UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TECNICA SUPERIOR DE INGENIERÍA
AGRONÓMICA, ALIMENTARIA Y DE BIOSISTEMAS

*Bioinformatic approaches to understand RNA biology in
the rice blast fungus*

TESIS DOCTORAL

Marco Marconi

Licenciado en Informática

Madrid, 2017

**Programa Oficial de Doctorado en Biotecnología y
Recursos Genéticos de Plantas y Microorganismos
Asociados**

DEPARTAMENTO DE BIOTECNOLOGÍA

UNIVERSIDAD POLITÉCNICA DE MADRID

TESIS DOCTORAL

*Bioinformatic approaches to understand RNA biology in the
rice blast fungus*

Autor:

Marco Marconi
Licenciado en Informática

Director:

Mark D. Wilkinson
B.Sc. Honors Genetics
Ph.D. Botany (Plant Molecular Biology)

Director:

Ane Sesma Galarraga
B.Sc. Chemistry, major in Biochemistry
& Molecular Biology
Ph.D. Agronomic Engineering (Crop
Protection)

Madrid, Junio de 2017

UNIVERSIDAD POLITÉCNICA DE MADRID

Tribunal nombrado por el Magfco. y Excmo. Sr. Rector de la
Universidad Politécnica de Madrid.

Presidente: Fernando Garcia-Arenal Rodriguez

Secretario: María Soledad Sacristán Benayas

Vocal: Victoriano Garre Mula

Vocal: Marcos Egea Gutiérrez-Cortines

Vocal internacional: David J. Studholme

Suplente: Isabel Cuesta de la Plaza

Suplente internacional: Darren Mark Soanes

Contents

Acknowledgements.....	ix
List of abbreviations	xi
Summary.....	xiii
Resumen	xv
1. GENERAL INTRODUCTION	1
OBJECTIVES.....	15
2. MATERIALS AND METHODS.....	17
3. RESULTS.....	25
Chapter 3.1 - Links between pre-mRNA 3'-end processing, nutritional stress response and plant pathogenesis in the rice blast fungus	27
Chapter 3.2 - small RNA sequencing provides new insights into RNA interference in the rice blast fungus <i>Magnaporthe oryzae</i>	53
Chapter 3.3 - Molecular characterization of a novel ssRNA ourmia-like virus from the rice blast fungus <i>Magnaporthe oryzae</i>	73
Chapter 3.4 - Large scale comparative analyses of RNA-related machineries in the fungal kingdom	83
4. GENERAL DISCUSSION.....	111
CONCLUSIONS.....	115
REFERENCES.....	117
ANNEXES	131
Annex Table 1.....	131
Annex Figure 1.....	168
Annex Figure 2.....	168

Acknowledgements

I would like to thank both Mark and Ane for the opportunity they gave me, doing my Ph.D. in the CBGP was a great experience. Their door was always open whenever I needed, and they consistently supervised me and suggested me.

I would like to express my deepest gratitude to my advisor, prof. Wilkinson, for accepting me into his group. During my tenure, he contributed to my personal skills by giving me intellectual freedom in my work, supporting my attendance to international conferences and scholarships, engaging me in new ideas, and demanding a high quality of work in all my endeavors.

I also thank prof. Sesma, for the patience she demonstrated to a non-biologist like me, and the enormous amount of knowledge that I was able to acquire through the continuous interaction with the “wet” lab.

I would also like to thank my lab colleagues: to Alejandro junior, who stayed years next to me and contributed to many details of this thesis, and of course for all the comics/gadgets he brought into the lab; to Alejandro senior, for bringing experience into the lab as a post-doc and for the discussions about politics; to Bea, for the wise advices she always had in every situation; to Mikel, that even for a short period he had been a great Basque in the lab; to Adrián, for all the important information about Spanish politics he shared and all the food leftovers the he also shared, and to Mario, for bringing a southern/African contribution to the lab and for the excellent group selfies.

Big thanks to Julio and Adriana, which have been the other side of the coin during my Ph.D. They did all the hard stuff in the lab struggling with *Magnaporthe* and other dreadful organisms allowing me to deal only with clean, aseptic data on my computer screen. The lab meetings have been incredibly interesting and useful, even when I practically understood nothing. I am much obliged also to the late members, Victor, Marta, Marie and Cristina. I witnessed the growth of the Sesma lab and I am sure it will achieve big things in the future.

A special thanks to Emilia, Pablo, Isa, Mariela, Miguel, Pedro, Saray, Chechu, Sandra, Martí and Paul for being our twin laboratory, nobody else would have accepted us, the computer-people, as real scientists. Many thanks to all CBGP staff, researches, technicians, the other Ph.D. students, the junior seminary committee (Amalia and Ana), the directors, the cleaning ladies and the coffee machine.

Finally, thanks to my family and friends, for supporting me and helping me in any moment and difficulty. Without them, nothing of this would have been possible.

List of abbreviations

aRNA	aberrant RNA
ATP	Adenosine Tri-Phosphate
APA	Alternative polyadenylation
Ago	Argonaute protein
PA	A-rich Position Element
APT	Associated with Pta1
EE	AU-rich Efficiency Element
CAGE	cap analysis of gene expression
CR	Chromatin Remodeling
CFIA	Cleavage Factor IA
CFIB	Cleavage Factor IB
CFII	Cleavage Factor II
CM	Complete Medium
CUT	cryptic unstable ncRNAs
CTD	C-Terminal Domain
DNA	Deoxyribonucleic acid
disiRNA	Dicer-independent siRNA
DPS	Differential Polyadenylation site Selection
DUE	Downstream U-rich Element
EXP5	Exportin-5
EST	Expressed Sequence Tag
GTF	General Transcription Factors
HRP1	Heteronuclear RNA binding protein 1
60S	Large ribosomal subunit
MOLV1	Magnaporthe oryzae Ourmia-like Virus 1
MPSS	massively parallel signature sequencing
Med	Mediator complex
MSUD	meiotic silencing by unpaired DNA
mRNA	Messenger RNA
miRNA	Micro RNA
MM	Minimal Medium
MM-C	Minimal Medium –Carbon
MM-N	Minimal Medium –Nitrogen
NCBI	National Center for Biotechnology Information
NGS	Next-Generation Sequencing
NTC	NineTeen Complex
No-go	No-go decay
ncRNA	non-coding RNA
NMD	Nonsense mediated decay

Non-stop	Non-stop decay
NNS	Nrd1–Nab3–Sen1 complex.
NPC	Nuclear Pore Complex
ORF	Open Reading Frame
PAP	Poly(A) Polymerase
PFI	Polyadenylation Factor I
PAS	Polyadenylation site
PCR	Polymerase Chain-Reaction
PCA	Principal Component Analysis
RIP	repeat-induced point mutation
RNA	Ribonucleic acid
rDNA	Ribosomal DNA
RNAi	RNA Interference
RBP35	RNA-Binding Protein 35
RdRp	RNA-dependent RNA polymerase
RISC	RNA-induced silencing complex
SAGE	Serial Analysis of Gene Expression
ssRNA	single-strand RNA
siRNA	small interfering RNA
snRNP	small nuclear ribonucleoprotein
snRNA	Small Nuclear RNA
snoRNA	Small Nucleolar RNA
40S	Small ribosomal subunit
sRNA	Small RNA
STF	Specific Transcription Factors
TF	Transcription Factor
TOR	Target Of Rapamycin
THO	THO complex
TSS	Transcription Start Site
tRNA	transfer RNA
TE	Transposable Element
TREX	TREX complex
TRAMP	Trf4/Air2/Mtr4p Polyadenylation complex
UTR	UnTranslated Region
UAS	Up-stream Activating Sequence
UUE	Upstream U-rich Element
WT	Wild-Type

Summary

The purpose of this thesis is to investigate various aspects of RNA metabolism in the rice blast fungus *Magnaporthe oryzae* using bioinformatic tools and approaches, and to attempt to extrapolate key features of the RNA life cycle over the whole fungal kingdom.

In the first results chapter we conducted a bioinformatic analysis of a poly(A)-specific RNA-sequencing dataset obtained from a *M. oryzae* wild-type strain and knock-out mutants of two genes involved in the polyadenylation machinery, *RBP35* and *HRP1*. This represents the first mapping of polyadenylation sites in the rice blast fungus, and it provided us with several insights regarding regulation of alternative polyadenylation. Alternative polyadenylation is a common feature in the *M. oryzae* transcriptome; it is a dynamic mechanism regulated by nutrient stresses such as carbon starvation, which provoke the selection of non-canonical poly(A) sites. Mutant analysis also indicate that Rbp35 and Hrp1 are involved in poly(A) site selection in a large number of genes.

In the second results chapter we performed the small RNA sequencing of *M. oryzae*, using a novel library construction protocol. Genomic repetitive sequences such as retrotransposons and non-genomic invasive nucleic acids such as viruses display clear evidence of an RNA silencing mechanism acting in the *M. oryzae* cell. These same elements are also subject to an extensive post-transcriptional modification not observed in prior studies. We also provide confirmation that *RBP35* and *EXP5* (a karyopherin) genes are involved in small RNA regulation and possibly in some RNA silencing pathway.

In the third results chapter we accomplish the molecular characterization of a novel ssRNA ourmia-like virus from *M. oryzae*, called MOLV1. This virus is specific to the strain Guy11. It is possibly polyadenylated, and contains a unique ORF coding for a putative RNA-dependent RNA polymerase.

In the fourth chapter we carried out the proteome-wide ortholog prediction for 734 RNA-related proteins, in 49 fungal species. The objective of this analysis was to provide a broader understanding of the conservation of RNA-associated processes in fungal species beyond *M. oryzae*. This evolutionary survey revealed an unexpected scenario; processes like mRNA translation and tRNA maturation are generally well conserved, while other processes like mRNA export and RNA silencing exhibit a variably degree of conservation among the selected fungal species. Significantly, core proteins previously described of fundamental importance can be disposed in some branches of the fungal tree.

Resumen

El propósito de esta tesis consiste en investigar algunos aspectos del metabolismo del ARN en el hongo *Magnaporthe oryzae* desde un punto de vista bioinformático, y extrapolar las características claves del ciclo de vida del ARN a todo el reino de los hongos.

En el primer capítulo de los resultados hemos conducido el análisis bioinformático de datos de secuenciación masiva de ARN específicos para poly(A) obtenido de la cepa wild-type de *M. oryzae* y dos mutantes knock-out de dos genes involucrados en la maquinaria de poliadenilación, *RBP35* y *HRP1*. Este ha sido el primer mapeo de los sitios de poliadenilación en *M. oryzae*, y nos ha proporcionado muchas nuevas informaciones relativamente al estudio de la poliadenilación alternativa. La poliadenilación alternativa es muy frecuente en *M. oryzae*, y es un mecanismo dinámico regulado por condiciones de estrés nutricional como la falta de carbón. El análisis de los mutantes también nos indica que *RBP35* y *HRP1* are profundamente involucrados en la poliadenilación alternativa de un gran número de genes.

En el segundo capítulo de los resultados hemos realizado la secuenciación masiva de ARN pequeños en *M. oryzae* utilizando un nuevo sistema de construcción de librerías. Resulta claro que algún sistema de silenciamiento de ARN está activo en contra de secuencias genómicas repetidas como los retrotransposones y ácidos nucleicos non-genómicos invasivo como los virus. Estos elementos están también sujetos a una modificación post-transcripcional muy extensa nunca observada antes. También hemos confirmado que los genes *RBP35* y *EXP5* (una carioferina) están implicadas en la regulación de ARN pequeños y posiblemente en alguna ruta de silenciamiento de ARN.

En el tercer capítulo de los resultados hemos conseguido la caracterización de un nuevo virus ssRNA de tipo ourmia de *M. oryzae*, llamado MOLV1. Este virus es específico de la cepa Guy11. Es un virus poliadenilado, y contiene una única ORF que codifica una polimerasa de RNA dependiente de RNA putativa.

En el cuarto capítulo de los resultados hemos llevado a cabo la predicción de ortólogos para alrededor de 700 proteínas relacionadas con ARN, en 49 especies. La idea detrás de este análisis consiste en proporcionar un conocimiento exhaustivo relativamente la conservación de procesos asociados a ARN en especies diferentes de *M. oryzae*. Esta investigación ha desvelado un escenario inesperado; procesos como traducción de ARN y maduración de ARNt están en general bien conservados, mientras otros procesos como la exportación de ARNm y en silenciamiento de ARN exhiben un nivel de conservación variable entre las especies fúngicas consideradas. Es preciso destacar que algunas de las proteínas anteriormente descritas como de importancia fundamental pueden ser descartadas en algunas ramas del árbol de los hongos.

1. General Introduction

The rice blast fungus

Magnaporthe oryzae (also known as the rice blast fungus) is an ascomycete fungus that causes disease in more than 40 species of Gramineae including a range of economically relevant crops like rice, barley wheat, finger and foxtail millet, and wild grasses [1]. *M. oryzae* is considered the most important pathogen of rice, and one of the top fungal plant pathogens [2]. Rice is the staple food for over half of the world population [3], which makes rice blast one of the worst threats to global food security. Blast was first described in China in 1637, and since then it has spread worldwide. Today, it exists in some 85 countries, causing common epidemics in China, Korea, Japan, Vietnam and the USA. Recent estimates suggest that losses of crops due to rice blast could feed 60 million people every year. *M. oryzae* is also an important wheat pathogen, and wheat blast is rapidly spreading in Bangladesh, India and South America, threatening to reach Europe [4].

The disease is favored by high humidity conditions and can cause losses of 10-30% of the grain yield per year. Long periods of rain and low solar radiation also favor disease establishment and spread [5]. The spores can survive the winter in rice grains and plant residues, therefore restarting the infection cycle the following year [6]. Fungal spores are spread by wind and water splashes causing the three-celled conidium to land on the leaf surface [5]. The spore apex attaches on the hydrophobic leaf surface and is blocked by the addition of an adhesive substance [7]. Almost immediately, the spore produces a narrow germ tube and attaches to the hydrophobic cuticle, before differentiating into the appressorium [6]. The appressorium is a specialized highly melanized cell which exploits turgor pressure generated by glycerol accumulation to penetrate the cell wall with a tiny infection peg [8]. After 4-8 hours, one of the nuclei of the three-celled conidium migrates into the germ tube and undergoes mitosis, next the two daughter cells separate, one reverts back to the conidium (which will then degrade through an autophagic process) while the other enters the appressorium [9–11]. After the penetration peg has crossed the cell cuticle, the fungus grows within the plant cell without getting in direct contact with the host cell cytoplasm due to the formation of a plant-derived membrane surrounding the fungal hyphae. Invasive hyphae can propagate to other plant cells through plasmodesmata [12].

Figure 1.1 | Infection cycle of *M. oryzae*. The infection starts with a spore landing on the leaf surface. Then, a germ tube emerges from the spore and matures into an appressorium, which in turn penetrates the leaf cuticle using turgor pressure by means of a penetration peg. A biotrophic cell invasion follows, and infecting hyphae propagate to adjacent cells through plasmodesmata. After five to six days, conidiospores appearing on the surface sporulate and cycle starts over again. In roots, colonization is thought to occur because of resting structures present on decaying plant matter. Root colonization takes place in a different manner than leaves, followed by systemic invasion and the development of typical disease symptoms on the aerial parts of the plant.

The fungus proliferation into the plant cells is initially biotrophic and symptomless and disease lesion become visible only 72-96 hours after infection and can occur on leaves, leaf collars, panicles, culms and culm nodes [12]. The fungus is able to evade plant immune responses as other plant pathogens, probably with the assistance of a set of fungal effectors [13,14]. Sporulation occurs under humid conditions, producing up to a thousand spores per night [15]. Rice blast infection of young rice seedlings is usually lethal to the plant, whereas lesions on stems, nodes or panicle of older plants results in nearly total loss of the rice grain, and sometimes the whole plant collapses [6]. Similar to other fungal pathogens *M. oryzae* reproduction is mainly asexual in all rice-growing regions, except for a small Asian population [16]. Sexual reproduction in contemporary *M. oryzae* population pathogenic to rice has been recently demonstrated in Southeast Asia, where *M. oryzae* is thought to have originated, and that sexual reproduction has been lost outside its center of origin [17].

The genome of *M. oryzae* strain 70-15 was sequenced in 2005. Its current assembly has a size of 41.03 Mb and encodes 12,593 protein-coding genes [18]. It contains 30 enzymes for

cellulose degradation and 44 enzymes for hemicellulose degradation of plant cell walls, including unusual carbohydrate-binding domains and novel G protein-coupled receptors. Approximately 9.7% of the *M. oryzae* 70-15 genome assembly comprises repetitive DNA sequences, but these numbers vary greatly among different isolates [19]. It has been suggested that transposable elements play an important part in host adaptation [1].

Several experiments using Next Generation Sequencing (NGS) have been performed on *M. oryzae* including transcriptome analysis during appressorium and mycelium formation [20], small RNA sequencing [21,22], whole-genome DNA methylation [23] and dual RNA-seq during plant infection [24]. Different *M. oryzae* isolates have been found to be infected by a variety of mycoviruses [25–27].

Evolution in the fungal kingdom

Together with the Animalia and the Viridiplantae clades, fungi seem to have originated from simple flagellated aquatic forms, diverging from other life forms around 1500 million years ago [28,29]. The earliest fossil records displaying fungal-like features belong to the benthic genus *Tappania*, found in sediments up to 1,630 million years old [30]. Reconstructing the phylogenetic tree of fungi has proven to be a hard task. “Higher” fungi like Ascomycota and Basidiomycota present a clear separation, but early divergence of “basal” fungi is harder to define [31]. A single loss of flagellum was initially suggested as the primal cause of divergence between flagellated fungi (Chytridiomycota) and non-flagellated fungi (Zygomycota, Glomeromycota, Ascomycota and Basidiomycota). This event was believed to be coincident with the adaptation to terrestrial habitats, and phylogenetic studies suggest that loss of the flagellum may have happened either only once during early stages of fungal evolution or four times [32], each one coinciding with evolving mechanisms in production and dispersal of spores. One of these innovations is the polar tube developed by the obligate endoparasitic Microsporidia unicellular organisms. The origin of Microsporidia has long been a subject of study, and the results obtained indicate a phylogenetic relationship between Microsporidia and an ancestor related to the Cryptomycota parasite *Rozella allomycis*, on the earliest diverging branch of the fungal tree [33]. It has been proposed that Cryptomycota and Microsporidia share a common endoparasitic ancestor, with the clade unified by a chitinous cell wall used to develop turgor pressure in the infection process [33]. Shifts between saprophytic and pathogenic nutritional strategies can be inferred based on the early phylogeny of fungi. After the divergence of the microsporidia and *Rozella* lineage, the rest of the fungal organisms developed the filamentous growth response [34], although a yeast-like growth has been reported for the early diverging ancestors of Ascomycota and Basidiomycota. The further lineage expansion probably took place in a solid ground prior to the colonization of plants. Fossilized hyphae and spores indicate that Glomales-like fungi originated

before the appearance of vascular plants [35], at a time when the land flora mainly consisted of bryophytes. This supports the theory that suggests that Glomalean (Glomeromycota) fungi may have played a crucial role in facilitating the colonization of land by plants thanks to the advent of mycorrhization.

Figure 1.2 | The phylogeny of fungi. Phylogenetic tree reconstruction of the fungal kingdom, according to [36]. Previous molecular phylogenies based on one or a few genes did not support the monophyly of the phylum Zygomycota, which has been abandoned in favor of a more diverse classification. Highlighted fungal clades have been used in our evolutionary analysis (see chapter 3.4).

The phylum Zygomycota was firstly defined in 1954, but recent phylogenetic analyses have shown its non-monophyly nature [37], which have progressively lead to the abandonment of the term Zygomycota in favor of two larger groups: "zygomycetes I", including Mucoromycotina, Mortierellomycotina and Glomeromycotina (MMG group) [38] and "zygomycetes II", including Entomophthoromycotina, Kickxellomycotina, and Zoopagomycotina (EKZ group) [39,40]. A more recent phylogenetic reconstruction has proposed the inclusion of two phyla, six subphyla, four classes and sixteen orders [36].

The Dikarya ("higher fungi") clade is formed by Basidiomycota and the larger Ascomycota phyla. Ascomycota includes three subphyla: the earliest diverging Taphrinomycotina, the

numerous Pezizomycotina subphylum and the Saccharomycotina subphylum, which contain the 'true-yeasts' hemiascomycetes group. Previous analysis of a number of hemiascomycetes genomes showed how an interplay of mechanisms such as segmental duplication, massive genome duplication and tandem gene repeat organization led to a differentiated evolution within these species [41]. As for the Basidiomycota, which also comprises the true mushrooms (Agaricomycotina), includes around 31,000 species divided into three subphyla, although the phylogenetic associations implied are still unclear.

mRNA life cycle

Messenger RNA (mRNA) represents the intermediate connection between DNA encoded genes and protein synthesis. Nuclear genes are transcribed into pre-mRNA, which, after several rounds of post-transcriptional modifications and export into the cytoplasm, is transformed into the mature form and translated by ribosomes. During its short life as a molecule in the eukaryotic cell, pre-mRNA is processed (spliced and polyadenylated), exported, localized, translated and finally degraded. These steps are highly regulated and achieved through the cooperation of hundreds of proteins, which are generally well conserved in the eukaryotic tree of life. A short introduction about mRNA life cycle is now presented, based on the knowledge obtained from studies made on fungi, especially the unicellular *Saccharomyces cerevisiae*.

mRNA life cycle starts with the transcription of DNA from protein-coding genes into pre-mRNA by RNA polymerase II (RNA Pol II) (whereas other non-coding RNAs are generally transcribed by RNA Pol I and RNA Pol III) [42]. RNA Pol II-dependent transcription is the best studied and *S. cerevisiae* is the best-known fungal system. mRNA transcription starts at the gene promoter usually around a sequence element called the TATA-box, 30 to 100 bp upstream of the transcription start site (TSS). This phase is called initiation and it is followed by two further phases, elongation and termination [43]. Initiation can be further subdivided into preinitiation complex (PIC) assembly, promoter DNA melting and early initiation events [44]. mRNA transcription results in the synthesis of a pre-mRNA molecule, roughly composed of three segments: the 5' untranslated region (5'UTR), the coding sequence (CDS), and the 3' untranslated region (3'UTR). The coding sequence is defined by an open reading frame (ORF) interrupted by non-coding elements called introns [45]. A great number of proteins and protein complexes participate in mRNA transcription, namely the specific transcription factors (STF), the chromatin remodeling (CR) complexes, the Mediator (Med) complex and general transcription factors (GTF) [46]. Several events are coupled to transcription through the modification of its C-terminal domain (CTD) [47–50]. The pre-mRNA is modified at the 5'-end by adding a 7-methylguanosine molecule at the first nucleotide to stabilize it during mRNA capping [51], introns

are spliced by a multiprotein complex called the *spliceosome* [52], the 3'UTR is cleaved and polyadenylated [53], and the export factors are recruited for its transport to the cytoplasm [54].

Introns and splicing are features of eukaryotic organisms. The spliceosome is composed of more than 100 proteins and five small nuclear RNAs (U1, U2, U4, U5 and U6 snRNAs). Some of these snRNAs associate with splicing proteins creating the small nuclear ribonucleoprotein (snRNP). The typical model of an intron is defined by three elements: the 5' splice site, the branchpoint and the 3' splice site [55,56]. These elements are recognized by the spliceosome at different stages of the splicing process and they allow the catalytic reaction to occur. At the beginning of spliceosome assembly, U1 snRNP associates with the pre-mRNA transcript through a base-pairing interaction between the 5'-end of U1 snRNA and the 5' ss of the intron [57]. Next, the branchpoint is recognized by the BBP-Mud2 heterodimer in an ATP-independent manner [58], and the U2 snRNP is later recruited [59]. The spliceosome is finally activated by the formation of the U4/U6•U5 tri-snRNP complex [60] and the acquisition of the Nineteen Complex (NTC) [61]. The splicing reaction consists of two sequential transesterification reactions separated by a period of spliceosomal remodeling [62]. After the intron has been spliced, the spliceosome disassembles [63,64]. The spliceosome assembly is also monitored by the SR proteins, a protein family that comprises a number of phylogenetically conserved and structurally related proteins with a characteristic domain rich in arginine and serine residues, known as the RS domain. They play significant roles in constitutive pre-mRNA splicing and are also important regulators of alternative splicing [65,66]. A small number of introns are spliced through a non-canonical pathway by the minor spliceosome. The minor spliceosome consists of U11, U12, U4atac, and U6atac, together with U5 and an unknown number of non-snRNP proteins. These introns (called U12-type introns) represent less than 1% of all introns in the human genome [67].

In another nuclear transcription-coupled mechanism, the pre-mRNA 3' untranslated region (3'UTR) is cleaved and a polyadenosine tail is added in a two-step process called polyadenylation [68]. It is another mechanism usually restricted to eukaryotic cells. These two chemical reactions, the endonucleolytic cleavage and the addition of an adenosine tail are performed by a small set of approximately 20 proteins, generally well conserved within the eukaryotic kingdom [69]. The polyadenylation machinery assembles onto a specific location at the 3'-end of the pre-mRNA called the polyadenylation site [70,71]. The specificity of the polyadenylation site is generally defined by a set of cis-elements and RNA secondary structure constraints [72]. A core upstream element, defined as the polyadenylation signal, was found to follow the consensus sequence AAUAAA [73]. However, this pattern has been shown to be less homogeneous in other species [73]. The majority of polyadenylation factors in *S. cerevisiae* are essential for viability and they assemble into three major complexes: Cleavage Factor IA (CFIA), Cleavage Factor IB (CFIB), and Cleavage and Polyadenylation Factor (CPF) [74].

Figure 1.3 | Overview of fungal mRNA life cycle. mRNA life cycle starts with its transcription from the nuclear genome by Pol II with the aid of several transcription factors. pre-mRNA is then processed with the splicing of introns, the capping of the 5'-end and the cleavage and polyadenylation of the 3' UTR. The mature mRNA is now ready to be exported to the cytoplasm where it can reach the ribosomes to be translated. A few cytoplasmic surveillance systems overlook this process. Translation occurs with the participation of tRNAs, which are also transcribed, processed and exported to the cytoplasm. mRNA silencing negatively regulates mRNA expression. mRNA finally undergoes degradation through different pathways, both in the nucleus and the cytoplasm. Abbreviations used; **Transcription** - *Med*: Mediation complex, *CR*: Chromatin remodeling, *STF*: Specific Transcription Factors, *GTF*: General Transcription Factors, *TFIID*: Transcription factor II D, *Pol II*: RNA Polymerase II, *CTD*: C-terminal domain, *UAS*: upstream activating sequence, *TATA*: TATA-box; **Splicing** - *U1*, *U2*, *U4*, *U5* and *U6*: spliceosomal snRNPs, *NTC*: Nineteen complex; **3'UTR processing** - *CFII*: Cleavage Factor II, *PFI*: Polyadenylation Factor I, *APT*: Associated with Pta1, *CFIA*: Cleavage Factor IA, *CFIB*: Cleavage Factor IB, *UUE*: Upstream U-rich Element, *DUE*: Downstream U-rich Element, *PA*: A-rich Position Element, *EE*: AU-rich Efficiency Element; **mRNA export** - *THO*: THO complex, *TREX*: TREX complex; **Silencing** - *aRNA*: aberrant RNA, *siRNA*: small interfering RNA; **Cytoplasmic surveillance** - *NMD*: Nonsense mediated decay, *Non-stop*: Non-stop decay, *No-go*: No-go decay; **Translation** - *40S*: Small ribosomal subunit, *60S*: Large ribosomal subunit; **Degradation** - *NNS*: Nrd1–Nab3–Sen1 complex, *CUT*: Cryptic Unstable Transcript

A great number of genes display more than one polyadenylation site, a phenomenon called alternative polyadenylation, which can result in the production of different mRNA isoforms [75,76]. The causes and effects of alternative polyadenylation are still the subject of investigation, but it is clear that developmental stage and environmental conditions can affect polyadenylation site selection, and that alternative polyadenylation can affect transcript stability, localization, and translation efficiency [77].

The mature mRNA, after being capped, spliced, cleaved and polyadenylated is later exported into the cytoplasm through the nuclear pore complex (NPC) by the action of several export factors [47]. Phosphorylation of the C-terminal domain (CTD) of RNA Pol II during transcription initiation attracts the shuttling serine/arginine (SR)-rich protein Npl3 onto the mRNA [78,79]. A second phosphorylation status change of the CTD during transcription elongation recruits the heterotetrameric THO (Hpr1, Tho2, Thp1, and Mft1) complex, which prevents DNA:RNA hybrid formation [80,81]. THO is later bound by Sub1 and Yra1 to form the TREX (**T**Ranscription **E**Xport) complex [82,83], a process that occurs in tandem with splicing. Two shuttling proteins, Gbp2 and Hrp1 are loaded onto the mRNA and Hrp1 (which is also participating to polyadenylation) also joins the other export factors [84]. During this stage poly(A) binding proteins Nab2 and Pab1 contribute to a controlled poly(A) tail length synthesis [85]. The mRNP (ribonucleoprotein complex) is now released from the transcription machinery and is ready to be exported through the NPC with the final association of multiple Mex67-Mtr2 heterodimers [86]. The NPC is an eight-fold symmetrical structure made of approximately 30 proteins, composed of a central area, a nuclear basket and cytoplasmic filaments [87]. The mRNP interacts and docks to the NPC, and it is eventually translocated into the cytoplasm,

where translation occurs [88]. mRNA can also be localized and asymmetrically distributed by active transport of motor proteins, an important mechanism that contributes to cell polarity and cellular asymmetry [89].

The ribosome is the largest non-membrane bound organelle of the cell, and the main actor in mRNA translation [90]. The eukaryotic ribosome is composed of two subunits: the 40S subunit (which consists of the 18S ribosomal RNA (rRNA) and ca.30 ribosomal proteins), and the 60S subunit (which consists of the 5S, 5.8S and 28S rRNAs, and 40–50 ribosomal proteins). The fully assembled 80S ribosome contains an mRNA channel and three tRNA binding sites, named A, P and E. Translation can be roughly divided into three steps: initiation, elongation and termination, where elongation is the most conserved of the three [91]. Translation initiation begins with the assembly of two large ribonucleoprotein complexes: the pre-initiation complex (PIC) binds the small ribosomal subunit, and the cap-binding complex binds the mRNA [92,93]. The main players in these complexes are the eukaryotic initiation factors (eIFs). The cap-binding complex is entrusted to remove the 5' cap, while the PIC scans the 5'UTR to remove any impeding secondary structures [94]. The start codon is then recognized and the tRNA^{Met} reaches the P-site [95]. Translation elongation is a loop process that starts with the decoding of the A-site codon, followed by the transfer of the P-site amino acid onto the A-site tRNA, the release of the tRNA on the E-site and the translocation of the ribosome onto the following codon [96]. The loop is broken when one of the three stop codons enters the A-site [97], the emerging peptide is released and the ribosome dissociates, while its subunits are recycled and reused for other translation processes [91,98].

Non-coding RNAs (ncRNAs) like transfer RNAs (tRNAs) play an essential role during translation [90,99]. After transcription by RNA Pol III, tRNAs undergo a number of chemical modifications that help to maintain the general shape of the molecule and improve its interactions with other molecules. tRNA 5' leader and 3' trailer ends are trimmed by the ribonucleoprotein RNaseP and the Rex1/RNaseZ/La complex, respectively [100,101]. Subsequently, the 3' CCA is added by the CCA nucleotidyl transferase Cca1 [102], and introns are spliced [103]. The pre-tRNA therefore is subject of an extensive round of posttranscriptional modifications, with more than 60 different tRNA chemical modifications described in eukarya alone, for which a database is available [104]. Nevertheless, out of the more than one hundred known modifications, only 17 are universally present in all domains of life, with a clear evolutionary propensity to increase the number and complexity of tRNA modifications [105].

Not all mRNA molecules reach the ribosomes and are translated. mRNA transcription and maturation is prone to errors, and a strict cytoplasmic surveillance systems exists to prevent translation of aberrant transcripts [47,106,107]. A number of quality-check mechanisms are available in the eukaryotic cell, which detect and degrade non-conforming mRNA transcripts, preventing their expression. The nonsense-mediated mRNA decay pathway (NMD) is entitled to

clear cells from mRNAs harboring premature in-frame stop codons [108]; the nonstop decay pathway (NSD) inspect and suppress mRNA lacking a stop codon [109]; and the No-Go decay (NGD) is assumed to degrade mRNAs on which ribosomes have stalled during elongation [110].

Additional mRNA degradation can occur through different mechanisms depending on the specific situation of the mRNA molecule [111]. Sometimes non-conforming mRNAs are never exported to the cytoplasm, and are degraded in the nucleus [112]. Transcripts that are not correctly capped are usually degraded 5' to 3' in the nucleus by Rat1 [113]. The nuclear exosome is a 3' to 5' exonucleolytic enzyme able to degrade any polymerase product. The TRAMP (Trf4/Air2/Mtr4p Polyadenylation complex) complex is a major functional nuclear co-factor of the exosome that adds poly(A) tails to nuclear RNA [114]. The NNS (Nrd1–Nab3–Sen1) complex acts upstream of TRAMP and the exosome and is involved in transcriptional termination of cryptic unstable ncRNAs (CUT) and of stable ncRNAs like small nucleolar RNAs (snoRNAs) and small nuclear RNAs (snRNAs) [115]. Degradation in the cytoplasm generally starts with the deadenylation of the mRNA (which is polyadenylated at this stage); this role is assigned to the Pan2/Pan3 and Ccr4/Not complexes [116]. The deadenylated transcript is then uncapped by the Dcp2/Dcp1 complex and degraded 5' to 3' by the cytoplasmic exonuclease Xrn1 (related to the nuclear Rat1) [117,118]. Degradation can also take place 3' to 5' by the cytoplasmic exosome together with Ski [119].

Many eukaryotic organisms also possess a distinctive mRNA regulation mechanism called RNA silencing (or RNA interference, RNAi), which acts to negatively regulate gene expression [120,121]. RNA silencing mechanisms have been extensively studied in animals and plants, which share many features, while in fungi this process is less homogeneous [122–124]. RNA silencing was firstly observed in *Neurospora crassa* in the form of a phenomenon happening during the asexual cycle called “quelling” [125]. The other two silencing processes named repeat-induced point mutation (RIP) and meiotic silencing by unpaired DNA (MSUD) occur during the sexual cycle [126,127]. RNA silencing in *N. crassa* is generally induced by the presence of repetitive sequences (like transposable elements), DNA damage, or unpaired DNA during meiosis. Aberrant transcripts are produced by these elements, which are cleaved by Dicer proteins and loaded into the RNA-induced silencing complex (RISC) complex, where the passenger strand is removed by the QIP exonuclease [128]. The RISC complex is made by an Argonaute (AGO) protein and a guide RNA called small interfering RNA (siRNA). The guide RNA recognizes a target molecule which is later cleaved by the Ago endonuclease [129]. RNA silencing activity with similar characteristics (but different pathways) has been observed in *Schizosaccharomyces pombe* [130], *Mucor circinelloides* [131] and *Cryptococcus neoformans* [132], while it is missing the model organism *S. cerevisiae*.

Transcriptomics, RNA-sequencing and Bioinformatics

Transcriptomics is the study of the transcriptome, which is the total set of all RNA molecules in the cell, generally examined in the context of varying conditions of interest to determine the effects of those conditions. The transcriptome can include both mRNAs and ncRNAs like tRNAs, snoRNAs, snRNAs and microRNAs (miRNAs). The purpose of transcriptomics is to study the population of RNA transcripts in order to better understand its role in development, differentiation, stress response, disease and other physiological or environmental conditions. Several technologies have been developed to understand and quantify the transcriptome, including hybridization and sequence-based approaches [133]. Hybridization-based approaches typically involve coupling fluorescently labeled probes with the cDNA of interested on custom-made microarrays, which are later washed and the hybridization intensity is measured with laser scanners [134]. These kinds of techniques are high throughput and generally inexpensive, but they rely upon existing knowledge of gene sequences. Conversely, sequence-based approaches aim at detecting the nucleotide sequences themselves. Sanger sequencing was developed in 1977 to sequence cDNA, and used continuously for 25 years to build a large number of cDNA and EST libraries [135]. Everyday use of this technology was however hindered due to inherent limitations of cost, low throughput, scarce scalability, low speed, and bad resolution. In the early 1990's several tag-based methods were introduced to overcome these limitations, including serial analysis of gene expression (SAGE) [136,137], cap analysis of gene expression (CAGE) [138,139] and massively parallel signature sequencing (MPSS) [140]. Tag-based sequencing approaches are high throughput and can provide more precise gene expression levels. Nevertheless, these techniques still experienced expensive costs and low resolution levels. The new millennium witnessed the appearance of Next Generation Sequencing (NGS): in 2004, 454 Life Sciences introduced pyrosequencing [141]; in 2005, Solexa (later Illumina) released the Genome Analyzer (GA), and in 2007 appeared the first SOLiD sequencing system. Illumina systems quickly began to be the reference for RNA-sequencing with the HiSeq platform series (including HiSeq 4000, HiSeq 3000, HiSeq 2500, HiSeq 2000, HiSeq 1500, HiSeq 1000 and MiSeq). HiSeq 4000 platform has the capacity to produce 5 billion 150 bp long pair-end reads in 1-3.5 days (<https://www.illumina.com/techniques/sequencing/rna-sequencing/mrna-seq.html>). RNA-sequencing made it possible to overcome several microarray limitations; for example RNA-seq presents a strong concordance between platforms, a higher sensitivity and dynamic range, lower technical variation, availability for all species, novel transcribed region detection, alternative splicing identification, allele-specific expression, and fusion gene events recognition [142]. In 2009 the first single-cell whole transcriptome sequencing was published [143]. Other

sequencing methods of NGS technologies include: Helicos sequencer released in 2009 (the company fell into bankruptcy shortly after), Life Technologies Ion Torrent sequencer [144], Pacific Biosciences [145], and Oxford Technologies Nanopore single molecule sequencer [146].

Bioinformatics was first described in 1970 as “the study of informatics processes in biotic systems”, but the activities that fell into that category were little related to what now it is understood as “bioinformatics”. By 1981, 579 human genes had been mapped and the need for integrated database resources became pressing. The first protein sequence resource was created by Margaret Dayhoff in 1965, containing all the 65 protein sequences known at the time. The Protein Data Bank (X-ray crystallographic protein structures) was created in 1972 containing the X-ray crystallographic structure of ten proteins [147], and the SWISSPROT protein sequence database was installed in 1987 [148]. The birth of the Internet and the initiation of the Human Genome Project were among the reasons that bioinformatics became recognized as a distinct research field [149]. The huge amount of biological information of all types (sequence, biochemical/metabolomic, phenotypic, and ecological) has led to the proliferation of often highly-specialized databases, containing vast quantities of information about many species. Nevertheless, sequence databases remain among the largest and most critical resources for the community.

NGS output is generally in the format of fastq files (fasta files including information about base quality), and requires an intensive bioinformatic analysis in order to produce interpretable results. Data quality control and preprocessing is the first step. Reads library quality needs to be evaluated removing contaminant (adapters and PCR primers) and low quality/complexity reads and bases. Typical software used for these purposes include FastQC [150], SAMstat [151](for data quality assessment); FASTX toolkit [152], cutadapt [153], trimmomatic [154] and fqtools [155](for data pre-processing). Reads are later mapped against a reference genome or transcriptome. Common alignment tools are: Tophat [156], bowtie [157], GSNAP [158] and STAR [159]. An important parameter to take into account when doing reads alignment is whether splicing should be considered (aligning against a genome) or not (aligning against a transcriptome). Reads alignment can be then be used for differential gene expression (DGE). Typical DGE tools are Cuffdiff [160], EdgeR [161], DESeq2 [162] and RSEM [163]. In case of missing genome or transcriptome data, NGS data can be used for *de novo* transcriptome assembly. Popular *de novo* transcriptome assembly software include: Trinity [164], Velvet [165] and ABySS [166].

NGS technologies are now routine in many laboratories, and have contributed to the attainment of a rich and deep view of gene expression and associated biological implications. NGS is today fundamental to the study of complex diseases and precision medicine. NGS accessibility also greatly benefited transcriptomic research: today, researchers are leveraging the power of NGS to resolve sequences down to single-transcript sensitivity [167], allowing the

investigation of the transcriptome in a highly quantitative manner at single nucleotide resolution. Furthermore, the rapid accumulation of data in public repositories will permit the integration of many different RNA-seq datasets, as well as distinct levels of omics data to unravel the complexity of biological systems and eventually ease work of scientists and professional all around the globe.

Objectives

1. Bioinformatic analysis of RNA sequencing data using polyadenylation site-specific libraries from three *M. oryzae* strains: wild-type, $\Delta rbp35$ and $\Delta hrp1$. $\Delta rbp35$ and $\Delta hrp1$ lack Rbp35 and Hrp1 proteins, respectively, both components of the polyadenylation machinery. The aim of the analysis is to identify the polyadenylation sites of genes expressed in a variety of growth conditions, and to observe the differences in the mutants in order i) to infer the role of these two proteins in 3' UTR processing and polyadenylation, and ii) to predict their target genes.
2. Bioinformatic analysis of RNA sequencing data of sRNA libraries from three *M. oryzae* strains: wild-type, $\Delta rbp35$ and $\Delta exp5$. $\Delta exp5$ is a knock-out mutant of a gene encoding for the karyopherin exportin-5. The aim of the analysis is to describe the small RNA (sRNA) population in the rice blast fungus and to observe the differences in the mutants in order to infer the possible function of these two genes in sRNA synthesis, processing, import/export and silencing pathways.
3. Molecular characterization of a novel ssRNA ourmia-like virus from *M. oryzae*. The virus was identified during the analysis of the small RNA sequencing described in the previous objective, and later fully characterized *in vitro*.
4. Evolutionary analysis of the major RNA processing machineries and RNA-related proteins in the fungal kingdom. The goal of this analysis is to identify the orthologs of the proteins involved in the metabolism of RNA in the typical fungal cell, and describe their conservation pattern. Transcriptomic data will also be analyzed in order to compare the results from evolutionary conservation analysis with the actual RNA population in the cell.

2. Materials and methods

Chapter 3.1

Fungal media, growth conditions and infection assays

Fungal strains were grown on different media composition: CM (complete medium) and MM (minimal medium) [168], and DCM (defined complex medium) [169]. Fungal growth on nitrogen and carbon starvation was carried out on MM depleted of nitrogen or carbon sources. Leaf and root infections were conducted as previously described [170].

RNA extraction and Library preparation

Fungal material was harvested from 60-hour-old fungal mycelia grown in liquid CM (24 hours) and then transferred for an additional 12 hours in a fresh CM, MM, MM-N or MM-C. RNA was extracted using a LiCl protocol [171,172]. Three biological replicates were independently harvested and extracted for each library preparation. Library preparation were performed by EMBL (Heidelberg) using a previously described method [173].

RNA-seq analyses and PASs detection

De-multiplexed samples were trimmed to remove poly(A) tails and Illumina adapter sequences using the fastq-mcf utility from the ea-utils package (<http://code.google.com/p/ea-utils/>), keeping only reads with at least 17 nucleotides after trimming. Read-mapping was performed using STAR [159] with default parameters. The alignments obtained were later filtered removing low-quality mapping (MAPQ <30, see <https://samtools.github.io/hts-specs/SAMv1.pdf>), mappings with a high-level of A/Ts (>80%) and potential internal primings (more than seven As in the first 5' nucleotides or more than twelve As in the whole alignment (~2.5% of the alignments)).

The resulting alignments were assigned to the genomic features listed in the Ensembl Fungi gene annotation for *M. oryzae* version 29 (<http://fungi.ensembl.org/>), containing 13218 annotated features. Each read was assigned to its overlapping feature (including protein-coding genes and the available ncRNAs), ambiguous cases were assigned to the closest 3' terminal end. To account for incomplete gene 3' UTR lengths, we added an extra 400 bp length to every feature. Reads that could not be assigned to any known features were treated separately.

A polyadenylation site (PAS) was considered to be called with high-confidence if it was detected in at least two of the three replicates according to the following rule: its expression is considered distinct from basal noise if its standard score calculated against the whole gene expression (summing up all PASs expression from that gene) has a confidence level > 99% (z-score > 2.58 or z-score < -2.58). The standard score was calculated as: $z\text{-score} = (\text{value} - \text{mean}) / \text{std}$, where "value" is the number of supporting reads for the PAS, "mean" is the mean value of supporting reads for all the PASs in the gene, and "std" is the standard deviation of supporting

reads for all the PASs in the gene. PASs with less than five supporting reads in every replicate were discarded. High-confidence PASs were subsequently merged if located less than 33bp from an adjacent one (roughly the span of the polyadenylation site regulatory region, including the A-rich and the U-rich regions upstream of the poly(A) cleavage), and only the most expressed PAS was retained.

Differential gene expression and differential poly(A) site usage

Differential gene expression analysis was performed using the R/Bioconductor package DESeq2 [162], using the whole reads count for each feature (the number of reads that have been assigned to that feature) in the three replicates as described before. A gene was considered differentially expressed if the p-value adjusted for false discovery rate control (the padj column in DESeq2) is less than 0.05.

Differential PAS expression analysis was performed using the R/Bioconductor package DEXSeq [174], using the reads count for high-confidence PASs for each feature in the three replicates as described before. A PAS was considered differentially expressed if the p-value adjusted for false discovery rate control (the padj column in DEXSeq) is less than 0.05.

Gene functional groups enrichment and sequence analysis

Roughly 60% of annotated genes in the *M. oryzae* genome have at least a single Gene Ontology (GO) annotation. We used the FungiFun2 utility to perform gene group functional profiling and statistical enrichment analysis of gene ontologies terms [175]. FungiFun2 utilizes a variety of annotation/classification methods such as FunCat (Functional Catalogue) [176], GO (Gene Ontology) [177] and KEGG (Kyoto Encyclopedia of Genes and Genomes) [178]. FungiFun2 categorizes genes and proteins of fungal species on different levels and conducts an enrichment analysis.

Enriched oligonucleotide motifs in the interesting regions were discovered using the DREME tool from the MEME suite [179,180].

Chapter 3.2

Fungal Strains, Growth Conditions and infections assays.

M. oryzae strains used in this study were the wild type strain Guy11 [181] and the deletion mutant $\Delta exp5$ [170]. *M. oryzae* strains were incubated in a controlled temperature room at 25°C with a 16h light/8h dark cycle. Growth media (complete medium-CM, defined complex medium-DCM), and leaf and root infection assays were carried out as previously described [168,170,172].

Cellular fractionation, RNA extractions and sRNA library preparation.

M. oryzae cellular fractionation was performed using a modified protocol from *Arabidopsis thaliana* [182]. Briefly, cellular fractionation was carried out using grounded, frozen mycelia obtained from liquid CM cultured during 48h at 25°C 120rpm in darkness. The same volume of cell wall-disrupting buffer (10 mM potassium phosphate/ pH7.0; 0.1M NaCl; 10 mM β -mercaptoethanol; 1M hexylene glycol/Sigma M-1408) was added to the mycelia. After centrifugation at 1,500 x g for 10 min at 4°C both, the supernatant and the pellet were recovered. Supernatant was re-centrifuged at 13,000 x g for 15 min at 4°C. The supernatant after the second centrifugation was saved as the cytoplasmic fraction. Pellet from the first centrifugation was washed with nuclei preparation buffer (10mM potassium phosphate/pH7.0; 0.1M NaCl; 10mM β -mercaptoethanol, 1M hexylene glycol; 10mM $MgCl_2$; 0.5% Triton X-100) and centrifuged at 1,500 x g for 10 min at 4°C. Washing and centrifugation were repeated 4-5 times, and the final pellet was saved as the nuclear fraction.

Total RNA isolation was done using a modified protocol of the LiCl method [172]. Two to three additional washes with phenol:chloroform were implemented to avoid RNA degradation and increase RNA purity. Using this protocol RNAs more than 500nt in length were detected in the precipitate, while sRNAs <500nt remained in the supernatant. To precipitate these sRNAs 2x volume of ethanol 100%, 1.5x volume of 7.5M Ammonium acetate and 2 μ l glycogen (5mg/ml) were added. Samples then were incubated during 1h at -80°C and centrifuged at 13.000 x g for 30 min at 4°C. The pellet was washed with EtOH 70% and resuspended with RNase-free water. A cleaning step with phenol-chloroform 1:1 volumes was carried out, incubating samples for three minutes at room temperature. After centrifugation at 13.000 x g 30 min at 4°C, supernatant was taken and added 0.1x volume 3M sodium acetate pH 5.2, 2.5x volumes of ethanol 100% and glycogen (5mg/ml), leaved at -80°C for an 1h, and centrifuged at 13.000 x g for 30min at 4°C. The final pellet was resuspended in 50 μ l RNase-free water. A second method for RNA isolation was carried out using Trizol extraction buffer (Invitrogen). After total RNA was extracted from the sample, RNAs were treated with the mirVana kit (Ambion) for enrichment in sRNAs up to 200nt length. sRNAs were enriched using at least 10 μ g of total RNA. Small cDNA library was performed as previously described [183].

sRNA sequencing analysis

Fastq reads were trimmed of their 3'-end adaptor TGGAATTCTCGGGTGCCAAGG, low quality reads were discarded and artifacts removed with the tools from the FASTX-Toolkit [152], the command line options used were: -c -a TGGAATTCTCGGGTGCCAAGG -i "fastq_file" |

fastq_quality_filter -q 30 -p 70 | fastx_artifacts_filter. Only reads containing the 3'-end adapter were retained.

Subsequently, HD adaptors were also removed trimming 4 nucleotides from both 5' and 3'-ends. Reads were aligned with bowtie [157] allowing no mismatches against *M. oryzae* 70-15 reference genome assembly 8, downloaded from the Ensembl Fungi website version 29 along with its annotation. The following command was used to align the reads against the genome: bowtie -v 0 -M 1 "genome_index" "fastq_file". Differential expression of sRNA was performed with DESeq2 [162].

Northern blotting assays of sRNAs.

10 µg of non-fractionated and cytoplasmic sRNAs, and 7.5 µg of nuclear sRNAs obtained from both sRNA isolation methods used from Guy11 and $\Delta exp5$ mutant were run on a denaturing 15% polyacrylamide gel with 7M urea, and transferred onto a Hybond-NX nylon membrane and fixed by chemical crosslinking. Complementary DNA oligonucleotide probes (Sigma) to targeted sRNAs were labelled with [γ -³²P]-ATP using T4 PNK (Roche) and detected using phosphorimager screens (Fujifilm). The Biorad molecular software, Image Lab. 3.0.1. (Beta 2), was used for signal visualization, and ImageJ software was used for quantification of signal strength and image processing (<http://rsb.info.nih.gov/ij/>).

Chapter 3.3

Northern blotting assays

Northern blots using 30 lg of total RNA extracted from *M. oryzae* Guy11 grown on complete media. MOLV1 and actin probes were PCR products amplified with MOLV1_fw (5'GCTTTTCGCTTCTTTGGT CGT3')/MOLV1_rv (5'GCTCCCTAACGTGGCGATAA3'), and Actin_fw (5'CTTCTCTCGGTTGGACTTGG3')/Actin-rv (5'TTCTACAACGAGCTGCGTGT3') primers, respectively. The actin probe was used as hybridization control. Ethidium bromide gel staining of 28S ribosomal RNA is shown as a loading control.

RNA ligase-mediated rapid amplification of cDNA ends (RLM-RACE)

5' and 3' RLM-RACE reactions were carried out using 5'Outer (5' ACTCTCAGATCAACCCTGAAGGT 3') / 5'Inner (5' TCACTAAGCAGTCCAGAGAACGA 3') and 3'Outer (5' AGGCGTATCGTCGAAAAGCCAA 3') / 3'Inner (5' TTAAAGTTGTCGGTGAACGCGA 3') primers, respectively. PCR products were loaded into a 2% agarose gel.

ePAT experiment

ePAT experiment using total RNA from *M. oryzae* Guy11. PCR amplification of cDNA derived from 1 lg of the RNA sample using gene-specific primers for the virus (ePAT Fw 5'-GCCGGTCTGCTGTTGGGAT-3' and reverse 5'-TTTTTTTTTTTTCTTAAACCTACATGCTGGTC-3'). This ePAT reaction generated a fragment of approx. 468 bp, which reflects the size of the 3'UTR plus the length of the poly(A) tail. The TVN-PAT reaction was used with the same forward primer and the universal ePAT reverse primer (5'-GCGAGCTCCGCGGCCGCG-3'), which generated a fragment of approx. 438 bp consistent with amplification of the 3'UTR without the poly(A) tail. PCR products from ePAT and TVN-PAT reactions were visualised in 2% agarose gels.

Phylogenetic tree reconstruction

Phylogenetic tree obtained by the Maximum Likelihood method with the RdRp sequences of 25 mycoviruses, including MOLV1, and 3 plant ourmiaviruses. The RdRp sequence of TMV (ABN79257.1) was included as outgroup. All bootstrap values (%) are represented at each node of the tree (test with 1000 replicates). Branch lengths are proportional to the number of amino acid substitutions and are measured by the scale bar. Sequence accession numbers of viruses used are: Botrytis ourmia-like virus (LN827955), Soybean-associated ourmiavirus 1 (KT598235.1), Soybean-associated ourmiavirus 2 (KT598247.1), Sclerotinia sclerotiorum ourmia-like virus 1 (ALD89138.1), Sclerotinia sclerotiorum ourmia-like virus 2 (ALD89139.1), Rhizoctonia solani ourmia-like virus 1 (ALD89131.1) Rhizoctonia solani ourmia-like virus 2 (ALD89132.1), Sclerotinia sclerotiorum mitovirus 6 (AHX84133.1), Ophiostoma mitovirus 3a (NP_660176.1), Sclerotinia sclerotiorum mitovirus 7 (AHE13866.1), Botrytis cinerea mitovirus 1 (YP_002284334.1), Cryphonectria parasitica mitovirus 1-NB631 (NP_660174.1), Cryphonectria cubensis mitovirus 2a (AAR01973.1), Ophiostoma mitovirus 1a (CAJ32466.1), Sclerotinia sclerotiorum mitovirus 2 (AEX91879.1), Cryphonectria cubensis mitovirus 1a (AAR01970.1), Sclerotinia sclerotiorum mitovirus 8 (AHE13867.1), Ophiostoma mitovirus 6 (NP_660181.1), Sclerotinia sclerotiorum mitovirus 5 (AHX84130.1), Ophiostoma mitovirus 4 (NP_660179.1), Phytophthora infestans RNA virus 4 (AEM89293.1), Saccharomyces 20S RNA narnavirus (NP_660178.1), Saccharomyces 23S RNA narnavirus (NP_660177.1), Cassava virus C (CsVC, YP_003104770.1), Ourmia melon virus (OuMV, YP_002019757.1) and Epirus cherry virus (EpCV, YP_002019754.1). Evolutionary analysis was conducted using the software MEGA6 [184]. Multiple sequence alignments of amino acid sequences of viral RdRps were obtained using MUSCLE (Multiple sequence comparison by log-expectation [185]) using default parameters. Phylogenetic relationships were inferred using the Maximum Likelihood method based on the WAG+G+F protein evolution model [186], and a bootstrap phylogeny test with 1,000 replicates. All positions with less than 50% site coverage were eliminated.

Chapter 3.4

Genomic and transcriptomic data

Assembled genomes, proteomes and transcriptomic data were retrieved from various sources as described in Supplemental Table 3.4.S1. In order to reduce redundancy we considered only the longest isoform for each annotated protein. Transcriptomic data was available either as EST or in raw reads from RNA NGS platforms, depending on the organism. EST sequences were downloaded either from Genbank [187] or the JGI [188] (depending on the species). RNA-seq data were downloaded from the DDBJ Sequence Read Archive (DRA) [189]. Gene annotation and protein prediction were retrieved from the Ensembl website (fungi.ensembl.org) [190], or the JGI website [188]. EST archives were processed using the PASA pipeline [191] in order to predict high confidence polyadenylation sites. When the data came from NGS RNA sequencing platforms, the RNA-seq was first assembled using the Trinity RNA-Seq De novo Transcriptome Assembly tool [164], in order to leverage the data for the PASA pipeline [191].

Sequence analysis and data visualization

Protein domains were predicted using the HMMER3 utility [192], against the Pfam database [193]. Sequence motif detection was performed with the Meme suite [179] and the Regulatory Sequence Analysis Tools RSAT [194]. Ortholog heatmaps and PCA plots were generated with the R software environment [195]. Nucleotides logos were generated with the WebLogo software utility [196]. The phylogenetic tree of the 52 species considered in this study was reconstructed from the concatenation of 7 highly conserved proteins (Ensembl unique protein name for *S. cerevisiae* is shown in parenthesis): actin (Act1), pyruvate kinase (Cdc19), diphthine synthase 5 (Dph5), glutamine tRNA synthetase (Gln4), Type V myosin (Myo4), Second-largest subunit of RNA polymerase III (Ret1), and Subunit 1 of heterotrimeric Replication Protein A (Rpa). The phylogenetic tree was built with the ETE3 toolkit [197] from the concatenated alignments of the aforementioned proteins. The amino-acid sequences of each orthologs group were aligned using ClustalOmega with default parameters [198] and the alignments were cleaned with trimAl [199]. The alignments were next concatenated by ETE3 and the tree was built using RAxML with 100 bootstraps [200].

Ortholog prediction

The protein orthology predictions are generated by a pipeline adapted from the Ensembl compara homology method [201]. The pipeline used in this study is composed of the following steps:

- Retrieve each species' proteome, merge them and perform a reciprocal blastp [202] of every protein sequence against self and nonself, with the default arguments.
- Build a sparse graph of protein relations based on BLAST e-values and generate clusters using the `hcluster_sg` utility from the `treeSoft` package (<http://treeSoft.sourceforge.net/>). The weight used as input to generate the graph are $\text{MIN}(100, \text{ROUND}(-\text{LOG}_{10}(\text{evalue})/2))$. The arguments used to create the graph are the following: `hcluster_sg -m 150 -w 0 -s 0.34 -O`
- For each cluster, build a multiple alignment based on the protein sequences using T-Coffee [203]: with methods `mafftgins_msa` `muscle_msa` `t_coffee_msa`
- For each cluster multiple alignment, build a phylogenetic tree using TreeBeST with the `phymI` command and providing the species tree as argument [201]. TreeBeST builds a maximum likelihood (ML) tree, based on the protein alignment with the WAG model [186] taking into the account the species tree. TreeBeST tries to build a gene tree that is consistent with the topology of the species tree. This "species-guided" `phymI` uses the original `phymI` tree-search code. However, the objective function maximised during the tree-search is multiplied by an extra likelihood factor not found in the original `phymI`. This extra likelihood factor reflects the number of duplications and losses inferred in a gene tree, given the topology of the species tree. The species-guided `phymI` allows the gene tree to have a topology that is inconsistent with the species tree if the alignment strongly supports this.
- From each gene tree, infer protein pairwise relations of orthology using the ETE3 utility [197]. Duplication and speciation events were automatically detected using species overlap (SO) between partitions algorithm [204], allowing a maximum duplication confidence score of 0.25, as described in the Ensembl pipeline.
- For each orthologs group, the protein sequences were aligned with MUSCLE [185], and a ML tree was reconstructed with TreeBeST providing the species tree, as described before. The tree's branches length were used to determine the evolutionary distance between protein orthologs.

3. Results

Chapter 3.1

**Links between pre-mRNA 3'-end processing,
nutritional stress response and plant pathogenesis
in the rice blast fungus**

Summary

Polyadenylation is a eukaryotic double-step process coupled to mRNA transcription that cleaves the 3' UTR and adds a polyadenosine tail. Polyadenylation is important for many mRNA-related processes such as transcript stability, export to the cytoplasm and translation efficiency. mRNAs can harbour more than one polyadenylation site, whose selection depends on many factors like nutrient stress and developmental stage. This phenomenon is known as alternative polyadenylation, and results in the transcription of different mRNA isoforms. Some proteins of the polyadenylation machinery are involved in the regulation of alternative polyadenylation. Two of them, Hrp1 and Rbp35, have been subject of study in the rice blast fungus *M. oryzae*, since they are fundamental for growth and pathogenicity. In this study, we present the first genome-wide polyadenylation site mapping of *M. oryzae*. We also analyze how nutrient stresses like carbon-starvation and the deletion of *HRP1* and *RBP35* genes affect alternative polyadenylation. The output of these studies not only provides the scientific community with a resource containing the genome-wide map of *M.oryzae* polyadenylation sites, but will help clarify the fine relationship between alternative polyadenylation and pathogenicity.

Background

In eukaryotes, messenger RNA maturation occurs through several interdependent and co-transcriptionally regulated steps that involve pre-mRNA formation, 5'-end capping, splicing, 3'-end polyadenylation and degradation [71]. The 3'-end formation of pre-mRNAs is an essential process for eukaryotic gene expression that involves cleavage of the mRNA 3'UTR and subsequent polyadenylation [53]. The polyadenylation machinery is a multi-subunit protein complex composed of about 20 core proteins extensively described in yeast, plants and metazoans [74]. The consensus metazoan 3'-end processing machinery includes the poly(A) polymerase (PAP), poly(A)-binding proteins (PABPs), the large subunits of RNA polymerase II (RNA Pol II) and four multi-subunits complexes: cleavage and polyadenylation specificity factor (CPSF), cleavage and stimulation factor (CstF), cleavage factor I (CFIm) and cleavage factor II (CFIIm). PAP and CPSF are sufficient for both cleavage and polyadenylation reactions. The other three complexes (CstF, CFIm and CFIIm) participate in the selection of a proper polyadenylation site (PAS) for cleavage. PABPs regulate poly(A) tail length and activate PAP.

Typically, a canonical or highly preferred cleavage site is present in all pre-mRNAs. However, it is common to find multiple potential PASs within pre-mRNAs. The selection of alternative PASs is regulated by multiple mechanisms during development and in response to environmental cues [76,77]. The use of different cleavage sites within a pre-mRNA is called alternative polyadenylation (APA), and is catalyzed by the presence of *cis* elements in the mRNA, or mRNA isoforms with different exon content or 3'UTR length. APA has been shown to affect cytoplasmic polyadenylation, subcellular localization, stability, translation and/or decay of the mRNA. Consequently, the selection of a proper 3'-end cleavage site represents an important regulatory step. Proteins involved in APA include CFIm [205,206], CstF64 [207] and Fip1 [208] in metazoans, and Hrp1 [209] in yeast. CFIm and Hrp1 also regulate splicing, mRNA export and nonsense-mediated decay (NMD) activation [210].

Compared to yeast or humans, very few studies on the polyadenylation machinery were previously carried out in filamentous fungi. The Sesma laboratory previously identified Rbp35 as a novel protein component of the 3'-end processing machinery present in some branches of the fungal kingdom, and displaying distant homology with human CFIm68 [172]. Three different subdomains are present in Rbp35: one N-terminal RNA recognition motif (RRM), six Arg-Gly-Gly (RGG) tripeptide repeats and a C-terminal Met-Asp-Gly rich region. Two Rbp35 isoforms, Rbp35A and Rbp35B, are found due to a proteolytic cleavage at the C-terminus of the full-length protein. Truncated forms of Rbp35 are expressed at much higher levels compared to the full-length protein in the rice blast fungus, suggesting a regulatory role of the C-terminal domain of Rbp35 in maintaining its cellular concentration. Rbp35A co-immunoprecipitates *in vivo* with Cfl25, the orthologue of mammalian CFIm25, a highly conserved protein in filamentous fungi but

absent in fission and budding yeasts [171]. Therefore, these studies place Rbp35 as the functional equivalent of metazoan CFI_m68. Rbp35 is not essential for *M. oryzae* viability, and it plays a significant role as a gene-specific polyadenylation factor, regulating alternative 3'UTR processing of infection-related mRNAs.

The Δ *rbp35* mutant displays significant alterations in nitrogen assimilation and grows better in the presence of rapamycin indicating a malfunction of the TOR (target of rapamycin) signalling pathway. This was corroborated using a transcriptomic approach that identified at least five mRNA targets of Rbp35 in *M. oryzae* including the 14–3–3 pre-mRNA, an important integrator of environmental cues and regulator of the TOR signalling cascade [172]. These mRNA targets presented altered 3'UTR lengths in the Δ *rbp35* mutant. A proteomics approach indicated that Rbp35 also controls cellular levels of protein subsets but it is not required for general splicing or translation [171].

The yeast Hrp1 is essential for cell viability. Yeast Hrp1 is a multifunctional protein involved in many cellular activities like pre-mRNA 3'UTR processing, APA, mRNA splicing, mRNA export, NMD and genome integrity maintenance. *S. cerevisiae* Hrp1 is required for the selection of a proper PAS by positioning itself upstream of the cleavage site [209], and regulates APA. Yeast Hrp1 is a vital nucleocytoplasmic shuttling protein, and arginine methylation is required for its efficient nuclear export [211]. In *M. oryzae*, Hrp1 shows a strong nuclear localization although it could also be localized in the cytoplasm, as similarly shown in *F. graminearum* Hrp1, where it can be observed in both compartments [212]. Δ *hrp1* is viable in *M. oryzae* but displays strong pleiotropic defects [171].

In order to elucidate the role of Rbp35 and Hrp1 in pre-mRNA 3'UTR processing, we performed the RNA sequencing of *M. oryzae* applying a new PAS-specific protocol. cDNA libraries generated from wild-type strain and the Δ *rbp35* and Δ *hrp1* mutants, under different nutritional conditions. This analysis allowed us to draw a detailed map of *M. oryzae* PASs and APA events; to depict a discrete representation of the nucleotide elements surrounding the typical *M. oryzae* polyadenylation site which are supposed to take part in the 3'UTR cleavage reaction; to describe the genes differentially expressed in the two mutant strains; and most importantly, to verify the effect of Δ *rbp35*, Δ *hrp1* and nutrient deprivation on PAS selection.

Results

Polyadenylation site identified for 9,327 protein-coding genes of *M. oryzae*

Using a mapping protocol based on deep sequencing of nucleotides located before the polyadenosine tail [213], we completed the genome-wide analysis of *M. oryzae* PASs. The analysis was performed on data from two distinct RNA-sequencing runs (see Materials and Methods). Reads were mapped against the complete genome of *M. oryzae* 70-15 MG8 [18]. The average distance between pair-end reads was approximately 100 bp. From all datasets, around 73% of the total reads were successfully mapped against the genome, of these about 90% could be assigned to known annotated features like protein-coding genes and ncRNAs (Suppl. Fig. 3.1.S1.A). In total, among all samples tested both in the wild-type and in the mutants we identified 27,563 PASs in the *M. oryzae* genome that could be reliably assigned to a known annotated feature. Out of 13,216 annotated genes (including annotated non-coding genes) in *M. oryzae*, we were able to assign at least one high-confidence PAS to 9,327 of them (~70%). More than 90% of the detected PASs are located in the 3'UTR, with the rest appearing in other transcript regions like the 5'UTR, the coding sequence and introns (Suppl. Fig. 3.1.S1B).

~50% of *M. oryzae* genes are alternatively polyadenylated

Combining the results from all growing conditions in the wild-type in both sequencing sessions, 4,587 genes have two or more PASs, which represents at least half of *M. oryzae* annotated genes (Fig. 3.1.1A). This is consistent with recent finding reporting high level of APA in many species [74]. The majority of APA events occur in the 3'UTR, indicating that two or more cleavage sites can occur on the same 3'UTR. The remaining types of APA are represented by genes where, apart from a 'canonical' PAS in the 3'UTR, additional PASs may appear in regions such as the 5'UTR, the coding sequence, and introns. More than 80% of the APA events revealed in our experiment consist of two PASs (Fig. 3.1.1A). To determine if APA occurs more frequently in specific functional gene groups, we performed a statistical enrichment analysis using the FungiFun2 tool (detailed in material and methods). Considering all alternatively polyadenylated genes, we detected that several FunCat categories appear to be enriched (Fig. 3.1.1B). Notably, APA occurs more frequently than expected in genes implicated in protein binding, phosphate metabolism, budding cell polarity, nucleotide binding, protein transport, actin cytoskeleton, vesicular transport, enzyme activator, general transcriptional activities and ATP binding among others.

A **B** **C** **D** **E** **F** **G**

Figure 3.1.1 | Analysis of *M. oryzae* PAS-seq. **A)** Approximately 50% of expressed genes are alternatively polyadenylated (APA). Most APA events involve only two polyadenylation sites (PASs). **B)** Enrichment analysis of functional groups: genes exhibiting APA are more frequently involved in biological processes like protein/ATP/nucleotides binding, phosphate metabolism and transcriptional control. **C)** The nucleotide profile of *M. oryzae* PASs differs slightly from that of *S. cerevisiae*. **D)** The AU-rich Enhancer Element is not found in *M. oryzae*, and two different elements are present in its place: UGUAH and UAGNH. **E)** Average base-pair probability for each nucleotide surrounding the PAS based on RNA secondary structure prediction, taking into account all the identified PASs. Higher base-pair probabilities indicate that a specific nucleotide is probably paired with another one in the same RNA molecule, lower probabilities indicate that that nucleotide is probably not base-paired. **F)** RNA secondary structure prediction considering only specific cleavage sites dinucleotides (CA, TA and GA). **G)** RNA secondary structure prediction, considering only PASs with a specific polyadenylation signal (AATAAA, AAGAAA, AATAGA, AATACA and AATTGA).

The nucleotide context of polyadenylation sites of *M. oryzae* differs from yeast

The nucleotide profile surrounding the cleavage site has been subject of study for many years [74]. The proteins composing the polyadenylation machinery are strongly conserved in eukaryotes, and the pre-mRNA features that determine the correct location of the PAS are also conserved in many species [70]. The central A-rich motif AAUAAA, also called the polyadenylation signal, was the first to be identified and later shown to require flanking, auxiliary elements for both 3'-end cleavage and polyadenylation of pre-mRNA as well as to promote downstream transcriptional termination [70]. In yeast, several components of the polyadenylation machinery have been described, and the nucleotide context of a generic PAS has been defined [214]. The polyadenylation signal motif (PE) is not well conserved across the whole set of yeast PASs, and many variations of the canonical AAUAAA motifs are allowed [215,216]. Rna15, the yeast homologue of the human CstF64, recognizes the A-rich PE located in the upstream region with the help of Rna14 [217,218]. Upstream of the polyadenylation signal there is a AU-rich region, called the efficiency element (EE), which interacts with the CFIB component [219]. The last two elements, called the positioning elements, are U-rich sequences that direct cleavage to a location approximately 20 nucleotides downstream (DUE) and upstream (UUE) of the PAS [74].

The organization of the region in *M. oryzae* differs slightly from that of *S. cerevisiae* (Fig. 3.1.1C-D). The polyadenylation signal motif is even less uniform than in yeast. The canonical AATAAA motif omnipresent in vertebrates is still the most represented but it accounts for only 7% of the PASs. Other alternative A-rich motifs are AATACA, AAGAAA and AATAGA. The upstream U-rich element (UUE) found in yeast and other species is clearly visible. On the other hand, the downstream U-rich element (DUE) found in yeast is less intense in *M. oryzae*. Moreover, the efficiency element (EE) is totally absent in *M. oryzae* (Suppl. Fig. 3.1.S1C). Inspecting more closely the nucleotide profile in *M. oryzae*, we identified two other specific motifs, UGUAAH and UAGNH, located upstream of the A-rich region, at -50 bp and -35 bp respectively (Suppl. Fig. 3.1.S1D). Finally, *M. oryzae* polyadenylation shows a preference for SA dinucleotides as a cut-site position, rather than YA dinucleotides observed in *S. cerevisiae*.

The RNA secondary structure of the polyadenylation site region has a defined configuration

The limitations of RNA secondary structure software prediction and the complexity of the polyadenylation machinery make it difficult to shed light on the polyadenylation region's secondary structure. The mechanisms regulating the polyadenylation event that depend on RNA structure, and the functional implication of this, are currently unknown, despite a history of knowledge that RNA secondary structure has an effect on polyadenylation site selection [220–222]. Genome-wide profiling of RNA secondary structure revealed novel regulatory features in *A. thaliana* [223], *S. cerevisiae* and human [224], where many of the PASs are predicted to adopt a similar structure. We analyzed our data and predicted the RNA secondary structure of the region surrounding the cleavage site (-40 nt upstream and +20 nt downstream) for every PAS identified in the previous section. Using the RNAfold tool from the Vienna package [225] we predicted the minimum folding energy (MFE) for every PAS region and compared it against a random control set. We noticed that the PAS region is significantly less structured (*Student's t test p-value* = $1.608E-015$). We also calculated base-pair probabilities for every nucleotide and plotted the average values against the random control set (Fig. 3.1.1E). We detected that the nucleotide preceding the cut-site has higher probability to be base-paired, while the nucleotide after it is generally not base-paired. The DUE and UUE elements tend to be base-paired, while the PE region containing the polyadenylation signal appears to be usually not base-paired. The reason behind this might be to locate the cut-site in a hairpin-like structure in order to allow a precise selection from the cleavage protein complex, while leaving the polyadenylation signal more accessible to be recognised by the complex.

A deeper analysis showed significant differences that were dependent on the preferred cleavage site. For example, the most common dinucleotide CA has a much lower base-pair probability compared to GA or TA (Fig. 3.1.1F). The predicted folding also seems to depend on

the specific polyadenylation signal. The canonical AAUAAA hexamer is associated with the loosest secondary structure, while other variants like AUAAAA, AAUACA and AAGAAA, are associated with increasingly more base-paired structures. Other hexamers display distinctly different folding predictions (Fig. 3.1.1G).

Lack of Rbp35 alters significantly the number of genes induced under all nutritional conditions

We carried out a differential expression study in four growing conditions: complete medium (CM), minimal medium (MM), nitrogen-starvation (MM-N) and carbon-starvation (MM-C). The analysis was performed with the wild-type and the two mutants (only CM and MM-C for $\Delta hrp1$). The highest number of differentially expressed genes in the wild-type occurs between CM and MM-C, with more than 4,500 genes involved (Fig. 3.1.2A). A change of growth medium for the mutant strains does not activate the same differential gene expression as observed in the wild-type. In fact, the number of differentially expressed genes in $\Delta rbp35$ between CM and other minimal media is clearly less than the wild-type, especially for the medium change from MM to MM-N, which does not produce any significant modification in genes expression in $\Delta rbp35$. This phenomenon is also observed during the *in vitro* growth on the fungus, where the $\Delta rbp35$ growth does not appear to be influenced by a nitrogen starvation medium [172]. Conversely, $\Delta hrp1$ displays a higher number of differentially expressed genes, at least for the only condition tested (CM $\rightarrow$ MM-C). These results were confirmed by the functional enrichment analysis (Fig. 3.1.2B-C), where we observed that enriched functional groups down/up-regulated in the wild-type during carbon starvation (compared to CM) like translation, ribosome biogenesis, mitochondrion, electron transport were not likewise down-regulated in $\Delta rbp35$. $\Delta hrp1$, in contrast, behaved similarly to the wild-type.

Notable differences were observed in the mutants when compared to the wild-type on the same medium (Fig. 3.1.2D). Disruption of Rbp35 results in the differential expression of a more limited number of genes when shifting to different growth media. Between wild-type and mutant, only 384 genes are differentially expressed in CM, 800 in MM, 386 in MM-N and 233 in MM-C. Disruption of Hrp1 on the other hand results in a larger number of differentially expressed genes in the growth media shifts, specifically, 2,580 in CM and 5,174 in MM-C. Regardless, for both $\Delta rbp35$ and $\Delta hrp1$ we did not find any enriched functional groups in CM, indicating that the effect of these mutations on gene expression is not specific to one overall function in the organism. However, in carbon starvation we were able to observe the enrichment of some functional groups, at least for $\Delta hrp1$, mainly related with protein/ATP/DNA/RNA binding, molecules transport and transcriptional control (Fig. 3.1.2E-F).

Figure 3.1.2 | Lack of Rbp35 significantly alters the number of genes induced under all nutritional conditions. **A)** Number of differentially expressed genes in each medium growth in the wild-type, $\Delta rbp35$ and $\Delta hrp1$ strains.* **B)** Enriched functional groups in down-regulated genes during carbon starvation (CM $\rightarrow$ MM-C only). **C)** Enriched functional groups in up-regulated genes during carbon starvation (CM $\rightarrow$ MM-C only). **D)** Number of differentially expressed genes in $\Delta rbp35$ and $\Delta hrp1$ strains, compared to the same medium growth condition in the wild-type.* **E)** Enriched functional groups in down-regulated genes in $\Delta rbp35$ (CM only). **F)** Enriched functional groups in down-regulated genes in $\Delta hrp1$ (CM only). * Conditions that were not tested are indicated as “NA”.

A

	DOWN			UP		
	WT	$\Delta rbp35$	$\Delta hrp1$	WT	$\Delta rbp35$	$\Delta hrp1$
CM $\rightarrow$ MM	314	508	NA	559	405	NA
CM $\rightarrow$ MM-N	630	461	NA	874	404	NA
CM $\rightarrow$ MM-C	2307	1241	2706	2342	1136	2571
MM $\rightarrow$ MM-N	48	0	NA	59	0	NA
MM $\rightarrow$ MM-C	1882	475	NA	1589	493	NA

B

C

D

	DOWN				UP			
	CM	MM	MM-N	MM-C	CM	MM	MM-N	MM-C
WT $\rightarrow$ $\Delta rbp35$	198	486	265	131	186	314	121	102
WT $\rightarrow$ $\Delta hrp1$	1276	NA	NA	2765	1304	NA	NA	2409

E

F

Transcripts associated with carbon starvation display longer 3'UTRs, with a preference for PASs depleted in the UUE element

In order to understand how nutritional stress affects PAS selection, we compared PAS usage levels between the tested growth conditions in the wild-type (Fig. 3.1.3A). During the switch from CM to MM-C, 1,067 genes (involving 2,032 PASs) display differential PAS selection (DPS), 80% of them preferring distal PASs in MM-C and 20% of them preferring more proximal PASs (Fig. 3.1.3B). The majority of DPS events occur between PASs within the 3'UTR, thus carbon starvation appears to induce the transcription of longer 3'UTR isoforms. The length of 3'UTR gained (or lost in the remaining 20% of genes with shorter 3'UTR) ranges from a few nucleotides to more than 2,000 nt (Fig. 3.1.3B). Some important virulence-related genes exhibit longer 3'UTRs in MM-C, including cell cycle control protein *cwf14* (MGG_06309), zinc finger protein *GIS2* (MGG_13778), a 14-3-3 family protein (MGG_13806) and the developmental regulator *VelB* (MGG_01620). Functional annotations enriched for genes associated with DPS in carbon-starvation are: protein binding, G-protein mediated signal transduction and small GTPase mediated signal transduction (Fig. 3.1.3C). Curiously, in a small number of cases, MM-C affects PAS selection in gene regions other than the 3'UTR, with most of these being sites within the 5'UTR. The effect of this kind of 5'UTR processing is unknown, and it could be related with regulation of translation [226]. A few genes that exhibit this kind of DPS are uricase (MGG_05283), an N amino acid transport system protein (MGG_06782), a glycosyl hydrolase (MGG_08616), the Fe-S cluster assembly protein *DRE2* (MGG_00776) and Rbp35 itself (MGG_02741).

A close inspection of the nucleotide profile surrounding the cleavage site of PAS regulated by carbon-starvation revealed that it slightly differs from the other unaffected PASs (Fig. 3.1.3D). The Upstream U-rich Element (UUE) located between the cleavage site and the polyadenylation signal is enriched in PAS down-selected during carbon starvation. In yeast, the UUE interacts with the Yth1 protein of the Polyadenylation Factor I complex (PFI). Expression of the rice blast ortholog of YTH1, MGG_00680, is down-regulated during carbon starvation. The reduced expression of this polyadenylation protein during carbon starvation could be related with the change in preference for PAS selection.

Figure 3.1.3 | Transcripts associated with carbon starvation display longer 3'UTRs, with a preference for PASs depleted in the UUE motif. A) Number of genes with DPS in each medium change, for both wild-type and mutant. Carbon starvation provides the highest number of genes with differential PAS selection (DPS). **B)** Distal PASs are generally preferred in carbon starvation compared to complete medium*. **C)** Enrichment analysis of functional groups for genes with DPS in carbon starvation. **D)** The U-rich Upstream Element (UUE) is enriched in down-selected PASs in carbon starvation.

*The preference for proximal or distal PASs is represented using the differential selection fold-change occurring on the distal PAS of affected genes (y axis). A positive fold-change means preference for distal PASs, a negative fold-change means preference for proximal PASs. The corresponding 3'UTR length is plotted in the "x" axis.

Rbp35 is required for correct PAS selection in ~37% of the genes, and proximal PASs are preferred in $\Delta rbp35$. The UGUAAH motif is enriched in PASs regulated by Rbp35.

In order to elucidate Rbp35 function in the rice blast fungus, and to understand how each PAS is affected in $\Delta rbp35$, we analyzed DPS between wild-type and mutant. 37% of all the expressed genes exhibit DPS in the $\Delta rbp35$ mutant, in any growth condition (Fig. 3.1.4A). Almost all the affected PASs are located in the 3' UTR. There is no notable difference in the number of genes exhibiting DPS between growth conditions. These results suggest that Rbp35 is deeply involved in the regulation of APA in *M. oryzae*. Interestingly, of the 422 known genes with virulence-related functions, 110 exhibit DPS in $\Delta rbp35$.

Nearly 75% of transcripts that display DPS in $\Delta rbp35$ also exhibit a preference for more proximal PAS usage compared to the wild-type, resulting in a global reduction in 3' UTR length (Fig. 3.1.4B), an observation opposite to that made previously with respect to carbon starvation. This 3' UTR length variation can range from a few nucleotides to more than 1,000 bp, but the majority of 3' UTR isoforms differ no more than 100 bp. This 3' UTR shortening has been observed in other organisms where some of their CFI protein components had been knocked-down [206,227]. With the purpose of clarifying what kind of biological processes are controlled by Rbp35 in *M. oryzae*, we looked at the enriched functional groups in genes with DPS in $\Delta rbp35$. Similar to generic APA genes, the enriched terms are protein binding/transport/targeting, cell polarity, transcriptional control and cell polarity (Fig. 3.1.4C); suggesting that Rbp35 could be a major regulator of APA in signaling-associated genes.

In metazoan, the cleavage factor Im (CFIm), consists of a 25 kDa subunit (CFIm25) and one of the three larger subunits (CFIm59, CFIm68, CFIm72) (Kim *et al.*, 2010). In human, it has been shown that the CFIm complex recognizes the upstream sequence of the PAS in a sequence-dependent manner. The RNA-bound structure obtained through X-ray crystallography shows that two UGUAA RNA sequences, with anti-parallel orientation, bind to one CFIm25-CFIm68 RRM heterotetramer [228]. We expect a similar behavior in *M. oryzae*, since Rbp35 has been found to interact with CFIm25 and it is considered to be the functional homologue of CFIm68 [172]. We performed a motif discovery analysis on nucleotide profile surrounding the cleavage site of PAS regulated by Rbp35. We found that the UGUAAH motif is strongly enriched in down-selected PAS in $\Delta rbp35$ (*p-value*: 6.6e-108 Fig. 3.1.4D). This could mean that, in *M. oryzae*, the recognition of a proper UGUA-like motif by the CFIm complex is essential for the cleavage reaction as it is seen in other species. In fact, the absence of Rbp35 in the mutant results in the down-selection of UGUAAH-associated PASs and to the up-selection of alternative and usually proximal PASs. These proximal PASs exhibit a distinct profile for the UGUAAH motif,

which in these cases tend to appear closer to the cleavage site than in the 'canonical' profile (Fig. 3.1.4D).

Figure 3.1.4 | Rbp35 is required for correct PAS selection in ~37% of the genes, and proximal PASs enriched in the UGUAH motif are preferred in Δ rbp35. A) Number of genes with DPS in Δ rbp35, different growth conditions have the same level of DPS. **B)** Proximal PASs are generally preferred in Δ rbp35 compared to wild-type (WT). The preference for proximal or distal PASs is represented using the differential selection fold-change occurring on the distal PAS of affected genes (y axis). A positive fold-change means preference for distal PASs, a negative fold-change means preference for proximal PASs. The corresponding 3'UTR length is plotted in the x axis. **C)** Enrichment analysis of functional groups for genes with DPS in Δ rbp35. **D)** The UGUAH motif is enriched in down-selected PASs in the Δ rbp35 mutant, compared to unaffected PASs, suggesting that this element is the possible binding site for the Rbp35 protein. Up-selected PASs also show an enrichment of the UGUAH motif, but closer to the cut-site.

Hrp1 is required for correct PAS selection in ~52% of genes, with no preference for proximal or distal PASs. The UAGNH motif is enriched in PASs regulated by Hrp1

In order to understand how each PAS selection is influenced by Hrp1, we performed DPS analysis on the $\Delta hrp1$ mutant. Roughly 65% of the PASs, corresponding to ~52% of all the expressed genes, suffer DPS in the $\Delta hrp1$ mutant (Fig. 3.1.5A). Comparing this result with $\Delta rbp35$ and carbon starvation, it is clear that Hrp1 plays a greater role in APA regulation. As discussed earlier, lack of Rbp35 affects DPS with a preference for proximal PASs. In terms of *in vivo* cellular function this suggests that Rbp35 is involved in selection of distal alternative PASs when specific conditions are met. This situation does not apply in $\Delta hrp1$, where no clear preference is observed, and either a distal or proximal PAS can be preferred compared to the wild-type (Fig. 3.1.5B). To better clarify what kind of biological processes are controlled by Hrp1 we looked at the functional annotations enriched for genes exhibiting DPS in $\Delta hrp1$ in CM. Similar to Rbp35, the term enrichment overlapped with those identified in generic APA genes (Fig. 3.1.5C).

Yeast Hrp1 recognizes an AU-rich region upstream of the cleavage site generally referred as Enhancer Element (EE). Bioinformatics analysis, RNA binding assays and crystallography have determined UAYRUA as the consensus motif for the Hrp1 binding site [219]. However, in *M. oryzae* this consensus sequence is not enriched as in *S. cerevisiae* (Suppl. Fig. 3.1.S1C), suggesting that *M. oryzae* Hrp1 could recognize another region. We performed a motif discovery analysis on nucleotide profile surrounding the cleavage site of PASs regulated by Hrp1. We found out that a UAGNH motif is strongly enriched in down-selected PASs in $\Delta hrp1$ (*p-value* 1.5e-191; Fig. 3.1.5D). The absence of Hrp1 in the mutant brings to the down-selection of UAGNH-enriched PASs and to the up-selection of alternative PASs. The motif map of these up-regulated PASs is distinct from the average PAS, in that the UAGNH motif tends to appear closer to the cleavage site in these loci, a situation similar to the UGUAAH motif in $\Delta rbp35$.

Figure 3.1.5 | Hrp1 is required for correct PAS selection in ~52% of genes, and the UAGNH motif is enriched in PASs regulated by Hrp1. **A)** Number of genes with DPS in $\Delta hrp1$, showing that different growth conditions have similar levels of DPS. **B)** Both proximal and distal PASs can be preferred in $\Delta hrp1$ compared to wild-type. The preference for proximal or distal PASs is represented using the differential selection fold-change occurring on the distal PAS of affected genes (y axis). A positive fold-change means preference for distal PASs, a negative fold-change means preference for proximal PASs. The corresponding 3'UTR length is plotted in the x axis. **C)** Enrichment analysis of functional groups for genes with DPS in $\Delta hrp1$. **D)** The UAGNH motif is enriched in down-selected PASs in the $\Delta hrp1$ mutant, compared to unaffected PASs, suggesting that this element is the possible binding site for the Hrp1 protein. Up-selected PASs also show an enrichment of the UAGNH motif, but closer to the cut-site.

3'-end processing and polyadenylation of transposable elements, ncRNAs and MOLV1 virus are also regulated under carbon starvation, Rbp35 and Hrp1

The sequencing of *M. oryzae* genome revealed that transposable elements (TEs) represent more than 5% of its total composition [18]. The overall repeat content varies depending on the isolate [1]. TEs found in the rice field isolate Guy11 are MAGGY [229], Inago1/2 [230] MGLR-3 [231], MINE [232], MGL [233], POT2/3 [234], Pyret [235] and retro5 [236,237]. Guy11 is also known to be infected by a ssRNA virus called MOLV1 [26]. We compared the expression in the wild-type, $\Delta hrp1$ and $\Delta rbp35$ of different repeated elements and ncRNAs (Fig. 3.1.6A), and found that transcription of rDNA, tRNAs and many TEs are up-regulated in the $\Delta rbp35$ mutant (especially Inago1, and MGLR-3), while the MOLV1 virus is dramatically down-regulated. On the other hand, $\Delta hrp1$ seems to display a slight down-regulation of some TEs, and the possible up-regulation of MOLV1. Therefore, deletion of these two genes generate opposite effects on the expression of *M. oryzae* repeated elements and ncRNAs.

We also checked for possible DPS with respect to these repeated elements. As mentioned earlier, DPS in *M. oryzae* can be due to carbon starvation, lack of Rbp35, or lack of Hrp1. Interestingly, we identified different DPS profiles for the repeated elements under these situations (Fig. 3.1.6B). All the examined elements present DPS in the $\Delta rbp35$ mutant. The majority of TEs also show DPS in the $\Delta hrp1$ mutant (but not rDNA and MOLV1); while in carbon starvation DPS is active on rDNA, MOLV1, tRNAs, Inago1 and MINE. tRNAs showed DPS in all three situations (to be considered a “positive”, at least one type of tRNA must exhibit DPS). 3'-end processing and polyadenylation of tRNAs is a mechanism that is not completely understood; for many yeast ncRNAs, polyadenylation is a way of marking the RNA for degradation [238]. This kind of polyadenylation, however, is done in the nucleus by the TRAMP complex, which only adds a tail of ca. 4 nt long [114], a characteristic that it is not detectable by our sequencing protocol.

Figure 3.1.6 | 3'-end processing and polyadenylation of transposable elements (TEs), ncRNAs and MOLV1 virus are also regulated under carbon starvation, Rbp35 and Hrp1. A) Expression profiles of rDNA, MOLV1, tRNAs and TEs. rDNA and tRNAs are up-regulated in Δ rbp35; MOLV1 is strongly down-regulated in Δ rbp35 and slightly up-regulated in Δ hrp1. TEs are generally up-regulated in Δ rbp35 and down-regulated in Δ hrp1. **B)** DPS events detected in carbon starvation, Δ rbp35 and Δ hrp1. Δ rbp35 induces DPS in all repeated elements. Note that tRNAs are marked when at least one type of tRNA exhibits DPS.

PAS selection is altered during plant root infection

M. oryzae can infect several rice tissues including roots [239], and plant infection strategies by the rice blast fungus have been deeply studied [5]. To our knowledge, the effect of APA in fungal plant infection has not been studied before. The Sesma laboratory has conducted a dual RNA-seq experiment of *M. oryzae* infection in both rice leaf and root tissues. The main purpose of the experiment was to examine differential gene expression profiles, but APA was also examined. The library was constructed for a typical RNA-seq analysis and lacks the PAS coverage depth achieved in our earlier sequencing experiments. Nevertheless, through the analysis of differential 3'UTRs expression we were able to predict DPS during plant root infection (Fig. 3.1.7A). The plot represents the genes that undergo DPS during a time course of 72 hours from inoculation, compared to spores. It can be observed that plant infection is associated with a global preference for proximal PASs and a reduction of 3'UTRs length. A total of 1149 genes display DPS during root infection. The enrichment analysis of functional groups for genes with DPS during root infection returned results similar to those obtained for generic APA genes (Fig. 3.1.7B); suggesting a global relationship between pathogenicity and APA. We were unable to reproduce the same analysis on data retrieved from leaf infection, because libraries obtained from infected leaves contain too little fungal RNA and the sequencing coverage of UTRs was too low to generate reliable results.

Figure 3.1.7 PAS selection is altered during plant root infection: A) Proximal PASs are generally preferred during root infection compared to spores*. **B)** Enrichment analysis of functional groups for genes with DPS during root infection.

*The preference for proximal or distal PASs is represented using the differential selection fold-change occurring on the distal PAS of affected genes (y axis). A positive fold-change means preference for distal PASs, a negative fold-change means preference for proximal PASs. The corresponding 3' UTR length is plotted in the x axis.

***Δrbp35*, *Δhrp1*, MM-C and root infection regulate polyadenylation of similar gene groups but in different ways**

Carbon starvation, root infection and the lack of polyadenylation machinery subunits Rbp35 and Hrp1 can directly affect PAS selection. Hrp1 regulates the highest number of genes (~52%), more than Rbp35 (~37%), carbon starvation (~12%) and root infection (~15%). Rbp35 is probably involved in regulating distal PASs. However, the global trend in *Δhrp1* showed no preference for either proximal or distal PASs. As a whole, 5,853 genes are regulated by at least one of the four conditions considered (*Δrbp35*, *Δhrp1*, MM-C and root infection). The majority of genes that display DPS during root infection are also dependent on the three APA regulators seen before (MM-C, Rbp35 and Hrp1) (Fig. 3.1.8A), suggesting a strong relationship between APA regulation and pathogenicity.

An interesting question is whether or not each of the four conditions affects APA in a similar way, that is, if DPS is always for distal/proximal PASs or it can be the opposite (Fig. 3.1.8B). There is very little correlation in DPS between *Δrbp35* and *Δhrp1* (p-value = 0.00841), suggesting that, for the same gene, these two proteins regulate PAS selection through distinct mechanisms. *Δhrp1* and MM-C exhibit positive correlation (p-value = 2.075e-10). Interestingly, *Δrbp35* and MM-C exhibit a negative correlation (p-value = 2.825e-09), meaning that when in *Δrbp35* we observe a shortening of the 3' UTR, for the same gene in carbon starvation the 3' UTR gets longer. This suggests that Rbp35 plays a role in the lengthening of 3' UTRs of those transcripts exhibiting DPS in carbon-starved cells. DPS during root plant infection is slightly positively correlated with MM-C and negatively correlated with *Δrbp35* (p-value 0.04771 and 0.005467 respectively), indicating a possible involvement of Rbp35 in APA regulation during plant root infection. There is no apparent correlation between DPS in root infection and *Δhrp1* (p-value 0.5036), suggesting that Hrp1 is probably not involved in APA regulation during root infection.

A common feature between *Δhrp1*, *Δrbp35* and MM-C is found with respect to the A-rich region, where the polyadenylation signal is located. This element appears to be depleted in down-selected PAS in *Δhrp1*, *Δrbp35* and MM-C. Most specifically, A-rich hexamers with a consensus motif of AAUAHA are under-represented (Fig. 3.1.8C). This could suggest that both Rbp35 and Hrp1 are involved in the selection of non-'canonical' PASs (those which do not possess the 'canonical' polyadenylation signal), while the opposite applies for carbon starvation.

Figure 3.1.8 | $\Delta rbp35$, $\Delta hrp1$, MM-C and root infection regulate polyadenylation of similar gene groups but in different ways. **A)** Venn diagram displaying the genes with DPS in common between the four groups of interest. **B)** Correlation plot of DPS in the four groups of interest. An inverse correlation exists for DPS between $\Delta rbp35$ and carbon starvation, meaning that when in $\Delta rbp35$ we observe a shortening of the 3'UTR, for the same gene in carbon starvation the 3'UTR gets longer. Instead, DPS in $\Delta hrp1$ and carbon starvation are positively correlated. DPS during root infection appears to be slight positively correlated with carbon starvation and inversely correlated with $\Delta rbp35$. **C)** The AAUAHA motif is depleted in down-selected PASs in $\Delta hrp1$, $\Delta rbp35$ and MM-C, compared to always unaffected PASs, suggesting that all these conditions force the selection of non-'canonical' PASs.

Discussion

A tight regulation of APA is essential for cellular development. APA is also known to affect mRNA stability, translation and translocation [76]. In eukaryotic organisms like yeast, worm and metazoan, APA is regulated by some components of the polyadenylation machinery, depending on specific signals surrounding the cleavage site. CFI proteins play an important role in APA regulation [240]. In this study we provide evidence that Rbp35 and Hrp1, as well as nutritional stresses like carbon starvation, are involved in regulation of APA in the rice blast fungus.

Using a PAS-specific mapping protocol we were able to identify at least one PAS for more than 65% of annotated *M. oryzae* genes. Roughly 50% of them displayed a clear APA event in at least one of the growing conditions considered. APA is not randomly represented among gene functional groups; it is more frequently observed in genes encoding proteins with binding and localization functions. The nucleotide context of PASs in *M. oryzae* is slightly different from the ones described for *H. sapiens* and *S. cerevisiae*. Also, in *M. oryzae*, the surrounding region displays a specific predicted RNA structure. In $\Delta rbp35$ and $\Delta hrp1$ mutants, 37% and 52% of total expressed genes are affected in PAS selection, respectively, meaning that Rbp35 and Hrp1 are important APA regulators. The way that PAS selection is influenced in both mutants is notably different. In $\Delta rbp35$, the vast majority of 3' UTRs are shorter respect to the wild-type, which indicates that Rbp35 is involved in distal poly(A) selection, as already observed in human for its functional orthologue CFI_m68, despite their lack of domain conservation with the exception of the RRM domain. Conversely, there is no clear preference for proximal or distal PASs in $\Delta hrp1$. Our work also revealed that nutritional stresses can affect APA. Under carbon starvation, more than 1,000 genes undergo a different polyadenylation processing respect to CM, and usually, distal PASs are preferred, resulting in longer 3' UTRs.

A detailed inspection of the surrounding areas of mutant-regulated PASs allowed us to identify possible nucleotide regions recognized by Rbp35 and Hrp1. The UGUAAH motif located 45 nt upstream of the cleavage site is significantly enriched in PASs regulated by Rbp35. These results confirm previous studies where UGUA was identified as the putative binding site for CFI_m68 [241]. Regarding Hrp1, its ortholog in *S. cerevisiae* was shown to have affinity with regions displaying the consensus motif UAYRUA, but this motif is missing in *M. oryzae* and instead the motif UAGNH is significantly enriched in Hrp1-dependent PASs. Lastly, APA regulation during carbon starvation seems to be correlated with the UUE motif.

Differential expression analysis was also conducted in order to identify which genes were down- or up-regulated in both mutants with respect to wild-type. The transcriptional profile of $\Delta rbp35$ in MM and nitrogen-starved cells is practically the same, corroborating the growth defects seen on this mutant on different nitrogen sources [171,172], and suggesting that $\Delta rbp35$ has lost the ability to sense nitrogen sources. Conversely, $\Delta hrp1$ mutant shows the greatest

discrepancies with wild-type during carbon starvation. This somehow correlates with the fact that most genes differentially expressed in $\Delta hrp1$ are associated with translation and ribosomes, two functional groups that are notably down-regulated during carbon starvation growth.

Repetitive elements of endogenous (like rDNA and tRNAs) or exogenous (like TEs) origin are also differentially regulated in the two mutants, as it is the PASs selection for many of them, especially in $\Delta rbp35$. TEs are known to be a major tool for genetic rearrangement and pathogenic evolution in the rice blast fungus [19], which suggests that Rbp35 could also play an important role in this adaptive response of *M. oryzae*.

Finally, we provided significant evidence that differential PAS selection is active *in planta*, at least during root infection, which strengthen the hypothesis that APA can regulate plant pathogenicity.

Supplemental Data

Supplemental Figure 3.1.S1. A) Workflow describing the steps followed to analyze poly(A)-seq data, 14593 poly(A)sites were identified in *M.oryzae* transcriptome. **B)** Bar plot showing the gene location of the identified PASs, classified as 3' UTR, 5' UTR, coding sequence (CDS), intronic or ncRNA (if the PAS belongs to a non coding RNA). The majority of poly(A) site are located in the 3' UTR of expressed genes. **D)** Distribution of UAYRUA motifs in the region surrounding the PAS. **D)** Distribution of UGUA and UAGA motifs in the region surrounding the PAS.

Chapter 3.2

**small RNA sequencing provides new insights into
RNA interference in the rice blast fungus
*Magnaporthe oryzae***

Summary

RNA interference (RNAi or RNA silencing) is a conserved eukaryotic mechanism for gene regulation and cell defense, which works through the interoperation between a set of core proteins and small RNA guides like small interfering RNAs (siRNAs) and microRNAs (miRNAs). RNAi in animals and plants has been studied thoroughly and both kingdoms share many features. In fungi, however, the situation is less homogeneous and research is still in progress. Here, we present small RNA (sRNA) sequencing data of *M. oryzae* using a novel library construction approach. Our results indicate that almost 50% of sRNAs are produced from parasitic repetitive elements like retrotransposons and the MOLV1 virus. Our analysis also revealed that a pervasive post-transcriptional modification of unknown origin occurs on these kinds of sRNAs. We sequenced sRNAs from two mutants, $\Delta rbp35$ and $\Delta exp5$, involved in RNA polyadenylation and export, respectively. Our analysis suggests that *RBP35* and *EXP5* genes could be involved in repetitive element regulation, particularly *RBP35*, whose deletion clearly affects RNAi regulation.

Background

Small RNA (sRNA) is a generic term referring to a heterogeneous group of RNA sequences usually shorter than 200 nt, which are typically non-coding RNA (ncRNA) molecules. Common sRNAs found in eukaryotic cells are transfer RNA (tRNAs), small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), microRNAs (miRNAs), small interfering RNAs (siRNA) and Piwi-interacting RNAs (piRNAs). miRNAs and siRNAs are involved in RNA interference (RNAi, also known as RNA silencing) or Post-Transcriptional Gene Silencing (PTGS), a highly conserved biological response to exogenous and endogenous parasitic nucleic acids as well as to a refined gene silencing process in eukaryotes [242–244]. RNAi was first observed as an inhibitory effect in gene expression of exogenously introduced antisense RNAs [245–247]. The discovery in *C. elegans* of miRNAs, ~22 nucleotide long small ncRNAs molecules that function in RNA silencing and post-transcriptional regulation of gene expression, gained the Nobel prize in 2006 [248]. In general, small sRNAs are 20 - 30 nt long, mostly derived from double stranded RNA (dsRNA), and recognize and cleave mRNAs bearing homologous sequences [249–252].

siRNAs and miRNAs differ in two key aspects. First, siRNAs are derived from exogenous dsRNA like viral RNAs or endogenous transcripts from repetitive sequences like transposable elements (TEs), or transcripts that can form long hairpins [253]. In many organisms, siRNA-induced silencing requires RNA-dependent RNA polymerases (RdRPs) to generate dsRNA from single-stranded RNA (ssRNA) or to amplify sRNA signals [254]. miRNAs, on the other hand, are generated from miRNA-encoding genes that generate ssRNA precursor transcripts that form hairpin structures. Second, siRNAs normally fully match their mRNA targets. siRNAs trigger RNA cleavage or transcriptional silencing mediated by the Argonaute (Ago) proteins and generally are part of the genome defense mechanism. In contrast, miRNAs can target mRNAs that are not fully complementary and cause mRNA degradation and translational repression [255]. siRNA existence and function has been demonstrated in many fungal species [253], while the main components of the miRNA pathway are missing in fungal world. miRNA-like genes have been observed in *Neurospora crassa* [256] and similar elements were later found in other fungi like *Aspergillus flavus* [257], *Antrodia cinnamomea* [258], *Penicillium marneffe* [259], *Metarhizium anisopliae* [260], *Nosema ceranae* [261], *Cochliobolus lunata* [262], *Penicillium chrysogenum* [263], *Trichoderma reesei* [264], *Sclerotinia sclerotiorum* [265] and *Fusarium oxysporum* [266]. However, the cellular role and processing pathways for miRNA-like elements in fungi remain obscure [253].

Biogenesis of most siRNAs shares a minimal common machinery consisting of an RNase III endonuclease called Dicer that processes dsRNA precursors into sRNA (sRNA) molecules, and an Ago endonuclease that binds sRNAs and uses them as a guide to identify and cleave

complementary target mRNA [242,267]. Additionally, some RNAi-competent organisms, including plants, nematodes and fungi, require RdRPs to generate dsRNA from ssRNA inducers or to amplify siRNA signals. RNAi components (Dicer, Ago, and RdRP) have been identified in all major branches of eukaryotes [254,268] suggesting that RNAi is an ancient defense and/or regulatory mechanism that existed in the common eukaryotic ancestor. Ago homologs (but not other core RNAi components), are found in some eubacteria and archaea [269]. Evolution has shaped the RNAi pathway adapting to novel functions and mechanisms, leading to novel biogenesis in different eukaryotic organisms and lineages [244,268,270]. Not every eukaryotic organism possesses the basic RNAi components, including fungal model species like *S. cerevisiae* and *U. maydis*, indicating that the RNAi pathway is dispensable in some organisms.

RNAi in fungi was first found to be required for heterochromatin formation in the fission yeast *Schizosaccharomyces pombe* [271–276], and this system remains the best understood pathway of RNAi-mediated transcriptional gene silencing. *N. crassa* has also been a model organism for RNA interference since the discovery of the phenomenon called *quelling* [125]. This Ascomycete fungus possesses at least three homology-based defense systems. The first, quelling, occurs during the vegetative phase; the second, meiotic silencing of unpaired DNA (MSUD), occurs during the sexual phase [127]; the third defense mechanism is known as repeat-induced point mutation (RIP) and it is initiated during sexual phase before karyogamy [277]. Quelling and MSUD are based on RNAi. *N. crassa* is known to produce several types of sRNAs through different biogenesis pathways: DNA damage-induced QDE-2 associated siRNA (qiRNA), miRNA-like sRNA (milRNA) and Dicer-independent siRNA (disiRNA) [256]. RNAi was also initially proposed to explain gene silencing observed in *Mucor circinelloides*, the basal fungus that is a causal agent for the rare but lethal fungal infection mucormycosis, an emerging infectious disease recognized as a prevalent fungal infection in patients with impaired immunity [278]. In *M. circinelloides f. lusitanicus* researchers identified a phenomenon similar to *N. crassa* quelling, where the lack of the carotenogenesis pathway gene *crgA* and its introduction as transgene gene both causes carotene overaccumulation in the dark [279]. Gene silencing was later demonstrated in *M. circinelloides* by using a simple visual reporter system to analyze transgene-induced gene silencing [280]. Later studies demonstrated that exogenous and endogenous RNAi generated by Dicer 2 and RdRP regulate mRNAs and protein production [280–282].

Magnaporthe oryzae, also known as rice blast fungus is a plant-pathogenic fungus belonging to the Ascomycota family that infects rice and other monocots [5,6]. The *M. oryzae* genome was the first to be published for a plant pathogenic fungus, and its analysis revealed that transposable elements (TEs) represent more than 5% of its total composition [18]. Gene silencing mediated by hairpin RNA structures also takes place in *M. oryzae* [283]. The basic RNAi elements, RdRp, Dicer and Ago, are present in three, two and three copies, respectively.

One of the two Dicer-like proteins is responsible for hairpin RNA-triggered RNA silencing and related siRNA accumulation [284], while the function of the second Dicer protein is unclear. Later, it was suggested that transcriptional control and protein specialization have roles in the functional diversification of these two Dicer-like proteins, and that they could participate in two different RNAi pathways in *M. oryzae* [285]. Transcriptional data analysis under physiological stresses indicates that sRNAs play a role in transcriptional regulation for a small subset of genes [21], and that sRNAs differentially accumulate in vegetative and specialized-infection tissues and may play an active role in genome integrity and regulating growth and development [22]. However, the lack of these two genes does not affect fungal development or plant pathogenicity [284].

Previous works showed that deletion of Exp5, the rice blast orthologs of human Exportin-5 and yeast Msn5, resulted in reduced pathogenicity [170]. In mammals this protein participates in the nuclear import/export of proteins and RNA molecules, including miRNA precursors [286]. In yeast, Msn5 is involved in tRNAs re-export to the cytoplasm and can also bind dsRNAs *in vitro* [287,288]. From an evolutionary point of view, Exportin-5 orthologues are functionally divergent among species [289]. The pathogenicity defects found in *M. oryzae* $\Delta exp5$ mutant suggest a possible link between sRNAs regulation and plant infection that could be revealed by RNA sequencing of $\Delta exp5$ in *M. oryzae*.

Rbp35 is a component of the fungal polyadenylation machinery and functional ortholog of human CFIm68, involved in the regulation of alternative polyadenylation [172]. Its deletion results in defective pathogenicity and altered 3' UTR processing [171,172]. Immunoprecipitation experiments performed by the Sesma laboratory using Rbp35 revealed that one of the three *M. oryzae* Ago proteins (MGG_01294) precipitates together with Rbp35. Ago is the catalytic component of the RNA-induced silencing complex (RISC) [290]. The interaction between Rbp35 and Ago may indicate a possible involvement of Rbp35 in RNAi regulation.

In order to elucidate the role of Exp5 and Rbp35 in sRNA metabolism in *M. oryzae* we conducted the sRNA sequencing of $\Delta exp5$ and $\Delta rbp35$ mutants and compared them to the wild-type strain. We used a novel method to build the sRNA library that reduces ligation bias of sRNAs using HD adapters [183]. The data was analyzed with bioinformatic tools and further validated by Northern blotting.

Results

sRNAs derived from TEs and MOLV1 virus display RNA silencing features

sRNAs were sequenced from mycelia tissue from wild-type Guy11, and the mutants $\Delta exp5$ and $\Delta rbp35$ using three biological replicates. Between thirteen and forty million reads were obtained from all samples, which were filtered to remove adapter contamination, reads without the 3' adapter, and low quality reads. After filtering, the average percentage of unique reads was ~9% (Suppl. Fig. 3.2.S1A). Read length distribution ranged from 18 to 25 nucleotides, with a peak at 19-20 nucleotides (Fig. 3.2.1A). Reads were classified by aligning them against annotated nucleotide sequences, including tRNAs, rRNAs, TEs, protein coding genes and the previously identified *M. oryzae* MOLV1 virus [26]. It was already shown that RNAi is functional in *M. oryzae* [283], and that some TEs like MAGGY are suppressed through specific RNAi metabolic pathways [291]. Not all the annotated (TEs) display a notable level of sRNAs; only MAGGY [229], Inago1/2 [230], MGLR-3 [231], MINE [232] and retro5 [236,237] show a significant level of expression (> 1 reads per million). This suggests that different *M. oryzae* (TEs) may be silenced through different pathways. With respect to MOLV1, the presence of siRNAs derived from mycoviruses in filamentous fungi was already shown in *Aspergillus* [292], *C. parasitica* [293], *C. higginsianum* [294] as well as in *M. oryzae* for MoV2 virus [295].

To discriminate if these sRNAs derived from RNA silencing or from RNA degradation we checked the average read length and sense/antisense transcription levels for each of the selected elements. Reads mapping to tRNAs and rRNAs usually align sense (Fig. 3.2.1B), indicating that the library captured the RNA transcript itself or some product of its degradation. Reads mapping to TEs, mRNAs or MOLV1 display a higher degree of antisense transcription up to 50%. This suggests that a strong silencing effect is acting on these elements; since antisense transcription is usually considering a sign of silencing [296].

Nucleotide composition of sRNAs is also an element of interest. Uracil is the most common first nucleotide, and this feature is generally associated with sRNAs derived from TEs, mRNAs and MOLV1 (Fig. 3.2.1C). Previous studies showed that the 5' terminal nucleotide of regulatory sRNAs is important for Dicer efficiency [297] and it is critical for the sorting and recruitment to Ago [298]. Immunoprecipitation of fungal Ago proteins frequently correlated with a great abundance of bound sRNAs starting with U [256,294]. A strong bias is also evident in the 3' nucleotide; TEs, mRNAs and MOLV1 present a visible preference for adenine. rRNAs generally present distinct bias for first and last nucleotide, with a preference for adenine and cytosine, respectively.

Figure 3.2.1 | sRNAs deriving from (TEs) and MOLV1 virus display signs of RNA silencing. A) Read length distribution for the three strains analyzed: wild-type (WT), $\Delta exp5$ and $\Delta rbp35$. The typical range is 18-24 nt, with a peak at 20 nt in the WT. There is a strong presence of 14 nt long reads, which disappears in $\Delta exp5$. In $\Delta rbp35$ we observe a decrease of reads with length 21 nt, 22nt and 23 nt, and an increase of reads with length 19 nt and 20 nt. **B)** Proportion of sense and antisense read alignment. sRNAs derived from TEs, mRNAs and MOLV1 virus present a higher level of antisense transcription. **C)** Read's first nucleotide frequency: uracil is the most common first nucleotide, generally associated with TEs, mRNAs and MOLV1. **D)** Read's last nucleotide frequency: adenine is the most common last nucleotide. Adenine as last nucleotide is generally associated with sRNAs derived from transposable elements (TEs), protein coding genes (mRNAs) and MOLV1 virus

The presence of 3'-end adenines in many sRNAs suggests a post-transcriptional modification event

Reads were aligned with no mismatches allowed against the complete genome of *M. oryzae* 70-15 [18], downloaded from the Ensembl Fungi website, annotation version 29 (<http://fungi.ensembl.org/>) (MG8). More than 70% of unique reads did not align on the published genome, meaning that they either come from degenerated copies of repetitive elements that do not appear on the assembled genome (or they are specific to Guy11 strain used in this experiment), or they arise from non-genomic nucleotide sequences. In the previous section we showed that reads display a strong bias for 3' adenines. Considering that polyadenylation is a common post-transcriptional modification performed by some protein complexes of the polyadenylation machinery [299] and the TRAMP complex [300], we investigated this issue more deeply. sRNAs were tested for possible 3'/5'-end post-transcriptional additions: reads were trimmed on both ends for each of the four nucleotides and tentatively aligned again on the genome. 3'-end trimming of adenines resulted in the successful alignment of 20% of previously unmapped reads. This is a compelling evidence that 3'-end adenines could be the result of post-transcriptional modifications (Fig. 3.2.2A). These oligo(A) ends are comprised only of one or two adenines in most cases. 5'-end trimming did not result in any improvement in alignment. Read length distribution of adenylated sRNAs is restricted to the range 17-23 (without the terminal As), meaning that this chemical modification occurs only on a specific subset of sRNAs (Fig. 3.2.2B). To understand which sRNAs were potential targets of adenylation we examined the types of elements showing improved read mapping after 3' adenine trimming (Fig. 3.2.2C). It was evident that TEs and MOLV1 showed dramatic increases in reads alignment. Globally, more than 30% of previously unmappable reads could be successfully aligned to some element after A-trimming, indicating that adenylation is a pervasive post-transcriptional modification, usually applied to sRNAs deriving from endogenous and exogenous invasive nucleic acids (which are also the preferred targets of RNAi).

The cause of this phenomenon is unknown. Apart from the typical mRNA polyadenylation, adenylation of sRNA in fungi has been described before, but only for tRNAs, rRNAs and snRNAs [301–304]. Evolutionary conservation of post-transcriptional 3'-end adenylation of sRNAs was previously discovered in *S. cerevisiae* and *Xenopus* signal recognition particle (SRP) RNA and U2 snRNA [305], which contain a post-transcriptionally added adenylic acid residue on their 3'-ends. Adenylation of plant miRNAs was observed in *Populus trichocarpa*, in the range of 1-7 long oligo(A), also indicating that adenylation is associated with the 3' to 5' miRNA degradation where the addition of adenylic acid residues on the 3'-end plays a negative role in miRNA degradation [306]. Adenylation of miRNA-like was also detected in *N. crassa* [307].

To confirm the evolutionary conservation of adenylation in other organisms, we inspected sRNA sequencing data of other fungal species. Generally, removal of a trailing A does not increase mapping success in common total sRNA libraries. However, when applied to RNA libraries obtained from Ago immunoprecipitation, the improvement is evident for libraries built from *S. pombe* [308], *N. crassa* [256] and *C. higginsianum* [294]. In each case, the mapping improvement is about 10%.

Figure 3.2.2 | The presence of 3'-end adenines in many sRNAs suggest a post-transcriptional modification event. A) Reads that failed to align on the reference *M. oryzae* genome were trimmed on their 3'-end for each of the four possible nucleotides and tentatively mapped again on the genome. 3' trimming of adenines resulted in the successful alignment of 20% of the previously non-aligned reads, suggesting that this is a likely post-transcriptional modification, composed usually of one or two adenines. **B)** Reads modified with an adenine are usually smaller than the background length distribution, suggesting a possible relationship with RNA silencing (see results). **C)** Read classification: After 3'-end A-trimming more than 30% of previously unknown reads could be successfully aligned to some sRNA group, mostly to TEs and MOLV1, suggesting that these two groups are the main targets of this kind of post-transcriptional modification. *other: sRNAs derived from introns, intergenic regions, and sRNAs that align to unclassified ESTs in the genbank database.

Expression of sRNAs derived from some TEs is slightly altered in $\Delta rbp35$ and $\Delta exp5$

The previous section showed that a large amount of sRNAs sequenced from the wild-type strain Guy11 derived from TEs and MOLV1 virus. The sRNA-seq data from the mutants were also analyzed and compared to the wild-type. We observed that globally, in $\Delta rbp35$ sRNAs derived from tRNAs, rRNAs and protein coding genes are up-regulated compare to the wild-type. In the $\Delta exp5$ mutant we observe the up-regulation of TEs and MOLV1 (Fig. 3.2.3A). A closer inspection of rDNA-derived sRNAs confirms that a global up-regulation occurs in $\Delta rbp35$ (Fig. 3.2.3B) while in a slight down-regulation is present in $\Delta exp5$ (Fig. 3.2.3C). Antisense transcription from rDNA is usually negligible, but surprisingly a strong induction of sRNAs derived from the Internal Transcribed Spacer 2 (ITS2) takes place in $\Delta rbp35$. The role of this kind of antisense transcription is unknown, but it could be involved in the regulation of rRNA itself. A miRNA and Argonaute-loaded sRNAs derived from ITS1 were previously found in *Drosophila* [309]. The most evident change in $\Delta exp5$ is related with ribosomal 5S, which goes from highly expressed to almost no expression.

Specific TEs were also analyzed (Fig. 3.2.3D). MAGGY and retro5 are up-regulated in $\Delta exp5$, while MGL [233], MINE [232], POT3 [234] and MgSINE [310] are up-regulated in $\Delta rbp35$. For each of these elements, both sense and antisense transcription was detected, but with no significant variation between wild-type and mutants. The fact that different TEs were found to be affected in each mutant could indicate that Exp5 and Rbp35 may be involved in different silencing pathways. MOLV1 expression was also examined in $\Delta rbp35$ and $\Delta exp5$. In $\Delta rbp35$ sRNAs derived from the virus are generally down-regulated from both the sense and antisense strand (Fig. 3.2.3F). Conversely, more sRNAs are produced from the sense strand and fewer from the antisense strand in $\Delta exp5$ (Fig. 3.2.3G). These opposite effects between the two mutants seem to support further the hypothesis that different silencing pathways are regulated by the two proteins.

About 5% of total sRNAs derived from protein-coding genes. These genes are therefore a potential target of mRNA silencing and degradation. As a whole, 1,424 annotated protein-coding genes display an average sRNA expression greater than 1 read per million in the wild-type. The FunCat enrichment analysis of this subset of genes indicates that they are related with biological processes like protein/ATP/nucleotide binding, phosphate metabolism, stress response, translation, cell growth/cell wall, signalling and reproduction (Suppl. Fig. 3.2.S1B). Consequently, RNA silencing acting on mRNAs in *M. oryzae* cells may help to regulate a variety of aspects of cell metabolism.

Figure 3.2.3 | Expression of sRNAs derived from some repetitive elements is altered in $\Delta rbp35$ and $\Delta exp5$. **A)** Expression levels of sRNAs from selected elements between the wild-type and the two mutants. **B)** Expression profile of the rDNA cassette. Downwards-pointing curves represent the negative strand. **C)** Expression levels of common *M. oryzae* transposable elements (see results for discussion). **D)** The expression profile of MOLV1. Downwards-pointing curves represent the negative strand. The expression is represented as average between the three replicates.

Typical Argonaute-loaded sRNAs are up-regulated in $\Delta rbp35$

As observed previously, sRNA-derived reads displayed a defined preference for uracil as first nucleotide and adenine as last nucleotide. sRNA 5' base is already known to be related with RNAi, since Ago proteins generally prefer guide sRNAs to target RNA transcripts where the uracil-beginning guide RNA matches with an adenine in the target RNA sequence [298,311,312]. In the wild-type, we detected that U-starting reads represent ~37% of the total. Lack of Rbp35 and Exp5 seems to be related with this class of sRNAs (Fig. 3.2.4A). Up-regulated sRNAs in $\Delta rbp35$ possess a stronger preference for a 5' uracil-end compared to the wild-type, with ~80% of them starting with U. Similar observations were made in $\Delta exp5$, but to a lesser degree, where the percentage of up-regulated sRNAs with a 5' uracil-end is ~50%. The greatest majority of these U-starting reads in $\Delta rbp35$ align to TEs, while in $\Delta exp5$ they align to MOLV1 (Fig. 3.2.4B), indicating that the absence of these proteins could be associated with an increase of RNAi acting on different elements.

Formerly, we presented several clues suggesting that the preference for adenine in *M. oryzae* sRNAs is presumably the result of some unknown type of post-transcriptional modification event. Analyzing differential expression in the mutants could help to elucidate if Rbp35 and Exp5 are involved in the regulation of this cellular mechanism. In the wild-type more than 50% of the reads end with adenine, while sRNAs up-regulated in $\Delta rbp35$ and $\Delta exp5$ present a 3' adenine in ~30% and ~65% of the total, respectively (Fig. 3.2.4C). Of the sRNAs ending with an adenine we checked which ones could be the result of a post-transcriptional modification aligning them against annotated features (Fig. 3.2.4D). In the wild-type, roughly 20% of adenine 3' endings are the possible result of this modification. Quite surprisingly, even if up-regulated sRNAs in $\Delta rbp35$ do not exhibit a strong preference for a 3'-end adenine, the majority of them are in fact modified. Conversely, for down-regulated sRNAs the trend is similar to the wild-type. A different situation was observed in $\Delta exp5$, where both induced and depleted sRNAs are generally adenylated, at least to a greater degree on average compared to wild-type.

Figure 3.2.4 | sRNAs differentially expressed in the mutants possess specific nucleotide preferences. **A)** sRNAs up-regulated in $\Delta rbp35$ present a strong preference for uracil as first nucleotide, even higher than the common uracil frequency in the wild-type. This could suggest that Rbp35 may be involved in the regulation of argonaute-associated sRNAs. **B)** Up-regulated sRNAs with a 5' uracil in $\Delta rbp35$ are usually derived from transposable elements, and in minor part from MOLV1 virus. **C)** sRNAs up-regulated in $\Delta rbp35$ do not display the typical preference for adenine as last nucleotide. On the other hand, sRNAs up-regulated in $\Delta exp5$ have a slight higher preference for adenines than wild-type. **D)** Even if up-regulated sRNAs in $\Delta rbp35$ usually do not end with A, the majority of them are post-transcriptionally modified, while in down-regulated RNAs the trend is similar to the wild-type. In $\Delta exp5$ both induced and depleted sRNAs are generally adenylated, at least to a higher degree respect to the average in the wild-type.

Deletion of Rbp35 and Exp5 affects sRNAs of a particular size

The sRNAs sequencing from *M. oryzae* produced reads in the range of 18 - 24 nt (Fig. 3.2.1A). It is noticeable the disappearance of a 14 nt peak from read-lengths in $\Delta exp5$, and an increase in reads with length 19 nt and 20 nt in $\Delta rbp35$. To further explore this, reads for each length were separated into specific RNA classes. sRNAs derived from rRNAs are notably affected in $\Delta exp5$; short reads ~14 nt long are down-regulated in $\Delta exp5$, accounting for the reduced 14 nt read-length peak (Fig. 3.2.5A). These 14 nt long reads are almost entirely constituted by the sequence 5'-ACATACGACCATAC-3', aligning to the 5'end of *M. oryzae* ribosomal 5S. 5S rRNA, like tRNAs, are transcribed by RNA Pol III [313]. It is unclear what biological reason lies behind the abundance of this specific 5S rRNA, it is likely the subproduct of a degradation process which could involve the role of Exp5. Overall, the uniformity of read length from rRNA is an indicator that we are observing the products of degradation.

Distribution of reads length in TEs appears to be anomalous in $\Delta rbp35$ (Fig. 3.2.5B). Reads of length 19 nt and 20 nt are generally increased at the expenses of reads 21-24 nt long. Together with the up-regulation of U-starting reads from TEs in $\Delta rbp35$, we could suppose that the deletion of Rbp35 may lead to a general intensification of RNA silencing acting on these TEs. Reads derived from protein coding genes display a mixed behaviour, with a down-regulation of very short reads (as in $\Delta exp5$) and an up-regulation of 19 nt and 20 nt long reads (as in $\Delta rbp35$) (Fig. 3.2.5C). This could be suggesting that several sRNA pathways may be regulating mRNA transcripts in rice blast fungus. Reads aligning to MOLV1 are affected in $\Delta rbp35$ similarly to TEs (Fig. 3.2.5D). There is an enhancement in the number of 19 nt reads, as in TEs, but no for 20 nt reads. This phenomenon could be the result of different silencing pathway acting on MOLV1 compared to TEs.

Figure 3.2.5 | Deletion of Rbp35 and Exp5 affects sRNAs of a particular size. A) sRNAs derived from rRNA are deeply affected in $\Delta exp5$: short reads ~14 nt long are down-regulated in $\Delta exp5$, accounting for the disappearing of the 14 nt peak visible in Fig. 3.2.1A. **B)** sRNAs derived from transposable elements are strongly affected in $\Delta rbp35$; 19 nt and 20 nt long reads are up-regulated, while ~23 nt long sRNAs are down-regulated. **C)** Similarly to rRNA, very short sRNAs derived from protein-coding genes are down-regulated in $\Delta exp5$, while there is an increase in production of sRNAs of regular size. In $\Delta rbp35$ we observe a similar behaviour as in transposable elements. **D)** sRNAs derived from the MOLV1 virus are strongly affected in $\Delta rbp35$; 19 nt long reads are up-regulated, while ~22 nt long sRNAs are down-regulated.

Northern blots validate most significant RNA-seq results

To confirm previous results we compared the accumulation of a subset of selected elements by Northern blot analyses in the wild-type and the two mutants. Probes were designed to detect antisense sRNA from ITS2, sense sRNA from retro5 and MAGGY retrotransposons, full-length sense MOLV1 virus and sRNA sense from MOLV1 virus (Suppl. Fig. 3.2.S1C).

We observed earlier that sRNAs antisenses to ITS2 are sharply increased in $\Delta rbp35$, mostly in the form of 21 nt sequences. We designed a 21 nt probe to detect variations in $\Delta rbp35$ (Fig. 3.2.6A). No clear differences are found for sRNA of ~25 nt, which could also be justified by the duration of exposure time. However, there is accumulation of ~150 nt RNAs in $\Delta rbp35$, which roughly matches the ITS2 size (177 nt), suggesting that antisense transcription from the ITS2 is real and variation can be observed in $\Delta rbp35$.

Two interesting retrotransposons, retro5 and MAGGY, were also subjected to Northern blot analysis (Fig. 3.2.6B). sRNAs derived from these two elements are up-regulated in $\Delta exp5$. Exp5 is a karyopherin known to shuttle protein and tRNAs in *S. cerevisiae* [314], so we also analyzed nuclear and cytoplasmic fractions separately in order to verify if sRNA were subject to nuclear/cytoplasmic retention. 50 nt sRNAs derived from retro5 appear retained in the nucleus in $\Delta exp5$, while no clear difference could be seen in cytoplasmic fraction and in total RNA. Regarding MAGGY, only a minimal increase of ~30 nt sRNAs could be seen in both cellular fractions.

As discussed before, MOLV1 derived sRNAs represented approx 10% of our RNA sequencing. A probe was designed to verify the presence of the full length MOLV1 ssRNA (Fig. 3.2.6C). MOLV1 RNA is strongly reduced in $\Delta rpb35$, and its presence is restored in the complemented strain. Conversely, the $\Delta exp5$ mutant is not affected in MOLV1 transcript levels. Similar to the full-length transcript, ~25 nt sRNAs from MOLV1 are also depleted in $\Delta rpb35$, which correlates with the low levels of the full length MOLV1 virus found in this mutant (Fig. 3.2.6D).

A **B** **C** **D**

Figure 3.2.6 | RNA-seq results validation by Northern blotting. A) Amount of ribosomal ITS2 antisense sRNAs in Δ rbp35. No clear difference could be observed in the range 18 nt-25 nt, but there is a noticeable overaccumulation of 150 nt long sRNAs in Δ rbp35. **B)** Amount of retro5 and MAGGY retrotransposons sRNAs in Δ exp5. Considering that Exp5 is a karyopherin that could shuttle sRNAs from the nucleus to the cytoplasm, we also separately analyzed the nuclear and cytoplasmic fractions. An increase of ~50 nt sRNAs derived from retro5 was found in Δ exp5 nucleus compared to the wild-type (WT) strain. Similarly, nuclear sRNAs ~30-40 nt in size of MAGGY accumulated in Δ exp5. However, both retro5 and MAGGY sRNAs were detected at the same level in the cytoplasmic fractions. **C)** Amount of MOLV1 virus full length in Δ exp5 and Δ rbp35 mutants. *M. oryzae* strain 70-15 (lacking MOLV1 virus) is added for comparison. MOLV1 RNA levels are clearly diminished in Δ rbp35, and restored in the complemented strain. No clear difference is appreciable in the Δ exp5 mutant. **D)** Amount of MOLV1 virus sRNAs in Δ exp5 and Δ rbp35. In Δ rbp35 we can observe a clear depletion of 25 nt sRNA compared to WT, which correlates with reduced levels of MOLV1 full length RNA, while no differences are found in Δ exp5.

Discussion

In this study, we explored and analyzed the sRNA population of the filamentous fungus *M. oryzae* in the wild-type strain and two mutants, $\Delta exp5$ and $\Delta rbp35$. These mutants display defects in mycelium growth and plant infection. Exp5 is homologous to the human Exportin-5 and the yeast Msn5, which are karyopherins required for the export of proteins, pre-miRNAs and tRNA cargos. It is therefore reasonable to hypothesize its involvement in sRNA regulation in *M. oryzae*. Rbp35 is a component of the polyadenylation machinery. Though not directly related to sRNA processes, we have observed that this protein immunoprecipitates with an Ago protein (A. Sesma, personal communication), one of the principal components of the RNA silencing pathway in many organisms. Moreover, several sRNAs and TE are polyadenylated which could lead to a possible connection between Rbp35 and RNA silencing.

A preliminary overview of our RNA-seq output immediately suggested that RNA silencing is active and the majority of reads probably derive from siRNA. Read size distribution is not uniform (which could be a sign of degradation) and centered at 20 nt; sRNAs derived from the negative strand are common in TEs, mRNAs and MOLV1 virus, all of which are transcripts known to be potential targets for RNAi. In these same elements, the first nucleotide is usually a uracil, which is the preferred 5'-end substrate for Dicer and the most common first nucleotide in sRNA incorporated by Argonaute. Transposons account for only about 5% of the *M. oryzae* genome, depending on the strain, but almost half of total reads aligns to either TEs or the virus, meaning that significant RNAi-related activity is engaged in suppressing these invasive nucleic acids.

Interestingly, reads length distribution is not the same in the three strains' datasets. We discovered that up to 20% of total sequences exhibit a form of post-transcriptional modification, consisting of the addition of 3' adenines. Polyadenylation is a frequent post-transcriptional modification occurring on mRNAs (3' UTR cleavage and polyadenylation processes of which Rbp35 is part), and on rRNAs and snoRNAs (by the TRAMP complex) [114]. 3' adenylation in animal and plant miRNAs has been suggested as a possible modulator of target effectiveness [315]. This is, however, the first time that such pervasive sRNA modification was observed in fungi. The A-adding correlates with typical RNAi targets like TEs, mRNAs and the virus. A similar modification was observed before in *S. cerevisiae* and *N. crassa*, but only for ribosomal sRNAs and miRNA-like [256]. rRNA adenylation is thought to trigger exosome-mediated degradation as a surveillance mechanism to remove improperly processed rRNAs [301].

Deletion of Rbp35 and Exp5 affected the *M. oryzae* sRNA population in different ways. Every major source of sRNAs, such as tRNAs, rRNAs, TEs, mRNAs and MOLV1, is affected in the mutants. Overall production of sRNAs is not strongly altered in $\Delta rbp35$; however we observe the relative up-regulation of U-starting 19 nt and 20 nt sRNAs, mainly from repetitive elements,

suggesting that they are a product of RNAi. Northern blot analysis of full length MOLV1 transcripts indicate that virus accumulation is dramatically reduced in $\Delta rbp35$. This observation can be related with the role of Rbp35 protein in RNAi pathway regulation. All these facts seems to indicate that this protein is in fact a major player of RNA silencing regulation in the rice blast fungus, and its absence leads to a global increase of RNA silencing. However, lack of Rbp35 can also affect viral replication. Regarding the role of Exp5, the $\Delta exp5$ mutant does not appear to be directly implicated in siRNA regulation; there is a general over-expression of sRNAs derived from some TEs, resulting from the accumulation of these elements in the nucleus, at least for the retrotransposon retro5. Some tRNAs are also retained in the nucleus, confirming that Exp5 is in fact directly involved in the export of some non-coding RNAs. Taken all these data together, these results suggest that Rbp35 and Exp5 play different roles in the regulation of sRNA pathways in the rice blast fungus.

Supplemental Data

A

Sample	Total reads	adapter contamination, no 3' adapter found, low quality and artifact filtering	%	unique	%
Δexp5 _1	25580902	24087237	5.84%	1962459	8.15%
Δexp5 _2	39423030	37807985	4.10%	2625645	6.94%
Δexp5 _3	16606918	12971720	21.89%	1242286	9.58%
Δrbp35 _1	14107800	8351051	40.81%	835354	10.00%
Δrbp35 _2	25965551	23961924	7.72%	2346343	9.79%
Δrbp35 _3	32233131	26039242	19.22%	2314513	8.89%
WT_1	13826696	11280697	18.41%	1067465	9.46%
WT_2	21412622	19757242	7.73%	1559790	7.89%
WT_3	14168509	12950535	8.60%	1266876	9.78%

B

C

Target	Sequence	Orientation
MAGGY	GGACGTCGTCCCCGGTCCA	antisense
MOLV1	AGAGTTCACGCTCAAGCCCA	sense
retro5	ACGCCGCACCTAGGGTTTA	antisense
ITS2	CCCGGGGCCCCCAAGTTCA	antisense

Supplemental Figure 3.2.S1. A) sRNA-sequencing protocol. **B)** Functional group enrichment for the 1424 genes with an sRNA expression greater than 1 read per million. **C)** Targets, nucleotide sequences and orientation of the probes designed for the northern blot analysis.

Chapter 3.3

Molecular characterization of a novel ssRNA ourmia-like virus from the rice blast fungus *Magnaporthe oryzae*

Results published in Archives of Virology, March 2017, Volume 162, Issue 3, pp 891–895
Illana, A., Marconi, M., Rodríguez-Romero, J. et al. Arch Virol (2017) 162: 891. doi:10.1007/s00705-016-3144-9

Summary

In this chapter we characterize a novel positive and single stranded RNA (ssRNA) mycovirus isolated from the rice field isolate of *M. oryzae* Guy11. The ssRNA is of 2,364 nucleotides in length and contains a single open reading frame (ORF) encoding an RNA-dependent RNA polymerase (RdRp) closely related to ourmiaviruses (plant viruses) and ourmia-like mycoviruses. Accordingly, we name this virus *M. oryzae* ourmia-like virus 1 (MOLV1). Although phylogenetic analysis suggests that MOLV1 is closely related to ourmia and ourmia-like viruses, it has some features never reported before within the *Ourmiavirus* genus. 3' RLM-RACE (RNA ligase-mediated rapid amplification of cDNA ends) and extension poly(A) tests (ePAT) suggest that the MOLV1 genome contains a poly(A) tail whereas the three cytosine and the three guanine residues present in 5' and 3' untranslated regions (UTRs) of ourmia viruses are not observed in the MOLV1 sequence. The discovery of this novel viral genome supports the hypothesis that plant pathogenic fungi may have acquired this type of viruses from their host plants.

Background

Mycoviruses are widespread in all taxonomic groups of fungi. Different outcomes result from the interaction between mycoviruses and their respective fungal hosts. Usually, a mycovirus infection remains latent and shows no symptoms in the fungal host [316]. However, some mycoviruses are able to cause severe lesions and impair the vegetative growth of their fungal host, causing economic losses, as in the case of the cultivated mushroom *Agaricus bisporus* [316,317]. The presence of a mycovirus can also be beneficial for all the partners in the tripartite interaction between virus, fungus and plant host. This is the case described for a tropical panic grass and its fungal endophyte when infected by a virus, a process which increases the grasses survival under stress temperature conditions [318]. The mycovirus can additionally cause fungal hypovirulence in the host plant making it a good option for biological control, as is the case for the well-known *Criphonectria hypovirus 1* (CHV1) and chestnut blight control in Europe [319,320].

Mycoviruses have a diverse genome organization. Most viruses have double stranded RNA (dsRNA) or linear positive (+) ssRNA genomes [317]. Unclassified linear negative ssRNAs genomes and circular ssDNA viruses have also been isolated [317,320–322]. Linear (+) ssRNA mycoviruses are classified into 5 families: *Alphaflexiviridae*, *Gammapflexiviridae*, *Hypoviridae*, *Narnaviridae* and *Barnaviridae* [317,323]. The *Narnaviridae* family includes the *Mitovirus* and *Narnavirus* genera.

The (+) ssRNA genomes from viruses classified in the *Narnaviridae* family are the simplest mycoviruses. Members of this family have a 2-3 Kb genome size, which usually encodes an RNA-dependent RNA polymerase (RdRp) needed for their own replication. Narnaviruses lack the coat protein (CP), that forms the capsid structure, as well as the movement protein (MP) [324]. Phylogenetically, the narnaviruses' closest relatives are the plant viruses classified in the *Ourmiavirus* genus [324,325]. Plant ourmiaviruses contain a tripartite ssRNA genome that encodes three proteins, the RdRp, the CP and the MP [325]. Phylogenetic analyses support the idea that the *Ourmiavirus* genus represents a link between mycoviruses and plant viruses [324,326].

The ascomycetous fungus *M. oryzae* causes the damaging rice blast disease and results in approximately a 30% yield loss in rice fields every year [327]. *M. oryzae* was the first plant pathogenic fungus where the presence of a polyhedral virus was reported [328,329]. Several viruses from different families have been found in this fungal species. *M. oryzae* virus 1, 2 and 3 (MoV1, MoV2 and MoV3) are dsRNA viruses belonging to Totiviridae family [25,27,330]. *M. oryzae* chrysovirus 1A and B (MoCV1-A and MoCV1-B) are dsRNA viruses that belong to the *Chrysoviridae* family. They can impair vegetative and invasive growth during *M. oryzae* host colonisation [331,332]. Recently, a (+) ssRNA virus has been described for the first time in this

fungal species [333], *M. oryzae* virus A (MoVA). The genome of MoVA is 3,246 nt in length and contains two in-frame ORFs, one of which encodes a protein with similarity to an RdRp observed in plant viruses of the Tombusviridae family.

In this work we describe a second and novel type of ssRNA virus found in the *M. oryzae* rice isolate Guy11, *M. oryzae* ourmia-like virus 1 (MOLV1), a polyadenylated (+) ssRNA virus phylogenetically related to plant viruses from the Ourmiavirus genus.

Results

Discovery of a new mycovirus in *M. oryzae*

During the analysis of sequences derived from small RNA (sRNA) libraries isolated from the rice field isolate Guy11 [181], we detected sequences that were not present in the published *M. oryzae* genome of the laboratory strain 70-15 [334], a near-isogenic strain of Guy11. We additionally confirmed this by downloading EST sequences from *M. oryzae* Guy11 and aligning them to the reference genome. EST sequences that failed to align were collected and assembled using the Cap3 DNA sequence assembly program [335]. Using these unknown ESTs, we predicted putative ORFs using the getorf tool from the EMBOSS suite [336], which were aligned against the NCBI non-redundant sequence collection [337] by blastp [202]. This allowed us to identify transcripts showing similarities with ourmia-like mycoviruses, and plant viruses of the *Ourmiavirus* genus. Consequently, we named this viral genome *M. oryzae* ourmialike virus 1 (MOLV1).

The presence of MOLV1 RNA virus in *M. oryzae* Guy11 strain was confirmed by Northern blotting using total RNA extracts (Fig. 3.3.1A). Almost all of the retrieved MOLV1 sequence was used as a DNA probe, and actin DNA was used as the control. A band of approximately 2.3 Kb in the *M. oryzae* Guy11 strain was observed, in addition to other smaller bands. These RNA molecules were not detected in the *M. oryzae* laboratory strain 70-15. This is consistent with the absence of these viral sequences from EST collections other than those of the fungal isolate Guy11.

3' RLM-RACE and ePAT experiments suggest that MOLV1 is a polyadenylated (+) ssRNA virus

To identify 5' and 3' untranslated regions (UTRs) of MOLV1 we carried out an RNA ligase-mediated rapid amplification of cDNA ends (RLM-RACE) (Fig. 3.3.1B). The 3' terminus is usually polyadenylated before conducting a 3' RLM-RACE in non-polyadenylated viruses. In our case we obtained a clear band without adding any adenosine to the RNA sample. We purified and sequenced the largest bands obtained from both 5' and 3' RLM-RACE experiments (asterisks, Fig. 3.3.1B). The 5' and 3'UTRs of MOLV1 are 116 nt and 430 nt in length, respectively. To confirm the presence of a poly(A) tail in the MOLV1 genome an extension poly(A) test (ePAT) experiment was carried out (Fig. 3.3.1C). This method allows measurement of the poly(A) tail length of polyadenylated RNA molecules [338]. The difference in band sizes between MOLV1 and the negative control TVN-PAT indicated that the poly(A) tail of MOLV1 is between 25-30 nt in length.

These experiments suggested that the MOLV1 genome is a positive and polyadenylated ssRNA molecule of 2,364 nt, excluding the poly(A) tail. We also performed the ePAT experiment to determine if the BOLV genome contained a poly(A) tail and found that it did not. With the exception of *S. sclerotiorum* mitovirus 1/KL-1 (SsMV-1/KL1) and 2a (CcMV-2a), no other members of the family *Narnaviridae* are polyadenylated at their 3'-end [324,339]. To our knowledge, this is the first described ourmia-like mycovirus that has a poly(A) tail. However, these results should be treated cautiously since it is possible to find amplification products with oligodT primers in non-polyadenylated RNA viruses due to the presence of small fractions of gRNA with poly(A) tails [325,340].

Figure 3.3.1 | Characterisation of a novel RNA virus in the rice blast fungus. A) Northern blots confirming the presence of MOLV1 in *M. oryzae* strain Guy11, and its absence from genome reference strain 70-15. **B)** 5' and 3' RLM-RACE reactions to identify 5' and 3' UTR regions of MOLV1. Bands selected for cloning and further sequencing are indicated with asterisks. **C)** ePAT experiment using total RNA from *M. oryzae* Guy11, indicating that MOLV1 poly(A) tail is between 25-30 nt in length. **D)** Schematic representation of MOLV1 RNA genome showing the ORF, UTRs and location of primers used for ePAT and Northern experiments

The MOLV1 genome has a unique open reading frame (ORF) from position 117 to 1931 (Fig. 3.3.1D). The translation of this 818 nt ORF generates a polypeptide of 605 aa with a theoretical molecular weight of 67.12 kDa. This protein has higher identity with the RdRp of *Rhizoctonia solani* ourmia-like virus 1 (RsOLV1; 40%), although it is also similar to other ourmia-like mycoviruses and plant ourmiaviruses (Fig. 3.3.2A). The MOLV1 viral protein contains the

conserved domains of the RdRp present in (+) ssRNA viruses [341], including the highly conserved core domain GDD in motif VI (Fig. 3.3.2A). These findings suggest that the ORF of MOLV1 encodes a putative RdRp. The alignment of MOLV1 with the BOLV sequence showed no conservation of the three cytosine residues at the 5'UTR and the three guanine residues at the 3'UTR. The predicted secondary structure indicated the presence of a stable terminal stem-loop structure only for the first 40 nt of MOLV1 (DG value of -10.00 kcal/mol) (Fig. 3.3.2C). The last 34 nt of the 3'UTR could be folded into a less stable stem-loop structure with DG value of -6.30 kcal/mol. These secondary structures at both UTRs could play a role in the replication of MOLV1 by providing a recognition site for RdRp and/or protecting the genome from degradation as has been proposed for BOLV and mitoviruses [324,339,342].

MOLV1 is closely related to plant ourmiaviruses and ourmia-like mycoviruses

Full-length amino acid sequences of published RdRps (29 proteins) were selected to construct a phylogenetic tree to infer the evolutionary history of MOLV1 (Fig. 3.3.2B). The RdRp of the tobamovirus tobacco mosaic virus (TMV) was used as outgroup in the tree. The maximum likelihood (ML) phylogenetic tree was inferred from the amino acid sequences of viral RdRps after removing positions with less than 50% site coverage. The phylogenetic analyses clearly showed two clades. The first clade grouped all mitoviruses. The second clade grouped the two *Saccharomyces* narnaviruses and *Phytophthora infestans* RNA virus (82% bootstrap support), and all the ourmia-like mycoviruses and the plant ourmiaviruses (97% bootstrap support). This ML analysis indicates that fungal ourmia-like viruses are more closely related to plant viruses of the *Ourmiavirus* genus than mycoviruses of the *Narnaviridae* family. As previously shown, this phylogenetic tree places ourmia-like viruses as more closely related to members of the genus Narnavirus than to members of the genus Mitovirus [325,326,343].

Figure 3.3.2 | Sequence analysis of MOLV1. A) Alignment showing RdRp conserved motifs I to VIII of MOLV1 (LT593139), RsOLV1 (ALD89131.1), BOLV (LN827955), SsOLV2 (ALD89139.1), RsOLV2 (ALD89132.1), SsOLV1 (ALD89138.1), SaOurV2 (KT598247.1) and SaOurV1 (KT598235.1). **B)** Phylogenetic tree obtained by the Maximum Likelihood method with the RdRp sequences of 25 mycoviruses, including MOLV1, and 3 plant ourmiaviruses. The RdRp sequence of TMV (ABN79257.1) was included as an outgroup. The gray box shows ourmia-like fungal viruses. All bootstrap values (%) are represented at each node of the tree (test with 1000 replicates). Branch lengths are proportional to the number of amino acid substitutions and are measured by the scale bar. **C)** Predicted secondary structure of the 5' (left) and 3' (right) UTRs of MOLV1 with their corresponding ΔG in kcal/mol.

Discussion

Our results suggest that MOLV1 is a novel (+) ssRNA virus isolated from the rice blast fungus closely related to ourmia-like mycoviruses and plant ourmia viruses that exhibits new features such as a poly(A) tail and showed no conservation of the three cytosine residues at the 5'UTR and the three guanine residues at the 3'UTR.

Different theories have been suggested for mycovirus evolution [344]. The plant ourmiavirus genomes have three RNA molecules which encode three proteins, the RdRp, the CP and the MP [325]. The identification of ourmia-like mycoviruses such as MOLV1, which contain a single ORF and are closely related to plant ourmia viruses suggest that they might have been originated by gene loss events during their adaptation to the fungal host as has been suggested for other ourmia-like mycoviruses [322,326,345]. However, we cannot discard the possibility of an ourmia-like mycovirus being the ancestor of plant ourmiaviruses as it has been proposed [325].

Chapter 3.4

Large scale comparative analyses of RNA-related machineries in the fungal kingdom

Summary

Although RNA-related protein machineries are well studied in model organisms such as budding yeast or human, these protein complexes have not been extensively analysed in other fungi. The recently sequenced genomes of distant fungal species provide an excellent platform for a broad comparative analysis of RNA-related proteins across the fungal kingdom. Here, we identified using a computational pipeline proteins involved in RNA metabolism in the most representative fungal species of diverging fungal phyla. We closely assessed 731 proteins related with nine RNA-associated protein machineries. Our analysis shows a general conservation of core proteins involved in splicing and translation across the fungal lineage, while protein related with mRNA export and RNA silencing diverged more along evolution. We also observed that core proteins of fundamental importance in model organisms can be disposed in some branches of the fungal tree. In addition, we identified conserved nucleotide elements in fungal mRNA from public available transcriptomic data. The identification of RNA-binding proteins in these fungal species can help discern the role played by these proteins in their corresponding biological context.

Background

Fungi are highly diverse eukaryotic organisms of great relevance for industry, medicine and agriculture and are taxonomically grouped as an independent eukaryotic kingdom. They are present in all terrestrial ecosystems and have common and distinctive features from plants and animals [346,347]. The number of fungal species on Earth is estimated to be more than one million, although recent data from environmental sequencing research estimate as many as five million species [348]. The relatively small size of their genomes and the wide range of lifestyles they undertake - from free-living and saprophytic to symbiotic and parasitic - have placed them in the centre of many studies to understand the molecular basis of their nutrient acquisition and adaptation to the environment [349]. Moreover, filamentous fungi represent powerful tools for understanding fundamental biological processes, since they are highly tractable organisms for experimental research and can be used to study important features of complex eukaryotes such as cellular differentiation, pathogenesis, DNA methylation, gene silencing, chromatin remodelling and programmed cell death. An increasing number of fungal genomes have already been sequenced, and initiatives exist to promote large-scale genome sequencing projects [350]. This has already enabled precise comparative genomics within this kingdom [13,351,352].

Cell survival depends on reliable mechanisms that both positively and negatively regulate gene expression to provide rapid and flexible responses to environmental signals, stress conditions and/or developmental stages. Synthesis and maturation of eukaryotic messenger RNA (mRNA) involves a series of coordinated steps including transcription, splicing and 3'-end processing [353,354]. mRNA export, localisation, translation and degradation guarantee the expression of a gene at the right time and place [47,89,106]. Typically, pre-mRNA synthesis is carried out by RNA polymerase II (Pol II) for protein-coding genes, whereas RNA polymerases I (Pol I) and RNA polymerase III (Pol III) transcribe non protein-coding genes like tRNAs, rRNAs, snRNAs and snoRNAs. The three enzymes contain a conserved core region and associate with specific transcription initiation factors [42]. During transcription, several co-transcriptionally regulated events take place. Pre-mRNA is modified at the 5'-end by adding a 7-methylguanosine molecule at the first nucleotide of the pre-mRNA to stabilise it (*mRNA capping*) [51], and the spliceosome removes the introns by splicing [52]. In addition, the 3' untranslated region (3' UTR) is cleaved in the nucleus and a polyadenosine tail is added in a two-step process called *polyadenylation* [68]. Once a mature mRNA is formed, it is exported from the nucleus to the cytoplasm where translation occurs. The mRNA export receptor Mex67-Mtr2 complex recruits the mRNAs, and contacts nucleoporins to mediate their transit through the nuclear pore complex (NPC) [47]. mRNA *export* is also tightly coupled with transcription and pre-mRNA maturation [355], and the entire process is overseen by the nuclear and cytoplasmic *surveillance* machinery to prevent translation of aberrant transcripts [47,107]. mRNA localization and asymmetric

distribution is achieved by active transport of motor proteins, an important mechanism that contributes to cell polarity and cellular asymmetry [89]. The synthesis of proteins from mRNAs is carried by the translational machinery in which the ribosome, the largest non-membrane bound organelle of the cell, is the central component [90]. Scanning of the ribosome through the 5' UTR of the mRNA regulates translation initiation, a highly regulated process [356], although protein synthesis control is also achieved during translation elongation and termination [357–359]. In addition, non-coding RNAs (ncRNAs) like transfer RNAs (tRNA) and ribosomal RNAs (rRNA) play an essential role during translation [90,99]. Aminoacyl-tRNAs are loaded into the ribosomes when matching the corresponding mRNA codon with the amino acid they code for during protein synthesis, whereas rRNAs are structural components of the ribosome. These ncRNAs experience several *post-transcriptional modifications* to reach their mature form [360,361]. All these steps that take place during the RNA life cycle have a strong influence on the final output of gene expression. Another way of controlling gene expression is by RNA *silencing*, which refers not only to a highly conserved biological response to parasitic nucleic acids (transposons and viruses) but also to posttranscriptional regulation of endogenous genes and heterochromatin formation in eukaryotes [120,121].

Few studies have been carried out regarding evolutionary aspects of fungal RNA-related machineries. Previous work has shown that the RNA-binding protein Rbp35 (MGG_02741) of the rice blast fungus *M. oryzae* has no clear orthologue in yeasts or metazoan cells [362]. *In vivo* immunoprecipitation assays revealed that Rbp35 interacts with CFI_m25 (MGG_01676), a conserved protein involved in poly(A) site recognition and cleavage of pre-mRNAs. Therefore, Rbp35 represents the functional ortholog of metazoan CFI_m68 in filamentous fungi, a subunit of the Cleavage Factor I complex. The altered protein domain conservation of Rbp35 with respect to its human counterpart, and the absence of a clear Rbp35 orthologue in yeast prompted our investigation towards understanding the protein composition of eukaryotic mRNA machineries within the different phyla of the fungal kingdom. Here, we analysed the degree of conservation of proteins involved in the most important RNA processes affecting its life cycle taking the advantage of the increased number of sequenced fungal genomes. We include in the study a broad range of species with a wide variety of ecological niches and lifestyles with the aim of identifying differences that can shed light on the evolution of fungal RNA metabolism. Deciphering these processes in fungi will have an important impact on health, environment and economy.

Results

mRNA export- and RNA silencing-related proteins diverged more rapidly in the fungal kingdom, while mRNA translation- and splicing-related proteins remained more conserved

Two considerations were applied to the selection of fungal species used in this study (Suppl. Table 3.4.S1): broad coverage of the fungal kingdom, and lifestyle diversity. A phylogenetic reconstruction with the 49 selected fungal species was executed using seven conserved proteins (Fig. 3.4.1A). Model plant *Arabidopsis thaliana*, *Homo sapiens*, and the oomycete *Phytophthora infestans* were included as outgroups. The 49 fungal species were organized in the following taxonomic clades: Microsporidia (four species: *Anncaliia algerae*, *Nematocida parisii*, *Encephalitozoon cuniculi* and *Vavraia culicis*), Chytridiomycota (three species: *Batrachochytrium dendrobatidis*, *Rhizoclostridium globosum* and *Gonapodia prolifera*), Kickxellomycotina (three species: *Coemansia reversa*, *Martensiomycetes pterosporus* and *Linderina pennispora*), Entomophthoromycotina (three species: *Basidiobolus meristosporus*, *Conidiobolus thomboides* and *Conidiobolus coronatus*), Glomeromycotina (one species: *Rhizophagus irregularis*), Mucoromycotina (three species: *Phycomyces blakesleeana*, *Mucor circinelloides* and *Rhizopus oryzae*), Basidiomycota (eight species: *Puccinia graminis*, *Microbotryum violaceum*, *Ustilago maydis*, *Cryptococcus neoformans*, *Gloeophyllum trabeum*, *Pleurotus ostreatus*, *Coprinopsis cinerea* and *Laccaria bicolor*), Taphrinomycotina (*Schizosaccharomyces pombe*, *Taphrina deformans* and *Saitoella complicata*), Pezizomycotina (sixteen species: *Tuber melanosporum*, *Aspergillus nidulans*, *Histoplasma capsulatum*, *Coccidioides posadasii*, *Verticillium dahliae*, *Fusarium oxysporum*, *Trichoderma reesei*, *Metarhizium acridum*, *Neurospora crassa*, *Podospora anserina*, *Grosmannia clavigera*, *Magnaporthe oryzae*, *Botrytis cinerea*, *Blumeria graminis*, *Leptosphaeria maculans* and *Zymoseptoria tritici*), and Saccharomycotina (five species: *Yarrowia lipolytica*, *Kluyveromyces lactis*, *Saccharomyces cerevisiae*, *Candida albicans* and *Candida tenuis*). We used the resulting phylogeny as the scaffold for many of the figures in this manuscript to facilitate interpretation of the results.

Figure 3.4.1 | Global overview of RNA-associated protein machineries conservation. A) The Phylogenetic tree of the 52 selected species, reconstructed from the multiple alignment of the orthologues of seven proteins: actin (Act1), pyruvate kinase (Cdc19), diphthine synthase (Dph5), glutamine tRNA synthetase (Gln4), myosin (Myo4), second-largest subunit of RNA Pol III (Ret1) and subunit of heterotrimeric Replication Protein A (Rfa1). **B)** Heatmap showing the degree of orthologs preservation along the selected 52 species, for each of the nine considered RNA-associated protein machineries. The color code represents the percentage of orthologs identified for each species, based on the proteins listed in Annex Table 1. Horizontal whitespaces break the heatmap into the same phylogenetic groups as in (A). **C)** Boxplots of average evolutionary distances among identified orthologs inside each RNA-associated process. Proteins composing the mRNA translation machinery are highly conserved among the fungal kingdom, while mRNA export and mRNA silencing are more divergent. The background control box has been build using the whole proteomes. Microsporidia clade was a highly distant outlier in every boxplot and was removed from this analysis.

Taphrinomycotina, Pezizomycotina and Saccharomycotina are part of the Ascomycota division, which together with Basidiomycota, the Ascomycota sister phylum, form the Dikarya subkingdom. These two phyla are the largest of terrestrial fungi [363]. The Ascomycota includes fungal species ranging from yeasts like *Saccharomyces cerevisiae* to molds including *Penicillium* and *Aspergillus*. Glomeromycotina are fungi that form arbuscular mycorrhizas with the roots of land plants, and are therefore of high ecological and agronomical importance [364]. Mucoromycotina are fast-growing colonizers of carbon-rich soils, saprobes and ectomycorrhizal [364], and they include the model organism *M. circinelloides*. The Kickxellomycotina fungi are united by the formation of disciform septal pores containing lenticular plugs, and encompasses a broad spectrum of life-styles, including the hindgut of arthropods, saprobes and mycoparasites [365]. Entomophthoromycotina (the “destroyer of insects”), are generally arthropods parasites or colonizers of decaying matter [366]. The Chytridiomycota clade was added since it includes early diverging fungal lineages, generally saprobic although some are important amphibian pathogens [367]. Microsporidia are eukaryotic parasites of animals with an obligate intracellular lifestyle, with extremely reduced genomes [368].

We selected proteins and complexes involved in nine RNA-associated processes, i.e. transcription, splicing, 5'- and 3'-end processing, export and localization, degradation and decay, cytoplasmic mRNA surveillance, RNA silencing, tRNA maturation, and mRNA translation. The initial analysis considered all the proteins within each of the RNA-related protein machineries as individual units, and examined their overall homology among fungal species, organized taxonomically. The list of components for each of these protein machineries was retrieved from the available literature, primarily from model organisms (Annex Table 1), and orthologs were identified in each of the remaining species (see Materials and Methods). In total, we worked with 731 proteins belonging to the 52 eukaryotic species included in this study.

Different percentages of orthology among the nine RNA-associated protein machineries were detected for each species (Fig. 3.4.1B), ranging from full preservation (every ortholog identified) to complete absence (no ortholog identified). In this chapter, we will use the term “preservation” to indicate the presence or absence of a protein ortholog, and “conservation” to denote the degree of evolutionary divergence between orthologs. The best-preserved proteins were the ones related with translation and tRNA maturation, with the majority of components having an ortholog in the majority of the species. Proteins involved in RNA silencing were poorly preserved among clades, being lost in *T. deformans* (Taphrinomycotina), *U. maydis* (Basidiomycota), *M. pterosporus* (Kickxellomycotina), *E. cuniculi* and *N. parisii* (Microsporidia). Our results also confirmed the loss of all the main RNA silencing components in Saccharomycotina, *U. maydis* and *T. deformans*. Microsporidia typically displayed the lowest number of orthologs, lacking almost the full protein set for mRNA splicing, export and localization. Globally, different degrees of preservation were observed among fungal species,

even within the same clade. For example, *R. globosum* (Chytridiomycota) and *T. deformans* (Taphrinomycotina) were below the average in terms of orthologs found, while *B. meristosporus* (Entomophthoromycotina), *R. oryzae* (Mucoromycotina) and several Pezizomycotina fungi like *N. crassa* or *M. oryzae* generally possessed orthologs for most components of the RNA-related processes examined. Surprisingly, species belonging to the same genus like *C. albicans*/*C. tenuis* and *C. coronatus*/*C. thomboides*, did not necessarily exhibit similar patterns of preservation, reflection of their different life-styles.

We also wanted to know if proteins participating in a given RNA-associated process have experienced comparatively higher or lower levels of modification/conservation with respect to the others. The corresponding phylogenetic tree was reconstructed for every orthologs group following the same procedure used during the orthologs prediction phase. The average evolutionary distance was calculated (in terms of number of amino acid changes per site) for each selected protein from every other protein inside the same orthologs tree. Considering that some fungal clades contain more species than others, we calculated a clade-wise mean of the distances and represented the average evolutionary distances for all the components of each protein machinery (Fig. 3.4.1C). We removed the Microsporidia clade from this analysis since it was an outlier in every box plot. Proteins within the mRNA translation machinery were highly conserved within the fungal kingdom, with an average evolutionary distance below one (less than one amino acid change per site). Interestingly, proteins involved in mRNA splicing were notably more conserved than proteins belonging to functionally coupled activities such as RNA transcription and pre-mRNA 5'/3' processing. The single top outlier of the mRNA splicing and mRNA 5'/3' processing box plots was Saccharomycotina, indicating that these machineries diverged significantly in this fungal subphylum. Proteins involved in RNA export/localization and RNA silencing were less well conserved a proof that they underwent more extensive adaptation and/or diversification.

Figure 3.4.2 | RNA-related proteins orthologs preservation. (A, B, C, D, E, F, G, H, I) Heatmaps showing the degree of orthologs preservation along the selected 52 species, for each of the nine considered RNA-related processes, classified by relevant complexes. Color code represents the percentage of orthologs identified for each species based on proteins listed in Annex Table 1. White indicates almost complete absence of orthologs (<25%) while navy blue indicates almost complete preservation of orthologs (>75%).

Table 3.4.1 | Summary of most divergent RNA-associated protein orthologs in fungal clades.

		Chy	Kic	Ent	Glo	Muc	Bas	Tap	Pez	Sac
RNA Transcription	Mediator	MED1	X	X	X	X	X	O	X	X
		MED2	X	X	X	X	X	X	X	
		MED4	X	X	X	X	X	X	X	
		MED8	X	X	X	X	X	X	X	
		MED9	X	X	X	X	X	X	X	X
		MED11	X	X	X	X	X	X	X	X
		MED15	X	X	X	X	X	X	X	X
		MED18	X	X	X	X	X	O	X	
		MED16	X	X	X	X	X	X	X	
		MED23	X	X	X	X	X	X	X	X
		MED24	X	X	X	X	X	X	X	X
		MED25	X	X	X	X	X	X	X	X
		MED26	X	X	X	X	X	X	X	X
		MED27	X	X	X	X	X	X	X	X
		MED28	X	X	X	X	X	X	X	X
		MED30	X	X	X	X	X	X	X	X
		CSE2	X	X	X	X	X	X	X	X
		NUT1	X	X	X	X	X	X	X	X
		PGD1	X	X	X	X	X	X	X	X
		ROX3	X	X	X	X	X	X	X	X
		SRB2	X	X	X	X	X	X	X	X
		SRB6	X	X	X	X	X	X	X	X
		RPA14	X	X	X	X	X	X	X	X
		RPA34	X	X	X	X	X	X	X	X
		RPB9	X	X	X	X	X	X	X	X
		NNF2	X	X	X	X	X	X	X	X
		RPC17	X	X	X	X	X	X	X	X
		RPC25	X	X	X	X	X	X	X	X
		RPC31	X	X	X	X	X	X	X	X
		RRN6	X	X	X	X	X	X	X	X
		RRN11	X	X	X	X	X	X	X	X
		TAF3	X	X	X	X	X	X	X	X
		TAF4	X	X	X	X	X	X	X	X
		TAF14	X	X	X	X	X	X	X	X
		TFB6	X	X	X	X	X	X	X	X
		TFC7	X	X	X	X	X	X	X	X
mRNA Splicing	TFIIH	LSM	X	X	X	X	X	X	X	X
		PRP42	X	X	X	X	X	X	X	X
		PRP39	X	X	X	X	X	X	X	X
	U4/U6	PRP2	X	X	X	X	X	X	X	X
		SPP381	X	X	X	X	X	X	X	X
		PRP38	X	X	X	X	X	X	X	X
	NTR	NTR2	X	X	X	X	X	X	X	X
		SNT309	X	X	X	X	X	X	X	X
		ZMAT5	X	X	X	X	X	X	X	X
	Minor spliceosome	SNRNP48	X	X	X	X	X	X	X	X
		PDCD7	X	X	X	X	X	X	X	X
		SRSF14/5/6/9	X	X	X	X	X	X	X	X
	SR proteins	SRSF3/7	X	X	X	X	X	X	X	X
		SRSF11	X	X	X	X	X	X	X	X
			X	X	X	X	X	X	X	X

		Chy	Kic	Ent	Glo	Muc	Bas	Tap	Pez	Sac
mRNA 5'/3' processing	CFIm complex	SYC1	X	X	X	X	X	X	X	O
		REF2	X	X	X	X	X	X	X	O
mRNA Export and Localization	mRNA shuttling	CFIm25	X	X	X	X	X	X	X	O
		RBP35	X	X	X	X	X	X	X	O
	TREX	MTR2	X	X	X	X	X	X	X	X
		THOC1	X	X	X	X	X	X	X	O
		THOC5	X	X	X	X	X	X	X	O
		THOC6	X	X	X	X	X	X	X	X
		THOC7	X	X	X	X	X	X	X	X
		THP2	X	X	X	X	X	X	X	X
		MET1	X	X	X	X	X	X	X	X
		HPR1	X	X	X	X	X	X	X	O
	TREX-2	SUS1	X	X	X	X	X	X	X	O
		SHE2	X	X	X	X	X	X	X	O
	Localization	SHE3	X	X	X	X	X	X	X	X
		LOC1	X	X	X	X	X	X	X	X
		RRM4	O	X	X	X	X	X	X	X
	NPC transmembrane ring	POM121	X	X	X	X	X	X	X	X
		pom34	X	X	X	X	X	X	X	X
mRNA degradation	Exosome	POM34	O	X	X	X	X	X	X	X
		NUP1	X	O	X	X	X	X	X	X
		NUP60	X	X	X	X	X	X	X	X
	NPC inner ring	NUP53	X	O	X	X	X	X	X	X
		NUP37	O	X	X	X	X	X	X	O
	NPC outer ring	NUP85	O	O	X	X	X	X	X	X
		NUP120	O	O	X	X	X	X	X	X
	NNS complex	AAAS	O	X	X	X	X	X	X	X
		cid1	O	X	X	X	X	X	X	X
		CSL4	X	X	X	X	X	X	X	X
	Ccr4Not complex	MPP6	X	X	X	X	X	X	X	X
		RRP43	X	X	X	X	X	X	X	X
	SK1 complex	NAB3	X	O	X	X	X	X	X	X
		NRD1	O	X	X	X	X	X	X	X
	Cytoplasmic 5' degradation	CAF130	X	X	X	X	X	X	X	X
		SKI7	X	X	X	X	X	X	X	O
mRNA Cytoplasmic Surveillance	RQC	EDC1	X	X	X	X	X	X	X	O
		EDC2	X	X	X	X	X	X	X	O
		EDC3	X	X	X	X	X	X	X	O
	NMD	LSM12	X	X	X	X	X	X	X	X
		PBP4	X	X	X	X	X	X	X	X
		TAE2	X	X	X	X	X	X	X	X
	NMD	EB51	X	X	X	X	X	X	X	X
		NMD4	X	X	X	X	X	X	X	X
		SMG5	X	X	X	X	X	X	X	X
	NMD	SMG6	X	X	X	X	X	X	X	X
		SMG7	X	X	X	X	X	X	X	X
		SMG8	O	X	O	X	X	X	X	X
	NMD	SMG9	O	X	O	X	X	X	X	X
			X	X	X	X	X	X	X	X
			X	X	X	X	X	X	X	X

		Chy	Kic	Ent	Glo	Muc	Bas	Tap	Pez	Sac
mRNA Silencing	Argonautes	qde-2	X	X	X	X	X	X	X	O
		sms-2	X	X	X	X	X	X	X	O
		AGO3	X	X	X	X	X	X	X	O
	Dicers	dcl-1	X	X	X	X	X	X	X	X
		dcl-2	X	X	X	X	X	X	X	X
		sad-1	O	X	X	X	X	X	X	X
	Rdps	qde-1	O	X	X	X	X	X	X	X
		rrp3	X	X	X	X	X	X	X	X
		aes1	X	X	X	X	X	X	X	X
	Chromatin Silencing	arb1	X	X	X	X	X	X	X	X
		arb2	X	X	X	X	X	X	X	X
		chp1	X	X	X	X	X	X	X	X
	RPAs and Helicases	dsh1	X	X	X	X	X	X	X	X
		eri1	X	X	X	X	X	X	X	X
		ers1	X	X	X	X	X	X	X	X
tRNA Maturation	5' Trimming	snd1	X	X	X	X	X	X	X	X
		stc1	X	X	X	X	X	X	X	X
		tas3	X	X	X	X	X	X	X	X
	3' Trimming	hrr1	O	X	O	X	O	O	X	X
		rpa3	O	X	O	X	O	O	X	X
		qip	O	X	O	X	O	O	X	X
	Import/Export	sad2	X	X	X	X	X	X	X	X
		sad4	X	X	X	X	X	X	X	X
		sad5	X	X	X	X	X	X	X	O
	Modification U	POP6	X	X	X	X	X	X	X	O
		POP7	X	X	X	X	X	X	X	O
		POP8	X	X	X	X	X	X	X	X
	Modification G	RPP25	X	X	X	X	X	X	X	X
		rpp40	O	X	X	X	X	X	X	O
		RNH70	X	X	X	X	X	X	X	O
	Modification A	RNY1	O	X	X	X	X	X	X	O
		MSN5	X	X	X	X	X	X	X	X
		ELP6	X	X	X	X	X	X	X	X
		PUS6	X	X	X	X	X	X	X	X
		RIB2	X	X	X	X	X	X	X	X
		SIT4	X	X	X	X	X	X	X	X
		TYW1	X	X	X	X	X	X	X	X
		TYW3	X	X	X	X	X	X	X	X
		TRM12	X	X	X	X	X	X	X	X
		TRM3	X	X	X	X	X	X	X	X
		TAD3	X	X	X	X	X	X	X	X
		TRM13	X	X	X	X	X	X	X	X

		Chy	Kic	Ent	Glo	Muc	Bas	Tap	Pez	Sac
mRNA translation	Elongation factors	EFB1	X	X	X	X	X	X	X	X
		YEF3	X	X	X	X	X	X	X	X
		int6	X	X	X	X	X	X	X	X
	Initiation factors	moe1	X	X	X	X	X	X	X	O
		mti3	O	X	X	X	X	X	X	O
		TIF5	X	X	X	X	X	X	X	X
	Mitochondrial large subunit	TMA64	X	X	X	X	X	X	X	X
		IMG2	X	X	X	X	X	X	X	X
		MRPL23	X	X	X	X	X	X	X	X
	Mitochondrial small subunit	MRPL25	X	X	X	X	X	X	X	X
		mrpl40	X	X	X	X	X	X	X	X
		mrpl44	X	X	X	X	X	X	X	X
	Mitochondrial small subunit	MRPL6	X	X	X	X	X	X	X	X
		MRP21	X	X	X	X	X	X	X	X
		mrp49	X	X	X	X	X	X	X	X
	Mitochondrial small subunit	MRP4	X	X	X	X	X	X	X	X
		MRPS5	O	X	X	X	X	X	X	X
		RSM23	O	X	X	X	X	X	X	X
		YMR31	O	X	X	X	X	X	X	X

Seed protein color coding:

Selected proteins display the most heterogeneous degree of preservation among fungal clades, e.g. proteins with orthologs completely missing in a fungal subphylum. Proteins are indicated by the relative gene locus identifier, according to the standard nomenclature for that species. Symbols used: “X”: protein with no orthologs identified in a specific clade. “O”, only one fungal species of the subphylum retains the ortholog. Situations where an ortholog has been successfully identified in two or more species are left blank. Background color indicates the model species from which the seed protein was used: *S. cerevisiae*, white; *H. sapiens*, orange; *S. pombe*, yellow; *N. crassa*, green; *M. oryzae*, red.

A number of core proteins and proteins complexes were lost in some branches of the fungal tree of life

We found significant protein preservation patterns within the studied RNA-associated processes (Fig. 3.4.2), although many exceptions can be observed (Table 3.4.1). To further deepen in our analysis, detailed heatmaps were built to visually display the orthologs found for every constituent of the nine RNA-associated protein machineries (Annex Fig. 1).

RNA Transcription

Eleven protein groups of the transcription machinery were studied, including RNA Pol core subunits, Pol I-, II- and III-specific and general subunits, Mediator, other Pol II regulators, transcription factor (TF) I holoenzyme (TFIH), TFIH regulators, TF II holoenzyme (TFIIH) and TF III holoenzyme (TFIIH). We observed that these protein groups are usually well-preserved with the exception of Mediator complex, Pol I-specific subunits and TFIH (Fig. 3.4.2A; Annex Fig. 1.1; Table 3.4.1).

mRNA Splicing

Within the splicing machinery, protein components reflected high levels of preservation, with the exception of Microsporidia (Fig. 3.4.2B). This may be consistent with the fact that the Microsporidia genome has few introns, which likely reduces the necessity for a well-preserved splicing machinery [369]. A number of proteins appear to be Saccharomycotina-specific: Snt309, Spp381 Prp38 and Ntr2 (Table 3.4.1). The 11 SR proteins described in *H. sapiens* [65,66] are not fully retained in the fungal kingdom. SRSF11 is lost in Entomophthoromycotina, Mucoromycotina and Ascomycota, while SRSF3 and SRSF7 are missing in Taphrinomycotina and Pezizomycotina (Table 3.4.1). Curiously, Glomeromycotina, Mucoromycotina and Entomophthoromycotina are the only subphyla inside the fungal kingdom that still retain snRNAs related with the minor spliceosome (namely, U4atac, U6atac, U11 and U12) (Suppl. Fig. 3.4.S1). This is also reflected in the preservation of U12-associated proteins like ZMAT, SNRNP48 and PDCD7 (Fig. 3.4.2A; Annex Fig. 1.1.20; Table 3.4.1) Elements of the minor spliceosome were already identified in *R. oryzae* [67], and our results could support the idea that this catalytic

machinery could have been retained in some branches of the fungal tree, mainly in species belonging to the phylum previously known as “Zygomycota”.

mRNA 5'/3'-ends processing

5' capping is performed by three proteins Cet1, Ceg1 and Abd1, well preserved among the studied species (Fig. 3.4.2C), with the exception of *L. bicolor*, where both Cet1 and Ceg1 appear to be absent (Table 3.4.1, Annex Fig. 1.22). Moreover, no Cet1 ortholog was found in *B. dendrobatidis*, *R. globosum* and *A. algerae*. Therefore, how 5' capping is performed in these species remains unclear. The core polyadenylation machinery components are generally well preserved. The most variable complex of the polyadenylation machinery is the Cleavage Factor I complex (CFIm), which is absent in Saccharomycotina (with the notable exception of *Y. lipolytica*), Taphrinomycotina, Kickxellomycotina, Microsporidia and some basal fungi (*M. circinelloides* and *C. coronatus*). Most Cleavage and Polyadenylation Factor (CPF) subunits are present in the fungal kingdom with the notable exception of yeasts-specific proteins Syc1 and Ref2.

mRNA export and localization

We observed a lesser degree of proteins preservation compared to the three machineries examined above, especially for the Nuclear Pore Complex (Fig. 3.4.2D). Microsporia lacked the majority of components. Surprisingly, the essential nuclear export receptor Mtr2 was not found outside of the budding yeast group (Table 3.4.1, Annex Fig. 1.25), raising questions about how mRNA export is achieved in every other fungal species. Out of the four members of the THO complex, Hpr1, Mtf1, Thp2, and Tho2, the first three proteins are specific to *S. cerevisiae* and *K. lactis* (Annex Fig. 1.27). We included the human THO complex proteins Thoc1, Thoc5 and Thoc7 in our analysis as well. We observed that orthologues of these were present in almost all fungal species except yeasts, suggesting that components of the THO complex have followed different evolutionary paths in Saccharomycotina. Sus1, a component of TREX-2 that also regulates transcription [370], was lost in Pezizomycotina with the exception of *T. melanosporum* (Table 3.4.1, Annex Fig. 1.28). TREX-2 components like Thp1 and Cdc31, were missing in Basidiomycota (in the class Agaricomycetes) and Mucoromycotina, respectively.

The NPC is comprised of more than 30 different proteins, termed nucleoporins (Nups) [87]. On its cytoplasmic side, Nup159 and Nup42 interact with the DEAD-box RNA helicase Dbp5 and its co-factor Gle1 [371]. These four proteins together with those comprising the NPC central tube were well preserved in the fungal kingdom. Instead, NPC nuclear basket main proteins Nup1 and Nup60 are not universally preserved (Annex Fig. 1.33). The NPC is also composed of an inner ring, an outer ring, and a transmembrane ring. The yeast transmembrane

ring subunit Pom34 only identifies an ortholog in *K. lactis*, while *S. pombe* annotated ortholog pom34 was well conserved in rest of Ascomycota (Annex Fig. 1.31). Inner and outer ring subunits Nup53, Nup85, Nup157, Nup37, Nup43 and AAS were sparsely missing in many fungal clades (Annex Fig. 1.34-35).

Regarding mRNA localization, it has been studied extensively in budding yeast with *ASH1*, a transcription inhibitor [89]. Localization of *ASH1* mRNA requires at least five RNA-binding proteins (She2, She3, Loc1, Khd1, Puf6). Our results confirmed the presence of She2 and She3 only in the Saccharomycotina clade while Loc1 is Ascomycota-specific (Table1; Annex Fig. 1.29). Transport and localisation of mRNA is also being studied in *U. maydis* [372]. The Rrm4 protein essential for polarity and membrane-coupled mRNA trafficking in *U. maydis* has been lost in Pezizomycotina.

mRNA degradation and decay

The mRNA degradation machinery displays an uneven degree of preservation (Fig. 3.4.2E), and we observed even bigger differences at single-protein level among fungal species (Table 3.4.1). Of the nine essential subunits known to be part of the nuclear and cytoplasmic exosomes in yeast [373], the non-catalytic components Rrp43 and Csl4 are only found in Saccharomycotina and Dikarya, respectively. Mpp6 is present exclusively in Saccharomycotina although a functional ortholog exists in humans [374]. However, we were able to identify proteins containing the yeast Mpp6 domain in several other fungi (Annex Fig. 2.36). The TRAMP proteins Air1/Air2 were not found in most Pezizomycotina and Kickxellomycotina species, suggesting a possible alternative function for the TRAMP complex [300]. The Ccr4/Not complex is comprised of nine subunits and Caf130 is the only one that has no orthologs outside Saccharomycotina (Annex Fig. 1.40). Nrd1 and Nab3, components of the NNS complex required for transcriptional termination of cryptic unstable ncRNAs (CUTs) and stable ncRNAs like snoRNAs and snRNAs [115], are absent in some basal fungi and most Basidiomycota (Annex Fig. 1.38). 5'-end cytoplasmic degradation of mRNA is poorly preserved (Fig. 3.4.2E, Table 3.4.1, Annex Fig. 1.42). Edc1 and Edc2 are only found in *S.cerevisiae*, and apart from yeasts, Dxo1 is only found in Mucoromycotina and Kickxellomycotina.

mRNA surveillance

The preservation of the three best-studied cytoplasmic surveillance pathways, Nonsense mediated decay (NMD), Nonstop mRNA Decay (NSD) and No-Go Decay (NGD) is diverse among fungi (Fig. 3.4.2F). The NMD-specific proteins Upf1, Upf2 and Upf3 and the DEAD-box RNA helicase Dbp2 are found in almost every fungal clade. However, Ebs1 is hemiascomycetes-specific (Table 3.4.1, Annex Fig. 1.45). When including metazoan NMD factors Smg1-8 in our analysis, we discovered that proteins such as Smg5, Smg6 and Smg7 are highly preserved in

Dikarya, with the exception of the Saccharomycotina. Regarding NGD, it was stated in previous studies that Hbs1 is present in every eukaryotic and archaeal organism [375] (and therefore the reason to believe that NGD was the first surveillance mechanism developed by life), we found however that Microsporidia is again the exception (Annex Fig. 1.46). The ribosome quality control complex (RQC) is highly preserved with the exception of Rqc1 and Rqc2 (Tae2), which is absent in Microsporidia and Basidiomycota, respectively (Table 3.4.1, Annex Fig. 1.43).

RNA silencing

The RNA silencing machinery is the least preserved in the fungal species selected (Fig. 3.4.2G). Saccharomycotina lacks almost the entire set of RNA interference (RNAi) proteins, as *T. deformans* (Taphrinomycotina), *U. maydis* (Basidiomycota) and the microsporidial species *E. cuniculi* and *N. parisii*. None of the RNA-dependent RNA polymerases (RdRPs) were found in individual species like *T. deformans*, *U. maydis*, *C. coronatus*, *B. dendrobatidis*, *E. cuniculi* and *N. parisii* (Annex Fig. 1.49). The RecQ helicase QDE-3 is only absent in *R. globosum*, and the exonuclease QIP is completely missing in Kickxellomycotina and Microsporidia. Saccharomycotina is the only fungal clade where no homolog of Dicer and Argonaute proteins were identified. One of the Dicer proteins, Dicer2, is absent in other species, including all Taphrinomycotina (Table 3.4.1, Annex Fig. 1.48). The Argonaute QDE-2 protein of *N. crassa* is absent in Saccharomycotina (except *C. albicans*) and Kickxellomycotina, while the Argonaute Sms-2 is absent in Basidiomycota in addition to Saccharomycotina (Table 3.4.1, Annex Fig. 1.47). Remarkably, we identified a novel Argonaute protein in our analysis, not annotated in any public database, generally less preserved than QDE-2 and SMS-2 (Table 3.4.1, Annex Fig. 1.47).

SAD-1 to SAD-5 proteins are components of the meiotic silencing by unpaired DNA (MSUD) pathway described in *N. crassa* [127]. Sad-1 is absent from Saccharomycotina and Kickxellomycotina. SAD-2, SAD-4 and SAD-5 are restricted to a few Pezizomycetes closely related with *N. crassa* (Table 3.4.1, Annex Fig. 1.50). SAD-3 was absent from Basidiomycota and Saccharomycotina (except *Y. lipolytica*).

tRNAs maturation pathways

Globally, proteins involved in tRNA maturation are relatively well preserved (Fig. 3.4.3H). The subunits Pop1, Pop3, Pop4, Pop5, Rpp1 and Rrp2 responsible for 5'-end tRNA processing were found in most fungal species, with the exception of Microsporidia where the majority of RNase P components are missing (Table 3.4.1, Annex Fig. 1.53). Pop6 and Pop8 subunits are only present in ascomycete yeasts whereas Pop7 displays a complex pattern. Pop7 are found annotated in *S. cerevisiae*, *S. pombe* and *H. sapiens* but they do not share a clear homology with each other. *S. cerevisiae* Pop7 is only found in Saccharomycotina and Pezizomycotina.

Remarkably, *S. pombe* pop7 is also in *R. irregularis*, and human POP7 is well preserved in Basidiomycota and many basal fungi. RPP25 is another human subunit with orthologs in some basal fungi.

The tRNAs 3'-end processing activity is performed by three proteins, Lhp1 (La), Trz1 and Rnh70 (Rex1), all well preserved in the fungal kingdom with the exception of Rnh70, which is not found in Kickxellomycotina, some chytrids and Microsporidia (Annex Fig. 1.54). After 3'-end removal, a CCA tri-nucleotide is added to the tRNA. Cca1 is universally preserved with the exception of *R. globosum*. During maturation, tRNAs also undergo an extensive round of chemical modifications, most of them described in the model organism *S. cerevisiae*. Some tRNA methyltransferases (TRM) like Trm3 and Trm13 are missing from higher fungal clades like Pezizomycotina and Basidiomycota (Annex Fig. 1.57-58). Deamination of adenosine in position 34 in all tRNAs (likely to influence codon-anticodon recognition) [360] is catalyzed by the heterodimeric enzyme ADAT2/ADAT3 (Tad2/Tad3). Strangely, Tad3 is absent from almost all Basidiomycota, raising questions regarding how the heterodimer is formed in these species. The addition of wybutosine (yW37) in yeast tRNA^{Phe} is catalyzed by four different enzymes Tyw1, Tyw2 (Trm12), Tyw3, and Tyw4 (Ppm2). Tyw1, Tyw2 and Tyw3 are missing from most of non-Ascomycota fungi (Annex Fig. 1.57).

GTPase Ran regulates the export and re-export of tRNA by its association with Los1, Mtr10 and Msn5 [287] The latter is the homolog of human Exportin-5, a protein also involved in the export of pre-miRNAs [376]. Msn5 is not found in Basidiomycota, Mucoromycotina, Microsporidia, *R. globosum* and the *Conidiobolus* genus (Annex Fig. 1.56), suggesting that tRNA re-export or transport of dsRNA-binding proteins work differently in these species.

mRNA translation

Proteins associated with translation are highly preserved throughout the fungal kingdom (Fig. 3.4.2I), with the exception of mitochondrial ribosomal proteins which are missing from Microsporidia, fungi known to lack mitochondria [377]. A sparse few proteins are absent from specific subfamilies (Table 3.4.1), including the translation initiation factor eIF5 (Tif5; Annex Fig. 1), a component of unknown function of the eIF3 complex (Clu1), the translation elongation factor 1 beta (Efb1; Annex Fig. 1.64), and the translation elongation factor 3 (Yef3). It is worth noting that Saccharomycotina lacks three proteins related to transcription initiation in metazoans: the eIF3e subunit Int6, the translation initiation factor eIF3d Moe1 and the mitochondrial translation initiation factor Mti3.

Higher and Basal fungi evolved different preferences for specific groups of domains related with RNA processing

Protein domains within each protein in our analytical set were predicted for every ortholog identified, using the Pfam database [193]. To identify the domains exhibiting interesting patterns of preservation within each RNA-related process, we calculated the percentage of presence/absence of every domain identified (Fig. 3.4.3). In these PCA (Principal Components Analysis) biplots each fungal clade that tended to co-locate close showed globally similar domain preservation considering all the RNA-related processes, as it was the case for the basal fungi Kickxellomycotina, Mucoromycotina and Entomophthoromycotina. Domains distal from the center of the graph had higher variability and tended to locate close to the fungal clade in which the domain was most highly preserved (Fig. 3.4.2), as in the case of NUP50 (top-left quarter) and Mtr2/SHE3 (top-right quarter) domains, which are only present in Chytridiomycota and Saccharomycotina, respectively. The basal fungal clades Mucoromycotina, Entomophthoromycotina and Kickxellomycotina located close to one another, indicating that they shared a similar domain preservation pattern. The same applied to the Dikarya fungi within Pezizomycotina, Basidiomycota and Taphrinomycotina. However, Saccharomycotina represented an exception since several domains were only present within this fungal subphylum, such as SAC3, PRP9_N and RNA_poll_A14. Chytridiomycota were located far from the other basal fungi, likely due to the effect of *R. globosum*, a fungus with less genomic variability in terms of domains identified.

Significantly, the majority of differently preserved domains were found in proteins related with the transcription machinery (orange colour, Fig. 3.4.3). Five Mediator domains (Med5, Med11, Med16, Med13_C, Med25_VWA and Rox3) were particularly variable in the fungal kingdom, and found mainly within Saccharomycotina, consistent with the orthologs preservation analysis (Fig. 3.4.2). Regarding mRNA splicing, Saccharomycotina was the only fungal group possessing both the N-terminus and SF3a60 binding domains of Prp9. The splicing factor 3b subunit 1 specific domain (SF3b1) was also absent in Saccharomycotina, suggesting that the SF3a/SF3b splicing factor complexes may have diverged differently during evolution in ascomycete yeasts. Domains associated with pre-mRNA 5' and 3'-end processing exhibited less conspicuous patterns. Symplekin_C was absent from Ascomycota, Taphrinomycotina and Mucoromycotina, although remarkably, all Basidiomycota species retained this domain. The polyadenylation-associated RBM39linker domain was only found in orthologs of Hrp1 except within the Saccharomycotina fungi and CFIA_Pcf11 also showed a heterogeneous level of preservation, being lost in Basidiomycota and most of basal fungi.

Figure 3.4.3 | Analysis of domain preservation in fungal clades. PCA (Principal Component analysis) plot reconstructed from the average percentage of preservation of every domain in each fungal clade. Clades that locate close to one another present similar domain preservation within the RNA-related process considered in this study, e.g. the basal fungi Kickxellomycotina, Mucoromycotina and Entomophthoromycotina. Only domains with a significant distance from the center of the graph are shown (see Materials and methods). Domains distant from the center of the graph have higher variability and tend to locate close to the fungal clade in which the domain is most highly preserved. Domains are indicated using their Pfam identifier [193]. Microsporidia and Glomeromycotina were not included.

With regard to mRNA export, it displayed variable preservation in domains associated with the Nuclear Pore Complex (NPC) such as NPCC, Nup88, NUP50, NUP214 and Nup192, confirming its heterogeneous nature [87]. Mtr2, SWI5-dependent HO expression protein 3 (SHE3) and SAC3 domains were only observed in Saccharomycotina within their respective homonym proteins Mtr2, She3 and Sac3. The TPR_MLP1_2 domain belonging to Mlp1, a myosin-like protein associated with the nuclear envelope, was missing in Mucoromycotina and Chytridiomycota.

Concerning domain preservation in proteins associated with the RNA silencing machinery, we checked for Argonaute/Dicer/Rdrp-related domains. Saccharomycotina, *T. deformans*, *U. maydis*, *E. cuniculi* and *N. parisii* seems to miss every fundamental RNA silencing domain, while other species lack only few of these domains: PAZ, a domain found in both Argonaute and Dicer proteins, missing in *P. graminis* and *V. dahliae*; the Dicer dimerization domain (Dicer_dimer), essential for Dicer activity, is not found in *C. reversa* and *A. algerae*; and

Rdrp, the main domain of RNA-dependent RNA polymerases, is missing in *C. coronatus*, *B. dendrobatidis* and the whole Kickxellomycotina subphylum.

tRNA maturation domains wyosine_form and TYW3 involved in wybutosine formation were absent from Basidiomycota, Kickxellomycotina and Chytridiomycota, suggesting that the nucleotide position 37 of the tRNA^{Phe} must be modified differently in non-Ascomycota fungi, or non-existent.

Despite that the translation machinery was well conserved in the fungal kingdom (Fig. 3.4.1), global domain analysis revealed that ribosomal proteins showed distinct levels of preservation (Fig. 3.4.2). We detected a clear separation between Dikarya and basal fungi, with many translation-related domains clustering far from higher fungi. These elements included ribosomal domain ribosomal_L24, Ribosomal_L33, Ribosomal_L6e_N, 40S_SA_C, initiation factor domains IF3_N and eIF-3_zeta, and elongation factor domain EFP_N.

Genomic and transcriptomic analysis reveals clade-specific features

In addition to their genome sequences, all the fungal species selected for this study had at least one transcriptomic dataset, which we downloaded and analyzed to find potential links with our protein conservation analyses.

Common elements are found in upstream genic regions in fungi

The upstream regions (250 bp from the ATG) of each fungal species total gene complement were retrieved and analyzed looking for *cis*-regulatory elements. The consensus TATA-box sequence TATAAA [378], which is usually located 25-35 bp upstream of the TSS, was found to be variably enriched across the fungal kingdom, without a clear evolutionary pattern (Fig. 3.4.4A). Conversely, the location of the TATA-box appeared to be clearly associated with the fungal clade. In Pezizomycotina and Kickxellomycotina this box was located quite far from the start codon (~125 bp), while in Microsporidia it was notably closer (~50 bp).

We identified two additional short motifs enriched in fungal genes always downstream of the TATA-box, meaning that these elements are possibly present in the 5'UTRs and can regulate translation initiation (Fig. 3.4.4A). A poly-pyrimidine trait was found located approximately 30bp-80bp upstream of the ATG, depending on the species. In mammals this element is recognized by the polypyrimidine-tract-binding protein and regulates translation initiation [379]. The second element is a CA-rich region usually restricted to fungi. The average position of the CA-rich region is very close to the start codon (~25 nt). To confirm the presence of these two elements in the transcribed mRNAs, we checked in the 5'UTR of the transcripts dataset. The polypyrimidine trait is especially enriched in Mucoromycotina and Kickxellomycotina (Fig. 3.4.4B), while the CA-rich element is generally enriched in Dikarya, but absent in Saccharomycotina with the exception of *Y. lipolytica* (Fig. 3.4.4C). To our knowledge the CA-rich region has not been described before, thus its role and relevance in gene expression is unknown.

mRNA polyadenylation sites presents common features and a few variable elements among fungi

Five key sequence elements are found in *S. cerevisiae* polyadenylation sites [299]: the A/U-rich Efficiency Element (EE) located at variable positions upstream of the cleavage site with the nucleotide consensus sequence UAYRUA; the A-rich Positioning Element (PE) also found in mammals [380], located 10–30 nucleotides upstream of the cleavage site; the Upstream U-rich Element (UUE); and the Downstream U-rich Element (DUE). In our analysis, we found a generally conserved nucleotide profile around the cleavage site, with the presence of different elements downstream and upstream of the PAS depending on the fungal clade (Fig. 3.4.4D). The consensus arrangement of polyadenylation elements in the fungal kingdom reveals presence of both UUE and DUE, with the exception of Kickxellomycotina where DUE was not clearly identified. The canonical mammalian polyadenylation signal is the hexamer AAUAAA, which has been shown to be conserved in other species but not universally necessary for cleavage. Basal fungi have a clear AAUAAA-like motif, likely mirroring the region defined in humans, while Dikarya fungi display a more heterogeneous set of sequences, especially the Basidiomycota.

Hrp1 recognises the A/U-rich efficiency element (EE) consensus UAYRUA sequence in yeast [219]. This sequence seems to be particularly enriched in Saccharomycotina, and less evident in other clades. Common alternative motifs identified in this region are UGUA (motif recognised by the mammalian CFIm complex, and predicted binding region for Rbp35 in Chapter 3.1) and UAGNH (the predicted binding region for Hrp1 in Chapter 3.1), both typically found in higher fungi. No enriched sequences upstream of the polyadenylation signal in basal fungal families like Chytridiomycota, Kickxellomycotina and Mucoromycotina were found. Nevertheless, they all display large U-rich regions, suggesting that other elements may be present. UGUA is also found downstream of the DUE in Taphrinomycotina, corroborating previous results found in *S. pombe* [381]. A different motif, HGUGA, is clearly discernable in Glomeromycotina, Mucoromycotina, Basidiomycota and Pezizomycotina.

Average 3'UTR lengths differ significantly between the selected fungal clades (Fig. 3.4.4E). Fungal 3'UTRs are usually shorter than their Metazoan or Plant counterparts, with Pezizomycotina having the longest 3'UTRs (~200 nt), and Microsporidia the shortest (~30 nt). We also predicted RNA secondary structures, which are known to regulate PAS recognition and cleavage [382]. The RNA secondary structure was computationally reconstructed for every predicted PAS, and a consensus base-pair probability plot was outlined (Fig. 3.4.4F). The base immediately before the cut-site appeared usually paired, followed by an unpaired base and flanked by paired regions, suggesting that the cut-site could be located in a hairpin structure. DUE and UUE were also usually paired, while the A-rich region was not.

Figure 3.4.4. | Prominent features of genic regions in fungi. A) Common elements found in upstream regions of the start codon. The TATA-box (blue box) is variably conserved and its location appears to be clearly defined based on the fungal clade. A poly-pyrimidine (py-RICH) trait is located 80bp-30bp upstream of the ATG and always downstream of the TATA-box. A CA-rich region (purple box) ~25bp upstream of the start codon is found specifically enriched in Dikarya. **B)** The polypyrimidine rich motif TYTYTY found in the 5'UTR*. **C)** The CA-rich motif CAHCAH found in the 5'UTR*. **D)** Elements found in 3'UTRs. The consensus arrangement of polyadenylation elements in fungi shows both UUE and DUE boxes. A search for enriched motifs in the A-rich region shows in fact that basal

fungi have a clear AAUAAA-like motif in this region, while Dikarya fungi display a more heterogeneous context, especially Basidiomycota. **E)** Average predicted 3' UTRs lengths based on the distances of the predicted polyadenylation sites (PASS) and the relative gene stop codons. **F)** RNA secondary structure prediction of the region surrounding the PAS. The graph shows the predicted base-pair probabilities calculated as average of all the species considered. A high base-pair probability means that in the predicted secondary structure that base is probably paired with another base in the molecule. The base right before the cut-site appears to be usually paired, followed by an unpaired base and flanked by paired regions, suggesting that the cut site could be located in a hairpin structure. DUE and UUE are usually paired, while the A-rich region is not.

*Motif levels are represented as the difference (in percentage) between the occurrence of the motif and the expected frequency of the motif in the sequences.

mRNA splicing features

Analysis of fungal introns suggested that evolution has shaped gene exon organization in different ways in the fungal kingdom (Fig. 3.4.5A). Saccharomycotina species are an exception since they tend to have few and larger introns per gene. The rest of the fungi generally have the same median intron size, but a different average number of introns per gene. Within each clade there is a degree of heterogeneity, particularly in Kickxellomycotina where, for example, *L. pennispora* has many short introns. On the other hand, introns number and size in Pezizomycotina are very uniform. Notably, fungi are distinct from plants and metazoans, which usually possess a greater number of introns per gene and of bigger size. Introns of all budding yeasts present a positional bias toward the 5'-end of the gene [55]. According to the average relative intron position (Fig. 3.4.5B), Saccharomycotina seem to be unique with respect to this 5' positional bias.

Spliceosomal introns present four defining features: the 5' splice site (5'ss, donor), the 3' splice site (3'ss, acceptor), the branch point and the polypyrimidine tract. The sequences defining these features are well conserved in all eukaryotes, where introns are flanked by a GT dinucleotide at the 5'-end and an AT dinucleotide at the 3'-end [56]. In our analysis, each feature followed a distinct evolutionary pathway within each fungal clade (Fig. 3.4.5C). In *H. sapiens* and *A. thaliana* there is also a preference for certain nucleotides on the exonic 3'-end [383]. This property is not found in fungi, where there is no uniformity among the exonic flanking 3' nucleotides. The preferred acceptor sequence was identified as YAG in *H. sapiens*, *A. thaliana* and *S. cerevisiae*. This is also true for most Dikarya, while in basal fungi there is a stronger preference for TAG with the exception of *L. pennispora*. In Saccharomycotina the acceptor site region is much less uniform than the rest of fungi and other species, though this general trend is influenced largely by the characteristics of *Candida albicans*. Regarding the polypyrimidine tract, a 15-20 nt sequence recognised by splicing factor U2AF and usually located about 5-40 nt before the 3'ss, there is scattered evidence of its presence in fungi. We found it only in Saccharomycotina and some basal fungi. The *S. cerevisiae* branch point recognised by Splicing

Factor 1 has been shown to be represented by the consensus motif UACUAAC [384]. This motif is not well conserved in other fungi, but they retain a recognizable CUNAC structure. Branch point distance from the 3' ss has been associated with splicing efficiency, lariat formation and alternative splicing [385]. Most of fungal clades have an average branch point distance from the 3'ss of approximately 35 nt. Taphrinomycotina have the shortest (~15 nt) and Kickxellomycotina the longest (~45 nt). Therefore, fungi branch points are usually closer to the 3'-end compared to metazoans and plants.

Figure 3.4.5 | mRNA splicing features in fungi. A) Average number of introns per genes for each fungal clade plotted against the median intron size. Saccharomycotina species possess usually one big intron per gene, while intron size in the rest of fungal kingdom is generally homogeneous. Taphrinomycotina usually has few introns per gene and Basidiomycota has many introns per gene. Inside each clade we can find a certain heterogeneity, especially in Kickxellomycotina (*L. pennispora* has many short introns), while Pezizomycotina are very uniform. We can observe that fungi diverge from Plants and Metazoan, which usually possess a higher number of introns per genes, especially in *H. sapiens* which is not included in this graph. **B)** Average introns position bias represented as percentage over the total intron length; only Saccharomycotina display intron position bias. **C)** Typical intronic feature conserved in the considered fungal clades; Donor 5' splice sites in Dikarya have an evident GT dinucleotide at position 5 and 6, while Mucoromycotina, Glomeromycotina and Entomophthoromycotina have an AA dinucleotide at position 3 and 4. Chytridiomycota somehow look like Dikarya, while Kickxellomycotina have a degenerate donor sequence, mostly due to *L. pennispora*. In *H. sapiens* and *A. thaliana* there is also a preference for some nucleotides on the exonic 3'-end side. This property is typically not found in fungi, where there is no uniformity among the exonic flanking 3 nucleotides. The preferred acceptor 3' splice site sequence was identified as YAG in most Dikarya, while in basal fungi there is a stronger preference for TAG. There is a scattered evidence of polypyrimidine trait's presence in fungi, where it is visible only in Saccharomycotina and some basal fungi. Branch point motif is highly homogeneous in Saccharomycotina, less so in other fungi. Entomophthoromycotina exhibits an uncommon consensus branch point motif.

tRNA features

The tRNA gene transcription termination signal consists of a stretch of thymine (T) nucleotides [386]. This stretch is generally longer in *S. cerevisiae* than *H. sapiens* [387]. We analyzed the 3' flanking sequence of tRNA of each organism in our dataset (Fig. 3.4.5A). As noted earlier, previous studies showed that fungi usually possess longer T-stretches compared to Metazoan, or even Oomycota and Plants. From our investigation we observe that Saccharomycotina possess the longest T-stretches, on average.

Eukaryotes tend to possess multiple tRNA copies for the same amino acid, but the anticodon is not always the same, and anticodon preferences can vary among different species. We observed that, in general, anticodon preferences are quite consistent among the selected species, but some tRNAs display certain variability (Fig. 3.4.5B). For example, the preferred anticodon for Lys in *H. sapiens* is CUU, while in fungi it is usually UUU. The PCA shows which amino acids tend to be represented by atypical anticodons in certain fungal clades as compared to the average fungal preference. Anticodons for Glutamine and Glutamic acid are generally different in Saccharomycotina, Entomophthoromycotina and Glomeromycotina compared to other groups. A similar distinction is noted for Valine in Microsporidia and Glycine in Mucoromycotina.

Figure 3.4.6 | tRNA features in fungi. A) Average tRNA genomic 3' T-stretches lengths. Fungi usually possess longer T-stretches respect to Metazoan, and even Oomycota and Plants. Saccharomycotina also presents the longest T-stretches on average. **B)** General anti codon preferences is quite constant among the selected species, but some tRNAs display certain variability. This PCA plot shows which amino acids tends to be represented by unusual anticodons in certain fungal clades in comparison with the average preference. It can be noticed that anticodons for amino acids like glutamine and glutamic acid are generally different in Saccharomycotina, Entomophthoromycotina and Glomeromycota than the rest of the groups. This also applies for valine in Microsporidia and glycine in Mucoromycotina.

Discussion

Many proteins of critical importance to human biology were first discovered by studying their homologs in *S. cerevisiae*, and other model fungal organisms that have contributed to a broad understanding of cellular biology including *A. nidulans*, *N. crassa*, *S. pombe* and *M. circinelloides*. Nevertheless, the startling diversity of fungal life forms suggests that a great deal of diversity might be identified within each fungal species. Indeed, we see that core molecular processes such as RNA metabolism - known to be highly conserved across all species - exhibits divergence, even total absence, with respect to some well-established components that would previously have been considered essential. Single proteins but also entire complexes with a clear homology encompassing high evolutionary distance from human to *S. cerevisiae* may have been disposed-of in intermediate fungal branches.

The results described in this paper represent a formal *in silico* evolutionary approach of the global conservation status of molecular machineries involved in RNA metabolism specifically within the fungal kingdom. We have examined the set of protein complexes that take part in these processes at two different levels of conservation: orthology and protein domain conservation. Elements involved in mRNA translation (namely, during the phase of translation initiation, elongation and termination) and mRNA splicing appear to be the best conserved, at least taking into account the presence of orthologs, the domains involved, and the overall similarity of the proteins. On the other hand, well studied multi-protein complexes like Mediator, Cleavage Factor Im and the Nuclear Pore Complex display many missing parts in some fungal clades. RNA silencing seems to be the most heterogeneous machinery, possibly completely missing with many organisms, implying that these organisms may have developed alternative strategies to cope with invasive repetitive nucleic acids like transposable elements and viruses. mRNA export-related structures also appears to have diverged significantly, especially the Nuclear Pore Complex. These results are sustained by the domains analysis, where it seems to be evident a strong dycotomy between basal and higher fungi regarding the preservation of RNA-associated domains.

The cross-species orthologs identification provided us with a comprehensive understanding about which of the fundamental building blocks of RNA metabolism are available for each fungus, even though this knowledge is still biased toward the model organisms where this information has been fetched. Similarly, the analysis of transcriptomic data delivers us with a representation of the outcome produced by the model envisioned by the previous examination. Transcriptomic analysis revealed some interesting characteristics; for example, even evolutionary close taxa where a certain machinery display a nearly identical homology, can results in strikingly different transcriptomic products. We noticed that each fungal group exhibits a unique polyadenylation profile; even well established 3'UTR processing signals are not

homogeneously conserved. Sometimes the change is more subtle, like for the splicing branch point, which usually presents a similar motif but with a different information content and intronic location.

This study unveiled a few surprising characteristics regarding the uniqueness of some fungal groups. Saccharomycotina, the sub-phylum that includes the top model organism *S. cerevisiae*, shows striking differences with the rest of Ascomycota, usually represented by a global simplification of the cellular toolset, probably as a consequence of the lost of true filamentous growth capabilities. On a wider scale, it appears evident that there is a clear separation between “higher” and “basal” fungi (Dikarya and non-Dikarya), giving strength to this informal denomination that has been in use for many decades. Some non-RNA related elements like the flagellum neatly separates them, but as we have appreciated in this study many proteins are either retained or lost in block in one or the other group.

We hope that this large-scale survey of the components of key RNA processing machineries in the fungal kingdom may create the groundwork for further analysis that could identify correlations between the evolutionary history of certain proteins or domains, and biological features such as lifestyle or pathogenicity.

Supplemental Data

Supplemental Table 3.4.S1 | Species used with genome assembly information and transcriptomic source.

Species	Strain	Genome assembly	Transcriptomic source
annalia algerae	pra109	GCA_000385855.2	SRA137156
arabidopsis thaliana	TAIR10	GCA_000001735.1	Genbank ESTs
aspergillus nidulans	FGSC A4	GCA_000149205.2	Genbank ESTs
basidiobolus meristosporus	CBS 931.73	GCA_000697375.1	JGI ESTs
batrachochytrium dendrobatidis	JAM81	GCA_000203795.1	Genbank ESTs
blumeria graminis	DH14	GCA_000151065.1	Genbank ESTs
botrytis cinerea	B05.10	GCA_000832945.1	Genbank ESTs
candida albicans	wo-1	GCA_000149445.2	SRA098283
candida tenuis	ATCC 10573	GCA_000223465.1	Genbank ESTs
coccidioides posadasii	Silveira	GCA_000170175.2	Genbank ESTs
coemansia reversa	NRRL 1564	JGI Project Id: 402455	Genbank ESTs
conidiobolus coronatus	NRRL28638	GCA_001566745.1	JGI ESTs
conidiobolus thromboides	FSU 785	JGI Project ID: 1015483	JGI ESTs
coprinopsis cinerea	okayama7#130	GCA_000182895.1	Genbank ESTs
cryptococcus neoformans	JEC21	GCA_000149385.1	Genbank ESTs
encephalitozoon cuniculi	GB-M1	GCA_000091225.1	SRA061149
fusarium oxysporum	FO2	GCA_000271705.2	Genbank ESTs
gloeophyllum trabeum	ATCC 11539	GCA_000344685.1	Genbank ESTs
gonapodya prolifera	JEL478	GCA_001574975.1	SRA050519
grosmannia clavigera	kw1407	GCA_000143105.2	Genbank ESTs
histoplasma capsulatum	NAm1	GCA_000149585.1	Genbank ESTs
homo sapiens		GCA_000001405.22	Genbank ESTs
kluyveromyces lactis	NRRL Y-1140	GCA_000002515.1	SRR1200527
laccaria bicolor	S238N-H82	GCA_000143565.1	Genbank ESTs
leptosphaeria maculans	JN3	GCA_000230375.1	Genbank ESTs
linderina pennispora	ATCC 12442	JGI Project Id: 1019509	JGI ESTs
magnaporthe oryzae	70-15	GCA_000002495.2	Genbank ESTs
martensiomycetes pterosporus	CBS 209.56	JGI Project Id: 1040168	JGI ESTs
metarhizium acridum	CQMa 102	GCA_000187405.1	Genbank ESTs
microbotryum violaceum	p1A1	GCA_000166175.1	Genbank ESTs
mucor circinelloides	1006PhL	GCA_000401635.1	Genbank ESTs / JGI ESTs
nematocida parisii	ERTm3	GCA_000190615.1	SRA052700
neurospora crassa	ATCC 24698	GCA_000182925.2	Genbank ESTs
phycomycetes blakesleeanus	NRRL1555	GCA_001638985.1	JGI ESTs
phytophthora infestans	T30-4	GCA_000142945.1	Genbank ESTs
pleurotus ostreatus	PC15	GCA_000697685.1	Genbank ESTs
podospira anserina	S mat+	GCA_000226545.1	Genbank ESTs
puccinia graminis	CRL 75-36-700-3	GCA_000149925.1	Genbank ESTs
rhizoclostridium globosum	JEL800	JGI Project Id: 1060049	JGI ESTs
rhizophagus irregularis	DAOM 181602	GCA_000439145.2	Genbank ESTs
rhizopus oryzae	RA 99-880	GCA_000697135.1	Genbank ESTs
saccharomyces cerevisiae	R64-1-1	GCA_000146045.2	Genbank ESTs
saitoella complicata	NRRL Y-17804	GCA_000227095.1	Genbank ESTs / JGI ESTs
schizosaccharomyces pombe	972h-	GCA_000002945.2	Genbank ESTs
taphrina deformans	PYCC 5710	GCA_000312925.1	Genbank ESTs
trichoderma reesei	QM6a	GCA_000167675.2	Genbank ESTs
tuber melanosporum	Mel28	GCA_000151645.1	Genbank ESTs

<i>ustilago maydis</i>	521	GCA_000328475.2	Genbank ESTs
<i>vavraia culicis</i>	floridensis	GCA_000192795.1	Genbank ESTs
<i>verticillium dahliae</i>	VdLs.17	GCA_000400815.2	ERA208100
<i>yarrowia lipolytica</i>	CLIB122	GCA_000002525.1	Genbank ESTs
<i>zymoseptoria tritici</i>	CBS 115943	GCA_000219625.1	Genbank ESTs

Supplemental Figure 3.4.S1 | Heatmap of spliceosomal snRNAs preservation.

4. General discussion

Fungi have occupied every environmental niche and they are of fundamental importance to many ecosystems. Interaction between fungi and humans range from culinary use to human diseases and from important mycorrhizal symbionts to crops pathogens. Knowledge of pathogen biology is paramount to the development of new methods of disease control, and understanding gene expression regulation is among the most studied aspect of cell physiology. *M. oryzae* is the top rice pathogen causing the rice blast disease, and one of the prime threats to global food security. The main purpose of this thesis was to investigate certain aspects of *M. oryzae* RNA metabolism using a bioinformatic approach, and to extrapolate these concepts to the whole fungal kingdom in order to get a more comprehensive evolutionary perspective.

The first results chapter showed the impact of alternative polyadenylation (APA) in *M. oryzae* transcriptome, and the importance of two pathogenicity-related proteins, Rbp35 and Hrp1, for a correct selection of the polyadenylation site in protein-coding genes. It was the first time that a genome-wide mapping of poly(A) sites was performed in the rice blast fungus, and it allowed us to identify at least one high-confidence cleavage site for about 65% of all genes currently annotated in *M. oryzae* genome. Knowing the exact nucleotide cleavage location is fundamental when dealing with 3' UTR processing analysis, since it has been shown that even a small difference of a few nucleotides could have huge consequences on the transcript's final fate. Half of annotated genes with a significant level of transcription during our experiment are subject to APA in one or more of the tested conditions. We demonstrated that nutrient stress conditions like carbon starvation can affect APA and thus nutritions represent an important factor in control of final gene expression. Carbon starvation is a typical situation faced by the fungus during the early stages of infection, and modulating 3' UTR length could be a way to provide more flexibility to gene regulation. We also provided more insight about the function of Rbp35 and Hrp1. Rbp35 is part of the polyadenylation machinery and a distant homolog of human CFI_{m68}. CFI_{m68} forms an heterodimer with CFI_{m25}, which recognizes UGUA tandem motifs and regulates APA, favoring the selection of distal poly(A) sites. We discovered a very similar situation in *M. oryzae*, where UGUAAH motifs are enriched in down-selected poly(A) sites in the Δ *rbp35* mutant (but they are not tandemly arranged as in human) and poly(A) selection favours more proximal poly(A) sites with respect to the wild-type. This suggests that the molecular function of the CFI heterodimer remained conserved from animals to fungi, despite many rearrangements of the global structure of the polyadenylation machinery. On the other hand, Hrp1 has no clear functional homolog in metazoans, and the most similar human protein, heterogeneous nuclear ribonucleoprotein A1, does not appear to be directly involved in 3' UTR processing and polyadenylation. The Δ *hrp1* mutant is severely affected in growth, development and pathogenicity, even more than Δ *rbp35*. This correlated with the fact that more genes display differential poly(A) site selection in Δ *hrp1* compared to Δ *rbp35*. The UAGNH motif, located a few nucleotides downstream of UGUAAH, is enriched in down-selected poly(A) sites in the Δ *hpr1*

mutant, indicating that this could be the RNA sequence recognized by Hrp1, which is itself a protein containing two RNA-binding motifs. The consequences of APA remain somehow obscure; previous studies have shown that changes in the 3' UTR length can be associated with transcript stability, translation efficiency and mRNA localization. In animals, the best known consequence of APA is to alter the availability of miRNA targets. miRNAs in animals and plants, modulate the stability of their target mRNA. miRNA targets are usually located in the 3' UTR of mRNA, and APA can potentially hide or reveal these target regions and therefore change the fate of the transcript. Nevertheless, miRNAs and their associated biogenesis proteins have not been found in fungi, suggesting that these mechanisms may not be present. Therefore, an alternative, plausible consequence of APA in fungi could be related to mRNA export, localization, stability/degradation, and/or translation efficiency. Nuclear export and cell localization signals are generally situated in the 3' UTR, and regulating poly(A) site selection could have an effect on how the transcripts should be exported into the cytoplasm and/or localized to a specific part of the cell.

Small RNAs (sRNAs) are currently a topic of great interest in fungal research, and mechanisms of gene expression regulation in fungi through RNA silencing are the subject of ongoing investigations. We analyzed sRNAs in *M. oryzae* and our results suggested that RNA silencing is the major actor in the control of transcription of repetitive elements in the rice blast fungus. Almost every retrotransposon present in the rice field isolate Guy11 (where the type and number of transposable elements are highly dependent on the strain) produces a considerable amount of sRNA with average length, closely resembling the typical small interfering RNA (siRNA). Bioinformatic analysis of *M. oryzae* sRNA data also allowed us to identify a novel mycovirus infecting *M. oryzae* strain Guy11, called MOLV1, which seems to be one of the prime targets of RNA silencing, given the high quantity of sRNAs derived from this virus. This ssRNA virus contains a single ORF encoding a putative RNA-dependent RNA polymerase, closely related to ourmia viruses (plant viruses) and ourmia-like mycoviruses. Similarly to earlier sRNA-seq experiments conducted on *M. oryzae*, only a small number of reads could be directly aligned to the reference genome. This raised the possibility that some sort of overlooked post-transcriptional modification could be acting on sRNAs in the rice blast fungus. We demonstrated that a pervasive modification involving the addition of a few trailing adenines is acting on sRNAs derived from repetitive elements. The cause and consequence of this modification are unknown. An analogous phenomenon occurs through the action of the TRAMP complex on snoRNA and rRNA, which serves as a target for degradation by the exosome. Based on the sRNA-seq analysis of $\Delta rbp35$ and $\Delta exp5$ mutants, we provided evidence that Rbp35 and Exp5 could be involved in sRNA regulation. Several transposable elements are differentially expressed in the mutants, and a specific sRNA population seems to be strongly affected, especially in $\Delta rbp35$. The MOLV1 virus in particular is dramatically depleted in $\Delta rbp35$, leading to the assertion that

Rbp35 (which immunoprecipitate with Ago, as demonstrated in another experiment from our lab), could have a part in RNA silencing.

RNA is a fundamental molecule to every organism and its metabolism is generally considered to be greatly conserved. There are, however, many components essential to the model organisms that are expendable in other close species. With the aim of giving a broader perspective to the conservation levels of RNA-related processes in the fungal kingdom, we performed a multi-species ortholog prediction based on hundreds of proteins described in the literature as being involved at some point within the 'RNA lifespan'. This global view indicates a high level of conservation, but nevertheless, dozens of genes of crucial importance in RNA biology in model organisms appear to be missing in specific species, or even in entire phyla. This evolutionary analysis also revealed a heterogeneous degree of conservation depending on the RNA-associated process considered; processes like mRNA translation and tRNA maturation present a high level of conservation, while other processes like mRNA export and RNA silencing are variably conserved among the selected fungal species. Regarding the polyadenylation machinery, we could observe that the CFIm complex (which includes Rbp35) has been lost in the majority of Saccharomycotina, Taphrinomycotina, Kickxellomycota and Microsporidia. A possible consequence of this loss is that these species cannot exploit the additional APA tuning capabilities provided by the CFIm complex, or perhaps that these capabilities have been taken-over by other proteins. The other CFI protein, Hrp1, is found in practically every fungal species. However, the RNA motif recognized by Hrp1 as described in yeast is generally not found in other species, suggesting a different role of this protein in polyadenylation process. Analogously, the core components of RNA silencing (a major regulator of repetitive elements as seen before) are poorly preserved in the fungal kingdom. RNA silencing seems to be completely absent in Saccharomycotina, and in some specific species like *Taphrina deformans* and *Ustilago maydis*. Moreover, paralogs of the three main components of RNA silencing, Ago, Dicer and RdRP, are present in different number depending on the species, denoting a broad diversification of this machinery. Exp5, studied in chapter 3.2 for its involvement in import/export and possibly RNA silencing, is completely missing in Basidiomycota. The enormous dataset of orthology relationships, domain predictions, and transcriptomic analysis resulting from this large-scale survey of the components of key RNA processing machineries in the fungal kingdom may provide a valuable source of information for other researchers and encourage further investigation in order to unveil the evolutionary history of gene expression and regulation in the fungal world.

Conclusions

1. The exact polyadenylation sites have been identified for 70% of annotated genes in the *M. oryzae* genome. Half of the genes of *M. oryzae* display alternative polyadenylation in the wild-type strain.
2. Polyadenylation site selection is affected during carbon starvation, resulting in the transcription of mRNA with longer 3'UTRs. The selection appears to be correlated with the enrichment of the Upstream U-rich Element (UUE).
3. Polyadenylation site selection is affected in the two mutants considered: $\Delta rbp35$ and $\Delta hrp1$. These results indicate that both Rbp35 and Hrp1 proteins are involved in the regulation of alternative polyadenylation, depending on two specific nucleotide elements located in the region upstream of the cleavage site.
4. A novel single stranded RNA ourmia-like virus from *M. oryzae*, MOLV1, has been identified.
5. The majority of small RNAs sequenced from *M. oryzae* derive from repetitive elements like retrotransposons and the MOLV1 virus. The 3'-end of these small RNAs is usually adenylated.
6. Lack of Rbp35 results in the up-regulation of many small interfering-like RNAs derived from repetitive elements, and the presence of MOLV1 virus RNA is strongly depleted in the $\Delta rbp35$ mutant, suggesting a possible involvement of Rbp35 in RNA silencing.
7. Orthologs of 731 RNA-related proteins have been identified in 49 fungal species and 3 non-fungal species. This evolutionary survey revealed that mRNA translation and tRNA maturation are generally well conserved, while mRNA export and RNA silencing exhibit a variable degree of conservation. Core proteins of fundamental importance are missing in some branches of the fungal tree of life.

References

1. Yoshida K, Saunders DGO, Mitsuoka C, Natsume S, Kosugi S, Saitoh H, et al. Host specialization of the blast fungus *Magnaporthe oryzae* is associated with dynamic gain and loss of genes linked to transposable elements. *BMC Genomics*. 2016;17: 370. doi:10.1186/s12864-016-2690-6
2. Dean R, Van Kan JAL, Pretorius ZA, Hammond-Kosack KE, Di Pietro A, Spanu PD, et al. The Top 10 fungal pathogens in molecular plant pathology. *Molecular Plant Pathology*. 2012. pp. 414–430. doi:10.1111/j.1364-3703.2011.00783.x
3. Khush GS. What it will take to Feed 5.0 Billion Rice consumers in 2030. *Plant Molecular Biology*. 2005. pp. 1–6. doi:10.1007/s11103-005-2159-5
4. Islam T, Croll D, Gladieux P, Soanes D, Persoons, Antoine Bhattacharjee P, Hossain S, et al. Emergence of wheat blast in Bangladesh was caused by a South American lineage of *Magnaporthe oryzae*. *bioRxiv*. 2016; doi:10.1101/059832
5. Wilson R a, Talbot NJ. Under pressure: investigating the biology of plant infection by *Magnaporthe oryzae*. *Nat Rev Microbiol*. 2009;7: 185–195. doi:10.1038/nrmicro2032
6. Talbot NJ. On the trail of a cereal killer: Exploring the biology of *Magnaporthe grisea*. *Annu Rev Microbiol*. 2003;57: 177–202. doi:10.1146/annurev.micro.57.030502.090957
7. Chumley FG, Valent B. A Mechanism for Surface Attachment. 1987;239: 288–290. doi:10.1126/science.239.4837.288
8. de Jong J, McCormack B, Smirnov N, Talbot N. Glycerol generates turgor in rice blast. *Nature*. 1997;389: 244. doi:10.1038/38418
9. Saunders DGO, Aves SJ, Talbot NJ. Cell cycle-mediated regulation of plant infection by the rice blast fungus. *Plant Cell*. 2010;22: 497–507. doi:10.1105/tpc.109.072447
10. Veneault-Fourrey C, Barooah M, Egan M, Wakley G, Talbot NJ. Autophagic Fungal Cell Death Is Necessary for Infection by the Rice Blast Fungus. *Science* (80-). 2006;312: 580–583. doi:10.1126/science.1124550
11. Saunders DGO, Dagdas YF, Talbot NJ. Spatial uncoupling of mitosis and cytokinesis during appressorium-mediated plant infection by the rice blast fungus *Magnaporthe oryzae*. *Plant Cell*. 2010;22: 2417–2428. doi:10.1105/tpc.110.074492
12. Kankanala P, Czymmek K, Valent B. Roles for rice membrane dynamics and plasmodesmata during biotrophic invasion by the blast fungus. *The Plant cell*. 2007;19: 706–724. doi:10.1105/tpc.106.046300
13. Soanes DM, Richards T a, Talbot NJ. Insights from sequencing fungal and oomycete genomes: what can we learn about plant disease and the evolution of pathogenicity? *Plant Cell*. 2007;19: 3318–3326. doi:10.1105/tpc.107.056663
14. Zhang S, Xu JR. Effectors and Effector Delivery in *Magnaporthe oryzae*. *PLoS Pathog*. 2014;10. doi:10.1371/journal.ppat.1003826
15. Cruz CD, Kiyuna J, Bockus WW, Todd TC, Stack JP, Valent B. *Magnaporthe oryzae* conidia on basal wheat leaves as a potential source of wheat blast inoculum. 2015; 1491–1498. doi:10.1111/ppa.12414
16. Saleh D, Xu P, Shen Y, Li C, Adreit H, Milazzo J, et al. Sex at the origin: An Asian population of the rice blast fungus *Magnaporthe oryzae* reproduces sexually. *Mol Ecol*. 2012;21: 1330–1344. doi:10.1111/j.1365-294X.2012.05469.x
17. Saleh D, Adreit H, Fournier E, Tharreau D. South-East Asia is the center of origin , diversity and dispersion of the rice blast fungus , *Magnaporthe oryzae*. 2013;1.
18. Dean R a, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, et al. The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature*. 2005;434: 980–986. doi:10.1038/nature03449
19. Shirke MD, Mahesh HB, Gowda M. Genome-wide comparison of *Magnaporthe* species reveals a host-specific pattern of secretory proteins and transposable elements. *PLoS One*. 2016;11: 1–19. doi:10.1371/journal.pone.0162458
20. Soanes DM, Chakrabarti A, Paszkiewicz KH, Dawe AL, Talbot NJ. Genome-wide transcriptional profiling of appressorium development by the rice blast fungus *Magnaporthe oryzae*. *PLoS Pathog*. 2012;8. doi:10.1371/journal.ppat.1002514
21. Raman V, Simon SA, Romag A, Demirci F, Mathioni SM, Zhai J, et al. Physiological stressors and invasive plant infections alter the small RNA transcriptome of the rice blast fungus, *Magnaporthe oryzae*. *BMC Genomics*. 2013;14: 326. doi:10.1186/1471-2164-14-326
22. Nunes CC, Gowda M, Sailsbery J, Xue M, Chen F, Brown DE, et al. Diverse and tissue-enriched small RNAs in the plant pathogenic fungus, *Magnaporthe oryzae*. *BMC Genomics*. 2011;12: 288. doi:10.1186/1471-2164-12-288
23. Jeon J, Choi J, Lee G-W, Park S-Y, Huh A, Dean R a, et al. Genome-wide profiling of DNA methylation provides insights into epigenetic regulation of fungal development in a plant pathogenic fungus, *Magnaporthe oryzae*. *Sci Rep*. 2015;5: 8567. doi:10.1038/srep08567
24. Kawahara Y, Oono Y, Kanamori H, Matsumoto T, Itoh T, Minami E. Simultaneous RNA-seq analysis of a mixed transcriptome of rice and blast fungus interaction. *PLoS One*. 2012;7: e49423. doi:10.1371/journal.pone.0049423
25. Maejima K, Himeno M, Komatsu K, Kakizawa S, Yamaji Y, Hamamoto H, et al. Complete nucleotide sequence of a new double-stranded RNA virus from the rice blast fungus, *Magnaporthe oryzae*. *Arch*

- Virol. 2008;153: 389–391. doi:10.1007/s00705-007-1101-3
26. Illana A, Marconi M, Rodríguez-Romero J, Xu P, Dalmay T, Wilkinson MD, et al. Molecular characterization of a novel ssRNA ourmia-like virus from the rice blast fungus *Magnaporthe oryzae*. Arch Virol. 2016; doi:10.1007/s00705-016-3144-9
27. Yokoi T, Yamashita S, Hibi T. The nucleotide sequence and genome organization of *Magnaporthe oryzae* virus 1. Arch Virol. 2007;152: 2265–2269. doi:10.1007/s00705-007-1045-7
28. Brundrett MC. Coevolution of roots and mycorrhizas of land plants. New Phytologist. 2002. pp. 275–304. doi:10.1046/j.1469-8137.2002.00397.x
29. Taylor JW, Berbee ML. Dating divergences in the Fungal Tree of Life: review and new analyses. Mycologia. 2006;98: 838–849. doi:10.3852/mycologia.98.6.838
30. Butterfield NJ. Probable Proterozoic fungi. Paleobiology. 2005;31: 165–182. doi:10.1666/0094-8373(2005)031<0165:PPF>2.0.CO;2
31. James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, Cox CJ, et al. Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature. 2006;443: 818–822. doi:10.1038/nature05110
32. Liu YJ, Hodson MC, Hall BD. Loss of the flagellum happened only once in the fungal lineage: phylogenetic structure of kingdom Fungi inferred from RNA polymerase II subunit genes. BMC Evol Biol. 2006;6: 74. doi:10.1186/1471-2148-6-74
33. James TY, Pelin A, Bonen L, Ahrendt S, Sain D, Corradi N, et al. Shared signatures of parasitism and phylogenomics unite cryptomycota and microsporidia. Curr Biol. 2013;23: 1548–1553. doi:10.1016/j.cub.2013.06.057
34. Stajich JE, Berbee ML, Blackwell M, Hibbett DS, James TY, Spatafora JW, et al. Primer – The Fungi. Curr Biol. 2009;19: R840–R845. doi:10.1016/j.cub.2009.07.004.Primer
35. Redecker D, Kodner R, Graham LE. Glomalean fungi from the Ordovician. Science. 2000;289: 1920–1921. doi:10.1126/science.289.5486.1920
36. Spatafora JW, Chang Y, Benny GL, Lazarus K, Smith ME, Berbee ML, et al. A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. Mycologia. 2016;108: 1028–1046. doi:10.3852/16-042
37. Chang Y, Wang S, Sekimoto S, Aerts AL, Choi C, Clum A, et al. Phylogenomic analyses indicate that early fungi evolved digesting cell walls of algal ancestors of land plants. Genome Biol Evol. 2015;7: 1590–1601. doi:10.1093/gbe/evv090
38. White MM, James TY, O'Donnell K, Cafaro MJ, Tanabe Y, Sugiyama J. Phylogeny of the Zygomycota based on nuclear ribosomal sequence data. Mycologia. 2006;98: 872–84. doi:10.3852/mycologia.98.6.872
39. Ebersberger I, De Matos Simoes R, Kupczok A, Gube M, Kothe E, Voigt K, et al. A consistent phylogenetic backbone for the fungi. Mol Biol Evol. 2012;29: 1319–1334. doi:10.1093/molbev/msr285
40. Sekimoto S, Rochon A, Long JE, Dee JM, Berbee ML. A multigene phylogeny of *Olpidium* and its implications for early fungal evolution. BMC Evol Biol. 2011;11: 331. doi:10.1186/1471-2148-11-331
41. Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, et al. Genome evolution in yeasts. Nature. 2004;430: 35–44. doi:10.1038/nature02579
42. Vannini A, Cramer P. Conservation between the RNA polymerase I, II, and III transcription initiation machineries. Mol Cell. 2012;45: 439–446. doi:10.1016/j.molcel.2012.01.023
43. Sung Rhee H, Franklin Pugh B. Erratum: Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature. 2012;487: 128–128. doi:10.1038/nature11266
44. Hahn S. Structure and mechanism of the RNA polymerase II transcription machinery. Nat Struct Mol Biol. 2004;11: 394–403. doi:10.1038/nsmb763
45. Kupfer DM, Drabenstot SD, Buchanan KL, Lai H, Zhu H, Dyer DW, et al. Introns and splicing elements of five diverse fungi. Eukaryot Cell. 2004;3: 1088–1100. doi:10.1128/EC.3.5.1088-1100.2004
46. Sainsbury S, Bernecky C, Cramer P. Structural basis of transcription initiation by RNA polymerase II. Nat Rev Mol Cell Biol. 2015;16: 129–143. doi:10.1038/nrm3952
47. Tutucci E, Stutz F. Keeping mRNPs in check during assembly and nuclear export. Nat Rev Mol Cell Biol. 2011;12: 377–384. doi:10.1038/nrm3119
48. Hsin J-P, Manley JL. The RNA polymerase II CTD coordinates transcription and RNA processing. Genes Dev. 2012;26: 2119–37. doi:10.1101/gad.200303.112
49. Heidemann M, Hintermair C, Voß K, Eick D. Dynamic phosphorylation patterns of RNA polymerase II CTD during transcription. Biochimica et Biophysica Acta - Gene Regulatory Mechanisms. 2013. pp. 55–62. doi:10.1016/j.bbagr.2012.08.013
50. Conaway RC, Conaway JW. Orchestrating transcription with the pol II CTD. Nat Rev Mol Cell Biol. Nature Publishing Group; 2015;16: 128–128. doi:10.1038/nrm3956
51. Gonatopoulos-Pournatzis T, Dunn S, Bounds R, Cowling VH. RAM/Fam103a1 Is Required for mRNA Cap Methylation. Mol Cell. 2011;44: 585–596. doi:10.1016/j.molcel.2011.08.041
52. Wahl MC, Will CL, Lührmann R. The Spliceosome: Design Principles of a Dynamic RNP Machine. Cell. 2009. pp. 701–718. doi:10.1016/j.cell.2009.02.009
53. Proudfoot NJ. Ending the message: poly(A) signals then and now. Genes Dev. 2011;25: 1770–1782. doi:10.1101/gad.172684.11
54. Wickramasinghe VO, Laskey RA. Control of mammalian gene expression by selective mRNA export. Nat Rev Mol Cell Biol. 2015;16: 431–442. doi:10.1038/nrm4010
55. Spingola M, Grate L, Haussler D, Ares M. Genome-wide bioinformatic and molecular analysis of introns in

- Saccharomyces cerevisiae. RNA. 1999;5: 221–34. doi:10.1017/S1355838299981682
56. Ivashchenko AT, Tauasaraova MK, Atambaeva SA. [Exon-intron structure of genes of fungi genomes]. Mol Biol (Mosk). Russia (Federation); 2009;43: 28–35.
57. Crawford DJ, Hoskins AA, Friedman LJ, Gelles J, Moore MJ. Single-molecule colocalization FRET evidence that spliceosome activation precedes stable approach of 5' splice site and branch site. Proc Natl Acad Sci U S A. United States; 2013;110: 6783–6788. doi:10.1073/pnas.1219305110
58. Wang Q, Zhang L, Lynn B, Rymond BC. A BBP-Mud2p heterodimer mediates branchpoint recognition and influences splicing substrate abundance in budding yeast. Nucleic Acids Res. England; 2008;36: 2787–2798. doi:10.1093/nar/gkn144
59. Matera AG, Wang Z. A day in the life of the spliceosome. Nat Rev Mol Cell Biol. 2014;15: 108–21. doi:10.1038/nrm3742
60. Stevens SW, Barta I, Ge HY, Moore RE, Young MK, Lee TD, et al. Biochemical and genetic analyses of the U5, U6, and U4/U6 x U5 small nuclear ribonucleoproteins from Saccharomyces cerevisiae. RNA. 2001;7: 1543–53. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1370197&tool=pmcentrez&rendertype=abstract>
61. Hoskins AA, Friedman LJ, Gallagher SS, Crawford DJ, Anderson EG, Wombacher R, et al. Ordered and dynamic assembly of single spliceosomes. Science. United States; 2011;331: 1289–1295. doi:10.1126/science.1198830
62. Padgett RA, Konarska MM, Grabowski PJ, Hardy SF, Sharp PA. Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. Science. 1984;225: 898–903. doi:10.1126/science.6206566
63. Tsai RT, Fu RH, Yeh FL, Tseng CK, Lin YC, Huang YH, et al. Spliceosome disassembly catalyzed by Prp43 and its associated components Ntr1 and Ntr2. Genes Dev. 2005;19: 2991–3003. doi:10.1101/gad.1377405
64. Tsai R-T, Tseng C-K, Lee P-J, Chen H-C, Fu R-H, Chang K, et al. Dynamic interactions of Ntr1-Ntr2 with Prp43 and with U5 govern the recruitment of Prp43 to mediate spliceosome disassembly. Mol Cell Biol. 2007;27: 8027–8037. doi:10.1128/MCB.01213-07
65. Long JC, Caceres JF. The SR protein family of splicing factors: master regulators of gene expression. Biochem J. 2009;417: 15–27. doi:10.1042/BJ20081501
66. Zhou Z, Fu XD. Regulation of splicing by SR proteins and SR protein-specific kinases. Chromosoma. 2013. pp. 191–207. doi:10.1007/s00412-013-0407-z
67. Russell AG, Charette JM, Spencer DF, Gray MW. An early evolutionary origin for the minor spliceosome. Nature. 2006;443: 863–6. doi:10.1038/nature05228
68. Shi Y, Manley JL. The end of the message: Multiple protein-RNA interactions define the mRNA polyadenylation site. Genes Dev. 2015;29: 889–897. doi:10.1101/gad.261974.115
69. Proudfoot N. New perspectives on connecting messenger RNA 3' end formation to transcription. Current Opinion in Cell Biology. 2004. pp. 272–278. doi:10.1016/j.ceb.2004.03.007
70. Chan S, Choi E-A, Shi Y. Pre-mRNA 3'-end processing complex assembly and function. Wiley Interdiscip Rev RNA. 2011;2: 321–35. doi:10.1002/wrna.54
71. Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR, et al. Molecular Architecture of the Human Pre-mRNA 3' Processing Complex. Mol Cell. 2009;33: 365–376. doi:10.1016/j.molcel.2008.12.028
72. Moqtaderi Z, Geisberg J V, Jin Y, Fan X, Struhl K. Species-specific factors mediate extensive heterogeneity of mRNA 3' ends in yeasts. Proc Natl Acad Sci U S A. 2013;110: 11073–8. doi:10.1073/pnas.1309384110
73. Salisbury J, Hutchison KW, Graber JH. A multispecies comparison of the metazoan 3'-processing downstream elements and the CstF-64 RNA recognition motif. BMC Genomics. 2006;7: 55. doi:10.1186/1471-2164-7-55
74. Millevoi S, Vagner S. Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. Nucleic Acids Res. 2010;38: 2757–2774. doi:10.1093/nar/gkp1176
75. Elkon R, Ugalde AP, Agami R. Alternative cleavage and polyadenylation: extent, regulation and function. Nat Rev Genet. Nature Publishing Group; 2013;14: 496–506. doi:10.1038/nrg3482
76. Tian B, Manley JL. Alternative polyadenylation of mRNA precursors. Nat Rev Mol Cell Biol. 2016;18: 18–30. doi:10.1038/nrm.2016.116
77. Manley JL, Nishida K, Di Giammartino DC. Mechanisms and Consequences of Alternative Polyadenylation. Molecular Cell. 2011. pp. 853–866. doi:10.1016/j.molcel.2011.08.017
78. Lei EP, Krebber H, Silver PA. Messenger RNAs are recruited for nuclear export during transcription. Genes Dev. 2001;15: 1771–1782. doi:10.1101/gad.892401
79. Dermody JL, Dreyfuss JM, Villén J, Ogundipe B, Gygi SP, Park PJ, et al. Unphosphorylated SR-like protein Npl3 stimulates RNA polymerase II elongation. PLoS One. 2008;3. doi:10.1371/journal.pone.0003273
80. Huertas P, Aguilera A. Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. Mol Cell. 2003;12: 711–721. doi:10.1016/j.molcel.2003.08.010
81. Rondón AG, Jimeno S, García-Rubio M, Aguilera A. Molecular evidence that the eukaryotic THO/TREX complex is required for efficient transcription elongation. J Biol Chem. 2003;278: 39037–39043. doi:10.1074/jbc.M305718200

82. Rodríguez-Navarro S, Hurt E. Linking gene regulation to mRNA production and export. *Current Opinion in Cell Biology*. 2011. pp. 302–309. doi:10.1016/j.ceb.2010.12.002
83. Strässer K, Hurt E. Splicing factor Sub2p is required for nuclear mRNA export through its interaction with Yra1p. *Nature*. 2001;413: 648–652. doi:10.1038/35098113
84. Hackmann A, Wu H, Schneider U-M, Meyer K, Jung K, Krebber H. Quality control of spliced mRNAs requires the shuttling SR proteins Gbp2 and Hrb1. *Nat Commun*. 2014;5: 1–14. doi:10.1038/ncomms4123
85. Brune C, Munchel SE, Fischer N, Podtelejnikov A V, Weis K. Yeast poly(A)-binding protein Pab1 shuttles between the nucleus and the cytoplasm and functions in mRNA export. *RNA*. 2005;11: 517–31. doi:10.1261/rna.7291205
86. Senay C, Ferrari P, Rocher C, Rieger KJ, Winter J, Platel D, et al. The Mtr2-Mex67 NTF2-like Domain Complex. *J Biol Chem*. 2003;278: 48395. doi:10.1074/jbc.M308275200
87. Strambio-De-Castillia C, Niepel M, Rout MP. The nuclear pore complex: bridging nuclear transport and gene regulation. *Nat Rev Mol Cell Biol*. 2010;11: 490–501. doi:nrm2928 [pii]10.1038/nrm2928
88. Tran EJ, Zhou Y, Corbett AH, Wentz SR. The DEAD-Box Protein Dbp5 Controls mRNA Export by Triggering Specific RNA:Protein Remodeling Events. *Mol Cell*. 2007;28: 850–859. doi:10.1016/j.molcel.2007.09.019
89. Buxbaum AR, Haimovich G, Singer RH. In the right place at the right time: visualizing and understanding mRNA localization. *Nat Rev Mol Cell Biol*. 2014;16: 95–109. doi:10.1038/nrm3918
90. de la Cruz J, Karbstein K, Woolford JL. Functions of ribosomal proteins in assembly of eukaryotic ribosomes in vivo. *Annu Rev Biochem*. 2015;84: 93–129. doi:10.1146/annurev-biochem-060614-033917
91. Dever TE, Green R. The elongation, termination, and recycling phases of translation in eukaryotes. *Cold Spring Harb Perspect Biol*. 2012;4: 1–16. doi:10.1101/cshperspect.a013706
92. Aitken CE, Lorsch JR. A mechanistic overview of translation initiation in eukaryotes. *Nat Struct Mol Biol*. 2012;19: 568–76. doi:10.1038/nsmb.2303
93. Topisirovic I, Svitkin Y V, Sonenberg N, Shatkin AJ. Cap and cap-binding proteins in the control of gene expression. *Wiley Interdiscip Rev RNA*. 2011;2: 277–298. doi:10.1002/wrna.52
94. Berthelot K, Muldoon M, Rajkowitsch L, Hughes J, McCarthy JEG. Dynamics and processivity of 40S ribosome scanning on mRNA in yeast. *Mol Microbiol*. 2004;51: 987–1001. doi:10.1046/j.1365-2958.2003.03898.x
95. Saini AK, Nanda JS, Lorsch JR, Hinnebusch AG. Regulatory elements in eIF1A control the fidelity of start codon selection by modulating tRNAⁱ Met binding to the ribosome. *Genes Dev*. 2010;24: 97–110. doi:10.1101/gad.1871910
96. Voorhees RM, Ramakrishnan V. Structural basis of the translational elongation cycle. *TL - 82. Annu Rev Biochem*. 2013;82 VN-r: 203–236. doi:10.1146/annurev-biochem-113009-092313
97. Jackson RJ, Hellen CUT, Pestova T V. Termination and post-termination events in eukaryotic translation. *Advances in Protein Chemistry and Structural Biology*. 2012. pp. 45–93. doi:10.1016/B978-0-12-386497-0.00002-5
98. Schagger H, Cramer WA, Vonjagow G. Structural aspects of translation termination on the ribosome. *Anal Biochem*. 1994;217: 220–230. doi:10.1261/rna.2733411
99. Huang H-Y, Hopper A. Multiple Layers of Stress-Induced Regulation in tRNA Biology. *Life*. 2016;6: 16. doi:10.3390/life6020016
100. Ellis JC, Brown JW. The RNase P family [Internet]. *RNA biology*. 2009. pp. 362–369. doi:10.4161/rna.6.4.9241
101. Hopper AK. Transfer RNA post-transcriptional processing, turnover, and subcellular dynamics in the yeast *Saccharomyces cerevisiae*. *Genetics*. 2013;194: 43–67. doi:10.1534/genetics.112.147470
102. Aebi M, Kirchner G, Chen JY, Vijayraghavan U, Jacobson A, Martin NC, et al. Isolation of a temperature-sensitive mutant with an altered tRNA nucleotidyltransferase and cloning of the gene encoding tRNA nucleotidyltransferase in the yeast *Saccharomyces cerevisiae*. *J Biol Chem*. 1990;265: 16216–16220.
103. Trotta CR, Miao F, Arn EA, Stevens SW, Ho CK, Rauhut R, et al. The yeast tRNA splicing endonuclease: a tetrameric enzyme with two active site subunits homologous to the archaeal tRNA endonucleases. *Cell*. 1997;89: 849–58. doi:10.1016/S0092-8674(00)80270-6
104. Cantara WA, Crain PF, Rozenski J, McCloskey JA, Harris KA, Zhang X, et al. The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Res*. 2011;39. doi:10.1093/nar/gkq1028
105. Jackman JE, Alfonzo JD. Transfer RNA modifications: Nature's combinatorial chemistry playground. *Wiley Interdisciplinary Reviews: RNA*. 2013. pp. 35–48. doi:10.1002/wrna.1144
106. Chávez S, García-Martínez J, Delgado-Ramos L, Pérez-Ortón JE. The importance of controlling mRNA turnover during cell proliferation. *Current Genetics*. 2016. pp. 701–710. doi:10.1007/s00294-016-0594-2
107. Kilchert C, Wittmann S, Vasiljeva L. The regulation and functions of the nuclear RNA exosome complex. *Nat Rev Mol Cell Biol*. 2016;17: 227–239. doi:10.1038/nrm.2015.15
108. Kervestin S, Jacobson A. NMD: a multifaceted response to premature translational termination. *Nat Rev Mol Cell Biol*. 2012;13: 700–712. doi:10.1038/nrm3454
109. Frischmeyer P a, van Hoof A, O'Donnell K, Guerrero AL, Parker R, Dietz HC. An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science*. 2002;295: 2258–61. doi:10.1126/science.1067338
110. Doma MK, Parker R. Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature*. 2006;440: 561–4. doi:10.1038/nature04530
111. Schoenberg DR, Maquat LE. Regulation of cytoplasmic mRNA decay. *Nat Rev Genet*. 2012;13: 246–59.

- doi:10.1038/nrg3160
112. Bregman A, Avraham-Kelbert M, Barkai O, Duek L, Guterman A, Choder M. Promoter elements regulate cytoplasmic mRNA decay. *Cell*. 2011;147: 1473–1483. doi:10.1016/j.cell.2011.12.005
 113. Jiao X, Xiang S, Oh C, Martin CE, Tong L, Kiledjian M. Identification of a quality-control mechanism for mRNA 5'-end capping. *Nature*. 2010;467: 608–11. doi:10.1038/nature09338
 114. Jia H, Wang X, Anderson JT, Jankowsky E. RNA unwinding by the Trf4/Air2/Mtr4 polyadenylation (TRAMP) complex. *Proc Natl Acad Sci*. 2012;109: 7292–7297. doi:10.1073/pnas.1201085109
 115. Jensen T, Jacquier A, Libri D. Dealing with pervasive transcription. *Molecular Cell*. 2013. pp. 473–484. doi:10.1016/j.molcel.2013.10.032
 116. Goldstrohm AC, Wickens M. Multifunctional deadenylase complexes diversify mRNA control. *Nat Rev Mol Cell Biol*. 2008;9: 337–344. doi:10.1038/nrm2370
 117. Fromm S a, Truffault V, Kamenz J, Braun JE, Hoffmann N a, Izaurralde E, et al. The structural basis of Edc3- and Scd6-mediated activation of the Dcp1:Dcp2 mRNA decapping complex. *EMBO J*. 2012;31: 279–90. doi:10.1038/emboj.2011.408
 118. Nagarajan VK, Jones CI, Newbury SF, Green PJ. XRN 5' → 3' exoribonucleases: Structure, mechanisms and functions. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. 2013. pp. 590–603. doi:10.1016/j.bbagrm.2013.03.005
 119. Synowsky SA, Heck AJR. The yeast Ski complex is a hetero-tetramer. *Protein Sci*. 2008;17: 119–25. doi:10.1110/ps.073155908
 120. Krol J, Loedige I, Filipowicz W. The widespread regulation of microRNA biogenesis, function and decay. *Nat Rev Genet*. 2010;11: 597–610. doi:10.1038/nrg2843
 121. Martienssen R, Moazed D. RNAi and Heterochromatin Assembly. *Cold Spring Harb Perspect Biol*. 2015;7: a019323. doi:10.1101/cshperspect.a019323
 122. Li L, Chang SS, Liu Y. RNA interference pathways in filamentous fungi. *Cellular and Molecular Life Sciences*. 2010. pp. 3849–3863. doi:10.1007/s00018-010-0471-y
 123. Nakayashiki H. RNA silencing in fungi: Mechanisms and applications. *FEBS Letters*. 2005. pp. 5950–5957. doi:10.1016/j.febslet.2005.08.016
 124. Schumann U, Ayliffe M, Kazan K, Wang M-B. RNA silencing in fungi. *Front Biol (Beijing)*. 2010;5: 478–494. doi:10.1007/s11515-010-0550-3
 125. Romano N, Macino G. Quelling: transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences. *Mol Microbiol*. 1992;6: 3343–3353. doi:10.1111/j.1365-2958.1992.tb02202.x
 126. Freitag M, Lee DW, Kothe GO, Pratt RJ, Aramayo R, Selker EU. DNA Methylation Is Independent of RNA Interference in *Neurospora*. *Science (80-)*. 2004;304: 1939. doi:10.1126/science.1099709
 127. Shiu PKT, Metzenberg RL. Meiotic silencing by unpaired DNA: Properties, regulation and suppression. *Genetics*. 2002;161: 1483–1495. doi:10.1016/S0092-8674(01)00609-2
 128. Liu Y, Ye X, Jiang F, Liang C, Chen D, Peng J, et al. C3PO, an endoribonuclease that promotes RNAi by facilitating RISC activation. *Science (80-)*. 2009;325: 750–3. doi:10.1126/science.1176325
 129. Nakanishi K, Weinberg DE, Bartel DP, Patel DJ. Structure of yeast Argonaute with guide RNA. *Nature*. 2012;486: 368–374. doi:10.1038/nature11211
 130. Goto DB, Nakayama JI. RNA and epigenetic silencing: Insight from fission yeast. *Development Growth and Differentiation*. 2012. pp. 129–141. doi:10.1111/j.1440-169X.2011.01310.x
 131. Trieu TA, Calo S, Nicolás FE, Vila A, Moxon S, Dalmay T, et al. A Non-canonical RNA Silencing Pathway Promotes mRNA Degradation in Basal Fungi. *PLoS Genet*. 2015;11. doi:10.1371/journal.pgen.1005168
 132. Liu H, Cottrell TR, Pierini LM, Goldman WE, Doering TL. RNA interference in the pathogenic fungus *Cryptococcus neoformans*. *Genetics*. 2002;160: 463–470.
 133. Barba M, Czosnek H, Hadidi A. Historical Perspective, Development and Applications of Next-Generation Sequencing in Plant Virology. *Viruses*. 2014;6: 106–136. doi:10.3390/v6010106
 134. Hoheisel JD. Microarray technology: beyond transcript profiling and genotype analysis. *Nat Rev Genet*. 2006;7: 200–210. doi:10.1038/nrg1809
 135. Boguski MS, Tolstoshev CM, Bassett DE. Gene discovery in dbEST. [Internet]. *Science (New York, N.Y.)*. 1994. pp. 1993–4. doi:10.1126/science.8091218
 136. Harbers M, Carninci P. Tag-based approaches for transcriptome research and genome annotation. *Nat Methods*. 2005;2: 495–502. doi:10.1038/nmeth768
 137. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial Analysis of Gene Expression. *Science (80-)*. 1995;270: 484–487. doi:10.1126/science.270.5235.484
 138. Kodzius R, Kojima M, Nishiyori H, Nakamura M, Fukuda S, Tagami M, et al. CAGE: cap analysis of gene expression. *Nat Methods*. 2006;3: 211–222. doi:10.1038/nmeth0306-211
 139. Shiraki T, Kondo S, Katayama S, Waki K, Kasukawa T, Kawaji H, et al. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc Natl Acad Sci U S A*. 2003;100: 15776–81. doi:10.1073/pnas.2136655100
 140. Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, et al. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol*. 2000;18: 630–634. doi:10.1038/76469
 141. Rothberg JM, Leamon JH. The development and impact of 454 sequencing. *Nat Biotechnol*. 2008;26: 1117–1124. doi:10.1038/nbt1485
 142. Mantione F KJ, Kream F RM, Kuzelova HF, Ptacek RF, Raboch JE, Samuel E JM, et al. Comparing

- Bioinformatic Gene Expression Profiling Methods: Microarray and RNA-Seq. *Med Sci Monit Basic Res.* 2014;20: 138–141. doi:10.12659/MSMBR.892101
143. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods.* 2009;6: 377–382. doi:10.1038/nmeth.1315
 144. Merriman B, Torrent I, Rothberg JM. Progress in Ion Torrent semiconductor chip based sequencing. *Electrophoresis.* 2012. pp. 3397–3417. doi:10.1002/elps.201200424
 145. Carneiro MO, Russ C, Ross MG, Gabriel SB, Nusbaum C, Depristo MA. Pacific biosciences sequencing technology for genotyping and variation discovery in human data. *BMC Genomics.* 2012;13: 1. doi:10.1186/1471-2164-13-375
 146. Laver T, Harrison J, O'Neill PA, Moore K, Farbos A, Paszkiewicz K, et al. Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomol Detect Quantif.* 2015;3: 1–8. doi:10.1016/j.bdq.2015.02.001
 147. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. *Nucleic Acids Res.* 2000;28: 235–242. doi:10.1093/nar/28.1.235
 148. Bairoch A, Boeckmann B. The {SWISS-PROT} protein sequence data bank. *Nucleic Acids Res.* 1992;20: 2019–2022. doi:Swiss-Prot
 149. Collins FS, Morgan M, Patrinos A. The Human Genome Project: lessons from large-scale biology. *Science (80-).* 2003;300: 286–90. doi:10.1126/science.1084564
 150. Andrews S. FastQC: A quality control tool for high throughput sequence data. In: <http://www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc/>. 2010 p. <http://www.bioinformatics.babraham.ac.uk/projects/>. doi:citeulike-article-id:11583827
 151. Lassmann T, Hayashizaki Y, Daub CO. SAMStat: Monitoring biases in next generation sequencing data. *Bioinformatics.* 2011;27: 130–131. doi:10.1093/bioinformatics/btq614
 152. HannonLab. FASTX toolkit. In: Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 2014.
 153. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal.* 2011;17: 10–12. doi:10.14806/ej.17.1.200
 154. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30: 2114–2120. doi:10.1093/bioinformatics/btu170
 155. Droop AP. Fqtools: An efficient software suite for modern FASTQ file manipulation. *Bioinformatics.* 2016;32: 1883–1884. doi:10.1093/bioinformatics/btw088
 156. Trapnell C, Pachter L, Salzberg SL. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics.* 2009;25: 1105–1111. doi:10.1093/bioinformatics/btp120
 157. Langmead B. Aligning short sequencing reads with Bowtie. *Curr Protoc Bioinforma.* 2010; doi:10.1002/0471250953.bi1107s32
 158. Wu TD, Reeder J, Lawrence M, Becker G, Brauer MJ. GMAP and GSNAP for genomic sequence alignment: Enhancements to speed, accuracy, and functionality. *Methods in Molecular Biology.* 2016. pp. 283–334. doi:10.1007/978-1-4939-3578-9_15
 159. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29: 15–21. doi:10.1093/bioinformatics/bts635
 160. Trapnell C. Cufflinks. cuffdiff Documentation, v6► Open Module on GenePattern Public Server. In: GenePattern [Internet]. 2013. Available: <http://www.broadinstitute.org/cancer/software/genepattern/modules/docs/Cufflinks.cuffdiff/6>
 161. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2009;26: 139–140. doi:10.1093/bioinformatics/btp616
 162. Love MI, Anders S, Huber W. Differential analysis of count data - the DESeq2 package. *Genome Biol.* 2014;15: 550. doi:10.1186/s13059-014-0550-8
 163. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011;12: 323. doi:10.1186/1471-2105-12-323
 164. Henschel R, Nista PM, Lieber M, Haas BJ, Wu L-S, LeDuc RD. Trinity RNA-Seq assembler performance optimization. *Proc 1st Conf Extrem Sci Eng Discov Environ Bridg from Extrem to campus beyond - XSEDE '12.* 2012; 8. doi:10.1145/2335755.2335842
 165. Zerbino DR. Using the Velvet de novo assembler for short-read sequencing technologies. *Current Protocols in Bioinformatics.* 2010. doi:10.1002/0471250953.bi1105s31
 166. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I. ABySS: A parallel assembler for short read sequence data. *Genome Res.* 2009;19: 1117–1123. doi:10.1101/gr.089532.108
 167. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet.* 2016;17: 333–351. doi:10.1038/nrg.2016.49
 168. Foster AJ, Jenkinson JM, Talbot NJ. Trehalose synthesis and metabolism are required at different stages of plant infection by *Magnaporthe oryzae*. *EMBO J.* 2003;22: 225–235. doi:10.1093/emboj/cdg018
 169. Sweigard J, Chumley F, Carroll A, Farrall L, Valent B. A series of vectors for fungal transformation. *Fungal Genet Newsl.* 2015;44: 52–53. Available: <http://www.fgsc.net/fgn44/sweig.html>
 170. Tucker SL, Besi MI, Galhano R, Franceschetti M, Goetz S, Lenhart S, et al. Common genetic pathways regulate organ-specific infection-related development in the rice blast fungus. *Plant Cell.* 2010;22: 953–972. doi:10.1105/tpc.109.066340
 171. Rodríguez-Romero J, Franceschetti M, Bueno E, Sesma A, Rodríguez-Romero J. Multilayer Regulatory

- Mechanisms Control Cleavage Factor I Proteins in Filamentous Fungi. *Nucleic Acid Res.* 2014;43: 179–95. doi:10.1093/nar/gku1297
172. Franceschetti M, Bueno E, Wilson R a, Tucker SL, Gómez-Mena C, Calder G, et al. Fungal virulence and development is regulated by alternative pre-mRNA 3'end processing in *Magnaporthe oryzae*. *PLoS Pathog.* 2011;7: e1002441. doi:10.1371/journal.ppat.1002441
 173. Wilkening S, Pelechano V, Järvelin AI, Tekkedil MM, Anders S, Benes V, et al. An efficient method for genome-wide polyadenylation site mapping and RNA quantification. *Nucleic Acids Res.* 2013;41: e65. doi:10.1093/nar/gks1249
 174. Reyes a., Anders S, Huber W. Analyzing RNA-seq data for differential exon usage with the DEXSeq package. *Gastroenterology.* 2011;138: 958–968. doi:10.1053/j.gastro.2009.11.005
 175. Priebe S, Kreisel C, Horn F, Guthke R, Linde J. FungiFun2: A comprehensive online resource for systematic analysis of gene lists from fungal species. *Bioinformatics.* 2014;31: 445–446. doi:10.1093/bioinformatics/btu627
 176. Ruepp A, Zollner A, Maier D, Albermann K, Hani J, Mokrejs M, et al. The FunCat, a functional annotation scheme for systematic classification of proteins from whole genomes. *Nucleic Acids Res.* 2004;32: 5539–5545. doi:10.1093/nar/gkh894
 177. The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. *Nat Genet.* 2000;25: 25–29. doi:10.1038/75556
 178. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research.* 1999. pp. 29–34. doi:10.1093/nar/27.1.29
 179. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite [Internet]. *Nucleic acids research.* 2015. pp. W39–49. doi:10.1093/nar/gkv416
 180. Bailey TL. DREME: Motif discovery in transcription factor ChIP-seq data. *Bioinformatics.* 2011;27: 1653–1659. doi:10.1093/bioinformatics/btr261
 181. Leung H, Borromeo ES, Bernardo MA, Notteghem JL. Genetic analysis of virulence in the rice blast fungus *Magnaporthe grisea*. *Phytopathology.* 1988;78: 1227–1233 LA–English.
 182. Chang HK, Park SY, Lee YH, Kang S. A dual selection based, targeted gene replacement tool for *Magnaporthe grisea* and *Fusarium oxysporum*. *Fungal Genet Biol.* 2005;42: 483–492. doi:10.1016/j.fgb.2005.03.004
 183. Sorefan K, Pais H, Hall AE, Kozomara A, Griffiths-Jones S, Moulton V, et al. Reducing ligation bias of small RNAs in libraries for next generation sequencing. *Silence.* 2012;3: 4. doi:10.1186/1758-907X-3-4
 184. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 2013;30: 2725–2729. doi:10.1093/molbev/mst197
 185. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32: 1792–1797. doi:10.1093/nar/gkh340
 186. Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol.* 2001;18: 691–699. doi:10.1093/oxfordjournals.molbev.a003851
 187. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. GenBank. *Nucleic Acids Res.* 2013;41. doi:10.1093/nar/gks1195
 188. Nordberg H, Cantor M, Dusheyko S, Hua S, Poliakov A, Shabalov I, et al. The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic Acids Res.* 2014;42. doi:10.1093/nar/gkt1069
 189. Kaminuma E, Kosuge T, Kodama Y, Aono H, Mashima J, Gojobori T, et al. DDBJ progress report. *Nucleic Acids Res.* 2011;39. doi:10.1093/nar/gkq1041
 190. Flicek P, Amode M, Barrell D. Ensembl 2012. *Nucleic acids* 2012;40: D84–90. doi:10.1093/nar/gkr991
 191. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* 2008;9: R7. doi:10.1186/gb-2008-9-1-r7
 192. Finn RD, Clements J, Eddy SR. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* 2011;39. doi:10.1093/nar/gkr367
 193. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 2015;44: D279–D285. doi:10.1093/nar/gkv1344
 194. Thomas-Chollier M, Defrance M, Medina-Rivera A, Sand O, Herrmann C, Thieffry D, et al. RSAT 2011: regulatory sequence analysis tools. *Nucleic Acids Res.* 2011;39: W86–W91. doi:10.1093/nar/gkr377
 195. R Development Core Team. R Language Definition V. 3.1.1 [Internet]. R Development Core Team. 2015. Available: <http://cran.us.r-project.org/doc/manuals/r-devel/R-lang.html>
 196. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: A sequence logo generator. *Genome Res.* 2004;14: 1188–1190. doi:10.1101/gr.849004
 197. Huerta-Cepas J, Serra F, Bork P. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol Biol Evol.* 2016;33: 1635–1638. doi:10.1093/molbev/msw046
 198. Sievers F, Higgins DG. Clustal Omega. *Curr Protoc Bioinforma.* 2014;2014: 3.13.1–3.13.16. doi:10.1002/0471250953.bi0313s48
 199. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 2009;25: 1972–1973. doi:10.1093/bioinformatics/btp348

200. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30: 1312–1313. doi:10.1093/bioinformatics/btu033
201. Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E. EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res*. 2009;19: 327–335. doi:10.1101/gr.073585.107
202. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10: 421. doi:10.1186/1471-2105-10-421
203. Di Tommaso P, Moretti S, Xenarios I, Orobitz M, Montanyola A, Chang JM, et al. T-Coffee: A web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res*. 2011;39. doi:10.1093/nar/gkr245
204. Huerta-Cepas J, Dopazo H, Dopazo J, Gabaldón T. The human phylome. *Genome Biol*. 2007;8: R109. doi:10.1186/gb-2007-8-6-r109
205. Yang Q, Coseno M, Gilmartin GM, Doublé S. Crystal structure of a human cleavage factor CFIm25/CFIm68/RNA complex provides an insight into poly(A) site recognition and RNA looping. *Structure*. 2011;19: 368–377. doi:10.1016/j.str.2010.12.021
206. Martin G, Gruber AR, Keller W, Zavolan M. Genome-wide Analysis of Pre-mRNA 3' End Processing Reveals a Decisive Role of Human Cleavage Factor I in the Regulation of 3' UTR Length. *Cell Rep*. 2012;1: 753–763. doi:10.1016/j.celrep.2012.05.003
207. Yao C, Biesinger J, Wan J, Weng L, Xing Y, Xie X, et al. Transcriptome-wide analyses of CstF64–RNA interactions in global regulation of mRNA alternative polyadenylation. *Proc Natl Acad Sci*. 2012;109: 18773–18778. doi:10.1073/pnas.1211101109
208. Lackford B, Yao C, Charles GM, Weng L, Zheng X, Choi EA, et al. Fip1 regulates mRNA alternative polyadenylation to promote stem cell self-renewal. *EMBO J*. 2014;33: 878–889. doi:10.1002/embj.201386537
209. Kim Guisbert KS, Li H, Guthrie C. Alternative 3' pre-mRNA processing in *Saccharomyces cerevisiae* is modulated by Nab4/Hrp1 in vivo. *PLoS Biol*. 2007;5: 0015–0022. doi:10.1371/journal.pbio.0050006
210. González CI, Ruiz-Echevarría MJ, Vasudevan S, Henry MF, Peltz SW. The Yeast hnRNP-like Protein Hrp1/Nab4 Marks a Transcript for Nonsense-Mediated mRNA Decay. *Mol Cell*. 2000;5: 489–499. doi:10.1016/S1097-2765(00)80443-8
211. Kessler MM, Henry MF, Shen E, Zhao J, Gross S, Silver PA, et al. Hrp1, a sequence-specific RNA-binding protein that shuttles between the nucleus and the cytoplasm, is required for mRNA 3'-end formation in yeast. *Genes Dev*. 1997;11: 2545–2556. doi:10.1101/gad.11.19.2545
212. Wang G, Wang C, Hou R, Zhou X, Li G, Zhang S, et al. The AMT1 arginine methyltransferase gene is important for plant infection and normal hyphal growth in *fusarium graminearum*. *PLoS One*. 2012;7. doi:10.1371/journal.pone.0038324
213. Wilkening S, Pelechano V, Järvelin AI, Tekkedil MM, Anders S, Benes V, et al. An efficient method for genome-wide polyadenylation site mapping and RNA quantification. *Nucleic Acids Res*. 2013; 1–8. doi:10.1093/nar/gks1249
214. Mandel CR, Bai Y, Tong L. Protein factors in pre-mRNA 3'-end processing. *Cellular and Molecular Life Sciences*. 2008. pp. 1099–1122. doi:10.1007/s00018-007-7474-3
215. Mata J. Genome-wide mapping of polyadenylation sites in fission yeast reveals widespread alternative polyadenylation. *RNA Biol*. 2013;10: 1407–1414. doi:10.4161/rna.25758
216. Ozsolak F, Kapranov P, Foissac S, Kim SW, Fishilevich E, Monaghan AP, et al. Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. *Cell*. 2010;143: 1018–1029. doi:10.1016/j.cell.2010.11.020
217. Gross S, Moore C. Five subunits are required for reconstitution of the cleavage and polyadenylation activities of *Saccharomyces cerevisiae* cleavage factor I. *Proc Natl Acad Sci U S A*. 2001;98: 6080–5. doi:10.1073/pnas.101046598
218. Paulson AR, Tong L. Crystal structure of the Rna14-Rna15 complex. *RNA*. 2012;18: 1154–62. doi:10.1261/rna.032524.112
219. Pérez-Cañadillas JM. Grabbing the message: structural basis of mRNA 3'UTR recognition by Hrp1. *EMBO J*. 2006;25: 3167–78. doi:10.1038/sj.emboj.7601190
220. Klasens BIF, Das AT, Berkhout B. Inhibition of polyadenylation by stable RNA secondary structure. *Nucleic Acids Res*. 1998;26: 1870–1876. doi:10.1093/nar/26.8.1870
221. Graveley BR, Fleming ES, Gilmartin GM. RNA structure is a critical determinant of poly(A) site recognition by cleavage and polyadenylation specificity factor. *Mol Cell Biol*. 1996;16: 4942–51. Available: <http://www.ncbi.nlm.nih.gov/pubmed/8756653> <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC231496>
222. Brown PH, Tiley LS, Cullen BR. Effect of RNA secondary structure on polyadenylation site selection. *Genes Dev*. 1991;5: 1277–1284. doi:10.1101/gad.5.7.1277
223. Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, Assmann SM. In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature*. 2014;505: 696–700. doi:10.1038/nature12756
224. Wan Y, Kertesz M, Spitale RC, Segal E, Chang HY. Understanding the transcriptome through RNA structure. *TL - 12. Nat Rev Genet*. 2011;12 VN-r: 641–655. doi:10.1038/nrg3049
225. Hofacker IL. RNA secondary structure analysis using the Vienna RNA package. *Curr Protoc Bioinformatics*. 2009;Chapter 12: Unit12.2. doi:10.1002/0471250953.bi1202s26

226. Araujo PR, Yoon K, Ko D, Smith AD, Qiao M, Suresh U, et al. Before it gets started: Regulating translation at the 5' UTR. *Comparative and Functional Genomics*. 2012. doi:10.1155/2012/475731
227. Kubo T, Wada T, Yamaguchi Y, Shimizu A, Handa H. Knock-down of 25 kDa subunit of cleavage factor Im in Hela cells alters alternative polyadenylation within 3'-UTRs. *Nucleic Acids Res*. 2006;34: 6264–6271. doi:10.1093/nar/gkl794
228. Yang L, Duff MO, Graveley BR, Carmichael GG, Chen L-L. Genomewide characterization of non-polyadenylated RNAs. *Genome Biol*. 2011;12: R16. doi:10.1186/gb-2011-12-2-r16
229. Farman ML, Tosa Y, Nitta N, Leong SA. MAGGY, a retrotransposon in the genome of the rice blast fungus *Magnaporthe grisea*. *Mol Gen Genet*. 1996;251: 665–674. Available: <http://www.ncbi.nlm.nih.gov/pubmed/8757397>
230. Sanchez E, Asano K, Sone T. Characterization of Inago1 and Inago2 retrotransposons in *Magnaporthe oryzae*. *J Gen Plant Pathol*. 2011;77: 239–242. doi:10.1007/s10327-011-0314-1
231. Kang S. Organization and distribution pattern of MGLR-3, a novel retrotransposon in the rice blast fungus *Magnaporthe grisea*. *Fungal Genet Biol*. 2001;32: 11–19. doi:10.1006/fgbi.2000.1246
232. Fudal I, Böhnert HU, Tharreau D, Lebrun MH. Transposition of MINE, a composite retrotransposon, in the avirulence gene ACE1 of the rice blast fungus *Magnaporthe grisea*. *Fungal Genet Biol*. 2005;42: 761–772. doi:10.1016/j.fgb.2005.05.001
233. Nishimura M, Hayashi N, Jwa NS, Lau GW, Hamer JE, Hasebe a. Insertion of the LINE retrotransposon MGL causes a conidiophore pattern mutation in *Magnaporthe grisea*. *Mol Plant Microbe Interact*. 2000;13: 892–4. doi:10.1094/MPMI.2000.13.8.892
234. Kang S, Lebrun MH, Farrall L, Valent B. Gain of virulence caused by insertion of a Pot3 transposon in a *Magnaporthe grisea* avirulence gene. *Mol Plant Microbe Interact*. 2001;14: 671–4. doi:10.1094/MPMI.2001.14.5.671
235. Nakayashiki H, Matsuo H, Chuma I, Ikeda K, Betsuyaku S, Kusaba M, et al. Pyret, a Ty3/Gypsy retrotransposon in *Magnaporthe grisea* contains an extra domain between the nucleocapsid and protease domains. *Nucleic Acids Res*. 2001;29: 4106–4113. doi:10.1094/MPMI.2002.15.1.6
236. Farman ML. *Pyricularia grisea* Isolates Causing Gray Leaf Spot on Perennial Ryegrass (*Lolium perenne*) in the United States: Relationship to *P. grisea* Isolates from Other Host Plants. *Phytopathology*. 2002;92: 245–54. doi:10.1094/PHYTO.2002.92.3.245
237. Farman ML, Eto Y, Nakao T, Tosa Y, Nakayashiki H, Mayama S, et al. Analysis of the structure of the AVR1-CO39 avirulence locus in virulent rice-infecting isolates of *Magnaporthe grisea*. *Mol Plant Microbe Interact*. 2002;15: 6–16. doi:10.1094/MPMI.2002.15.1.6
238. Reinisch KM, Wolin SL. Emerging themes in non-coding RNA quality control. *Current Opinion in Structural Biology*. 2007. pp. 209–214. doi:10.1016/j.sbi.2007.03.012
239. Sesma A, Osbourn AE. The rice leaf blast pathogen undergoes developmental processes typical of root-infecting fungi. *Nature*. 2004;431: 582–586. doi:10.1038/nature02880
240. Kim S, Yamamoto J, Chen Y, Aida M, Wada T, Handa H, et al. Evidence that cleavage factor Im is a heterotetrameric protein complex controlling alternative polyadenylation. *Genes to Cells*. 2010;15: 1003–1013. doi:10.1111/j.1365-2443.2010.01436.x
241. Yang Q, Gilmartin GM, Doublé S. Structural basis of UGUA recognition by the Nudix protein CFI(m)25 and implications for a regulatory role in mRNA 3' processing. *Proc Natl Acad Sci U S A*. 2010;107: 10062–7. doi:10.1073/pnas.1000848107
242. Carthew RW, Sontheimer EJ. Origins and Mechanisms of miRNAs and siRNAs. *Cell*. 2009. pp. 642–655. doi:10.1016/j.cell.2009.01.035
243. Buchon N, Vauray C. RNAi: a defensive RNA-silencing against viruses and transposable elements. *Heredity (Edinb)*. 2006;96: 195–202. doi:10.1038/sj.hdy.6800789
244. Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. *Nat Rev Genet*. 2009;10: 94–108. doi:10.1038/nrg2504
245. Crowley TE, Nellen W, Gomer RH, Firtel RA. Phenocopy of discoidin I-minus mutants by antisense transformation in *Dictyostelium*. *Cell*. 1985;43: 633–641. doi:10.1016/0092-8674(85)90235-1
246. Ecker JR, Davis RW. Inhibition of gene expression in plant cells by expression of antisense RNA. *Proc Natl Acad Sci U S A*. 1986;83: 5372–6. doi:10.1073/pnas.83.15.5372
247. Kim SK, Wold BJ. Stable reduction of thymidine kinase activity in cells expressing high levels of antisense RNA. *Cell*. 1985;42: 129–138. doi:10.1016/0092-8674(85)80108-2 [pii]
248. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*. 1998;391: 806–811. doi:10.1038/35888
249. Hamilton AJ, Baulcombe DC, Cogoni C, Macino G, Kennerdell JR, Carthew RW, et al. A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science*. 1999;286: 950–2. doi:10.1126/science.286.5441.950
250. Zamore PD, Tuschl T, Sharp P a, Bartel DP. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell*. 2000;101: 25–33. doi:10.1016/S0092-8674(00)80620-0
251. Bernstein E, Caudy AA, Hammond SM, Hannon GJ. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*. 2001;409: 363–366. doi:10.1038/35053110
252. Elbashir SM, Elbashir SM, Lendeckel W, Lendeckel W, Tuschl T, Tuschl T. RNA interference is mediated 1- and 22-nucleotide RNAs. *Genes Dev*. 2001;15: 188–200. doi:10.1101/gad.862301.vents
253. Nicolás FE, Ruiz-Vázquez RM. Functional diversity of RNAi-associated sRNAs in fungi. *International*

- Journal of Molecular Sciences. 2013. pp. 15348–15360. doi:10.3390/ijms140815348
254. Dang Y, Yang Q, Xue Z, Liu Y. RNA interference in fungi: Pathways, functions, and applications. *Eukaryotic Cell*. 2011. pp. 1148–1155. doi:10.1128/EC.05109-11
 255. Valencia-Sanchez MA, Liu J, Hannon GJ, Parker R. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes and Development*. 2006. pp. 515–524. doi:10.1101/gad.1399806
 256. Lee HC, Li L, Gu W, Xue Z, Crosthwaite SK, Pertsemliadis A, et al. Diverse Pathways Generate MicroRNA-like RNAs and Dicer-Independent Small Interfering RNAs in Fungi. *Mol Cell*. Elsevier Ltd; 2010;38: 803–814. doi:10.1016/j.molcel.2010.04.005
 257. Bai Y, Lan F, Yang W, Zhang F, Yang K, Li Z, et al. SRNA profiling in *Aspergillus flavus* reveals differentially expressed miRNA-like RNAs response to water activity and temperature. *Fungal Genet Biol*. 2015;81: 113–119. doi:10.1016/j.fgb.2015.03.004
 258. Lin YL, Ma LT, Lee YR, Lin SS, Wang SY, Chang TT, et al. MicroRNA-like small RNAs prediction in the development of antrodia cinnamomea. *PLoS One*. 2015;10. doi:10.1371/journal.pone.0123245
 259. Lau SKP, Chow WN, Wong AYP, Yeung JMY, Bao J, Zhang N, et al. Identification of MicroRNA-Like RNAs in Mycelial and Yeast Phases of the Thermal Dimorphic Fungus *Penicillium marneffeii*. *PLoS Negl Trop Dis*. 2013;7. doi:10.1371/journal.pntd.0002398
 260. Zhou Q, Wang Z, Zhang J, Meng H, Huang B. Genome-wide identification and profiling of microRNA-like RNAs from *Metarhizium anisopliae* during development. *Fungal Biol*. 2012;116: 1156–1162. doi:10.1016/j.funbio.2012.09.001
 261. Huang Q, Evans JD. Identification of microRNA-like small RNAs from fungal parasite *Nosema ceranae*. *J Invertebr Pathol*. 2016;133: 107–109. doi:10.1016/j.jip.2015.12.005
 262. Liu T, Hu J, Zuo Y, Jin Y, Hou J. Identification of microRNA-like RNAs from *Curvularia lunata* associated with maize leaf spot by bioinformation analysis and deep sequencing. *Mol Genet Genomics*. 2016;291: 587–596. doi:10.1007/s00438-015-1128-1
 263. Dahlmann TA, Kück U. Dicer-dependent biogenesis of small RNAs and evidence for microRNA-like RNAs in the penicillin producing fungus *Penicillium chrysogenum*. *PLoS One*. 2015;10. doi:10.1371/journal.pone.0125989
 264. Kang K, Zhong J, Jiang L, Liu G, Gou CY, Wu Q, et al. Identification of microRNA-Like RNAs in the Filamentous Fungus *Trichoderma reesei* by Solexa Sequencing. *PLoS One*. 2013;8. doi:10.1371/journal.pone.0076288
 265. Zhou J, Fu Y, Xie J, Li B, Jiang D, Li G, et al. Identification of microRNA-like RNAs in a plant pathogenic fungus <i>Sclerotinia sclerotiorum</i> by high-throughput sequencing. *Mol Genet Genomics*. 2012;287: 275–282. doi:10.1007/s00438-012-0678-8
 266. Chen R, Jiang N, Jiang Q, Sun X, Wang Y, Zhang H, et al. Exploring microRNA-like small RNAs in the filamentous fungus *Fusarium oxysporum*. *PLoS One*. 2014;9. doi:10.1371/journal.pone.0104956
 267. Wilson RC, Doudna JA. Molecular Mechanisms of RNA Interference. *Annu Rev Biophys*. 2013;42: 217–239. doi:10.1146/annurev-biophys-083012-130404
 268. Shabalina SA, Koonin E V. Origins and evolution of eukaryotic RNA interference. *Trends in Ecology and Evolution*. 2008. pp. 578–587. doi:10.1016/j.tree.2008.06.005
 269. Song J-J, Smith SK, Hannon GJ, Joshua-Tor L. Crystal structure of Argonaute and its implications for RISC slicer activity. *Science*. 2004;305: 1434–7. doi:10.1126/science.1102514
 270. Obbard DJ, Gordon KHJ, Buck AH, Jiggins FM. The evolution of RNAi as a defence against viruses and transposable elements. *Philos Trans R Soc Lond B Biol Sci*. 2009;364: 99–115. doi:10.1098/rstb.2008.0168
 271. Volpe TA, Kidner C, Hall IM, Teng G, Grewal SI, Martienssen RA. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* (80-). 2002;297: 1833–1837. doi:10.1126/science.1074973
 272. Martienssen RA, Zaratiegui M, Goto DB. RNA interference and heterochromatin in the fission yeast *Schizosaccharomyces pombe*. *Trends in Genetics*. 2005. pp. 450–456. doi:10.1016/j.tig.2005.06.005
 273. Smialowska A, Djupedal I, Wang J, Kylsten P, Swoboda P, Ekwall K. RNAi mediates post-transcriptional repression of gene expression in fission yeast *Schizosaccharomyces pombe*. *Biochem Biophys Res Commun*. 2014;444: 254–259. doi:10.1016/j.bbrc.2014.01.057
 274. Volpe T, Martienssen RA. RNA interference and heterochromatin assembly. *Cold Spring Harb Perspect Biol*. 2011;3: 1–11. doi:10.1101/cshperspect.a003731
 275. Grewal SIS. RNAi-dependent formation of heterochromatin and its diverse functions. *Current Opinion in Genetics and Development*. 2010. pp. 134–141. doi:10.1016/j.gde.2010.02.003
 276. Holloch D, Moazed D. RNAi in fission yeast finds new targets and new ways of targeting at the nuclear periphery. *Genes Dev*. 2012;26: 741–745. doi:10.1101/gad.191155.112
 277. Galagan JE, Selker EU. RIP: The evolutionary cost of genome defense. *Trends in Genetics*. 2004. pp. 417–423. doi:10.1016/j.tig.2004.07.007
 278. Ibrahim AS, Spellberg B, Walsh TJ, Kontoyiannis DP. Pathogenesis of mucormycosis. *Clin Infect Dis*. 2012;54. doi:10.1093/cid/cir865
 279. Navarro E, Ruiz-Pérez VL, Torres-Martínez S. Overexpression of the *crgA* gene abolishes light requirement for carotenoid biosynthesis in *Mucor circinelloides*. *Eur J Biochem*. 2000;267: 800–807. doi:10.1046/j.1432-1327.2000.01058.x
 280. Cervantes M, Vila A, Nicolás FE, Moxon S, de Haro JP, Dalmay T, et al. A Single Argonaute Gene Participates in Exogenous and Endogenous RNAi and Controls Cellular Functions in the Basal Fungus

- Mucor circinelloides. PLoS One. 2013;8. doi:10.1371/journal.pone.0069283
281. Nicolás FE, Vila A, Moxon S, Cascales MD, Torres-Martínez S, Ruiz-Vázquez RM, et al. The RNAi machinery controls distinct responses to environmental signals in the basal fungus *Mucor circinelloides*. BMC Genomics. 2015;16: 237. doi:10.1186/s12864-015-1443-2
 282. Torres-Martínez S, Ruiz-Vázquez RM. RNAi pathways in *Mucor*: A tale of proteins, small RNAs and functional diversity. Fungal Genet Biol. 2015;90: 44–52. doi:10.1016/j.fgb.2015.11.006
 283. Kadotani N, Nakayashiki H, Tosa Y, Mayama S. RNA silencing in the phytopathogenic fungus *Magnaporthe oryzae*. Mol Plant Microbe Interact. 2003;16: 769–76. doi:10.1094/MPMI.2003.16.9.769
 284. Kadotani N, Nakayashiki H, Tosa Y, Mayama S. One of the two Dicer-like proteins in the filamentous fungi *Magnaporthe oryzae* genome is responsible for hairpin RNA-triggered RNA silencing and related small interfering RNA accumulation. J Biol Chem. 2004;279: 44467–44474. doi:10.1074/jbc.M408259200
 285. Kadotani N, Murata T, Nguyen BQ, Adachi Y, Nakayashiki H. Transcriptional control and protein specialization have roles in the functional diversification of two dicer-like proteins in *Magnaporthe oryzae*. Genetics. 2008;180: 1245–1249. doi:10.1534/genetics.108.093922
 286. Bohnsack MT, Czaplinski K, Gorlich D. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. RNA. 2004;10: 185–91. doi:10.1261/rna.5167604
 287. Murthi A, Shaheen HH, Huang H-Y, Preston M a, Lai T-P, Phizicky EM, et al. Regulation of tRNA Bidirectional Nuclear-Cytoplasmic Trafficking in *Saccharomyces cerevisiae*. Mol Biol Cell. 2010;21: 639–649. doi:10.1091/mbc.E09-07-0551
 288. Huang HY, Hopper AK. In vivo biochemical analyses reveal distinct roles of β -importins and eEF1A in tRNA subcellular traffic. Genes Dev. 2015;29: 772–783. doi:10.1101/gad.258293.115
 289. Shibata S, Sasaki M, Miki T, Shimamoto A, Furuichi Y, Katahira J, et al. Exportin-5 orthologues are functionally divergent among species. Nucleic Acids Res. 2006;34: 4711–4721. doi:10.1093/nar/gkl663
 290. Hall TMT. Structure and function of argonaute proteins. Structure. 2005. pp. 1403–1408. doi:10.1016/j.str.2005.08.005
 291. Murata T, Kadotani N, Yamaguchi M, Tosa Y, Mayama S, Nakayashiki H. siRNA-dependent and -independent post-transcriptional cosuppression of the LTR-retrotransposon MAGGY in the phytopathogenic fungus *Magnaporthe oryzae*. Nucleic Acids Res. 2007;35: 5987–5994. doi:10.1093/nar/gkm646
 292. Hammond TM, Andrews MD, Roossinck MJ, Keller NP. Aspergillus mycoviruses are targets and suppressors of RNA silencing. Eukaryot Cell. 2008;7: 350–357. doi:10.1128/EC.00356-07
 293. Zhang X, Segers GC, Sun Q, Deng F, Nuss DL. Characterization of hypovirus-derived small RNAs generated in the chestnut blight fungus by an inducible DCL-2-dependent pathway. J Virol. 2008;82: 2613–2619. doi:10.1128/JVI.02324-07
 294. Campo S, Gilbert KB, Carrington JC. Small RNA-Based Antiviral Defense in the Phytopathogenic Fungus *Colletotrichum higginsianum*. PLoS Pathog. 2016;12: e1005640. doi:10.1371/journal.ppat.1005640
 295. Himeno M, Maejima K, Komatsu K, Ozeki J, Hashimoto M, Kagiwada S, et al. Significantly low level of small RNA accumulation derived from an encapsidated mycovirus with dsRNA genome. Virology. 2010;396: 69–75. doi:10.1016/j.virol.2009.10.008
 296. Pelechano V, Steinmetz LM. Gene regulation by antisense transcription. Nat Rev Genet. 2013;14: 880–93. doi:10.1038/nrg3594
 297. MacRae IJ, Zhou K, Doudna JA. Structural determinants of RNA recognition and cleavage by Dicer. Nat Struct Mol Biol. 2007;14: 934–940. doi:10.1038/nsmb1293
 298. Mi S, Cai T, Hu Y, Chen Y, Hodges E, Ni F, et al. Sorting of Small RNAs into Arabidopsis Argonaute Complexes Is Directed by the 5' Terminal Nucleotide. Cell. 2008;133: 116–127. doi:10.1016/j.cell.2008.02.034
 299. Proudfoot NJ. Ending the message : poly (A) signals then and now. Genes Dev. 2011;25: 1770–1782. doi:10.1101/gad.17268411.Freely
 300. LaCava J, Houseley J, Saveanu C, Petfalski E, Thompson E, Jacquier A, et al. RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. Cell. 2005;121: 713–724. doi:10.1016/j.cell.2005.04.029
 301. Kuai L, Fang F, Butler JS, Sherman F. Polyadenylation of rRNA in *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A. 2004;101: 8581–6. doi:10.1073/pnas.0402888101
 302. Düvel K, Pries R, Braus GH. Polyadenylation of rRNA- and tRNA-based yeast transcripts cleaved by internal ribozyme activity. Curr Genet. 2003;43: 255–262. doi:10.1007/s00294-003-0401-8
 303. Fleischmann J, Liu H, Wu C-P. Polyadenylation of ribosomal RNA by *Candida albicans* also involves the small subunit. BMC Mol Biol. 2004;5: 17. doi:10.1186/1471-2199-5-17
 304. Morlando M, Greco P, Dichtl B, Fatica A, Keller W, Bozzoni I. Functional analysis of yeast snoRNA and snRNA 3'-end formation mediated by uncoupling of cleavage and polyadenylation. Mol Cell Biol. 2002;22: 1379–89. doi:10.1128/MCB.22.5.1379-1389.2002
 305. Perumal K, Gu J, Reddy R. Evolutionary conservation of post-transcriptional 3' end adenylation of small RNAs: *S. cerevisiae* signal recognition particle RNA and U2 small nuclear RNA are post-transcriptionally adenylated. Mol Cell Biochem. 2000;208: 99–109. Available: <http://www.ncbi.nlm.nih.gov/pubmed/10939633>
 306. Lu S, Sun YH, Chiang VL. Adenylation of plant miRNAs. Nucleic Acids Res. 2009;37: 1878–1885. doi:10.1093/nar/gkp031

307. Xue Z, Yuan H, Guo J, Liu Y. Reconstitution of an Argonaute-Dependent Small RNA Biogenesis Pathway Reveals a Handover Mechanism Involving the RNA Exosome and the Exonuclease QIP. *Mol Cell*. 2012;46: 299–310. doi:10.1016/j.molcel.2012.03.019
308. Holoch D, Moazed D. Small-RNA loading licenses Argonaute for assembly into a transcriptional silencing complex. *Nat Struct Mol Biol*. 2015;22: 328–35. doi:10.1038/nsmb.2979
309. Chak L-L, Mohammed J, Lai EC, Tucker-Kellogg G, Okamura K. A deeply conserved, noncanonical miRNA hosted by ribosomal DNA. *RNA*. 2015;21: 375–84. doi:10.1261/rna.049098.114
310. Kachroo P, Leong SA, Chattoo BB. Mg-SINE: a short interspersed nuclear element from the rice blast fungus, *Magnaporthe grisea*. *Proc Natl Acad Sci USA*. 1995;92: 11125–11129. doi:10.1073/pnas.92.24.11125
311. Chak LL, Okamura K. Argonaute-dependent small RNAs derived from single-stranded, non-structured precursors. *Frontiers in Genetics*. 2014. doi:10.3389/fgene.2014.00172
312. Wu L, Zhang Q, Zhou H, Ni F, Wu X, Qi Y. Rice MicroRNA effector complexes and targets. *Plant Cell*. 2009;21: 3421–35. doi:10.1105/tpc.109.070938
313. White RJ. Transcription by RNA polymerase III: more complex than we thought. *Nat Rev Genet*. 2011;12: 459–463. doi:10.1038/nrg3001
314. Yoshida K, Blobel G. The karyopherin Kap142p/Msn5p mediates nuclear import and nuclear export of different cargo proteins. *J Cell Biol*. 2001;152: 729–739. doi:10.1083/jcb.152.4.729
315. Burroughs AM, Ando Y, De Hoon MJL, Tomaru Y, Nishibu T, Ukekawa R, et al. A comprehensive survey of 3' animal miRNA modification events and a possible role for 3' adenylation in modulating miRNA targeting effectiveness. *Genome Res*. 2010;20: 1398–1410. doi:10.1101/gr.106054.110
316. Ghabrial SA, Suzuki N. Viruses of Plant Pathogenic Fungi. *Annu Rev Phytopathol*. Annual Reviews; 2009;47: 353–384. doi:10.1146/annurev-phyto-080508-081932
317. Ghabrial SA, Castón JR, Jiang D, Nibert ML, Suzuki N. 50-plus years of fungal viruses. *Virology*. Elsevier {BV}; 2015;479–480: 356–368. doi:10.1016/j.virol.2015.02.034
318. Marquez LM, Redman RS, Rodriguez RJ, Roossinck MJ. A Virus in a Fungus in a Plant: Three-Way Symbiosis Required for Thermal Tolerance. *Science* (80-). American Association for the Advancement of Science ({AAAS}); 2007;315: 513–515. doi:10.1126/science.1136237
319. Xie J, Jiang D. New Insights into Mycoviruses and Exploration for the Biological Control of Crop Fungal Diseases. *Annu Rev Phytopathol*. Annual Reviews; 2014;52: 45–68. doi:10.1146/annurev-phyto-102313-050222
320. Nuss DL. Hypovirulence: Mycoviruses at the fungal{textendash}plant interface. *Nat Rev Microbiol*. Springer Nature; 2005;3: 632–642. doi:10.1038/nrmicro1206
321. Liu L, Xie J, Cheng J, Fu Y, Li G, Yi X, et al. Fungal negative-stranded {RNA} virus that is related to bornaviruses and nyaviruses. *Proc Natl Acad Sci*. Proceedings of the National Academy of Sciences; 2014;111: 12205–12210. doi:10.1073/pnas.1401786111
322. Marzano S-YL, Nelson BD, Ajayi-Oyetunde O, Bradley CA, Hughes TJ, Hartman GL, et al. Identification of Diverse Mycoviruses through Metatranscriptomics Characterization of the Viromes of Five Major Fungal Plant Pathogens. Simon A, editor. *J Virol*. American Society for Microbiology; 2016;90: 6846–6863. doi:10.1128/jvi.00357-16
323. ICTV. Virus Taxonomy: 2012 Release. In: International Committee on Taxonomy of Viruses [Internet]. 2012. Available: <http://ictvonline.org/virustaxonomy.asp?version=2012>
324. Hillman BI, Cai G. The Family Narnaviridae. Simplest of RNA Viruses. *Adv Virus Res*. 2013;86: 149–176. doi:10.1016/B978-0-12-394315-6.00006-4
325. Rastgou M, Habibi MK, Izaadpanah K, Masenga V, Milne RG, Wolf YI, et al. Molecular characterization of the plant virus genus Ourmiavirus and evidence of inter-kingdom reassortment of viral genome segments as its possible route of origin. *J Gen Virol*. Microbiology Society; 2009;90: 2525–2535. doi:10.1099/vir.0.013086-0
326. Donaire L, Rozas J, Ayllón MA. Molecular characterization of Botrytis ourmia-like virus, a mycovirus close to the plant pathogenic genus Ourmiavirus. *Virology*. Elsevier {BV}; 2016;489: 158–164. doi:10.1016/j.virol.2015.11.027
327. Skamnioti P, Gurr SJ. Against the grain: safeguarding rice from rice blast disease. *Trends Biotechnol*. Elsevier {BV}; 2009;27: 141–150. doi:10.1016/j.tibtech.2008.12.002
328. Hunst PL. Variation in Double-Stranded {RNA} from Isolates {ofPyricularia} oryzae. *Phytopathology*. Scientific Societies; 1986;76: 674. doi:10.1094/phyto-76-674
329. YAMASHITA S, DOI Y, YORA K. A Polyhedral Virus Found in Rice Blast Fungus, *Pyricularia oryzae* Cavara. *Japanese J Phytopathol*. The Phytopathological Society of Japan; 1971;37: 356–359. doi:10.3186/jjphytopath.37.356
330. Tang L, Hu Y, Liu L, Wu S, Xie J, Cheng J, et al. Genomic organization of a novel victorivirus from the rice blast fungus *Magnaporthe oryzae*. *Arch Virol*. Springer Nature; 2015;160: 2907–2910. doi:10.1007/s00705-015-2562-4
331. Urayama S, Kato S, Suzuki Y, Aoki N, Le MT, Arie T, et al. Mycoviruses related to chrysovirus affect vegetative growth in the rice blast fungus *Magnaporthe oryzae*. *J Gen Virol*. Microbiology Society; 2010;91: 3085–3094. doi:10.1099/vir.0.025411-0
332. Urayama S, Sakoda H, Takai R, Katoh Y, Le TM, Fukuhara T, et al. A {dsRNA} mycovirus, *Magnaporthe oryzae* chrysovirus 1-B, suppresses vegetative growth and development of the rice blast fungus. *Virology*. Elsevier {BV}; 2014;448: 265–273. doi:10.1016/j.virol.2013.10.022

333. Ai Y-P, Zhong J, Chen C-Y, Zhu H-J, Gao B-D. A novel single-stranded {RNA} virus isolated from the rice-pathogenic fungus *Magnaporthe oryzae* with similarity to members of the family Tombusviridae. *Arch Virol.* Springer Nature; 2015;161: 725–729. doi:10.1007/s00705-015-2683-9
334. Chao C-CT, Ellingboe AH. Selection for mating competence {in *Magnaporthe*} griseopathogenic to rice. *Can J Bot.* Canadian Science Publishing; 1991;69: 2130–2134. doi:10.1139/b91-267
335. Huang X. {CAP}3: A {DNA} Sequence Assembly Program. *Genome Res.* Cold Spring Harbor Laboratory Press; 1999;9: 868–877. doi:10.1101/gr.9.9.868
336. Rice P, Longden I, Bleasby A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* 2000;16: 276–277.
337. Pruitt KD. {NCBI} Reference Sequence ({RefSeq}): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* Oxford University Press ({OUP}); 2004;33: D501--D504. doi:10.1093/nar/gki025
338. Janicke A, Vancuylenberg J, Boag PR, Traven A, Beilharz TH. {ePAT}: A simple method to tag adenylated {RNA} to measure poly(A)-tail length and other 3' applications. *RNA.* Cold Spring Harbor Laboratory Press; 2012;18: 1289–1295. doi:10.1261/rna.031898.111
339. Xie J, Ghabrial SA. Molecular characterizations of two mitoviruses co-infecting a hyovirulent isolate of the plant pathogenic fungus *Sclerotinia sclerotiorum*. *Virology.* Elsevier {BV}; 2012;428: 77–85. doi:10.1016/j.virol.2012.03.015
340. Li W, Zhang Y, Zhang C, Pei X, Wang Z, Jia S. Presence of poly(A) and poly(A)-rich tails in a positive-strand {RNA} virus known to lack 3' poly(A) tails. *Virology.* Elsevier {BV}; 2014;454–455: 1–10. doi:10.1016/j.virol.2014.02.002
341. Koonin E V. The phylogeny of {RNA}-dependent {RNA} polymerases of positive-strand {RNA} viruses. *J Gen Virol.* Microbiology Society; 1991;72: 2197–2206. doi:10.1099/0022-1317-72-9-2197
342. Hong Y, Dover SL, Cole TE, Brasier CM, Buck KW. Multiple Mitochondrial Viruses in an Isolate of the Dutch Elm Disease {*FungusOphiostoma*} novo-ulmi. *Virology.* Elsevier {BV}; 1999;258: 118–127. doi:10.1006/viro.1999.9691
343. Cai G, Krychiw JF, Myers K, Fry WE, Hillman BI. A new virus from the plant pathogenic oomycete *Phytophthora infestans* with an 8 kb {dsRNA} genome: The sixth member of a proposed new virus genus. *Virology.* Elsevier {BV}; 2013;435: 341–349. doi:10.1016/j.virol.2012.10.012
344. Ghabrial SA. Origin, adaptation and evolutionary pathways of fungal viruses. *Virus Genes.* 1998. pp. 119–131. doi:10.1023/A:1007966229595
345. Marzano S-YL, Domier LL. Novel mycoviruses discovered from metatranscriptomics survey of soybean phyllosphere phytobiomes. *Virus Res.* Elsevier {BV}; 2016;213: 332–342. doi:10.1016/j.virusres.2015.11.002
346. Hawksworth DL. The magnitude of fungal diversity: the 1.5 million species estimate revisited. *Mycol Res.* 2001;105: 1422–1432. doi:10.1017/S0953756201004725
347. Tedersoo L, Bahram M, Polme S, Koljalg U, Yorou NS, Wijesundera R, et al. Global diversity and geography of soil fungi. *Science* (80-). 2014;346: 6213. doi:10.1126/science.1256688
348. Blackwell M. The fungi: 1, 2, 3 ... 5.1 million species? *Am J Bot.* 2011;98: 426–438. doi:10.3732/ajb.1000298
349. Aguileta G, Refrégier G, Yockteng R, Fournier E, Giraud T. Rapidly evolving genes in pathogens: Methods for detecting positive selection and examples among fungi, bacteria, viruses and protists. *Infect Genet Evol.* 2009;9: 656–670. doi:10.1016/j.meegid.2009.03.010
350. Grigoriev I V., Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, et al. MycoCosm portal: Gearing up for 1000 fungal genomes. *Nucleic Acids Res.* 2014;42. doi:10.1093/nar/gkt1183
351. Wapinski I, Pfeffer A, Friedman N, Regev A. Natural history and evolutionary principles of gene duplication in fungi. *Nature.* 2007;449: 54–61. doi:10.1038/nature06107
352. Fitzpatrick D a, Logue ME, Stajich JE, Butler G. A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. *BMC Evol Biol.* 2006;6: 99. doi:10.1186/1471-2148-6-99
353. Hieronymus H, Silver PA. A systems view of mRNP biology. *Genes and Development.* 2004. pp. 2845–2860. doi:10.1101/gad.1256904
354. Mitchell SF, Parker R. Principles and Properties of Eukaryotic mRNPs. *Molecular Cell.* 2014. pp. 547–588. doi:10.1016/j.molcel.2014.04.033
355. Elbarbary RA, Maquat LE. Coupling pre-mRNA splicing and 3' end formation to mRNA export: Alternative ways to punch the nuclear export clock. *Genes Dev.* 2016;30: 487–488. doi:10.1101/gad.278937.116
356. Hinnebusch AG, Ivanov IP, Sonenberg N. Translational control by 5' -untranslated regions of eukaryotic mRNAs. *Science* (80-). 2016;352: 1413–1416. doi:10.1126/science.aad9868
357. Proudfoot NJ, Platt T, Epshtein V, Cardinale CJ, Ruckenstein AE, Borukhov S, et al. Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science.* 2016;352: aad9926. doi:10.1126/science.aad9926
358. Chu D, Kazana E, Bellanger N, Singh T, Tuite MF, Von Der Haar T. Translation elongation can control translation initiation on eukaryotic mRNAs. *EMBO J.* 2014;33: 21–34. doi:10.1002/embj.201385651
359. Beznosková P, Cuchalová L, Wagner S, Shoemaker CJ, Gunišová S, von der Haar T, et al. Translation Initiation Factors eIF3 and HCR1 Control Translation Termination and Stop Codon Read-Through in Yeast Cells. *PLoS Genet.* 2013;9. doi:10.1371/journal.pgen.1003962
360. Novoa EM, Pavon-Eternod M, Pan T, Ribas De Pouplana L. A role for tRNA modifications in genome

- structure and codon usage. *Cell*. 2012;149: 202–213. doi:10.1016/j.cell.2012.01.050
361. Falaleeva M, Stamm S. Processing of snoRNAs as a new source of regulatory non-coding RNAs: SnoRNA fragments form a new class of functional RNAs. *BioEssays*. 2013;35: 46–54. doi:10.1002/bies.201200117
 362. Berbee ML. The phylogeny of plant and animal pathogens in the Ascomycota. *Physiol Mol Plant Pathol*. 2001;59: 165–187. doi:10.1006/pmpp.2001.0355
 363. Schüßler A, Schwarzott D, Walker C. A new fungal phylum, the Glomeromycota: phylogeny and evolution. *Mycol Res*. 2001;105: 1413–1421. doi:10.1017/S0953756201005196
 364. Field KJ, Rimington WR, Bidartondo MI, Allinson KE, Beerling DJ, Cameron DD, et al. First evidence of mutualism between ancient plant lineages (Haplomitriopsida liverworts) and Mucoromycotina fungi and its response to simulated Palaeozoic changes in atmospheric CO₂. *New Phytol*. 2015;205: 743–756. doi:10.1111/nph.13024
 365. Tretter ED, Johnson EM, Benny GL, Lichtwardt RW, Wang Y, Kandel P, et al. An eight-gene molecular phylogeny of the Kickxellomycotina, including the first phylogenetic placement of Asellariales. *Mycologia*. 2014;106: 912–35. doi:10.3852/13-253
 366. Humber R a. <|>Entomophthoromycota</|>: a new phylum and reclassification for entomophthoroid fungi. *Mycotaxon*. 2012;120: 477–492. doi:10.5248/120.477
 367. Longcore JE. Chytridiomycota. *J Wildl Dis*. 2006; 1–8. doi:10.1038/npg.els.0004251
 368. Corradi N. Microsporidia: Eukaryotic Intracellular Parasites Shaped by Gene Loss and Horizontal Gene Transfers. *Annu Rev Microbiol*. 2015;69: 150720190645000. doi:10.1146/annurev-micro-091014-104136
 369. Cuomo CA, Desjardins CA, Bakowski MA, Goldberg J, Ma AT, Becnel JJ, et al. Microsporidian genome analysis reveals evolutionary strategies for obligate intracellular growth. *Genome Res*. 2012;22: 2478–2488. doi:10.1101/gr.142802.112
 370. Rodríguez-Navarro S, Fischer T, Luo MJ, Antúnez O, Brettschneider S, Lechner J, et al. Sus1, a Functional Component of the SAGA Histone Acetylase Complex and the Nuclear Pore-Associated mRNA Export Machinery. *Cell*. 2004;116: 75–86. doi:10.1016/S0092-8674(03)01025-0
 371. Bolger TA, Folkmann AW, Tran EJ, Went SR. The mRNA Export Factor Gle1 and Inositol Hexakisphosphate Regulate Distinct Stages of Translation. *Cell*. 2008;134: 624–633. doi:10.1016/j.cell.2008.06.027
 372. Koepke J, Kaffarnik F, Haag C, Zarnack K, Luscombe NM, König J, et al. The RNA-binding protein Rrm4 is essential for efficient secretion of endochitinase Cts1. *Mol Cell Proteomics*. 2011;10: M111.011213. doi:10.1074/mcp.M111.011213
 373. Houseley J, Tollervey D. The Many Pathways of RNA Degradation. *Cell*. 2009. pp. 763–776. doi:10.1016/j.cell.2009.01.019
 374. Schilders G, Raijmakers R, Raats JMH, Pruijn GJM. MPP6 is an exosome-associated RNA-binding protein involved in 5.8S rRNA maturation. *Nucleic Acids Res*. 2005;33: 6795–6804. doi:10.1093/nar/gki982
 375. Atkinson GC, Baldauf SL, Hauryliuk V. Evolution of nonstop, no-go and nonsense-mediated mRNA decay and their termination factor-derived components. *BMC Evol Biol*. 2008;8: 290. doi:10.1186/1471-2148-8-290
 376. Brownawell AM, Macara IG. Exportin-5, a novel karyopherin, mediates nuclear export of double-stranded RNA binding proteins. *J Cell Biol*. 2002;156: 53–64. doi:10.1083/jcb.200110082
 377. Keeling P. Five questions about microsporidia. *PLoS Pathog*. 2009;5. doi:10.1371/journal.ppat.1000489
 378. Hampsey M. Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiol Mol Biol Rev*. 1998;62: 465–503. doi:10.1158/0008-5472.CAN-07-2721
 379. Bushell M, Stoneley M, Kong YW, Hamilton TL, Spriggs KA, Dobbyn HC, et al. Polypyrimidine Tract Binding Protein Regulates IRES-Mediated Gene Expression during Apoptosis. *Mol Cell*. 2006;23: 401–412. doi:http://doi.org/10.1016/j.molcel.2006.06.012
 380. Gross S, Moore CL. Rna15 interaction with the A-rich yeast polyadenylation signal is an essential step in mRNA 3'-end formation. *Mol Cell Biol*. 2001;21: 8045–55. doi:10.1128/MCB.21.23.8045-8055.2001
 381. Tian B, Graber JH. Signals for pre-mRNA cleavage and polyadenylation. *Wiley Interdisciplinary Reviews: RNA*. 2012. pp. 385–396. doi:10.1002/wrna.116
 382. Brown PH, Tiley LS, Cullen BR. Effect of RNA secondary structure on polyadenylation site selection. *Genes Dev*. 1991;5: 1277–1284. doi:10.1101/gad.5.7.1277
 383. Iwata H, Gotoh O. Comparative analysis of information contents relevant to recognition of introns in many species. *BMC Genomics*. 2011;12: 45. doi:10.1186/1471-2164-12-45
 384. Berglund JA, Chua K, Abovich N, Reed R, Rosbash M. The Splicing Factor BBP Interacts Specifically with the Pre-mRNA Branchpoint Sequence UACUAAC. *Cell*. 1997;89: 781–787. doi:10.1016/S0092-8674(00)80261-5
 385. Reed R, Maniatis T. The role of the mammalian branchpoint sequence in pre-mRNA splicing. *Genes Dev*. 1988;2: 1268–1276. doi:10.1101/gad.2.10.1268
 386. Nielsen S, Yuzenkova Y, Zenkin N. Mechanism of eukaryotic RNA polymerase III transcription termination. *Science*. 2013;340: 1577–80. doi:10.1126/science.1237934
 387. Arimbasseri AG, Rijal K, Maraia RJ, Bhargava P, Reese JC. Transcription termination by the eukaryotic RNA polymerase III. *Biochim Biophys Acta - Gene Regul Mech*. 2013;1829: 318–30. doi:10.1016/j.bbagr.2012.10.006