza; Hartig net in ecto and ectendo mycorrhiza; coil-like structure in ericoid and arbutoid mycorrhizas; fungal peloton in orchid mycorrhiza; fungal peg in monotropoid mycorrhiza). The groups of mycorrhiza also differentiated in their favored symbiotic partners and ecosystems which determine their distribution (Smith and Read 2008).

Arbuscular mycorrhizal fungi (AMF) are the oldest and most important group of mycorrhiza and they establish symbioses over 80% of land plant species including major crops such as rice, wheat, maize and potato (Oehl, et al. 2011). AMF are considered as an early diverged fungal lineage assigned to phylum Glomeromycota (Smith and Read 2008). Fossil records and molecular clock analysis have revealed that AMF evolved to be the obligatory symbionts of

their plant partners since approximately 460 million years ago (Luc Simon 1993; Redecker, et al. 2000). AMF establish symbiosis with host plant root cells, where they form special structures called "arbuscules" in which nutrients are exchanged between partners (Bonfante and Genre 2010). AMF colonize host cells in the cortex of roots and elongate the network of hyphae along and beyond the rhizosphere. Through the hyphae, AMF absorb various nutrients from the soil (including phosphate and nitrogen), that are used as an exchange for carbon supplied by the host plant's photosynthesis (Nouri, et al. 2014). Moreover, not only helping plants to uptake essential nutrients from the soil, AMF also have functions of degrading large range of soil chemicals, giving tolerance for plants during drought, enhancing host plant immunity against various plant pathogens, and modulating the soil microbial community structure (Jansa, et al. 2013; Smith and Read 2008). Thus, the study of AMF is essential to understand the soil microbial community as well as the land ecosystem.

1.2. Importance of sRNA regulation system in symbiosis

The regulation of gene expression is one of the most crucial mechanisms in understanding the biological processes of an organism (Aravind, et al. 2012; Damiani, et al. 2012). The regulation of gene expression becomes more crucial and complex in the cases of organisms which live in symbiotic association (Aravind, et al. 2012; Damiani, et al. 2012; Hou and Lin 2009; Yigit, et al. 2006). The processing and expression of small non-coding RNAs (sRNAs) is among the crucial biological processes in organisms that form symbiosis (Formey, et al. 2016). A maintenance of symbiosis mediated by the sRNA processing mechanism has been previously reported in other symbiotic organisms, such as in the symbiosis between soybean and *Rhizobium*, or between maize and diazotrophic bacteria (Formey, et al. 2016; Thiebaut, et al. 2014; Yan, et

al. 2015). It was also revealed that mycorrhizal plants actively express and regulate symbiosis-related sRNAs in the establishment and development of AM symbiosis (Branscheid, et al. 2011).

1.3. RNAi system in fungi

RNA interference (RNAi) is the sRNA processing and regulation mechanism in eukaryotes. RNAi and gene silencing has been reported in a broad range of eukaryotic organisms, including Fungi (Dang, et al. 2011; Torres-Martinez and Ruiz-Vazquez 2016), where RNAi-related pathways play a role in various functions including genome surveillance, gene regulation and genome defense. In Fungi, the discovery of a phenomenon called "Quelling" in *Neurospora crassa* (Pickford, et al. 2002) initiated the study of the RNAi system. Quelling is post-transcriptional gene silencing guided by sRNAs resulting from the RNAi process. The principle of the fungal RNAi process is to form sRNAs which can act as regulators for cellular processes such as development, RNA stability and processing, host defense, chromosome segregation, transcription, and translation (Dang, et al. 2011). There are three components comprising the fungal RNAi system: Dicer, Argonaute/Piwi (Ago), and RNA-dependent RNA polymerase (RdRP) (Nicolas and Ruiz-Vazquez 2013). sRNA biosynthesis is mediated by Dicer and RdRP. RdRP produces double-stranded RNAs (dsRNAs) from single-stranded RNAs (ssRNAs) as templates which can originate from endogenous transcripts (e.g. transposable elements) or foreign RNAs (e.g. viral RNAs). Dicer is a class IV eukaryotic ribonuclease III enzyme (EC 3.1.26.3) which chops target dsRNAs into sRNAs. In Fungi, together with Dicer, an additional ribonuclease III enzyme belonging to class II, which shows differences in protein architecture, also contributes to digestion of dsRNAs. The identical function is mediated by class I ribonuclease III enzyme in prokaryotes (Lamontagne, et al. 2000; Liang, et al. 2014). After the

biogenesis of sRNAs by RdRP and Dicer, synthesized sRNAs are loaded to an RNA-induced silencing complex (RISC) or an RNA-induced transcriptional silencing (RITS) complex, which have AGOs as core components. By loading of sRNAs, RISC or RITS complexes can recognize and approach target RNAs to trigger gene silencing. Interestingly, since it is essential for signal crosstalk between host and symbiont, RNAi systems of symbiotic organisms often underwent selective pressure during evolution. Indeed, previous study revealed that AGO evolution in nematode family was correlated with symbiosis by showing additional AGO protein diversity found in parasitic nematodes compared to free-living relatives (Buck and Blaxter 2013). However, even though the importance of RNAi system in the gene regulation mechanism of AM symbiosis can be assumed, the RNAi system in AMF has never been considered.

1.4. Importance of light perception in biology of fungi

Another important gene regulation mechanism of fungi which was poorly studied in AMF is the light perception mechanism. In nature, all organisms interact with their environment. The environment, which includes biotic and abiotic stresses, is changing dynamically (Corrochano 2011). To follow these changes, organisms continuously evolve mechanisms and strategies in order to adapt themselves for changing environments. Each organism has its own way to sense the factors related with environmental signals including humidity, pressures, light and chemicals (Corrochano 2011; Purschwitz, et al. 2006). This is particularly important for the organisms that are not able to move around and change their habitat such as plants or fungi. Therefore, plants and fungi have evolved diverse mechanisms to sense environmental signals. Environmental change perception is more important if the sensing of certain factors is related to an organism's reproduction and metabolism (Purschwitz, et al. 2006). Moreover, light is the most useful

indicator for fungi in nature since it is highly related to humid and dry soil conditions. Light modulates various physiological and biological processes including sporulation, the germination of spores, growth of hyphae and synthesis of metabolites, thus, it is considered as the most important stimuli in fungi (Corrochano 2011; Purschwitz, et al. 2006; Rodriguez-Romero, et al. 2010). Until now, different light wavelength perception mechanisms were reported in kingdom Fungi (Corrochano 2011; Purschwitz, et al. 2006). Since the first report of light perception derived from a random mutation of *N. crassa*, this mechanism was largely studied in Fungi, particularly in the phylum Ascomycota and Basidiomycota (Sancar, et al. 2015). Fungal light sensing mechanisms vary within and between species, along the corresponding different wavelengths of the light. However, according to Purschwitz, et al. (2006), blue light perception was suggested to be the oldest mechanism acquired by fungi during their evolution.

1.5. Blue light perception in fungi

In fungi, three different blue light perception mechanisms exist with different core proteins called White collar (WC) proteins, Vivid (VVD) and Cryptochrome (CRY), respectively. However, the most general and widely distributed blue light perception is mediated by two proteins WC1 and WC2, which were first reported in *Neurospora crassa* (Corrochano 2011). WC1 and WC2 proteins have light-dependent phosphorylation domains which induce the inter-cellular signaling pathways (Takemiya, et al. 2005). Both proteins together act as a transcription factor, making the signaling cascades in various output pathways including spore production and hyphae growth. WC1 binds FAD (Flavin Adenine Dinucleotide) in its PAS (Per-Arnt-Sim) domain. When the blue light (wavelength of 400-470 nm) excites the electrons in FAD, it affects the chemical structure of FAD, causing entire confirmation changes in the WC1 protein. WC1

protein with a changed structure becomes active and makes the further meta-structure with WC2 protein, and that activates the further signal pathways downstream with their characteristic as transcriptional regulator (Schwerdtfeger and Linden 2000). In fungi, the genes encoding WC1 and WC2 proteins were reported and named *wc1* and *wc2* (Purschwitz, et al. 2006; Schwerdtfeger and Linden 2000).

1.6. Arbuscular Mycorrhizal Fungi (AMF) and light response

Despite the important role light sensing plays in gene regulation of an organism, the sensing mechanism can degenerate during evolution, for example in the case of an organism which lives without light (Sumner-Rooney 2018). This could explain the scarce investigation of the light sensing mechanism of AMF which lives its entire lifecycle underground as obligatory plant root symbiont. Only few studies reported that AMF respond to the light stimuli. The first report was for the AMF species *Glomus fasciculatus* where light affected the infection and sporulation processes (Ferguson and Menge 1982), followed by Nagahashi and Douds (2003) who showed that blue light (390~550nm) affected early germination of spores and hyphae branching in *Gigaspora gigantea*. Both sporulation and host infection are critical processes for AMF to undergo their life cycle. Thus, investigating the possible existence of blue light perception and mechanism of blue light sensing is essential for further understanding of AMF's function and physiology. To date, there is no study conducted on AMF's blue light perception mechanism using genetics or molecular biology.

1.7. Circadian rhythms and gene regulation

One of the most important and interesting mechanism of gene regulation is related to circadian rhythms. Circadian rhythms are ubiquitous in nature. The earth's energy flow fluctuates over every 24 hours as day and night because of the rotation of the earth. Thus, organisms have evolved rhythmical oscillations of metabolism to adapt them to the changing environment (Pittendrigh et al, 1960; Deborah et al, 1996; Roenneberg et al, 1999). Circadian rhythms exist in prokaryotes (ex, cyanobacteria or microbes in human intestine) (Johnson, et al. 1996; Kondo and Ishiura 2000) and also in eukaryotes (ex, plants, animals or fungi) (Johnston, et al. 2016; Liu and Bell-Pedersen 2006). The rhythms are produced and coordinated by an endogenous mechanism called a circadian clock. It varies among the organisms, but in general, the circadian clock is acknowledged to be responsible for regulating between 2 and 15% of the total transcriptome of an organism. (Dibner, et al. 2010; McDonald and Rosbash 2001; Nagel and Kay 2012; Storch, et al. 2002). There are many reports describing land autotrophs (plants) that maintain their circadian rhythm mainly with light and temperature (McClung 2006; Simon and Dodd 2017). Another report also showed that cyanobacteria have circadian rhythms in their nitrogenase activity, as well as in transcription, translation, and enzyme activity levels (Chen, et al. 1998). It is also reported that yeast and *N. crassa* have well-defined circadian rhythm-related genes (Liu and Bell-Pedersen 2006). Interestingly, a recent field study showed that AMF have a diurnal rhythm in their hyphal growth in the field (Hernandez and Allen 2013). However, the genes or mechanism involved in this phenomenon remain unexplored. Diurnal/Nocturnal (Circadian) rhythms are an extension of circadian rhythms (McClung 2006; Zisapel 2018).

1.8. Fungal circadian clock mechanism and core genes.

The first circadian oscillation in fungi was reported in growth of *N. crassa*, followed by many reports on the fungal circadian clock mechanism, and successfully revealed the fundamental mechanism consisted of three core proteins, Frequency (FRQ), White collar 1 (WC1), and White collar 2 (WC2) (Loros and Dunlap 2001). Interestingly, WC1 and WC2 also take in charge of blue light perception in fungi. Thus, in fungi blue light perception mechanism is also connected with circadian clock. Indeed, the activity of FRQ and its linkage to WC proteins were tested by blue light stimulation of *N. crassa* (Dunlap and Loros 2005; Feldman and Hoyle 1973; Lewis and Feldman 1996). These clock components produce a rhythmic oscillation, which is called a free-running rhythm, based on a feedback loop (Dunlap and Loros 2005). FRQ feeds back to shut down its own expression by suppressing the activity of the positive-acting elements, WC1 and WC2. The white collar complex (WCC), consisting both WC1 and WC2, activates the rhythmic transcription of circadian components, including its repressor protein FRQ. In the morning, FRQ inhibits the activity of WCC by mediating the phosphorylation of WCC in a process that requires FRQ-interacting RNA helicase (FRH), Casein kinase 1 (CKI) and CKII. Later in the day, FRQ supports the cytosolic accumulation of WCC. At night, hyper-phosphorylated FRQ is degraded through the ubiquitin-proteasome pathway and WCC is again active (Gallego and Virshup 2007). WC1 and WC2 proteins contain PAS domains, which are specific domains for protein-protein interaction and are similar to those found in the mammalian clock proteins Period, Clock, and BMAL1 (Dunlap and Loros 2005; Loros and Dunlap 2001). Since the phosphorylation of FRQ protein is especially important for the feedback loop of fungal circadian clock mechanism, and FRQ and WCC's phosphorylation is highly affected by light changes during the day and night, one could speculate that soil fungi which are not directly exposed to light may have different circadian clock mechanism. Would the general fungal clock

still be conserved inside of AMF which evolved approximately 460 million years as underground symbiotic partner of plant root?

1.9. Circadian rhythms related to nutrition assimilation of various organisms

Another aspect of the circadian clock that needs to be highlighted regarding AM symbiosis is that the clock orchestrates the nutrition fluxes. Carbon fixation related to plant circadian rhythms is intensively studied because autotrophic organisms depend on sunlight which has 24-hour fluctuating rhythm (McClung 2006; Nagel and Kay 2012). Also, the nitrogenase related circadian rhythm system was widely studied in cyanobacteria (mainly, in species *Synechococcus elongates* which has a well-defined biological rhythm, clock genes and mutants (Johnson, et al. 1998; Kondo and Ishiura 2000). A large body of research has shown that, in AM symbiosis, phosphorus, nitrogen, and carbon exchanges are the three core processes related with establishment and maintenance of mycorrhizal symbiosis (Blanke, et al. 2005; Bonfante and Genre 2010; Harley and Smith 1983; Schüßler, et al. 2007; Smith and Read 2008). If a circadian mechanism is conserved in AMF, it is likely to be related with these nutrition fluxes.

1.10. The objectives, hypotheses and thesis presentation

Even though AMF have significant roles in ecosystem, their genetics and evolution are far from being understood. The regulation of gene expression is one of the most crucial factors in understanding the biological mechanisms of an organism (Aravind, et al. 2012; Damiani, et al. 2012). Gene regulation is a complex mechanism maintained by endless crosstalk between an organism and its surroundings, rather than a simple mechanical process. The regulation of gene

expression becomes more crucial and complex in the case of organisms which are intimately forming symbiosis with others (Aravind, et al. 2012; Damiani, et al. 2012; Hou and Lin 2009; Yigit, et al. 2006). Thus, mechanisms were often exposed to selective pressures during evolution, especially in the genomes of symbiotic organisms. In the case of the RNAi system, which is essential for signal crosstalk between host and symbiont, the core protein (Argonaute, AGO) diversity in nematodes was related to symbiosis since additional AGO protein diversity was found in parasitic nematodes compared to free-living relatives (Buck and Blaxter 2013). This is similar to the case of light perception. Despite the fact that light sensing has an important role in gene regulation of an organism, the sensing mechanism can degenerate during evolution in the case of an organism which lives without light (Sumner-Rooney 2018). AMF are different from other fungi with respect to their light sensing mechanisms, since AMF evolved to be obligatory underground plant root symbionts. Several studies showed that blue light can affect the spore germination and hypha growth of AMF, but its mechanism was not understood (Ferguson and Menge 1982; Nagahashi and Douds 2003, 2004). The circadian clock can also be subject to a selective pressure. The clock is known to regulate between 2 and 15% of the total transcriptome in numerous organisms (Dibner, et al. 2010; McDonald and Rosbash 2001; Nagel and Kay 2012; Storch, et al. 2002). However, the clock's components could be lost during evolution in the case of organisms which do not need to synchronize their biological rhythms with the environment (Halberg, et al. 2005; Loza-Correa, et al. 2010; Mori and Johnson 2000).

The RNAi system, blue light sensing mechanisms and the circadian clock are important mechanisms inside fungi which regulate the expression of various genes. However, prior to my study, we had no idea what kind of systems were in AMF nor if they were conserved. The objective of my Ph.D. project was therefore to study the evolution of the RNAi system, blue

light sensing mechanisms and the circadian clock in the *R. irregularis* genome using both bioinformatic and molecular biological approaches.

The specific objectives were:

1. To investigate if the RNAi system was conserved in *R. irregularis* and to explore the evolutionary traits of its core proteins.

2. To determine whether a blue light sensing mechanism exists in *R. irregularis*.

3. To investigate if the general fungal circadian clock components exist in *R. irregularis*.

I have addressed and tested the following hypotheses in my Ph.D. project:

1. The general fungal RNAi system is conserved with its core components in the *R. irregularis* genome.

2. The conserved RNAi system in *R. irregularis* possesses evolutionary traits such as gene gain or loss which reflect its evolutionary history.

3. The fungal blue light sensing mechanism is conserved with its core components in the *R. irregularis* genome.

4. The conserved blue light sensing mechanism in *R. irregularis* has evolutionary traits such as gene gain or loss which reflect its evolutionary history.

5. The blue light-sensing related genes in *R. irregularis* are transcribed and their expression is independent of light.

6. The general fungal circadian clock components are conserved in *R. irregularis* with FRQ as its core oscillator.

7. The conserved circadian clock in *R. irregularis* has evolutionary traits such as gene gain or loss which reflect its evolutionary history.

8. The expression of the conserved FRQ homolog can be induced by light as for the circadian model fungus, *Neurospora crassa.*

In the first investigation (second chapter, first article), I reported that the general fungal RNAi system was highly conserved and expressed in the model AMF *R. irregularis* with respect to all three of its core proteins (Dicer, RdRP, and AGO). We found several interesting evolutionary traits such as expansion of one of the core protein (AGO) homologs, a shrinking of other core protein (Dicer, class IV ribonuclease) homologs and a loss of class II ribonuclease homologs. Surprisingly, we also found *R. irregularis* and its relative AMF harbor class I ribonuclease homologs which have never been reported in eukaryotes. Further investigation revealed these genes were horizontally transferred from autotrophic cyanobacteria.

The third chapter of my thesis (second article) describes the study of the blue light sensing mechanism and the circadian clock inside *R. irregularis*. We found the conserved proteins of the blue light sensing mechanism in *R. irregularis* which were also reported to be members of the core components in the fungal circadian clock. We further found that all of the circadian clock components, as well as several output genes involved in nutrition metabolism, were well conserved and expressed in *R. irregularis*. Our finding provides new insight into the evolution of circadian clock components in fungi, especially with the finding of the *frq* gene which was known to have evolved after the divergence of Dikarya (Ascomycota and Basidiomycota).

Interestingly, not only at the stage of spore germination, but also after forming the mycorrhizal symbiosis, the expression of circadian core oscillator *frq* can be induced by blue light.

Finally, the fourth chapter of my thesis (third article) is an opinion letter based on the finding of the hypotheses developed in the third chapter. We expanded and emphasized the importance of chronobiology in the context of holobionts which is a new concept in understanding biological individual in nature.

My thesis is presented under the scientific article form for the Doctorate in Biological Science Program at the Université de Montreal. Chapter 1 of the thesis (General Introduction) introduces the current knowledge about the arbuscular mycorrhizal fungal biology, the RNAi system, blue light sensing mechanisms and the circadian clock in fungi as well as their importance in understanding of AM symbiosis. The experiments, methods and results are then introduced in the chapters 2 and 3 and are followed by chapter 4 which is a submitted manuscript containing a literature review and an opinion letter. Chapter 5, the last section of the thesis, serves as a general discussion and conclusion of all the results obtained, followed by the future perspectives. During my study, the scientific name of AMF, *Rhizophagus irregularis*, was changed from *Rhizoglomus irregulare* to *Rhizophagus irregularis*. In the cases of published manuscripts which use *Rhizoglomus irregulare* as the official nomenclature, the explanation of synonyms was noted.

Chapter 2 – Components of RNAi system conserved inside *R. irregularis*

Presentation of article 1

This article was published in Genome Biology and Evolution. I have conceived and designed the experiment with supervision of Dr. M. Hijri. I have conducted bioinformatics and molecular biological experiments with some assistance from Miss M. Kong and Dr. P. Harrison. I have written this article with revisions from Dr. P. Harrison and Dr. M. Hijri. During the time of this study conducted and published, *Rhizophagus irregulare* was the official nomenclature of *Rhizophagus irregularis* and was explained in the manuscript.

Article 1 - Conserved proteins of the RNA interference system in the arbuscular mycorrhizal fungus *Rhizoglomus irregulare* provide new insight into the evolutionary history of Glomeromycota

Soon-Jae Lee[1], Mengxuan Kong[1], Paul Harrison[2] and Mohamed Hijri[1]

[1] *Institut de Recherche en Biologie Végétale, Université de Montréal and Jardin botanique de Montréal, 4101 Sherbrooke est, Montréal (Québec) H1X 2B2, Canada*

[2] *Department of Biology, McGill University, 845 Rue Sherbrooke Ouest, Montréal (Quebec) H3A 1B1, Canada*

Published in : Genome Biology and Evolution (GBE), January 10, 2018. 10(1):328–343. doi:10.1093/gbe/evy002

2.1. Abstract

Horizontal gene transfer (HGT) is an important mechanism in the evolution of many living organisms particularly in Prokaryotes where genes are frequently dispersed between taxa. Although, HGT has been reported in Eukaryotes, its accumulative effect and its frequency has been questioned. Arbuscular mycorrhizal fungi (AMF) are an early diverged fungal lineage belonging to phylum *Glomeromycota*, whose phylogenetic position is still under debate. The history of AMF and land plant symbiosis dates back to at least 460 million years ago. However, *Glomeromycota* are estimated to have emerged much earlier than land plants. In this study, we surveyed genomic and transcriptomic data of the model arbuscular mycorrhizal fungus *Rhizoglomus irregulare* (synonym *Rhizophagus irregularis*) and its relatives to search for evidence of HGT that occurred during AMF evolution. Surprisingly, we found a signature of putative HGT of class I ribonuclease III protein-coding genes that occurred from autotrophic cyanobacteria genomes to *R. irregulare*. At least one of two HGTs was conserved among AMF species with high levels of sequence similarity. Previously, an example of intimate symbiosis between AM fungus and cyanobacteria was reported in the literature. Ribonuclease III family enzymes are important in small RNA (sRNA) regulation in Fungi together with two additional core proteins (Argonaute/piwi and RdRP). The eukaryotic RNA interference system found in AMF was conserved and showed homology with high sequence similarity with *Mucoromycotina*, a group of fungi closely related to *Glomeromycota*. Prior to this analysis, class I ribonuclease III has not been identified in any eukaryotes. Our results indicate that a unique acquisition of class I ribonuclease III in AMF is due to a HGT event that occurred from cyanobacteria to *Glomeromycota,* at least before the divergence of the two *Glomeromycota* orders *Diversisporales* and *Glomerales*.

2.2. Keywords

RNA interference, Ribonuclease, Argonaute, Dicer, Arbuscular mycorrhizal fungi, cyanobacteria, symbiosis, Horizontal gene transfer, Gene expression, Gene evolution.

2.3. Introduction

Horizontal gene transfer (HGT) is an acquisition of genetic material between two or more organisms that may be unrelated by evolutionary descent. HGT is distinct from the vertical transmission of DNA from parent to offspring. HGTs were commonly reported in prokaryotes (Pinto-Carbo, et al. 2016; Soucy, et al. 2015). Although, HGTs were also reported in eukaryotic organisms including Fungi, their frequency and cumulative effects on eukaryotic genomes are under debate (Martin 2017; Soanes and Richards 2014; Soucy, et al. 2015). The events of HGT occur more frequently, if partners live in symbiotic association (Chaib De Mares, et al. 2015; Davis and Xi 2015; He, et al. 2016; Pinto-Carbo, et al. 2016). During evolution, partners involved in close symbiosis could have exchanged their genetic material in both directions by HGT. These transferred genes can be used to trace back their evolutionary history (Chaib De Mares, et al. 2015). Many examples of HGT between endosymbiotic prokaryotes and their hosts have been reported, including *Mollicutes-related endobacteria*, the bacterium which forms endosymbiosis with *Glomeromycota* and *Mucoromycotina* (Kuo 2015; Naito, et al. 2015; Torres-Cortes, et al. 2015). These studies found between 3% and 5% of protein-coding genes in *Mollicutes-related endobacterial* genomes, were horizontally transferred from the ancestral fungi of *Glomeromycota* and *Mucoromycotina.* The finding supports the hypothesis of ancestral

symbiosis between a fungal ancestor and a bacterial ancestor. However, whether HGTs in *Glomeromycota* have occurred is unclear.

Mycorrhiza is a mutualistic association between fungi and plants and it is one of the most intimate and ubiquitous symbioses found on earth (Smith and Read 2008). Arbuscular mycorrhizal fungi (AMF) form mycorrhizal symbioses known as endomycorrhiza. AMF are considered as an early diverged fungal lineage assigned to phylum *Glomeromycota*, and they are obligatory symbionts that establish symbiosis with host plant root cells, where they form special structures called "arbuscules" in which nutrients are exchanged between partners (Bonfante and Genre 2010). After colonization of host cell in the cortex of roots, AMF elongate the network of hyphae along and beyond the rhizosphere. Through the hyphae, AMF absorbs various nutrients from the soil (including phosphate and nitrogen), that are used as an exchange of carbon supply on the host plant's photosynthesis (Nouri, et al. 2014). Up to date, over 80% of land plant species are known to establish the symbiosis with AMF (Oehl, et al. 2011). The history of AMF and autotrophic land plant's symbiosis was dated back to at least 460 million years ago, according to fossil records and molecular clock analysis (Luc Simon 1993; Redecker, et al. 2000). On the other hand, detailed studies by molecular clocks indicated that the appearance of *Glomeromycota* is much older than the emergence of land plant on earth (Berbee and Taylor 2010; Bidartondo, et al. 2011). At the same time, it was recently reported that fungus-like mycelial fossils were found in a 2.4-billion-year-old basalt (Bengtson, et al. 2017), which showed that the entire fungal lineage is much older than previously thought, and their early evolution may have occurred in deep ocean. It is still unclear at what time did the AMF ancestor arise and whether they had previous symbiotic partners before the land plants, but we may find some evidence to clarify their evolution by tracking HGT events.

The processing and expression of small non-coding RNAs (sRNAs) is one of the crucial biological processes in symbiotic organisms (Formey, et al. 2016). sRNAs can originate from bioprocessing of small interference RNAs (siRNAs) or microRNAs (miRNAs) in the cell. Symbiosis involving the sRNA processing mechanism has been previously reported in other symbiotic organisms, such as in the symbiosis between soybean and *Rhizobium*, or between maize and diazotrophic bacteria (Formey, et al. 2016; Thiebaut, et al. 2014; Yan, et al. 2015). Branscheid, et al. (2011) revealed that mycorrhizal plants actively express and regulate symbiosis-related miRNAs (microRNAs) in the establishment and development of AMF mycorrhization. However, the possible related regulation mechanism of sRNA in AMF was yet far from understanding.

In eukaryotes, the sRNA processing and regulation mechanism is called RNA interference (RNAi). RNAi and gene silencing has been reported in a broad range of eukaryotic organisms, including Fungi (Dang, et al. 2011; Torres-Martinez and Ruiz-Vazquez 2016), where RNAi-related pathways play a role in various functions comprising genomic surveillance, gene regulation and genomic defense. The study of the RNAi system in Fungi has been expanded by the discovery of a phenomenon called “Quelling” in *Neurospora crassa* (Pickford, et al. 2002). Quelling is post-transcriptional gene silencing guided by sRNAs resulting from the RNAi process. The principle of the fungal RNAi process is to form sRNAs which can act as regulators for cellular processes such as development, RNA stability and processing, host defense, chromosome segregation, transcription, and translation (Dang, et al. 2011). There are three components known in the RNAi system: Dicer, Argonaute/piwi (Ago), and RNA dependent RNA polymerase (RdRP) (Nicolas and Ruiz-Vazquez 2013). Dicer and RdRP play roles in sRNA biosynthesis. RdRP produces double-stranded RNAs (dsRNAs) from single-stranded

RNAs (ssRNAs) which can originate from endogenous transcripts (e.g. transposable elements) or foreign RNAs (e.g. viral RNAs). Dicer is a class IV eukaryotic ribonuclease III enzyme (BRENDA 3.1.26.3) which chops target dsRNAs into sRNAs. In Fungi, together with Dicer, an additional ribonuclease III enzyme belonging to class II which shows differences in protein architecture, also contributes to digestion of dsRNAs. The same function is mediated by class I ribonuclease III enzyme in prokaryotes (Lamontagne, et al. 2000; Liang, et al. 2014). sRNA biogenesis could also directly function as a defense mechanism against viral infection. Biosynthesized sRNAs are loaded to RNA-induced silencing complex (RISC) or RNA-induced transcriptional silencing (RITS) complex, which have AGOs as core components. By loading of sRNAs, RISC or RITS complex can recognize and approach to target RNAs to trigger the silencing.

Our study aims to investigate the evolutionary history of the RNAi system in *Rhizoglomus irregulare*. We asked the question: does HGT occur between *Glomeromycota* and their associated microbes? The rationale was that HGT might have higher chance to occur between symbiotic partners particularly in genes involved in sRNA processing mechanisms during evolution. To address our question, we searched for the RNAi system in the model arbuscular mycorrhizal fungus, *R. irregulare* whose expressed sequence tags (ESTs), transcriptome and genome are publicly available, using three RNAi core genes of three model fungi *Neurospora crassa, Mucor circinelloides* and *Schizosccharomyces pombe* as queries (Table S1). BLAST-based analyses with ESTs and transcriptomic data was conducted. In the case mRNA level expression was not confirmed, polymerase chain reaction (PCR) followed by Sanger sequencing was conducted. All of the three putative core proteins of the AMF RNAi system showed high sequence identity with those proteins of *Mucoromycotina*. Surprisingly,

not only does the fungal RNAi system operate with eukaryotic (class IV) ribonuclease III proteins, but we also discovered that AMF possess two additional prokaryotic (class I) ribonuclease III protein-coding genes arising by putative HGT from cyanobacteria. This HGT was not observed in *Mucoromycotina* and other basal fungi, while one of the two enzymes resulted from HGT was highly conserved among five different AMF species along with two different orders (*Diversisporales* and *Glomerales*) in *Glomeromycota*, showing that HGT of the enzyme-coding gene occurred at least in the ancestor of *Diversisporales* and *Glomerales*. The transcriptome of horizontally transferred cyanobacterial ribonuclease III enzymes were expressed in both germination and mycorrhization phases of AMF life cycle.

2.4. Materials and Methods

2.4.1. Sequence identification of RNAi core genes

Core RNAi genes of *Neurospora crassa, Mucor circinelloides* and *Schizosccharomyces pombe,* were used as reference sequences (Table S2.1) to identify relevant homologs in the transcriptomic and genomic data of *R. irregulare* using TBLASTN (Altschul, et al. 1997). These transcriptomic and genomic data include published expressed sequence tags (ESTs) and the genome of *R. irregulare* isolate DAOM-197198 Version 2.0 available at the JGI Genome Portal (Tisserant, et al. 2012; Tisserant, et al. 2013). Orthology of the *R. irregulare* genes was first assessed using the reciprocal best hits approach in conjunction with domain architecture analysis. Hits of e-value lower than $1E^{-04}$ with a query coverage higher than 60%, were compared with all proteins of *N. crassa, M. circinelloides* and *S. pombe* using BLASTX. Putative open reading frames of *R. irregulare* candidate sequences were annotated by pairwise comparisons using BLASTX against GenBank sequences. Accession numbers of all sequences of *R. irregulare*

used in this study, are shown in Table S2.2. To confirm the identity of these homologs, retrieved sequences from reciprocal BLAST were also examined for functional domains conserved in the core RNAi genes using SMART 7 (Simple Modular Architecture Research Tool) (http://smart.embl.de/) and InterProScan v. 61.0 (https://www.ebi.ac.uk/interpro/). For AGO homologs, ArgoN (IPR032474), PAZ (Piwi-Argonaute-Zwille) domain (IPR003100) and Piwi (IPR003165) domains were searched for. RNA dependent RNA polymerase domain (IPR007855) was searched for in RDRP homologs. For DICER, ATP-binding helicase domains (DEXDc (IPR014001), HELICc (IPR001650) and ribonuclease III (RIBOc) domains (IPR000999) were investigated. Candidates were considered as hits if they contained all domains and the cut-off e-value of each domain hit was $1E^{-04}$. With retrieved candidates of *R. irregualre* as a query, the RNAi genes of a broad range of fungal taxa, including *Ascomycota*, *Basidiomycota* and *Mucoromycotina* were searched for using EMBL databases of orthologous groups and functional annotations using EggNOG v. 4.5.1 (http://eggnogdb.embl.de). For the EggNOG orthologous sequence group search, we applied cut-off e-value = E^{-30} with the query coverage over 60%. Finally, to confirm homology, *R. irregulare* sequences were compared with orthologous sequences in representative Fungi and, when it is possible, with representatives of well-defined outgroups of Fungi (such as *Arabidopsis thaliana* or *Drosophila melanogaster*), using phylogeny. Full species name and related accession codes of all sequences used for phylogenetic analyses were noted with each phylogeny. Protein multiple sequence alignments (MUSCLE v. 3.8 (Edgar 2004)) were inspected and divergent or ambiguous positions were removed using BioEdit v. 7. For each protein, the best evolutionary models with the lowest Akaike information criterion (AIC) were determined using ProtTest 3 (Darriba, et al. 2011) and the selected model for each phylogeny were noted with each phylogenetic tree. The phylogenetic

trees were inferred by PhyML v. 3.0 (Guindon, et al. 2010), using the best model with 1,000 bootstrap replicates.

2.4.2. Sequence identification of class I ribonuclease III genes

Two sequences of class I (prokaryotic) ribonuclease III genes (*rirnc 2* and *rirnc 3*) were found during the identification of ribonuclease III homologs in *R, irregulare* described in the previous section. Domain architectures were examined using SMART 7 (Simple Modular Architecture Research Tool) (http://smart.embl.de/) and InterProScan v. 61.0 (https://www.ebi.ac.uk/interpro/) with each domain's cut-off e-value = $1E^{-04}$. Candidate containing both ribonuclease III (RIBOc) domain (IPR000999) and double stranded RNA-binding domain (DSRM) (IPR014720) were considered as a hit. Sequences were validated for mRNA level expression by having EST contigs as well as RNA extraction followed by PCR and Sanger sequencing described in PCR and Sanger sequencing section. The corresponding genome scaffold for the two sequences were found using BLASTN. Similar sequences of these genes were also found in the *R. irregulare* genome data published in another study using single nucleus sequencing of *R. irregulre* (coverage depth 60x) (Lin, et al. 2014). In addition, we found other similar sequences in the GenBank database (Table S2.2). In order to investigate the conservation of the core catalytic (RIBOc) domain from class I ribonuclease III homologs among different species of *Glomeromycota*, a TBLASTN search was performed with protein sequences of RIBOc in *rirnc 2* and *rirnc 3* as queries against the genomes of four other AMF species (*Rhizoglomus clarum* (syn, *Rhizophagus clarus*), *Rhizoglomus diaphanum* (syn, *Rhizophagus diaphanous), Claroideoglomus etunicatum* and *Scutelospora calospora* of which genomic data are available in our lab. *R. clarum* and *R. diaphanum* were sequenced by Roche

454 and *C. etunicatum* and *S. calospora* were sequenced by Illumina Mi-Seq. Quality trimming of the raw sequences applied before the assembly that was conducted with Geneious 8.0. Contigs with coverage lower than 6x were discarded. Hits of e-value lower than $1E^{-10}$ with the query coverage higher than 60%, were examined for domain architecture using SMART 7 (Simple Modular Architecture Research Tool) (http://smart.embl.de/) and InterProScan v. 61.0 (https://www.ebi.ac.uk/interpro/). Retrieved sequences of *Glomeromycota* were used to build protein multiple sequence alignment by MUSCLE v. 3.8 (Edgar 2004) and inspected to investigate conserved amino acid residues together with RIBOc domains of class II ribonuclease III enzymes which were known to have the most similarity with class I ribonuclease III enzymes (Liang, et al. 2014). The four Glomeromyco sequences used in our analysis were deposited in GenBank database (accession numbers from MF926648 to MF926651).

2.4.3. Phylogenetic analysis of class I ribonuclease III amino acid sequences of *R. irregulare*

Phylogenetic analysis of core RIBOc domains of class I and II ribonuclease III enzyme was conducted. RIBOc domains from class I ribonuclease III homologs found in *R. irregulare* and its relatives as well as two well-defined class II ribonuclease III enzymes from yeast (PAC1 and RNT1), were used as queries. RIBOc domains from the homologs of each query were compared in GenBank database using strict cut-off threshold (e-value = E^{-30} and query coverage over 60%), to examine their sequence similarities. Species name and accession numbers of all sequences used in phylogenetic analyses were indicated in the phylogenetic tree. RIBOc homologs from four Glomeromycota species: *R. clarum, R. diaphanum, C. etunicatum* and *S.*

calospora, were also included in phylogenetic analyses. The phylogenetic trees were inferred by PhyML v. 3.0 (Guindon, et al. 2010), using the best model with 1,000 bootstrap replicates.

Finally, to confirm the class I ribonuclease III sequence similarity, the full length amino acid sequences of RIRNC 2 and 3 were compared with the orthologous sequences from Bacteria using phylogeny in which a sequence of Archaea was used as an outgroup. Query sequences of RIRNC 2 and 3 and bacterial orthologs were compared with datasets from EggNOG v. 4.5.1. For the EggNOG orthologous sequence group finding, we applied cut-off e-value = E^{-30} with the query coverage over 60%. Species name and accession numbers of all sequences used to phylogenetic analyses are shown in the phylogenetic trees, which were inferred by PhyML v. 3.0 (Guindon, et al. 2010), as described above.

2.4.4. AMF isolate and growth condition

The arbuscular mycorrhizal fungal isolate used in this study was *Rhizoglomus irregulare* (syn. *Rhizophagus irregularis* previously named *Glomus irregulare*), deposited in the Fungal Collection of the Department of Agriculture Canada under the accession number DAOM-197198 (Ottawa, ON). *R. irregulare* isolate was cultured *in vitro* with Ri-T-DNA transformed *Daucus carota* roots in Petri dishes containing minimal medium (M) solidified with 0.4 % Gellan gum (Sigma Aldrich, Canada). The cultures were kept in an incubator at 25 °C in the dark. An AMF isolate is a culture that was originally started from a single spore and subcultured by transferring 1 to 2 cm^2 of gel containing a mixture of mycorrhized roots, spores and hyphae to a new Petri dish. Spores were harvested from four months old *in vitro* cultures. *R. irregulare* isolate DAOM-197198 spore suspension was also provided by Premier Tech Biotechnologies (Rivière-du-Loup, QC).

2.4.5. Culture preparation for germination and mycorrhization stages

60,000 fresh spores (20,000 spores per replicate) of *R. irregulare* were inoculated on the 1 mL of liquid minimal medium. The spores were germinated under dark condition in an incubator at 25℃ for three weeks. Culture preparation for mycorrhization stage of the isolate of *R. irregulare* was prepared as described above.

2.4.6. Extraction of RNA and cDNA library construction

Spores and hyphae of *R. irregulare* were harvested in a 1.5 ml microtube and quickly frozen with liquid nitrogen. Samples were grounded using a sterilized pestle. RNA extraction and DNase treatment were performed using E.Z.N.A Fungal RNA extraction kit (Omega-Biotek, Canada) according to the manufacturer's instructions. The amount of RNA per each reaction was normalized to 100 ng for further analyses. RT-PCR was done to produce cDNA library using iScript reverse transcription kit (Bio-Rad, Canada) following manufacturer's instructions.

2.4.7. PCR and Sanger sequencing

PCR was used to validate *RIRNC 2* mRNA sequences. cDNA samples produced from germinating spore were used as template. PCR primers were designed by Primer 3 software (Table S3). PCR reactions were performed using *Taq* DNA polymerase (Qiagen, Canada) following the manufacturer's recommendations. PCRs were performed in a volume of 20 µl under the following conditions: denaturation at 95 °C for 2 min, followed by 35 cycles of 94 °C for 30 sec, 48 °C for 30 sec and 70°C for 2 min. Final elongation was performed at 72 °C for 10

min. PCRs were run on a Mastercycler Pro S thermocycler (Eppendorf, Canada). PCR amplicons were visualized on a 1% agarose gels stained with GelRed (Invitrogen, Canada). Successful PCR amplicons were extracted from gels, purified and sequenced using a commercial service at the Genome Quebec Innovation Center at McGill University (Montreal, QC).

2.4.8. TaqMan qPCR assay

qPCR was performed to analyze the expression level changes of fungal-origin ribonuclease III (*Ridcl 1*) and bacterial-origin ribonuclease III (*Rirnc 2*) in two different phases of *R. irregulare* life cycle (spore germination and post mycorrhization phases). Two housekeeping genes (*18S rRNA* and *ubiquitin*) were used as reference of expression level. TaqMan probes and primer sets were designed by Primer 3 software (details are described in Table S2). Probes were labelled with FAM at the 5' end and BHQ-1 at the 3' end. Fluorescence data were collected with the ViiA^{TM} 7 Real-Time PCR System (Life Technologies, Canada). qPCR was performed in three biological replicates and three technical triplicates.

2.5. Results

2.5.1. Conservation of RNAi system in model AMF, *R. irregulare*

The genome of the model AMF species *R. irregulare* isolate DAOM-197198 possessed all of the three core genes (Dicer, RdRP, Argonaute/Piwi) involved in the RNAi system. RNAi core homologs retrieved using reciprocal BLAST and conserved domain searching, are summarized in Table S2.2, where all homologs were examined and compared using BLAST searches against the GenBank database. All homologs of RNAi system core proteins in *R.*

irregulare showed high sequence similarity with orthologs of Mucoromycotina fungi. Table 1 summarizes the number of homologs for each core protein of *R. irregulare* and three query species (*N. crassa, M. circinelloides* and *S. pombe*) with the comparison of their genome sizes. *R. irregulare* possessed the largest genome of 136.81 Mb which is 3.3-fold and 9.7-fold larger that the genomes of *N. crassa* (41.04 Mb) and *S. pombe* (14.1 Mb), respectively. *R. irregulare* had a large number of homologs of Argonaute proteins (26 homologs). On the contrary, despite its large genome size, *R. irregulare* didn't show any difference in number of homologs of RdRP (3 homologs) compared to phylogenetically close queries (*N. crassa* and *M. circinelloides*). Interestingly, all of the three query species had two eukaryotic ribonuclease III homologs (two IV ribonuclease III homologs in the case of *N. crassa* and *M. circinelloides* and one class II ribonuclease III with one class IV ribonuclease III in the case of *S. pombe*), while only one eukaryotic ribonuclease III homologue (named as RIDCL 1, from *R. irregulare* dicer-like protein) was detected in *R. irregulare* genome.

Table 2.1. The number of homologs of core genes of RNAi system in *R. irregulare* and three query species (*N. crassa*, *S. pombe* and *M. circinelloides*) with their genome sizes between brackets in mega base pairs (Mb). The number of copies in *R. irregulare* was determined based on BLAST and protein structure analyses. *R. irregulare* genome showed an expansion of Argonaute/Piwi genes (26 copies), and it had two class I ribonuclease III (prokaryotic) enzymes and only one class IV ribonuclease III (eukaryotic) enzyme.

Name of core component of RNAi system	**The number of homologs In species (with genome size published)**			
	R. irregulare (136.81 Mb)	*N. crassa* (41.04 Mb)	*S. pombe* (14.1 Mb)	*M. circinelloides* (36.6 Mb)
RNA dependent RNA polymerase (RDRP)	**3**	**3**	**1**	**3**

Argonaute/Piwi Protein (AGO)	**26**	**2**	**1**	**3**
Ribonuclease III	**1 (Class IV)** **2 (Class I)**	**2 (Class IV)**	**1 (Class IV)** **1 (Class II)**	**2 (Class IV)**

We analyzed the domain architecture of each protein homolog. Figure 2.1 (a, b and c) shows the conserved domains in representative homologs of each protein of the fungal RNAi system in *R. irregulare*. All of the homologs of the fungal RNAi system in *R. irregulare* (Figure 2.1 a, b and c) showed protein domain architectures identical to those of the original query sequences of *N. crassa, M. circinelloides* and *S. pombe*.

Figure 2.1

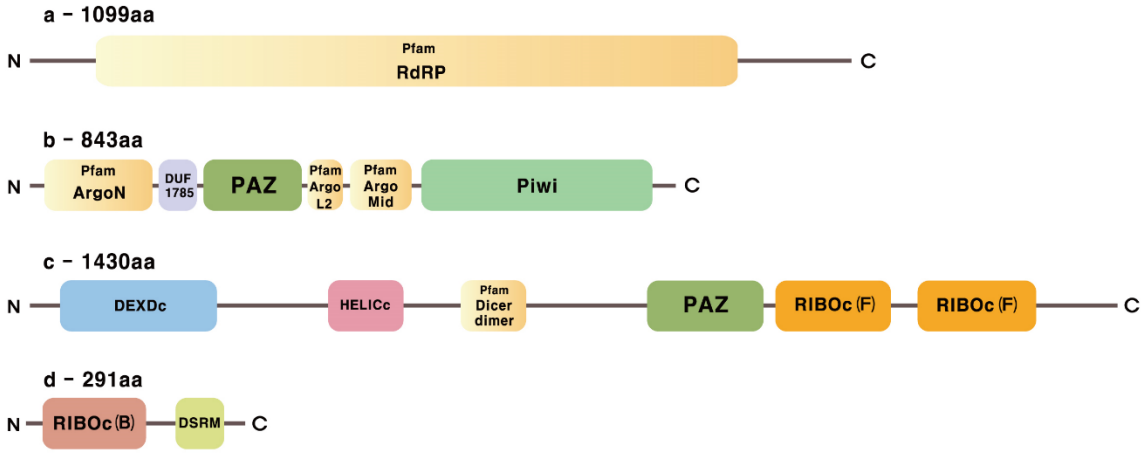


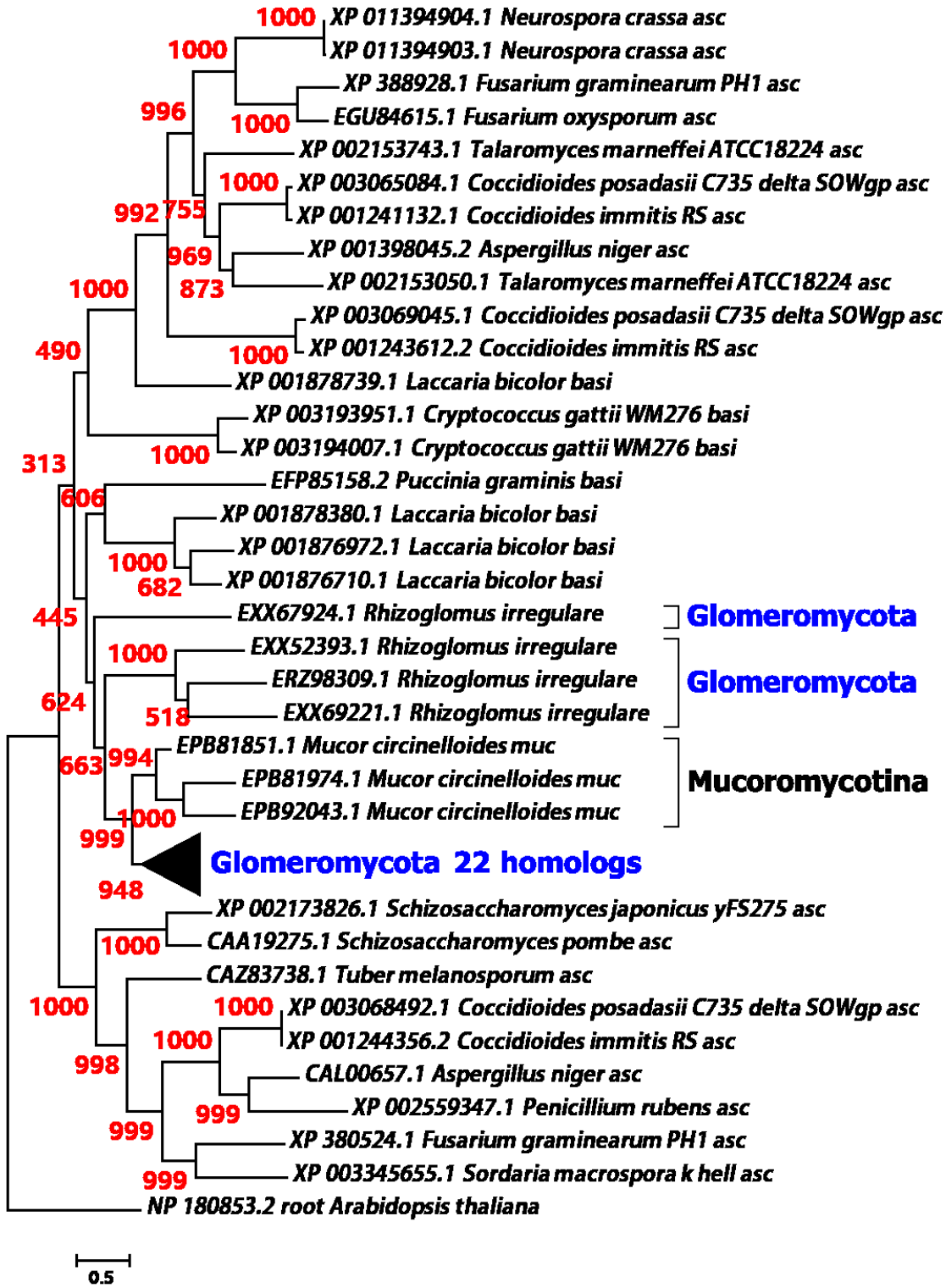
Figure 2.1. Domain based protein structures of three core genes forming RNAi system in *Rhizoglomus irregulare*. Four representative sequences of three core proteins (RdRP (a), Argonaute (b), and Dicer(c)) were analyzed by SMART (Leunic, et al, 2014) and InterProScan v. 61.0. (a) RNA dependent RNA polymerase domain (IPR007855) was found in *R. irregulare* RdRP. (b) Putative AGO homologs, N terminal domain of argonaute protein (ArgoN, Pfam domain) was detected with PAZ (Piwi-Argonaute-Zwille) domain (IPR003100) in the middle and Piwi domain (IPR003165) C terminal. (c) Class IV ribonuclease III enzyme (RIDCL) contains at its N terminal an ATP-binding helicase domains (DEXDc (IPR014001) and HELICc (IPR001650)), PAZ (IPR003100), and two catalytic ribonuclease III (RIBOc) domains (IPR000999) in C terminal. (d) Class I ribonuclease III enzyme (RIRNC) contains one ribonuclease III (RIBOc) domain (IPR000999) starting at 18 a.a. from N terminal and one double stranded RNA-binding domain (DSRM) (IPR014720) in C terminal.

To investigate the detailed protein homology, we inferred phylogenies for the three core genes using Maximum Likelihood. Figures 2.2 and S2.1 shows homology of three core proteins of RNAi system in *R. irregulare* with other fungal species including Ascomycota, Basidiomycota and Mucoromycotina. As expected, all of the three putative core proteins of RNAi system found in *R. irregulare* clustered together with Mucoromycotina in a single clade in each phylogenetic tree. Sequences of *R. irregulare* showed high sequence similarity with orthologs of the mucoromycete *M. circinelloides* (bootstrap values: 624/1000 for AGO, 966/1000 for RDRP, and 1000/1000 for RIDCL 1). On the contrary, the orthologs from *N. crassa* and *S. pombe* showed relatively low sequence similarity with *R. irregulare* and *M. circinelloides,* therefore they were distributed in different clades (Figure 2.2). Two sub-clusters were observed in the cluster of AGO homologs from *M. circinelloides* and *R, irregulare* (624/1000) (Figure 2.2a). Twenty-two sequences of *R. irregulare* clustered within three orthologs of *M. circinelloides* (1000/1000). However, four homologs of *R. irregulare* clustered separately from the 22 AGO homolog sequence. In RDRP (Figure 2.2b), all of three homologs from *R. irregulare* clustered with an orthologue from *M. circinelloides* (966/1000). However, two other orthologs of *M. circinelloides* clustered separately. In DICER homologs (Figure 2.2c), two homologs of *M. circinelloides* and one homolog of *R. irregulare* clustered in a single clade (1000/1000), showing their high sequence similarity. Apparently, with a divergence of the number of homologs and amino acid residues, all of the AMF RNAi core gene candidates showed the highest similarities with proteins of *Mucoromycotina*, a basal fungal subphylum suggested to be the closest neighbor of *Glomeromycota* in other studies using mitochondrial genomes (Nadimi, et al. 2012) and SSU rDNA sequences (Redecker, et al. 2013). Our findings

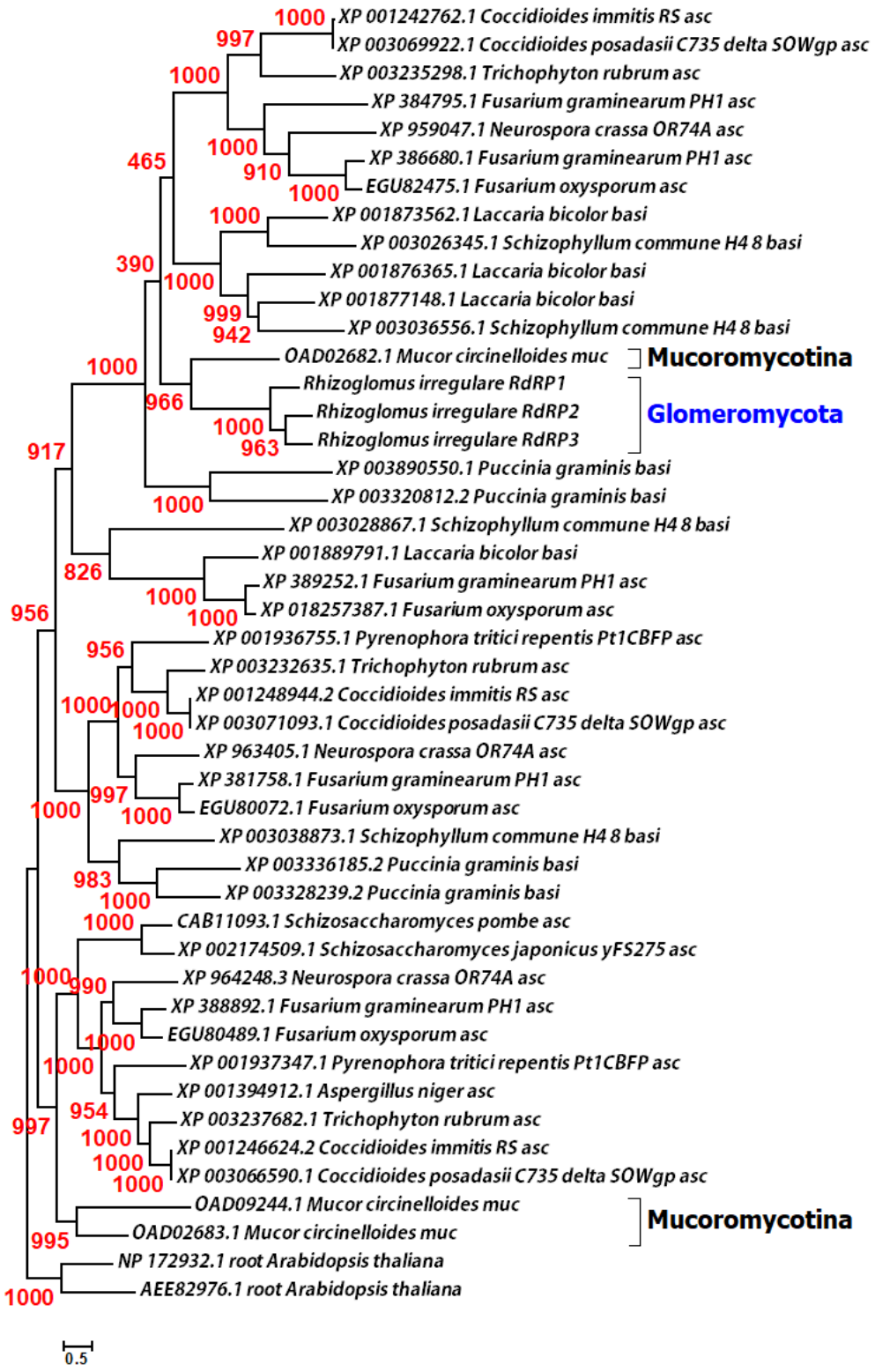
are in line with GenBank BLAST best-hit results of *R. irregulare* RNAi core protein sequences (Table S2.2).

Figure 2.2

(a) **ARGONAUTE/PIWI**


1000
XP 011394904.1 Neurospora crassa asc
1000
XP 011394903.1 Neurospora crassa asc
XP 388928.1 Fusarium graminearum PH1 asc
996
1000
EGU84615.1 Fusarium oxysporum asc
XP 002153743.1 Talaromyces marneffei ATCC18224 asc
1000
XP 003065084.1 Coccidioides posadasii C735 delta SOWgp asc
992
755
XP 001241132.1 Coccidioides immitis RS asc
969
XP 001398045.2 Aspergillus niger asc
1000
873
XP 002153050.1 Talaromyces marneffei ATCC18224 asc
XP 003069045.1 Coccidioides posadasii C735 delta SOWgp asc
490
1000
XP 001243612.2 Coccidioides immitis RS asc
XP 001878739.1 Laccaria bicolor basi
XP 003193951.1 Cryptococcus gattii WM276 basi
1000
XP 003194007.1 Cryptococcus gattii WM276 basi
313
EFP85158.2 Puccinia graminis basi
606
XP 001878380.1 Laccaria bicolor basi
XP 001876972.1 Laccaria bicolor basi
1000
682
XP 001876710.1 Laccaria bicolor basi
445
EXX67924.1 Rhizoglomus irregulare
Glomeromycota
1000
EXX52393.1 Rhizoglomus irregulare
ERZ98309.1 Rhizoglomus irregulare
Glomeromycota
624
518
EXX69221.1 Rhizoglomus irregulare
EPB81851.1 Mucor circinelloides muc
663
994
EPB81974.1 Mucor circinelloides muc
Mucoromycotina
1000
EPB92043.1 Mucor circinelloides muc
999
Glomeromycota 22 homologs
948
XP 002173826.1 Schizosaccharomyces japonicus yFS275 asc
1000
CAA19275.1 Schizosaccharomyces pombe asc
CAZ83738.1 Tuber melanosporum asc
1000
1000
XP 003068492.1 Coccidioides posadasii C735 delta SOWgp asc
1000
XP 001244356.2 Coccidioides immitis RS asc
998
CAL00657.1 Aspergillus niger asc
999
XP 002559347.1 Penicillium rubens asc
999
XP 380524.1 Fusarium graminearum PH1 asc
999
XP 003345655.1 Sordaria macrospora k hell asc
NP 180853.2 root Arabidopsis thaliana
0.5

(b) **RDRP**


XP 001242762.1 Coccidioides immitis RS asc
XP 003069922.1 Coccidioides posadasii C735 delta SOWgp asc
XP 003235298.1 Trichophyton rubrum asc
XP 384795.1 Fusarium graminearum PH1 asc
XP 959047.1 Neurospora crassa OR74A asc
XP 386680.1 Fusarium graminearum PH1 asc
EGU82475.1 Fusarium oxysporum asc
XP 001873562.1 Laccaria bicolor basi
XP 003026345.1 Schizophyllum commune H4 8 basi
XP 001876365.1 Laccaria bicolor basi
XP 001877148.1 Laccaria bicolor basi
XP 003036556.1 Schizophyllum commune H4 8 basi
OAD02682.1 Mucor circinelloides muc
Mucoromycotina
Rhizoglomus irregulare RdRP1
Rhizoglomus irregulare RdRP2
Rhizoglomus irregulare RdRP3
Glomeromycota
XP 003890550.1 Puccinia graminis basi
XP 003320812.2 Puccinia graminis basi
XP 003028867.1 Schizophyllum commune H4 8 basi
XP 001889791.1 Laccaria bicolor basi
XP 389252.1 Fusarium graminearum PH1 asc
XP 018257387.1 Fusarium oxysporum asc
XP 001936755.1 Pyrenophora tritici repentis Pt1CBFP asc
XP 003232635.1 Trichophyton rubrum asc
XP 001248944.2 Coccidioides immitis RS asc
XP 003071093.1 Coccidioides posadasii C735 delta SOWgp asc
XP 963405.1 Neurospora crassa OR74A asc
XP 381758.1 Fusarium graminearum PH1 asc
EGU80072.1 Fusarium oxysporum asc
XP 003038873.1 Schizophyllum commune H4 8 basi
XP 003336185.2 Puccinia graminis basi
XP 003328239.2 Puccinia graminis basi
CAB11093.1 Schizosaccharomyces pombe asc
XP 002174509.1 Schizosaccharomyces japonicus yFS275 asc
XP 964248.3 Neurospora crassa OR74A asc
XP 388892.1 Fusarium graminearum PH1 asc
EGU80489.1 Fusarium oxysporum asc
XP 001937347.1 Pyrenophora tritici repentis Pt1CBFP asc
XP 001394912.1 Aspergillus niger asc
XP 003237682.1 Trichophyton rubrum asc
XP 001246624.2 Coccidioides immitis RS asc
XP 003066590.1 Coccidioides posadasii C735 delta SOWgp asc
OAD09244.1 Mucor circinelloides muc
OAD02683.1 Mucor circinelloides muc
Mucoromycotina
NP 172932.1 root Arabidopsis thaliana
AEE82976.1 root Arabidopsis thaliana
0.5

(c) DICER

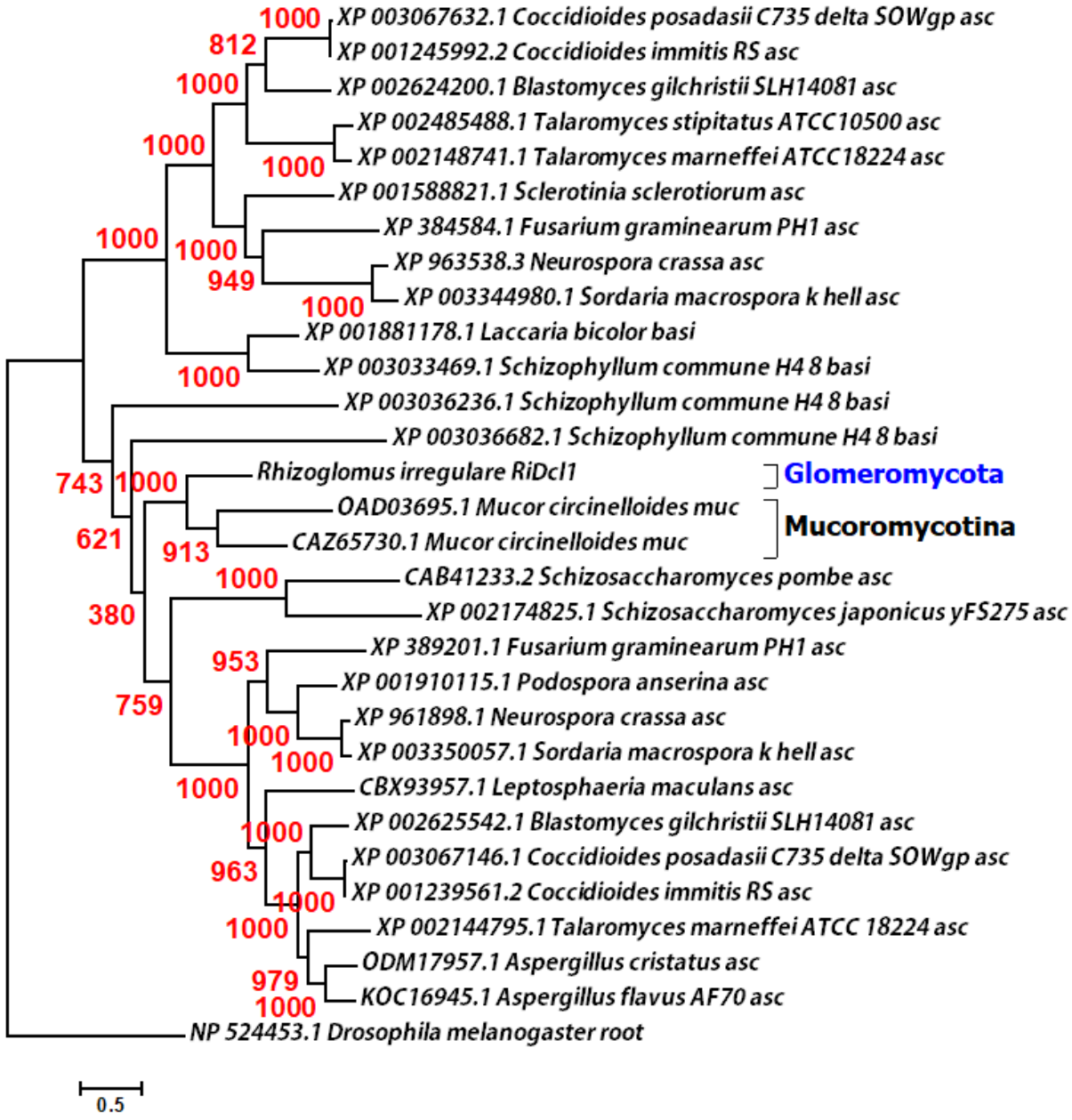


Figure 2.2. Phylogeny of three core proteins of RNAi system in AMF with other fungal species. Maximum likelihood of amino acid sequences of AGO (a), RDRP (b) and DICER (Class IV ribonuclease III) (c) were analyzed with the WAG+I+G (with four distinct gamma categories) phylogenetic model which showed the lowest AIC value. The *R. irregulare* (synonym of *Rhizophags irregularis*) and *Mucoromycotina* sequences were annotated in right side of taxon names, respectively. The numbers at branches correspond to bootstrap support values generated with 1,000 bootstrap replicates. All of the putative *R. irregulare* AGOs,

RDRPs and DICER showed the most sequence similarities with orthologs of *Mucor circinelloides* (624/1000 for AGO, 966/1000 for RDRP, and 1000/1000 for RIDCL 1) by forming single clade in each analysis. Trees (a) and (b) were rooted using *Arabidopsis thaliana*, while tree (c) was rooted using *Drosophila melanogaster*.

2.5.2. Existence of two class I ribonuclease III enzymes in *R. irregulare*

Surprisingly, we found two additional homologs of ribonuclease III (RIRNC 2 and RIRNC 3) in *R. irregulare*, using reciprocal BLAST (Table S2.2). RIRNC 2 and 3 showed relatively low e-values and low sequence similarities when they were compared with their reciprocal hits as queries (Table S2.2). Moreover, both of best BLAST hits of RIRNC 2 and RIRNC 3 were cyanobacterial ribonuclease III, while RIDCL1 had mucoromycete ortholog as the best BLAST hit. We analyzed the domain architecture of each protein homolog and we found that RIRNC 2 and RIRNC 3 have the typical protein architecture (Figure 2.1d) of prokaryotic ribonuclease III reported in bacteria (Class I ribonuclease III) (Court, et al. 2013; Liang, et al. 2014). It has been known that Fungi only possess two variants of ribonuclease III (class II and IV), which are clearly distinguished by the protein architecture from class I ribonuclease III in prokaryotes (Court, et al. 2013; Lamontagne, et al. 2001; Liang, et al. 2014). Interestingly, the structurally closest variant of eukaryotic ribonuclease III (PAC1) was detected in query species, *S. pombe* (Table S2.2). In a previous study, PAC1 showed the highest structural homology with class I ribonuclease III among other eukaryotic ribonuclease III by having one ribonuclease III (RIBOc) domain and one double stranded RNA-binding domain (DSRM) C-terminal to it (Lamontagne, et al. 2000). However, PAC1 and related homologs are clearly different in domain architecture from a typical class I ribonuclease III (RIRNC 2 and 3) detected in *R. irregulare* and other prokaryotes (Figure 2.1d) by having an additional domain in the N-

terminus (NTD domain, PDB ID: 400g), which was proven to be crucial for the enzyme's function (Lamontagne, et al. 2001). PAC1 is classified as class II eukaryotic ribonuclease III with similar proteins found in other yeasts (RNT1) (Liang, et al. 2014).

In *R. irregulare* genome, we found in total three ribonuclease III homologs which were classified into two types: class I (RIRNC 2 and 3) and class IV (RIDCL I). Full protein architecture between eukaryotic (class IV ribonuclease III) and prokaryotic (class I ribonuclease III) homologs found in *R. irregulare* were apparently different (Figure 2.1c and 2.1d). However, to clarify the differences between core RIBOc domains of class I (RIRNC2 and 3) and class IV (RIDCL1) ribonuclease III enzymes found in *R. irregulare*, we further investigated their RIBOc domains. The amino acid sequences of RIBOc domains of RIRNC 2 and 3 and RIDCL 1 of *R. irregulare* were used as queries in BLASTP searches that were performed with the entire non-redundant (nr) protein database of GenBank to investigate the taxonomical distributions of their hits. As expected, we found clear differences between class I and class IV RIBOc groups of *R. irregulare* (Table S2.4 and Figure S2.2). The taxonomical results of BLASTP of two class IV RIBOc domains (domains A and B, Figure 2.1c) from RIDCL 1 were 100% eukaryotic sequences (domain A: 1304/1304 hits and domain B: 1368/1368 hits). In contrast, 100% of the BLAST hit results of both class I RIBOc domains (RIRNC 2 and 3) were prokaryotic sequences (RIRNC 2: 9779/9779 hits, RIRNC 3: 9515/9515 hits). This is an evidence showing that class I and IV RIBOc domains from *R. irregulare* ribonuclease III homologs are clearly different in amino acid sequence level, which is in agreement with previous reports (Court, et al. 2013; Lamontagne, et al. 2001; Liang, et al. 2014). To identify the full length protein homologies of class I ribonuclease III (RIRNC 2 and 3), BLASTP-based analysis was conducted with translated *rirnc 2* and *rirnc 3* sequences (RIRNC 2 and 3, later, identical proteins were also

found in *R. irregulare* proteome dataset) as queries (Table S2.5 and Figure S2.3). The query coverage of top 100 hits ranged from 83% to 98% (RIRNC 2) and 55% to 87% (RIRNC 3) and e-values of top 100 hits were lower than $5E^{-38}$ (RIRNC 2) and $3E^{-35}$ (RIRNC 3). Expectedly, 100% of top 5,000 hits of BLASTP results for both RIRNC 2 and 3 were class I prokaryotic ribonuclease III enzymes, and top 100 hits of BLASTP results were mostly from cyanobacteria clades, implying the high level of sequence similarity between homolog from *R. irregulare* and cyanobacteria. To diagnose the possibilities of mis-identification of RIRNC 2 and 3 homologs, we did detailed inspection of genes of RIRNC 2 and 3 by cross-checking all of the sequence datasets publicly available for *R. irregulare* whose genome was sequenced two times in different studies using different approaches (Lin, et al. 2014; Tisserant, et al. 2013). Later, Tisserant, et al. (2013) resequenced and updated the dataset of *R. irregulare* genome (version 2.0 in JGI) with a coverage of 50X. The RIRNC 2 and 3 sequences we found in two versions 1.0 and 2.0 datasets published by (Tisserant, et al. 2013) as well as in the dataset reported by (Lin, et al. 2014) where the genome sequencing was done using four single nuclei of *R. irregulare* (coverage of 60X). Moreover, we retrieved hits of ESTs of *R. irregulare* published in previous studies (Tisserant, et al. 2012; Tisserant, et al. 2013). Both coding genes of RIRNC 2 and 3 were located in assembled genome scaffolds of *R. irregulare* genome version 2.0 (Tisserant, et al. 2013). The coding gene of RIRNC 2 was located from 722662 to 723453 bp in genome scaffold 4 (scaffold length: 939,913 bp in version 2.0 of *R. irregulare* genome available at JGI website) without intron, while *rirnc 3* was located from 391302 to 392174 bp of genome scaffold 82 (length: 423,271 bp, in version 2.0) without intron. The well-defined neighboring genes in both sides (5' upstream and 3' downstream) of *rirnc 2* and *3* were also investigated respectively. Phylogenetic analysis was conducted together with BLAST searches in GenBank of each

sequence (threshold e-value = E^{-20} and query coverage over 80%). The phylogenetic analysis of each gene with comparison of other prokaryotic or eukaryotic orthologs as outgroups showed that all of the tested genes were typical fungal genes (Figure S2.4). Taken together, these evidences clearly confirmed that class I homolog sequences are inserted in *R. irregulare's* genome.

2.5.3. Expression confirmation of class I ribonuclease III enzyme

We found expression evidence of *rirnc 3* in EST datasets. However, at the time with initial EST dataset of *R. irregulare*, we didn't find any transcripts of *rirnc 2* matching with these EST libraries. We therefore experimentally verified its expression status using RT-PCR followed by Sanger sequencing on *R. irregulare* mRNA. The primer set was designed to target open reading frame (ORF) region of *rirnc 2* sequence (Table S2.1). We amplified a partial sequence of *rirnc 2* (110 bp) from cDNA of *R. irregulare*. The obtained sequence was identically matched with candidate sequence of *rirnc 2* retrieved by sequence analysis. The updated *R. irregulare* EST dataset available at the JGI had an identical match with *rirnc 2*, confirming the expression of the gene. Furthermore, we tested the gene expression of two types of ribonuclease III family enzymes (class IV ribonuclease III (*ridcl 1*) and class I ribonuclease III (*rirnc 2*)) in *R. irregulare* under two different life cycle stages (spore germination stage and mycorrhization stage) of AMF. Relative expression of RNA was tested with qRT-PCR using TaqMan assay and two housekeeping genes (*18s rrna* and *ubiquitin*). The primers and probes were designed and successfully tested (Table S2.2). Figure 3 shows the different levels of the expressions of *ridcl 1* and *rirnc 2* in *R. irregulare* isolate DAOM197198. The patterns of normalized results using both housekeeping genes as reference were similar. Interestingly, the

expression of the gene encoding class I type ribonuclease III enzyme (*rirnc 2*) was significantly higher than the expression of the gene encoding class IV type ribonuclease III enzyme (*ridcl 1*) during both spore germination (approximately 1.9-fold increase) (*P=0.024253* (with reference gene: *18s rrna*) and *P=0.027964* (with *ubiquitin*), Student's two-tailed T-test) and mycorrhization stages (approximately 2.7-fold increase) (*P=0.017255* (*18s rrna*) and *P=0.021482* (*ubiquitin*), Student's two-tailed T-test). The difference between expressions of *rirnc 2* and *ridcl 1* becomes larger in mycorrhization phase compared to spore germination phase. However, the expression level differences of same gene under different lifecycle stages were not significant in our experiments.

Figure 3

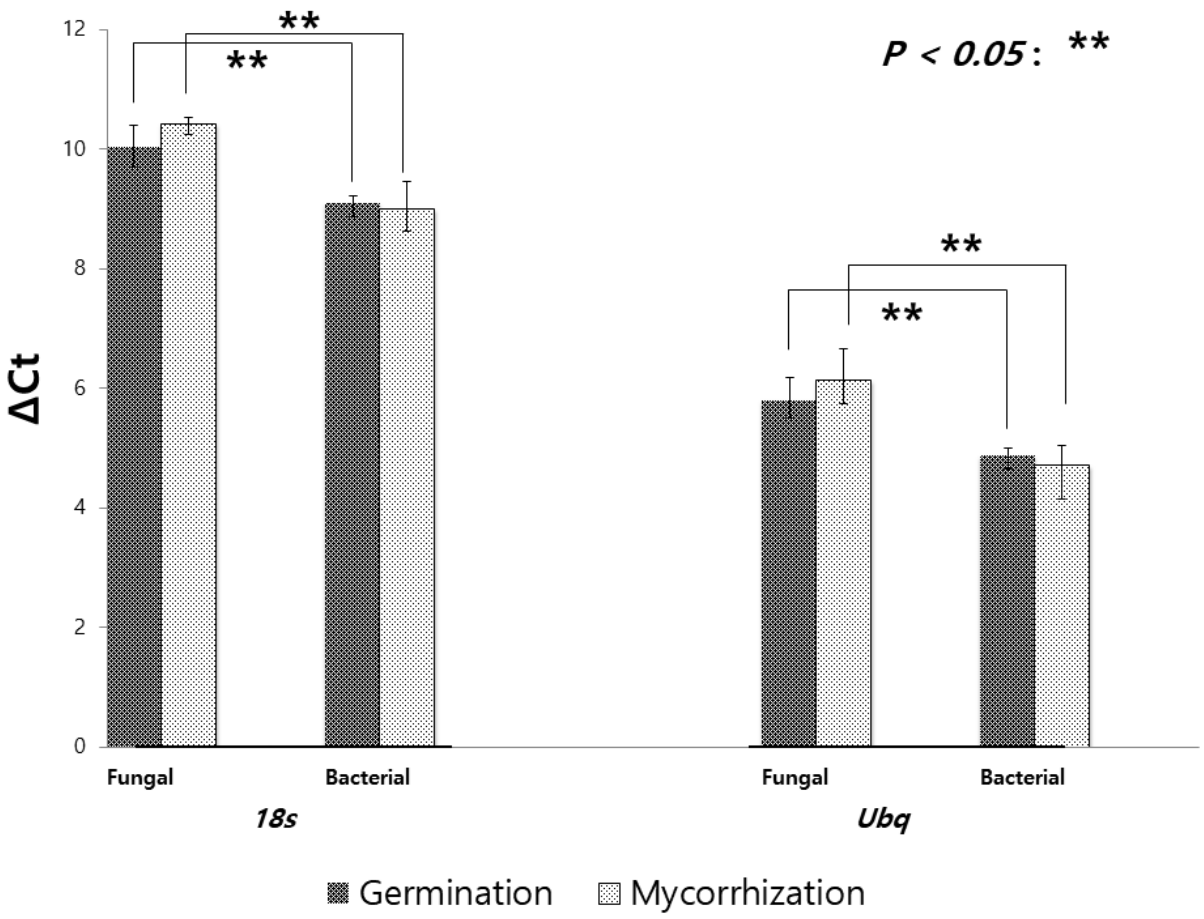


Figure 2.3. Expressions of class I (bacterial) and class IV (fungal) ribonuclease III family proteins coding mRNA (*ridcl 1* and *rirnc 2*) in two different phases of *R. irregulare* life cycle. Relative quantification of gene expression of *ridcl l* and *rirnc 2* using TaqMan assay and two different housekeeping genes (*18 rrna* (left), *ubiquitin* (right), in germination and mycorrhizations phases. Y-axis shows threshold cycle difference between dicer genes and reference genes (calculated by $Ct_{Target} - Ct_{Reference}$ and converted to fold changes by formula $2^{-\Delta Ct}$). Black bars represent spore germination phase, while gray bars represent mycorrhization phase. The error bar indicates minimum and maximum values among the replicates. The statistical significance ($P < 0.05$) was marked on figure with star.

2.5.4. Unique and high level of conservation of cyanobacterial ribonuclease III enzymes in AMF

Up to date, there has been no report of the existence of class I ribonuclease III in Eukaryotes (Lamontagne, et al. 2001; Liang, et al. 2014). To confirm our result, we performed repeated sequence searches in all of the fungal genome and proteome databases available including GenBank and we didn't find any evidence of the existence of the class I ribonuclease III enzyme coding genes (*rirnc 2* and *3*) in any fungal genome except that of *R. irregulare*. No homologous sequence was found even in the basal fungal subphylum *Mucoromycotina*, which yielded the closest homology for all three fungal core proteins related with RNAi. Because class I type ribonuclease III enzyme encoding genes were uniquely found in *R. irregulare*, we investigated whether catalytic domains (RIBOc, ribonuclease III domain) of RIRNC are in other AMF species (Figure 2.4). Four other AMF species (*Rhizoglomus clarum* (syn, *Rhizophagus clarus*), *Rhizoglomus diaphanum* (syn, *Rhizophagus diaphanous), Claroideoglomus etunicatum, Scutelospora calospora*) of which genomic data are available in our lab (data not published) were analyzed. We were unable to detect the hits of RIBOc domain of RIRNC 2 because of the incompletion of other species' genome data sets. On the other hand, partial sequences of RIBOc domain of RIRNC 3 were detected in each of the four AMF species genome. All of four sequences of *rirnc 3* homologs were deposited and are publicly available in GenBank under the accession numbers MF926648-MF926651. Figure 4 shows conserved RIBOc domain of class I ribonuclease III enzymes from different AMF species (> 80 a.a.) with comparison of same domains in class II ribonuclease III enzymes of *S. pombe* and *S. cerevisiae*, which were reported to be the most similar fungal ribonuclease III enzymes to class I ribonuclease III enzymes (Liang, et al. 2014). The shared top hit from BLASTP analysis of *R. irregulare* class I ribonuclease III enzyme (*Microcoleus sp. PCC7113* (cyanobacteria), E-value = $2E^{-34}$ (in BLAST result of RIRNC 2) and $4E^{-38}$ (in BLAST result of RIRNC3)), was also compared for their sequence

similarity). Interestingly, amino acid residues Glu-Phe-Leu-Gly-Asp (from 48 to 52 a.a. in figure 2.4) were highly conserved among all of the species analyzed so far for bacterial ribonuclease III (total 52 species). AMF species sequences showed high levels of sequence similarities between each species compared to their similarity with yeasts or cyanobacteria. Three species of the *Rhizoglomus* genus showed 100% identical matching along the amino acid residues from 17 to 96, while *C. etunicatum* (at a single point, two amino acid residues) and *S. calospora* (at three points, four amino acid residues) showed minor divergences. *rirnc 2* of *R. irregulare* showed a higher total number of differences at amino-acid positions (22/80) compared with *rirnc 3* of *R. irregulare*, while highest total differences at amino-acid positions for a *rirnc 3* homolog was 4/80 for *rirnc 3* from *S. calospora*.

Figure 2.4

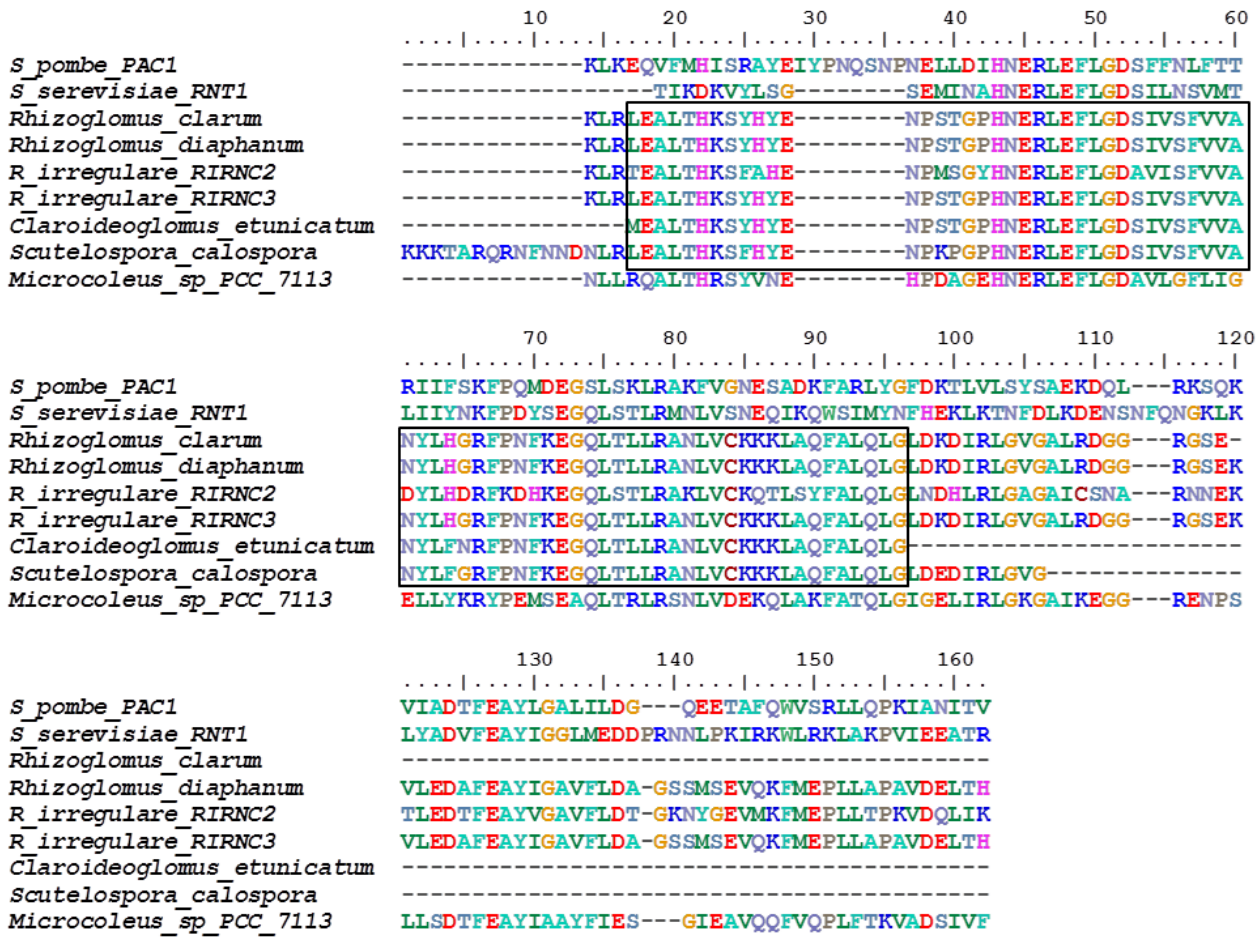


Figure 2.4. Multiple Sequence Alignment of conserved RIBOc domain from class I ribonuclease III among Glomeromycota and cyanobacteria, comparing with those of the protein structurally closest eukaryotic class II ribonuclease III from yeasts. Protein sequences from five Glomeromycota species (*R. irregularis*, *Rhizoglomus clarum, Rhizoglomus diaphanum, Claroideoglomus etunicatum, Scutelospora calospora*), two yeast species (*S. pombe* and *S. cerevisiae*) and top hit of BLASTP analysis (*Microcoleus sp. PCC7113* (cyanobacteria), $4E^{-38}$) were used. Similar amino acid residues were marked with same color. Glomeromycota species were labeled with black rectangle (from residue 17 to 96, 80 a.a.). Amino acid residue Glu-Phe-Leu-Gly-Asp (from 48 to 52 a.a. in the figure) were highly conserved among all the species analyzed so far for class I ribonuclease III (total 52 species).

We further conducted detailed phylogenetic analysis with these RIBOc domains from class I and class II ribonuclease III (Figure S2.5). All of the class I ribonuclease III sequences were clearly separated from those of class II ribonuclease III enzyme (high bootstrap supporting:

994/1000). The sequence variation was also observed in the core catalytic domain (RIBOc). Taxonomical BLAST results described above (Table S2.5 and Figure S2.3), along with phylogenetic and amino acid sequence analyses support that discovery of RIRNC 2 and 3 in *R. irregulare* was not a result of false detection of class II ribonuclease III homologs. In Figure S5, class I ribonuclease III sequences of Glomeromycota taxa were clustered as a monophyletic clade with high level of sequence similarity and they separated from other class I ribonuclease III sequences (927/1000). Inside of the monophyletic clade of Glomeromycota sequences, internal monophyletic clustering of RIRNC 3 homologs was observed, apart from RIRNC 2 sequence from *R. irregulare* (908/1000). Three species of the *Rhizoglomus* genus (*R. irregulare*, *R. clarum* (syn, *Rhyzophagus clarus*), *R. diaphanum* (syn, *Rhyzophagus diaphanous*)) showed high percentage of sequence similarity of RIBOc domain of RIRNC 3, and they formed a phylogenetic clade (738/1000), distinguished from sequences of *C. etunicatum* and *Scutelospora calospora* which belong to different families in Glomeromycota.

We further inferred a phylogenetic tree of full length sequence of RIRNC 2 and 3 of *R. irregulare* and those of class I ribonuclease III orthologs from Bacteria, where Archaea sequence was used as an outgroup (Figure 2.5). RIRNC 3 homologous sequences from other AMF species were excluded from this analysis because they were partial sequences (Figure S2.5). As expected, both RIRNC 2 and 3 clustered with Cyanobacteria class I ribonuclease III sequences with a high bootstrap support (945/1000). Within the *R. irregulare*-cyanobacterial group, RIRNC 2 and 3 formed a single clade which showed 100% bootstrap support (Figure 2.5). This result supports that the event of HGT was ancient.

Figure 2.5

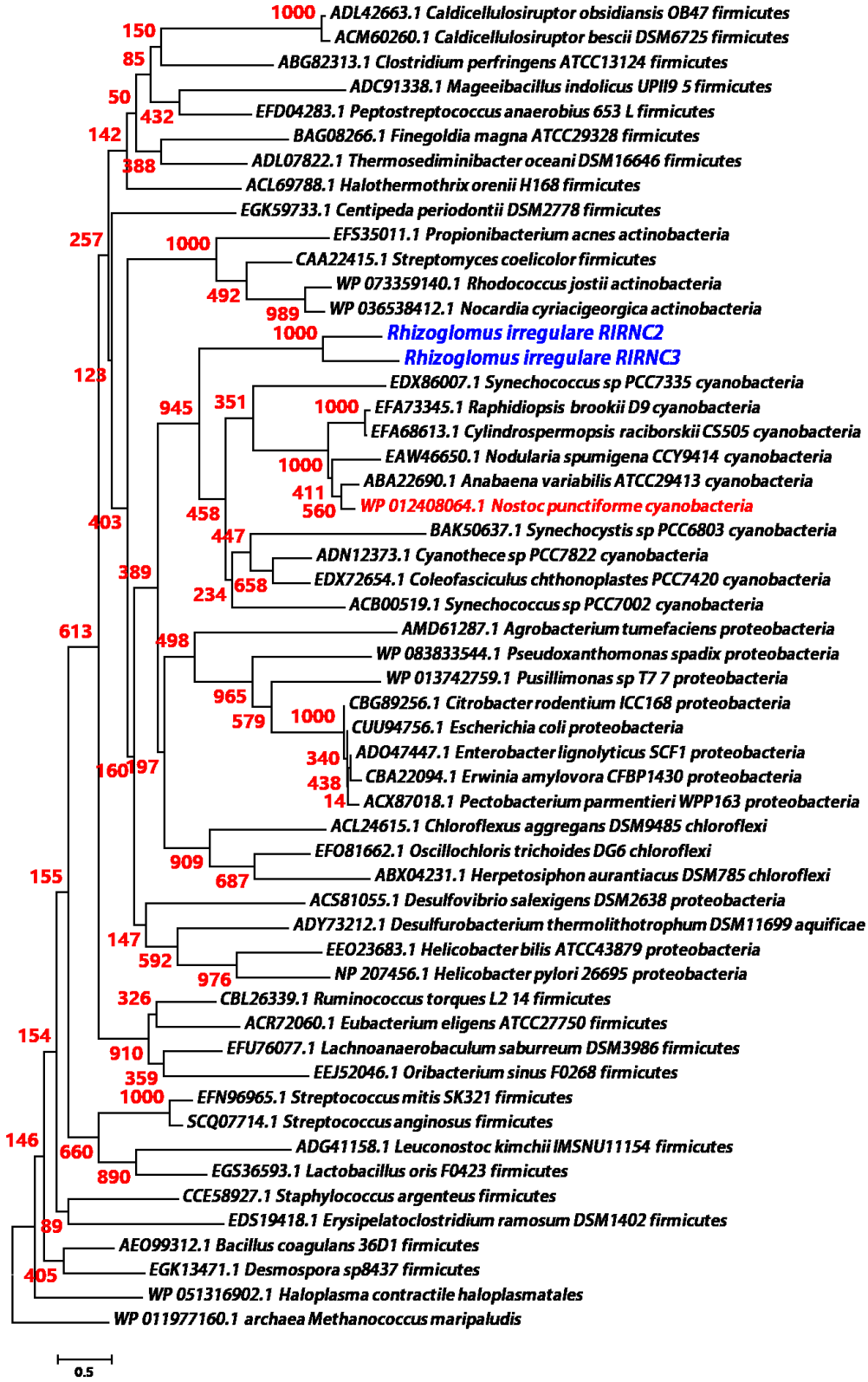

1000 ADL42663.1 Caldicellulosiruptor obsidiansis OB47 firmicutes
150 ACM60260.1 Caldicellulosiruptor bescii DSM6725 firmicutes
85 ABG82313.1 Clostridium perfringens ATCC13124 firmicutes
50 ADC91338.1 Mageeibacillus indolicus UPII9 5 firmicutes
432 EFD04283.1 Peptostreptococcus anaerobius 653 L firmicutes
142 BAG08266.1 Finegoldia magna ATCC29328 firmicutes
388 ADL07822.1 Thermosediminibacter oceani DSM16646 firmicutes
ACL69788.1 Halothermothrix orenii H168 firmicutes
EGK59733.1 Centipeda periodontii DSM2778 firmicutes
257 EFS35011.1 Propionibacterium acnes actinobacteria
1000 CAA22415.1 Streptomyces coelicolor firmicutes
492 WP 073359140.1 Rhodococcus jostii actinobacteria
989 WP 036538412.1 Nocardia cyriacigeorgica actinobacteria
1000 Rhizoglomus irregulare RIRNC2
Rhizoglomus irregulare RIRNC3
128 EDX86007.1 Synechococcus sp PCC7335 cyanobacteria
945 351 1000 EFA73345.1 Raphidiopsis brookii D9 cyanobacteria
EFA68613.1 Cylindrospermopsis raciborskii CS505 cyanobacteria
1000 EAW46650.1 Nodularia spumigena CCY9414 cyanobacteria
411 ABA22690.1 Anabaena variabilis ATCC29413 cyanobacteria
458 560 WP 012408064.1 Nostoc punctiforme cyanobacteria
403 447 BAK50637.1 Synechocystis sp PCC6803 cyanobacteria
ADN12373.1 Cyanothece sp PCC7822 cyanobacteria
389 658 EDX72654.1 Coleofasciculus chthonoplastes PCC7420 cyanobacteria
234 ACB00519.1 Synechococcus sp PCC7002 cyanobacteria
613 AMD61287.1 Agrobacterium tumefaciens proteobacteria
498 WP 083833544.1 Pseudoxanthomonas spadix proteobacteria
WP 013742759.1 Pusillimonas sp T7 7 proteobacteria
965 CBG89256.1 Citrobacter rodentium ICC168 proteobacteria
579 1000 CUU94756.1 Escherichia coli proteobacteria
340 ADO47447.1 Enterobacter lignolyticus SCF1 proteobacteria
160 197 438 CBA22094.1 Erwinia amylovora CFBP1430 proteobacteria
14 ACX87018.1 Pectobacterium parmentieri WPP163 proteobacteria
ACL24615.1 Chloroflexus aggregans DSM9485 chloroflexi
909 EFO81662.1 Oscillochloris trichoides DG6 chloroflexi
155 687 ABX04231.1 Herpetosiphon aurantiacus DSM785 chloroflexi
ACS81055.1 Desulfovibrio salexigens DSM2638 proteobacteria
ADY73212.1 Desulfurobacterium thermolithotrophum DSM11699 aquificae
147 EEO23683.1 Helicobacter bilis ATCC43879 proteobacteria
592 NP 207456.1 Helicobacter pylori 26695 proteobacteria
976
326 CBL26339.1 Ruminococcus torques L2 14 firmicutes
ACR72060.1 Eubacterium eligens ATCC27750 firmicutes
154 910 EFU76077.1 Lachnoanaerobaculum saburreum DSM3986 firmicutes
359 EEJ52046.1 Oribacterium sinus F0268 firmicutes
1000 EFN96965.1 Streptococcus mitis SK321 firmicutes
SCQ07714.1 Streptococcus anginosus firmicutes
146 660 ADG41158.1 Leuconostoc kimchii IMSNU11154 firmicutes
890 EGS36593.1 Lactobacillus oris F0423 firmicutes
CCE58927.1 Staphylococcus argenteus firmicutes
89 EDS19418.1 Erysipelatoclostridium ramosum DSM1402 firmicutes
AEO99312.1 Bacillus coagulans 36D1 firmicutes
405 EGK13471.1 Desmospora sp8437 firmicutes
WP 051316902.1 Haloplasma contractile haloplasmatales
WP 011977160.1 archaea Methanococcus maripaludis
0.5

Figure 2.5. Phylogenetic analysis of class I ribonuclease III enzymes in of *R. irregulare*. Maximum likelihood of amino acid sequences of class I ribonuclease III enzymes were analyzed with the LG+I+G (with four distinct gamma categories) phylogenetic model which showed the lowest AIC value. The tree was rooted with class I ribonuclease sequence from archaea (*Methanococcus maripaludis*).The *R. irregulare* sequences were colored in blue. A cyanobacterial species (*Nostoc punctiforme*) which were known to form symbiosis with one basal AMF species (*Geosiphon pyriforme*) was colored in red. Cyanobacteria sequences were annotated in right side of taxon names. The numbers at branches correspond to bootstrap support values generated with 1,000 bootstrap replicates. Both RIRNC 2 and 3 of *R. irregulare* strongly formed single clade with orthologs of cyanobacteria group with high bootstrap supporting (945/1000), clearly separated from protein groups of other bacteria.

2.6. Discussion

2.6.1. Conservation of RNAi system in the model AMF genome

The RNAi system was highly conserved in the model AMF *R. irregulare* with respect to all three of its core proteins (Dicer, RdRP, and Agos). Three core proteins showed high levels of sequence similarity with orthologs from *Mucoromycotina*, a basal fungal subphylum which was suspected to be a closely related with *Glomeromycota* in other studies (Nadimi, et al. 2012; Nadimi, et al. 2016; Redecker, et al. 2013).

It has been well described that natural selection ensures that each organism is pressured to optimize its RNAi system to counter the particular challenges it faces (Ketting 2011). In spite of the close homology of core proteins with sister subphylum Mucoromycotina, RNAi system of *R. irregulare* had two interesting points. First, as shown in Table 1, the *R. irregulare* genome harbors large numbers of AGO homologs compared to the general fungal species including the model fungus *N. crassa* (24 more in number), or its closest neighbor mucoromycete *M. circinelloides* (23 more in number) (Dang, et al. 2011; Torres-Martinez and Ruiz-Vazquez 2016). In contrast to the cases in non-eukaryotic genomes, eukaryotic genome size and the

number of genes are not correlated (Hou and Lin 2009). Thus, the expansion of the number of AGO homologs couldn't be simply explained by the large genome size of *R. irregulare*. Different with the case of AGO, the number of homologs of the other crucial core protein (RDRP) did not show differences compared to *N. crassa* or *M. circinelloides.* The number of AGO homologs in the *R. irregulare* genome was 26, close to that reported in the model nematode worm, *Caenorhabditis elegans*, which had 27 homologs of AGOs (Yigit, et al. 2006). Interestingly, in the case of *C. elegans* and its relative nematodes, AGO diversity was related with symbiosis since there was additional AGO protein diversity found in parasitic nematodes compared to free-living relatives (Buck and Blaxter 2013). The variety of symbioses, including parasitism, require complex sRNA regulation in both symbiotic partners, and AGO is one of the crucial enzymes in this process (Formey, et al. 2016; Katiyar-Agarwal and Jin 2010; Weiberg, et al. 2015). Considering its complex symbiotic relationship with other fungi, bacteria and obligatory symbiosis with a plant host, we hypothesized that the high number (26) of AGO homologs in *R. irregulare* reflects the complex sRNAs regulation mechanism related with symbiosis, as in nematode worms. Further investigation of protein function of each AGO of *R. irregulare* related with its symbiosis will shed light on this hypothesis. At the same time, it has been shown that *R. irregulare* isolates harbor abnormally high number of gene copies compared to other fungi, at least 76 homologs of Mating-type-high-mobility group (*Sex M* and *Sex P* genes) (Halary, et al. 2013; Riley, et al. 2014). These copies also showed a high intraspecific genetic polymorphism among *R. irregulare* isolates. Riley, et al. (2014) explained that intragenomic gene duplications play an important role in the expansion of copy number of Mating-type-high-mobility group in *R. irregulare* isolates. Thus, we hypothesize that the high copy number of AGO genes in *R. irregulare* could also be explained by gene duplications originating from intra-

isolate genetic polymorphism in Glomeromycota (Hijri and Sanders 2005; Marleau, et al. 2011; Riley, et al. 2014; Stukenbrock and Rosendahl 2005).

The second interesting feature observed in *R. irregulare* RNAi system is the number of DICER (class IV ribonuclease III) homologs. In contrast to the large number of AGO homologs, *R. irregulare* genome only harbored a single class IV ribonuclease III homolog (RIDCL 1), similar to the cases of *S. cerevisiae* or *S. pombe* genomes. However, in yeasts, an additional variant of eukaryotic class II ribonuclease III that digests double strand RNA was reported (Lamontagne, et al. 2001). We couldn't find any class II homologs in *R. irregulare*. Strikingly, instead, we discovered two additional class I ribonuclease III homologs (RIRNC 2 and 3) in the *R. irregulare* genome. Domain architectures of RIRNC 2 and 3 were identical with typical prokaryotic class I ribonuclease III enzymes which have the function to digest dsRNA and they are involved in viral defense mechanism as well as transcriptome processing in Bacteria (Nicholson 1999). ESTs based sequence analysis and Sanger sequencing confirmed the expression of all class I and class IV ribonuclease III mRNAs (*ridcl 1, rirnc 2*, and *rirnc 3*) in *R. irregulare*. It is well known that class I ribonuclease III has a function in dimeric form, while class IV ribonuclease III acts as monomeric form (Court, et al. 2013; Lamontagne, et al. 2001). RT-qPCR experiment showed that a representative of class I ribonuclease III mRNA (*rirnc 2*) had higher expression level than the class IV ribonuclease III mRNA (*ridcl 1*) in two contrasting phases of *R. irregulare* life cycle (spore germination and mycorrhization). The detected expression level difference could be due to the homodimer characteristic of class I ribonuclease III which needs double of the population of protein chains for forming homodimer to have catalytic function compared to monomeric class IV ribonuclease III (Sun and Nicholson 2001). At the same time, the heterodimerization of ribonuclease III in *E. coli* was

also reported and demonstrated (Conrad, et al. 2002; Meng and Nicholson 2008). The further investigation of protein level interaction between two class I ribonuclease III (RIRNC 2 and 3) in *R. irregulare* is required for understanding their function. The expression levels difference between class I and class IV ribonuclease III mRNAs increased in symbiosis phase (Figure 2.3).

Figure 2.6

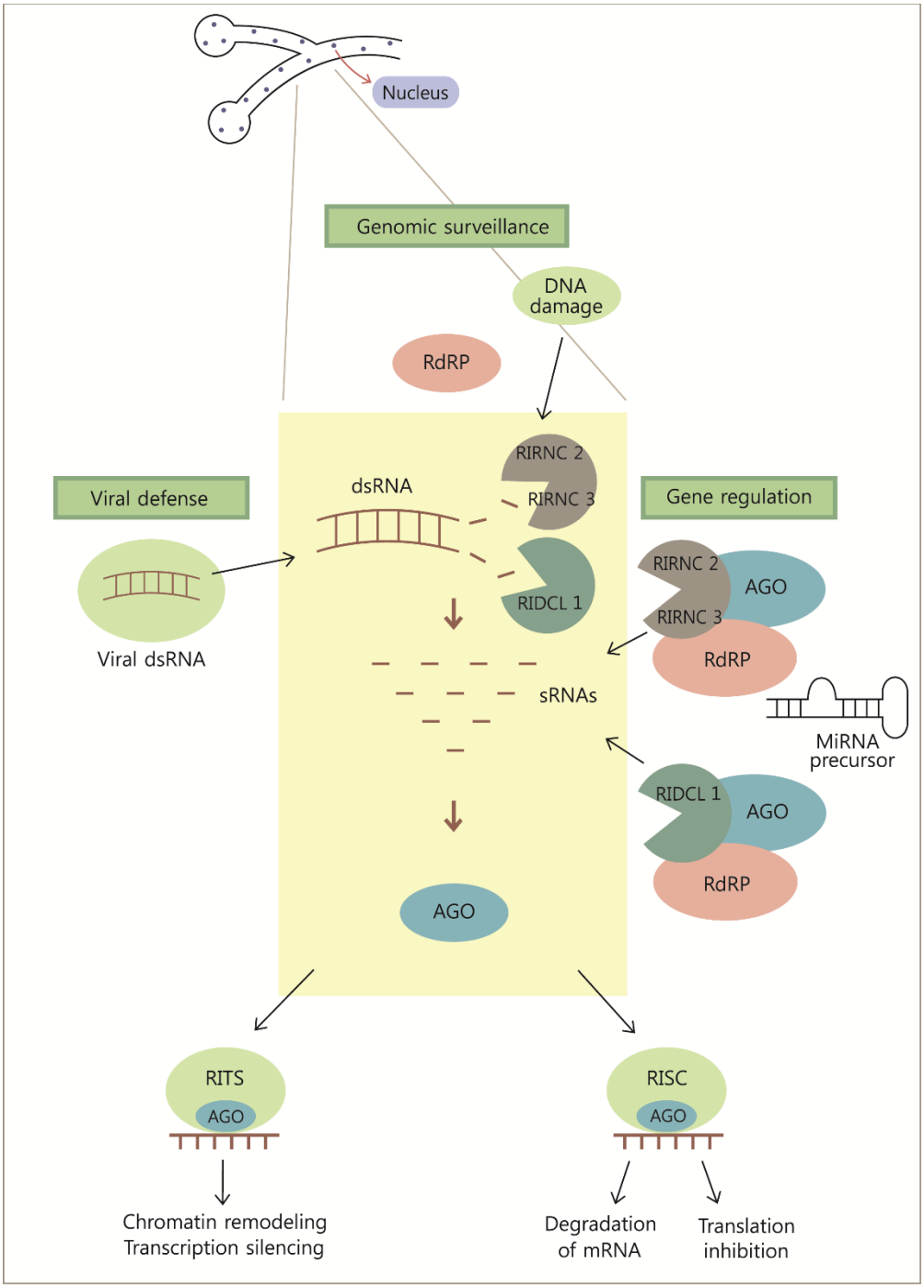


Figure 2.6. Schematic diagram of *R. irregulare* putative RNAi system consisted on three core proteins. This diagram was built based on the study of RNAi system in the ascomycete *N. crassa* (Dang, et al. 2011). Two different types of ribonuclease III proteins were noted as RIRNC 2 and 3 (class I ribonuclease III) and RIDCL 1 (class IV ribonuclease III).

Considering the core gene conservation and related two interesting features described above, we propose a putative mechanism of RNAi system in the model AMF, *R. irregulare* (Figure 2.6). Based on known functions in prokaryotic orthologous (Knip, et al. 2014; Lamontagne, et al. 2001; Nicholson 1999), we hypothesize that class I ribonuclease III enzymes of *R. irregulare* are also functional in the processing of dsRNA originated from exogenous RNAs or endogenous transcripts. sRNAs can be generated by class I and class IV ribonuclease III enzymes (RIDCL 1, RIDCL 2 or RIDCL 3) with or without putative RDRPs. This sRNA biogenesis could directly function as a defense mechanism against viral attack by degrading their dsRNA. Biosynthesized sRNAs could be loaded to the RISC or RITS complexes, which have putative AGOs as core components. By loading of sRNAs, RISC or RITS complex can recognize and approach to target RNAs which are complementary to guide sRNAs to trigger the silencing.

2.6.2. Class I ribonuclease III found in *R. irregulare* and HGT event

So far, there has been no report for the existence of class I (prokaryotic) ribonuclease III in Eukaryotes (Lamontagne, et al. 2001; Liang, et al. 2014). In the kingdom Fungi, the closest type of class I ribonuclease III enzyme is class II which was reported in *S. serevisiae* (RNT1) and *S. pombe* (PAC1) (Lamontagne, et al. 2000). RNT1 and PAC1 showed structural homology with class I ribonuclease III by having one ribonuclease III (RIBOc) domain and one double stranded RNA-binding domain (DSRM) in C-terminal. However, RNT1 and PAC1 are different in domain architecture of class I ribonuclease III by having an additional domain in the N-terminus (NTD, PDB ID: 400g), which was proven to be crucial for the enzyme's function (Lamontagne, et al. 2001). RIRNC 2 and 3 found in *R. irregulare* were clearly

distinguished from RNT1 or PAC1 by lacking eukaryotic N-terminal domain and they showed identical protein architecture of prokaryotic class I ribonuclease III. Moreover, the RIBOc domains of class II and class I ribonuclease III were also clearly distinguished (Figure S2.5). Thus, we concluded that RIRNC 2 and 3 encoding genes found in *R. irregulare* are likely acquired by horizontal gene transfer from a prokaryotic ancestor related to Cyanobacteria (Table S2.2 and Figure 2.5). At the same time, even though both RIRNC 2 and 3 showed their high level of sequence similarity with cyanobacterial ribonuclease III enzymes respectively, they were different in length (RIRNC 2 - 264 a.a., RIRNC 3 - 291 a.a.), and sequence identity is only 52% between each. Moreover, the core catalytic domains (RIBOc) of RIRNC 2 and 3 were also divergent as shown by the phylogenetic tree (Figure S5). At the same time, the coding genes for RIRNC 2 and 3 were located in two different genome scaffolds (*rirnc 2*: scaffold 4 (939,913 bp) and *rirnc 3*: scaffold 82 (423,271 bp)) and there is no evident linkage between these two scaffolds. Two possible scenarios can be proposed: 1) a gene duplication may have occurred after HGT of a gene from an ancestral cyanobacterium, followed by the diversification and translocation of the *rirnc* copies; 2) two genes were transferred from a cyanobacterial ancestor to *R. irregulare* by two ancient independent events of HGT.

2.6.3. HGT between intimately interacting species

Events of HGT are ubiquitous in eukaryotic microorganisms which are closely interacting. In the case of diploid commercial wine yeast, *Saccharomyces cerevisiae EC1118,* a region of the genome involved in key wine fermentation function was originated from non-*Saccharomyces* species, *Zygosaccharomyces bailii* which shares habitat (Novo, et al. 2009). Another case of HGT occurred between oomycetes and fungi, which related with a

phytopathogenic function to colonize their host plant (Soanes and Richards 2014). HGT also takes place frequently between symbiotic partners. Many examples of HGT between endosymbiotic prokaryotes and their hosts have been reported, including *Mollicutes-related endobacteria*, the bacterium which forms endosymbiosis with Glomeromycota and Mucoromycotina (Kuo 2015; Naito, et al. 2015; Torres-Cortes, et al. 2015). Authors found between 3% and 5% of protein coding genes in *Mollicutes-related endobacterial* genomes, were horizontally transferred from the ancestors of Glomeromycota and Mucoromycotina (Torres-Cortes, et al. 2015). This example involving *Mollicutes-related endobacteria* and AMF ancestor is an example of HGT from AMF ancestor to endosymbiotic bacterial ancestor. However, we cannot ignore the possibility of HGTs in opposite direction: from bacteria to AMF. Indeed, in nature, HGT frequently take place in bi-directional way between symbiotic partners (Soucy, et al. 2015).

In our study, we found the event of HGT in core enzyme (ribonuclease III) of sRNA regulation which is reported to be linked with symbiosis as well as gene regulation of the organism (Formey, et al. 2016; Thiebaut, et al. 2014; Yan, et al. 2015). Moreover, the HGT has likely occurred long time ago (with cyanobacterial ancestor) resulting in the occurrence of two additional enzymes of class I ribonuclease III in *R. irregulare* genome. At the same time, class I ribonuclease III homologs have never been reported in other fungi or other eukaryotes. Our finding of RIRNC 2 and RIRNC 3 in *R. irregulare* which both show the closest sequence similarities with cyanobacterial class I ribonuclease III enzymes (Figure 2.5) could probably be related with an intimate relationship of *R. irregulare* ancestor with cyanobacterial ancestor.

Observation of class I ribonuclease III in *R. irregulare* is unique, considering that the enzyme's distribution has been known to be limited within prokaryotes (Lamontagne, et al. 2001;

Liang, et al. 2014). However, this observation seems not to be restricted to *R. irregulare* within the phylum Glomeromycota. The homologs of class I ribonuclease III (RIRNC 3) were found among various AMF species in phylum Glomeromycota (*Rhizoglomus irregulare*, *Rhizoglomus clarum, Rhizoglomus diaphanum, Claroideoglomus etunicatum, Scutelospora calospora*) (Figure S2.4). On the contrary, the two novel homologs (RIRNC 2 and 3) of class I ribonuclease III found in *R. irregulare* do not exist in other fungal groups including *Mucoromycotina*, which has the closest homology of RNAi protein components with AMF. Moreover, our result of RIRNC 3's RIBOc domain sequence comparison and phylogenetic analysis (Figure 2.5 and Figure S2.5) showed a high level of sequence similarities between AMF species in same genus or order as well as minor variation of sequences of AMF species belonging to different orders. The result may reflect the long time evolution of *rirnc 3* gene along the diversification of AMF. Indeed, in phylogenetic analysis of class I ribonuclease III (Figure 2.5), RIRNC 2 and 3 were still separated from the group of sequences from cyanobacteria, implying the event of HGT has not recently occurred (which would result in higher sequence similarity with modern cyanobacterial ribonuclease III).

Based on RIRNC 3 homologs distribution among five AMF species, a hypothetical HGT event that may occur during the evolution of Glomeromycota is inferred and shown in Figure 2.7.

Figure 2.7

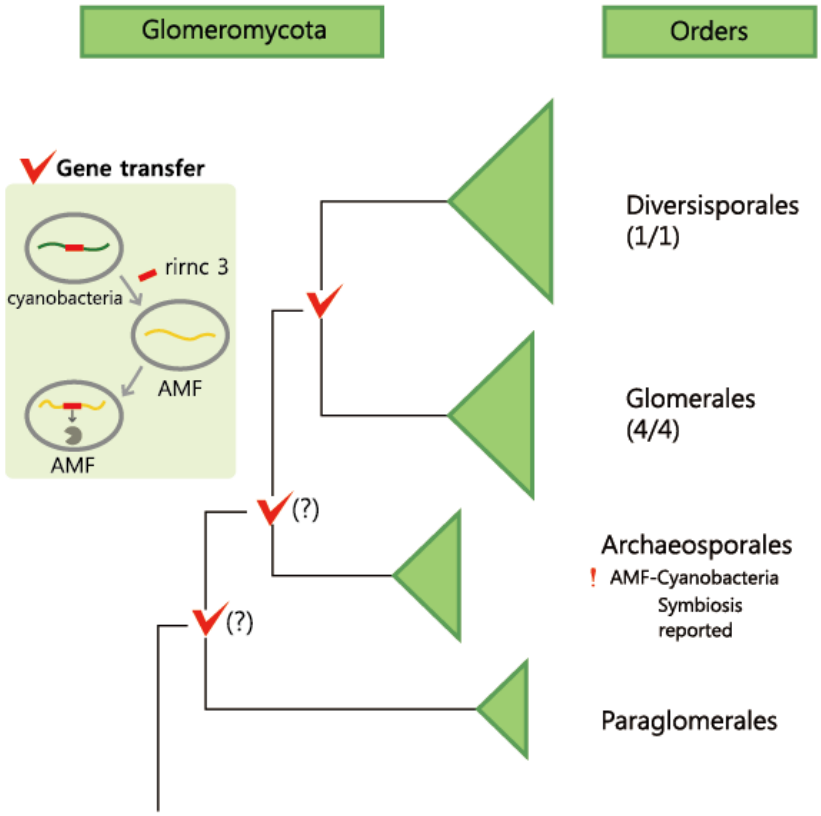


Figure 2.7. Summary of putative HGT of cyanobacterial ribonuclease III (*rirnc 3*) that may occur during the diversification of phylum *Glomeromycota*. In our study, four species from a order, *Glomerales* (*R. irregular*, *R. clarum* (syn, *R. clarus*), *R. diaphanum* (syn, *R. diaphanous)* and *C. etunicatum*) and one species (*Scutelospora calospora*) from another order, Diversisporales were investigated. The numbers indicated below the order names represent (Number of species possesses *rirnc 3* homolog / Number of species genome analyzed). All of 5 different AMF species analyzed possessed cyanobacterial ribonuclease III (*rirnc 3*) homologs. Based on the dataset used in our analyses, the event of HGT may have occurred at least before the common ancestor of orders *Diversisporales* and *Glomerales* during *Glomeromycota* evolution (indicated by red symbol). Interestingly, AMF-cyanobacteria symbiosis was reported in basal order: *Archaeosporales*.

In our study, four species from an order, *Glomerales* (*R. irregulare*, *R. clarum*, *R. diaphanum* and *C. etunicatum*) and one species (*S. calospora*) from another order, *Diversisporales* were searched for RIRNC 3 homologs possession, and all of five investigated species possessed sequences of RIRNC 3 homologs. We couldn't specify if HGT occurred in the common ancestor of Glomeromycota, because of the limited availability of AMF genomes of other two basal orders (*Paraglomerales* and *Archaeosporales*) in Glomeromycota. However, according to our result, HGT of RIRNC 3 coding gene from cyanobacterial ancestor occurred at least before the divergence of orders *Diversisporales* and *Glomerales* from their common ancestor. What would be the corresponding knowledge for this unique HGT regarding the conservation of a crucial gene of sRNA regulation in AMF?

2.6.4. Putative origin of HGT from cyanobacteria to AMF

As an obligatory symbiont of its host plant, AMF depends for its supplying of carbon nutrition from the photosynthesis of the plant (Harley and Smith 1983). The history of AMF and autotrophic land plant's symbiosis was dated back to at least 460 million years ago, according to fossil records and molecular clock analysis (Luc Simon 1993; Redecker, et al. 2000). On the other hand, investigations of molecular clocks indicated that the appearance of Glomeromycota is much older than the emergence of land plant on earth (Berbee and Taylor 2010; Bidartondo, et al. 2011). It was recently reported that fungus-like mycelial fossils were found in a 2.4-billion-year-old basalt (Bengtson, et al. 2017), which showed that the fungal lineage is much older and their early evolution may have occurred in the deep ocean. So far, it is uncertain when AMF ancestor lost its ability to assimilate carbon by itself or if AMF ancestor

had former symbiotic partnership with an autotrophic cyanobacterium ancestor. However, the studies of recent molecular data of Glomeromycota does not reject the hypothesis that some basal lineages of Glomeromycota (at least *Paraglomerales* and *Archaeosporales*) might predate the land plant species (Berbee and Taylor 2010; Bidartondo, et al. 2011; Pöggeler and Wöstemeyer 2011). .

Interestingly, an evidence of symbiosis between a basal species of *Glomeromycota* (*Geosiphon pyriforme,* assigned to basal order *Archaeosporales*) and Cyanobacteria (*Nostoc punctiforme*) has been documented and characterized (Gehrig, et al. 1996). Instead of the photosynthetic land plant as a symbiotic partner, *G. pyriforme* forms symbiosis with photosynthetic cyanobacteria. An analogy can be made between *Geosiphon-Nostoc* symbiosis and Glomeromycota-*Mollicutes-related endobacteria* endosymbiosis where a substantial amount of HGT events were reported (Naito, et al. 2015; Torres-Cortes, et al. 2015). The *Geosiphon-Nostoc* symbiosis is characterized by the formation of photosynthetically active bladder-like cellular structure that is made of a swollen fungal hypha of up to 2 mm in size, containing *Nostoc* filaments and heterocyst in the upper 2/3 of the bladder and lipid droplets in the lower part (Alexopolous, et al. 2004). *Geosiphon* produces large spores resembling those of *Glomeromycota*. Phylogenetic analysis based on SSU rDNA sequence assigned *Geosiphon pyriforme* into order *Archaeosporales* of phylum *Glomeromycota* (Redecker, et al. 2013). Compared to other eukaryotes including fungi, the possession of class I ribonuclease III enzyme in AMF is unique and this seems to be related with HGT from cyanobacterial ancestor. It is unclear whether the symbiotic form of *G. pyriforme* is exceptional to this specific species or it is the primitive symbiotic form which was maintained throughout AMF evolution like a living fossil. However, the interaction of *G. pyriforme* and *N. punctiforme* is a clear evidence of

proximity of AMF with cyanobacteria. The gene exchange with its symbiotic partner was reported multiple times in Glomeromycota (Kuo 2015; Naito, et al. 2015; Torres-Cortes, et al. 2015). The evolution of RNAi system is related with the symbiosis in eukaryotic organisms including nematode according to many published studies (Buck and Blaxter 2013; Weiberg, et al. 2015; Yigit, et al. 2006). At the same time, our finding of RNAi system in *R. irregulare* clearly showed its complexity and uniqueness comparing to other fungi, and this could be related to the symbiotic nature with AMF. It is likely that the acquisition of key enzyme in RNAi (cyanobacterial ribonuclease III) by HGT into AMF genomes may reflect an ancient symbiosis history with cyanobacteria. Further analyses of the distribution of cyanobacterial ribonuclease III homologs (RIRNC 2 and 3) in AMF orders: *Archaeosporales* and *Paraglomerales* will bring insights on our hypothesis of HGT event in *Glomeromycota* as well as their complex evolution of symbiosis.

2.6. Acknowledgements

This work was supported by NSERC Discovery fund to MH which is gratefully acknowledged. We thank Dr Yves Terrat for his assistance in bioinformatics and Yerim Heo for assistance to build and edit some figures. We also thank two anonymous reviewers for their helpful comments.

Chapter 3 – Blue light sensing mechanism and circadian clock components conserved in *R. irregularis*

Presentation of article 2

This article was published in Mycorrhiza. I have conceived and design the experiment with supervision of Dr. D. Morse and Dr. M. Hijri. I have conducted all bioinformatic analyses. I also did the molecular biological experiment with some assistances from Miss M. Kong. I have written this article with revisions from Dr. D. Morse and Dr. M. Hijri. During the time of this study conducted and published, *Rhizophagus irregulare* was the official nomenclature of *Rhizophagus irregularis* and was explained in the manuscript.

Article 2 - Expression of putative circadian clock Components in the arbuscular mycorrhizal fungus *Rhizoglomus irregulare*

Soon-Jae Lee, Mengxuan Kong, David Morse and Mohamed Hijri

Institut de Recherche en Biologie Végétale, Université de Montréal, 4101 Rue Sherbrooke Est, Montréal (Québec) H1X 2B2, Canada

Published in : Mycorrhiza, June 21, 2018. https://doi.org/10.1007/s00572-018-0843-y

3.1. Abstract

Arbuscular mycorrhizal fungi (AMF) are obligatory plant symbionts that live underground, so few studies have examined their response to light. Responses to blue light by other fungi can be mediated by White Collar-1 (WC-1) and WC-2 proteins. These *wc* genes, together with the frequency gene (*frq*), also form part of the endogenous circadian clock. The clock mechanism has never been studied in AMF, although circadian growth of their hyphae in the field has been reported. Using both genomic and transcriptomic data, we have found homologs of *wc-1*, *wc-2*, and *frq* and related circadian clock genes in the arbuscular mycorrhizal fungus *Rhizoglomus irregulare* (synonym *Rhizophagus irregularis*). Gene expression of *wc-1*, *wc-2*, and *frq* was analyzed using RT-qPCR on RNA extracted from germinating spores and from fungal material cultivated *in vitro* with transformed carrot-roots. We found that all three core clock genes were expressed in both pre- and post-mycorrhizal stages of *R. irregulare* growth. Similar to the model fungus *Neurospora crassa*, the core circadian oscillator gene *frq* was induced by brief light stimulation. The presence of circadian clock and output genes in *R. irregulare* opens the door to the study of circadian clocks in the fungal partner of plant-AMF symbiosis. Our finding also provides new insight into the evolution of the circadian *frq* gene in fungi.

3.2. Keywords

Arbuscular Mycorrhizal Fungi (AMF), symbiosis, light sensing, white collar, circadian clock, FRQ gene evolution

3.3. Introduction

Circadian rhythms are ubiquitous, endogenously generated rhythms in various biological processes with periods of about 24 hours under constant conditions. A broad range of prokaryotes and eukaryotes have circadian rhythms, which have presumably evolved to better adapt cellular physiology to the predictable changing environment of day and night. The endogenous mechanism that produces these rhythmical oscillations of biological activity is called the circadian clock. In the model fungus *Neurospora*, the circadian clock involves the daily synthesis of the *frequency* (*frq*) transcript by a heterodimeric transcription factor and blue light receptor formed from two White collar (WC) proteins, WC-1 and WC-2 (Loros and Dunlap 2001; Roenneberg and Merrow 1999). So far, all fungal species in which the *frq* gene was found, had a circadian clock (Liu and Bell-Pedersen 2006; Lombardi and Brody 2005). Not all fungi with circadian rhythms, however, have a *frq* gene. Examples include the circadian rhythm of spore production by the mucoromycotan genus *Pilobolus* (Uebelmesser 1954), and the circadian rhythm in sclerotia development in *Aspergillus* (Greene, et al. 2003), as neither of these fungi have *frq*.

Light is among the most important stimuli in fungi, where it modulates various physiological and biological processes including sporulation, the germination of spores, growth of hyphae, synthesis of metabolites and resetting the timing of the circadian clock (Corrochano 2011; Purschwitz, et al. 2006; Rodriguez-Romero, et al. 2010). Light sensing mechanisms in fungi differ between species and wavelengths of light, with perception of blue light mediated by WC proteins thought to be the oldest mechanism (Purschwitz, et al. 2006; Rodriguez-Romero, et al. 2010). In fungi, WC protein-mediated blue light perception has been widely conserved in Mucoromycota but also in Ascomycota and Basidiomycota, although several species in those

lineages have lost the *wc* genes (Idnurm, et al. 2010; Rodriguez-Romero, et al. 2010). Cryptochrome (CRY) and Vivid (VVD) proteins also were reported as additional blue light perception mediators. CRY is not part of the circadian clock in *Neurospora*, however, and VVD has only been found in the Ascomycota. Interestingly, the circadian clock component *frq* was previously thought to exist only in the ascomycete fungal lineage (Salichos and Rokas 2010). Recently, however, *frq* was annotated in a multi-omic study of *Rhodosporidium toruloides* (Basidiomycota) but has not yet been characterized in depth (Zhu, et al. 2012).

The blue light sensing pathway using the white collar complex (WCC) has been extensively studied in *Neurospora crassa* (Grimaldi, et al. 2006). The WCC acts as a transcription factor, as both WC-1 and WC-2 have a DNA binding zinc-finger domain in their C-terminal ends (Takemiya, et al. 2005). WC-1 has three PAS (Per-Arnt-Sim) domains, and it is the N-terminal PAS domain, also called a LOV domain (for light, oxygen, and voltage sensing) that binds the chromophore FAD (Flavin Adenine Dinucleotide) essential for blue light sensing. Stimulation by blue light enhances the transcriptional activation mediated by the WCC (Liu and Bell-Pedersen 2006; Olmedo, et al. 2010).

One of the genes regulated by the WCC is *frq* (Liu and Bell-Pedersen 2006; Loros, et al. 1989; Loros and Dunlap 2001; Olmedo, et al. 2010). The protein FRQ participates in a feedback loop inhibiting its own transcription by interfering with the transcriptional activity of the WCC. In the *Neurospora* model, FRQ recruits an FRQ-interacting RNA helicase (FRH), as well as the protein kinases Casein Kinase I (CKI) and II (CKII)) that together enter the nucleus and promote phosphorylation of WCC. The phosphorylated WCC is exported to the cytoplasm, thus inhibiting new *frq* transcription. At the same time, FRQ also becomes progressively phosphorylated over time, eventually leading to its recognition by a ubiquitin ligase and

proteolysis, after which WCC again becomes active thus restarting the circadian cycle (Dunlap and Loros 2017; Gallego and Virshup 2007; Loros and Dunlap 2001). In fungal species with a conserved *frq* gene, the periodical fluctuations of *frq* by this mechanism constitute the circadian clock (Loros and Dunlap 2001).

The circadian clock made up of *frq* and WCC as core components controls the expression of a number of other genes. These clock-controlled genes (*ccgs*) include some involved in glucose-repression, fungal hydrophobin production, asexual reproduction as well as metabolism (Bell-Pedersen, et al. 1996; Hurley, et al. 2014; Shinohara, et al. 2002). Among the *ccgs* related to metabolism is the gene *ccg-7* which encodes glyceraldehyde-3-phosphate dehydrogenase, a key enzyme in glycolysis and gluconeogenesis (Bell-Pedersen, et al. 1996). This gene is regulated by the clock but is not light-induced, indicating that expressions of at least some of the *ccg*s are not directly mediated by WCC.

So far, the fungal circadian clock is modelled on studies in *Neurospora crassa* (Bell-Pedersen, et al. 1996; Loros, et al. 1989; Loros and Dunlap 2001). There have not been any previous studies on the mechanism of light perception or the circadian clock in the arbuscular mycorrhizal fungi (AMF), which mostly belong to an early diverging fungal phylum, the Glomeromycota, with only a few species reported in its related sub-phylum Mucoromycotina. Even though the mycorrhizal symbiosis with plant roots is tightly linked with nutrition, metabolism and signaling (Bago, et al. 2002; Jansa, et al. 2013; Schüβler, et al. 2001), there have been no studies of the plant circadian clock in the context of mycorrhizal symbiosis. Indeed, most work in plant rhythms has been done using leaf tissue, and only a small number of studies have involved plant roots (James, et al. 2008; Simon and Dodd 2017; Yazdanbakhsh, et al. 2011).

The potential for generating circadian rhythms and for light sensing in the belowground symbiotic partners, AMF, has not previously been examined.

Although no mechanism was previously identified, there is a clear blue light response in AMF. For example, light affected spore germination and mycorrhiza establishment in the arbuscular mycorrhizal fungus species *Glomus fasciculatus* (Ferguson and Menge 1982), and blue light (340~450 nm) affected early germination of spores and hyphal branching in *Gigaspora gigantea* (Nagahashi and Douds 2003) with other light wavelengths having no effect. Interestingly, these last authors also found a synergy between host plant chemicals and blue light on the germination of AMF spores (Nagahashi and Douds 2004), which suggested to them that the blue light response in AMF may relate to host pre-symbiotic signaling processes and root colonization. As described above, light is an important stimulus which allows AMF to orient their hyphal growth toward the surface layers of the soil where seed germination occurs, therefore increasing their chance of meeting their future hosts. These observations on pre-mycorrhizal AMF, however, raise many questions. How does light sensing actually occur? Might blue light perception be linked to a circadian clock as is the case in *Neurospora crassa* (Liu and Bell-Pedersen 2006; Lombardi and Brody 2005)? If there is a conserved circadian clock mechanism in these plant-root symbionts, would it be similar to the general circadian clock found in many free-living fungi?

To take a first step for addressing these questions, we have searched the *R. irregularis* sequence databases for the *wc-1* and *wc-2* genes, which encode the blue light sensor, and for *frq*, to see if the core circadian clock genes of the model fungus *N. crassa* are present. Furthermore, we have tested if the core clock genes *wc-1*, *wc-2*, and *frq*, are expressed, and if *frq* expression can be induced with a short (20 min) exposure to blue light as described for *N. crassa*.

3.4. Materials and Methods

3.4.1. Identification and phylogenetic analyses of blue light sensing and circadian clock genes

The three core clock genes (*wc-1, wc-2* and *frq*) as well as three additional clock genes (*frh, ck-1* and *ck-2*) of *Neurospora crassa* (Table 3.1) were used as queries to search the *R. irregulare* transcriptomic and genomic sequence databases (Tisserant, et al. 2012; Tisserant, et al. 2013) for orthologs using TBLASTN (Lee, et al. 2018a). In addition, 12 *N. crassa ccgs* (Bell-Pedersen, et al. 1996; Loros, et al. 1989) were also used as queries to look for homologs in *R. irregulare*. Orthology of the *R. irregulare* genes was assessed using the reciprocal best hits approach together with domain architecture analysis. First, with sequences from *N. crassa* as queries, potential homologs in *R. irregulare* were retrieved using an e-value lower than $1E^{-10}$ and a sequence identity over 30% as selection criteria. All candidate sequences were then compared with all proteins of *N. crassa* using BLASTX. Putative open reading frames of *R. irregulare* candidate sequences were annotated by pairwise comparisons using BLASTX against GenBank sequences and multiple sequence alignments were conducted using MUSCLE v. 3.8 (Edgar 2004). The domain architecture of all candidates remaining after the reciprocal BLAST was examined using SMART 7 (Simple Modular Architecture Research Tool) (http://smart.embl.de/) (Letunic, et al. 2015) and InterProScan v. 61.0 (https://www.ebi.ac.uk/interpro/) (McDowall and Hunter 2011) with a cut-off E-value of $1E^{-04}$ for domain searching. Only candidates containing domain architectures identical to the *N. crassa* orthologs were retained.

To conduct phylogenetic analyses, clock genes were recovered from a broad range of fungal taxa, including Ascomycota, Basidiomycota and Mucoromycotina, in the EMBL databases of orthologous groups and functional annotation using EggNOG v. 4.5.1 (http://eggnogdb.embl.de) (Huerta-Cepas, et al. 2016) and *R. irregulare* sequences as queries. Again, sequences with an e-value lower than $1E^{-10}$ and a sequence identity over 30% were collected. Multiple amino acid sequence alignments (MUSCLE v. 3.8 (Edgar 2004)) were inspected and divergent or ambiguous positions were removed using BioEdit v. 7. For each protein, the best evolutionary models with the lowest Akaike information criterion (AIC) were determined using ProtTest 3 (Darriba, et al. 2011). The phylogenetic trees were inferred by the maximum likelihood phylogenetic analysis program, PhyML v. 3.0 (Guindon, et al. 2010), using the best model of each multiple sequence alignment and 1,000 bootstrap replicates.

3.4.2. AMF isolate and growth conditions

The arbuscular mycorrhizal fungal isolate used in this study was *Rhizoglomus irregulare* (syn. *Rhizophagus irregularis* and *Glomus irregulare*), deposited in the Fungal Collection of the Department of Agriculture Canada under the accession number DAOM-197198 (Ottawa, ON). An *R. irregulare* isolate DAOM-197198 spore suspension, provided by Premier Tech Biotechnologies (Rivière-du-Loup, QC), was used in the spore germination experiment which required a large number of fresh spores. For the mycorrhizal phase experiment, an *R. irregulare* isolate was cultured *in vitro* with Ri-T-DNA transformed *Daucus carota* roots in Petri dishes containing minimal (M) medium solidified with 0.4 % Gellan gum (Sigma Aldrich, Canada) kept in an incubator at 25 °C in the dark for four months. Each AMF isolate is a culture that was originally started from a single spore and subcultured by transferring

1 to 2 cm^2 of gel containing a mixture of mycorrhizal roots, spores and hyphae to a new Petri dish.

3.4.3. Culture preparation for germination and mycorrhizal phase experiments

100,000 fresh spores (20,000 spores per replicate for a total of five replicates) of *R. irregulare* were inoculated in 10 mL of liquid minimal medium. The spores were placed in an incubator and allowed to germinate in the dark at 25°C for two weeks. Culture preparations for mycorrhizal stages of the *R. irregulare* isolate were prepared as described above with modified M media containing 0.04% Gellan gum (Sigma Aldrich, Canada). Five replicates were prepared for each treatment.

3.4.4. Blue light exposure

All cultures were wrapped with aluminum foil and boxed in a 13.5 cm (height) x 13.5 cm (width) x 12.5cm (depth) paper dark box to prevent penetration of light during the initial incubation in the dark at 25°C. Five replicates were used for each of control (dark) and treatment (blue light exposure). Experiments were conducted with both germination and mycorrhizal stages. Before the harvesting and extraction of RNA, two-week old dark incubation cultures (controls and treatments) were transferred to another incubator adjusted to 25°C and equipped with a blue light source of 2,650 lumens producing a light intensity of 190 lux. A Marine-Glo actinic blue marine bulb (Hagen, Japan) with an emission spectrum of 380-580 nm (peak at 430 nm) was used as the source of blue light. Boxes containing plates were opened and the aluminum foil was removed to allow exposure to light. All plates were arranged to be equidistant from the

light source. The temperature was kept at 25°C for both control and treatment plates. After 20 minutes of light exposure, all samples were harvested for RNA extraction.

3.4.5. Extraction of RNA and cDNA library construction

Germinating spores of *R. irregulare* were harvested by filtration through a plastic sieve of 10 µm mesh in a 1.5 mL microtube and quickly frozen with liquid nitrogen. Hyphae and spores of *R. irregulare* cultivated *in vitro* were cut from the media and harvested by dissolving Gellan Gum using an extraction buffer solution of 0.82 mM sodium citrate and 0.18 mM citric acid. Dissolved medium was filtered using a 10 µm mesh sieve and fungal material was transferred to a 1.5 mL microtube and quickly frozen with liquid nitrogen. Roots were removed from all fungal material. Samples were ground using a sterilized pestle. RNA extractions and DNase I treatments for removing any possible DNA contamination were performed using E.Z.N.A Fungal RNA extraction kit (Omega-Biotek, Canada) following the manufacturer's instructions. The amount of RNA per each reaction was normalized to 100 ng for further analyses. RT-PCR was performed to produce cDNA libraries using an iScript reverse transcription kit (Bio-Rad, Canada) following the manufacturer's instructions.

3.4.6. Polymerase Chain Reaction (PCR)

PCR was used to confirm the mRNA expression and to validate primers before designing TaqMan probes. cDNA samples produced from germinating spores were used as templates. PCR primers were designed using Primer 3 software (Supplementary Information Table S1). PCR reactions were performed using *Taq* DNA polymerase (Qiagen, Canada) following the manufacturer's recommendations. Reactions were performed in a 20 µL volume

under the following conditions: denaturation at 95°C for 2 min, followed by 35 cycles of 94°C for 30 sec, 48°C for 30 sec and 70°C for 2 min. Final elongation was performed at 72°C for 10 min. Reactions were performed using a Mastercycler Pro S thermocycler (Eppendorf, Canada). PCR amplicons were visualized on 1% agarose gels stained with GelRed (Invitrogen, Canada).

3.4.7. TaqMan qPCR assay

Quantitative PCR (qPCR) was performed to analyze the expression level changes of *wc-1*, *wc-2* and *frq* genes on two-week old germinated spores and four-month old mycorrhizal cultures, which represent two different phases of the *R. irregulare* life cycle. 18S rRNA and *ubiquitin* genes were used as references to determine the expression level of *wc-1*, *wc-2* and *frq* genes. TaqMan probes and primer sets were designed using Primer 3 software (Supplementary Information Table S1). Probes were labeled with FAM at the 5' end and BHQ-1 at the 3' end. Fluorescence data were collected with the ViiA^{TM} 7 Real-Time PCR System (Life Technologies, Canada). qPCR was performed with five biological replicates and three technical replicates for each biological replicate. Threshold cycle difference (ΔCt) between target genes and reference genes was calculated by the formula: $Ct_{Target} - Ct_{Reference}$, and further comparison of gene expression between the blue light and the dark conditions were obtained by ΔΔCt with a formula: $\Delta Ct_{dark} - \Delta Ct_{light}$. In each set of reactions, a negative control of qPCR with sterilized water was used to detect possible false fluorescence signal.

3.4.8. Statistical analysis

Statistical analysis was conducted with the software, XLSTAT (Addinsoft, France). Two independent experiments (spore germination and mycorrhizal phase) were performed in

this study. In each experiment, means of the control group (dark) and the treatment group (light) were compared by using a two-tailed Welch's t-test. Sample sizes were five biological replicates comprising three technical replicates each for each treatment.

3.5. Results

3.5.1. Identification of blue light sensing and circadian clock genes in *R. irregulare*

We have found that the genome of *R. irregulare* isolate DAOM-197198 has the three core circadian clock genes (*wc-1*, *wc-2* and *frq*) as well as two additional clock genes (*frh* and *ck*). Homologs were recovered by reciprocal BLAST using authentic clock genes from the model fungus *N. crassa* as queries and verified by protein domain conservation analysis. All of the homologs retrieved identical sequences from all the EST datasets publically available (Tisserant, et al. 2012; Tisserant, et al. 2013). The *R. irregulare* homologs of the *N. crassa* blue light perception and circadian clock genes are summarized in Table 3.1.

Table 3.1. The number of copies of three core clock genes (*wc-1*, *wc-2* and *frq*), two additional clock genes (*frh*, *casein kinase*) and circadian clock-controlled genes in *R. irregulare*. The number of copies was determined based on reciprocal BLAST and conserved domain analysis via SMART 7 (Simple Modular Architecture Research Tool) (http://smart.embl.de/) (Letunic, et al. 2015) and InterProScan v. 61.0 (https://www.ebi.ac.uk/interpro/) (McDowall and Hunter 2011). Orthologues from the model fungi (*N. crassa*) were used as queries. For the multiple FRH and CK copies, the accession numbers of *N. crassa* query sequences were noted only once at the top of the list.

Name of component	Number of homologs in	Name of homolog(s) in *R.irregulare*	Best reciprocal hit in query (N. crassa)	Best BLAST hit E-value / (%identity)	BLAST hit information [organism name]

	R. irregulare		E-value / (%identity)		
Core components					
White collar 1 (WC 1)	1	WC1 (EXX50758.1)	XP_011395150.1 0.0 / (45%)	ORE20722.1 0.0 / (52%)	putative white-collar-1a protein [*Rhizopus microsporus*]
White collar 2 (WC 2)	1	WC2 (EXX67146.1)	XP_963819.3 3E-72 / (37%)	ACT46736.1 2E-82 / (37%)	white collar-2 [*Phaeosphaeria avenaria f. sp. tritici*]
Frequency (FRQ)	1	FRQ (LC350022)	XP_011395125.1 3E-19 / (31%)	CCX09772.1 7E-17 / (33%)	Similar to Frequency clock protein [*Pyronema omphalodes*]
Additional components					
FRQ related RNA helicase (FRH)	2	FRH1 (EXX70985.1)	XP_956298.1 0.0 / (46%)	XP_019025319 0.0 / (52%)	putative ATP dependent RNA helicase [*Saitoella complicata*]
		FRH2 (EXX52311.1)	4E-164/(33%)	ORX93574.1 0.0 / (55%)	antiviral helicase [*Basidiobolus meristosporus*]
Casein kinase (CK)	6	CK1 (EXX58190.1)	ESA42258.1 0.0 / (78%)	KFH67286.1 0.0 / (69%)	protein kinase [*Mortierella verticillata*]
		CK2 (EXX58189.1)	0.0 / (80%)	KFH67286.1 0.0 / (70%)	protein kinase [*Mortierella verticillata*]
		CK3 (EXX58188.1)	0.0 / (80%)	KFH67286.1 0.0 / (72%)	protein kinase [*Mortierella verticillata*]
		CK4 (EXX58191.1)	0.0 / (87%)	KNE98836.1 0.0 / (91%)	protein kinase [*Puccinia striiformis*]

		CK5 (EXX58187.1)	0.0 / (80%)	KFH67286.1 0.0 / (70%)	protein kinase [*Mortierella verticillata*]
		CK6 (EXX59340.1)	6E-66 / (39%)	GAN01887.1 8E-80 / (43%)	casein kinase I isoform [*Mucor ambiguus*]
Clock controlled genes (ccgs)					
CCG-7 (glyceraldehyde-3-phosphate dehydrogenas)	1	(EXX77548.1)	XP_956977.1 7E-175/(69%)	OAQ36335.1 0.0 / (77%)	Glyceraldehyde 3-phosphate dehydrogenase [*Mortierella elongata*]
CCG-8 (transcriptional regulator Opi1)	1	(EXX69239.1)	XP_960618.2 5E-17/(31%)	GBC19065.1 0.0/(99%)	Transcription factor Opi1-domain-containing protein [*Lobosporangium transversale*]
CCG-9 (trehalose synthase)	1	(EXX56092.1)	XP_959484.3 8E-163/(39%)	CDH48244.1 0.0/(51%)	Trehalose synthase [*Lichtheimia corymbifera*]

Homologs of *wc-1, wc-2* and *frq* were also found in the *R. irregulare* genome. The *R. irregulare* genome contains two homologs of *frh* and six homologs of *ck,* however, in contrast to *N. crassa* where the genome has only one copy of *frh* and two copies of *ck*. The domain architecture (McDowall and Hunter 2011) of these later showed that the FRH of *N. crassa* and both *R. irregulare* homologs were classified as ATP-dependent RNA helicases (IPR016438) with the same core domain, Helicase superfamily 1/2 ATP-binding domain (IPR014001). All six CK homologs in *R. irregulare* had a protein kinase domain (IPR000719), an ATP binding site (IPR017441) and a serine/threonine-protein kinase active site (IPR008271) with sequence architecture identical to the *N. crassa* CK.

3.5.2. Protein domain architecture and phylogenetic analysis of blue light perception genes in *R. irregulare*

The domain architecture of *R. irregulare* and *N. crassa* WC-1 and WC-2 proteins are shown in Figure 3.1. WC-1 from both *R. irregulare* and *N. crassa* has three PAS domains (IPR000014) interspersed with two PAS-C terminal (PAC) domains (IPR001610), and the C-terminal end has a GATA-type zinc finger domain (IPR000679). Similarly, WC-2 from both species has a single PAS domain (IPR000014) and one GATA-type zinc finger domain (IPR000679) at the C-terminal. *R. irregulare* sequences for both WC-1 and WC-2, however, were slightly shorter in length than their *N. crassa* counterparts.

Figure 3.1

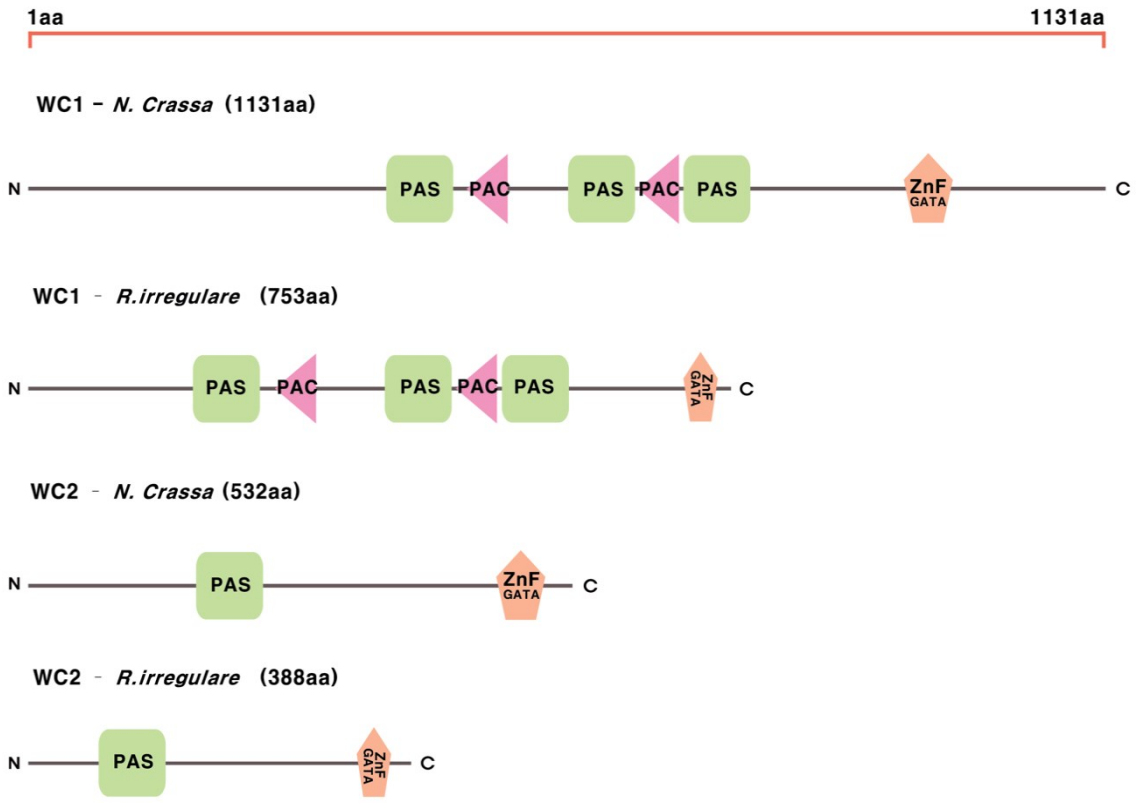


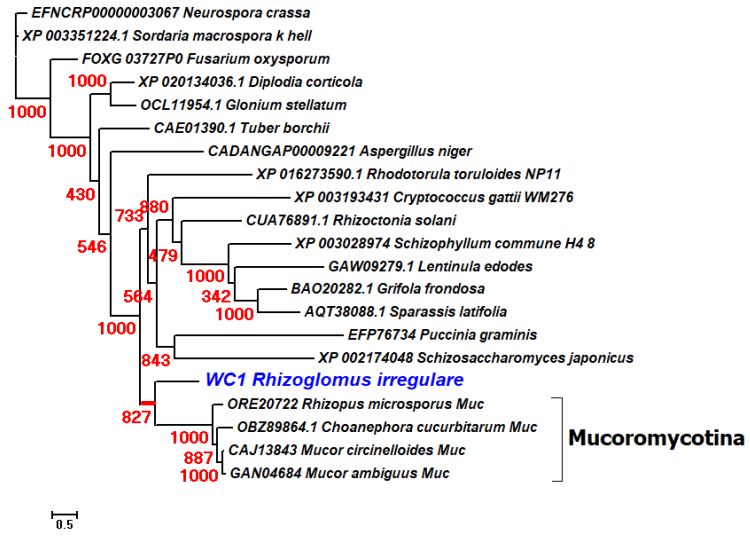
Figure 3.1. Domain structure of WC-1 and WC-2 proteins from *Rhizoglomus irregulare* and *N. crassa*. Domain structures were determined using SMART and InterProScan v. 61.0. The length of each protein is indicated in parentheses. WC-1 contains three PAS domains (IPR000014) interspersed by two PAS-C terminal (PAC) domains (IPR001610), and a single GATA-type zinc finger (ZnF GATA) domain (IPR000679) at the C-terminal. WC-2 contains one PAS domain (IPR000014) and one GATA-type zinc finger domain (IPR000679).

The phylogenies for WC-1 and WC-2 were inferred using a Maximum Likelihood method using orthologues from other fungal taxa retrieved from EMBL databases coupled with

functional annotation (EggNOG v. 4.5.1, http://eggnogdb.embl.de) (Huerta-Cepas, et al. 2016). Figure 3.2 shows phylogenetic trees for WC-1 and WC-2 sequences from major fungal groups including the Ascomycota, Basidiomycota and Mucoromycotina. *R. irregulare* WC-1 and WC-2 are found in the same clade as their Mucoromycotina orthologs (bootstrap values of 827/1000 for WC-1 and 825/1000 for WC-2), but, the topology of the two trees was different (Figure 3.2). Genomic copies of *wc-1* and *wc-2* were found in *R. irregulare* DAOM-197198 genome data version 2.0 recently released by the Joint Genome Institute (JGI). The *wc-1* gene was located in scaffold 106 (382,850 bp) and consisted of ten exons with nine introns, while the *wc-2* gene contained three exons and two introns and was found in scaffold 96 (402,815 bp).

Figure 3.2

a) WC1



b) WC2

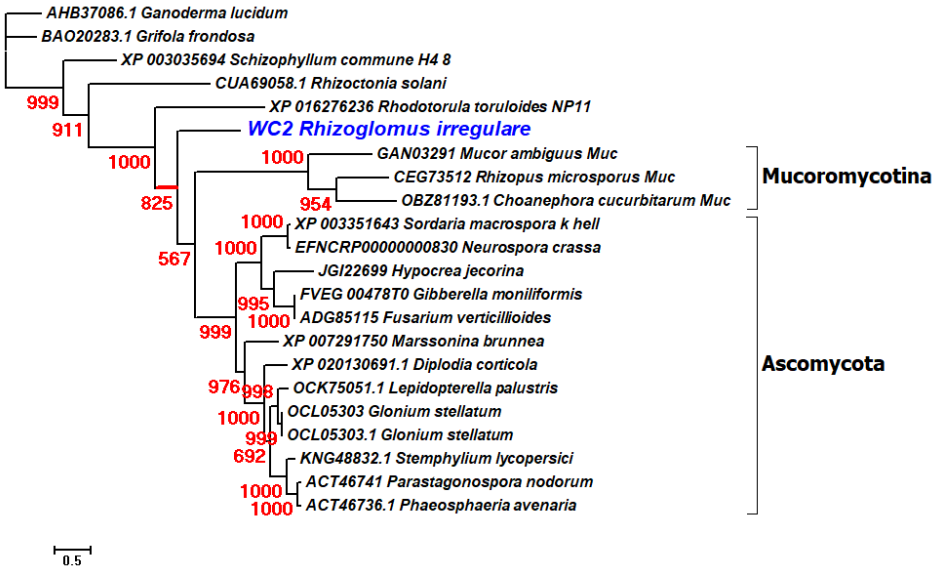


Figure 3.2. Phylogenetic analysis of the WC-1 and WC-2 proteins in Fungi. Amino acid sequences of WC-1 (a) and WC-2 (b) were aligned and the phylogeny reconstructed using a Maximum likelihood method with an LG+I+G (with four distinct gamma categories) phylogenetic model which showed the lowest AIC value. The numbers at branches correspond to bootstrap support values generated with 1,000 replicates.

3.5.3. Characterization of the *R. irregulare* circadian *frq* gene

A single *frq* homolog was found in the *R. irregulare* genome (Table 3.1). The domain architecture of the *N. crassa* and *R. irregulare* homologs is very similar (Figure S2) and both are classified as "Frequency clock protein (IPR018554)" by InterProScan v. 61.0. Both proteins contain a coiled-coil region (analyzed with COILS (Lupas, et al. 1991)) followed by a winged helix-like DNA-binding domain (IPR036388). The coiled-coil region is a crucial domain for the functioning of FRQ and has been was reported to be the most conserved region of FRQ (Cheng, et al. 2001; Querfurth, et al. 2011).

Figure 3.3

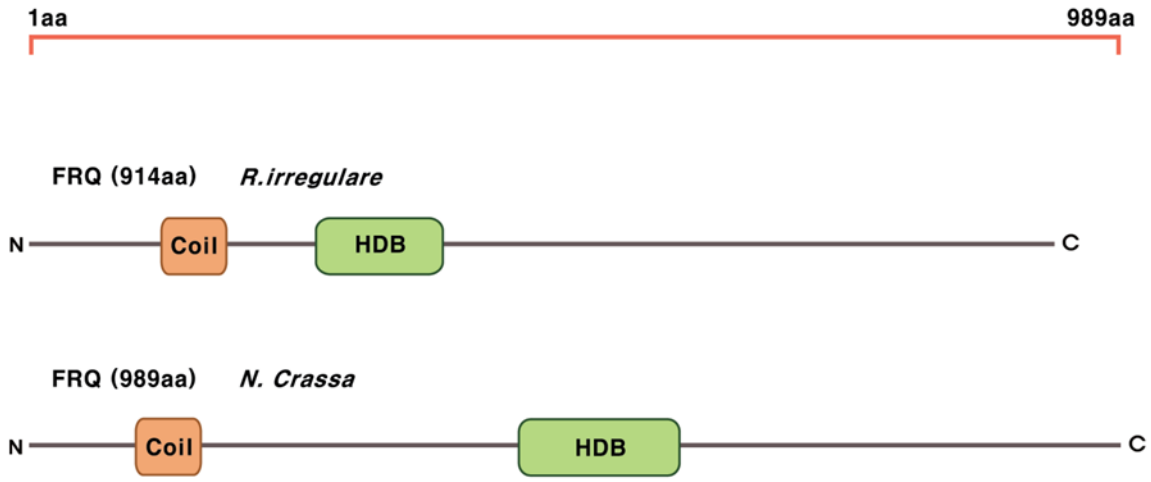


Figure 3.3. Conserved domains in FRQ from *R. irregulare* and *N. crassa*. The sequences of the circadian core oscillator protein FRQ from *R. irregulare* and *N. crassa* were analyzed using SMART and InterProScan v. 61.0. The length of each protein (in amino acids) is indicated by numbers in parentheses. A coiled-coil region (detected by COILS (Lupas, et al. 1991)) and a winged helix-like DNA-binding domain (HDB, IPR036388) were found in both.

This region was thus examined further using multiple sequence alignments of full-length FRQ sequences from 20 species including *R. irregulare.* The entire coiled-coil region (68 amino acid residues) is shown in Figure 3.4, together with the amino acid conservation, the percentage of identity and the alignment quality score (based on BLOSUM 62 substitution matrix). Interestingly, an arginine residue at position 8 in the alignment, as well as an Ile-Asp-Asp-Leu-Thr motif (position 11 to 15), and two leucine residues at positions 21 and 25, were completely conserved in all fungal species analysed so far. The FRQ homolog of *R. irregulare* did show some minor differences, such as a leucine instead of tyrosine at position 28, although

both these amino acids are hydrophobic. The coding region of FRQ in the *R. irregulare* genome contains six exons and five introns, and it was located in scaffold 70 (468,957 bp).

Figure 3.4

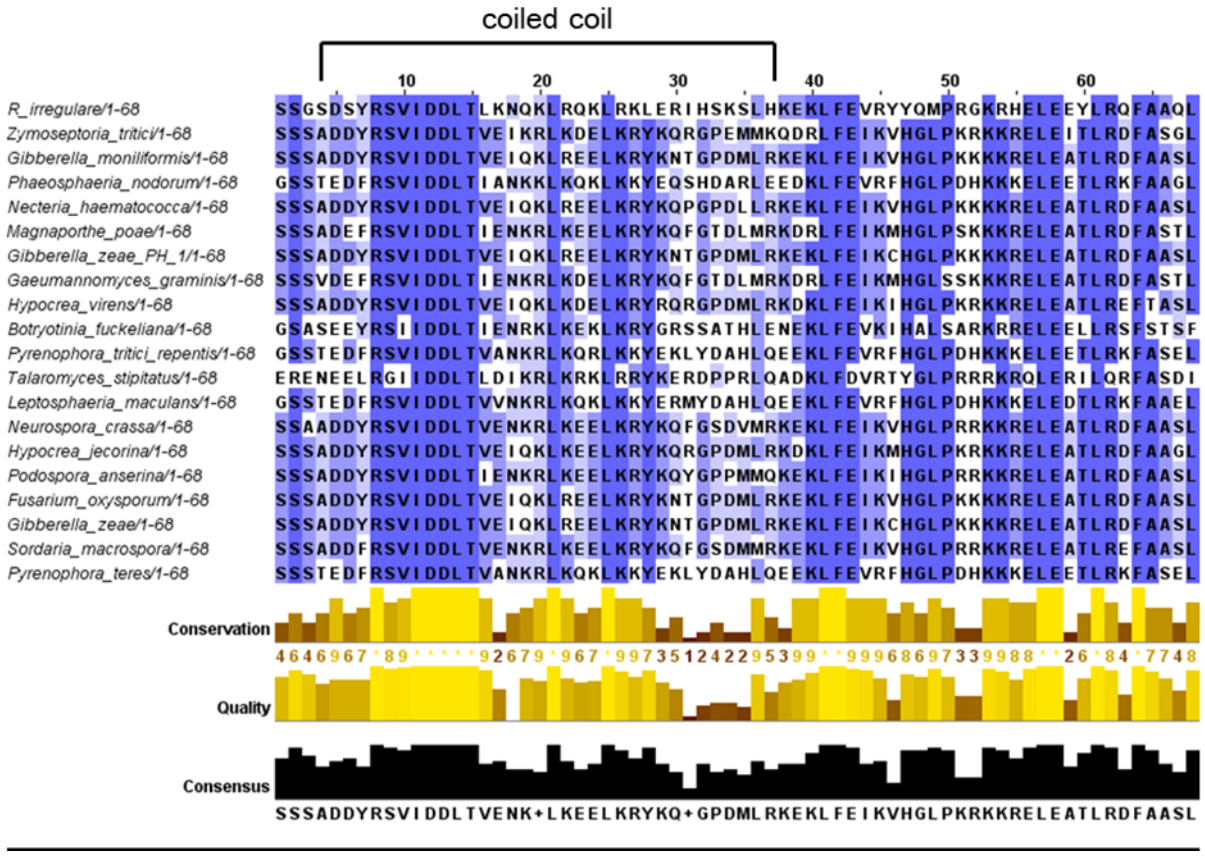


Figure 3.4. Multiple Sequence Alignment (MSA) of the conserved coiled-coil region of FRQ from different fungal species. Protein sequences were obtained from *R. irregulare* and the FRQ orthologs group in the EggNOG v. 4.5.1 database (Huerta-Cepas, et al. 2016) and aligned. Conserved amino acid residues are shaded in blue in the MSA and on a brown to yellow color scale below the MSA (% identity represented as a number from one to ten). The alignment quality score (based on BLOSUM 62 substitution matrix) and a consensus sequence are also shown. The region of the coiled-coil is marked above the MSA. Interestingly, in the coiled-coil region, an arginine (at position 8), the motif Ile-Asp-Asp-Leu-Thr (from position 11 to 15), and two leucines (at positions 21 and 25) were highly conserved among all the species analyzed.

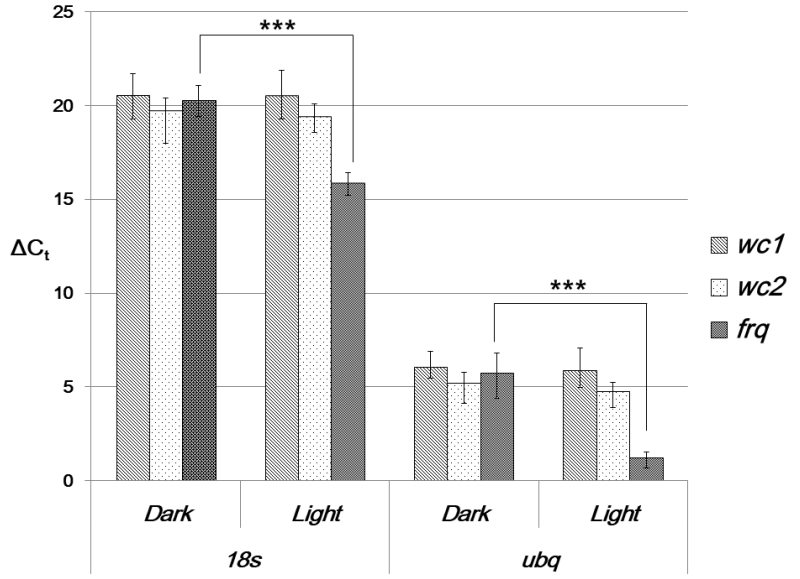
However, *R. irregulare* has leucine (at position 28) instead of the tyrosine found in all other species.

3.5.4. Expression of the *R. irregulare frq* gene in response to blue light

We next tested the expression of three core clock genes (*wc-1*, *wc-2* and *frq*) in both germinating spores and the mycorrhizal phase of *R. irregulare* in response to blue light. All primers and probes were successfully tested (Table S3.1) before use and no fluorescence signal was detected from the negative controls (sterilized water). We found the expression of both reference genes to be the same under blue light and dark conditions (Welch's two-tailed t-test, $p>0.05$), and there was no significant difference in the expression of the white collar genes *wc-1* and *wc-2* (Welch's two-tailed t-test, $p>0.05$) (Figure 3.5). In contrast, a highly significant difference in *frq* expression between dark and blue light conditions was found in both germinating spores and mycorrhizal phases. In pre-mycorrhizal, germinating spore phase, the amount of *frq* mRNA was significantly higher after exposure to blue light, with an average ΔΔCt of 4.39 (Welch's two-tailed t-test, $p=9.630E^{-06}$, normalized with *18S rRNA*) and 4.56 (Welch's two-tailed t-test, $p=2.292E^{-04}$, normalized with *ubiquitin*). A similar increase in the amount of *frq* was observed in blue light-treated mycorrhizal phase samples, with an average ΔΔCt of 5.41 (Welch's two-tailed t-test, $p=1.420E^{-04}$, normalized with *18S rRNA*) and 4.68 (Welch's two-tailed t-test, $p=9.390E^{-05}$, normalized with *ubiquitin*).

Figure 3.5

(a) Spore germination phase



(b) Mycorrhizal phase

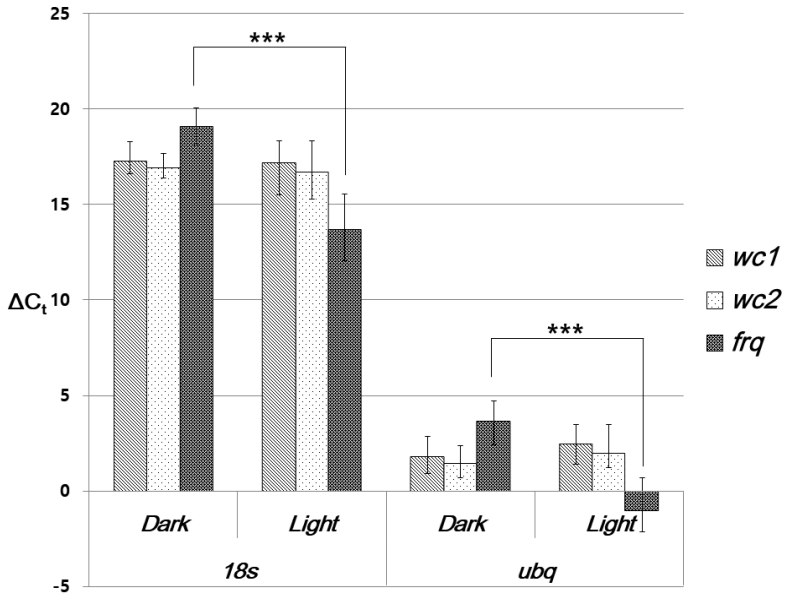


Figure 3.5. Light response of circadian clock gene transcripts (wc-*1*, wc-*2*, *frq*) in two different stages of the *R. irregulare* life cycle. Relative quantification of the amount of *wc-1*, *wc-2* and *frq,* using TaqMan assays with two different housekeeping genes (18S rRNA (left) and *ubiquitin* (right)), was conducted in spore germination (a) and mycorrhizal (b) phases. Each bar represents the mean of five biological replicates (n=5) each containing three technical replicates. The threshold cycle difference (y axis) between target genes and reference genes was calculated as $Ct_{Target} - Ct_{Reference}$. A lower Ct means more expressed mRNA. Welch's two-tailed t-test was conducted between each set, and significance value $p<0.001$ marked as *** (in (a),

p=9.630E^{-06} (*18s rrna*) and p=2.292E^{-04} (*ubiquitin*), and in (b), p=1.420E^{-04} (*18s rrna*) and p=9.390E^{-05} (*ubiquitin*)). The error bars indicate the minimum and maximum values among the replicates.

3.5.5. Identification of circadian output gene homologs in *R. irregulare*

After confirming the expression of circadian clock genes in *R. irregulare*, we next investigated if *R. irregulare* possessed any of the twelve circadian clock output genes (*ccgs*) reported in *N. crassa* (Bell-Pedersen, et al. 1996; Loros, et al. 1989). Among the twelve *N. crassa ccgs*, only three (*ccg-7, ccg-8* and *ccg-9*) were identified in *R. irregulare* (Table 3.1). Furthermore, one of the twelve (*ccg-7)* was also found in another AMF species, *Gigaspora rosea* using the GenBank database of NCBI (Franken, et al. 1997). All of the sequences found in *R. irregulare* had functional domains identical with their *N. crassa* homologs. Interestingly, all the *R. irregulare ccgs* were involved in carbon source metabolism. CCG-7 was identified as glyceraldehyde-3-phosphate dehydrogenase (IPR020829), a key enzyme in glycolysis and gluconeogenesis (Bell-Pedersen, et al. 1996). CCG-8 was classified as a transcription factor Opi1 (IPR013927) which regulates phospholipid biosynthesis (White, et al. 1991). Lastly, CCG-9 was reported as trehalose synthase, and contains a conserved family 1 glycosyl transferase domain (IPR001296). Trehalose synthase catalyzes the synthesis of the disaccharide trehalose, which plays a crucial role in osmotic stress response and glucose starvation (Bell-Pedersen, et al. 1996; Shinohara, et al. 2002). Expression of *ccg-7, ccg-8* and *ccg-9* was confirmed by identical hits in the Expressed Sequence Tags (ESTs) dataset (Tisserant, et al. 2012; Tisserant,

et al. 2013). Thus, three of the twelve circadian clock output genes previously reported in *N. crassa* have conserved homologs in *R. irregulare*.

3.6. Discussion

3.6.1. The belowground niche of glomeromycotan fungi and blue light sensing

The mechanism of blue light perception has not been previously investigated in the AMF, even though a response to blue light has been reported in early spore germination and hyphae branching processes (Ferguson and Menge 1982; Nagahashi and Douds 2003). A number of fungal proteins can be involved in blue light perception, including the WC proteins, Vivid (VVD) and Cryptochrome (CRY). Database searches with *N. crassa* CRY and VVD as query sequences (Rodriguez-Romero, et al. 2010), however, did not detect any homologs in the *R. irregulare* genome. Thus, VVD and CRY are unlikely to be involved in the blue light response of *R. irregulare*. In contrast, our results clearly support the use of a WC-1/WC-2 fungal blue light sensing mechanism in *R. irregulare*. Both these two proteins in *R. irregulare* are most closely related to orthologues in the Mucoromycotina (Figure 3.2), a basal fungal subphylum suggested to be the closest neighbor of Glomeromycota in studies using mitochondrial genomes (Nadimi, et al. 2012) and SSU rRNA sequences (Redecker, et al. 2013). Previous work on the blue light perception mechanism in fungi (Corrochano 2011; Idnurm, et al. 2010; Purschwitz, et al. 2006; Salichos and Rokas 2010) indicated that WC-1 and WC-2 were absent from the early diverging Chytrids and Microsporidia, but were present in all other major clades including Mucoromycotina and Dikarya (Ascomycota and Basidiomycota) (Salichos and Rokas 2010). None of the previous studies, however, included Glomeromycota. It seems likely that the presence of WC-1 and WC-2 homologs in *R. irregulare* genome are sufficient to explain the

blue light response previously reported in other studies (Ferguson and Menge 1982; Nagahashi and Douds 2003).

It has been proposed that the blue light response of germinating AMF spores may be related to host-finding and colonization, because light induces hyphae branching in germinating spores (Nagahashi and Douds 2004). We found that *wc-1* and *wc-2* were expressed in *R. irregulare,* however, even after the establishment of mycorrhizal symbiosis (Figure 3.5). One possible role for WC-1 and WC-2 in mycorrhizal stages may be to provide information about spatial orientation in the soil. Light penetration in the soil is limited to 4-5 mm, and beyond 1 cm light intensity is extremely low and is insignificant (Tester and Morris 1987), while the natural habitat of AMF is far below the one centimeter depth. One typical AMF host plant, *Gossypium,* has roots that can descend 90 cm over a 50 day period (Oosterhuis 1990). Moreover, the presence of AMF in deep roots has been reported for many mycorrhizal trees and shrubs (Stone and Kalisz 1991). At the same time, so far, all the fungal species with conserved *frq* and WC proteins had a circadian clock (Liu and Bell-Pedersen 2006; Lombardi and Brody 2005; Salichos and Rokas 2010). Interestingly, circadian rhythms also were reported in fungi with conserved WC proteins that lacked *frq*. These include the mucoromycotan fungus, *Pilobolus,* where spore production showed a circadian rhythm (Uebelmesser 1954), and in *Aspergillus* sp., where the development of sclerotia exhibited a circadian rhythm (Greene, et al. 2003). It thus seems most likely that WC proteins of *R. irregulare* function primarily, along with conserved FRQ, FRH and CK, in a putative circadian clock similar to what has been found other fungi with FRQ, such as *N. crassa* (Corrochano 2011; Idnurm, et al. 2010; Loros and Dunlap 2001).

3.6.2. Putative conserved circadian clock genes in *R. irregulare* and *frq* gene response to blue light

Based on previous studies in fungi with conserved FRQ and WC proteins, we have summarized the putative mechanisms of blue light sensing and the circadian clock in *R. irregulare* (Figure 3.6). Blue light thus affects the timing of the circadian clock by an effect on the levels of the *frq* transcript.

Figure 3.6

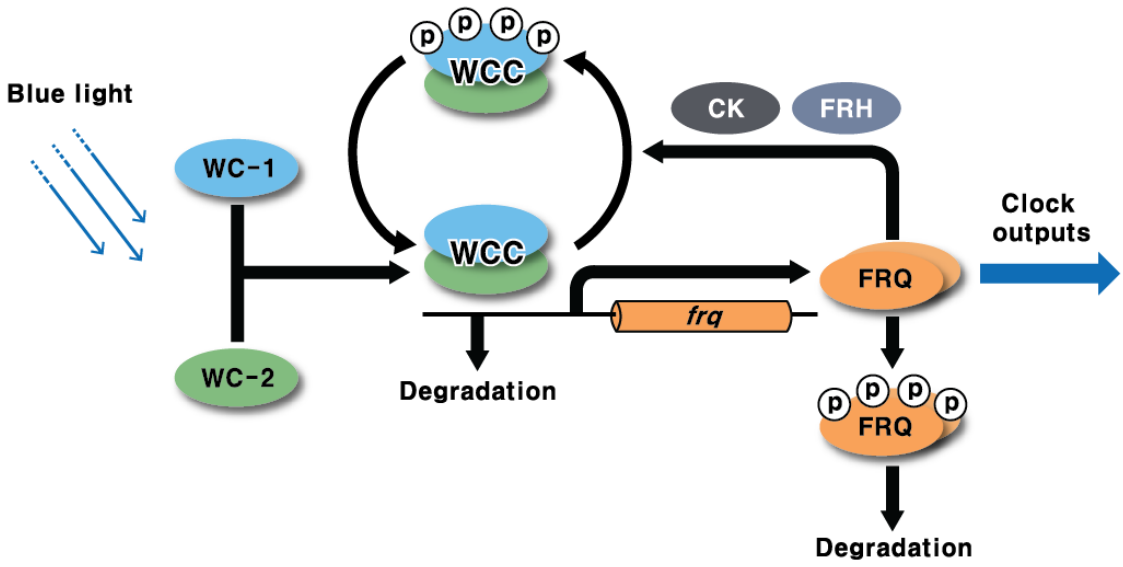


Figure 3.6. Schematic diagram of the putative blue light perception mechanism and the circadian clock in AMF. This diagram, taken from **(Lombardi and Brody 2005; Loros and Dunlap 2001; Tseng, et al. 2012)** was based on studies on the circadian clock in the model fungus *N. crassa*. The three core clock proteins (WC-1, WC-2 and FRQ), as well as two additional proteins FRQ-interacting RNA helicase (FRH) and casein kinase-1a (CK) are conserved in *R. irregulare.*

The effect of a brief light stimulus on *wc-1* or *wc-2* mRNA accumulation has been examined in many fungal species with variable results. The levels of *wc-1* and *wc-2* mRNA were not affected by light in *Fusarium fujikuroi*, *Cryptococcus neoformans* and *Phycomyces blakesleeanus* (Estrada and Avalos 2008; Idnurm and Heitman 2005; Idnurm, et al. 2006), but were affected by light in *Tuber borchii* and *Ophiocordyceps sinensis* (Ambra, et al. 2004; Yang, et al. 2014). A light-dependent induction of *frq,* however, was reported in all cases (Idnurm and Heitman 2005; Loros and Dunlap 2001). We show here that in *R. irregulare*, *wc-1* and *wc-2* expression was not affected by brief light stimulation in either of the two phases of the AMF life cycle studied, while *frq* was induced in both stages (Figure 3.5). This suggests there could be a light-dependent resetting of the putative circadian clock by *frq* accumulation in *R. irregulare* similar to what has been found in other fungi (Gallego and Virshup 2007; Hurley, et al. 2016a; Loros and Dunlap 2001).

3.6.3. Possible clock mechanism in a belowground obligatory root-symbiont

The circadian clock is a mechanism by which an organism synchronizes its biological rhythms to the periodic fluctuation of day and night (Liu and Bell-Pedersen 2006; Loros, et al. 1989; Loros and Dunlap 2001; Olmedo, et al. 2010). It has been shown that the growth of the plant root and carbon allocation is orchestrated by the circadian clock in *Arabidopsis thaliana* (Yazdanbakhsh, et al. 2011) and that circadian rhythms in the root are synchronized by the circadian clock in the photosynthetic upper shoots (James, et al. 2008). Interestingly, the diurnal pattern of hyphae growth of AMF *in situ* has a peak in the day time (from 12:00 to 18:00) which coincides with the highest photosynthetic productivity of plants as well as soil temperature and moisture (Hernandez & Allen, 2013). Considering that all the fungal species with a conserved

frq gene have a circadian clock (Liu and Bell-Pedersen 2006; Lombardi and Brody 2005), the results of both our study and those of Hernandez & Allen, (2013) suggest that AMF possess a circadian oscillator.

While light resetting of the AMF clock seems possible, given our observed induction of *frq* by light, many AMF species are associated with host roots deep in the soil where light cannot penetrate (Oosterhuis 1990; Shukla, et al. 2013). In model organisms, both temperature and nutrition have been found to provide the information of natural periodical fluctuation, and thus affect circadian timing (Hurley, et al. 2016b; Liu and Bell-Pedersen 2006; Lombardi and Brody 2005). Temperature becomes more and more constant as the depth in soil increases (Ciani, et al. 2005; Smith 1932), however, and plant roots can grow deep enough to have a constant temperature in soil (Maeght, et al. 2013; Oosterhuis 1990; Shukla, et al. 2013; Stone and Kalisz 1991). Thus, it seems unlikely that temperature alone acts as a periodic information signal to the putative clock of AMF. In contrast, nutrition seems a reasonable hypothesis for circadian timing signals capable of synchronizing AMF. In *N. crassa,* there is a tight and reciprocal relationship between the circadian timing mechanism and cellular metabolism (Hurley, et al. 2016b). We have found that three of the *N. crassa* clock output genes, *ccg-7* (glyceraldehyde-3-phosphate dehydrogenase), *ccg-8* (transcriptional regulator of phospholipid biosynthesis) and *ccg-9* (trehalose synthase), are conserved and expressed in *R. irregulare*. These three genes encode key enzymes of carbohydrate or phospholipid metabolism (Bell-Pedersen, et al. 1996; Shinohara, et al. 2002; White, et al. 1991), so it is possible that the putative AMF clock may regulate carbon flow in AMF hyphae. At the same time, AMF are obligatory symbionts, and depend entirely on the host plant as a source of carbon. Plants typically display circadian rhythm in photosynthesis (McClung 2006; Serikawa, et al. 2008; Simon and Dodd 2017). We thus hypothesize that timing

signals generated by the plant host can be passed to the AMF where they function in assuring synchrony between the plant and AMF clocks.

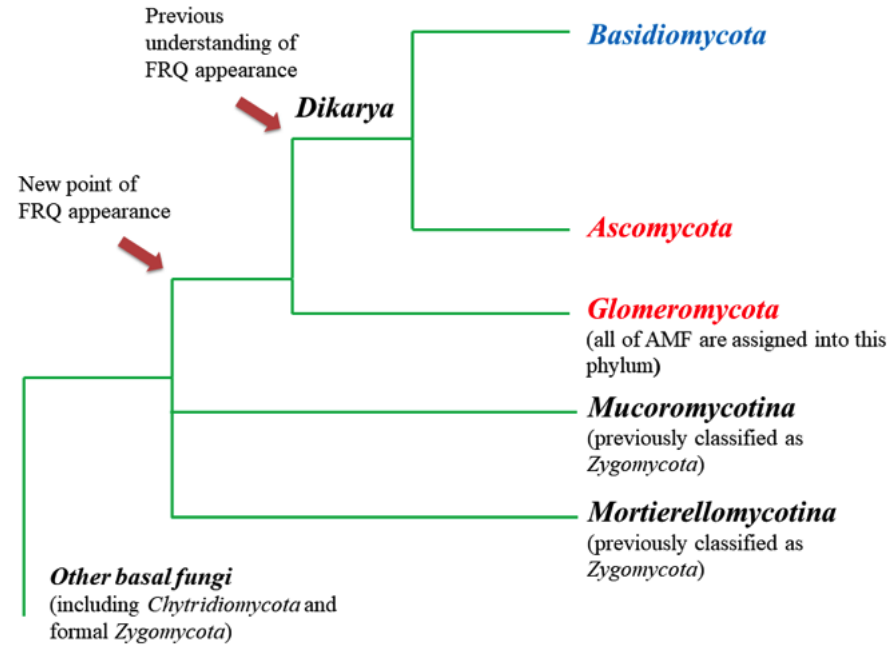
3.6.4. Evolution of the circadian core gene (*frq*) in Fungi

Previous work has indicated that *frq* evolved differently among different fungal lineages, and it was proposed that *frq* was acquired after the divergence of Ascomycota and Basidiomycota from their common ancestor (Salichos and Rokas 2010). Subsequent work revealed a homolog of *frq* in the Basidiomycotan fungus *Rhodosporidium toruloides,* and confirmed the lack of *frq* in non-Dikarya fungi (Zhu, et al. 2012). This suggested the gain of *frq* occurred prior to the divergence of Dikarya. We have searched for homologs of *frq* in available databases (GenBank and JGI) of fungi that do not belong to Dikarya and Glomeromycota, and were unable to find any. The presence of a *frq* homolog in *R. irregulare* is thus surprising because the Glomeromycota phylum diverged before the Dikarya. This suggests that *frq* acquisition in fungi is much older than previously thought.

Our finding of a FRQ homolog in *R. irregulare* changes our understanding of *frq* gene evolution in the fungi. Previously, based on rRNA sequences and morphological identification, the AMF were classified into the phylum Glomeromycota (Hibbett, et al. 2007; Schüβler, et al. 2001). This phylogeny would fit the observed *frq* gene distribution by assuming a single appearance of *frq* before the divergence of Dikarya and Glomeromycota (Figure 3.7a). In contrast to this, a recent study using a set of orthologous proteins (Spatafora, et al. 2016) proposed a sub-phylum Glomeromycotina as a sister group of Mucoromycotina. This phylogeny seems less parsimonious because it would require the extra step of *frq* gene loss in the Mucoromycotina (Figure 3.7b).

Figure 3.7

a. Conventional tree (reference (Hibbett, et al. 2007))



b. Conventional tree (reference (Spatafora, et al. 2016))

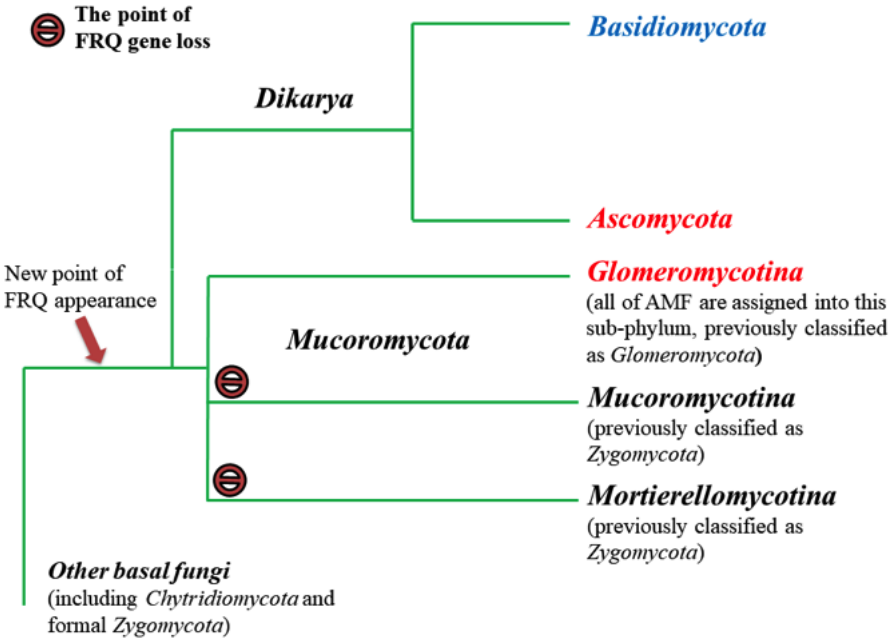


Figure 3.7. Phylogenetic distribution of FRQ in the kingdom Fungi. Two conventional fungal phylogenetic trees regarding *Glomeromycota* were built based on two references (a) (Hibbett, et al. 2007) and (b) (Spatafora, et al. 2016). The formal phylogenetic analysis of FRQ

homologs were used as references to indicate FRQ conservation (Salichos and Rokas 2010). The clades with defined FRQ homologs were colored in red. Basidiomycota was colored in blue, since the *frq* homolog recently reported in this clade has not yet been studied in-depth (Zhu, et al. 2012). The classical and new points of FRQ appearance are indicated by arrows in (a) and (b).

3.7. Acknowledgements

This work was supported by an NSERC Discovery grant to MH which is gratefully acknowledged. We thank Miss Yerim Heo for assistance in designing and editing some figures. We also thank Dr Dave Janos and two anonymous reviewers for their helpful comments.

Chapter 4 – Importance of Chronobiology in AM symbiosis

Presentation of article 3

In this opinion letter, we reviewed our current understanding of chronobiology regarding the intimate symbiosis and pointed out there is no consideration of circadian system in holobiont. We further emphasize the importance chronobiological study of AM symbiosis which may advance our understanding of the circadian clock and its role in holobionts, especially in plant holobiont which is crucial for understanding the ecosystem. Furthermore, we explained why AM symbiosis can be the potential model system of this field of research.

This manuscript was prepared for journal submission. I have written this manuscript with revisions from Dr. D. Morse and Dr. M. Hijri.

Article 3 - On the chronobiology of holobionts, can mycorrhizal symbiosis be used as a model?

Soon-Jae Lee, David Morse and Mohamed Hijri

Institut de Recherche en Biologie Végétale, Université de Montréal, 4101 Rue Sherbrooke Est, Montréal (Québec) H1X 2B2, Canada

Manuscript was prepared for journal submission

4.1. Abstract

Circadian clocks are an endogenous and virtually ubiquitous mechanism that orchestrate up to 15% of the transcriptome in numerous organisms. Thus, study of circadian clock is essential to understand the function of an organism in nature. Clocks are becoming well understood in plants and animals, yet both these types of organisms live with a complex and diverse microbiota which must be taken into account when examining their biology. Symbiotic relationships between multicellular organisms and their microbiota have evolved together over time and form a holobiont. Interrelated regulation of holobiont circadian rhythms may enable species to temporally synchronize or shift their activities in the community. Mechanisms that may regulate interspecies temporal adaptations or how the circadian clock might evolve in holobionts is far from being understood, although some studies have shown linkages between the circadian rhythms of partners in several symbioses. Arbuscular mycorrhizal (AM) symbiosis is one of the oldest and most studied associations and is formed by plant roots and fungi belonging to Glomeromycota. By mediating the flux of nutrition between the plant and a myriad of other microbes in the soil, AM symbiosis comprise backbone of plant holobionts. Even though the importance of the AM symbiosis has been well recognized in agriculture and terrestrial ecology, the chronobiology of the AM partners remains almost unknown. We have begun to study the circadian clock of arbuscular mycorrhizal fungi and present and discuss initial data reported in the literature. We believe that analyzing the interrelated temporal organization of AM symbiosis and determining its underlying mechanisms will advance our understanding of the circadian clock and its role in holobionts.

4.2. Keywords

Arbuscular mycorrhizal fungi (AMF), Mycorrhizal symbiosis, Evolution, Circadian clock, Chronobiology, Holobiont, Frequency (FRQ) gene

4.3. Main body

4.3.1. Circadian rhythms in organisms

Circadian rhythms are ubiquitous and endogenously generated rhythms involved in various biological processes with periods of about 24 hours under constant conditions. The rhythms are generated and maintained by an endogenous timer termed a circadian clock. Both eukaryotes and prokaryotes have been shown to have circadian clocks which orchestrate various biological processes over time. For example, in numerous organisms, between 2 and 15% of the transcriptome varies in a circadian manner (Dibner, et al. 2010; McDonald and Rosbash 2001; Nagel and Kay 2012; Storch, et al. 2002). The biological rhythms in RNA levels were shown to be involved in a range of phenomenon from basal metabolic pathways to cellular growth and reproduction. Thus, understanding the circadian clock is an important part of understanding how an organism functions (Harmer, et al. 2001; Kondo and Ishiura 2000; Lewis and Feldman 1996; Millar and Kay 1997; Piggins 2002; Subramanian, et al. 2003).

Circadian clocks have been widely studied in various organisms, both unicellular and multicellular (Halberg, et al. 2005; Loza-Correa, et al. 2010; Mori and Johnson 2000). Regardless of the types of organism, unicellular or multicellular, each cell possesses its own cellular clock which regulates the biological rhythms of the individual cells (Harmer, et al. 2001). Because all cells in a multicellular organism share the same genomic information, all have essentially identical molecular clocks (Halberg, et al. 2005; James, et al. 2008; Millar and Kay 1997). However, they do not always behave similarly. For example, in vertebrates such as mice

and humans, a master clock in the brain synchronizes the circadian clocks in other cells of the body, and only the master clock is able to sustain its rhythm without external cues (Piggins 2002). The resulting coordinated rhythms can be conveyed through the endocrine system such as melatonin fluctuation which affects the sleeping cycle, blood pressure and metabolism (Czeisler and Gooley 2007; Johnston, et al. 2016; Zisapel 2018). In the case of plants, light and temperature signals interact with the circadian clocks in aerial parts of the organisms to regulate the rhythmicity of physiological and developmental processes including photosynthesis, leaf movement as well as flowering time (Halaban 1968; Mas and Yanovsky 2009; McClung 2006). Even though a "master clock" has not been reported in plants, clocks in the non-photosynthetic root cells of *Arabidopsis thaliana* receive timing cues from the circadian clock in photosynthetic cells (James, et al. 2008). Numerous studies have reported the existence of multiple clocks within an organism, but all of these can ultimately be synchronized to environmental changes and help the organism to adjust its timing (Michael, et al. 2003; Piggins 2002; Subramanian, et al. 2003; Xu, et al. 2007). Thus, regardless of the number of cells composing an organism, it is believed that the circadian clocks in a given organism are interacting and sharing crucial roles to maintain the perpetuity of life of that organism.

The vast majority of circadian clock studies have been conducted with free-living model organisms because they are easy to work with (e.g. easy to cultivate and rapid reproduction) and there is an enormous amount of basic information available in public databases (e.g. genomes, transcriptomes, metabolomes, mutants, etc.) (Halberg, et al. 2005; Harmer, et al. 2001; James, et al. 2008; Kondo and Ishiura 2000; Lewis and Feldman 1996; Piggins 2002; Subramanian, et al. 2003). However, the concepts of "free-living" and "individual" are currently challenged as

meta-organismal studies have begun to question the validity of these concepts (Greer, et al. 2016; Haag 2018; Walter, et al. 2013).

4.3.2. Holobionts and the concept of an organism

Increasing numbers of meta-organismal studies in plants and animals together with their microbiota have revealed that the biology of an individual organism is not simply a function of the biology of one species (Greer, et al. 2016; Walter, et al. 2013). In essence, the body of an organism does not consist only of its own cells, but also includes those of its microbiome (Gilbert, et al. 2012; Haag 2018; Rosenberg, et al. 2010). It has been shown that an organism can be considered as a system comprised of its own species as well as its symbiotic organisms (Greer, et al. 2016; Margulis and Fester 1991; Walter, et al. 2013). In the case of humans, the body contains as many bacterial cells as its own cells (Sender, et al. 2016). The studies of microbiomes interacting with human body have revealed that the signal cross-talk between microbiomes and human cells can modulate gene expression in each partner with respect to the various biological process including metabolism and immune responses (Gilbert, et al. 2012; Greer, et al. 2016; Walter, et al. 2013). For plants, meta-organismal studies including animals (herbivore invertebrates such as nematode worms) as well as microbiomes, provided a better understanding of plant physiology, ecology and evolution (Beans 2017; Rosenberg, et al. 2010). For instance, a massive number of studies have been conducted on the rhizosphere, the soil region around the roots, considered as hotspots of microbial biodiversity. The rhizosphere contains a variety of microorganisms, including fungi, capable of interacting with the host plant. The host plant modulates and shapes the rhizosphere microbial communities by releasing root-exudates acting as a source of carbon and other chemicals as well as by secreting various

signaling molecules such as phytohormones (Dodd, et al. 2015; Dodd and Ruiz-Lozano 2012). In turn, the rhizosphere microbes affect soil and plant nutrition status, as well as plant growth and immunity, a process known as plant-soil feedback (Bever 1994; Gopalakrishnan, et al. 2015; Jansa, et al. 2013).

Symbiotic meta-organisms, or holobionts, have evolved together under natural selection (Margulis and Fester 1991; Mindell 1992). The term holobiont was first coined by Margulis and Fester (1991) to describe the symbiotic associations throughout a significant portion of an organism's lifetime, with the prefix holo- derived from the Greek word *holos* meaning whole or entire (Bordenstein and Theis 2015; Margulis and Fester 1991). The term remained dormant for about a decade until the advancements that have been made in next-generation sequencing which resulted in an increasing number of host and microbiome studies, where term holobiont resurfaced to describe such assemblages. High-throughput sequencing and bioinformatics revealed the complementarity of various functional genes in the genomes of hosts and their associated microbes, which were considered as hologenome (Bordenstein and Theis 2015; Gilbert, et al. 2012; Rosenberg, et al. 2010). In nature, the organisms of the holobiont are not only physically connected resulting in changes of the morphology, but also at molecular level allowing cross-talk. Moreover, the hologenomes of the holobionts have evolved interlocking metabolic or immunological pathways (Gilbert, et al. 2012; Haag 2018; Margulis and Fester 1991; Rosenberg, et al. 2010; Walter, et al. 2013). In short, in both physiological and genomic context, the organisms comprising holobionts have evolved to function together. Interestingly though, the chronobiology of the holobionts, and the regulation of biological processes in each species over time, has not yet been fully described.

4.3.3. Symbiosis and circadian rhythms

Circadian rhythms have been studied in several organisms now considered as holobionts. For example, a number of investigations have examined the linkage of circadian rhythmicity between the intestinal microbiome and its host. In mammals such as mice and humans, the composition and functional profiles of intestinal microbiomes exhibited diurnal oscillations that are influenced by host feeding rhythms (Thaiss, et al. 2014). It was also reported that the total number of bacterial cells in the mouse gut fluctuate over the day/night cycle, and that this fluctuation involved the mammalian core clock protein BMAL1 (Liang, et al. 2015). At the same time, signaling molecules of microbes were required for the functioning of the circadian clock of host intestinal epithelial cells (Mukherji, et al. 2013).

Corals in the oceans are one of the most famous symbiotic associations, formed between anthozoan hosts and photosynthetic dinoflagellate symbionts. The symbionts live enclosed within "symbiosomes" which lie within the host cells of the gastroderm. Corals acquire photosynthates from the symbionts to fuel their metabolic needs, while symbionts receive from the host essential inorganic nutrients required for their growth (Blackall, et al. 2015). Over 210 million years of evolution of coral symbiotic partners, their genomes also underwent the selective pressure on genes involved in the symbiosis (Schwarz, et al. 2008). Linkages between the circadian clock of the two partners of the coral holobiont have also been reported. Dinoflagellate symbiont activities were regulated simultaneously over the day/night cycle (Levy, et al. 2011; Sorek, et al. 2014). The expression of symbiosis-related genes such as those encoding stress-related chaperones and antioxidants of coral hosts began to be transcribed before the peak oxygen production of the symbiont (Levy, et al. 2011; Sorek, et al. 2013).

The symbiosis between the squid *Euprymna scolopes* and the bioluminescent bacterium *Vibrio fischeri* is also an interesting model to study holobionts. The association between the host and symbionts is mediated by the unique light-producing organ formed in host epithelia following the establishment of the symbiosis. The host uses the light produced by *V. fischeri* to avoid predators during the nights out, while the bacteria can be fed and live protected inside the epithelial cells of the host (Heath-Heckman, et al. 2013; Wier, et al. 2010). The population size of *V. fischeri* has been shown to be under the control of host circadian rhythm (Wier, et al. 2010). It was also reported that the daily rhythmicity of the light produced by *V. fischeri* modulates the rhythmic expression of host transcriptome (Heath-Heckman, et al. 2013).

Another popular example of a holobiont is lichen, a symbiotic association between heterotrophic fungi with internal autotrophic photosynthetic algal and/or cyanobacteria as well as other bacteria (Honegger, et al. 2013; Vancurova, et al. 2018). The morphology of lichen can vary dramatically depending on the taxonomy of the partners, particularly fungi. In general, the thallus of lichen comprises a stratified fungal structure, with peripheric layers formed by a fungal mesh glued together in an extracellular matrix that shelters an internal layer dominated by the algal partner as well as various bacterial partners (Ahmadjian 1994). The role of each partner in this holobiont is highly specific (Ahmadjian 1994; del Campo, et al. 2013; Eymann, et al. 2017). A recent meta-proteomic approach using *Lobaria pulmonaria* revealed that various partner-specific proteins are expressed at the same time. For example, vesicle transporter proteins were expressed from the fungi, photosynthesis-related proteins were mostly from algae, nitrogenase and nitrogen fixation proteins came from cyanobacteria and the most of methanol metabolism as well as CO-detoxification enzymes were produced by bacteria (Eymann, et al. 2017). Yet, unfortunately, there are no reports of circadian rhythms in lichen holobionts, even

though each individual symbiont can have its own circadian rhythm (Kondo and Ishiura 2000; Lewis and Feldman 1996; Mittag 2001).

It seems clear that there are coordination and synchronization between members involved in the symbiosis of the holobiont. However, in the most cases, the systematic understanding of the chronobiology in the holobiont remains in the dark and, as observed in the case of lichen holobiont, is sometimes not even considered.

4.3.4. Plant and arbuscular mycorrhizal (AM) symbiosis

In general, circadian clock research in plants has rarely been concerned with symbiosis, even though the daily fluctuations in the light that affect photosynthesis and the plant's circadian rhythms have been well studied (Millar and Kay 1997). Plants are unique in that their aerial parts and roots are essentially two different ecosystems (aboveground and belowground). This organization elevates the importance of understanding the chronobiology of the plant as well and its microbiome.

Interestingly, it has been reported that the circadian rhythms in the roots are coordinated by the aerial part which synchronizes the rhythms to sunshine (James, et al. 2008), just as one might expect for a single free-living organism. However, when we consider the plant holobiont, the plant is far from being a single free-living organism, because its organs are associated with diverse microorganisms (Beans 2017; Hill, et al. 2018). One of the most important contributors to the plant holobiont is the arbuscular mycorrhizal symbiosis. This AM symbiosis is formed by plant roots and arbuscular mycorrhizal fungi (AMF) which are the oldest symbiotic partners of plants and have evolved over 460 million years ago together with the plant ancestors during their initial emergence and colonization of land (Simon, et al. 1993). They are also the most

ubiquitous symbiotic partner of plants as they are able to form symbiosis with more than 80% of vascular plant species (Smith and Read 2008). AMF are obligatory plant root symbionts which depend on photosynthesis of host plants for their source of carbon. In exchange, AMF help plants to absorb the various essential soil nutrients, especially phosphorus and nitrogen, by transferring these nutrients through their hyphae which have grown into and colonized plant root cells in which they form a structure called an arbuscule (Smith and Read 2008; Thirkell, et al. 2016). The intimate symbiosis between plants and AMF is called arbuscular mycorrhizal symbiosis and is an essential element of plant physiology. The intimacy of this mycorrhizal symbiosis is not only observed by the arbuscular structure, but also by the obligate biotrophy od the fungal partner. The nutrient allocation and interlocked metabolic pathways between the fungus and the host underwent selective pressure as symbiotic partners (Jansa, et al. 2013; Schüβler, et al. 2001; Smith and Read 2008). Furthermore, AMF hyphae act as an ecological niche for various soil bacteria and other fungi (Jansa, et al. 2013), thus forming the backbone of the rhizospheric part of the plant holobiont. Considering the previous cases of human and its intestinal microbiome or coral symbiosis where circadian clocks are linked between partners, we raise the question of whether mycorrhizal symbiosis can also be controlled by a circadian clock. A prerequisite question would then be: do AMF possess a circadian clock?

4.3.5. AMF and circadian clock

In the circadian model fungus, *Neurospora crassa*, circadian rhythms are mediated by a fungal circadian clock whose molecular components include the circadian oscillator *frequency* (*frq*) and white-collar complex (WCC) genes as core components (Dunlap and Loros 2005; Lewis and Feldman 1996; Loros and Dunlap 2001). The periodical fluctuation of both *frq*

mRNA and FRQ protein are tightly linked due to the inhibition of *frq* transcription when FRQ interferes with the transcriptional activity of its transcription activator, WCC. FRQ recruits an FRQ-interacting RNA helicase (FRH), as well as protein kinases (Casein Kinases I (CKI) and II (CKII)) that together enter the nucleus and promote phosphorylation of WCC. The phosphorylated WCC is exported to the cytoplasm, resulting in a decrease in *frq* transcription. At the same time, FRQ also becomes progressively phosphorylated over time, leading to its eventual recognition by a ubiquitin ligase and proteolysis, after which WCC again becomes active thus restarting *frq* transcription and a new circadian cycle (Gallego and Virshup 2007; Loros and Dunlap 2001). Interestingly, AMF express the conserved circadian clock components *frq* and the white-collar complex (WCC) (Lee, et al. 2018b). The conservation of *frq* is especially notable because there is no other function reported so far, other than that of circadian oscillator of this gene with its unique protein structure (Lewis and Feldman 1996; Loros and Dunlap 2001; Sancar, et al. 2012). Indeed, the transcription of *frq* was induced by blue light exposure of AMF, which is mediated by WCC, similar to the function of the circadian clock in *N. crassa* (Collett, et al. 2002; Lee, et al. 2018b; Loros and Dunlap 2001).

The circadian clock is thought to be beneficial because it allows the biology of an organism to be in resonance with its surrounding environment (Johnson, et al. 1998; Ouyang, et al. 1998). In *N. crassa* and other free-living fungal species belonging to genus *Aspergillus*, it was found that the principal environmental factors that can influence the timing of the fungal circadian clock are light, temperature and nutrition (Harmer, et al. 2001; Hurley, et al. 2016b; Lewis and Feldman 1996; Sancar, et al. 2012). Because AMF live underground, light and temperature are unlikely to be the main environmental timing cues. Depending on the soil composition, light penetrates only a few centimeters into the soil and the temperature becomes

more and more constant as the depth increases (Ciani, et al. 2005; Shukla, et al. 2013; Smith 1932). Plant roots can easily grow deep enough to preclude light signals and to have a constant temperature (Maeght, et al. 2013; Oosterhuis 1990; Stone and Kalisz 1991). One typical AMF host plant, *Gossypium,* has roots that grow down 90 cm over a 50 day period (Oosterhuis 1990). Another host plant, *Sorghastrum nutans,* has roots 270 cm below the surface. In what may be a record, roots of the AMF host plant *Juniperus monosperma* are up to 61 m underground (Stone and Kalisz 1991). AMF are likely to be present even quite deep, with many different AMF found in rhizospheric soil acquired from a depth of 40 cm (Shukla, et al. 2013). Thus, any putative circadian clock system in AMF is likely to have undergone a selective pressure to favor nutrition as a timing cue. As briefly introduced previously, the time of feeding is a circadian timing cue of various eukaryotes including mammals and fungi (Hurley, et al. 2016b; Mendoza, et al. 2005; Sancar, et al. 2012). In the case of mammals, the master clock in the brain is affected by energy metabolism and affects in turn pineal melatonin levels and vasopressin mRNA cycles (Challet, et al. 2003; Mendoza, et al. 2005). In the case of the fungi *N. crassa,* the level of glucose uptake can modulate the circadian feedback-loop via expression of the circadian repressor CSP1 which affects the core oscillator (Sancar, et al. 2012). Interestingly, the diurnal pattern of hyphae growth of AMF has a peak in the daytime which coincides with the highest photosynthetic productivity of plants (Hernandez and Allen 2013).

4.3.6. Presence of a circadian clock allows for resonance with other clocks

Is having a circadian clock in AMF important if they receive timing cues from plant roots? Having a functional circadian clock in cells in a multicellular organism can be crucial for the synchronization of circadian rhythms between the cells (D'Alessandro, et al. 2015; Pando,

et al. 2002). For example, mammals have a central (master) circadian clock in the suprachiasmatic nucleus of the brain and peripheral oscillators in all other cells. These peripheral clocks, while not self-sustaining, are nonetheless required for entrainment to the signals provided by the master clock. Mouse embryo fibroblasts (MEF) lacking a functioning clock are not rhythmic after subcutaneous transplantation in a wild-type host, while MEF with a functional clock are successfully synchronized even if their endogenous period differs from that of the host (Pando, et al. 2002). Possession of a circadian clock is essential for providing a sense of time to an organism, however, the clock's components could be lost during evolution in the cases of the organisms which does not need to sync its biological rhythm with environment (Halberg, et al. 2005; Loza-Correa, et al. 2010; Mori and Johnson 2000). It is not the case of AMF, because the circadian clock components and several output genes were all conserved in AMF genome during its long time evolution as plant root symbiont (Lee, et al. 2018b). We propose that the presence of a circadian clock components in AMF may have an important functional role to interact with the clock of host plant.

4.3.7. Role of AMF in organizing rhizosphere components

Despite the paucity of research in this area to date, a better understanding of holobionts will include an understanding of their chronobiology. In the case of the AM symbiosis, understanding how different events are synchronized in time will help to understand the intimate connections which link the different rhizosphere components of the plant holobiont. AMF play a key role in organizing the plant rhizosphere by acting as liaison between plant roots and the soil microbial community (Smith and Read, 2008;Jansa et al., 2013;Dodd et al., 2015). There are non-photosynthetic prokaryotes in which circadian rhythms have been found, including

Klebsiella pneumonia (Halberg et al., 2003) and Pseudomonas putida, a well-known soil bacterium whose colony growth pattern seems to be affected and synchronized by direct exposure to light (Soriano et al., 2010). However, in nature, these microbes are in the soil, not at the surface, and should not be able to directly sense the general circadian entrainment cues such as light or temperature fluctuations, especially if they are deep underground. For this reason, nutrition fluctuations initiated from above-ground regions of the plants are likely to be important for the soil microbes, especially those who depend on the plants as a carbon source. Initial light information received by the above-ground parts of the plants is transduced into a carbon flux. Indeed, this is the main reason why a large variety of soil microbes interact intimately with autotrophs, especially plants (Bever, 1994;Gopalakrishnan et al., 2015;Beans, 2017). At the same time, AMF are the major nutrition and signaling mediators between plants and the soil microbial community, so they will ultimately shape the community structure (Solis-Dominguez et al., 2011;Cruz and Ishii, 2012;Jansa et al., 2013;Meglouli et al., 2018;Pankova et al., 2018). Thus, to understand true chronobiological connection between above- and below-ground component of plant holobiont, it is essential to understand circadian system of AM symbiosis. What kind of mechanisms might connect the circadian systems between the partners of AM symbiosis? One potential link between AMF and the plant is likely to be carbon, which was shown to synchronize the fungal circadian clock (Figure 4.1). AMF rely entirely on the carbon source supplied by the host plant (Smith and Read, 2008), where the carbon fluxes are under control of the circadian clock (Millar and Kay, 1997;Muller et al., 2014), so it is plausible that the clock system in AMF could have evolved to be entrained by photosynthetic carbon flux from the host plant. Carbon is not the only possible link, though. Numerous studies have reported that the level of phosphorus or nitrogen uptake in host plant can modulate root colonization by AMF

(Alarcon et al., 2002;Blanke et al., 2005;Mohammad et al., 2005;Ohtomo and Saito, 2005;Chen et al., 2007;Wang et al., 2018), and the host genes coding phosphorus transporters protein or nitrogen assimilatory proteins were also clock controlled (Gutierrez et al., 2008;Wang et al., 2011). In plant-AMF symbiosis, phosphorus, nitrogen, and carbon exchanges are the three core processes related to establishment and maintenance of mycorrhizal symbiosis (Bever, 1994;Blanke et al., 2005;Mohammad et al., 2005;Ohtomo and Saito, 2005;Smith and Read, 2008;Jansa et al., 2013) so all of these are potential candidates for crosstalk between the circadian clocks of AMF and its host plants.

Figure 4.1

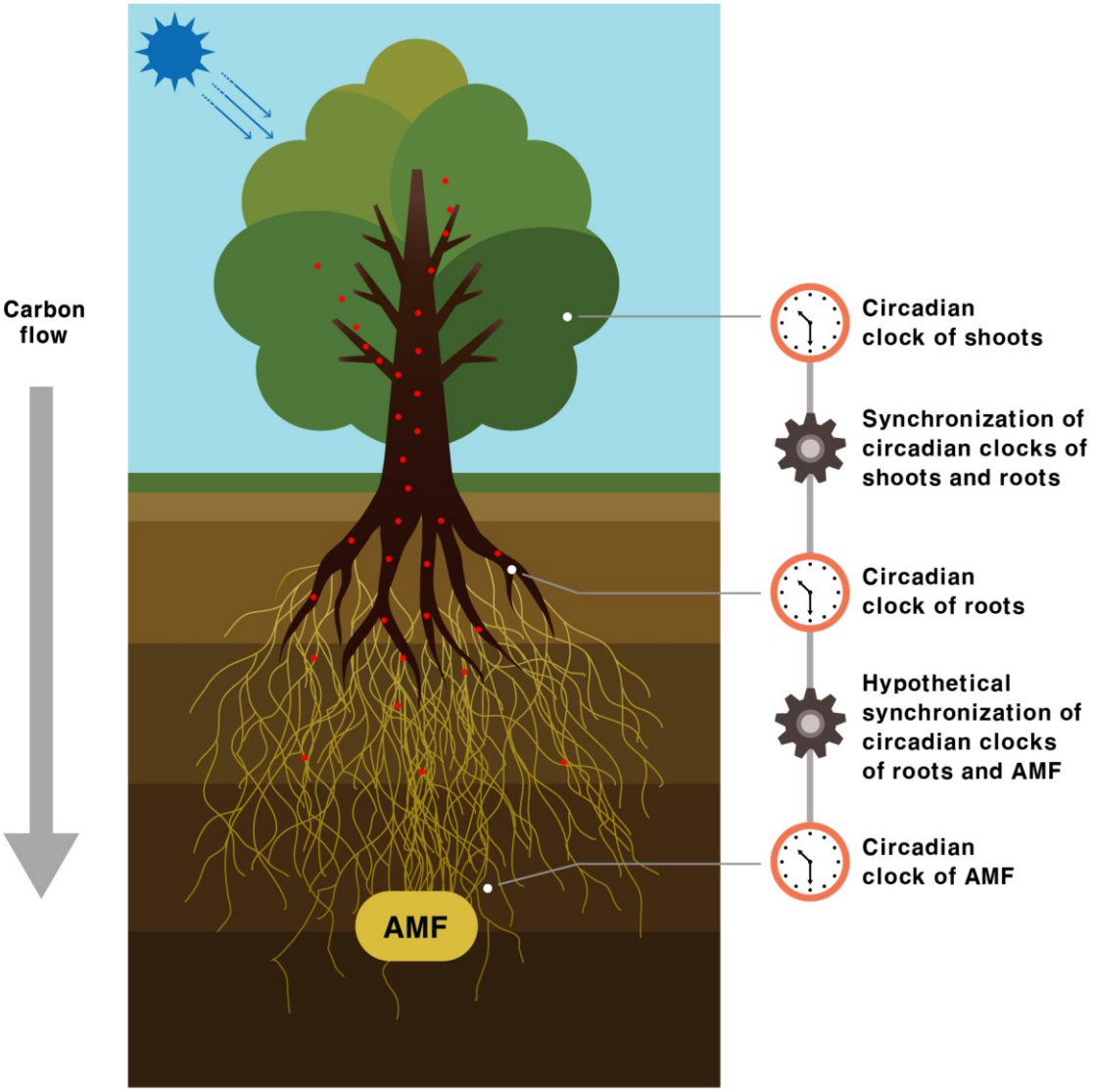


Figure 4.1. Hypothetical mechanism for synchronizing plant-AMF circadian rhythms. The flux of carbon is indicated at left and presented as red dots in the figure. The synchronization of circadian clocks in plant roots with those in shoots was reported previously (James, et al. 2008).

4.3.8. Conclusion

To date, most studies of signal transduction as well as the network studies of an organism or a symbiosis have been based on the concept that there are linear linkages between the components. However, we now know links between components can be modulated by circadian clocks. Considering the intimacy and complexity of signals between the symbionts comprising holobionts, understanding the chronobiology of the system as a whole should be emphasized.

As circadian systems in individual cells of multicellular organisms function as one system, the hypothesis that the circadian clocks of different holobiont components to function as one system should be addressed. It is becoming clear in intimate symbiosis such as that of a holobiont that the biological activity of one of the partners could produce circadian cues for the other partner (Heath-Heckman et al., 2013;Thaiss et al., 2014;Liang et al., 2015). It remains to be determined to what degree linking or interlocking of circadian clocks between organisms within holobiont will be necessary and achieved.

Despite the importance of chronobiology in plant or plant interacting microbes, chronobiological understanding of plant holobiont far from understanding. There have been few circadian clocks reported in fungal and bacterial species which interact with plants (Lewis and Feldman, 1996;Halberg et al., 2003;Soriano et al., 2010), and all of them are free-living and opportunistic symbionts which have not necessarily evolved close interactions with their plant hosts. In contrast, not only have crucial impact on the nutrition flux in plant holobiont, but AMF have also evolved over 460 million years as obligatory root symbionts which intimately share their biological processes with host plant (Simon et al., 1993;Alarcon et al., 2002;Blanke et al., 2005;Smith and Read, 2008;Jansa et al., 2013). Plants live in an environment whose conditions vary over a 24 hour period, and this is likely to be reflected in nutritional requirements met by

the rhizosphere. Thus, the chronobiology of AMF can be a good model for studying the circadian clock in the obligatory symbionts which have colonized the majority of land plant species. Considering the importance of plants in agriculture and terrestrial ecology, there is a need for increased research into the chronobiology of mycorrhizal symbioses. The addition of the time dimension would help us to reveal the core connections between aboveground and underground function, thus linking underground oscillations with the sky.

4.4. Acknowledgements

We thank Miss Yerim Heo for assistance in designing and editing of the figure. We also thank the helpful comments from Dr. Marc St-Arnaud.

Chapter 5 - General discussion, Conclusion and Perspectives

My Ph.D. thesis contributes to expanding our fundamental knowledge on the evolution of three important mechanisms of gene regulation (RNAi system, the blue light sensing mechanism and the circadian clock) inside AMF, the oldest and most ubiquitous symbiotic partner of plants. It clearly demonstrates the conservation and expression of these mechanisms in the model AMF *R. irregularis*. The RNAi system in *R. irregularis* (Chapter 2) has evolved to harbor the enormously large numbers of AGO homologs compared to the general free-living fungal species, which reflects the complex sRNAs regulation mechanisms related with symbiosis. The most striking finding was in another core protein of RNAi system, DICER (class IV ribonuclease III enzyme). Unlike other fungal species which have at least two fungal ribonuclease III enzymes, *R. irregularis* only possessed one fungal ribonuclease III enzyme (RIDCL 1). Instead, I discovered two bacterial ribonuclease III enzymes (RIRNC 2 and 3) which were horizontally transferred from cyanobacteria. I further proved that at least one of the transferred gene (*rirnc 3*) was conserved among other relative AMF species with high sequence similarities. The conservation reflects the long-time evolution of *rirnc 3* along the diversification of AMF. Interestingly, despite its well acknowledged characteristic of depending carbon from autotrophic land plant, recent investigations of molecular clocks indicated that the appearance of Glomeromycota is much older than the emergence of land plants on earth from the ocean (Berbee and Taylor 2010; Bidartondo, et al. 2011). At the same time, there is a basal Glomeromycota species reported to make the symbiosis with autotrophic cyanobacteria, instead

of land plant (Gehrig, et al. 1996). Due to the lack of the available AMF genome to cover entire Glomeromycota, I was not able to prove the conservation of cyanobacterial ribonuclease III in entire Glomeromycota. However, the conservation of the genetic fossil is the first molecular proof that shows intimacy between Glomeromycotan fungi and cyanobacteria to transfer the core enzyme of gene regulation. The genome sequencing projects of AMF are on-going. In the future, with increasing genomic data of AMF taxa, we could design the degenerated primer of *rirnc 3* to investigate the larger coverage of AMF species or Glomeromycota related species and clarify the point of HGT in Glomeromycota. Since the basal glomeromycotan fungus *Geosiphon* is reported as a living example of symbiosis between AMF and cyanobacteria, future investigation on this species is a priority. The finding of conserved RNAi system in AMF has opened the possibility for investigating this organism with the application of RNAi mechanism, as well as the future research regarding the mycovirus which can trigger the RNAi system of the organism.

In my Ph.D. study, I also proved that blue light perception mechanism with WC proteins and linked circadian clock components were conserved inside of *R. irregularis* (Chapter 3). Among the three different types of blue light sensing mechanisms of fungi, only WC proteins-based mechanism was conserved in *R. irregularis*, proteins which also play a role in the fungal circadian clock. The fungal circadian core oscillator FRQ is a unique protein found only in fungal kingdom which has never been reported with any other function except as part of the circadian oscillator. In *R. irregularis*, this gene is expressed regardless of the mycorrhizal symbiosis, and its expression can be induced by blue light just as in the model circadian fungus *N. crassa*. The study is the first finding of a circadian clock system conserved in non-free-living fungi, which opens the door to do the research of chronobiology in AM symbiosis. I further

emphasized the importance and necessity of changing our view of chronobiology for biological individuals (Chapter 4). The concept of biological individuals is now shifting from an organism to a holobiont. Up to date, there were few studies conducted on circadian clocks in symbiotic organisms. However, the systematic understanding of circadian clocks inside symbiosis is yet to be understood. Moreover, the plant holobiont was even not considered in this context. Chronobiological understanding of AM symbiosis has important impacts on its biology. Because (1) plants are the initial photosynthetic provider of land ecosystem; (2) plants connect two different ecosystems (aboveground and belowground) by its shoots and roots; (3) AMF are obligatory symbionts which cannot survive without a host; both partners of symbiosis appear tightly linked by nutrition signaling rather than other circadian cues such as light or temperature. The circadian clock is basically the evolved system of an organism to coordinate its biological rhythms with the environmental rhythms. Up to date, light, temperature and nutrition were acknowledged as the important circadian cues. However, it is hard to separate these cues when studying the circadian clocks in symbiotic organisms. For example, in the coral symbiosis, the dinoflagellate symbiont and host were exposed to the light or temperature changes in nature at the same time due to the form of the symbiosis. This is also true in the case of symbiosis between the squid *Euprymna scolopes* and the bioluminescent bacterium *Vibrio fischeri.* In the case of human and gut microbiome, light can be removed from circadian cue since the symbiont grows under light-free conditions. However, the temperature of the host is still shared with the symbiont. AM symbiosis is unique, however, since the reception part of light and temperature is limited to the aboveground and surface area of the soil while the actual environment that symbiont faces is deep soil which is out of range of light or temperature changes. At the same time, the carbon flux from the plant host to the symbiont is tightly linked as AMF absolutely

depend on carbon derived from the host. Thus, the study of the chronobiology in AM symbiosis will help us to understand the function of nutrition in the circadian clock as well as carbon flux in the rhizosphere mediated by AMF hyphae which is crucial to understand the function and diversity of soil microbes.

The pattern of *frq* gene gain and loss found in my Ph.D. study also give insights on not only the evolution of this gene, but also the evolution of Glomeromycota. Importantly, the finding of *frq* inside *R. irregularis* is a clear counterexample of our previous understanding that *frq* appeared after the diversification of Dikarya during the evolution of fungi. The unique occurrence of prokaryotic ribonuclease III inside of *R. irregularis* and its relative AMF is a great counterexample of the statement that prokaryotic ribonuclease III does not exist in eukaryote. Classically, based on 18s rDNA sequences and morphological identification, all AMF were classified into the phylum Glomeromycota (Hibbett, et al. 2007; Schüβler, et al. 2001). However, the phylogenetic position of Glomeromycota and one species classification in this group is still under debate. A recent study using a set of orthologous proteins (Spatafora, et al. 2016) suggested merging Glomeromycota with Mucoromycota to form the subphylum Glomeromycotina. At the same time, research of fine endophytes using only partial sequence of 18S rDNA (260bp) have claimed that *Glomus tenue* should be classified as Mucoromycotina, not Glomeromycota (Orchard, et al. 2017). Because of this study, the classification of AMF in fungal kingdom became no longer monophyletic. However, interestingly, recent morphological investigation regarding spores, vesicles and arbuscule structures found there is no suitable genus inside Mucoromycotina for classifying *Glomus tenue.* Thus, new genus *Planticonsortium* was created under Mucoromycotina to classify this fungus (Walker, et al. 2018). However, none of these studies has considered one of the most crucial indicator of the evolution, that of gene gain

and loss. Even though I was not able to examine the existence of *rirnc 3* in all the Glomeromycota, this gene was conserved among all of the available AMF species. Thus, further investigation of *rirnc 3* inside a broader range of AMF species has great potential to be an indicator to distinguish the Glomeromycota from other fungal species. The evolution of circadian clock genes in *R. irregularis* also gives insight into the evolution of Glomeromycota. Considering the appearance of *frq* gene, Glomeromycota has closer relationship with Dikarya than Mucoromycotina, which supports the classical assignment of Glomeromycota as an independent phylum distinguished from Mucoromycota. Even though I am facing a lack of genetic information covering the entire Glomeromycota, I believe the findings from my Ph.D. study have great potential to clarify the evolution of Glomeromycota and its phylogenetic position in fungal kingdom. As time goes on, an accumulation of the genomes of AMF and related fungal taxa such as Mucoromycotina will let us approach the real answer of Glomeromycota evolution.

In conclusion, my thesis reports the evolutionary insights of AM symbiosis drawn from the conserved systems of three important gene regulation mechanisms (RNAi, blue light perception and circadian clock) of *R. irregularis.* The findings of my study open the door for deeper understanding of AM symbiosis which is the oldest and most ubiquitous symbiosis on earth.

Bibliographie

Ahmadjian V 1994. The Lichen Symbiosis. Nord. J. Bot. 14: 588. doi: https://doi.org/10.1111/j.1756-1051.1994.tb00653.x

Alexopolous C, Mims C, Blackwell M. 2004. Introductory Mycology, 4thed.: John Wiley and Sons, Hoboken NJ.

Altschul SF, et al. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research 25: 3389-3402.

Ambra R, et al. 2004. Photomorphogenesis in the hypogeous fungus Tuber borchii: isolation and characterization of Tbwc-1, the homologue of the blue-light photoreceptor of Neurospora crassa. Fungal genetics and biology : FG & B 41: 688-697.

Aravind L, Anantharaman V, Zhang D, de Souza RF, Iyer LM 2012. Gene flow and biological conflict systems in the origin and evolution of eukaryotes. Frontiers in cellular and infection microbiology 2: 89.

Bago B, et al. 2002. Translocation and utilization of fungal storage lipid in the arbuscular mycorrhizal symbiosis. Plant physiology 128: 108-124.

Beans C 2017. Core Concept: Probing the phytobiome to advance agriculture. Proceedings of the National Academy of Sciences of the United States of America 114: 8900-8902.

Bell-Pedersen D, Shinohara ML, Loros JJ, Dunlap JC 1996. Circadian clock-controlled genes isolated from Neurospora crassa are late night- to early morning-specific. Proceedings of the National Academy of Sciences of the United States of America 93: 13096-13101.

Bengtson S, et al. 2017. Fungus-like mycelial fossils in 2.4-billion-year-old vesicular basalt. Nature Ecology & Evolution 1: 0141. doi: 10.1038/s41559-017-0141

Berbee ML, Taylor JW 2010. Dating the molecular clock in fungi – how close are we? Fungal Biology Reviews 24: 1-16. doi: 10.1016/j.fbr.2010.03.001

Bever JD 1994. Feeback between Plants and Their Soil Communities in an Old Field Community. . Ecology 75: 1965–1977. doi: DOI: 10.2307/1941601

Bidartondo MI, et al. 2011. The dawn of symbiosis between plants and fungi. Biology letters 7: 574-577.

Blackall LL, Wilson B, van Oppen MJH 2015. Coral-the world's most diverse symbiotic ecosystem. Molecular ecology 24: 5330-5347.

Blanke V, et al. 2005. Nitrogen supply affects arbuscular mycorrhizal colonization of Artemisia vulgaris in a phosphate-polluted field site. The New phytologist 166: 981-992.

Bonfante P, Genre A 2010. Mechanisms underlying beneficial plant-fungus interactions in mycorrhizal symbiosis. Nature communications 1: 48.

Bordenstein SR, Theis KR 2015. Host Biology in Light of the Microbiome: Ten Principles of Holobionts and Hologenomes. PLoS biology 13: e1002226.

Branscheid A, Devers EA, May P, Krajinski F 2011. Distribution pattern of small RNA and degradome reads provides information on miRNA gene structure and regulation. Plant Signaling & Behavior 6: 1609-1611.

Buck AH, Blaxter M 2013. Functional diversification of Argonautes in nematodes: an expanding universe. Biochemical Society transactions 41: 881-886.

Chaib De Mares M, et al. 2015. Horizontal transfer of carbohydrate metabolism genes into ectomycorrhizal Amanita. The New phytologist 205: 1552-1564.
Challet E, Caldelas I, Graff C, Pevet P 2003. Synchronization of the molecular clockwork by light- and food-related cues in mammals. Biological chemistry 384: 711-719.
Chen YB, Dominic B, Mellon MT, Zehr JP 1998. Circadian rhythm of nitrogenase gene expression in the diazotrophic filamentous nonheterocystous cyanobacterium Trichodesmium sp. strain IMS 101. Journal of bacteriology 180: 3598-3605.
Cheng P, Yang Y, Heintzen C, Liu Y 2001. Coiled-coil domain-mediated FRQ-FRQ interaction is essential for its circadian clock function in Neurospora. The EMBO journal 20: 101-108.
Ciani A, Goss KU, Schwarzenbach RP 2005. Light penetration in soil and particulate minerals. European Journal of Soil Science 56: 561-574. doi: 10.1111/j.1365-2389.2005.00688.x
Collett MA, Garceau N, Dunlap JC, Loros JJ 2002. Light and clock expression of the Neurospora clock gene frequency is differentially driven by but dependent on WHITE COLLAR-2. Genetics 160: 149-158.
Conrad C, Schmitt JG, Evguenieva-Hackenberg E, Klug G 2002. One functional subunit is sufficient for catalytic activity and substrate specificity of Escherichia coli endoribonuclease III artificial heterodimers. FEBS letters 518: 93-96.
Corrochano LM 2011. Fungal photobiology: a synopsis. IMA fungus 2: 25-28.
Court DL, et al. 2013. RNase III: Genetics and function; structure and mechanism. Annual review of genetics 47: 405-431.
Czeisler CA, Gooley JJ 2007. Sleep and circadian rhythms in humans. Cold Spring Harbor symposia on quantitative biology 72: 579-597.
D'Alessandro M, et al. 2015. A tunable artificial circadian clock in clock-defective mice. Nature communications 6: 8587.
Damiani I, et al. 2012. Plant genes involved in harbouring symbiotic rhizobia or pathogenic nematodes. The New phytologist 194: 511-522.
Dang Y, Yang Q, Xue Z, Liu Y 2011. RNA interference in fungi: pathways, functions, and applications. Eukaryotic cell 10: 1148-1155.
Darriba D, Taboada GL, Doallo R, Posada D 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics (Oxford, England) 27: 1164-1165.
Davis CC, Xi Z 2015. Horizontal gene transfer in parasitic plants. Current opinion in plant biology 26: 14-19.
del Campo EM, et al. 2013. The genetic structure of the cosmopolitan three-partner lichen Ramalina farinacea evidences the concerted diversification of symbionts. FEMS microbiology ecology 83: 310-323.
Dibner C, Schibler U, Albrecht U 2010. The mammalian circadian timing system: organization and coordination of central and peripheral clocks. Annual review of physiology 72: 517-549.
Dodd IC, et al. 2015. The importance of soil drying and re-wetting in crop phytohormonal and nutritional responses to deficit irrigation. Journal of experimental botany 66: 2239-2252.
Dodd IC, Ruiz-Lozano JM 2012. Microbial enhancement of crop resource use efficiency. Current opinion in biotechnology 23: 236-242.
Dunlap JC, Loros JJ 2005. Analysis of circadian rhythms in Neurospora: overview of assays and genetic and molecular biological manipulation. Methods in enzymology 393: 3-22.
Dunlap JC, Loros JJ 2017. Making Time: Conservation of Biological Clocks from Fungi to Animals. Microbiology spectrum 5.

Edgar RC 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC bioinformatics 5: 113.
Estrada AF, Avalos J 2008. The White Collar protein WcoA of Fusarium fujikuroi is not essential for photocarotenogenesis, but is involved in the regulation of secondary metabolism and conidiation. Fungal genetics and biology : FG & B 45: 705-718.
Eymann C, et al. 2017. Symbiotic Interplay of Fungi, Algae, and Bacteria within the Lung Lichen Lobaria pulmonaria L. Hoffm. as Assessed by State-of-the-Art Metaproteomics. Journal of proteome research 16: 2160-2173.
Eyre-Walker A 1996. Synonymous codon bias is related to gene length in Escherichia coli: selection for translational accuracy? Molecular biology and evolution 13: 864-872.
Feldman JF, Hoyle MN 1973. Isolation of circadian clock mutants of Neurospora crassa. Genetics 75: 605-613.
Ferguson JJ, Menge JA 1982. Factors that affect production of endomycorrhizal inoculum. Proc. Fla. State Hort. Soc. 95: 37-39.
Formey D, et al. 2016. Regulation of Small RNAs and Corresponding Targets in Nod Factor-Induced Phaseolus vulgaris Root Hair Cells. International journal of molecular sciences 17.
Franken P, Lapopin L, Meyer-Gauen G, Gianinazzi-Pearson V 1997. RNA Accumulation and Genes Expressed in Spores of the Arbuscular Mycorrhizal Fungus, Gigaspora rosea. Mycologia 89: 293-297. doi: 10.2307/3761085
Gallego M, Virshup DM 2007. Post-translational modifications regulate the ticking of the circadian clock. Nature reviews Molecular cell biology 8: 139-148.
Gehrig H, Schussler 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 43: 71-81.
Gilbert SF, Sapp J, Tauber AI 2012. A symbiotic view of life: we have never been individuals. The Quarterly review of biology 87: 325-341.
Gopalakrishnan S, et al. 2015. Plant growth promoting rhizobia: challenges and opportunities. 3 Biotech 5: 355-377.
Greene AV, Keller N, Haas H, Bell-Pedersen D 2003. A circadian oscillator in Aspergillus spp. regulates daily development and gene expression. Eukaryotic cell 2: 231-237.
Greer R, Dong X, Morgun A, Shulzhenko N 2016. Investigating a holobiont: Microbiota perturbations and transkingdom networks. Gut microbes 7: 126-135.
Grimaldi B, et al. 2006. The Neurospora crassa White Collar-1 dependent blue light response requires acetylation of histone H3 lysine 14 by NGF-1. Molecular biology of the cell 17: 4576-4583.
Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic biology 59: 307-321.
Haag KL 2018. Holobionts and their hologenomes: Evolution with mixed modes of inheritance. Genetics and molecular biology: 0.
Halaban R 1968. The Circadian Rhythm of Leaf Movement of Coleus blumei x C. frederici, a Short Day Plant. II. The Effects of Light and Temperature Signals. Plant physiology 43: 1887-1893.
Halary S, et al. 2013. Mating type gene homologues and putative sex pheromone-sensing pathway in arbuscular mycorrhizal fungi, a presumably asexual plant root symbiont. PloS one 8: e80729. doi: 10.1371/journal.pone.0080729

Halberg F, et al. 2005. Prokaryotic and eukaryotic unicellular chronomics. Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie 59 Suppl 1: S192-202.
Harley JL, Smith SE. 1983. Mycorrhizal Symbiosis: Academic Press, London.
Harmer SL, Panda S, Kay SA 2001. Molecular bases of circadian rhythms. Annual review of cell and developmental biology 17: 215-253.
He D, Fu C-J, Baldauf SL 2016. Multiple Origins of Eukaryotic cox15 Suggest Horizontal Gene Transfer from Bacteria to Jakobid Mitochondrial DNA. Molecular biology and evolution 33: 122-133.
Heath-Heckman EAC, et al. 2013. Bacterial bioluminescence regulates expression of a host cryptochrome gene in the squid-Vibrio symbiosis. mBio 4.
Hernandez RR, Allen MF 2013. Diurnal patterns of productivity of arbuscular mycorrhizal fungi revealed with the Soil Ecosystem Observatory. The New phytologist 200: 547-557.
Hibbett DS, et al. 2007. A higher-level phylogenetic classification of the Fungi. Mycological Research 111: 509-547.
Hijri M, Sanders IR 2005. Low gene copy number shows that arbuscular mycorrhizal fungi inherit genetically different nuclei. Nature 433: 160-163.
Hill EM, et al. 2018. Arbuscular Mycorrhizal Fungi and Plant Chemical Defence: Effects of Colonisation on Aboveground and Belowground Metabolomes. Journal of chemical ecology 44: 198-208.
Honegger R, Edwards D, Axe L 2013. The earliest records of internally stratified cyanobacterial and algal lichens from the Lower Devonian of the Welsh Borderland. The New phytologist 197: 264-275.
Hou Y, Lin S 2009. Distinct gene number-genome size relationships for eukaryotes and non-eukaryotes: gene content estimation for dinoflagellate genomes. PloS one 4: e6978.
Huerta-Cepas J, et al. 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic acids research 44: D286-293.
Hurley JM, et al. 2014. Analysis of clock-regulated genes in Neurospora reveals widespread posttranscriptional control of metabolic potential. Proceedings of the National Academy of Sciences of the United States of America 111: 16995-17002.
Hurley JM, Loros JJ, Dunlap JC 2016a. Circadian Oscillators: Around the Transcription-Translation Feedback Loop and on to Output. Trends in biochemical sciences 41: 834-846.
Hurley JM, Loros JJ, Dunlap JC 2016b. The circadian system as an organizer of metabolism. Fungal genetics and biology : FG & B 90: 39-43.
Idnurm A, Heitman J 2005. Light controls growth and development via a conserved pathway in the fungal kingdom. PLoS biology 3: e95.
Idnurm A, et al. 2006. The Phycomyces madA gene encodes a blue-light photoreceptor for phototropism and other light responses. Proceedings of the National Academy of Sciences of the United States of America 103: 4546-4551.
Idnurm A, Verma S, Corrochano LM 2010. A glimpse into the basis of vision in the kingdom Mycota. Fungal genetics and biology : FG & B 47: 881-892.
James AB, et al. 2008. The circadian clock in Arabidopsis roots is a simplified slave version of the clock in shoots. Science (New York, N Y) 322: 1832-1835.
Jansa J, Bukovska P, Gryndler M 2013. Mycorrhizal hyphae as ecological niche for highly specialized hypersymbionts - or just soil free-riders? Frontiers in plant science 4: 134.

Johnson CH, Golden SS, Ishiura M, Kondo T 1996. Circadian clocks in prokaryotes. Molecular microbiology 21: 5-11.
Johnson CH, Golden SS, Kondo T 1998. Adaptive significance of circadian programs in cyanobacteria. Trends in microbiology 6: 407-410.
Johnston JD, Ordovas JM, Scheer FA, Turek FW 2016. Circadian Rhythms, Metabolism, and Chrononutrition in Rodents and Humans. Advances in nutrition (Bethesda, Md) 7: 399-406.
Katiyar-Agarwal S, Jin H 2010. Role of small RNAs in host-microbe interactions. Annual review of phytopathology 48: 225-246.
Ketting RF 2011. The many faces of RNAi. Developmental cell 20: 148-161.
Knip M, Constantin ME, Thordal-Christensen H 2014. Trans-kingdom cross-talk: small RNAs on the move. PLoS genetics 10: e1004602.
Kondo T, Ishiura M 2000. The circadian clock of cyanobacteria. BioEssays : news and reviews in molecular, cellular and developmental biology 22: 10-15.
Kuo C-H 2015. Scrambled and not-so-tiny genomes of fungal endosymbionts. Proceedings of the National Academy of Sciences of the United States of America 112: 7622-7623.
Lamontagne B, Larose S, Boulanger J, Elela SA 2001. The RNase III family: a conserved structure and expanding functions in eukaryotic dsRNA metabolism. Current issues in molecular biology 3: 71-78.
Lamontagne B, Tremblay A, Abou Elela S 2000. The N-terminal domain that distinguishes yeast from bacterial RNase III contains a dimerization signal required for efficient double-stranded RNA cleavage. Molecular and cellular biology 20: 1104-1115.
Lee S-J, Kong M, Harrison P, Hijri M 2018a. Conserved Proteins of the RNA Interference System in the Arbuscular Mycorrhizal Fungus Rhizoglomus irregulare Provide New Insight into the Evolutionary History of Glomeromycota. Genome Biology and Evolution 10: 328-343. doi: https://doi.org/10.1093/gbe/evy002
Lee S-J, Kong M, Morse D, Hijri M 2018b. Expression of putative circadian clock components in the arbuscular mycorrhizal fungus Rhizoglomus irregulare. Mycorrhiza.
Letunic I, Doerks T, Bork P 2015. SMART: recent updates, new developments and status in 2015. Nucleic acids research 43: D257-260.
Levy O, et al. 2011. Complex diel cycles of gene expression in coral-algal symbiosis. Science (New York, N Y) 331: 175.
Lewis MT, Feldman JF 1996. Evolution of the frequency (frq) clock locus in Ascomycete fungi. Molecular biology and evolution 13: 1233-1241.
Liang X, Bushman FD, FitzGerald GA 2015. Rhythmicity of the intestinal microbiota is regulated by gender and the host circadian clock. Proceedings of the National Academy of Sciences of the United States of America 112: 10479-10484.
Liang Y-H, Lavoie M, Comeau M-A, Abou Elela S, Ji X 2014. Structure of a eukaryotic RNase III postcleavage complex reveals a double-ruler mechanism for substrate selection. Molecular cell 54: 431-444.
Lin K, et al. 2014. Single nucleus genome sequencing reveals high similarity among nuclei of an endomycorrhizal fungus. PLoS genetics 10: e1004078.
Liu Y, Bell-Pedersen D 2006. Circadian rhythms in Neurospora crassa and other filamentous fungi. Eukaryotic cell 5: 1184-1193.
Lombardi LM, Brody S 2005. Circadian rhythms in Neurospora crassa: clock gene homologues in fungi. Fungal genetics and biology : FG & B 42: 887-892.

Loros JJ, Denome SA, Dunlap JC 1989. Molecular cloning of genes under control of the circadian clock in Neurospora. Science (New York, N Y) 243: 385-388.
Loros JJ, Dunlap JC 2001. Genetic and molecular analysis of circadian rhythms in Neurospora. Annual review of physiology 63: 757-794.
Loza-Correa M, Gomez-Valero L, Buchrieser C 2010. Circadian clock proteins in prokaryotes: hidden rhythms? Frontiers in microbiology 1: 130.
Luc Simon JB, Roger C. Levesque and Maurice Lalonde 1993. Origin and diversification of endomycorrhizal fungi and coincidence with vascular land plants. Nature Vol. 363: 67-69.
Lupas A, Van Dyke M, Stock J 1991. Predicting coiled coils from protein sequences. Science (New York, N Y) 252: 1162-1164.
Maeght J-L, Rewald B, Pierret A 2013. How to study deep roots-and why it matters. Frontiers in plant science 4: 299.
Margulis L, Fester R. 1991. Symbiosis as a Source of Evolutionary Innovation: MIT Press.
Marleau J, Dalpe Y, St-Arnaud M, Hijri M 2011. Spore development and nuclear inheritance in arbuscular mycorrhizal fungi. BMC Evol Biol 11: 51. doi: 10.1186/1471-2148-11-51
Martin WF 2017. Too Much Eukaryote LGT. BioEssays : news and reviews in molecular, cellular and developmental biology 39.
Mas P, Yanovsky MJ 2009. Time for circadian rhythms: plants get synchronized. Current opinion in plant biology 12: 574-579.
McClung CR 2006. Plant circadian rhythms. The Plant cell 18: 792-803.
McDonald MJ, Rosbash M 2001. Microarray analysis and organization of circadian gene expression in Drosophila. Cell 107: 567-578.
McDowall J, Hunter S 2011. InterPro protein classification. Methods in molecular biology (Clifton, N J) 694: 37-47.
Mendoza J, Graff C, Dardente H, Pevet P, Challet E 2005. Feeding cues alter clock gene oscillations and photic responses in the suprachiasmatic nuclei of mice exposed to a light/dark cycle. The Journal of neuroscience : the official journal of the Society for Neuroscience 25: 1514-1522.
Meng W, Nicholson AW 2008. Heterodimer-based analysis of subunit and domain contributions to double-stranded RNA processing by Escherichia coli RNase III in vitro. The Biochemical journal 410: 39-48.
Michael TP, Salome PA, McClung CR 2003. Two Arabidopsis circadian oscillators can be distinguished by differential temperature sensitivity. Proceedings of the National Academy of Sciences of the United States of America 100: 6878-6883.
Millar AJ, Kay SA 1997. The genetics of phototransduction and circadian rhythms in Arabidopsis. BioEssays : news and reviews in molecular, cellular and developmental biology 19: 209-214.
Mindell DP 1992. Phylogenetic consequences of symbioses: Eukarya and Eubacteria are not monophyletic taxa. Bio Systems 27: 53-62.
Mittag M 2001. Circadian rhythms in microalgae. International review of cytology 206: 213-247.
Mori T, Johnson CH 2000. Circadian control of cell division in unicellular organisms. Progress in cell cycle research 4: 185-192.
Mukherji A, Kobiita A, Ye T, Chambon P 2013. Homeostasis in intestinal epithelium is orchestrated by the circadian clock and microbiota cues transduced by TLRs. Cell 153: 812-827.

Nadimi M, Beaudet D, Forget L, Hijri M, Lang BF 2012. Group I intron-mediated trans-splicing in mitochondria of Gigaspora rosea and a robust phylogenetic affiliation of arbuscular mycorrhizal fungi with Mortierellales. Molecular biology and evolution 29: 2199-2210.
Nadimi M, Daubois L, Hijri M 2016. Mitochondrial comparative genomics and phylogenetic signal assessment of mtDNA among arbuscular mycorrhizal fungi. Molecular phylogenetics and evolution 98: 74-83.
Nagahashi G, Douds DD, Jr. 2003. Action spectrum for the induction of hyphal branches of an arbuscular mycorrhizal fungus: exposure sites versus branching sites. Mycological Research 107: 1075-1082.
Nagahashi G, Douds DD, Jr. 2004. Synergism between blue light and root exudate compounds and evidence for a second messenger in the hyphal branching response of Gigaspora gigantea. Mycologia 96: 948-954.
Nagel DH, Kay SA 2012. Complexity in the wiring and regulation of plant circadian networks. Current biology : CB 22: R648-657.
Naito M, Morton JB, Pawlowska TE 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: 7791-7796.
Nicholson AW 1999. Function, mechanism and regulation of bacterial ribonucleases. FEMS microbiology reviews 23: 371-390.
Nicolas FE, Ruiz-Vazquez RM 2013. Functional diversity of RNAi-associated sRNAs in fungi. International journal of molecular sciences 14: 15348-15360.
Nouri E, Breuillin-Sessoms F, Feller U, Reinhardt D 2014. Phosphorus and nitrogen regulate arbuscular mycorrhizal symbiosis in Petunia hybrida. PloS one 9: e90841.
Novo M, et al. 2009. Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast Saccharomyces cerevisiae EC1118. Proceedings of the National Academy of Sciences of the United States of America 106: 16333-16338.
Oehl F, Sieverding E, Palenzuela J, Ineichen K, Alves da Silva G 2011. Advances in Glomeromycota taxonomy and classification. IMA fungus 2: 191-199.
Olmedo M, Ruger-Herreros C, Corrochano LM 2010. Regulation by blue light of the fluffy gene encoding a major regulator of conidiation in Neurospora crassa. Genetics 184: 651-658.
Oosterhuis DM 1990. Growth and development of a cotton plant. Publications of the American Society of Agronomy, Madison. doi: MP332-4M-9-92R
Orchard S, et al. 2017. Fine endophytes (Glomus tenue) are related to Mucoromycotina, not Glomeromycota. The New phytologist 213: 481-486.
Ouyang Y, Andersson CR, Kondo T, Golden SS, Johnson CH 1998. Resonating circadian clocks enhance fitness in cyanobacteria. Proceedings of the National Academy of Sciences of the United States of America 95: 8660-8664.
Pöggeler S, Wöstemeyer J. 2011. Evolution of Fungi and Fungal-Like Organisms: Springer.
Pando MP, Morse D, Cermakian N, Sassone-Corsi P 2002. Phenotypic rescue of a peripheral clock genetic defect via SCN hierarchical dominance. Cell 110: 107-117.
Pickford AS, Catalanotto C, Cogoni C, Macino G 2002. Quelling in Neurospora crassa. Advances in genetics 46: 277-303.
Piggins HD 2002. Human clock genes. Annals of medicine 34: 394-400.
Pinto-Carbo M, et al. 2016. Evidence of horizontal gene transfer between obligate leaf nodule symbionts. The ISME journal 10: 2092-2105.

Purschwitz J, Muller S, Kastner C, Fischer R 2006. Seeing the rainbow: light sensing in fungi. Current opinion in microbiology 9: 566-571.
Querfurth C, et al. 2011. Circadian conformational change of the Neurospora clock protein FREQUENCY triggered by clustered hyperphosphorylation of a basic domain. Molecular cell 43: 713-722.
Redecker D, Kodner R, Graham LE 2000. Glomalean fungi from the Ordovician. Science (New York, N Y) 289: 1920-1921.
Redecker D, et al. 2013. An evidence-based consensus for the classification of arbuscular mycorrhizal fungi (Glomeromycota). Mycorrhiza 23: 515-531.
Riley R, et al. 2014. Extreme diversification of the mating type-high-mobility group (MATA-HMG) gene family in a plant-associated arbuscular mycorrhizal fungus. The New phytologist 201: 254-268.
Rodriguez-Romero J, Hedtke M, Kastner C, Muller S, Fischer R 2010. Fungi, hidden in soil or up in the air: light makes a difference. Annual review of microbiology 64: 585-610.
Roenneberg T, Merrow M 1999. Circadian clocks--from genes to complex behaviour. Reproduction, nutrition, development 39: 277-294.
Rosenberg E, Sharon G, Atad I, Zilber-Rosenberg I 2010. The evolution of animals and plants via symbiosis with microorganisms. Environmental microbiology reports 2: 500-506.
Salichos L, Rokas A 2010. The diversity and evolution of circadian clock proteins in fungi. Mycologia 102: 269-278.
Sancar C, et al. 2015. Combinatorial control of light induced chromatin remodeling and gene activation in Neurospora. PLoS genetics 11: e1005105.
Sancar G, Sancar C, Brunner M 2012. Metabolic compensation of the Neurospora clock by a glucose-dependent feedback of the circadian repressor CSP1 on the core oscillator. Genes & Dev 26: 2435-2442. doi: 10.1101/gad.199547.112
Schüßler A, Martin H, Cohen D, Fitz M, Wipf D 2007. Arbuscular Mycorrhiza. Plant Signaling & Behavior 2: 431-434. doi: 10.4161/psb.2.5.4465
Schüßler A, Schwarzott D, Walker C 2001. A new fungal phylum, the Glomeromycota: phylogeny and evolution. Mycological research 105: 1413-1421. doi: 10.1017/S0953756201005196
Schwarz JA, et al. 2008. Coral life history and symbiosis: functional genomic resources for two reef building Caribbean corals, Acropora palmata and Montastraea faveolata. BMC genomics 9: 97.
Schwerdtfeger C, Linden H 2000. Localization and light-dependent phosphorylation of white collar 1 and 2, the two central components of blue light signaling in Neurospora crassa. European journal of biochemistry / FEBS 267: 414-422.
Sender R, Fuchs S, Milo R 2016. Revised Estimates for the Number of Human and Bacteria Cells in the Body. PLoS biology 14: e1002533.
Serikawa M, Miwa K, Kondo T, Oyama T 2008. Functional conservation of clock-related genes in flowering plants: overexpression and RNA interference analyses of the circadian rhythm in the monocotyledon Lemna gibba. Plant physiology 146: 1952-1963.
Sharp PM, Li WH 1987. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. Nucleic acids research 15: 1281-1295.
Shinohara ML, Correa A, Bell-Pedersen D, Dunlap JC, Loros JJ 2002. Neurospora clock-controlled gene 9 (ccg-9) encodes trehalose synthase: circadian regulation of stress responses and development. Eukaryotic cell 1: 33-43.

Shukla A, Vyas D, Anuradha J 2013. Soil depth: an overriding factor for distribution of arbuscular mycorrhizal fungi. Journal of Soil Science and Plant Nutrition 13: 23-33. doi: 10.4067/S0718-95162013005000003
Simon L, Bousquet J, Levesque CR, Lalonde M 1993. Origin and diversification of endomycorrhizal fungi and coincidence with vascular land plants. Nature Vol. 363: 67-69.
Simon NML, Dodd AN 2017. A new link between plant metabolism and circadian rhythms? Plant, cell & environment 40: 995-996.
Smith A 1932. Seasonal subsoil temperature variations. Journal of Agricultural Research 44: 421-428.
Smith SE, Read DJ. 2008. Mycorrhizal Symbiosis (Third Edition): Academic Press, New York.
Soanes D, Richards TA 2014. Horizontal gene transfer in eukaryotic plant pathogens. Annual review of phytopathology 52: 583-614.
Sorek M, Diaz-Almeyda EM, Medina M, Levy O 2014. Circadian clocks in symbiotic corals: the duet between Symbiodinium algae and their coral host. Marine genomics 14: 47-57.
Sorek M, Yacobi YZ, Roopin M, Berman-Frank I, Levy O 2013. Photosynthetic circadian rhythmicity patterns of Symbiodinium, [corrected] the coral endosymbiotic algae. Proceedings Biological sciences 280: 20122942.
Soucy SM, Huang J, Gogarten JP 2015. Horizontal gene transfer: building the web of life. Nature reviews Genetics 16: 472-482.
Spatafora JW, et al. 2016. A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. Mycologia 108: 1028-1046.
Stone EL, Kalisz PJ 1991. On the maximum extent of tree roots. Forest Ecology and Management 46: 59-102. doi: https://doi.org/10.1016/0378-1127(91)90245-Q
Storch K-F, et al. 2002. Extensive and divergent circadian gene expression in liver and heart. Nature 417: 78-83.
Stukenbrock EH, Rosendahl S 2005. Clonal diversity and population genetic structure of arbuscular mycorrhizal fungi (Glomus spp.) studied by multilocus genotyping of single spores. Molecular ecology 14: 743-752.
Subramanian P, Balamurugan E, Suthakar G 2003. Circadian clock genes in Drosophila: recent developments. Indian journal of experimental biology 41: 797-804.
Sumner-Rooney L 2018. The kingdom of the blind: disentangling fundamental drivers in the evolution of eye loss. Integrative and comparative biology.
Sun W, Nicholson AW 2001. Mechanism of action of Escherichia coli ribonuclease III. Stringent chemical requirement for the glutamic acid 117 side chain and Mn2+ rescue of the Glu117Asp mutant. Biochemistry 40: 5102-5110.
Takemiya A, Inoue S-I, Doi M, Kinoshita T, Shimazaki K-I 2005. Phototropins promote plant growth in response to blue light in low light environments. The Plant cell 17: 1120-1127.
Tester M, Morris C 1987. The penentration of light through soil. Plant, cell & environment 10: 281-286. doi: 10.1111/j.1365-3040.1987.tb01607.x
Thaiss CA, et al. 2014. Transkingdom control of microbiota diurnal oscillations promotes metabolic homeostasis. Cell 159: 514-529.
Thiebaut F, et al. 2014. Genome-wide identification of microRNA and siRNA responsive to endophytic beneficial diazotrophic bacteria in maize. BMC genomics 15: 766.
Thirkell TJ, Cameron DD, Hodge A 2016. Resolving the 'nitrogen paradox' of arbuscular mycorrhizas: fertilization with organic matter brings considerable benefits for plant nutrition and growth. Plant, cell & environment 39: 1683-1690.

Tisserant E, et al. 2012. The transcriptome of the arbuscular mycorrhizal fungus Glomus intraradices (DAOM 197198) reveals functional tradeoffs in an obligate symbiont. The New phytologist 193: 755-769.
Tisserant E, et al. 2013. Genome of an arbuscular mycorrhizal fungus provides insight into the oldest plant symbiosis. Proceedings of the National Academy of Sciences of the United States of America 110: 20117-20122.
Torres-Cortes G, Ghignone S, Bonfante P, SchuSsler 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 112: 7785-7790.
Torres-Martinez S, Ruiz-Vazquez RM 2016. RNAi pathways in Mucor: A tale of proteins, small RNAs and functional diversity. Fungal genetics and biology : FG & B 90: 44-52.
Tseng Y-Y, Hunt SM, Heintzen C, Crosthwaite SK, Schwartz J-M 2012. Comprehensive modelling of the Neurospora circadian clock and its temperature compensation. PLoS computational biology 8: e1002437.
Uebelmesser E 1954. The endogenous daily rhythm of conidiospore formation of Pilobolus. Arch Mikrobiol 20: 1-33.
Vancurova L, Muggia L, Peksa O, Ridka T, Skaloud P 2018. The complexity of symbiotic interactions influences the ecological amplitude of the host: a case study in Stereocaulon (lichenized Ascomycota). Molecular ecology.
Walker C, Gollotte A, Redecker D 2018. A new genus, Planticonsortium (Mucoromycotina), and new combination (P. tenue), for the fine root endophyte, Glomus tenue (basionym Rhizophagus tenuis). Mycorrhiza 28: 213-219.
Walter J, Martinez I, Rose DJ 2013. Holobiont nutrition: considering the role of the gastrointestinal microbiota in the health benefits of whole grains. Gut microbes 4: 340-346.
Weiberg A, Bellinger M, Jin H 2015. Conversations between kingdoms: small RNAs. Current opinion in biotechnology 32: 207-215.
White MJ, Hirsch JP, Henry SA 1991. The OPI1 gene of Saccharomyces cerevisiae, a negative regulator of phospholipid biosynthesis, encodes a protein containing polyglutamine tracts and a leucine zipper. The Journal of biological chemistry 266: 863-872.
Wier AM, et al. 2010. Transcriptional patterns in both host and bacterium underlie a daily rhythm of anatomical and metabolic change in a beneficial symbiosis. Proceedings of the National Academy of Sciences of the United States of America 107: 2259-2264.
Xu X, et al. 2007. Distinct light and clock modulation of cytosolic free Ca2+ oscillations and rhythmic CHLOROPHYLL A/B BINDING PROTEIN2 promoter activity in Arabidopsis. The Plant cell 19: 3474-3490.
Yan Z, et al. 2015. Identification of microRNAs and their mRNA targets during soybean nodule development: functional analysis of the role of miR393j-3p in soybean nodulation. The New phytologist 207: 748-759.
Yang T, Xiong W, Dong C 2014. Cloning and analysis of the Oswc-1 gene encoding a putative blue light photoreceptor from Ophiocordyceps sinensis. Mycoscience 55: 241-245. doi: 10.1016/j.myc.2013.09.003
Yazdanbakhsh N, Sulpice R, Graf A, Stitt M, Fisahn J 2011. Circadian control of root elongation and C partitioning in Arabidopsis thaliana. Plant, cell & environment 34: 877-894.
Yigit E, et al. 2006. Analysis of the C. elegans Argonaute family reveals that distinct Argonautes act sequentially during RNAi. Cell 127: 747-757.

Zhu Z, et al. 2012. A multi-omic map of the lipid-producing yeast Rhodosporidium toruloides. Nature communications 3: 1112.
Zisapel N 2018. New perspectives on the role of melatonin in human sleep, circadian rhythms and their regulation. British journal of pharmacology.

Annex 1 : Supplementary Information (Chapter 2)

Table S2.1. Query sequences for BLAST.

	N. crassa	*S. pombe*	*M. circinelloides*
DICER	XP_961898.1	CAB41233.2	OAD03695.1
	XP_963538.3		CAZ65730.1
ARGONAUTE/PIWI	XP_011394904.1	CAA19275.1	EPB81851.1
	XP_011394903.1		EPB81974.1
			EPB92043.1
RDRP	XP_963405.1	CAB11093.1	OAD02682.1
	XP_959047.1		OAD02683.1
	XP_964248.3		OAD09244.1

Table S2.2. Summarized result of homologs of core proteins of RNAi system in *R. irregulare*. Genbank accession codes for *R. irregulare* sequences used in this study can be found with each homolog. Related information such as each CDS and their related genome region could be found via these accession codes. All of the sequences were successfully tested via reciprocal BLAST as well as protein domain architecture analysis. Surprisingly, there were two homologs of ribonuclease III (RIRNC2 and RIRNC3, shaded in yellow) whose best Genbank BLAST hits were cyanobacterial ribonuclease III, while all other homologs (three homologs of RDRP and 26 homologs of AGO) have fungal orthologs as their best Genbank BLAST hits.

Name of component (Genbank Accession code)	Comparison with top reciprocal hit (N. crassa) E-value / (%identity)	Comparison with top reciprocal hit (S. pombe) E-value / (%identity)	Comparison with top reciprocal hit (M.circinelloides) E-value / (%identity)	Best Genbank BLAST hit E-value / (%identity)	Best BLAST hit information [organism name]
RDRP1 (EXX60053.1)	XP_959047.1 3e-51/(31%)	CAB11093.1 2e-33/(25%)	OAD02682.1 1e-116/(31%)	XP_018290421.1 7e-115/(31%)	RNA-dependent RNA polymerase RdRP [*Phycomyces blakesleeanus*]
RDRP2 (EXX62399.1)	XP_959047.1 3e-42/(27%)	CAB11093.1 1e-30/(25%)	OAD02682.1 2e-115/(28%)	CDH49104.1 2e-132/(33%)	RNA-directed RNA polymerase [*Lichtheimia corymbifera*]
RDRP3 (EXX74509.1)	XP_959047.1 6e-50/(26%)	CAB11093.1 1e-29/(27%)	OAD02682.1 1e-119/(31%)	XP_018293829.1 2e-129/(33%)	RNA-dependent RNA polymerase RdRP [*Phycomyces blakesleeanus*]
AGO1 (EXX56889.1)	XP_958586.1 8e-125/(31%)	CAA19275.1 0.0/(37%)	EPB81851.1 0.0/(62%)	OBZ91340.1 0.0/(64%)	Protein argonaute 1 [*Choanephora cucurbitarum*]
AGO2 (EXX56891.1)	XP_958586.1 1e-124/(31%)	CAA19275.1 0.0/(37%)	EPB81851.1 0.0/(62%)	OBZ91340.1 0.0/(63%)	Protein argonaute 1 [*Choanephora cucurbitarum*]
AGO3 (EXX56890.1)	XP_958586.1 1e-116/(31%)	CAA19275.1 5e-177/(37%)	EPB81851.1 0.0/(63%)	OBZ91340.1 0.0/(63%)	Protein argonaute 1 [*Choanephora cucurbitarum*]
AGO4 (EXX59523.1)	XP_958586.1 2e-119/(31%)	CAA19275.1 5e-178/(35%)	EPB81851.1 0.0/(58%)	OBZ91340.1 0.0/(60%)	Protein argonaute 1 [*Choanephora cucurbitarum*]
AGO5 (EXX72566.1)	XP_958586.1 2e-105/(29%)	CAA19275.1 4e-179/(38%)	EPB81851.1 0.0/(58%)	OBZ91340.1 0.0/(60%)	Protein argonaute 1 [*Choanephora cucurbitarum*]

AGO6 (EXX58072.1)	XP_958586.1 5e-110/(29%)	CAA19275.1 0.0/(37%)	EPB81851.1 0.0/(63%)	OBZ91340.1 0.0/(64%)	Protein argonaute 1 [*Choanephora cucurbitarum*]
AGO7 (EXX59521.1)	XP_958586.1 5e-106/(29%)	CAA19275.1 0.0/(38%)	EPB81851.1 0.0/(56%)	OBZ91340.1 0.0/(57%)	Protein argonaute 1 [*Choanephora cucurbitarum*]
AGO8 (EXX62045.1)	XP_958586.1 2e-105/(28%)	CAA19275.1 0.0/(38%)	EPB81851.1 0.0/(60%)	OBZ91340.1 0.0/(60%)	Protein argonaute 1 [*Choanephora cucurbitarum*]
AGO9 (EXX60142.1)	XP_958586.1 8e-124/(31%)	CAA19275.1 0.0/(39%)	EPB81851.1 0.0/(62%)	OBZ91340.1 0.0/(62%)	Protein argonaute 1 [*Choanephora cucurbitarum*]
AGO10 (EXX71200.1)	XP_958586.1 6e-112/(30%)	CAA19275.1 0.0/(38%)	EPB81851.1 0.0/(61%)	OBZ91340.1 0.0/(62%)	Protein argonaute 1 [*Choanephora cucurbitarum*]
AGO11 (EXX68513.1)	XP_958586.1 1e-108/(30%)	CAA19275.1 7e-166/(34%)	EPB81851.1 0.0/(55%)	OBZ91340.1 0.0/(56%)	Protein argonaute 1 [*Choanephora cucurbitarum*]
AGO12 (EXX53458.1)	XP_958586.1 4e-111/(30%)	CAA19275.1 7e-176/(36%)	EPB81851.1 0.0/(53%)	OBZ91340.1 0.0/(54%)	Protein argonaute 1 [*Choanephora cucurbitarum*]
AGO13 (EXX59522.1)	XP_958586.1 1e-115/(31%)	CAA19275.1 2e-174/(36%)	EPB81851.1 0.0/(54%)	OBZ91340.1 0.0/(55%)	Protein argonaute 1 [*Choanephora cucurbitarum*]
AGO14 (EXX76428.1)	XP_958586.1 2e-112/(29%)	CAA19275.1 0.0/(37%)	EPB81851.1 0.0/(58%)	OBZ91340.1 0.0/(59%)	Protein argonaute 1 [*Choanephora cucurbitarum*]
AGO15 (EXX73036.1)	XP_958586.1 7e-108/(28%)	CAA19275.1 0.0/(37%)	EPB81851.1 0.0/(57%)	OBZ91340.1 0.0/(57%)	Protein argonaute 1 [*Choanephora cucurbitarum*]
AGO16 (EXX54101.1)	XP_958586.1 7e-96/(28%)	CAA19275.1 2e-154/(33%)	EPB81851.1 0.0/(51%)	OBZ91340.1 0.0/(52%)	Protein argonaute 1 [*Choanephora cucurbitarum*]
AGO17 (EXX60325.1)	XP_958586.1 3e-94/(28%)	CAA19275.1 2e-162/(34%)	EPB81851.1 0.0/(52%)	OBZ91340.1 0.0/(53%)	Protein argonaute 1 [*Choanephora cucurbitarum*]
AGO18 (EXX75353.1)	XP_958586.1 7e-91/(27%)	CAA19275.1 2e-152/(34%)	EPB81851.1 0.0/(52%)	OBZ91340.1 0.0/(53%)	Protein argonaute 1 [*Choanephora cucurbitarum*]
AGO19 (EXX76429.1)	XP_958586.1 8e-98/(28%)	CAA19275.1 6e-176/(37%)	EPB81851.1 0.0/(54%)	OBZ91340.1 0.0/(55%)	Protein argonaute 1 [*Choanephora cucurbitarum*]
AGO20 (EXX68714.1)	XP_958586.1 5e-81/(27%)	CAA19275.1 4e-124/(32%)	EPB81851.1 0.0/(52%)	OBZ91340.1 0.0/(53%)	Protein argonaute 1 [*Choanephora cucurbitarum*]
AGO21 (EXX62871.1)	XP_958586.1 2e-63/(25%)	CAA19275.1 1e-108/(30%)	EPB81851.1 0.0/(51%)	OBZ91340.1 0.0/(52%)	Protein argonaute 1 [*Choanephora cucurbitarum*]
AGO22 (EXX71201.1)	XP_958586.1 5e-71/(27%)	CAA19275.1 5e-135/(35%)	EPB81851.1 0.0/(59%)	OBZ91340.1 0.0/(60%)	Protein argonaute 1 [*Choanephora cucurbitarum*]

AGO23 (EXX52393.1)	XP_958586.1 7e-80/(26%)	CAA19275.1 4e-129/(31%)	EPB81851.1 0.0/(39%)	OBZ91340.1 0.0/(40%)	Protein argonaute 1 [*Choanephora cucurbitarum*]
AGO24 (EXX67924.1)	XP_958586.1 5e-86/(27%)	CAA19275.1 7e-125/(32%)	EPB81851.1 0.0/(37%)	OBZ91340.1 0.0/(37%)	Protein argonaute 1 [*Choanephora cucurbitarum*]
AGO25 (EXX57432.1)	XP_958586.1 1e-80/(28%)	CAA19275.1 6e-113/(30%)	EPB81851.1 0.0/(39%)	OBZ91340.1 0.0/(39%)	Protein argonaute 1 [*Choanephora cucurbitarum*]
AGO26 (EXX69221.1)	XP_958586.1 7e-63/(27%)	CAA19275.1 9e-118/(30%)	EPB81851.1 7e-168/(36%)	OBZ91340.1 7e-164/(36%)	Protein argonaute 1 [*Choanephora cucurbitarum*]
RIDCL1 (EXX59830.1)	XP_961898.1 0.0/(34%)	CAB41233.2 9e-154/(29%)	OAD03695.1 0.0/(38%)	XP_018289885.1 0.0/(44%)	dsRNA-specific ribonuclease III dicer [*Phycomyces blakesleeanus*]
RIRNC2 (EXX75321.1)	XP_961898.1 1e-07/(24%)	AAU05314.1 1e-20/(27%)	OAD03695.1 8e-19/(35%)	WP_039725904.1 2e-46/(39%)	ribonuclease III [*Lyngbya confervoides*]
RIRNC3 (EXX54746.1)	XP_961898.1 4e-08/(26%)	AAU05314.1 1e-21/(31%)	OAD03695.1 3e-16/(33%)	WP_096577203.1 4e-49/(43%)	ribonuclease III [*Anabaenopsis circularis*]

Table S2.3. Primers and probes used in PCR and TaqMan qPCR assays.

Name	Target gene	Sequence (5'->3')	Length (bp)	modification	
Forward	*ridcl1* (fungal)	5'-CCTGAGGATCGTAAATTACAAATG-3'	24		
Reverse	*ridcl1* (fungal)	5'-TAACTCAATAATGCACCTGTTTCA-3'	24		
Probe	*ridcl1* (fungal)	5'-CGGATGATTCGGATGATGAAGAAGGACA-3'	28	5'-6-FAM	3'-BHQ-1
Forward	*rirnc2* (bacterial)	5'-CAACTTATTAAGGAAAGTCTTGCC-3'	24		
Reverse	*rirnc2* (bacterial)	5'-GGATCTGATTCAGGAGGTAAATC-3'	23		
Probe	*rirnc2* (bacterial)	5'-ACTCCCATTGATCCTAGCCGTCCTAC-3'	26	5'-6-FAM	3'-BHQ-1
Forward	*18s rrna*	5'-GCTGAAACTTAAAGGAATTGACG-3'	23		

Reverse	*18s rrna*	5'-TGTCAATCCTTACTATGTCTGGA-3'	23		
Probe	*18s rrna*	5'-TGCGGCTTAATTTGACTCAACACGG-3'	25	5'-6-FAM	3'-BHQ-1
Forward	*ubiquitin*	5'-AGACCAAGTTAAAGCAAAGATTC-3'	23		
Reverse	*ubiquitin*	5'-TATAGTCTGACAATGTACGACCA-3'	23		
Probe	*ubiquitin*	5'-ATCAACAACGCTTGATCTTCGCTGGT-3'	26	5'-6-FAM	3'-BHQ-1

Table S2.4. Comparison of eukaryotic and prokaryotic RIBOc domains by BLAST. Two RIBOc domains (Domain A and B) from eukaryotic ribonuclease III homolog (RIDCL 1) and two RIBOc domains from two different prokaryotic ribonuclease III homologs (RIRNC 2 and 3) of *R. irregulare* were analyzed with BLAST with maximum number of target sequences 5,000.

(a) Eukaryotic RIBOc domain (two domains from RIDCL 1)

A. A domain

Taxonomy	Number of hits	Number of Organisms
Total	1304	566
Eukaryota	1304	566
1. Opisthokonta	1018	489
1.1. Fungi	1014	485
1.2. Metazoa	4	4
2. Embryophyta	286	77

B. B domain

Taxonomy	Number of hits	Number of Organisms
Total	1368	509
Eukaryota	1368	509
1. Opisthokonta	620	399
1.1. Fungi	379	273
1.2. Eumetazoa	241	126
2. Streptophyta	748	110

(b) Prokaryotic RIBOc domain

A. Domain from RIRNC 2

Taxonomy	Number of hits	Number of Organisms
Total	9779	5598
Bacteria	9778	5597

1. Terrabacteria group	8613	4809
1.1. Cyanobacteria	295	227
1.2. Firmicutes	8244	4536
1.3. Chloroflexi	31	20
1.4. Actinobacteria	41	25
1.5. Acholeplasma brassicae	2	1
2. unclassified Bacteria	296	291
3. uncultured bacterium	4	1
4. Proteobacteria	840	477
5. Spirochaetes	7	4
6. unclassified Synergistales	2	2
7. Nitrospirae	7	5
8. Bacteroidetes/Chlorobi group	6	5
9. Rubritalea squalenifaciens DSM 18772	1	1
10. Calditerrivibrio	2	2
uncultured prokaryote	1	1

B. Domain from RIRNC 3

Taxonomy	Number of hits	Number of Organisms
Total	9515	5471
Bacteria	9514	5470
1. Terrabacteria group	8782	4938
1.1. Cyanobacteria	348	263
1.2. Chloroflexi	50	32
1.3. Firmicutes	1	4594
1.4. Actinobacteria	93	49
2. unclassified Bacteria	200	198
3. uncultured bacterium	3	1
4. Proteobacteria	515	322

5. Rhodothermus	4	3
6. PVC group	6	4
7. Chloracidobacterium	2	2
8. Thermovirga	2	2
uncultured prokaryote	1	1

Table S2.5. Top 100 and 5000 hits of BLAST of RIRNC 2 and 3. Two different prokaryotic ribonuclease III homologs (RIRNC 2 and 3) of R. irregulare were analyzed with BLAST with maximum number of target sequences 100 and 5,000, respectively.

(a) RIRNC 2 (Max target sequences: 100)

Taxonomy	Number of hits	Number of Organisms
Bacteria	194	149
1. Terrabacteria group	158	113
1.1. Cyanobacteria	113	92
1.2. Bacilli	42	18
1.3. Chloroflexi	3	3
2. unclassified Bacteria	32	32
3. unclassified Nitrospirae	3	3
4. uncultured bacterium	1	1

(b) RIRNC 2 (Max target sequences: 5,000)

Taxonomy	Number of hits	Number of Organisms
Total	3315	1454
Bacteria	3314	1453
1. Terrabacteria group	2894	1108
1.1. Cyanobacteria/Melainabacteria group	323	247
1.2. Firmicutes	2521	822
1.3. Chloroflexi	41	32
1.4. Actinobacteria	7	6
1.5. Acholeplasma brassicae	2	1
2. unclassified Bacteria	163	162
3. Nitrospirae	31	26
4. uncultured bacterium	5	1
5. Proteobacteria	193	129

6. Calditerrivibrio	2	2
7. Bacteroidetes/Chlorobi group	11	10
8. unclassified Synergistales	2	2
9. unclassified Spirochaetes	7	7
10. unclassified Nitrospinae	2	2
11. Caviibacter abscessus	1	1
12. unclassified Lentisphaerae (miscellaneous)	2	2
13. Elusimicrobia bacterium	1	1
uncultured prokaryote	1	1

(c) RIRNC 3 (Max target sequences: 100)

Taxonomy	**Number of hits**	**Number of Organisms**
Bacteria	156	125
1. Terrabacteria group	143	113
1.1. Cyanobacteria	131	104
1.2. Chloroflexi	7	6
1.3. Firmicutes	5	3
2. Patescibacteria group	11	11
3. uncultured bacterium	2	1

(d) RIRNC 3 (Max target sequences: 5,000)

Taxonomy	**Number of hits**	**Number of Organisms**
Total	3597	1578
Bacteria	3596	1577
1. Terrabacteria group	3062	1107

1.1. Cyanobacteria/Melainabacteria group	367	267
1.2. Chloroflexi	60	44
1.3. Firmicutes	2607	776
1.4. Actinobacteria	28	20
2. unclassified Bacteria	297	294
3. uncultured bacterium	5	1
4. Nitrospirae	19	16
5. Elusimicrobia	13	12
6. Proteobacteria	179	127
7. Bacteroidetes/Chlorobi group	3	3
8. Synergistales	4	4
9. PVC group	8	8
10. Calditerrivibrio	2	2
11. Brachyspira	4	3
uncultured prokaryote	1	1

Table S2 6. Codon Adaptation Index (CAI) score calculated. A set of 200 Coding Sequences (CDSs) from model AMF species, R. irregulare, used as reference set. Top BLAST hits of eukaryotic and prokaryotic ribonuclease III enzymes (RIDCL1, RIRNC 2 and RIRNC 3), were compared with CDSs from *R. irregulare.* Calculations were conducted by two different method ((a) and (b)) (Eyre-Walker 1996; Sharp and Li 1987). All of the CDSs examined showed high CAI scores (*>0.5*), indicating their higher proportion of the most abundant codons in *R. irregulare*.

(a) Sharp & Li (1987)

Gene No.	**Description**	**Length (bp)**	**CAI**

1	*Rhizoglomus_irregulare_ridcl1*_CDS	4293	0.503
2	*Phycomyces_blakesleeanus*_CDS_closest_to_RIDCL1	4707	0.502
3	*Rhizoglomus_irregulare_rirnc2*_CDS	795	0.503
4	*Planktothricoides_sp.SR001*_CDS_closest_to_RIRNC2	1221	0.500
5	*Rhizoglomus_irregulare_rirnc3*_CDS	876	0.502
6	*Anabaenopsis_circularis_NIES21*_CDS_closest_to_RIRNC3	1185	0.503

(b) Eyre-Walker (1996)

Gene No.	**Description**	**Length (bp)**	**CAI**
1	*Rhizoglomus_irregulare_ridcl1*_CDS	4293	0.507
2	*Phycomyces_blakesleeanus*_CDS_closest_to_RIDCL1	4707	0.505
3	*Rhizoglomus_irregulare_rirnc2*_CDS	795	0.505
4	*Planktothricoides_sp.SR001*_CDS_closest_to_RIRNC2	1221	0.507
5	*Rhizoglomus_irregulare_rirnc3*_CDS	876	0.503
6	*Anabaenopsis_circularis_NIES21*_CDS_closest_to_RIRNC3	1185	0.505

Figure S2.1

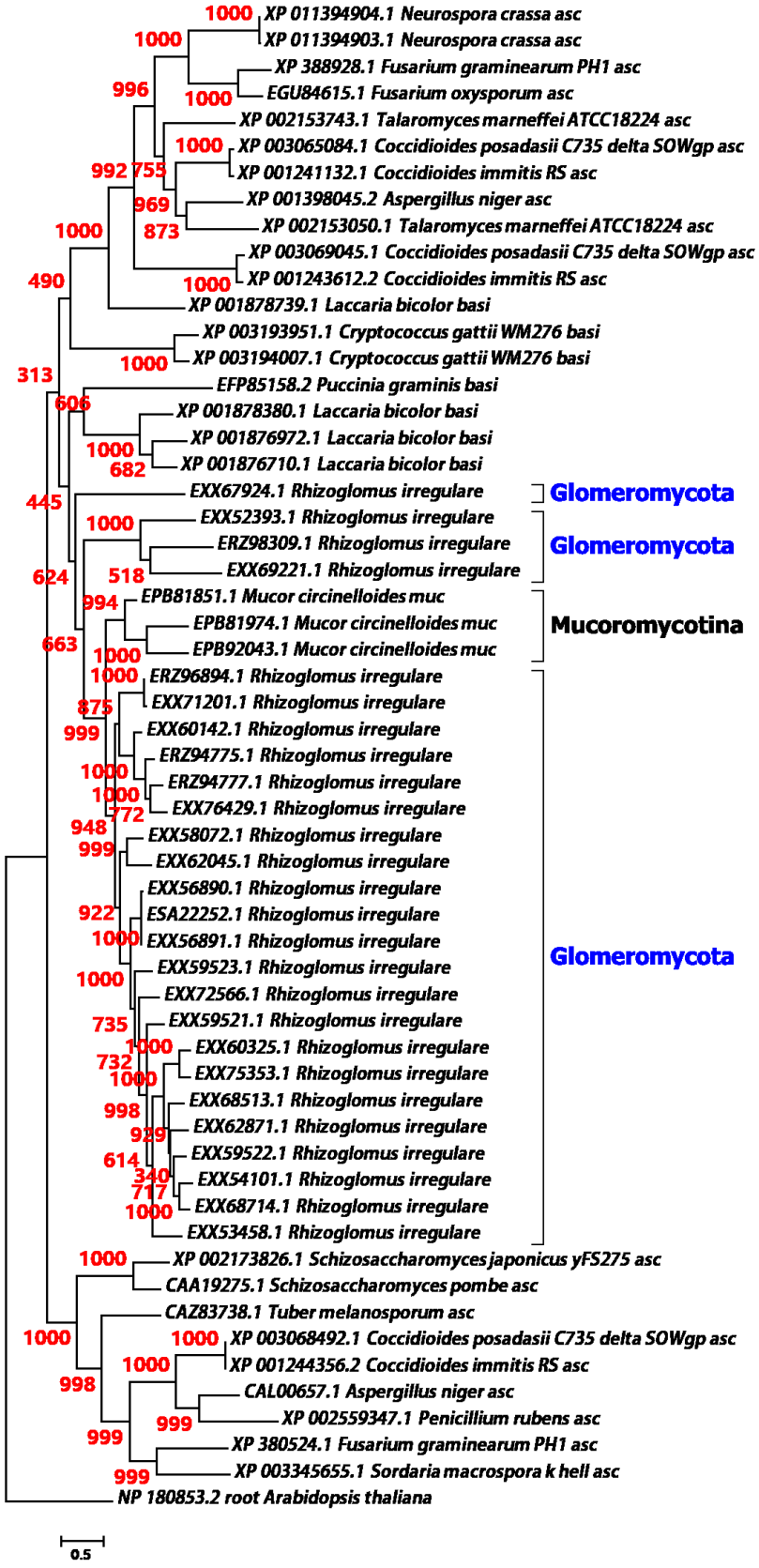
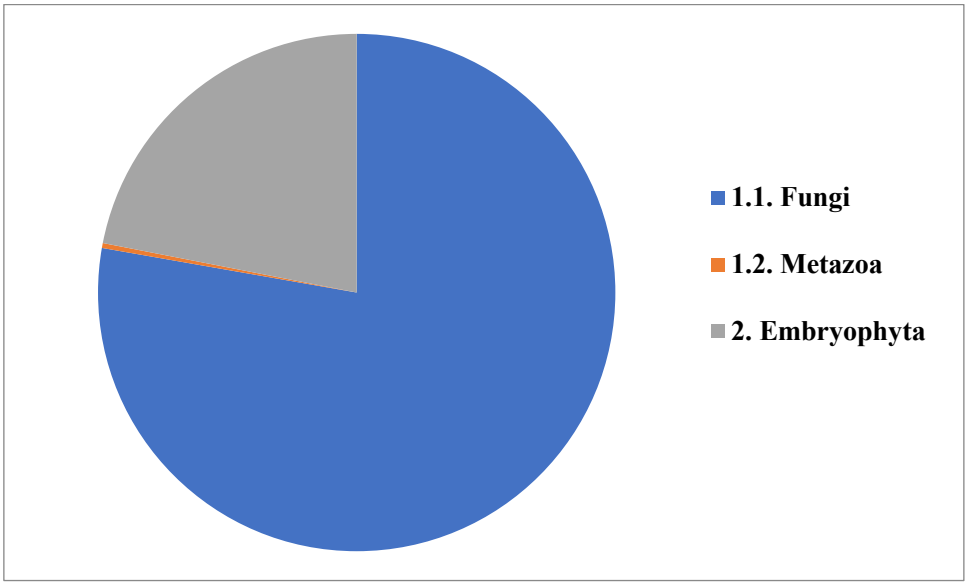

XP 011394904.1 Neurospora crassa asc
XP 011394903.1 Neurospora crassa asc
XP 388928.1 Fusarium graminearum PH1 asc
EGU84615.1 Fusarium oxysporum asc
XP 002153743.1 Talaromyces marneffei ATCC18224 asc
XP 003065084.1 Coccidioides posadasii C735 delta SOWgp asc
XP 001241132.1 Coccidioides immitis RS asc
XP 001398045.2 Aspergillus niger asc
XP 002153050.1 Talaromyces marneffei ATCC18224 asc
XP 003069045.1 Coccidioides posadasii C735 delta SOWgp asc
XP 001243612.2 Coccidioides immitis RS asc
XP 001878739.1 Laccaria bicolor basi
XP 003193951.1 Cryptococcus gattii WM276 basi
XP 003194007.1 Cryptococcus gattii WM276 basi
EFP85158.2 Puccinia graminis basi
XP 001878380.1 Laccaria bicolor basi
XP 001876972.1 Laccaria bicolor basi
XP 001876710.1 Laccaria bicolor basi
EXX67924.1 Rhizoglomus irregulare
EXX52393.1 Rhizoglomus irregulare
ERZ98309.1 Rhizoglomus irregulare
EXX69221.1 Rhizoglomus irregulare
EPB81851.1 Mucor circinelloides muc
EPB81974.1 Mucor circinelloides muc
EPB92043.1 Mucor circinelloides muc
ERZ96894.1 Rhizoglomus irregulare
EXX71201.1 Rhizoglomus irregulare
EXX60142.1 Rhizoglomus irregulare
ERZ94775.1 Rhizoglomus irregulare
ERZ94777.1 Rhizoglomus irregulare
EXX76429.1 Rhizoglomus irregulare
EXX58072.1 Rhizoglomus irregulare
EXX62045.1 Rhizoglomus irregulare
EXX56890.1 Rhizoglomus irregulare
ESA22252.1 Rhizoglomus irregulare
EXX56891.1 Rhizoglomus irregulare
EXX59523.1 Rhizoglomus irregulare
EXX72566.1 Rhizoglomus irregulare
EXX59521.1 Rhizoglomus irregulare
EXX60325.1 Rhizoglomus irregulare
EXX75353.1 Rhizoglomus irregulare
EXX68513.1 Rhizoglomus irregulare
EXX62871.1 Rhizoglomus irregulare
EXX59522.1 Rhizoglomus irregulare
EXX54101.1 Rhizoglomus irregulare
EXX68714.1 Rhizoglomus irregulare
EXX53458.1 Rhizoglomus irregulare
XP 002173826.1 Schizosaccharomyces japonicus yFS275 asc
CAA19275.1 Schizosaccharomyces pombe asc
CAZ83738.1 Tuber melanosporum asc
XP 003068492.1 Coccidioides posadasii C735 delta SOWgp asc
XP 001244356.2 Coccidioides immitis RS asc
CAL00657.1 Aspergillus niger asc
XP 002559347.1 Penicillium rubens asc
XP 380524.1 Fusarium graminearum PH1 asc
XP 003345655.1 Sordaria macrospora k hell asc
NP 180853.2 root Arabidopsis thaliana
Glomeromycota
Glomeromycota
Mucoromycotina
Glomeromycota
0.5

Figure S2.1. Phylogeny of three core proteins of RNAi system in AMF with other fungal species. Maximum likelihood of amino acid sequences of AGO were analyzed with the WAG+I+G (with four distinct gamma categories) phylogenetic model which showed the lowest AIC value. The *R. irregulare* (synonym of *Rhizophags irregularis*) and *Mucoromycotina* sequences were annotated in right side of taxon names, respectively. The numbers at branches correspond to bootstrap support values generated with 1,000 bootstrap replicates. Tree was rooted using *Drosophila melanogaster*.

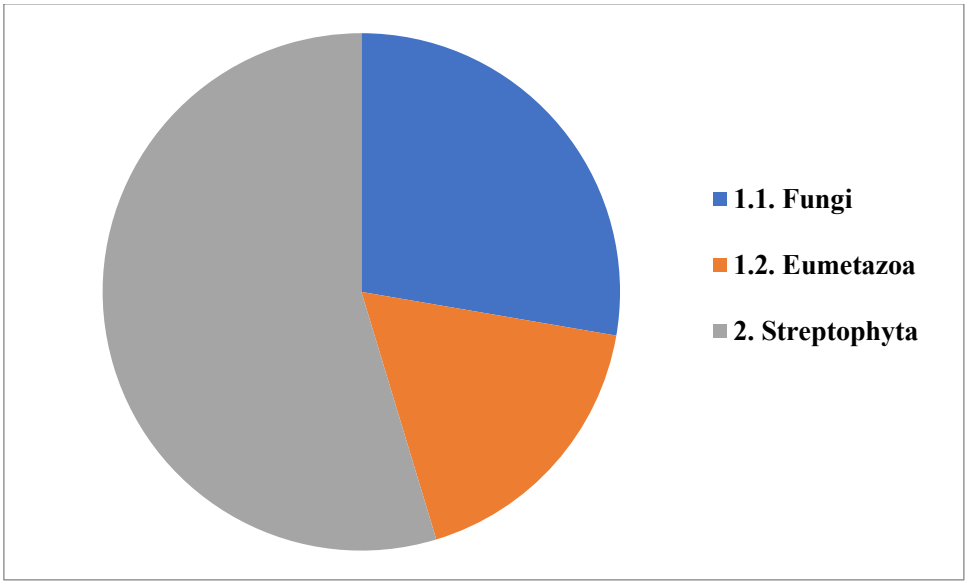
Figure S2.2

(a) Eukaryotic RIBOc domain (two domains from RIDCL 1)

A. A domain

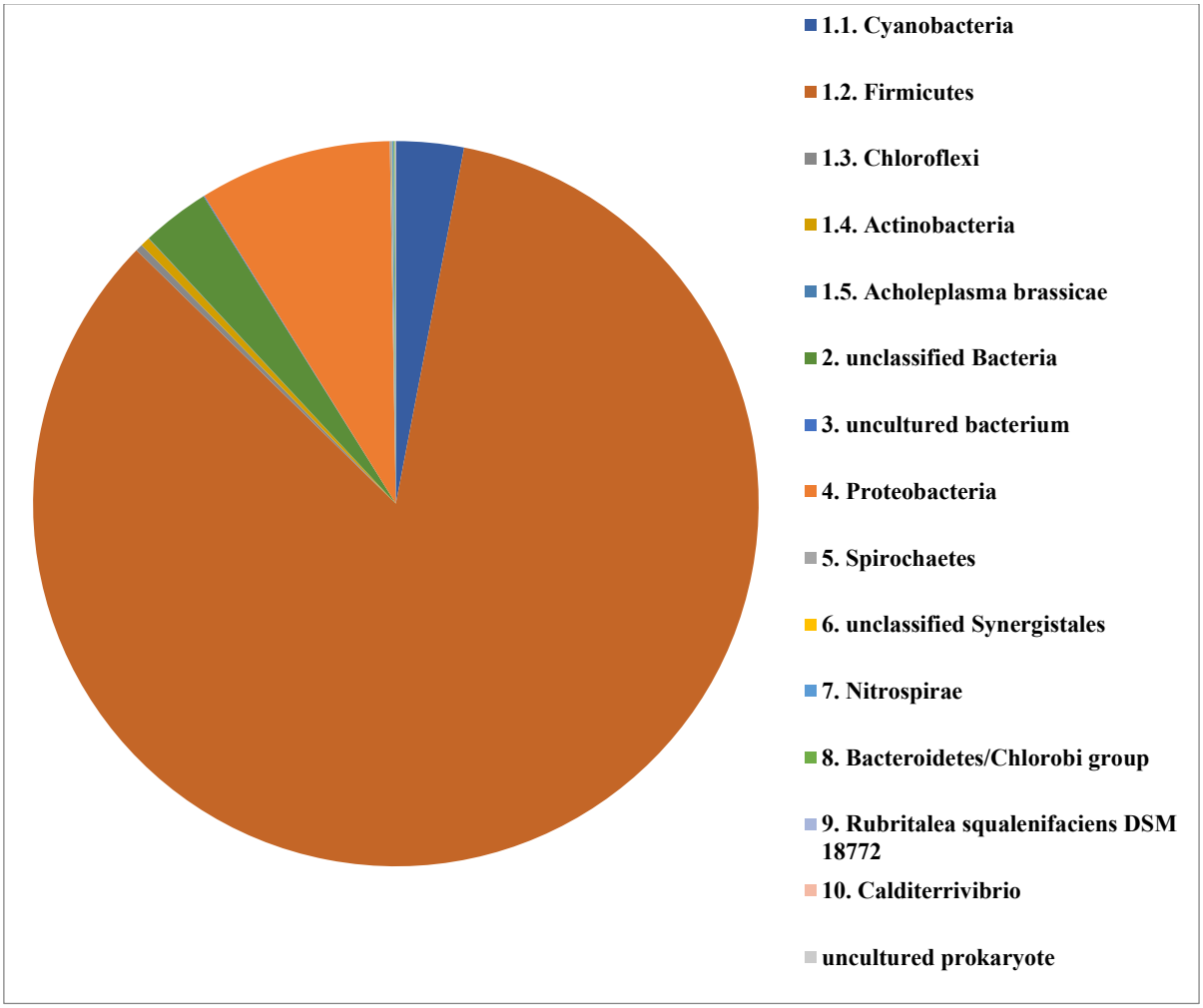


B. B domain



(b) Prokaryotic RIBOc domains (domains from RIRNC 2 and 3)

A. Domain from RIRNC 2



B. Domain from RIRNC 3

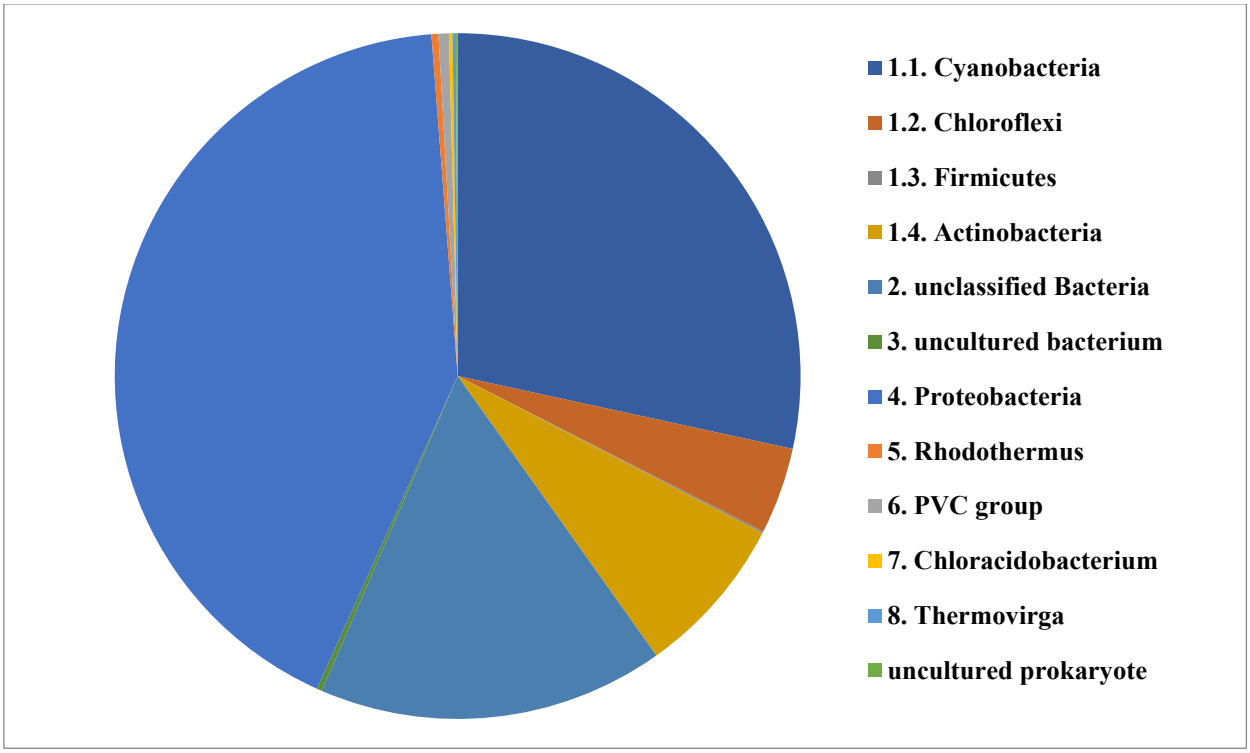
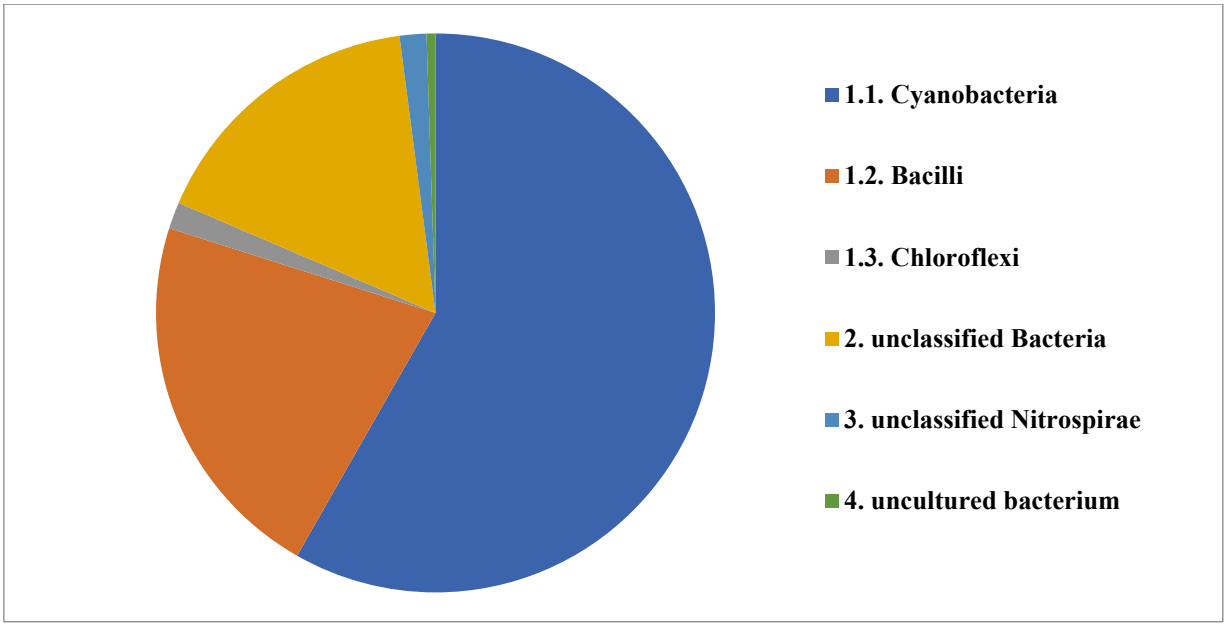


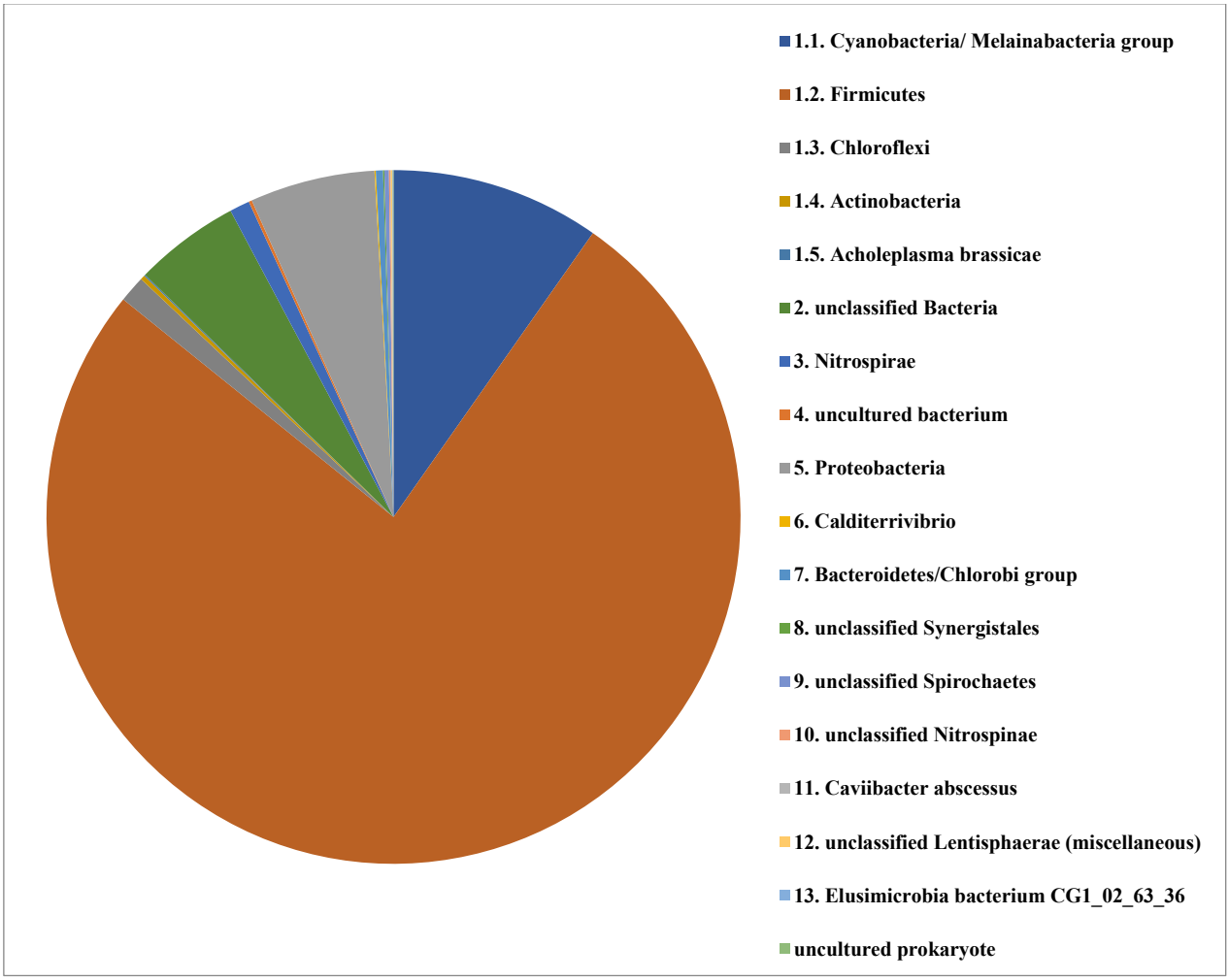
Figure S2.2. Summary of BLAST results of eukaryotic and prokaryotic RIBOc domains. Two RIBOc domains (Domain A and B) from eukaryotic ribonuclease III homolog (RIDCL 1) and two RIBOc domains from two different prokaryotic ribonuclease III homologs (RIRNC 2 and 3) of *R. irregulare* were analyzed with BLASTP with maximum number of target sequences 5,000.

Figure S2.3

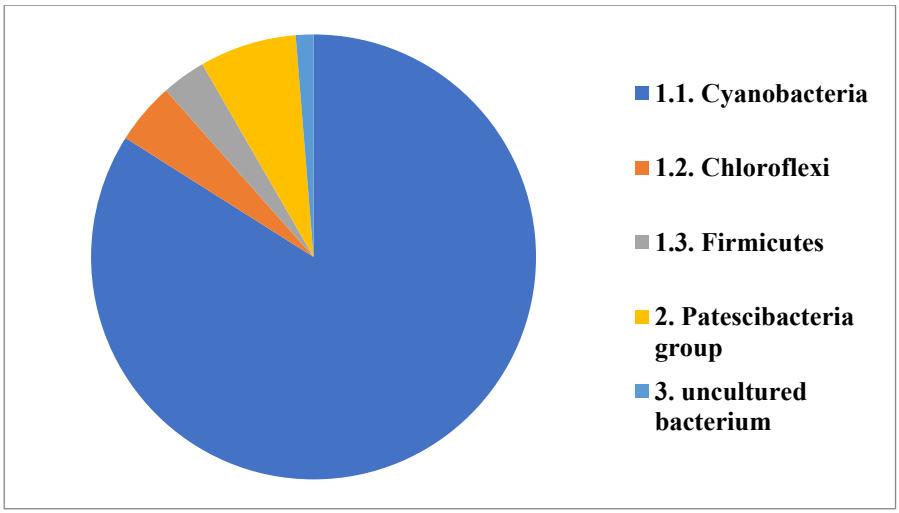
(a) RIRNC 2 (Max target sequences: 100)



(b) RIRNC 2 (Max target sequences: 5,000)



(c) RIRNC 3 (Max target sequences: 100)



(d) RIRNC 3 (Max target sequences: 5,000)

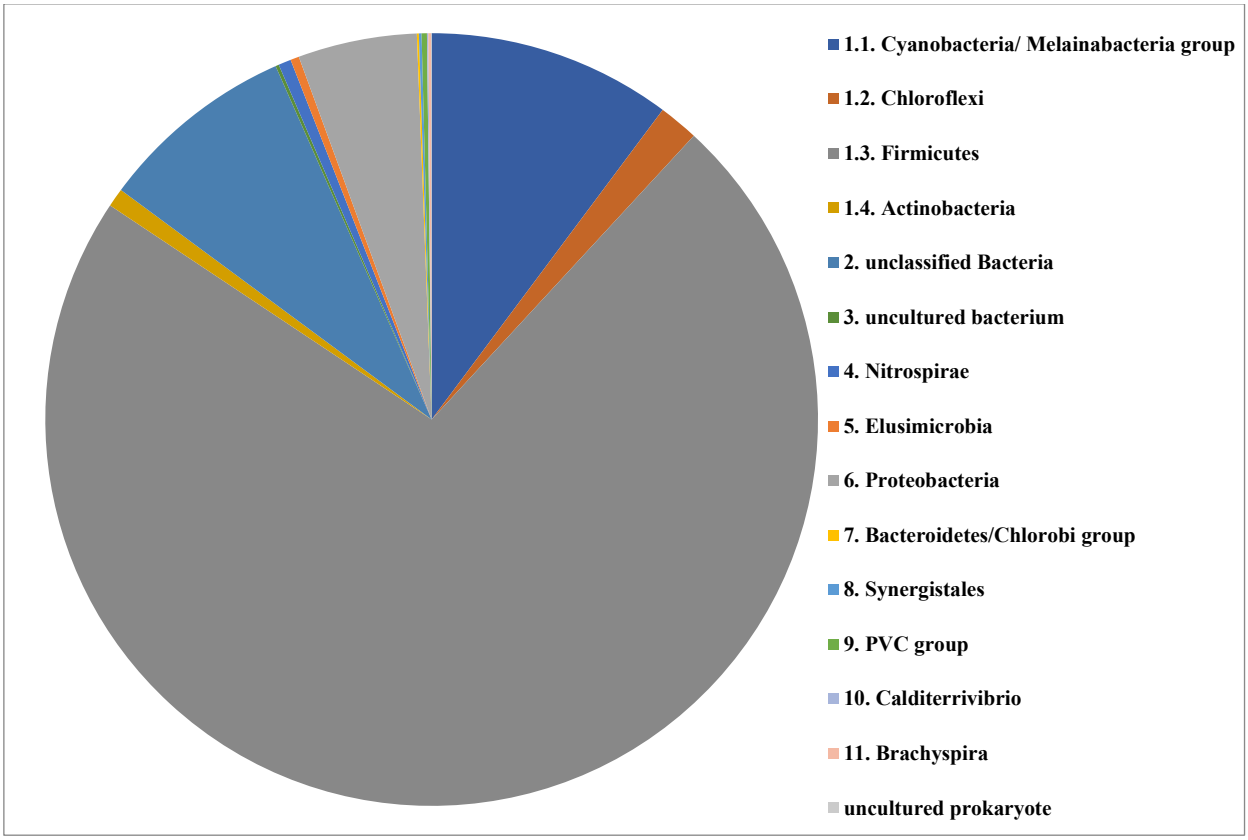
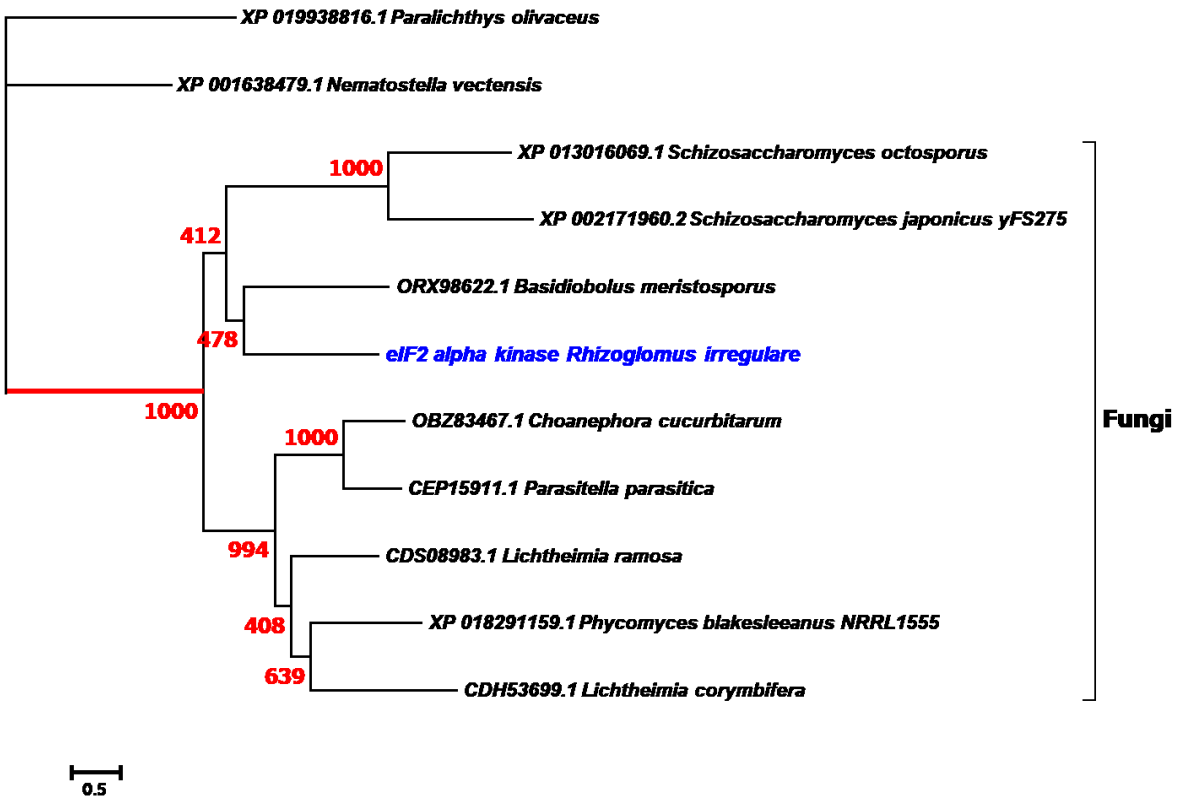


Figure S2.3. Summary of Top 100 and 5000 hits of BLAST of RIRNC 2 and 3. Two different prokaryotic ribonuclease III homologs (RIRNC 2 and 3) of *R. irregulare* were analyzed with BLAST with maximum number of target sequences 100 and 5,000, respectively.

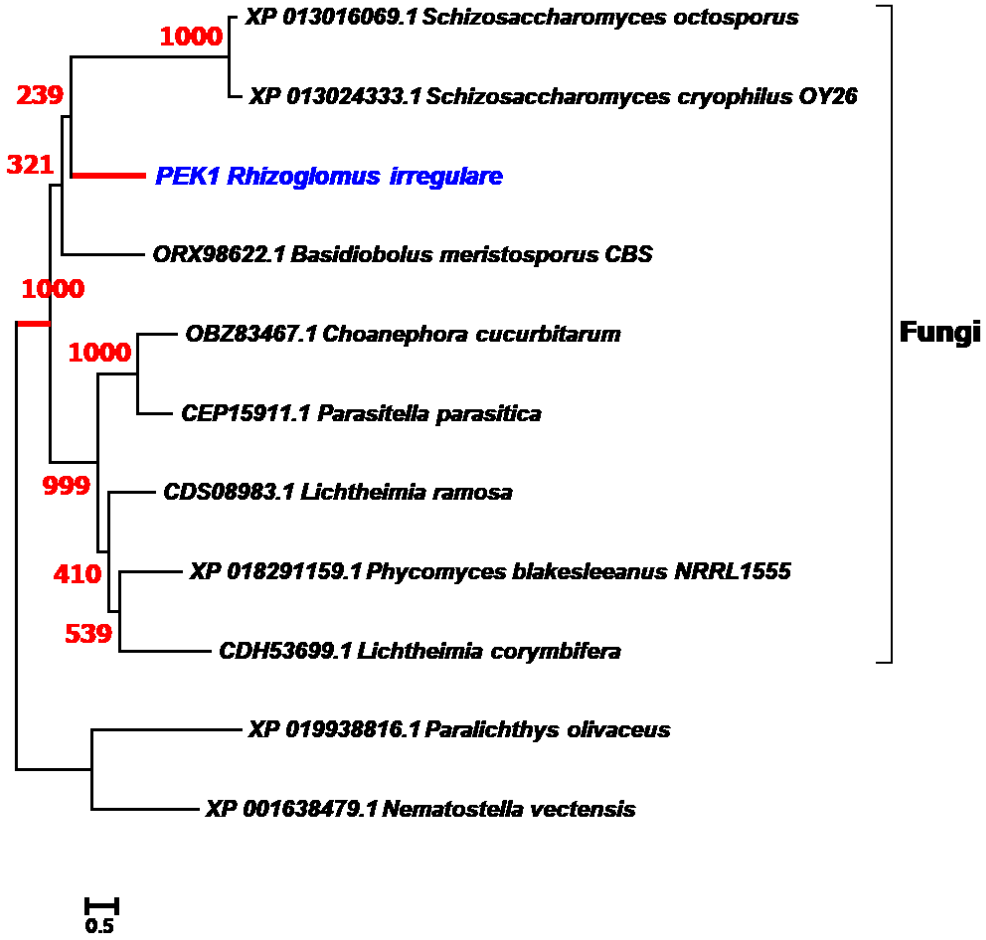
Figure S2.4

(a) The neighboring genes of *rirnc 2*

A. 5' upstream neighbor (PEK1 ,8 intron)

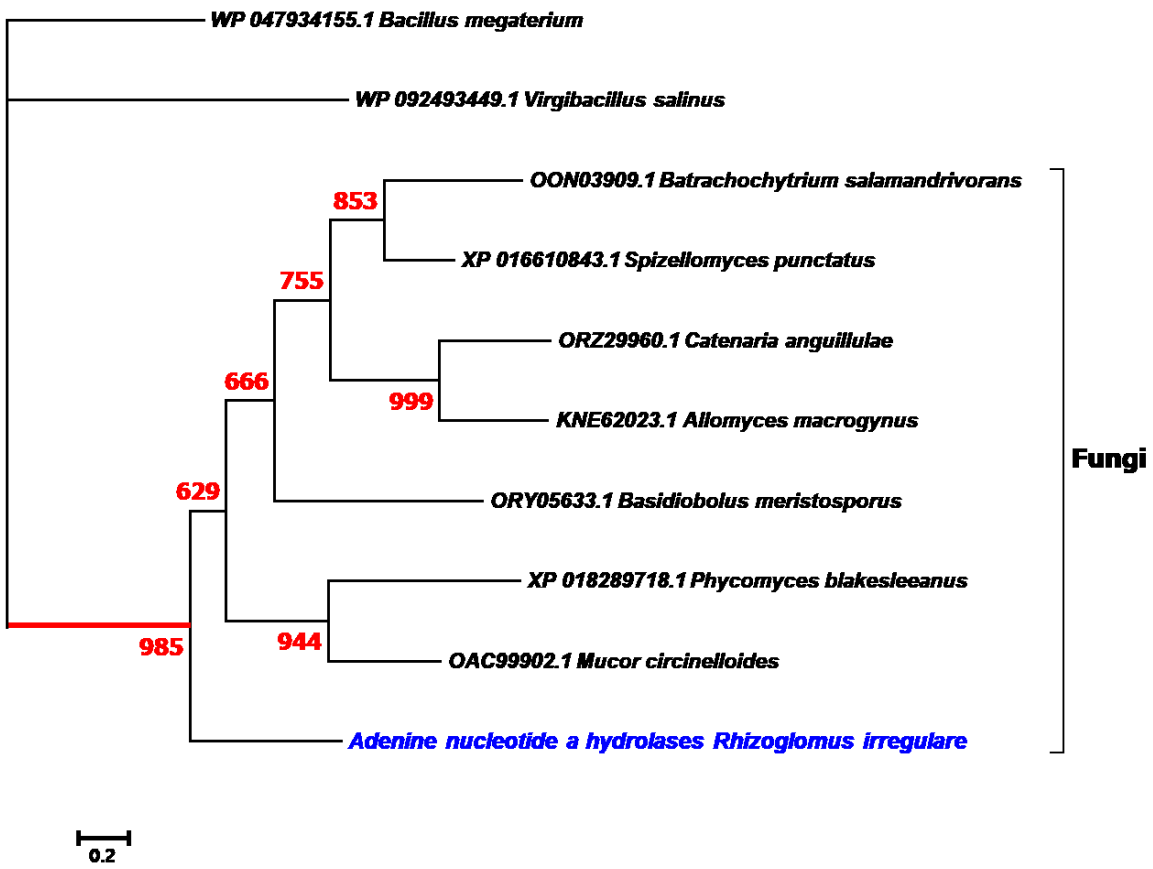


B. 3' downstream neighbor (PEK2, 10 introns)



(b) The neighboring genes of *rirnc 3*

A. 5' upstream neighbor (adenine nucleotide alpha hydrolases-like protein, 4 introns)



B. 3' downstream neighbor (cell devision controlling protein 73, no intron)

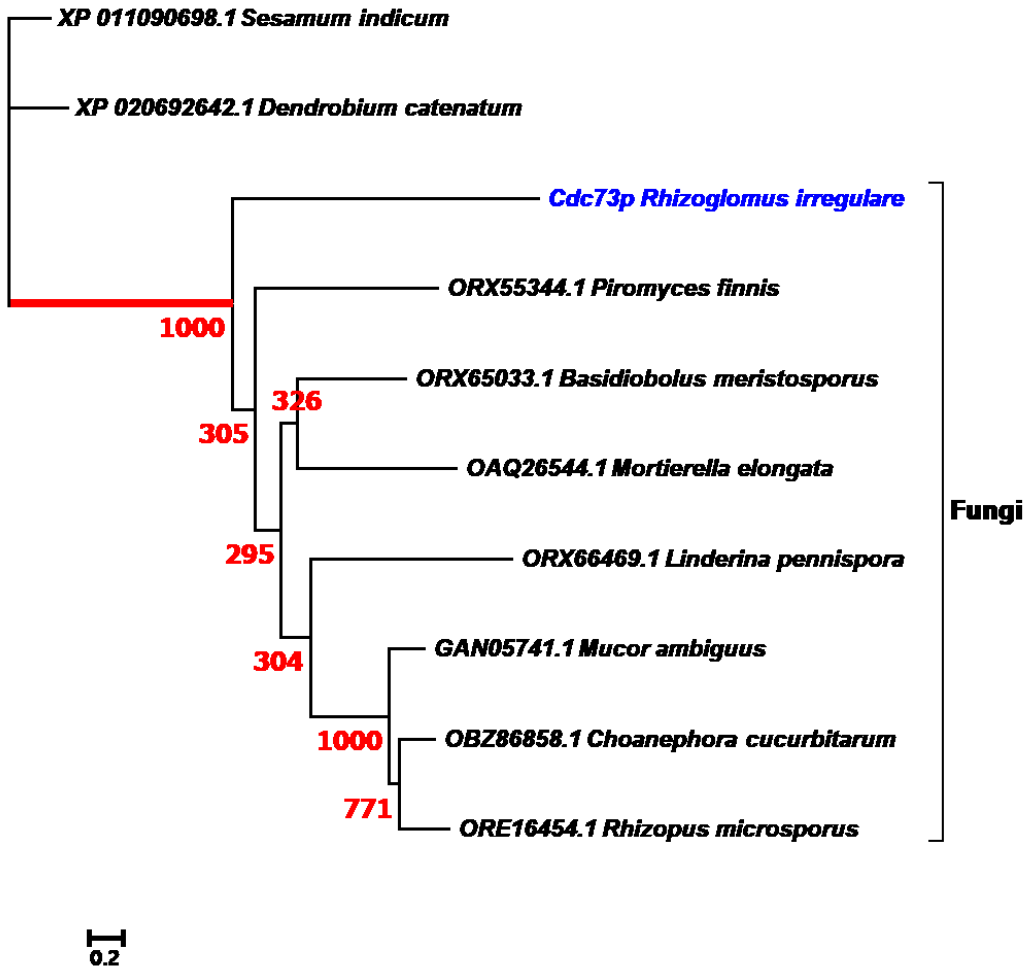


Figure S2.4. Phylogenetic analysis of the proteins encoded by upstream/downstream neighboring genes of *rirnc 2* and *3*. Maximum likelihood of amino acid sequences were analyzed with the LG+I+G (with four distinct gamma categories) phylogenetic model which showed the lowest AIC value. AMF sequences were colored in blue. The numbers at branches correspond to bootstrap support values generated with 1,000 bootstrap replicates. All of four genes were showing their closest homology with sequences from fungi. If there is no prokaryotic homolog available as the gene is eukaryotic specific (PEK1 and PEK2 protein kinases and CDC73), closest homologs from other eukaryotes (plants or animal) were chosen as outgroups. All sequences from *R. irregulare* clearly grouped with other fungal homologs, with clear separation from outgroups.

Figure S2.5

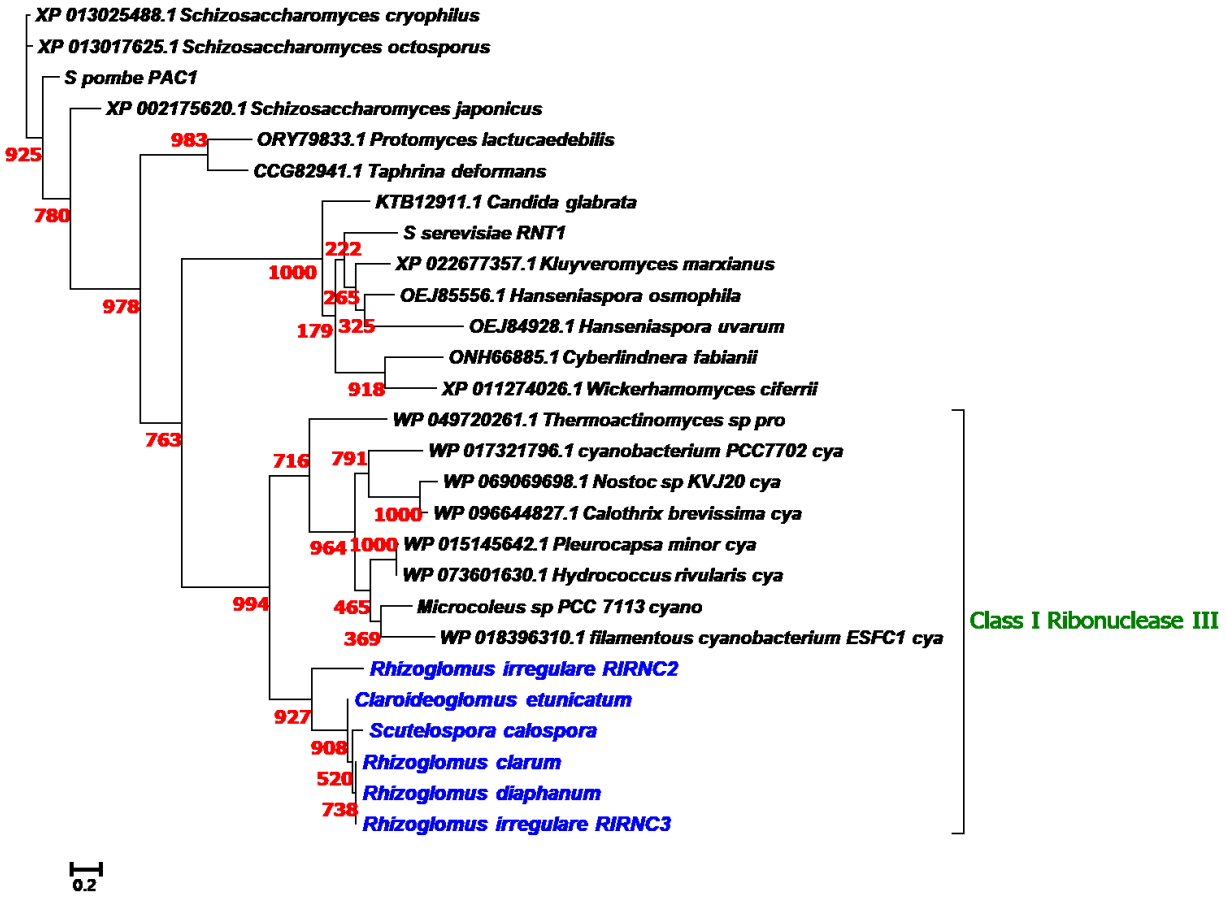


Figure S2.5. Phylogenetic analysis of RIBOc domains of class I (prokaryotic) ribonuclease III enzymes and the most structurally similar class II (eukaryotic) ribonuclease III. Maximum likelihood of amino acid sequences of Class I and II ribonuclease III enzymes were analyzed with the LG+I+G (with four distinct gamma categories) phylogenetic model which showed the lowest AIC value. AMF sequences were colored in blue. The numbers at branches correspond to bootstrap support values generated with 1,000 bootstrap replicates. All of the class I ribonuclease III enzymes were clearly separated with class II ribonuclease III enzymes with high bootstrap supporting (994/1000) indicating the sequence differences in their core catalytic domain (RIBOc). Internal difference between *R. irregulare* RIRNC 2 and 3 were also detected

by monophyletic clustering of sequences from RIRNC 3 homologs, apart from RIRNC 2 (908/1000). Three species of the *Rhizoglomus* genus show their high level of sequence similarity in RIBOc domain of RIRNC 3 (738/1000), compared with other two species of non-*Rhizoglomus* genus.

Figure S2.6

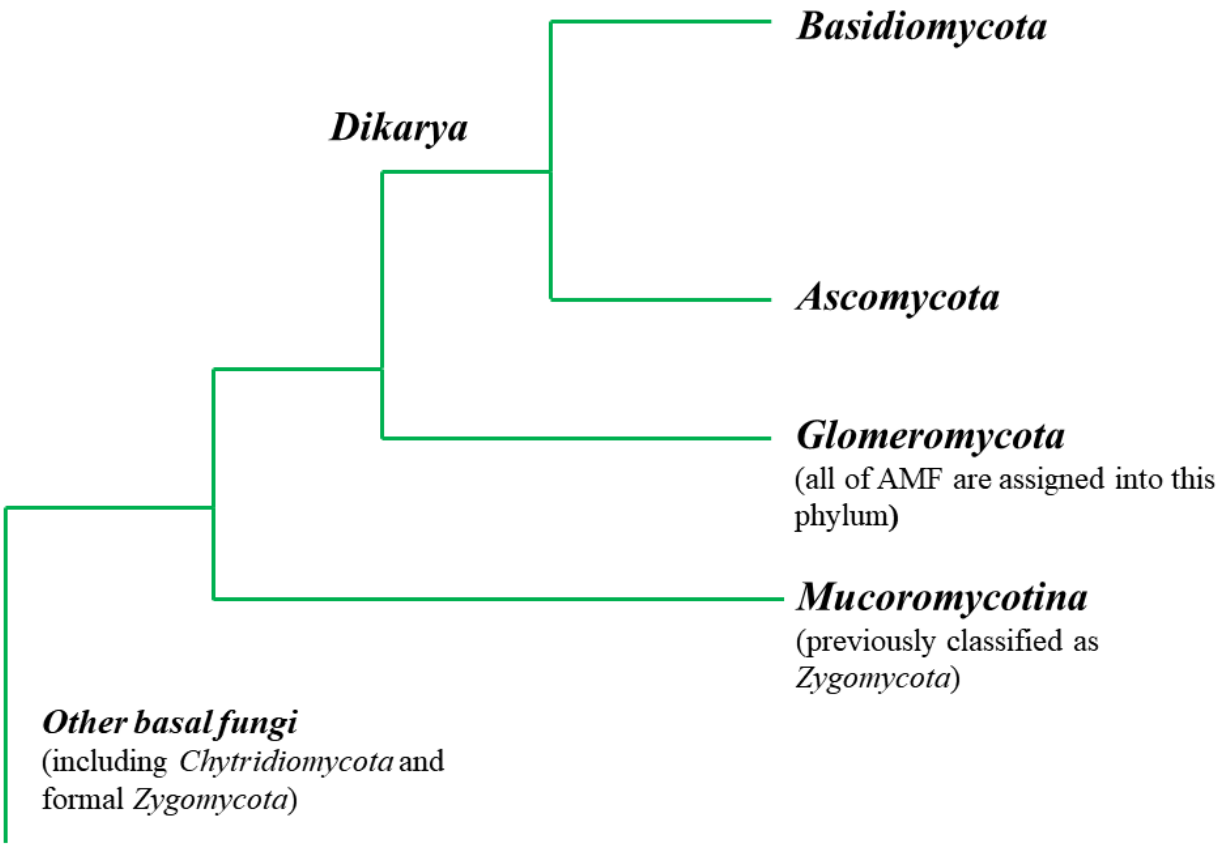


Figure S2.6. Conventional fungal phylogenetic tree regarding *Glomeromycota.* The conventional tree was created with reference fungal phylogenetic tree from (Hibbett, et al. 2007). The query species, *N. crassa*, *S. pombe* and *M. circinelloides* were assigned to *Ascomycota*, *Ascomycota* and *Mucoromycotina*, respectively.

Annex 2 : Supplementary Information (Chapter 3)

Table S3.1. Primers and probes used in TaqMan qPCR assays. Primers and probes were designed using Primer 3 software and oligonucleotides were analyzed by Oligoanalyzer (https://www.idtdna.com/calc/analyzer).

Primer					
Name	**Target gene**	**Sequence (5'->3')**	**length**	**Tm**	**GC%**
WC1-F	***wc1***	TCG GAA ATT ATT GGG CGA AA	20	59.8	40
WC1-R	***wc1***	GAT AAA CCG CGT TGT TAT CG	20	59.2	45
WC2-F	***wc2***	ACA ACG GCA ATG TTA GAT TCA	21	60.2	38.1
WC2-R	***wc2***	GAG TTT CCG ATG GTC CAA TTA	21	59.7	42.9
FRQ-F	***frq***	AAC AGC CAT CTT TAA TAT CAA GAC	24	59.2	33.3
FRQ-R	***frq***	CTT CCC GAA GAT TCT TCA CAT	21	59.7	42.9
18S-F	***18s rrna***	GCT GAA ACT TAA AGG AAT TGA CG	23	60.4	39.1
18S-R	***18s rrna***	TGT CAA TCC TTA CTATGT CTG GA	23	60.4	39.1
UBQ-F	***ubiquitin***	AGA CCA AGT TAA AGC AAA GAT TC	23	59.2	34.8
UBQ-R	***ubiquitin***	TAT AGT CTG ACA ATG TAC GAC CA	23	60.4	39.1
Probe					
Name	**Target gene**	**Sequence (5'->3')**	**length**	**Tm**	**GC%**
WC1-P	***wc1***	ATG GAC ATG TGG CTC TTG GTT CGC	24	69.5	54.2
WC2-P	***wc2***	AAA TCT GCG TCG TCA ACT TCG CGA G	25	69.7	52
FRQ-P	***frq***	TCT CAA CGT ACT GGA CAT CAT GCG GT	26	69.7	50
18S-P	***18s rRNA***	TGC GGC TTA ATT TGA CTC AAC ACG G	25	68.1	48
UBQ-P	***ubiquitin***	ATC AAC AAC GCT TGA TCT TCG CTG GT	26	69.1	46.2