Evolution and Diversity of Sexually-Related Genes in the Arbuscular Mycorrhizal Fungus *Rhizophagus irregularis*

Philippe Charron

Supervisor
Dr. Nicolas Corradi

Faculty of Graduate and Postdoctoral Studies
University of Ottawa

A Thesis Submitted in Partial Fulfillment
of the Requirements for the M.Sc. Degree
in the Ottawa-Carleton Institute of Biology

Thèse soumise en vue de l'obtention
de la Maîtrise en Science à
l'Institut de Biologie d'Ottawa-Carleton

©Philippe Charron, Ottawa, Canada, 2015

Abstract

Arbuscular mycorrhizal fungi (AMF) are ancient organisms that form symbioses with more than 80% of land plants. Fossil evidence of this partnership dates back 460Ma, when land was first colonized by plants. The mutualistic relationship between host roots and the fungus consists of an exchange of essential nutrients for the proliferation of both organisms, highlighting the importance of the mycorrhizal symbiosis. Despite their extreme longevity, a lack of evidence supporting sexual reproduction has labelled AMF as asexual organisms. However, recent evidence seems to point towards the potential of a cryptic sexuality. Specifically, AMF genomes encode for homologues of proteins that have a role in sexual processes in other fungi, including several typically involved in partner recognition, such as mating-type high mobility group (*MATA*-HMG) proteins found in mating-type loci. In my thesis, I expanded our analyses to five isolates of the AMF model organism *Rhizophagus irregularis*, through the acquisition of novel genome data. Some key findings consist of an expansion of *MATA*-HMG proteins, their unique organizations throughout the genome and the presence of a conserved fungal pheromone pathway. In retrospect, this research uncovers an unprecedented number of AMF genes that are homologues to sex-related genes of other fungi and reveals for the first time their atypical genomic architecture, providing valuable information towards the identification of a cryptic sexuality in these ecologically and economically important organisms.

Résumé

Les champignons mycorhiziens à arbuscules (CMA) sont des organismes anciens qui forment des symbioses avec plus de 80% des plantes terrestres. Ce partenariat remonte à 460Ma, lorsque la terre a été colonisée par les plantes. Cette relation mutualiste consiste en un échange de nutriments essentiels pour la prolifération des deux organismes, soulignant l'importance de la symbiose mycorhizienne. Malgré leur longévité surprenante, une absence de preuves à l'appui de la reproduction sexuée a conduit à supposer que les CMA sont des organismes asexués. Toutefois, des études récentes semblent pointer vers le potentiel d'une sexualité cryptique. Plus précisément, les génomes de CMA codent pour des homologues de protéines qui jouent un rôle dans les processus sexuels dans d'autres champignons, dont plusieurs généralement impliquées dans la reconnaissance des partenaires, comme les groupes de haute mobilité reliés aux processus sexuels (*MATA*-HMG). Dans ma thèse, mes analyses focalisaient sur cinq isolats du CMA modèle *Rhizophagus irregularis,* grâce à l'acquisition de nouvelles données génomiques. Certains résultats consistent en une expansion des protéines *MATA*-HMG, leurs organisations uniques au travers du génome et la présence d'une voie de phéromone fongique conservée. En conclusion, cette recherche présente une quantité élevée de gènes qui sont homologues à des gènes sexuels appartenant à d’autres champignons et révèle leurs architectures génomiques atypiques, fournissant de l’information essentielle sur la potentialité d'une sexualité cryptique dans ces organismes écologiquement et économiquement importants.

Table of Contents

Abstract ii

Résumé iii

Acknowledgements v

List of Figures and Tables vi

Chapter 1 - Introduction 1

- Arbuscular mycorrhizal fungi 2
- AMF Life Cycle 3
- Fungal signatures of sex in AMF 4
- Cellular biology of AMF 7
- AMF Classification 8
- AMF Genomics 9
- Research Goals 10

Chapter 2 - Inventory and genomic organization of MATA High-mobility group proteins and potential Sex-Related Genes in a Supposedly Asexual Fungus 12

- Abstract 12
- Introduction 13
- Methods 17
- Results 22
- Discussion 31
- Concluding remarks 33
- Supplementary Information 35

Chapter 3 – Discussion, Future Directions and Concluding Remarks 56

- Importance of Research 56
- Summary of findings 57
- Future Directions 59
- Concluding Remarks 62

References 63

Acknowledgements

I am thankful for Dr. Nicolas Corradi who gave me the chance to start working in his lab during my third year of undergraduate studies and later on supervise me during my Masters degree. For more than 3 years, Nic has been a mentor and a friend, guiding me through both professional and personal endeavours. It has been a great pleasure working with you and I will cherish all the moments we have shared together. I would also like to thank my committee members, Dr. Stéphane Aris-Brosou and Dr. Keith Seifert for the advice and feedback they provided throughout my project.

I would like to thank all former and current members of the Corradi lab who have assisted me throughout my time in the lab; Adrian Pelin, Steve Nkidumana, Timea Marton, Jeanne Ropars, Denis Beaudet, Jessica Noël, Manuela Krüger and especially Rohan Riley for his guidance and patience when I first started in the lab. Also, a big thanks to Benoit Tremblay from Dr. Rolland-Lagan's lab who has been a great friend since my first year of undergraduate studies. You guys are a second family to me and I will always remember the great moments we've shared together.

A special thanks goes out to my parents, sister and girlfriend. You have always encouraged me and believed in me. You have supported me throughout my entire journey and I will forever be grateful for everything you have done for me. I love you all very much and thank you.

List of Figures and Tables

Figure 2.1. Number of genes encoding *MATA*-HMG domains in five isolates of the AMF model *R. irregularis* and various representatives of all major fungal phyla…………….23

Figure 2.2. Schematic representations of three *R. irregularis* isolate A4 genomic regions harbouring tandem-repeated mating type high mobility group (MATA-HMG) genes….25

Figure 2.3. A) Comparison of sex locus structures of Zygomycota phylum members with *R. irregularis.* B) Comparison of sex locus structures of the Pezizomycotina subphylum members (phylum=Ascomycota) with *R. irregularis*…………………………………………………26

Figure 2.4. Gene expression levels of meiosis-specific genes in *R. irregularis* isolates A4, B3, and C2 in standard *in vitro* standalone conditions and *in plantae* conditions measured using qRT-PCR. ………………………………………………………………………….…….28

Figure 2.5. Maximum likelihood phylogenetic tree of cytosine-specific methyltransferase domains (PTHR10629) from DNMT1 fungal proteins…………………………………..…30

Table S2.1. *MATA*-HMG fungal representatives used as query sequences to identify *MATA*-HMGs in *R. irregularis* strains……………….…………..…………………35

Table S2.2. Primers used for qRT-PCR reactions of meiosis-specific genes and controls.

……………………………………………………………………………………36

Table S2.3. List of MAT-HMG domain-containing proteins found in *R. irregularis* isolates and the reciprocal BLAST first hit of each MAT-HMG query sequence….…………………………37

Table S2.4. List of meiosis-specfic genes, mitogen-activated protein kinase (MAPK) pathway genes and repeat-induced point mutations (RIP) genes in *R. irregularis* isolates with their accession numbers. …………………………………………………………………………45

Figure S2.1. Maximum likelihood phylogenetic tree of *MATA*-HMGs amino acid sequences *in R. irregularis* isolate A1. …………………………………………………46

Figure S2.2. Amino acid alignment of HMG122 showing the divergence of sequences among isolates of *R. irregularis*……………………………………………………………52

Figure S2.3. Amino acid alignment of HMG130 showing the divergence of sequences among isolates of *R. irregularis*……………………………………………………………53

Figure S2.4. Amino acid alignment of HMG21 showing the divergence of sequences among isolates of *R. irregularis*..……………………………………………………...……54

Figure S2.5. RIP mutation frequency (C <--> T) across a multiple alignment of an AMF transposable element……………………………………………………………………..55

Chapter 1 - Introduction

Sexual reproduction has been an important factor in the evolution of most eukaryotes, including animals, plants and fungi (Grishkan et al. 2003; Hamilton, Axelrod, Tanese 1990; Kondrashov 1988). Genetic shuffling of parental DNA can lead to genetically diverse progeny through recombination following meiosis. In the absence of sexual reproduction, no mixing of genetic material with compatible mating partners will occur, thus leading to the acquisition of extensive deleterious mutations (Ni et al. 2011). This phenomenon, known as Muller's ratchet, implies that over an extended period of time mutations will accumulate, leading to the eventual extinction of a species in the absence of mechanisms that purge detrimental mutations (Felsenstein 1974). However, a select group of organisms, known as ancient asexuals, are suspected to have evolved through millions of years with no evidence of sexual reproduction (Judson and Normark 1996; Schurko, Neiman, Logsdon 2009). These biological oddities, which include only two groups of organisms (bdelloid rotifers and arbuscular mycorrhizal fungi (AMF)), seem to defy evolutionary theories by persisting through an exclusively clonal propagation, questioning the biological advantages of sex.

The last few decades have seen numerous supposed uniquely asexual fungi become labelled as facultative sexual organisms, with the presence of sexual processes under specific growth conditions (Geiser, Pitt, Taylor 1998; Lucas, Dyer, Murray 2000; Paoletti et al. 2005; Tzung et al. 2001). Under ideal growth conditions, fungi tend to resort to asexual reproduction because it is quicker and more energy efficient (Heitman 2006).

When they are subject to stressful environments, they may resort to mating in order to increase their survival and adaptability (Grishkan et al. 2003).

Arbuscular mycorrhizal fungi

AMF are part of the Glomeromycota phylum, which include organisms characterized for their obligate symbiotic relationships with land plants (Sanders 2003; Schussler, Schwarzott, Walker 2001). The mycorrhizal symbiosis is believed to have been crucial in land colonization by plants, since fossilized remains of AMF date back approximately 460 million years (Brundrett 2002; Redecker, Kodner, Graham 2000). Both AMF and plants benefit from the presence and interactions of their symbiotic partner (Bonfante and Genre 2010). In order to prosper, AMF acquire energy from the plant's photosynthetically-produced carbohydrates and exchange water and key nutrients (such as phosphorous) to the plant in return (Auge 2001; Bucher 2007). AMF grow out from the plant roots through its mycelium and increase the root absorptive network (Friese and Allen 1991), allowing for more efficient mineral extraction from the soil, thus leading to an increase in plant development (Smith and Read 2008). Studies performed on plant development report that AMF-induced plants show improvement in overall growth, highlighting the importance of these ecological fertilizing organisms (Marulanda, Barea, Azcon 2006; Zhang et al. 2005). The ecological effects of AMF are not ubiquitous among all strains, where genetically different individuals can enhance plant growth better than others (Koch, Croll, Sanders 2006). Establishing ways to create an optimal strain that exhibits the most favourable phenotypic traits represents a key step in improving the use of this natural fertilizer in agricultural settings because breeding programs of some crops

and animals are commonly used to increase productivity when specific phenotypes are required (Fraser 2008; Radovanovic and Cloutier 2003). However, this is presently not a feasible option in AMF due to their atypical asexual life cycle.

AMF Life Cycle

The life cycle of AMF starts with spore germination, where hyphae grow out searching for host plant roots to colonize (Parniske 2008). Because AMF are obligate symbionts, their germinated spores have a shortened life span in the absence of host roots (Bago *et al*. 2000). When a potential host is in close-proximity, morphological and chemical changes occur in both members of the mycorrhizal symbiosis, leading to the eventual colonization of the plant by AMF through the growth of appressoria (Besserer et al. 2006). The penetration of the plant roots is followed by the intraradical growth of tree-like structures, known as arbuscules (Bago *et al.* 2000). It is in these arbuscules that the exchange of carbon materials and mineral nutrients takes place (Parniske 2008). Arbuscule formation leads to extraradical growth of AMF into the soil (Bago *et al.* 2000). These filaments, known as mycelia, will expand the absorptive network of the plant roots, gather nutrients and water from the soil, and begin forming spores (Rillig 2004). The spores germinate and the AMF life cycle restarts (Bago *et al.* 2000).

The conventional life cycle of AMF shows no evidence of any mating apparatus, sexual stage or meiotic events, leading to the assumption that these organisms have evolved in a purely clonal matter (Sanders 1999). This unusual lifestyle and their long-term existence has placed AMF in a group of unique organisms known as ancient asexuals: individuals

who have propagated for millions of years with no sexual means to purge deleterious mutations (Judson and Normark 1996; Normark, Judson, Moran 2003). However, is there evidence of fungal sexual traits in AMF genomes?

Fungal signatures of sex in AMF

Meiosis and Recombination

In many eukaryotic species, meiosis is a necessary step in reproduction (Halary et al. 2011). The presence of the "core meiotic machinery" is necessary in the genome in order for the processes to occur (Corradi and Lildhar 2012). We would assume that in genomes of ancient asexuals, these genes would be lost or pseudogenized over time since they have lost their original functions (Riley and Corradi 2013). However, recent evidence suggests that AMF possess some of the necessary genes to take part in sexual life cycle that would involve meiosis (Halary et al. 2011). More specifically, AMF possess genes that encode orthologues of meiosis-specific proteins involved in recombination, such as Rec8, Spo11 and Dmc1 (Halary et al. 2011). The identification of a complete set of meiosis-specific genes in AMF has raised the intriguing possibility that these organisms could be able to undergo a conventional sexual reproduction, and challenge the idea that they are ancient asexuals. However, it is still possible that these genes have been maintained in AMF for other non-sexual purposes, such as DNA repair.

With the presence of the "core meiotic machinery", which is involved in meiotic recombination, found in its genome, one has to wonder if recombination is taking place in AMF. Recombination is an essential process with a primary function of shuffling

homologous alleles and potentially reduces the amount of deleterious mutations (Felsenstein 1974). Presence of recombination has been identified in *Rhizophagus irregularis* (formerly *Glomus intraradices*), a fungus that is part of the Glomeromycota phylum, by sequencing polymorphic nuclear loci and performing phylogenetic relationships among genotypes (Croll and Sanders 2009). The results found in that study seem to contrast all prior beliefs that AMF are ancient asexuals. However, as mentioned previously, without the presence of any evidence of meiosis, it is difficult to specify whether the recombination that was observed is due to meiosis or an independent asexual method, such as mitosis.

Mating-type loci and High Mobility Group proteins

The *MAT*-locus is a genomic region responsible for determining a fungus' sexual identity (Lee et al. 2010b). It contains genes with functions linked to sexual reproduction such as producing DNA binding transcription factors that regulate sexual structure development, partner recognition and nuclear fusion (Hull, Davidson, Heitman 2002). Syngamy is only possible between opposite mating-types (Ni et al. 2011). Two main categories of sexual reproduction have so far been reported in fungi, heterothallism and homothallism (Ni et al. 2011). Heterothallic fungi possess only one copy of the *MAT*-locus and therefore need to mate with an individual that harbours the compatible locus. In contrast, homothallic fungi have two copies of the mating-type locus (one from each "sex") within a single organism's genome, meaning they are able to self-fertilize and produce gametes without needing a compatible partner. Cases of pseudo-homothallism are also known in *Neurospora tetrasperma* (Callac et al. 2006), *Podospora anserina* (Grognet et al. 2014)

and *Agaricus bisporus* (Merino et al. 1996). This type of mating seems to be a hybrid of both homo- and heterothallism, where multinucleate cells made up of nuclei containing either one copy of the *MAT*-locus or the other, leading them to be able to reproduce by means of self-fertilizing even though their nuclear genomes do not contain both mating types (Ni et al. 2011). Presently, how or if sex is performed in AMF is completely unknown.

The structure and gene composition of mating-type loci vary considerably between fungal phyla. However, one key factor of many mating-type loci is the presence of *MATA*-HMG, serving as a regulator of sexual identity (Ni et al. 2011). Previous work done on *R. irregularis* failed to identify the presence of syntenic regions between other fungal phyla's mating-type loci and AMF (Halary et al. 2013; Riley et al. 2014). *MATA*-HMG proteins are characterized by inter-isolate polymorphisms (idiomorphic alleles) that are believed to determine if an individual is one mating-type or another (Ni et al. 2011). In Zygomycota, idiomorphic *MATA*-HMG proteins, SexM and SexP, are present between isolates of different mating-types to function as master regulators of sexual functions (Idnurm 2011). In *Mucor circinelloides*, a mutation in the SexM protein led to sterility (Li et al. 2011). Also, in *Phycomyces blakesleeanus*, there was an up-regulation of the SexM and SexP genes when isolates of different mating types were grown together (Idnurm et al. 2008).

Pheromone-sensing pathways

Many pathways are involved in establishing sexual reproduction between fungal species (Ni et al. 2011). Pheromones play a crucial role when it comes to identifying potential sexual partners (Jones and Bennett 2011) and performing anastomosis (hyphal fusion). In members of the Zygomycota phylum, sexual development is mediated by a pheromone known as trisporic acid (Lee et al. 2010a), which triggers the formation of sexual hyphae (Burmester et al. 2007). In order to complete trisporic acid synthesis, there must be an exchange of precursors in between mating types (Lee et al. 2010a). The presence of the genes responsible for the trisporic acid synthesis pathway in AMF could be supporting evidence for sexuality. In members of the Ascomycota and Basidiomycota phyla, pheromones involved in interactions between mating partners during sexual reproduction are regulated by a mitogen-activated protein kinase (MAPK) pathway (Jones and Bennett 2011). Previous work done on the presence of MAPK genes in AMF revealed that a homologous pathway is functional in members of the *Rhizophagus* family (Halary et al. 2013). However, this analysis was performed on incomplete data, therefore a more thorough look into how these genes have evolved is necessary.

Cellular biology of AMF

In addition to being atypical from other eukaryotes through the absence of obvious sexual processes, AMF differ from conventional eukaryotes through morphological features as well. AMF cells consist of coenocytic hyphae, composed of numerous nuclei flowing together within the same cytoplasmic environment (Marleau et al. 2011). These multinucleated cells have raised many debates regarding whether or not they are

homokaryotic (all nuclei are identical) or heterokaryotic (genetically different nuclei) (Boon et al. 2015; Kuhn, Hijri, Sanders 2001; Lanfranco, Delpero, Bonfante 1999; LloydMacgilp et al. 1996; Ropars and Corradi 2015; Young 2015). Some propose that heterokaryotic cells are created by the accumulation of mutations in the absence of sexuality (Sanders, Clapp, Wiemken 1996), whereas others question the idea of heterokaryosis due to the presence of low quantities of polymorphism in recent genome analyses of the AMF *R. irregularis* (Lin et al. 2014; Tisserant et al. 2013). The identification of the genetic diversity's origin within AMF may be hindered through their capacity of undergoing anastomoses, a process that involves the fusion of hyphae, thus allowing the exchange of cytoplasmic material (Croll et al. 2009; Giovannetti, Azzolini, Citernesi 1999). This mechanism is believed to enhance resource sharing, increasing the performance of AMF hyphal networks. Through these linkages, nuclei of divergent strains could potentially mix, creating a population of genetically different nuclei within the same cell.

AMF Classification

The phylogenetic placement of AMF within the fungal kingdom is another important question that has puzzled researchers. Initially AMF were part of the Zygomycota phylum due to their morphological similarities (i.e. presence of coenocytic hyphae) (Cavalier-Smith 1998). However, further analyses conducted on the small subunit of the ribosomal RNA (SSU rRNA) showed that AMF should be classified under its own phylum known as the Glomeromycota (Schussler, Schwarzott, Walker 2001). The exact location of this phylum has been debated because different results have been obtained

according to the approach taken. For instance, their association with fungi of higher order (Ascomycota and Basidiomycota) was the result of the ribosomal gene analysis (Schussler, Schwarzott, Walker 2001). In other studies performed on nuclear and mitochondrial genes, findings reveal that AMF are more closely related to Zygomycota (Halary et al. 2011; Pelin et al. 2012).

AMF Genomics

Previous efforts to sequence AMF genomes have been laborious and difficult (Martin et al. 2008). A genome sequencing project which started in 2004 recently released a draft genome assembly of *R. irregularis* isolate DAOM197918 and revealed that its genome has an estimated size of 153 Mb, a low GC content (~28%) and a large quantity of transposable elements (~36% according to fosmid Sanger sequencing), all factors weighing in on the complexity of the assembly (Tisserant et al. 2013). However, the assembly was only 101 Mb (66% of the estimated size) and highly fragmented with a poor quality N50 value and number of scaffolds (4186 bp and 28731 scaffolds respectively). This draft genome was useful for preliminary analyses but a better genome sequence would be necessary to understand the nature of these AMF.

Another paper published shortly after investigated nuclear diversity within an isolate by sequencing four individual nuclei from *R. irregularis* DAOM197198, as well as two samples of mycelium from the same isolate (Lin et al. 2014). The analysis revealed that the four nuclear genome assemblies were similar, providing evidence towards the homokaryosis hypothesis. Additionally, the assemblies were pooled together in order to

create a reference genome of all samples. This reference assembly was 140.9 Mb, covering most of the estimated genome size. However, it was made up of a large quantity of small scaffolds (30628 scaffolds with an N50 of 16360 bp). Although this genome assembly was more complete, its poor quality limited its usefulness because the high fragmentation of scaffolds resulted in difficult analyses of large genomic regions.

Research Goals

In our lab, we aim to obtain information towards uncovering the presence of sexual signatures in AMF to understand how these enigmatic organisms have not succumbed to extinction in the absence of sex. Studies on AMF have brought some hints to understanding how these ancient asexuals have been able to prosper for such a long period of time (Halary et al. 2013; Riley and Corradi 2013; Riley et al. 2014). However, these previous works were performed on partial genomic and transcriptomic datasets. In this thesis, I set out to increase our knowledge of sexuality-related genes in AMF through the acquisition and analysis of new genomic data. We sequenced the genomes of five *R. irregularis* isolates that were initially collected from a field site in Taenikon, Switzerland. These isolates, which were previously shown to be genetically different (Croll et al. 2008), were used to evaluate inter-isolate variation in order to infer how genes related to sexuality have evolved in this AMF model species.

These analyses will be covered in the following chapter, entitled "Inventory and genomic organization of *MATA* High-mobility group proteins and potential Sex-Related Genes in a Supposedly Asexual Fungus". I will focus on identifying the presence and diversity of

genomic signatures of sexual reproduction among the five sequenced isolates by answering specific questions: (1) do better genome assemblies reveal an increase in the number of total *MATA*-HMGs found in AMF, (2) is their allelic diversity typical of idiomorphic alleles, (3) does their genomic arrangement resemble those of other fungal mating-type loci, (4) are pheromone sensing pathways and other sexually-related mechanisms conserved among different *R. irregularis* isolates, and (5) do meiosis-specific genes expression patterns vary under different growth conditions. Overall, the presence of a cryptic sexuality seems plausible. Establishing whether typical signatures of sexuality are present in this peculiar organism and how the genes involved in these processes have evolved are fundamental steps in exposing AMF reproduction.

Chapter 2 - Inventory and genomic organization of MATA High-mobility group proteins and potential Sex-Related Genes in a Supposedly Asexual Fungus

Philippe Charron[1], Timea Marton[1], Manuela Kruger [1, 2] and Nicolas Corradi[1]

[1] Canadian Institute for Advanced Research, Department of Biology, University of Ottawa, Ottawa, ON, K1N 6N5, Canada

[2] Academy of Sciences of the Czech Republic, Institute of Botany, Prague, Czech Republic

PC conducted the literature review, performed data analyses in exception for meiosis-specific genes and wrote the initial draft of the manuscript. TM performed all analyses linked to meiosis-specific genes. MK assisted with identification of fungal mating-type loci. NC was the principal investigator of this study, providing important supervision in the data analyses and critical advice during the writing and editing process.

Abstract

Arbuscular mycorrhizal fungi (AMF) are unique organisms that form mutualistic relationships with the majority of land plants and are presently recognized as being ancient asexuals. However, recent genome analyses are starting to challenge the notion of ancient asexuality in AMF, because their genomes have been found to encode for a large number of homologues of proteins that are linked to sexual processes in other eukaryotes. The present study expands the current knowledge of sex-related genes in AMF by

exploring new genome data obtained from several isolates of the AMF model *Rhizophagus irregularis*. These new investigations reveal that a single isolate can potentially code 223 mating-type high mobility group (*MATA*-HMG) proteins, 53% more than previously reported. This gene family also differs in both number and structure among isolates of *R. irregularis*, indicating that genome plasticity is substantial in this species. Some *MATA*-HMGs display unique genomic organizations found in close tandem-repeated regions, whereas others structurally resemble MAT-loci of the Zygomycota and the Ascomycota phyla. We also found that several meiosis-specific genes are not constitutively expressed in AMF mycelia, suggesting that these may be functional in very specific processes that may include meiosis. We also show that AMF genomes harbour a conserved set of homologues for a fungal pheromone pathway, indicating the presence of a potential mating identification system. Overall, the present study expands the current catalogue of genomic signatures potentially linked to sexual reproduction in AMF and reveals their genomic organization, stepping stones in our search for potential mating-type locus in these organisms.

Introduction

Arbuscular mycorrhizal fungi (AMF) are ecologically significant organisms that are obligatorily associated with the roots of over 80% of land plants, including essential agricultural crops such as maize, wheat and tomato (Smith 2008). This mutualistic fungal-plant association, called the mycorrhizal symbiosis, improves the fitness of the plant partner via an expansion of its absorptive network and the subsequent increase in the availability of nutrients from the soil provided by the fungus (primarily phosphates and nitrates)(Auge 2001; Bonfante and Genre 2010; Bucher 2007). In return, the plant

provides photosynthetically produced carbohydrates to the fungus. The mycorrhizal symbiosis has been important in the evolution of plants because it has been reported that fossilized remains of AMF date back approximately 460 million years, during the Ordovician period (Redecker, Kodner, Graham 2000; Brundrett 2002). Because of their obvious benefits for plants, AMF are recognized as keystone organisms within ecosystems, and are used to improve the yield of important plants in organic agriculture and environmental practices (Zhang et al. 2005; Marulanda, Barea, Azcon 2006).

The ecological relevance of AMF is almost certainly a reflection of their genetic content. In fact, there is growing evidence that the "genetic make-up" of AMF individuals (defined here as one mycelium) can drastically impact how the plant will grow – i.e. some AMF individuals of one species can make one plant grow better than others (Koch, Croll, Sanders 2006). This suggests that not all AMF spores are equal when it comes to their potential for agricultural applications and that, perhaps, optimal strains may be "selected" for such purposes. However, in the case of AMF the idea of breeding appears unfeasible, as their life cycle appears to be clonal. In particular, no sexual apparatus, meiosis, or sexual stage has ever been observed (Sanders 1999). This clonal lifestyle, coupled with their long evolutionary history, has placed AMF in a group of unique organisms presently referred to as "ancient asexuals" (Judson and Normark 1996). Importantly, long-term asexuality in eukaryotes is considered an evolutionary oddity, because sexual mechanisms necessary to purge deleterious mutations are absent (Felsenstein 1974).

The widespread notion that AMF are very ancient clones has so far been based on the absence of evidence for sex. However, this idea is now being challenged by analyses of sequence data from these organisms. In particular, previous studies have identified the presence of recombination events in AMF, demonstrating that these organisms possess mechanisms to shuffle genetic material and reduce their load of deleterious mutations (Croll and Sanders 2009; Vandenkoornhuyse, Leyval, Bonnin 2001; den Bakker et al. 2010). Furthermore, the discovery of a expanded and conserved "core meiotic machinery" – i.e. a set of genes that are essential for a conventional meiotic cycle – in many AMF species suggests that recombination events could result from meiosis (as opposed to mitotic recombination) (Halary et al. 2011). Importantly, although the meiotic machinery is present in AMF, the actual process of meiosis has never been formally observed in this lineage.

In the fungal kingdom, sexual identity and partner recognition is determined by specific genomic regions known as the MAT-loci (Lee et al. 2010b). A MAT-locus contains genes responsible for establishing cell identity and producing DNA binding transcription factors, which regulate sexual structure development, partner recognition and nuclear fusion (Hull, Davidson, Heitman 2002). The structure and gene composition of mating-type loci vary considerably between fungal phyla (Ni et al. 2011). However, one key factor of many mating-type loci is the presence of a mating-type high mobility group (*MATA*-HMG) protein serving as a regulator of sexual identity. Surprisingly, previous studies show that the genome of *R. irregularis* harbours at least 146 *MATA*-HMG

homologues, but none of these seems to be located in regions that resemble mating-type loci (Tisserant et al. 2013).

Other pathways involved in partner recognition include genes that encode for specific pheromones and their receptors (Ni et al 2011). Members of the Ascomycota and Basidiomycota phyla regulate the interactions between mating partners during sexual reproduction using a mitogen-activated protein kinase (MAPK) pathway (Jones and Bennett 2011). This pathway is tightly linked to sexual processes in Dikarya, where major cellular responses will promote mating once the pheromones have been detected. Although some genes involved in these pathways have been identified in *R. irregularis*, it is presently unknown whether these are involved in sexual processes (Halary et al. 2013).

In some fungi, cellular processes have also been associated with sexual reproduction in fungi without having a direct impact on mating. A prime example includes repeat-induced point mutations (RIP); a defense mechanism that serves against transposable element (TE) expansions by targeting repeated sequences (Cambareri 1989). In ascomycetes, this process occurs during their sexual stage and is commonly used as a tool to detect sexual reproduction. RIP operates by identifying duplicated DNA sequences and introducing C:G to T:A mutations at target CpN dinucleotide locations (often CA to TA). The genes defective in methylation-2 (DIM2) and RIP defective (RID) are essential for RIP to initiate (Jurkowski et Jeltsch 2011). Another defense mechanism called methylation induced premeiotically (MIP) is responsible for cytosine methylation during meiosis in the ascomycete *Ascobolus immersus* and is controlled by the genes MASC1

(RID homologue) and MASC2 (DIM2 homologue) (Irelan and Selker 1996). All four genes belong to a class of cytosine-methyltransferases known as DNMT1 containing the PTHR10629 domain (Amselem, Lebrun, Quesneville 2015). Certainly, identifying orthologues of these genes in *R. irregularis* and related genomic signatures would suggest that this sex-related process exists in this model AMF.

In the present study, we aim to identify new genomic signatures of sexual reproduction in AMF. This has been done by exploring newly acquired genome data from 5 different isolates of the AMF model *R. irregularis* for genes that are often related to sexual reproduction in other fungi. The availability of better assemblies for this species allowed us to reassess the existence and extent of gene diversity of *MATA*-HMG genes in *R. irregularis*, obtain new information about their genomic organization, and search for potential idiomorphs. This sequence data was also used to verify if a complete MAPK cascade and RIP-like mechanisms exist in these organisms, and to search for existing signs of RIP processes in AMF. Furthermore, expression data from 7 meiosis-specific genes was also investigated to understand how this key sexual process is controlled in *R. irregularis*. Overall, these analyses represent important, additional stepping-stones in understanding the AMF's mode of reproduction.

Methods

Fungal Isolates and Mycelium Extraction

Five isolates from the AMF *Rhizophagus irregularis (*SwiA1, SwiA4, SwiA5, SwiB3 and SwiC2) were cultured and analyzed. These isolates were previously collected from a field

in Taenikon, Switzerland and used to create monoxenic *in vitro* cultures with Ri-T-DNA-transformed *Daucus carota* roots (Croll et al. 2008). They were grown in 90 mm two compartment split plate Petri dishes with minimal medium (M) in order to optimize mycelium and spore formation. Isolates used for the following analyses were chosen for genomic and transcriptomic analyses because they have been shown to be genetically different (Croll et al. 2008). After roughly 40 days of inoculation, mycelium and spores were extracted from the split plates by dissolving the media with a 0.0083 N sodium citrate and 0.0017 N citric acid solution.

DNA Extraction, Sequencing and Genome Assembly

Genomic DNA was extracted from the mycelium and spores of all isolates using an in-house protocol combining a phenol-chloroform extraction with the MasterPure^{TM} Complete DNA and RNA purification kit from Epicentre Biotechnologies (Madison, WI, USA). Extracted DNA was sent to Fasteris S.A. (Geneva, Switzerland), where Next-Generation DNA sequencing was performed using Illumina technology on a HiSeq 2500 sequencing platform. Libraries consisted of mate-pair and paired end reads, resulting in a 70x average coverage for each isolate. Following inspection of sequences for bacterial contamination, *de novo* draft genomes were assembled using SPAdes v3.1 (Bankevich et al. 2012) and scaffolded with SSPACE v3.0 (Boetzer et al. 2011). The assemblies are deposited in Genbank (Bioproject accession numbers PRJNA287285).

In silico identification of AMF MATA-HMG homologues

In common with previous studies (Tisserant et al. 2013; Riley et al. 2014), homologues of *MATA*-HMG genes were found by searching for gene representatives from other fungi (Table S2.1) across our genomes using tBLASTn (Altschul et al. 1990). Due to high divergence levels of mating-specific genes amongst fungal phyla, the cut-off used to identify *MATA*-HMGs was very high. To validate orthology, identified *MATA*-HMGs were validated using reciprocal BLAST procedures against the GenBank nr database. Sequences were also manually inspected to discard redundant *MATA*-HMGs. *MATA*-HMGs were screened for inter-isolate variability by aligning each orthologue using MUSCLE (Edgar 2004). A phylogeny was constructed with PhyML (Guindon et al. 2010) using the conserved amino acid sequences of all *MATA*-HMG isolated genes from isolate A1. The alignment was produced with MUSCLE and trimmed using TrimAl (Capella-Gutierrez, Silla-Martinez, Gabaldon 2009) with the Gappyout option. The evolutionary model implemented for the analysis was LG, as suggested by ProtTest (Abascal, Zardoya, Posada 2005), and 100 bootstraps were used to assess branch support.

In silico identification of meiosis-specific genes and MAPK cascade pathway

Orthologues of meiosis-specific genes and those involved in the MAPK cascade mating-pheromone response pathway were identified in the genome sequence data of our *R. irregularis* isolates using tBLASTx procedures. Initial queries for our searches included published partial gene sequences belonging to both these gene families from AMF (Halary et al. 2011; Halary et al. 2013). The putative complete open reading frames of those genes were recovered using comparisons with BLASTx and manual inspections

against orthologues from other fungi represented in GenBank. Gene orthology was confirmed by performing reciprocal BLASTp searches against the GenBank nr database on NCBI.

RNA extraction and qPCR procedures

Total RNA extractions were conducted using the QIAGEN RNeasy Plant Mini Kit (Qiagen, Venio, Netherlands) and manufacturer's recommended procedures. Each RNA sample originated from roughly 50 mg of *R. irregularis* mycelium. The quantification and quality check of extracted RNA was conducted using Nanodrop spectrophotometer (Fisher Scientific, Waltham, MA, USA) and gel electrophoresis. RNA purification using PROMEGA DNase (Madison, WI, USA) followed the manufacturer's protocol for 30 minutes, with a DNase re-spike of 15 minutes. A total of 100 ng of RNA template was then treated with the iScript^{TM} Reverse Transcriptase Supermix (Bio-Rad Laboratories, Hercules, CA, USA) following the manufacturer's conditions. No-RT (reverse transcriptase) controls were run to assure absence of genomic DNA contamination and success of reverse transcription.

Expression profiles of the seven meiosis-specific genes in AMF were examined using reverse transcriptase PCR (RT-PCR) and quantitative real-time PCR (qPCR). Housekeeping genes (α-tubulin and Ef1a) were used as positive controls, whereas an isoform of GAPDH was used as a negative control. This isoform was initially found non-expressed in a preparatory RNA-seq study of our isolates. All 10 genes (7 MSGs, 2 positive control and 1 negative control) underwent a preliminary study investigating their

presence and absence in cDNA samples by RT-PCR. The 20 µL reactions consisted of 10 µL EconoTaq® PLUS GREEN 2X Master Mix (Lucigen, Middleton, WI, USA), 1µL cDNA template, 10ng of each primer and 7 µL of water. Subsequently qPCR reactions were carried out on a CFX 96 thermal cycler (Bio-Rad Laboratories) and analyzed with the Bio-Rad CFX MANAGER software V3.0 (Bio-Rad Laboratories). Each reaction was performed using the manufacturer's protocol, and consisted of 10 µL of SsoFast 2x Master Mix (Bio-Rad Laboratories), 8 µL of 80x diluted cDNA, 0.5 µL each of forward and reverse primers and 1 µL of H_2O. Primer sequences used can be found in Table S2.2. For all primer sets, optimal annealing temperatures were determined by means of a temperature gradient. Ideal quantification was assured using amplification efficiencies ranging between 90% and 110% with R^2 values of at least 0.98 during standard curve analyses (Taylor et al. 2010). Primer specificity was measured by confirming the presence of a single melt-curve peak. Conditions for qPCR reactions were: initial denaturing at 98 °C for 30 sec, followed by 40 cycles of 98 °C for 5 sec, annealing temperatures ranging from 50.0 °C to 56.6 °C (depending on primer set), then 60 °C elongation temperature for 60 seconds, and a final elongation for 5 min at 60°C. Melt curves were performed at the end of each qPCR reaction with temperatures going from 65 °C to 95 °C at 0.5 °C increments.

Identification of RIP-like events

Homologue detection of genes involved in RIP/MIP in *R. irregularis* was done by using tBLASTn procedures with gene representatives from *Neurospora crassa* (RID = AAM27408 and DIM2 = XP_959891) and *Ascobolus immersus* (MASC1 = AAC49849

and MASC2 = AAC03766). As previously suggested by others (Amselem, Lebrun, Quesneville 2015), the orthology was confirmed by performing a phylogenetic analysis using the cytosine-specific methyltransferase domains of these genes (PTHR10629). Ambiguous and non-informative sites were removed by using the gappyout option in TrimAl. A phylogenetic tree was built using PhyML with 100 bootstraps to assess branch support. The substitution model used for the tree (as determined with ProtTest) was LG with estimated invariable sites and gamma-distributed rates.

A dataset of TEs from *R. irregularis* was created using TransposonPSI (Haas 2011) to test the potential effect of RIP/MIP on AMF transposons. TEs were separated in different categories according to annotations and duplicated copies were extracted. These copies were processed using the software RIPCAL (Hane and Oliver 2008), which looks for RIP-like footprints (transitions from C:G to T:A at specific dinucleotide sites).

Results

Vast expansion of mating-type high mobility group domains

Searching homologues of *MATA*-HMGs in our newly acquired assemblies resulted in the identification of 223 genes, 77 more than previously reported. All homologues contain the fungal domain (CDD ID: cd01389) typical of *MATA*-HMGs, and reciprocal BLASTx procedures revealed their similarity with a vast array of genes responsible for mating in members of the Ascomycota, Basidiomycota, and Zygomycota phyla (Table S2.3). This number of *MATA*-HMGs in *R. irregularis* is of extreme proportions compared to any other fungal species with a sequenced genome (Figure 2.1). A few *MATA*-HMGs harbour

frameshift mutations in some isolates. More precisely, the percentage of pseudogenized *MATA*-HMGs varies between isolates from 4.6% (n=10; isolate A5) to 8.1% (n=18; isolate C2).

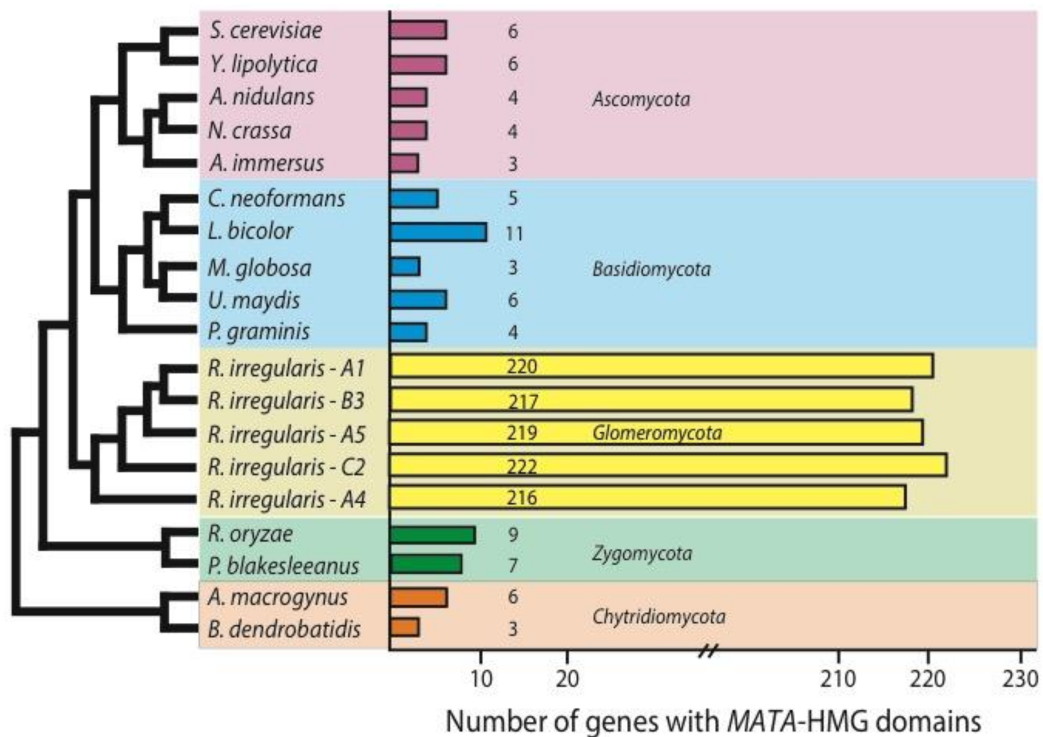


Figure 2.1. Number of genes encoding *MATA*-HMG domains in five isolates of the AMF model *R. irregularis* and various representatives of all major fungal phyla. Schematic phylogenetic tree used to show relationships between fungi (branch lengths are not to scale and are not representative of divergence). Species names in order from top to bottom: *Saccharomyces cerevisiae, Yarrowia lipolytica, Aspergillus nidulans, Neurospora crassa, Ascobolus immersus, Cryptococcus neoformans, Laccaria bicolor, Malassezia globosa, Ustilago maydis, Puccinia graminis, Rhizophagus irregularis, Rhizopus oryzae, Phycomyces blakesleeanus, Allomyces macrogynus* and *Batrachochytrium dendrobatidis*.

Most *MATA*-HMGs diverge substantially among each other, suggesting that their expansion is probably ancient. Although extreme divergence is present between AMF

MATA-HMGs, they are usually more similar to each other than to other fungal mating-type genes Figure S1). In some cases, recent events of gene duplications could also be found (see blue highlighted accession numbers in Table S2.2). Inter-isolate comparisons showed that *MATA*-HMGs genes are usually very conserved in sequence among isolates (with 95% of *MATA*-HMGs being over 90% similar). However, a few showed evidence of substantial divergence (3 *MATA*-HMGs had amino acid similarities below 80%) (Figure S2.2-S2.4), but investigation of flanking areas did not reveal structural similarities with known fungal MAT-loci. Interestingly some *MATA*-HMGs were found to be absent in some isolates, indicating the presence of genome plasticity for this gene family.

Some MATA-HMGs are found in unique genome arrangements

Identification of syntenic conservation between the regions surrounding *R. irregularis* *MATA*-HMGs and known fungal mating loci is an essential step to determine their potential involvement in mating. The genome assemblies we explored provide an optimal framework for such analyses because their scaffolds are much larger than those previously obtained by others. Our investigation reveals that *MATA*-HMGs are organized in very complex and unique ways. Many of these genes (48%) are present as tandem-repeats in the genome. A total of 31 tandem-repeated regions have been found to exist in all isolates, with the number of genes per region ranging from 2 to 13 (Figure 2.2). The total number of these repeated regions also varies among isolates of *R. irregularis* and their genomic locations are not always conserved, another indication that intra-specific genome plasticity is frequent in this species.

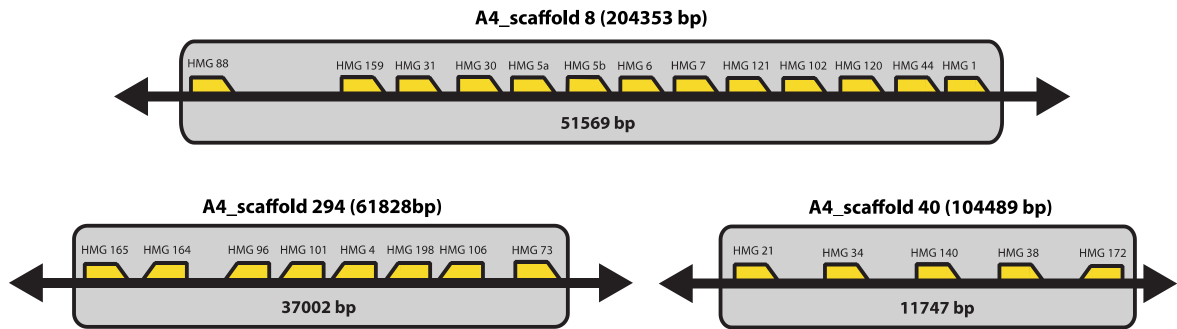


Figure 2.2. Schematic representations of three *R. irregularis* isolate A4 genomic regions harbouring tandem-repeated mating type high mobility group (MATA-HMG) genes. In total, up to 31 regions containing two tandem MATA-HMG genes are found throughout the genome.

While many *MATA*-HMGs are found as tandem repeats, we also found that a few are located next to orthologues found in the mating type locus of other fungi. In one case, an *MATA*-HMG is adjacent to an orthologue of the RNA helicase that resides in the sex-locus of Mucorales (phylum Zygomycota) (Figure 2.3a). The latter locus consists of a *MATA*-HMG flanked by a downstream RNA helicase, as well as an upstream triose phosphate transporter (TPT) or alginate lyase. Importantly, in *R. irregularis* orthologues of the TPT and alginate lyase genes are not linked, and are found in separate scaffolds. Our search also revealed another region of the AMF genome that mirrors the genomic organization of the mating-locus from the Ascomycota subphylum Pezizomycotina. In this case, the *MATA*-HMG is found in close proximity to an orthologue of the SLA2 gene (Figure 2.3b). The overall conserved synteny of these loci with other fungi is intriguing because it suggests their involvement in mating. However, until the function of these genes is revealed, it could also be assumed that their organization is simply the result of stochastic processes.

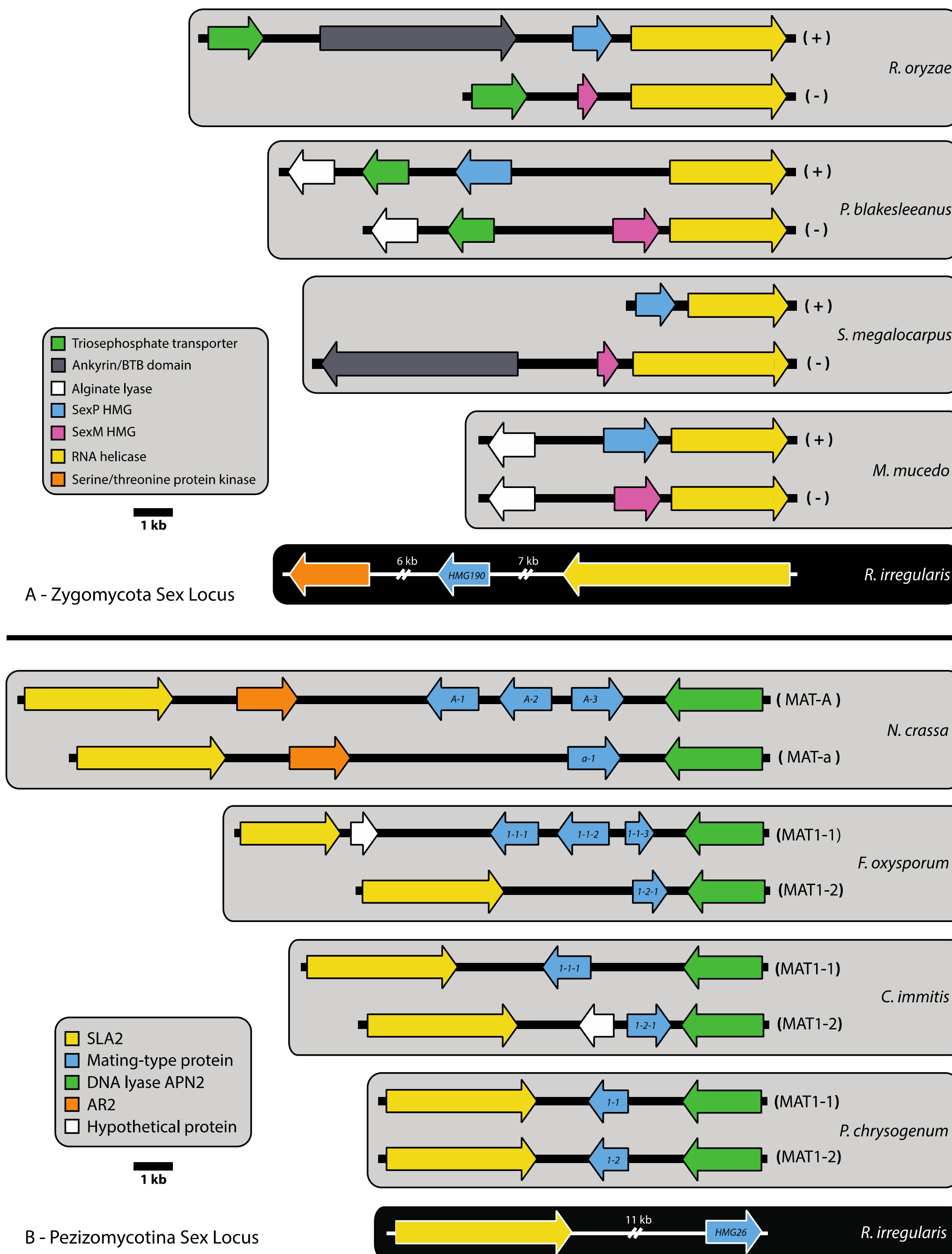
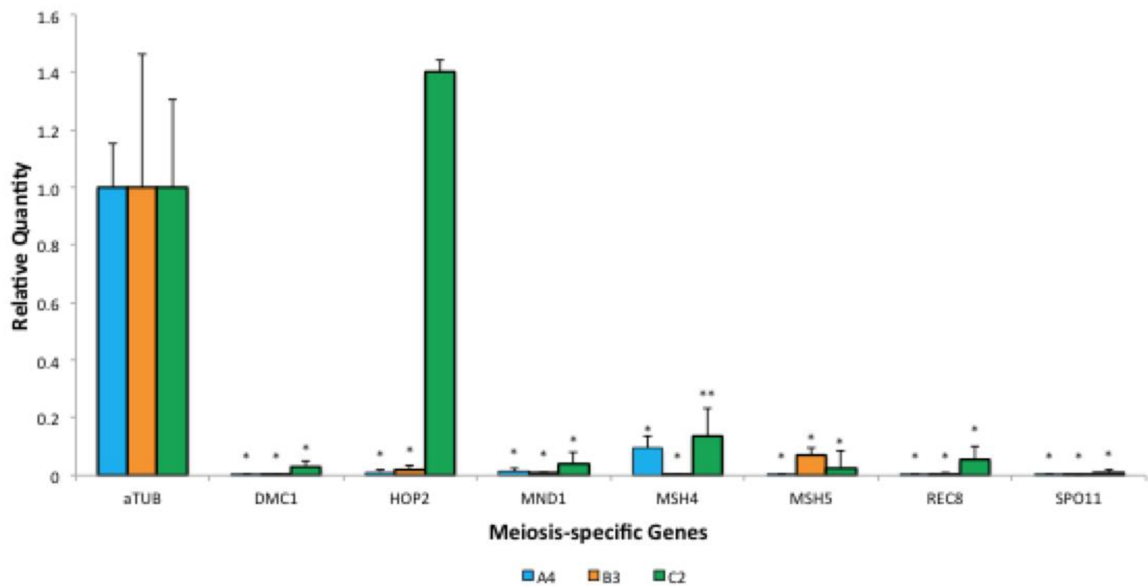


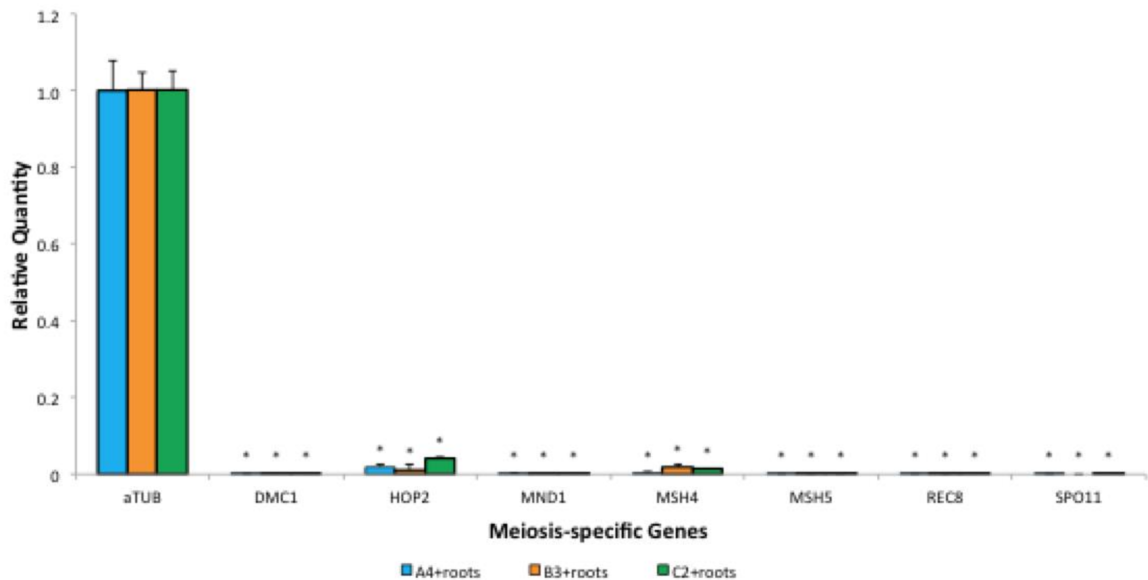
Figure 2.3. A) Comparison of sex locus structures of Zygomycota phylum members with *R. irregularis*. Zygomycete species used are: *Rhizopus oryzae, Phycomyces blakesleeanus, Syzygites megalocarpus* and *Mucor mucedo*. B) Comparison of sex locus structures of the Pezizomycotina subphylum members (phylum=Ascomycota) with *R. irregularis*. Species used are: *Neurospora crassa, Fusarium oxysporum, Coccidioides immitis* and *Penicillium chrysogenum*. Distance between genes in both figures are to scale.

Expression of meiosis-specific genes in AMF in-vitro cultures and in-plantae

All isolates studied contain a complete set of core eukaryotic meiosis-specific genes (Table S2.4). The core-meiotic proteome is composed of 51 genes that are essential for the initiation and termination of meiosis. Within the core, a total of 7 are presently recognized as meiosis specific genes (MSG). These genes are presently known to function solely during the process of meiosis in model organisms. MSG include the genes for sister-chromatid cohesion (Rec8), double strand DNA breaks (Spo11), inter-homolog recombination (Mnd1, Hop2, Dmc1) and class II crossovers (Msh4 and Msh5). Investigation of their expression using Real-Time PCR revealed that most these genes are either weakly or unexpressed under *in vitro* cultures and *in plantae* (Figure 2.4). Hop2 represents the only notable exception, but this gene has been found to be conventionally highly overexpressed in other microbial organisms (Desjardins et al. 2015). The absence of ubiquitous expression for many MSG suggests these may be activated under very specific conditions that may possibly include meiosis.



A) Standalone Conditions



B) *In plantae* Conditions

Figure 2.4. Gene expression levels of meiosis-specific genes in *R. irregularis* isolates A4, B3, and C2 in standard *in vitro* standalone conditions and *in plantae* conditions measured using qRT-PCR. Biological replicates (n=3) for every experimental condition and technical replicates (n=2-3) for all biological replicates. Asterisks represent groups that are significantly dissimilar ($P<0.05$) to other groups for each condition. Standalone conditions were samples consisting of mycelium. *In plantae* conditions were samples consisting of mycelium and the associated plant roots.

Pheromone pathway, RIP-related genes and repeat induced point mutations in R. irregularis

Previous investigations on AMF reported the presence of orthologues necessary to reconstruct the MAPK pathway (Table S2.4), with one notable absence, the scaffolding protein STE5 (Halary et al. 2013). Our analyses confirmed that the 12 genes are present and highly conserved in all investigated *R. irregularis* strains, with 99% amino acid similarity). We also failed to find the pheromone receptor for the α-factor (STE2), but found the receptor for the a-factor (STE3).

While mining for genes related to RIP/MIP, potential homologues of DIM2 and MASC2 were identified (Figure 2.5) (Table S2.4). However, no orthologue of RID was found. This latter gene is also absent in Basidiomycota, and does not necessarily impede the occurrence of MIP, raising the question if other genes are responsible for this process. Among 189 clusters of transposable elements tested in this study, only 1 showed potential signs of RIP-like signatures (Figure S2.5). Other clusters showed no clear bias for dinucleotide substitutions usually associated with RIP.

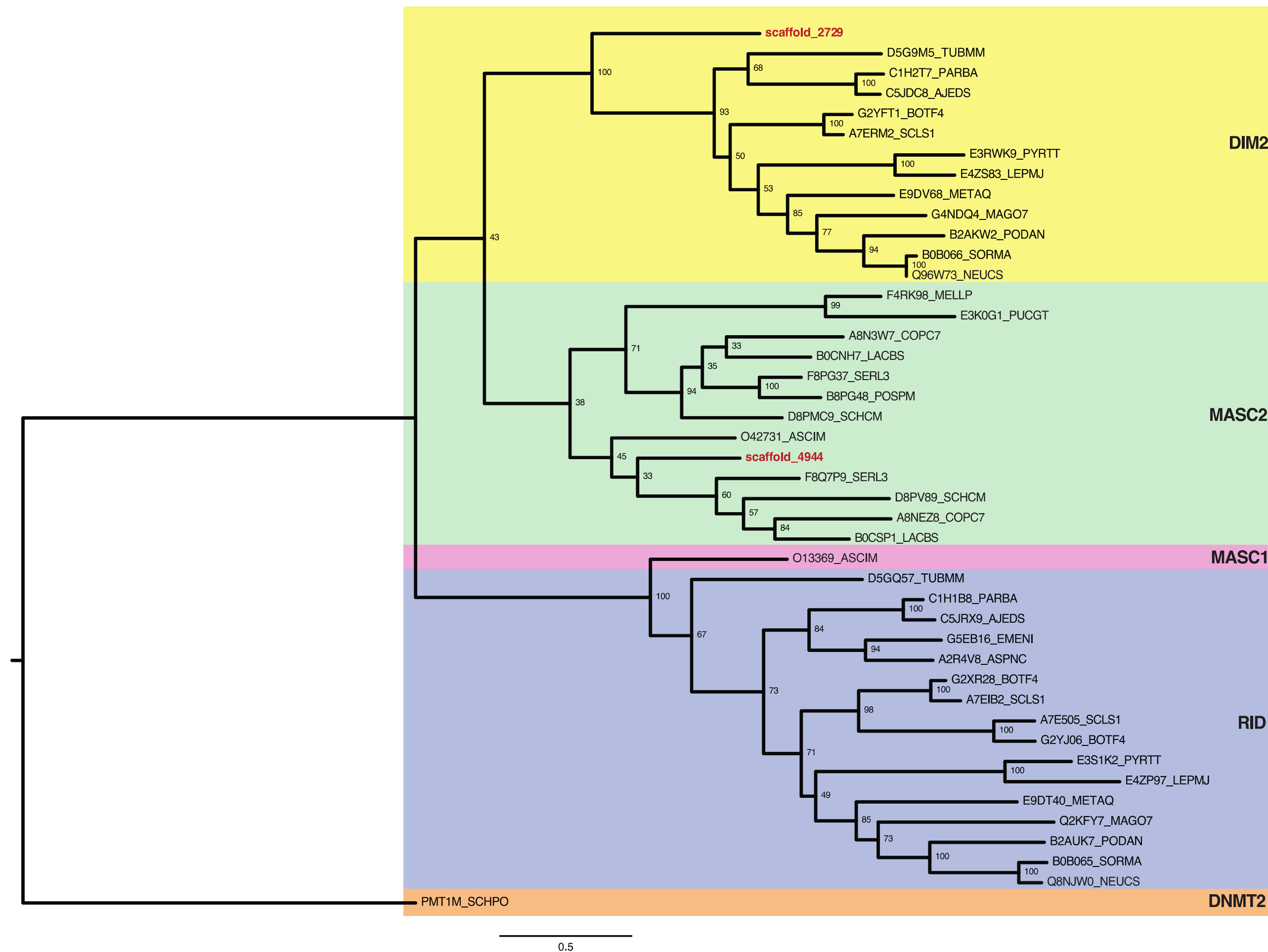


Figure 2.5. Maximum likelihood phylogenetic tree of cytosine-specific methyltransferase domains (PTHR10629) from DNMT1 fungal proteins. In red are homologs from *R. irregularis*. A gene from the DNMT2 family was used as an outgroup. Branch support was confirmed using 100 bootstraps. Genes used for analysis can be found in study by Amselem et al. (2015).

Discussion

A good genome assembly is essential to acquire appropriate information regarding the biology of an organism. In this study, the genomes of 5 isolates were sequenced using a combination of 3 kbp mate pairs and 125 bp paired end reads, resulting in assemblies with statistics that compare very favourably with those previously published. These improved assemblies have been used in this study to optimize searches for genes potentially involved in sexual reproduction in AMF. Our investigations revealed an even greater expansion of *MATA*-HMG domain than previously proposed. The cause of this extreme number of copies is still currently unknown, but the excessive abundance of these transcription factors suggests these probably play a key role in these multinucleate organisms. Certainly, there is presently little evidence that their function is related to mating. Indeed, so far no evidence for the presence of idiomorphs has been found, even for those regions where some synteny is present with known *MAT*-loci. Monitoring the expression patterns of these *MAT*-HMGs under different growth conditions appears necessary to determine whether these genes are potentially involved in mating (i.e. crossing experiments, “stressful” environments) or not.

Perhaps the most intriguing finding of this study is the presence of several tandem-repeated *MATA*-HMGs in the genome of all isolates. This structure appears to be unique in the fungal kingdom, and is somehow reminiscent of an operon. Indeed, in most cases, genes are transcribed in the same direction, so it is interesting to speculate that they may act synergistically. However, preliminary inspections of their expression using RNA-seq found no evidence for their co-regulation (unpublished data). Tandem-repeated *MATA*-

HMGs tend to show higher sequence similarity in comparison to other genes in this library, indicating that these repetitions may be driven by recombination events (mitotic or meiotic).

In AMF, the absence of observational evidence of meiosis has played a key role in their classification as ancient asexuals. However, the identification of 7 MSG orthologues in AMF has demonstrated that AMF could theoretically complete a typical meiotic cycle. Studying the expression patterns of this gene-set will allow us to identify whether these genes have evolved to fill secondary functions, such as DNA repair. Specifically, a constitutive expression of most MSG would indicate that their products are always required, and not necessarily during meiotic processes. The reduced or absence of expression of most MSG suggests that these are regulated under specific conditions, and not all the time. The conditions that trigger their expression are presently unknown, but these could be potentially induce meiosis.

The identification of a conserved MAPK pheromone pathway in all isolates of *R. irregularis* serves as evidence that a putative mechanism of mating compatibility can exist in AMF. Although mating processes have not been observed in AMF, a total inactivity of sexuality would be assumed to lead to pseudogenization of these genes. An identical situation is found in meiosis-specific genes. Overall, these key regulators of mating in other organisms are all present in AMF, and very conserved in sequence.

The absence of a RID gene in *R. irregularis*, and of clear footprints of RIP indicates that this process is most likely absent in this model AMF. Certainly, these finding correlate well with the over-abundance of transposable elements in the *R. irregularis* genome. However, this should not serve as evidence towards the absence of sexuality, as RIP have not been identified in fungal lineages where sexual reproduction is common (i.e. Zygomycota).

Concluding remarks

The long-term absence of sexual processes in AMF has puzzled scientists for some time; this notion has now been challenged by the discovery of many genomic regions in AMF involved in sex or mating in other fungi.

The present study builds on past discoveries of sex-related gene in AMF by providing a better inventory of these regions in AMF genomes. Although no clear sign of a MAT-locus were found, our study sheds a little more light on our understanding of their potential for sex. Indeed, our genome explorations reveal that the number of *MATA*-HMG genes present in AMF is even greater than previously proposed, and we confirm that none of these harbour evidence of idiomorphic divergence. The acquisition of larger assemblies allowed us to determine the genomic organization of this massive gene family, and found that almost half of their members are found in unique genomic tandem arrangements. Hopefully, future analyses based on expression data will shed light onto function/regulation of these atypical genomic regions. Future investigations of gene expression should focus on meiosis-specific genes, as a mean to identify the life cycle or

condition that triggers meiosis. The conservation of the MAPK pheromone-sensing cascade among various isolates of *R. irregularis* is also good indicator that a pheromone-based dialogue (possibly between partners) exists in AMF.

Supplementary Information

Table S2.1. *MATA*-HMG fungal representatives used as query sequences to identify *MATA*-HMGs in *R. irregularis* strains.

Accession	**Organism**	**Protein**
AAK15315	*Candida albicans*	Rfg1
AAF00498	*Ceratocystis eucalypti*	MAT2-1
Q02991	*Cochliobolus heterostrophus*	MAT2
BAA33018	*Coprinopsis cinerea*	Pcc1
AAK83343	*Cryphonectria parasitica*	MAT1-2-1
AAK83344	*Cryphonectria parasitica*	MAT1-1-3
AAC71053	*Fusarium fujikuroi*	MAT1-3
AAG42810	*Gibberella zeae*	MAT1-2-1
AAG42812	*Gibberella zeae*	MAT1-1-3
BAC66503	*Isaria tenuipes*	MAT1-2-1
CAD21099	*Neurospora crassa*	Hypothetical
P36981	*Neurospora crassa*	MAT a-1
Q10116	*Neurospora crassa*	MAT a-3
AAP13349	*Pneumocystis carinii*	Ste11
P35693	*Podospora anserina*	Fpr1
Q08143	*Podospora anserina*	Smr2
CAA06843	*Pyrenopeziza brassicae*	HMG-box protein
CAA06846	*Pyrenopeziza brassicae*	HMG-box protein
P25042	*Saccharomyces cerevisiae*	Rox1
C7U331	*Schizosaccharomyces pombe*	MAT-Mc
P36631	*Schizosaccharomyces pombe*	Ste11
AET35419	*Syzygites megalocarpus*	SexM
AET35422	*Syzygites megalocarpus*	SexP
Q99101	*Ustilago maydis*	Prf1
AAL30836	*Zymoseptoria tritici*	MAT1-2

Table S2.2. Primers used for qRT-PCR reactions of meiosis-specific genes and controls.

Primer name	**Sequence (5' – 3')**
RT-DMC1_F	GAG TTG ACT GCT CGA TTT GCA
RT-DMC1_R	ACC AGC ATA ATC GGT TCG GA
RT-HOP2_F	TAA GAC GGT CGC ACA GAA GG
RT-HOP2_R	CTT CCT GTG AAG GTG ACT CGA
RT-MND1_F	GGA CGA TCG AGC CAC GTT AT
RT-MND1_R	GGC CTC TAA AAG TGC CGG AT
RT-MSH4_F	TTG GAG GCT CTC GAA GAA CG
RT-MSH4_R	GCC CTA ACT GCG TAA CAT CG
RT-MSH5_F	GCT TGT GCG GAA CTT GAC TG
RT-MSH5_R	CGT TTA CTA AGT GGG GCC GA
RT-REC8_F	ACA CAG CCA GAA GGA CAA CA
RT-REC8_R	TCA AGA CCA GCT TCG TGT TCA
RT-SPO11_F	GTG ATT GCT TCC GCC AAA GG
RT-SPO11_R	CCT GAC AAC TCG TTT TTG AGC A
RT-EF1a_F	GCT GGC ACG GTG ATA ACA TG
RT-EF1a_R	GCC CTT AAC TTT GGC ACC AC
RT-aTub_F	AAT CGT CGA CCT CGT CCT TG
RT-aTub_R	AAA CCA GAT CCG GTA CCA CC
RT-GAPDH_F	AAG ATT GGA GAC GCG GTA GG
RT-GAPDH_R	GGA ATG ATC CTA CCA ACG GCT

Table S2.3. List of MAT-HMG domain-containing proteins found in *R. irregularis* isolates and the reciprocal BLAST first hit of each MAT-HMG query sequence. The query sequence accession for each isolate along with the pairwise identity percentage between all isolates are found in the table.

MAT-HMG #	A1 Acc.	A4 Acc.	A5 Acc.	B3 Acc.	C2 Acc.	PI % Between Isolates	Organism	Protein	BLAST E-value	First BLAST Hit Accession
1	KT212333	KT212334	KT212335	KT212336	KT212337	96.9%	*Agaricus bisporus*	Hypothetical	2.00E-16	XP_007328424.1
2	KT211902	KT211903	KT211904	KT211905	KT211906	99.2%	*Xanthoria polycarpa*	MAT1-2-1	1.00E-03	CAI59768.2
3	KT212848	KT212849	KT212850	KT212851	KT212852	99.6%	*Fusarium oxysporum*	MAT-2	1.00E-19	BAA28611.1
4	KT211953	KT211954	KT211955	KT211956	Pseudo	98.9%	*Metarhizium acridum*	HMG box protein	8.00E-05	EFY86728.1
5	KT212012	KT212015	KT212013	KT212014	KT212016	95.1%	*Grosmannia clavigera*	MAT1-2-1	4.00E-03	EFX05114.1
6	KT211992	KT211993	KT211994	KT211995	KT211996	92.2%	*Verticillium alboatrum*	Predicted protein	4.00E-06	XP_003007798.1
7	KT212922	KT212923	KT212924	KT212925	KT212926	86.6%	*Melampsora larici-populina*	Hypothetical	9.00E-07	EGF99649.1
8	KT211987	KT211988	KT211989	KT211990	KT211991	99.2%	*Talaromyces marneffei*	MAT1-2-1	3.00E-03	ABC68485.1
9	KT212601	KT212602	KT212603	KT212604	KT212605	99.6%	*Schizosaccharomyces pombe*	MAT-Mc2	8.00E-18	NP_595867.1
10	KT211917	KT211918	KT211919	KT211920	KT211921	99.3%	*Penicillium chrysogenum*	Hypothetical	4.00E-05	XP_002564591.1
11	Pseudo	Pseudo	Pseudo	Pseudo	Pseudo	-	*Erysiphe necator*	MAT1-2-1	7.90E-02	AEB33764.1
12	KT212348	KT212349	KT212350	KT212351	KT212352	99.8%	*Diaporthe sp.*	MAT1-2-1	5.00E-11	BAE93753.1
13	KT211997	KT211998	KT211999	KT212000	KT212001	99.6%	*Penicillium chrysogenum*	HMG box protein	3.00E-10	CCE33026.1
14	KT212150	KT212151	KT212152 KT212155	KT212154	KT212153	98.6%	*Mucor mucedo*	SexM	4.00E-10	AFA26123.1
15	KT212773	KT212774	KT212775	KT212776	KT212777	98.6%	*Cadophora finlandica*	MAT1-2-1	2.00E-09	ADJ38503.1
16	KT212156	KT212157	KT212158	KT212159	KT212160	99.6%	*Trametes versicolor*	HMG box protein	5.00E-04	EIW55118.1
17	KT212002	KT212003	KT212004	KT212005	KT212006	99.5%	*Cladonia galindezii*	MAT1-2	1.00E-04	AAT48651.1
18	KT212161	KT212162	KT212163	KT212164	KT212165	99.7%	*Fusarium oxysporum*	MAT-2	4.00E-14	BAA28611.1
19	KT212007	KT212008	KT212009	KT212010	KT212011	99.7%	*Rhynchosporium secalis*	HMG box protein	3.00E-04	CAD62166.1
20	KT212166	KT212167	KT212168	KT212169	KT212170	99.2%	*Schizosaccharomyces japonicus*	MAT-Mc	1.00E-17	AFM85245.1
21	KT212017	KT212018	KT212019	KT212020	KT212021	77.3%	*Fibroporia radiculosa*	Predicted protein	2.00E-03	CCM00606.1
22	KT211957	KT211958	KT211959	KT211960	KT211961	99.6%	*Trichoderma atroviride*	Hypothetical	6.00E-03	EHK50111.1
23	KT212022	KT212023	KT212024	KT212025	KT212026	99.4%	*Cercospora apiicola*	MAT1-2	3.00E-10	ABB83710.1
24	Pseudo	Pseudo	Pseudo	Pseudo	Pseudo	-	*Rhynchosporium secalis*	HMG box protein	2.00E-10	CAD62166.1
25	KT212027	KT212028	KT212029	KT212030	KT212031	99.7%	*Fibroporia radiculosa*	Predicted protein	2.00E-02	CCM01306.1
26	KT212181	KT212182	KT212183	KT212184	KT212185	99.0%	*Talaromyces marneffei*	HMG box protein	6.00E-12	XP_002151220.1

Table S2.3. (continued)

27	KT211922	KT211923	KT211924	KT211925	KT211926	96.6%	*Rhynchosporium secalis*	HMG box protein	3.00E-04	CAD62166.1
28	KT212186	KT212187	KT212188	KT212189	KT212190	99.9%	*Trametes versicolor*	Hypothetical	2.00E-09	EIW63176.1
29	Absent	Pseudo	Absent	Absent	KT211927	-	*Talaromyces marneffei*	MAT1-2-1	9.70E-01	ABC68485.1
30	KT212191	KT212192	KT212193	KT212194	KT212195	99.3%	*Piriformospora indica*	Hypothetical	5.00E-11	CCA67490.1
31	KT211928	KT211929	KT211930	KT211931	KT211932	96.2%	*Paracoccidioides brasiliensis*	MAT1-2	3.00E-05	AEI83491.1
32	KT212196	KT212197	KT212198	KT212199	KT212200	99.6%	*Penicillium chrysogenum*	HMG box protein	6.00E-03	XP_002564591.1
33	KT211933	KT211934	KT211935	KT211936	KT211937	96.4%	*Ustilago hordei*	Prf1	2.00E-11	CCF52951.1
34	KT212201	KT212202	KT212203	KT212204	KT212205	91.0%	*Talaromyces marneffei*	MAT1-2-1	1.00E-04	XP_002152469.1
35	KT211938	KT211939	KT211940	KT211941	KT211942	98.7%	*Sordaria fimicola*	MAT a-1	8.00E-04	CAB63226.1
36	KT212206	KT212207	KT212208	KT212209	KT212210	98.8%	*Rhizopus oryzae*	SexP	2.80E-01	ADU04732.1
37	KT212176	KT212177	KT212178	KT212179	KT212180	98.3%	*Trametes versicolor*	Hypothetical	1.80E+00	EIW55066.1
38	KT212211	KT212212	KT212213	KT212214	KT212215	99.4%	*Schizosaccharomyces pombe*	MAT-Mc2	5.00E-05	NP_595867.1
39	KT211943	KT211944	KT211945	KT211946	KT211947	98.9%	*Fibroporia radiculosa*	Hypothetical	8.00E-05	CCM01306.1
40	KT212768	KT212769	KT212770	KT212771	KT212772	98.8%	*Diaporthe sp.*	MAT1-2-1	1.00E-06	BAE93759.1
41	KT211948	KT211949	KT211950	KT211951	KT211952	99.6%	*Aspergillus kawachii*	Hypothetical	2.40E+00	GAA93066.1
42	KT212606	KT212607 KT212611	KT212608	KT212609	KT212610 KT212612	92.8%	*Zymoseptoria tritici*	Hypothetical	4.00E-06	EGP90006.1
43	KT212171	KT212172	KT212173	KT212174	KT212175	99.8%	*Talaromyces marneffei*	MAT1-2-1	8.00E-03	ABC68485.1
44	KT212216	KT212217	KT212218	KT212219	KT212220	96.9%	*Cryphonectria parasitica*	MAT1-1-3	1.00E-07	AAK83344.1
45	KT211907	KT211908	KT211909	KT211910	KT211911	98.5%	*Talaromyces stipitatus*	MAT1-2-1	2.00E-06	XP_002488738.1
46	KT212221	KT212222	KT212223	KT212224	KT212225	99.2%	*Trametes versicolor*	Hypothetical	1.00E-08	EIW52457.1
47	KT211962	KT211963	KT211964	KT211965	KT211966	97.8%	*Schizophyllum commune*	Hypothetical	3.00E-05	XP_003029891.1
48	KT212226	KT212227	KT212228	KT212229	KT212230	97.1%	*Piriformospora indica*	Hypothetical	1.00E-04	CCA72393.1
49	KT211969 KT211972	KT211967	KT211968	KT211971	KT211970 KT211973	94.0%	*Syzygites megalocarpus*	SexP	9.30E-02	AET35404.1
50	KT212062	KT212063	KT212064	KT212065	KT212066	99.6%	*Mycosphaerella populorum*	Hypothetical	1.00E+00	EMF17374.1
51	KT211974	KT211975	KT211976	KT211977	KT211978	98.4%	*Xanthoria polycarpa*	MAT1-2-1	2.00E-03	CAI59768.2
52	KT212067	KT212068	KT212069	KT212070	KT212071	95.4%	*Rhynchosporium secalis*	HMG box protein	2.00E-09	CAD62166.1
53	KT211912	KT211913	KT211914	KT211915	KT211916	99.3%	*Talaromyces marneffei*	MAT1-2-1	3.00E-07	ABC68485.1
54	KT212072	KT212073	KT212074	KT212075	KT212076	99.6%	*Metarhizium acridum*	HMG box protein	3.00E-07	EFY86728.1
55	Pseudo	Absent	KT211979	KT211980	KT211981	99.5%	*Tremella mesenterica*	Hypothetical	7.90E-02	EIW72397.1

Table S2.3. (continued)

56	KT212077	KT212078	KT212079	KT212080	KT212081	99.8%	*Debaryomyces hansenii*	HMG box protein	7.00E-03	XP_459717.2
57	KT212082	KT212083	KT212084	KT212085	KT212086	99.2%	*Metarhizium anisopliae*	MAT1-1-3	5.00E-03	BAE93596.1
58	KT212087	KT212088	KT212089	KT212090	KT212091	97.5%	*Phomopsis sp.*	MAT1-2-1	5.80E-02	AFP89369.1
59	KT212912	KT212913	KT212914	KT212915	KT212916	99.5%	*Pyrenophora teres*	Hypothetical	1.00E-16	XP_003297060.1
60	KT212037	KT212038	KT212039	KT212040	KT212041	99.8%	*Salpingoeca sp.*	Hypothetical	3.40E-02	EGD75508.1
61	KT212092	KT212093	KT212094	KT212095	Pseudo	95.8%	*Schizosaccharomyces japonicus*	MAT-Mc	6.00E-03	AFM85245.1
62	KT212042	KT212043	KT212044	KT212045	KT212046	97.6%	*Piriformospora indica*	Hypothetical	1.00E-04	CCA72393.1
63	KT212047	KT212048	KT212049	KT212050	KT212051	96.4%	*Acyrthosiphon pisum*	Hypothetical	1.00E-05	XP_003245920.1
64	KT212052	KT212053	KT212054	KT212055	KT212056	93.8%	*Metarhizium anisopliae*	MAT1-1-3	6.10E+00	EFZ01123.1
65	KT212917	KT212918	KT212919	KT212920	KT212921	99.4%	*Colletotrichum higginsianum*	HMG box protein	2.00E-04	CCF38267.1
66	Pseudo	Pseudo	Pseudo	Pseudo	Pseudo	-	*Schizosaccharomyces pombe*	MAT-Mc	7.50E-01	AAB28876.1
67	KT212130	KT212131	KT212132	KT212133	KT212134	99.8%	*Colletotrichum higginsianum*	HMG box protein	1.60E-02	CCF38267.1
68	KT212135	KT212136	KT212137	KT212138	KT212139	99.8%	*Phycomyces blakesleeanus*	SexM	9.00E-06	ABX27909.1
69	KT212116	KT212117	KT212118	KT212119	KT212120	99.3%	*Verticillium longisporum*	MAT1-1-2	9.60E-02	AEA29200.1
70	KT212121	KT212122	KT212123	KT212124	KT212125	99.5%	*Gaeumannomyces graminis*	Hypothetical	2.00E-03	EJT79613.1
71	Pseudo	KT212126	KT212127	KT212128	KT212129	99.8%	*Verticillium dahlia*	Hypothetical	8.00E-06	EGY15843.1
72	KT212823	KT212824	KT212825	KT212826	KT212827	99.6%	*Hymenoscyphus pseudoalbidus*	MAT1-2-1	1.00E-03	AFQ90566.1
73	KT212145	KT212146	KT212147	KT212148	KT212149	98.8%	*Colletotrichum higginsianum*	HMG box protein	4.00E-04	CCF38267.1
74	KT212101	KT212102	KT212103	KT212104	KT212105	95.1%	*Baudoinia compniacensis*	MAT1-2-1	2.00E-05	EMC98166.1
75	KT212140	KT212141	KT212142	KT212143	KT212144	99.7%	*Beauveria bassiana*	HMG box protein	1.00E-08	EJP65574.1
76	KT211982	Pseudo	KT211983 KT211985	KT211984	KT211986	99.9%	*Xanthoria polycarpa*	MAT1-2-1	1.20E-02	CAI59768.2
77	KT212654	Pseudo	KT212655	KT212656	KT212657	97.2%	*Fusarium sp.*	MAT1-2-1	5.00E-06	ACJ70020.1
78	KT212658	KT212659	KT212660	KT212661	KT212662	98.4%	*Ophiocordyceps sinensis*	MAT1-2-1	8.00E-04	ACV60399.1
79	KT212663	KT212664	KT212665	KT212666	KT212667	99.5%	*Trametes versicolor*	Hypothetical	2.00E-06	EIW55066.1
80	KT212586	KT212587	KT212588	KT212589	KT212590	95.6%	*Erysiphe necator*	MAT1-2-1	3.00E-05	AEB33764.1
81	KT212668	KT212669	KT212670	KT212671	KT212672	99.1%	*Colletotrichum higginsianum*	HMG box protein	6.00E-08	CCF38267.1
82	KT212673	KT212674	KT212675	KT212676	KT212677	99.5%	*Xanthoria polycarpa*	MAT1-2-1	3.00E-10	CAI59768.2
83	KT212678	KT212679	KT212680	KT212681	KT212682	95.7%	*Trametes versicolor*	Hypothetical	3.00E-07	EIW55066.1
84	KT212683	KT212684	KT212685	KT212686	KT212687	98.1%	*Aspergillus terreus*	Predicted protein	4.00E-06	EAU36401.1
85	KT212688	KT212689	KT212690	KT212691	KT212692	98.6%	*Neurospora pannonica*	MAT a-1	1.00E-06	CAB63345.1

Table S2.3. (continued)

86	KT212693	KT212694	KT212695	KT212696	KT212697	99.8%	*Cladonia galindezii*	MAT1-2	6.00E-06	AAT48651.1
87	KT212698	KT212699	KT212700	KT212701	KT212702	98.7%	*Schizosaccharomyces pombe*	MAT-Mc	9.00E-06	AAB28876.1
88	KT212703	KT212704	KT212705	KT212706	Pseudo	99.2%	*Pyrenophora teres*	Hypothetical	6.00E-11	EFQ94844.1
89	KT212707	KT212708	KT212709	KT212711	KT212710 KT212712	86.0%	*Cadophora finlandica*	MAT1-2-1	4.00E-05	ADJ38503.1
90	Pseudo	KT212714 KT212716	KT212713	KT212715	KT212717	77.8%	*Diaporthe phoenicicola*	MAT1-2-1	1.00E-04	ADD92633
91	KT212718	KT212719	KT212720	KT212721	KT212722	87.9%	*Trichoderma atroviride*	Hypothetical	2.00E-05	EHK50111.1
92	KT212723	KT212724	KT212725	KT212726	KT212727	99.0%	*Talaromyces marneffei*	MAT1-2-1	3.00E-04	ABC68485.1
93	KT212728	KT212729	KT212730	KT212731	KT212732	99.6%	*Zymoseptoria tritici*	HMG box protein	2.00E-14	EGP92003.1
94	KT212733	Pseudo	KT212734	KT212735	KT212736	97.0%	*Verticillium alfalfae*	Predicted protein	5.00E-07	EEY15877.1
95	Pseudo	KT212737	Pseudo	Pseudo	Pseudo	-	*Diaporthe sp.*	MAT1-2-1	4.00E-05	BAE93753
96	KT212738	KT212739	KT212740	KT212741	KT212742	96.8%	*Trametes versicolor*	Hypothetical	3.00E-07	EIW55066.1
97	KT212743	KT212744	KT212745	KT212746	KT212747	99.3%	*Trametes versicolor*	Hypothetical	7.00E-13	EIW61440.1
98	KT212748	KT212749	KT212750	KT212751	KT212752	93.6%	*Aspergillus niger*	Hypothetical	9.00E-08	EHA21813
99	KT212753	KT212754	KT212755	KT212756	KT212757	99.6%	*Fusarium poae*	MAT-2	7.00E-15	CAD59614.3
100	KT212758	KT212759	KT212760	KT212761	KT212762	98.4%	*Schizophyllum commune*	Hypothetical	3.00E-08	EFI94988.1
101	KT212763	KT212764	KT212765	KT212766	KT212767	97.4%	*Trichoderma sp.*	MAT1-2-1	1.00E-07	ADB28879.1
102	KT212907	KT212908	KT212909	KT212910	KT212911	96.0%	*Pisolithus tinctorius*	Hypothetical	3.00E-10	KIO04486.1
103	KT212528	KT212529	KT212530	KT212531	KT212532	99.2%	*Piriformospora indica*	Hypothetical	4.30E-01	CCA69882.1
104	KT212778	KT212779	KT212780	KT212781	KT212782	99.9%	*Talaromyces marneffei*	HMG box protein	1.00E-25	EEA20220.1
105	KT212783	KT212784	KT212785	KT212786	KT212787	99.8%	*Talaromyces stipitatus*	MAT1-2-1	2.00E-08	EED11557.1
106	KT212788	KT212789	KT212790	KT212791	KT212792	98.2%	*Xanthoria polycarpa*	MAT1-2-1	2.00E-05	CAI59768.2
107	KT212793	KT212794	KT212795	KT212796	KT212797	99.2%	*Cryphonectria parasitica*	MAT1-1-3	3.00E-12	AAK83344.1
108	KT212798	KT212799	KT212800	KT212801	KT212802	97.4%	*Fusarium aff. xylarioides*	MAT-2	1.00E-13	CAI45464.3
109	KT212803	KT212804	KT212805	KT212806	KT212807	96.8%	*Fusarium oxysporum*	MAT-2	7.00E-22	BAA28611.1
110	Pseudo	Absent	Pseudo	Pseudo	Pseudo	-	*Grosmannia clavigera*	HMG box protein	1.00E-05	EFX04658.1
111	Absent	Absent	KT212808	Absent	Absent	-	*Metarhizium acridum*	MAT1-2-1	3.00E-08	EFY88585.1
112	KT212591	KT212592	KT212593	KT212594	KT212595	99.6%	*Fusarium oxysporum*	MAT-2	5.00E-09	BAA28611.1
113	KT212809	KT212810	KT212811	KT212812	KT212813	99.7%	*Neurospora terricola*	MAT a-1	3.00E-05	CAB63346.1
114	KT212814	KT212815	KT212816	KT212817	KT212818	98.7%	*Aspergillus nidulans*	Hypothetical	4.00E-13	XP_661271.1

Table S2.3. (continued)

115	Absent	Absent	Absent	Absent	Absent	-	*Neosartorya fischeri*	HMG box protein	8.00E-04	EAW24898.1
116	KT212819	Absent	KT212820	KT212821	KT212822	89.0%	*Neurospora sp.*	MAT a-1	3.00E-05	AAL28013.1
117	KT212828	KT212829	KT212830	KT212831	KT212832	99.6%	*Phycomyces blakesleeanus*	SexM	1.00E-05	ABX27909.1
118	KT212833	KT212834	KT212835	KT212836	KT212837	95.2%	*Schizosaccharomyces pombe*	MAT-Mc	9.00E-08	A49103
119	Absent	Absent	Absent	Absent	Absent	-	*Schizosaccharomyces kambucha*	MAT-Mc	1.00E-03	AAQ82722.1
120	KT212838	KT212839	KT212840	KT212841	KT212842	96.2%	*Talaromyces marneffei*	MAT1-2-1	1.00E-16	XP_002152469.1
121	KT212843	KT212844	KT212845	KT212846	KT212847	95.0%	*Fusarium oxysporum*	Hypothetical	4.00E-07	EGU84400.1
122	KT212523	KT212524	KT212525	KT212526	KT212527	76.9%	*Pseudogymnoascus pannorum*	Hypothetical	1.00E-16	KFY26829.1
123	KT212853	KT212854	KT212855	KT212856	KT212857	95.7%	*Schizosaccharomyces pombe*	MAT-Mc	1.00E-02	A49103
124	KT212858	KT212859	KT212860	KT212861	KT212862	98.7%	*Rhizopus delemar*	Hypothetical	1.00E-12	EIE88156.1
125	KT212863	KT212864	KT212865	KT212866	KT212867	99.3%	*Metarhizium acridum*	HMG box protein	2.00E-09	EFY86728.1
126	KT212868	KT212869	KT212870	KT212871	KT212872	99.1%	*Schizosaccharomyces pombe*	MAT-Mc	2.00E-15	CAA30481.1
127	KT212873	KT212874	KT212875	KT212876	KT212877	99.1%	*Colletotrichum graminicola*	HMG box protein	9.00E-15	EFQ29499.1
128	KT212878	KT212879	KT212880	KT212881	KT212882	98.1%	*Erysiphe necator*	MAT1-2-1	7.00E-07	AEB33764.1
129	KT212883	KT212884	KT212885	KT212886	KT212887	89.1%	*Fusarium avenaceum*	MAT-2	2.00E-10	CAD59612.3
130	KT212888	KT212889	KT212890	KT212891	KT212892	78.4%	*Erysiphe necator*	MAT1-2-1	2.00E-08	AEB33764.1
131	KT212893	KT212894	KT212895	KT212896	KT212897	98.9%	*Verticillium dahliae*	MAT1-1-2	2.00E-06	AEA29202.1
132	KT212898	KT212899	KT212900	KT212901	KT212902	99.7%	*Ascosphaera apis*	STE11	7.00E-15	ABN11480.1
133	Absent	Absent	Absent	Absent	Absent	-	*Schizosaccharomyces pombe*	MAT-Mc	5.00E-10	A49103
134	KT212903	KT212904	KT212905	KT212906	Pseudo	96.0%	*Trametes versicolor*	HMG box protein	8.00E-09	EIW65046.1
135	KT212596	KT212597	KT212598	KT212599	KT212600	99.8%	*Sordaria fimicola*	MAT a-1	2.00E-07	CAB63226.1
136	Pseudo	Pseudo	Pseudo	Pseudo	Pseudo	-	*Trichoderma reesei*	MAT1-1-3	7.00E-08	ACR78246.1
137	KT212111	KT212112	KT212113	KT212114	KT212115	99.5%	*Verticillium longisporum*	MAT1-1-2	2.00E-09	AEA29192.1
138	Pseudo	Pseudo	Pseudo	Pseudo	Pseudo	-	*Schizosaccharomyces pombe*	MAT-Mc	1.00E-04	A49103
139	KT212106	KT212107	KT212108	KT212109	KT212110	92.9%	*Gymnopus luxurians*	Hypothetical	2.00E-09	KIK66800.1
140	KT212613	KT212614	KT212615	KT212616	KT212617	86.8%	*Chaetomium globosum*	Hypothetical	4.00E-10	XP_001230155.1
141	Pseudo	Pseudo	Pseudo	Pseudo	KT212618	-	*Mixia osmundae*	Hypothetical	1.00E-04	GAA95608.1
142	KT212619	KT212620	KT212621	KT212622	KT212623	99.8%	*Erysiphe necator*	MAT1-2-1	4.00E-06	AEB33764.1
143	KT212624	KT212625	KT212626	KT212627	KT212628	99.2%	*Mucor mucedo*	SexM	4.00E-07	AFA26123.1
144	KT212629	KT212630	KT212631	KT212632	KT212633	99.7%	*Aspergillus fumigatus*	MAT1-2-1	1.00E-06	EAL92951.2
145	KT212634	KT212635	KT212636	KT212637	KT212638	99.8%	*Mixia osmundae*	Hypothetical	2.00E-08	GAA98854.1

Table S2.3. (continued)

146	KT212639	KT212640	KT212641	KT212642	KT212643	87.1%	*Epichloe festucae*	MAT A-1-3	7.00E-07	AEI72616.1
147	KT212644	KT212645	KT212646	KT212647	KT212648	91.3%	*Trametes versicolor*	HMG box protein	1.00E-11	EIW55118.1
148	Pseudo	Absent	Absent	Absent	Absent	-	*Davidsoniella eucalypti*	MAT-2	1.00E-05	AAF00498.1
149	KT212649	KT212650	KT212651	KT212652	KT212653	99.4%	*Postia placenta*	Predicted protein	2.00E-05	EED82531.1
150	KT212231	KT212232	KT212233	KT212234	KT212235	99.1%	*Neurospora cerealis*	MAT a-1	4.70E-01	AAL28011.1
151	KT212236	KT212237	KT212238	KT212239	KT212240	97.9%	*Fomitiporia mediterranea*	HMG box protein	1.00E-04	XP_007270408.1
152	KT212241	KT212242	KT212243	KT212244	KT212245	99.7%	*Cladophialophora carrionii*	Hypothetical	6.10E-01	XP_008724511.1
153	KT212246	KT212247	KT212248	KT212249	KT212250	94.9%	*Phaeosphaeria avenaria*	MAT-2	5.00E-03	AAK51442.1
154	KT212251	KT212252	KT212253	KT212254	KT212255	99.4%	*Botryobasidium botryosum*	Hypothetical	1.00E-03	KDQ21312.1
155	KT212256	KT212257	KT212258	KT212259	KT212260	98.7%	*Mixia osmundae*	Hypothetical	1.00E-05	GAA97244.1
156	KT212261	KT212262	KT212263	KT212264	KT212265	99.3%	*Pseudogymnoascus pannorum*	Hypothetical	4.00E-04	KFY24209.1
157	KT212266 KT212270	KT212267	KT212268	KT212269	KT212271	99.8%	*Trichoderma virens*	MAT a-1	2.60E-01	EHK23194.1
158	KT212272	KT212273	KT212274	KT212275	KT212276	98.1%	*Gaeumannomyces graminis*	Hypothetical	4.40E-02	XP_009220656.1
159	KT212277	KT212278	KT212279	KT212280	Pseudo	98.4%	*Wallemia ichthyophaga*	MAT-Mc	3.00E-11	XP_009270306.1
160	KT212281	Pseudo	KT212282	KT212283	Pseudo	95.7%	*Coprinopsis cinerea*	Pcc1	3.00E-05	XP_001830477.1
161	KT212284	KT212285	KT212286	KT212287	KT212288	99.5%	*Postia placenta*	Predicted protein	1.00E-03	XP_002472274.1
162	KT212289	KT212290	KT212291	KT212292	KT212293	99.2%	*Postia placenta*	Predicted protein	1.00E-03	XP_002472274.1
163	KT212294	KT212295	KT212296	KT212297	KT212298	99.6%	*Plicaturopsis crispa*	Hypothetical	7.00E-05	KII92044.1
164	KT212299	KT212300	KT212301	KT212302	KT212303	97.0%	*Blakeslea trispora*	SexM	1.00E-06	CDN67533.1
165	KT212304	KT212305	KT212306	KT212307	KT212308	98.8%	*Schizosaccharomyces pombe*	MAT-Mc	7.00E-03	A49103
166	KT212309	KT212310	KT212311	KT212312	KT212313	98.6%	*Fusarium pseudograminearum*	Hypothetical	1.00E-03	XP_009257944.1
167	KT212314	KT212315	KT212316	KT212317	KT212318	99.9%	*Ophiocordyceps robertsii*	MAT1-2-1	1.20E-01	AGJ84380.1
168	KT212319	KT212320	KT212321	KT212322	KT212323	98.9%	*Botryobasidium botryosum*	Hypothetical	1.00E-03	KDQ21312.1
169	KT212324	KT212325	KT212326	KT212327	KT212328	99.3%	*Kazachstania naganishii*	Hypothetical	1.10E-01	CCK68154.1
170	KT212329	KT212330	KT212331	KT212332	Pseudo	99.1%	*Botryobasidium botryosum*	Hypothetical	5.50E-02	KDQ21312.1
171	KT212096	KT212097	KT212098	KT212099	KT212100	99.5%	*Pneumocystis murina*	Hypothetical	2.00E-08	XP_007874266.1
172	KT212338	KT212339	KT212340	KT212341	KT212342	96.9%	*Arthrobotrys oligospora*	Hypothetical	1.00E-05	XP_011122279.1
173	KT212343	KT212344	KT212345	KT212346	KT212347	97.5%	*Paxillus rubicundulus*	Hypothetical	5.00E-10	KIK99891.1
174	KT212057	KT212058	KT212059	KT212060	KT212061	99.7%	*Rhizopus oryzae*	SexM	1.30E-01	ADT91556.1
175	KT212353	KT212354	KT212355	KT212356	KT212357	99.8%	*Fomitiporia mediterranea*	HMG box protein	5.90E-02	XP_007270408.1

Table S2.3. (continued)

176	KT212358	KT212359	KT212360	KT212361	KT212362	100.0%	*Lichtheimia corymbifera*	Predicted protein	9.00E-04	CDH51762.1
177	KT212363	KT212364	KT212365	KT212366	KT212367	100.0%	*Schizosaccharomyces pombe*	MAT-Mc	2.00E-03	A49103
178	KT212368	KT212369	KT212370	KT212371	KT212372	99.6%	*Fusarium oxysporum*	MAT-2	6.00E-03	BAA28611.1
179	KT212373	KT212374	KT212375	KT212376	KT212377	99.3%	*Pyrenopeziza brassicae*	HMG box protein	8.00E-09	CAA06843.1
180	KT212378	KT212379	KT212380	KT212381	KT212382	98.3%	*Cladophialophora carrionii*	Hypothetical	4.10E-01	XP_008724511.1
181	KT212383	KT212384	KT212385	KT212386	KT212387	87.7%	*Fomitopsis pinicola*	Hypothetical	2.00E-04	EPS99600.1
182	KT212388	KT212389	KT212390	KT212391	KT212392	95.7%	*Puccinia graminis*	Hypothetical	6.00E-04	XP_003335886.2
183	KT212393	Pseudo	KT212394	KT212395	KT212396	94.1%	*Auricularia delicata*	Hypothetical	5.00E-04	XP_007341626.1
184	KT212397	KT212398	KT212399	KT212400	KT212401	99.1%	*Fusarium graminearum*	Hypothetical	9.80E-02	XP_389069.1
185	KT212402	KT212403	KT212404	KT212405	KT212406	97.3%	*Rhizoctonia solani*	Hypothetical	8.30E-01	CC_36097.1
186	KT212407	KT212408	KT212409	KT212410	KT212411	99.8%	*Postia placenta*	Predicted protein	6.00E-07	XP_002472274.1
187	KT212412	KT212413	KT212414	KT212415	KT212416	98.7%	*Trametes versicolor*	Hypothetical	2.00E-04	XP_008042321.1
188	Pseudo	KT212417	KT212418	Pseudo	Pseudo	96.9%	*Fomitiporia mediterranea*	HMG box protein	4.00E-03	XP_007270408.1
189	KT212419	KT212420	KT212421	KT212422	KT212423	99.2%	*Coprinopsis cinerea*	HMG box protein	3.00E-03	XP_001828950.2
190	KT212424	KT212425	KT212426	KT212427	KT212428	99.7%	*Fusarium oxysporum*	Hypothetical	5.00E-04	EXA30486.1
191	Absent	Absent	Absent	Absent	Absent	-	*Cladophialophora psammophila*	Hypothetical	4.50E+00	XP_007751801.1
192	Absent	Absent	Absent	Absent	Absent	-	*Talaromyces marneffei*	MAT1-2-1	2.20E-01	ABC68485.1
193	Pseudo	KT212429	KT212430	Pseudo	KT212431 KT212432 KT212433	94.4%	*Aspergillus nidulans*	Hypothetical	4.00E-04	XP_661271.1
194	KT212434	KT212435	KT212436	KT212437	Pseudo	99.1%	*Tulasnella calospora*	Hypothetical	4.00E-12	KIO16416.1
195	KT212438	KT212439	KT212440	KT212441	KT212442	99.4%	*Melampsora larici-populina*	Hypothetical	1.00E-07	XP_007415639.1
196	KT212443	KT212444	KT212445	KT212446	KT212447	99.7%	*Botryobasidium botryosum*	Hypothetical	3.00E-03	KDQ21312.1
197	KT212448	KT212449	KT212450	KT212451	KT212452	99.3%	*Coprinopsis cinerea*	Hypothetical	3.00E-03	XP_001837119.2
198	KT212453	KT212454	KT212455	KT212456	KT212457	98.8%	*Moniliophthora roreri*	HMG box protein	6.00E-03	XP_007849956.1
199	KT212458	KT212459	KT212460	KT212461	KT212462	93.3%	*Talaromyces stipitatus*	MAT1-2-1	9.00E-05	XP_002488738.1
200	KT212463	KT212464	KT212465	KT212466	KT212467	99.7%	*Dichomitus squalens*	Hypothetical	6.00E-08	XP_007364800.1
201	KT212468	KT212469	KT212470	KT212471	KT212472	96.2%	*Lichtheimia corymbifera*	SexM	3.00E-04	CDH60400.1
202	KT212473	KT212474	KT212475	KT212476	KT212477	95.0%	*Melampsora larici-populina*	Hypothetical	5.10E-01	XP_007417104.1
203	KT212478	KT212479	KT212480	KT212481	KT212482	99.2%	*Fusarium oxysporum*	Hypothetical	4.00E-03	EXA30486.1
204	KT212483	KT212484	KT212485	KT212486	KT212487	99.3%	*Lichtheimia corymbifera*	SexM	6.70E-02	CDH60400.1

Table S2.3. (continued)

205	KT212488	KT212489	KT212490	KT212491	KT212492	99.1%	*Bipolaris sorokiniana*	MAT-2	1.40E-02	Q9P445.1
206	KT212493	KT212494	KT212495	KT212496	KT212497	92.6%	*Coprinopsis cinerea*	Hypothetical	1.70E-01	XP_001837119.2
207	KT212498	KT212499	KT212500	KT212501	KT212502	99.9%	*Postia placenta*	Predicted protein	5.00E-04	XP_002473014.1
208	KT212503 KT212508	KT212504 KT212509	KT212505 KT212510	KT212506 KT212511 KT212513	KT212507 KT212512	97.0%	*Schizophyllum commune*	Hypothetical	5.00E-05	XP_003035536.1
209	Pseudo	KT212514	KT212515	KT212516	KT212517	99.6%	*Postia placenta*	Predicted protein	2.00E-04	XP_002472274.1
210	KT212518	KT212519	KT212520	KT212521	KT212522	99.7%	*Talaromyces marneffei*	MAT1-2-1	1.00E-03	ABC68485.1
211	KT212533	KT212534	KT212535	KT212536	KT212537	99.4%	*Colletotrichum gloeosporioides*	Hypothetical	1.60E+00	EQB44894.1
212	KT212538	KT212539	KT212540	KT212541	KT212542	100.0%	*Tulasnella calospora*	Hypothetical	2.00E-03	KIO23466.1
213	KT212543	KT212544	KT212545	KT212546	KT212547	97.1%	*Fibroporia radiculosa*	Predicted protein	3.00E-04	CCM01306.1
214	KT212548	KT212549	KT212550	KT212551	KT212552	98.5%	*Suillus luteus*	Hypothetical	2.00E-08	KIK47880.1
215	KT212553	KT212554	KT212555	KT212556	KT212557	98.9%	*Postia placenta*	Predicted protein	9.00E-07	XP_002472274.1
216	KT212558	KT212559	Pseudo	Pseudo	KT212560	85.8%	*Stereum hirsutum*	Hypothetical	9.70E+00	XP_007309566.1
217	KT212563 KT212565	KT212561	KT212562	KT212564	KT212566	98.7%	*Phialocephala subalpina*	MAT1-1-3	8.40E-02	ADJ38447.1
218	KT212567	KT212568	KT212569	KT212570	KT212571	99.7%	*Pisolithus microcarpus*	Hypothetical	4.50E-02	KIK19531.1
219	KT212572	KT212573	KT212574	KT212575	KT212576	98.3%	*Penicillium italicum*	HMG box protein	2.00E-04	KGO71276.1
220	Pseudo	Pseudo	Pseudo	Pseudo	Pseudo	-	*Ophiocordyceps robertsii*	HMG box protein	2.00E-03	EPE03784.1
221	KT212577	KT212578	KT212579	KT212580	KT212581	83.5%	*Gaeumannomyces graminis*	Hypothetical	4.00E-01	XP_009220656.1
222	KT212582	KT212583	KT212584	KT212585	Pseudo	82.9%	*Pyrenopeziza brassicae*	HMG box protein	7.00E-08	CAD79639.1
223	KT212032	KT212033	KT212034	KT212035	KT212036	99.3%	*Leptosphaeria bigblosa*	MAT1-2-1	4.00E-10	AAV69628.1

Table S2.4. List of meiosis-specfic genes, mitogen-activated protein kinase (MAPK) pathway genes and repeat-induced point mutations (RIP) genes in *R. irregularis* isolates with their accession numbers.

Protein	A1 Acc.	A4 Acc.	A5 Acc.	B3 Acc.	C2 Acc.
Meiosis-specific genes					
Dmc1	KT694213	KT694214	KT694215	KT694216	KT694217
Hop2	KT694228	KT694229	KT694230	KT694231	KT694232
Mnd1	KT694238	KT694239	KT694240	KT694241	KT694242
Msh4	KT694243	KT694244	KT694245	KT694246	KT694247
Msh5	KT694248	KT694249	KT694250	KT694251	KT694252
Rec8	KT694253	KT694254	KT694255	KT694256	KT694257
Spo11	KT694258	KT694259	KT694260	KT694261	KT694262
MAPK pathway					
Bem1	KT694193	KT694194	KT694195	KT694196	KT694197
Cdc24	KT694198	KT694199	KT694200	KT694201	KT694202
Cdc42	KT694203	KT694204	KT694205	KT694206	KT694207
Fus3	KT694218	KT694219	KT694220	KT694221	KT694222
Gpa1	KT694218	KT694219	KT694220	KT694221	KT694222
Ste3	KT694263	KT694264	KT694265	KT694266	KT694267
Ste4	KT694268	KT694269	KT694270	KT694271	KT694272
Ste7	KT694273	KT694274	KT694275	KT694276	KT694277
Ste11	KT694278	KT694279	KT694280	KT694281	KT694282
Ste12	KT694283	KT694284	KT694285	KT694286	KT694287
Ste18	KT694288	KT694289	KT694290	KT694291	KT694292
Ste20	KT694293	KT694294	KT694295	KT694296	KT694297
RIP genes					
Dim2	KT694208	KT694209	KT694210	KT694211	KT694212
Masc2	KT694233	KT694234	KT694235	KT694236	KT694237

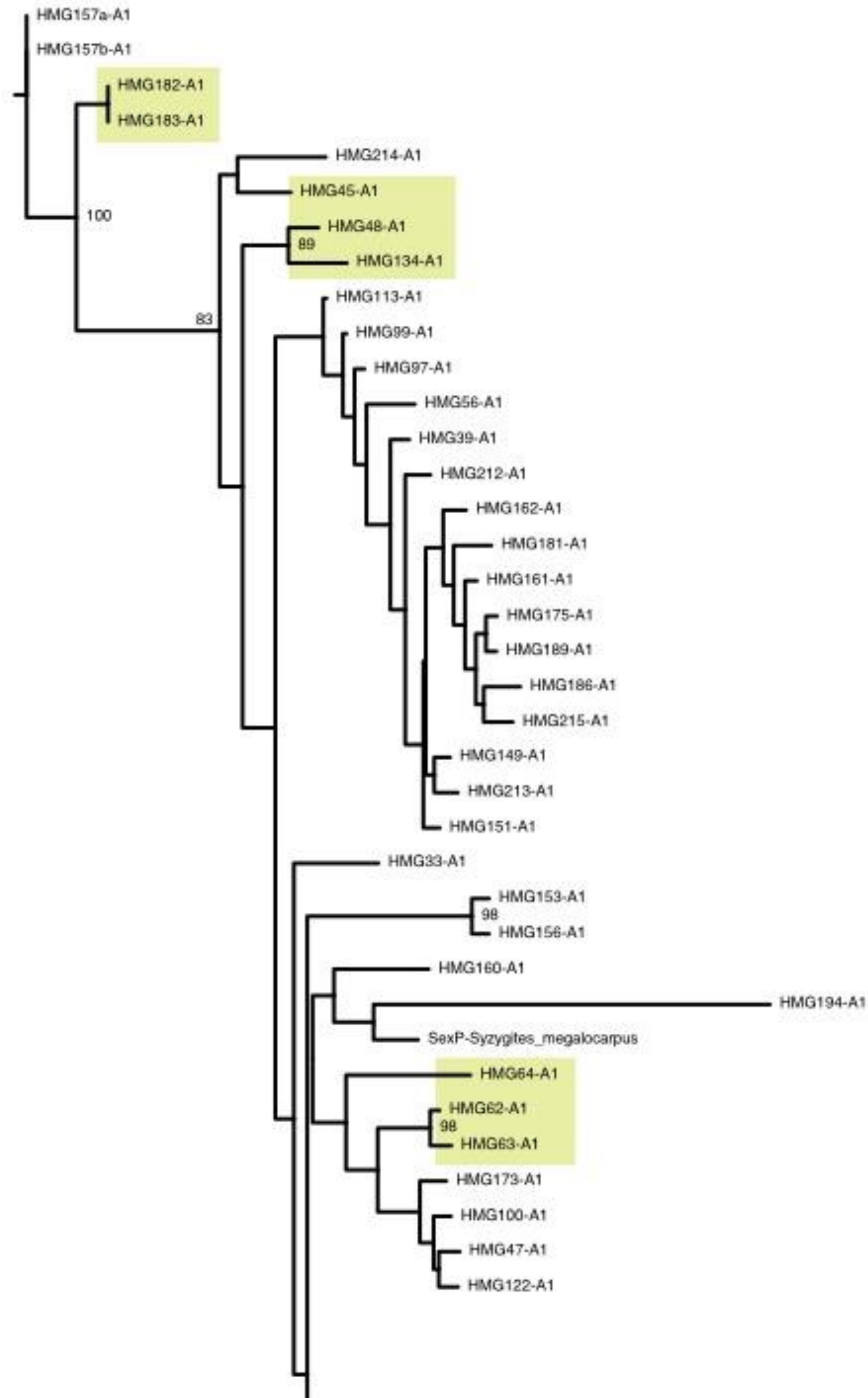


Figure S2.1. Maximum likelihood phylogenetic tree of *MATA*-HMGs amino acid sequences *in R. irregularis* isolate A1. List of *MATA*-HMG representatives used as query for identification were included in tree. Only branch supports over 80 are shown. Coloured boxes represent groups of *MATA*-HMGs that are found on the same scaffold and that are phylogenetically similar.

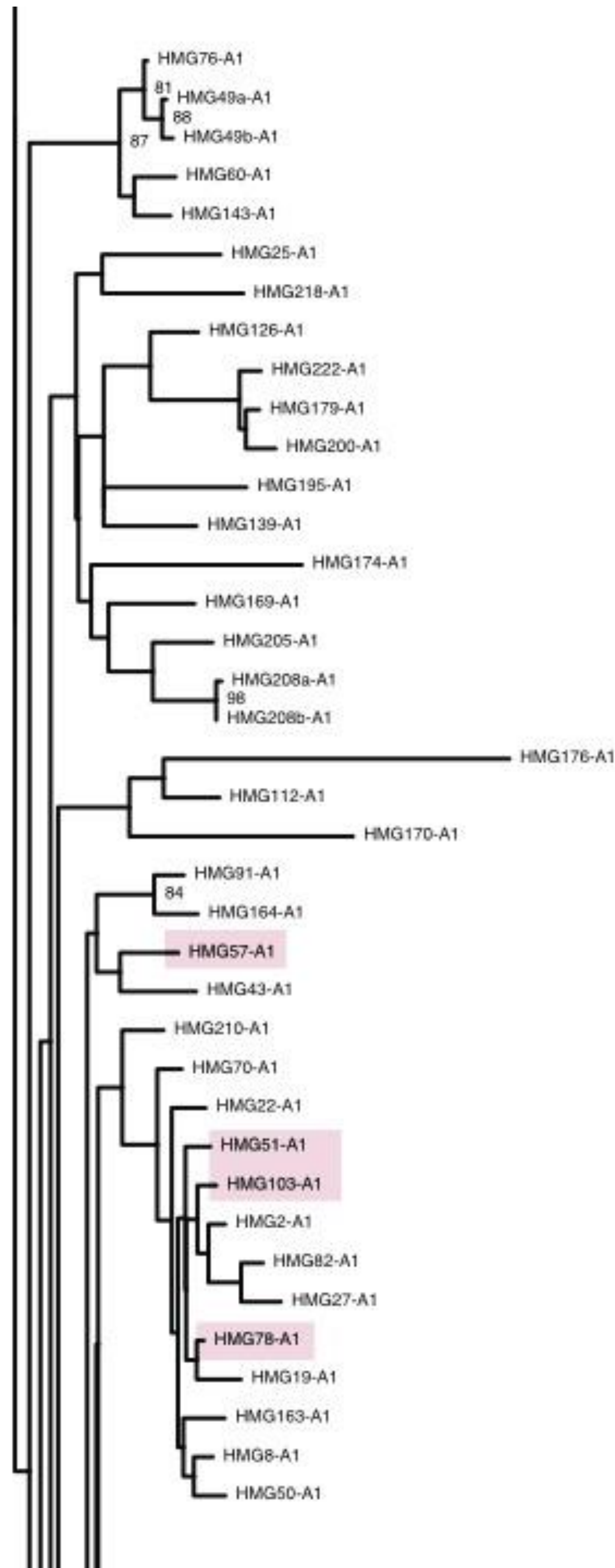


Figure S2.1. (continued)

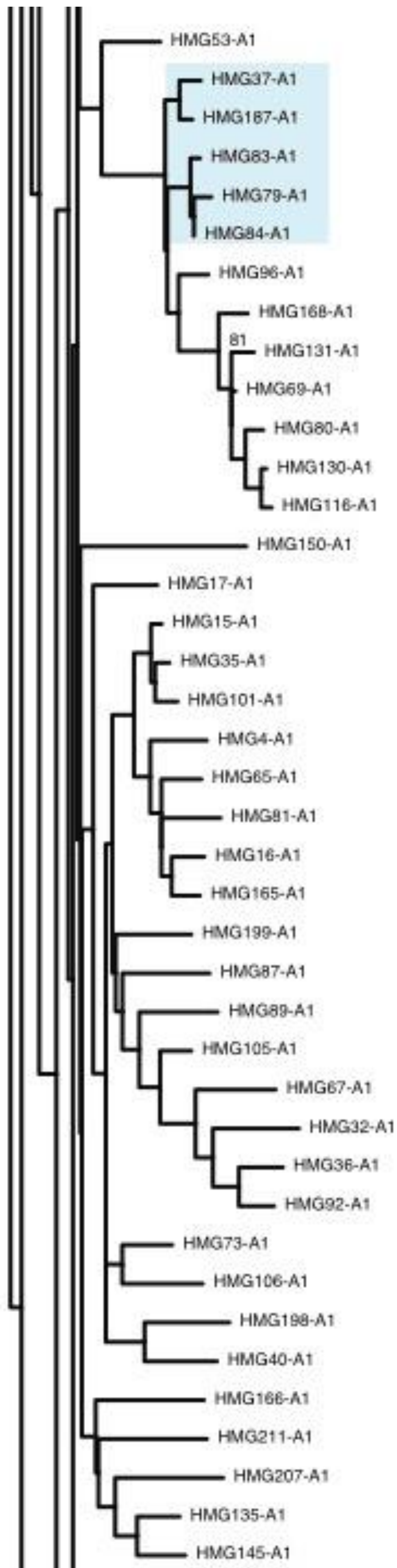


Figure S2.1. (continued)

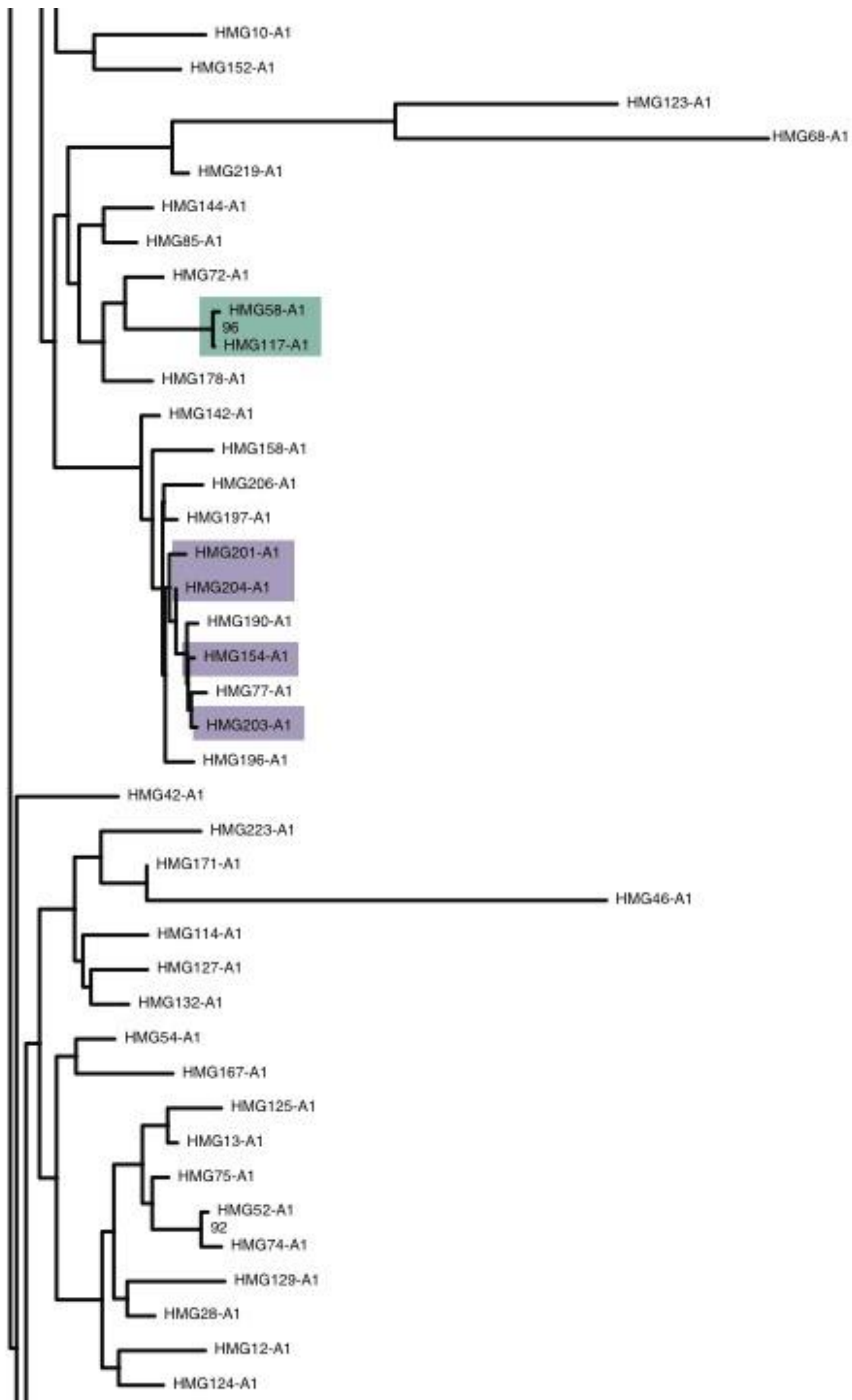


Figure S2.1. (continued)

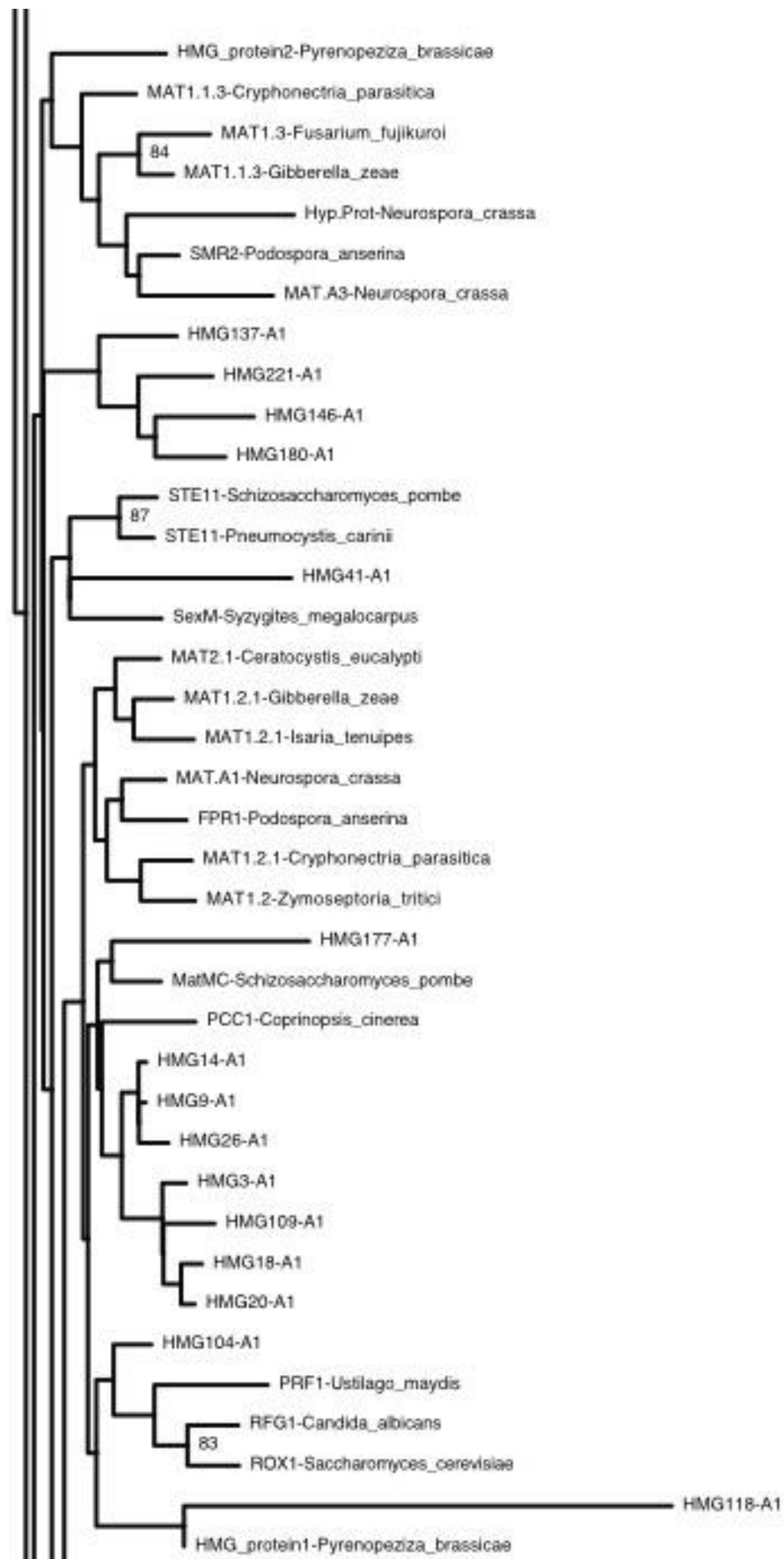


Figure S2.1. (continued)

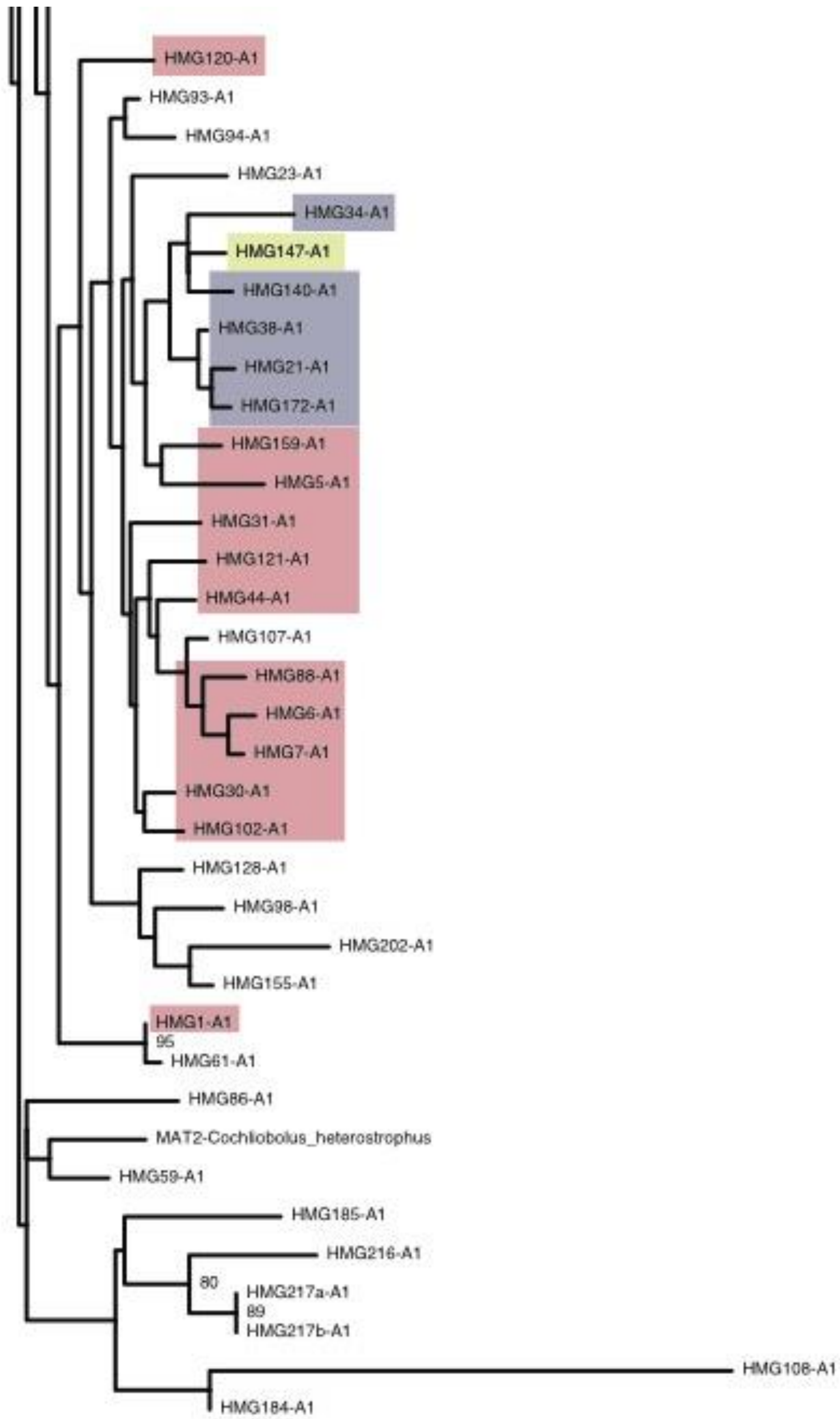


Figure S2.1. (continued)

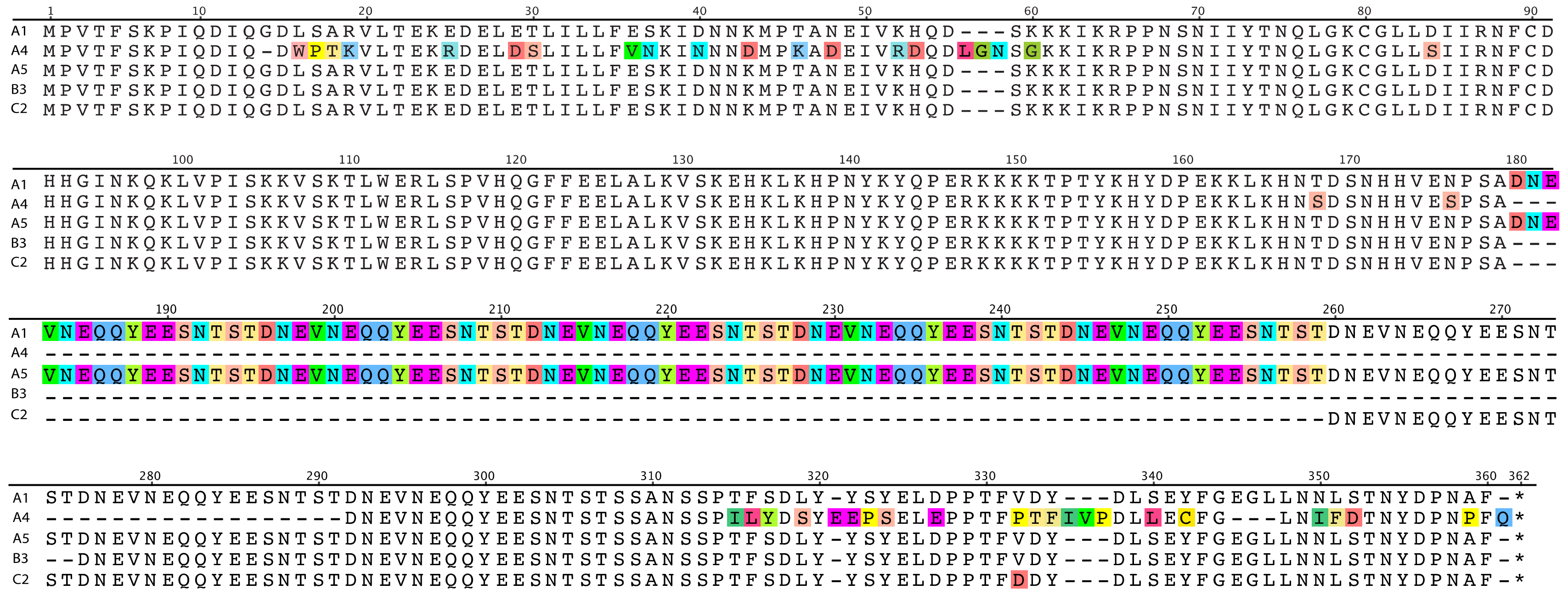


Figure S2.2. Amino acid alignment of HMG122 showing the divergence of sequences among isolates of *R. irregularis*.

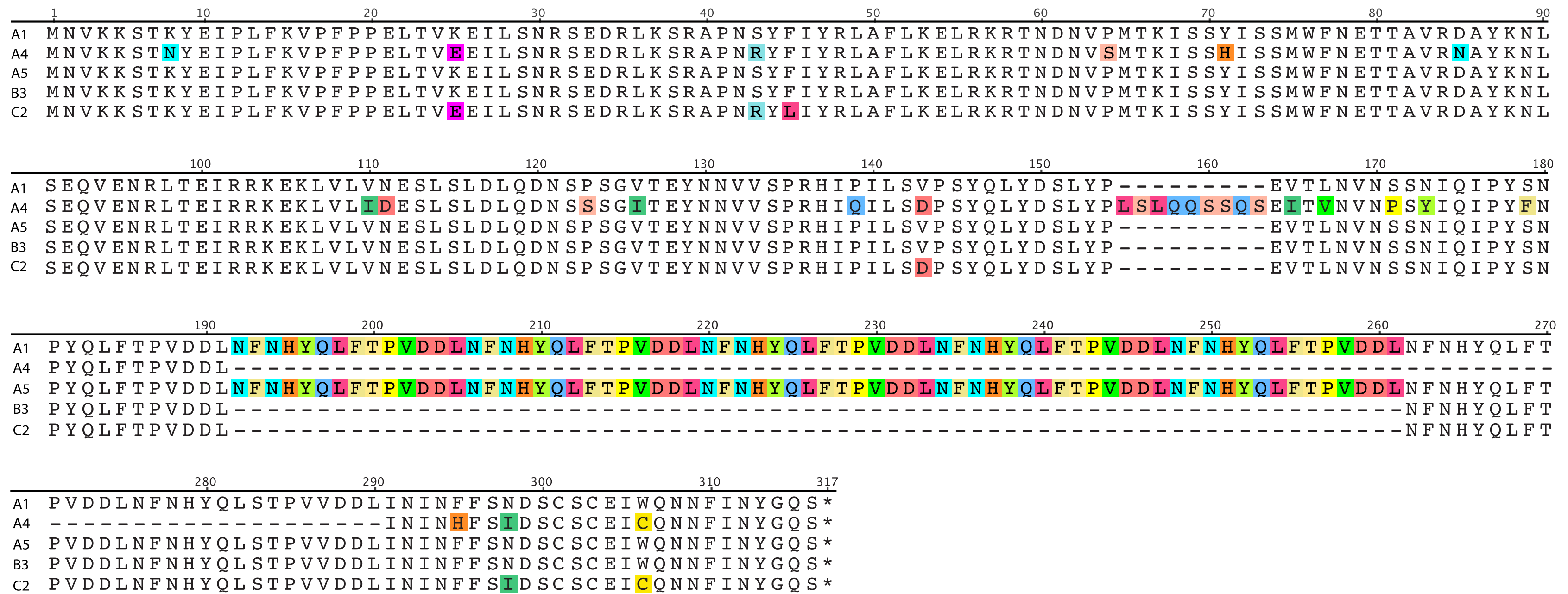


Figure S2.3. Amino acid alignment of HMG130 showing the divergence of sequences among isolates of *R. irregularis.*

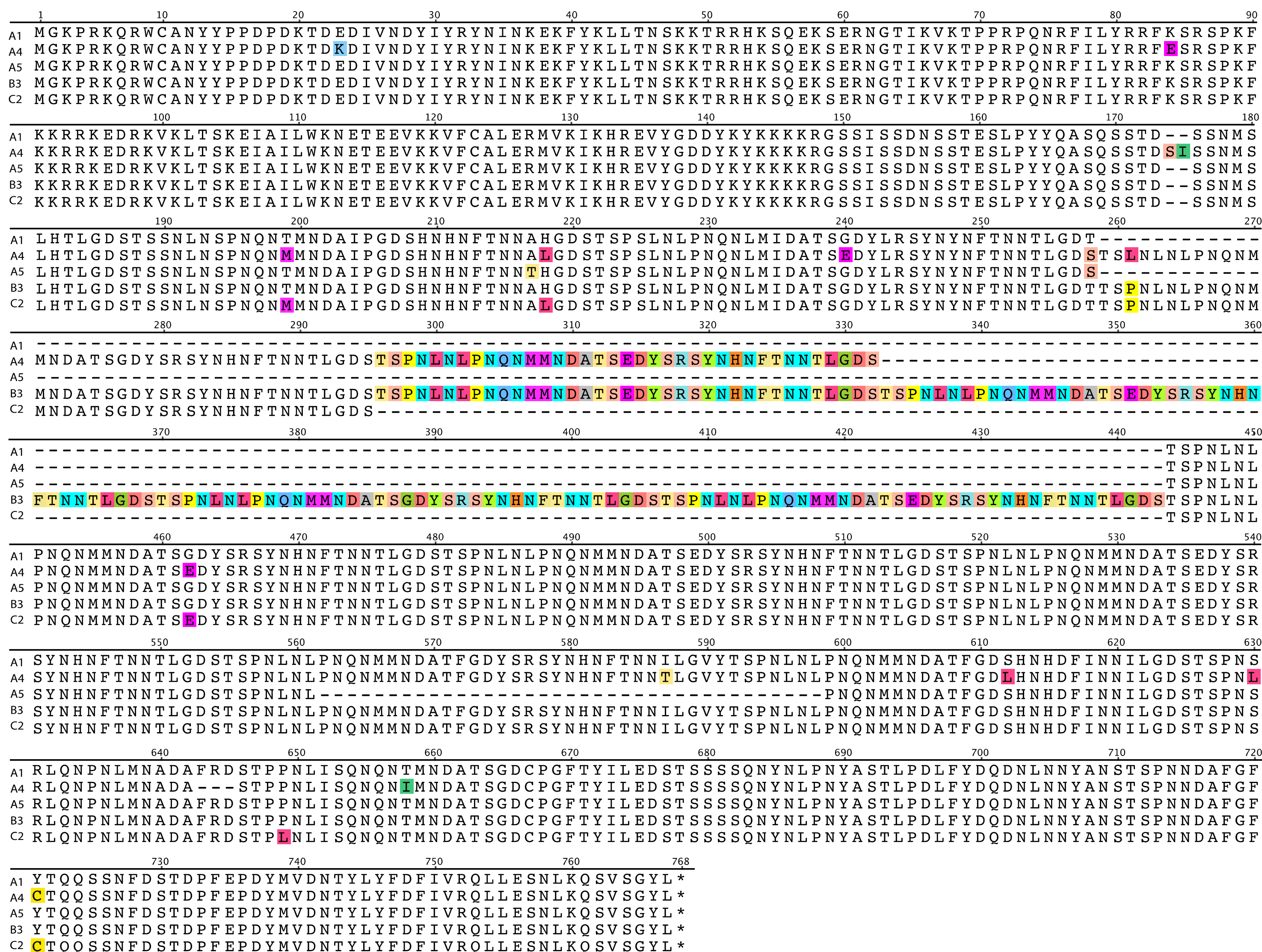


Figure S2.4. Amino acid alignment of HMG21 showing the divergence of sequences among isolates of *R. irregularis*.

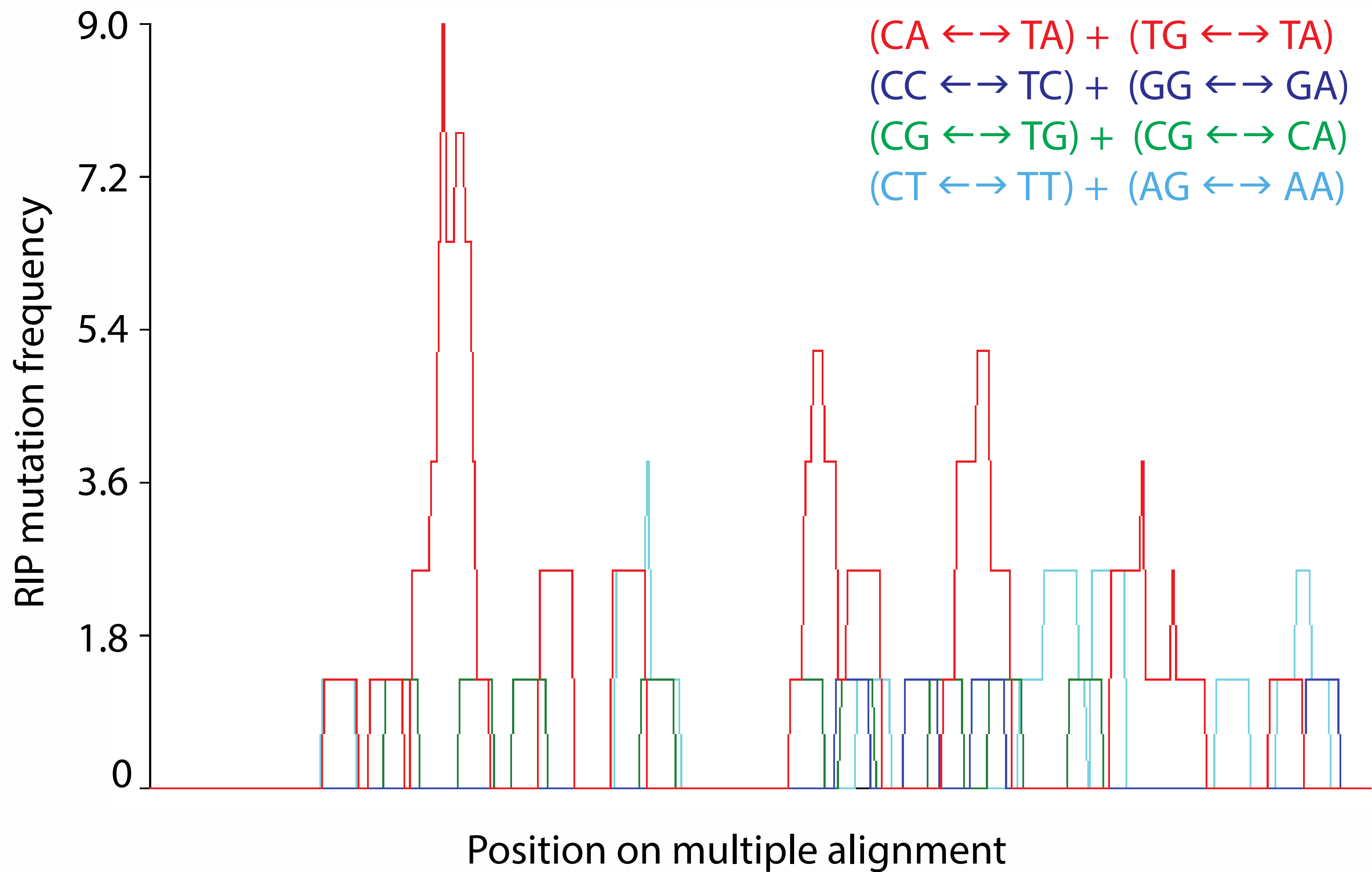


Figure S2.5. RIP mutation frequency (C <--> T) across a multiple alignment of an AMF transposable element.

Chapter 3 – Discussion, Future Directions and Concluding Remarks

Importance of Research

AMF are keystone organisms in the development and growth of land plants. Some key features of the mycorrhizal symbiosis involve increase in plant growth and increase in the plant's tolerance to pathogens and environmental stresses (Smith and Read 2008). For this reason, many agricultural companies have begun selling AMF-inoculated soil. AMF serve as an ecological alternative to harsh fertilizers and have been shown to improve commercial plant crops (Marulanda, Barea, Azcon 2006; Smith and Read 2008; Zhang et al. 2005). However, some complications are linked to mass-producing AMF because only a few species are capable of growing under *in vitro* cultures, a necessary step to produce great quantities of material. This limited spectrum of species could result in an overall loss of diversity and plant growth benefits. By developing breeding programs for these fungi, engineering hybrid species of AMF could create an organism reaching its maximum potential as a fertilizing agent. However, little is known about their basic biology and how these fungi reproduce.

The assumptions of total asexuality are caused by the lack of evidence towards the opposing alternative. Therefore, understanding how these organisms have prospered for such an extended period of time is necessary in order to develop effective breeding programs. My thesis focussed on finding hints of sexuality through the analysis of sexually-related genes.

Summary of findings

In order to fully understand the biological processes undergoing in AMF, the acquisition of better genome assemblies was necessary. During my project, the genomes of five isolates were sequenced and assembled, which favourably compare to previously published versions. Specifically, higher quality read coverage and read length allowed better scaffolding. These genomes revealed a greater expansion of *MATA*-HMG domain containing genes than previously reported (Riley et al. 2014; Tisserant et al. 2013), clearly standing out as being the most abundant ever identified in a fungal lineage. Whether their function is related to mating remains to be seen, but given their inter-isolate conservation and similarity with genes known to be involved in sex in other fungi, one can only theorize about these genes' relationship with reproduction since their link is evident in other fungi. However, no extreme allelic variation has been identified, which is a typical characteristic of idiomorphic alleles responsible for sexual identity.

This gene family is also intriguing from a genomic arrangement point of view with many *MATA*-HMGs being in close proximity to one another. Tandem-repeated *MATA*-HMGs tend to show higher sequence similarity in comparison to other genes in this library. Since rare cases of recombination have been reported in *R. irregularis*, unequal crossovers during recombination events could potentially drive the expansion of these duplicated genomic regions. Additionally, the conservation of syntenic regions with mating-type loci from other fungal phyla could indicate their potential link to sexuality. However, these regions do not comprise the entire *MAT*-locus found in other fungi, raising the question on whether or not these are remnants of sexual loci that have lost

their configuration due to lack of use. We could also hypothesize that the mating-type locus of AMF, if present, could have a unique arrangement of genes surrounding a *MATA*-HMG since these loci differ substantially between different fungal phyla. Also, *MATA*-HMGs may not be responsible for sexual identity in AMF. A situation similar to members of the Basidiomycota phylum where homeodomain transcription factors govern sexuality could be present.

Pheromones are tightly linked with sexual reproduction in other fungi due to their role in mating compatibility. The presence of a theoretically functional pheromone pathway in *R. irregularis*, along with its conservation among all studied isolates, hints towards a mechanism used for "communication" between these fungi. However, whether their role in AMF is directly linked to sexual recognition or has evolved for other uses such as partner compatibility for anastomosis or interacting with the plant host remains a mystery.

In addition to these sexuality-related genes and pathways, other processes have shown to be indirectly linked with sexual reproduction. For instance, repeat-induced point mutations are known to occur during the sexual stage of ascomycetes. Therefore, their presence (or the identification of genes used in the process) has been used as proof of sexuality. Since the hallmark gene for RIP is absent in the *R. irregularis* genome and no apparent evidence of this mechanism is present, it can be assumed that this process is not found in AMF. However, this should not serve as an indication of asexuality since these processes are usually associated with Dikarya fungi, with no cases being found in the

basal fungi phyla Zygomycota (where sexuality has been observed). The abundance of transposable elements and repetitive regions in AMF genomes could be linked to this defence mechanism's absence. RIP-related genes are absent in *Blumeria graminis*, an ascomycete fungus known to have a TE genome content of 76.4% (Amselem, Lebrun, Quesneville 2015). A similar scenario could be found in AMF, where the apparent absence of the RID gene (the only gene known to be necessary for RIP process) has rendered this defense mechanism obsolete, leading to an uncontrolled expansion of TEs and repeated regions.

Future Directions

Fungi tend to resort to asexual reproduction under ideal conditions (Heitman 2006). When evaluating the presence of sexual versus asexual fungi in various levels of stressful environments, an increase in sexual fungi were found as the conditions worsened (Grishkan et al. 2003). Under different conditions, genetic shuffling through recombination may lead to advantageous mutations, increasing the overall fitness of the population, thus giving an advantage to those who reproduce sexually. Therefore, establishing growth conditions that could potentially induce mating is necessary. Identifying these conditions, if they are even present, will prove to be a challenging step since the conditions that AMF may favour for reproduction are completely unknown. These could be linked to mating compatibility of different strains. Our current efforts to identify idiomorphic *MATA*-HMG alleles have not been able to find strains that show substantial variation at typical loci in other fungi, increasing the complexity of establishing whether two strains are compatible due to their genomic composition. By

growing these fungi in pairs, allowing them to anastomose, we could potentially notice distinct morphological features depending on the combinations. Modifications in the media compositions in which the fungi are grown could also trigger sexual reproduction. For instance, modifying the quantity of vitamins or sugar in the media would cause stress on the fungi, forcing them to adapt to their new environment by potentially resorting to sexual processes. Finally, modifying the environment in which they are cultured could have an impact on their life cycle as well. For example, increasing the temperature could activate certain pathways, which could affect how the fungi interact with each other. Monitoring the expression patterns of sexually related genes (*MATA*-HMGs, meiosis genes and pheromone pathways) under these different growth conditions could serve as an essential next step to try and elucidate their roles in AMF's functioning and survival. If these genes are activated under certain conditions, we could potentially pinpoint a moment where some type of cryptic sexuality is happening. However, stressful conditions such as infection by pathogens, gamma radiation, heat shock and oxidative stress have been shown to induce genetic shuffling in other fungi through the increase of transposable element activity (Anaya and Roncero 1996; de Lima Favaro et al. 2005; Gregory et al. 2009). This has been hypothesized to be a method of introducing genetic variation in species that do not conventionally resort to sexuality. Since a large number of transposable elements are found in AMF genomes, the introduction of genetic variability through stressful conditions would have to be heavily scrutinized since these could be linked to transposition rather than sexual processes.

Most members of the Zygomycota and Ascomycota phyla have *MATA*-HMGs as regulators for sexual identity in their MAT-loci (Ni et al. 2011). However, other fungi, notably members of the Basidiomycota phylum, have developed a mating type locus mainly composed of homeodomain transcription factors instead (Metin, Findley, Heitman 2010; Raudaskoski and Kothe 2010). Recent evidence in our lab suggests the presence of homeodomain genes that are organized like those found in the basidiomycete sex locus, hinting towards the presence of a similar mating system. Different alleles are found across some isolates with substantial variation, typical of idiomorphs. Expanding the analyses to a greater amount of strains and AMF species could help understand this intriguing genomic region since this represents our best candidate of a bona fide mating locus thus far.

An important next step in understanding AMF genomics is trying to elucidate the evolutionary trajectory of their coexisting nuclei. Determining whether or not AMF are homokaryotic or heterokaryotic has been an ongoing debate for many years (Young 2015). It has been shown that nuclei can flow freely between anastomosing fungi (Giovannetti, Azzolini, Citernesi 1999). Whether these divergent nuclei are maintained throughout generations still remains a mystery. The creation of “hybrid” strains containing a wide variety of nuclei from different strains through anastomosis could help identify whether the maintenance of heterokaryotic cells are possible. By following the evolutionary trajectory of the nuclei, we could measure different types of selection that are undergoing within cells. Whether or not some nuclei are “lost” due to the presence of mutations or inferiority in fitness could lead us to understand the nature of AMF cells.

Concluding Remarks

Very little is known regarding this fungus' genomic characteristics and hopefully the genomes that were assembled during my project will help reveal key information. Although no clear sexual locus has been identified, our study sheds light on understanding this enigmatic fungus' potential for sex through the identification of common fungal signatures of sex. Whether these genes are activated under specific growth conditions inducing sexuality, or which specific compatible partners are needed in order for them to engage in mating, are some questions that remain to be answered in the future. Developing breeding programs for AMF remains an unfeasible concept due to the absence of an apparent sexuality. However, the findings presented in this thesis hint toward the potential of sexual reproduction, taking us one step closer to achieving this goal. Further analyses will hopefully provide essential information that will help answer the long-lasting question of asexuality in AMF.

References

Abascal F, Zardoya R, Posada D. 2005. ProtTest: Selection of best-fit models of protein evolution. Bioinformatics 21(9):2104-2105.

Altschul S, Gish W, Miller W, Myers E, Lipman D. 1990. Basic local alignment search tool. Journal of Molecular Biology 215(3):403-410.

Amselem J, Lebrun M, Quesneville H. 2015. Whole genome comparative analysis of transposable elements provides new insight into mechanisms of their inactivation in fungal genomes. BMC Genomics 16:141.

Anaya N and Roncero MIG. 1996. Stress-induced rearrangement of *Fusarium* retrotransposon sequences. Molecular & General Genetics 253(1-2):89-94.

Auge RM. 2001. Water relations, drought and vesicular-arbuscular mycorrhizal symbiosis. Mycorrhiza 11(1):3-42.

Bago B, Pfeffer P, Shachar-Hill Y. 2000. Carbon metabolism and transport in arbuscular mycorrhizas. Plant Physiology 124(3):949-57.

Bankevich A, Nurk S, Antipov D, et al. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology 19(5):455-477.

Besserer A, Puech-Pages V, Kiefer P, Gomez-Roldan V, Jauneau A, Roy S, Portais J, Roux C, Becard G, Sejalon-Delmas N. 2006. Strigolactones stimulate arbuscular mycorrhizal fungi by activating mitochondria. PLOS Biology 4(7):1239-47.

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27(4):578-579.

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

Boon E, Halary S, Bapteste E, Hijri M. 2015. Studying genome heterogeneity within the arbuscular mycorrhizal fungal cytoplasm. Genome Biology and Evolution 7(2):505-21.

Brundrett MC. 2002. Coevolution of roots and mycorrhizas of land plants. New Phytologist 154(2):275-304.

Bucher M. 2007. Functional biology of plant phosphate uptake at root and mycorrhiza interfaces. New Phytologist 173(1):11-26.

Burmester A, Richter M, Schultze K, Voelz K, Schachtschabel D, Boland W, Woestemeyer J, Schimek C. 2007. Cleavage of beta-carotene as the first step in sexual hormone synthesis in zygomycetes is mediated by a trisporic acid regulated beta-carotene oxygenase. Fungal Genetics and Biology 44(11):1096-108.

Callac P, Spataro C, Caille A, Imbernon M. 2006. Evidence for outcrossing via the buller phenomenon in a substrate simultaneously inoculated with spores and mycelium of *Agaricus bisporus*. Applied and Environmental Microbiology 72(4):2366-72.

Cambareri E, Jensen B, Schabtach E, Selker E. 1989. Repeat-induced G-C to A-T mutations in *Neurospora*. Science 244(4912):1571-1575.

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25(15):1972-1973.

Cavalier-Smith T. 1998. A revised six-kingdom system of life. Biological Reviews 73:203-266.

Corradi N and Lildhar L. 2012. Meiotic genes in the arbuscular mycorrhizal fungi: What for? Communicative & Integrative Biology 5(2):187-9.

Croll D and Sanders IR. 2009. Recombination in *Glomus intraradices*, a supposed ancient asexual arbuscular mycorrhizal fungus. BMC Evolutionary Biology 9:13.

Croll D, Giovannetti M, Koch AM, Sbrana C, Ehinger M, Lammers PJ, Sanders IR. 2009. Nonself vegetative fusion and genetic exchange in the arbuscular mycorrhizal fungus *Glomus intraradices*. New Phytologist 181(4):924-37.

Croll D, Wille L, Gamper HA, Mathimaran N, Lammers PJ, Corradi N, Sanders IR. 2008. Genetic diversity and host plant preferences revealed by simple sequence repeat and mitochondrial markers in a population of the arbuscular mycorrhizal fungus *Glomus intraradices*. New Phytologist 178(3):672-87.

de Lima Favaro LC, de Araujo WL, de Azevedo JL, Paccola-Meirelles LD. 2005. The biology and potential for genetic research of transposable elements in filamentous fungi. Genetics and Molecular Biology 28(4):804-13.

den Bakker HC, VanKuren NW, Morton JB, Pawlowska TE. 2010. Clonality and

recombination in the life history of an asexual arbuscular mycorrhizal fungus. Molecular Biology and Evolution 27(11):2474-2486.

Desjardins CA, Sanscrainte ND, Goldberg JM, et al. 2015. Contrasting host-pathogen interactions and genome evolution in two generalist and specialist microsporidian pathogens of mosquitoes. Nature Communications 6:7121.

Edgar R. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32(5):1792-1797.

Felsenstein J. 1974. Evolutionary advantage of recombination. Genetics 78(2):737-56.

Fraser DJ. 2008. How well can captive breeding programs conserve biodiversity? A review of salmonids. Evolutionary Applications 1(4):535-86.

Friese CF and Allen MF. 1991. The spread of va mycorrhizal fungal hyphae in the soil - inoculum types and external hyphal architecture. Mycologia 83(4):409-18.

Geiser DM, Pitt JI, Taylor JW. 1998. Cryptic speciation and recombination in the aflatoxin-producing fungus *Aspergillus flavus*. PNAS 95(1):388-93.

Giovannetti M, Azzolini D, Citernesi A. 1999. Anastomosis formation and nuclear and protoplasmic exchange in arbuscular mycorrhizal fungi. Applied and Environmental Microbiology 65(12):5571-5.

Gregory PJ, Johnson SN, Newton AC, Ingram JSI. 2009. Integrating pests and pathogens into the climate change/food security debate. Journal of Experimental Biology 60(10):2827-38.

Grishkan I, Korol AB, Nevo E, Wasser SP. 2003. Ecological stress and sex evolution in soil microfungi. Proceedings of the Royal Society B-Biological Sciences 270(1510):13-8.

Grognet P, Bidard F, Kuchly C, Tong LCH, Coppin E, Benkhali JA, Couloux A, Wincker P, Debuchy R, Silar P. 2014. Maintaining two mating types: Structure of the mating type locus and its role in heterokaryosis in *Podospora anserina*. Genetics 197(1):421-32.

Guindon S, Dufayard J, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Systematic Biology 59(3):307-321.

Haas BJ. 2011. TransposonPSI. Accessed from <http://transposonpsi.sourceforge.net>.

Halary S, Daubois L, Terrat Y, Ellenberger S, Woestemeyer J, Hijri M. 2013. Mating type gene homologues and putative sex pheromone-sensing pathway in arbuscular mycorrhizal fungi, a presumably asexual plant root symbiont. PLOS One 8(11).

Halary S, Malik S, Lildhar L, Slamovits CH, Hijri M, Corradi N. 2011. Conserved meiotic machinery in *Glomus spp.*, a putatively ancient asexual fungal lineage. Genome Biology and Evolution 3:950-8.

Hamilton WD, Axelrod R, Tanese R. 1990. Sexual reproduction as an adaptation to resist parasites (A review). PNAS 87(9):3566-73. 38.

Hane JK, Oliver RP. 2008. RIPCAL: A tool for alignment-based analysis of repeat-induced point mutations in fungal genomic sequences. BMC Bioinformatics 9:478.

Heitman J. 2006. Sexual reproduction and the evolution of microbial pathogens. Current Biology 16(17):R711-25.

Hull CM, Davidson RC, Heitman J. 2002. Cell identity and sexual development in *Cryptococcus neoformans* are controlled by the mating-type-specific homeodomain protein Sxi1alpha. Genes & Development 16(23):3046-60.

Idnurm A. 2011. Sex determination in the first-described sexual fungus. Eukaryotic Cell 10(11):1485-91.

Idnurm A, Walton FJ, Floyd A, Heitman J. 2008. Identification of the sex genes in an early diverged fungus. Nature 451(7175):193-U8.

Irelan J, Selker E. 1996. Gene silencing in filamentous fungi: RIP, MIP and quelling. Journal of Genetics 75(3):313-324.

Jones SK, Jr. and Bennett RJ. 2011. Fungal mating pheromones: Choreographing the dating game. Fungal Genetics and Biology 48(7):668-76.

Judson O and Normark B. 1996. Ancient asexual scandals. Trends in Ecology & Evolution 11(2):A41-6.

Jurkowski TP, Jeltsch A. 2011. On the evolutionary origin of eukaryotic DNA methyltransferases and Dnmt2. PLOS One 6(11):e28104.

Koch AM, Croll D, Sanders IR. 2006. Genetic variability in a population of arbuscular mycorrhizal fungi causes variation in plant growth. Ecology Letters 9(2):103-10.

Kondrashov AS. 1988. Deleterious mutations and the evolution of sexual reproduction. Nature 336(6198):435-40.

Kuhn G, Hijri M, Sanders IR. 2001. Evidence for the evolution of multiple genomes in arbuscular mycorrhizal fungi. Nature 414(6865):745-8.

Lanfranco L, Delpero M, Bonfante P. 1999. Intrasporal variability of ribosomal sequences in the endomycorrhizal fungus *Gigaspora margarita*. Molecular Ecology 8(1):37-45.

Lee SC, Ni M, Li W, Shertz C, Heitman J. 2010. The evolution of sex: A perspective from the fungal kingdom. Microbiology and Molecular Biology Reviews 74(2):298,+.

Lee SC, Corradi N, Doan S, Dietrich FS, Keeling PJ, Heitman J. 2010. Evolution of the sex-related locus and genomic features shared in microsporidia and fungi. PLOS One 5(5):e10539.

Li CH, Cervantes M, Springer DJ, Boekhout T, Ruiz-Vazquez RM, Torres-Martinez SR, Heitman J, Lee SC. 2011. Sporangiospore size dimorphism is linked to virulence of *Mucor circinelloides*. PLOS Pathogens 7(6):e1002086.

Lin K, Limpens E, Zhang Z, Ivanov S, Saunders DGO, Mu D, Pang E, Cao H, Cha H, Lin T, et al. 2014. Single nucleus genome sequencing reveals high similarity among nuclei of an endomycorrhizal fungus. PLOS Genetics 10(1):e1004078.

LloydMacgilp SA, Chambers SM, Dodd JC, Fitter AH, Walker C, Young JPW. 1996. Diversity of the ribosomal internal transcribed spacers within and among isolates of *Glomus mosseae* and related mycorrhizal fungi. New Phytologist 133(1):103-11.

Lucas JA, Dyer PS, Murray TD. 2000. Pathogenicity, host-specificity, and population biology of *Tapesia spp.*, causal agents of eyespot disease of cereals. Advances in Botanical Research Incorporating Advances in Plant Pathology 33(33):225-58.

Marleau J, Dalpe Y, St-Arnaud M, Hijri M. 2011. Spore development and nuclear inheritance in arbuscular mycorrhizal fungi. BMC Evolutionary Biology 11:51.

Martin F, Gianinazzi-Pearson V, Hijri M, Lammers P, Requena N, Sanders IR, Shachar-Hill Y, Shapiro H, Tuskan GA, Young JPW. 2008. The long hard road to a completed *Glomus intraradices* genome. New Phytologist 180(4):747-50.

Marulanda A, Barea JM, Azcon R. 2006. An indigenous drought-tolerant strain of *Glomus intraradices* associated with a native bacterium improves water transport and root development in *Retama sphaerocarpa*. Microbial Ecology 52(4):670-8.

Merino ST, Nelson MA, Jacobson DJ, Natvig DO. 1996. Pseudohomothallism and evolution of the mating-type chromosome in *Neurospora tetrasperma*. Genetics 143(2):789-99.

Metin B, Findley K, Heitman J. 2010. The mating type locus (MAT) and sexual reproduction of *Cryptococcus heveanensis*: Insights into the evolution of sex and sex-determining chromosomal regions in fungi. PLOS Genetics 6(5).

Ni M, Feretzaki M, Sun S, Wang X, Heitman J. 2011. Sex in fungi. Annual Review Genetics, Vol 45 45:405-30.

Normark BB, Judson OP, Moran NA. 2003. Genomic signatures of ancient asexual lineages. Biological Journal of the Linnean Society 79(1):69-84.

Paoletti M, Rydholm C, Schwier EU, Anderson MJ, Szakacs G, Lutzoni F, Debeaupuis JP, Latge JP, Denning DW, Dyer PS. 2005. Evidence for sexuality in the opportunistic fungal pathogen *Aspergillus fumigatus*. Current Biology 15(13):1242-8.

Parniske M. 2008. Arbuscular mycorrhiza: The mother of plant root endosymbioses. Nature Reviews Microbiology 6(10):763-75.

Pelin A, Pombert JF, Salvioli A, Bonen L, Bonfante P, Corradi N. 2012. The mitochondrial genome of the arbuscular mycorrhizal fungus *Gigaspora margarita* reveals two unsuspected trans-splicing events of group I introns. New Phytologist 194(3):836-45.

Radovanovic N and Cloutier S. 2003. Gene-assisted selection for high molecular weight glutenin subunits in wheat doubled haploid breeding programs. Molecular Breeding 12(1):51-9.

Raudaskoski M and Kothe E. 2010. Basidiomycete mating type genes and pheromone signaling. Eukaryotic Cell 9(6):847-59.

Redecker D, Kodner R, Graham LE. 2000. Glomalean fungi from the ordovician. Science 289(5486):1920-1.

Riley R and Corradi N. 2013. Searching for clues of sexual reproduction in the genomes of arbuscular mycorrhizal fungi. Fungal Ecology 6(1):44-9.

Riley R, Charron P, Idnurm A, Farinelli L, Dalpe Y, Martin F, Corradi N. 2014. Extreme diversification of the mating type-high-mobility group (MATA-HMG) gene family in a plant-associated arbuscular mycorrhizal fungus. New Phytologist 201(1):254-68.

Rillig MC. 2004. Arbuscular mycorrhizae and terrestrial ecosystem processes. Ecology Letters 7(8):740-54.

Ropars J and Corradi N. 2015. Homokaryotic vs heterokaryotic mycelium in arbuscular mycorrhizal fungi: Different techniques, different results? New Phytologist 208(3):638-41.

Sanders IR. 2003. Preference, specificity and cheating in the arbuscular mycorrhizal symbiosis. Trends in Plant Science 8(4):143-5.

Sanders IR. 1999. Evolutionary genetics - no sex please, we're fungi. Nature 399(6738):737-9.

Sanders IR, Clapp JP, Wiemken A. 1996. The genetic diversity of arbuscular mycorrhizal fungi in natural ecosystems - A key to understanding the ecology and functioning of the mycorrhizal symbiosis. New Phytologist 133(1):123-34.

Schurko AM, Neiman M, Logsdon JM, Jr. 2009. Signs of sex: What we know and how we know it. Trends in Ecology & Evolution 24(4):208-17.

Schussler A, Schwarzott D, Walker C. 2001. A new fungal phylum, the glomeromycota: Phylogeny and evolution. Mycological Research 105:1413-21.

Smith SE and Read DJ. 2008. *Mycorrhizal Symbiosis (Third Edition)*, Academic Press, London, ISBN 9780123705266.

Taylor S, Wakem M, Dijkman G, Alsarraj M, Nguyen M. 2010. A practical approach to RT-qPCR-publishing data that conform to the MIQE guidelines. Methods 50(4):S1-S5.

Tisserant E, Malbreil M, Kuo A, Kohler A, Symeonidi A, Balestrini R, Charron P, Duensing N, Frey NFD, Gianinazzi-Pearson V, et al. 2013. Genome of an arbuscular

mycorrhizal fungus provides insight into the oldest plant symbiosis. PNAS 110(50):20117-22.

Tzung KW, Williams RM, Scherer S, Federspiel N, Jones T, Hansen N, Bivolarevic V, Huizar L, Komp C, Surzycki R, et al. 2001. Genomic evidence for a complete sexual cycle in *Candida albicans*. PNAS 98(6):3249-53.

Vandenkoornhuyse P, Leyval C, Bonnin I. 2001. High genetic diversity in arbuscular mycorrhizal fungi: Evidence for recombination events. Heredity 87:243-253.

Young JPW. 2015. Genome diversity in arbuscular mycorrhizal fungi. Current Opinions in Plant Biology 26:113-9.

Zhang XH, Zhu YG, Chen BD, Lin AJ, Smith SE, Smith FA. 2005. Arbuscular mycorrhizal fungi contribute to resistance of upland rice to combined metal contamination of soil. Journal of Plant Nutrition 28(12):2065-77.