Virulence- and signaling-associated genes display a preference for long 3'UTRs during rice infection and metabolic stress in the rice blast fungus

Julio Rodríguez-Romero[1,2], Marco Marconi[1,2], V. Ortega-Campayo[1,2], Marie Demuez[1,2], Mark D. Wilkinson[1,2] and Ane Sesma[1,2,*]

[1]Centre for Plant Biotechnology and Genomics (CBGP UPM-INIA), Universidad Politécnica de Madrid & Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, Campus de Montegancedo, 28223 Pozuelo de Alarcón, Madrid (Spain); [2]Dept. Biotecnología y Biología Vegetal, UPM.

*Correspondence and requests for materials should be addressed to A.S. (ane.sesma@upm.es).

Abstract

- Generation of mRNA isoforms by alternative polyadenylation (APA) and their involvement in regulation of fungal cellular processes, including virulence, remains elusive. Here, we investigate genome-wide polyadenylation site (PAS) selection in the rice blast fungus to understand how APA regulates pathogenicity.
- More than half of *Magnaporthe oryzae* transcripts undergo APA and show novel motifs in their PAS region. Transcripts with shorter 3'UTRs are more stable and abundant in polysomal fractions, suggesting they are being translated more efficiently. Importantly, rice colonization increases the use of distal PASs of pathogenicity genes, especially those participating in signaling pathways like *14-3-3B*, whose long 3'UTR is required for infection.
- Cleavage factor-I (CFI) Rbp35 regulates expression and distal PAS selection of virulence and signalling-associated genes, tRNAs and transposable elements, pointing its potential to drive genomic rearrangements and pathogen evolution. We propose a non-canonical PAS selection mechanism for Rbp35 that recognizes UGUAH, unlike humans, without CFI25.
- ts involved in fungal growth and environmental adaptation. Furthermore, these data provides useful information for enhancing genome annotations and for cross-species comparisons of PASs and PAS usage within the fungal kingdom and the tree of life.

INTRODUCTION

Eukaryotic mRNAs are stabilized by the addition of a polyadenosine tail at their 3' end (Proudfoot, 2016). Typically, a preferred or canonical PAS is present at the 3' untranslated region (UTR) of all pre-mRNAs. However, multiple potential polyadenylation sites (PASs) are often found within a pre-mRNA. The selection among these preferred PASs is regulated by multiple mechanisms during development and in response to environmental cues (Tian & Manley, 2017). The use of different cleavage sites within a pre-mRNA can result in the generation of mRNA isoforms with different exon content or 3' UTR length. *Cis* elements present in the 3'UTRs such as microRNA target sites modulate gene expression by affecting cytoplasmic polyadenylation, subcellular localization, stability, translation and/or decay of the mRNA (Berkovits & Mayr, 2015; Shi & Manley, 2015; Mayr, 2017; Tian & Manley, 2017). Therefore, in addition to alternative splicing, alternative polyadenylation (APA) is an important process for creating different mRNA isoforms of a given pre-mRNA precursor and for regulating gene expression. While selection of a proper 3´end cleavage site is observed in yeast (Shalem *et al.*, 2013), it is less well characterized in filamentous fungi (Sesma, 2016). Polyadenylation of pre-mRNAs is carried out in two consecutive steps: a cleavage reaction to free a 3' hydroxyl, followed by addition of the poly(A) tail, which makes the RNA molecule more stable by preventing its degradation, and promotes mRNA transport and translation (Eckmann *et al.*, 2011; Weill *et al.*, 2012). Poly(A) tail length can also regulate the fate of an mRNA, and varies between 50-80 residues in yeast to 200-300 residues in higher eukaryotes.

Most pre-mRNA 3' end processing factors and polyadenylation signals of mammals and yeast have been identified and reviewed extensively (Shi *et al.*, 2009; Yang & Doublie, 2011; Shi & Manley, 2015). A basic metazoan 3' end processing machinery contains ~20 proteins and includes the poly(A) polymerase (PAP), poly(A)-binding proteins (PABPs), the large subunits of RNA polymerase II (RNAP II), and four multi-subunit complexes: cleavage and polyadenylation specificity factor (CPSF), cleavage and stimulation factor (CstF), cleavage factor I (CFI_m) and cleavage factor II ($CFII_m$). PAP and CPSF are sufficient for both cleavage and polyadenylation reactions. The other three complexes (CstF, CFI_m and $CFII_m$) participate in the selection of a proper site for cleavage. Poly(A) tail length is regulated by PABPs, which in turn activate PAP. Many yeast 3' processing factors have counterparts in mammals, including Hrp1 for CFI_m, CFIA for CstF and CFII for CPSF (Chan *et al.*, 2011).

Very little is known about other fungal polyadenylation machineries. We previously identified and characterised the Rbp35 protein in the rice blast fungus, a novel CFI protein exclusive to filamentous fungi (Franceschetti *et al.*, 2011; Rodríguez-Romero *et al.*, 2015). Rbp35 co-

immunoprecipitates *in vivo* with CFI25, the orthologue of mammalian CFI_m25. The CFI25/Rbp35 complex is absent in fission and budding yeasts. Rbp35 is the functional equivalent of metazoan CFI_m68, although their protein domain structure is not conserved (Franceschetti *et al.*, 2011). The crystal structure of a human complex CFI_m25/CFI_m68/RNA identified the residues within the CFIm68 RNA recognition motif (RRM) domain critical for the interaction with the UGUA motif and mentioned Rbp35 RRM as the likely ortholog of CFI_m68 RRM (Yang *et al.*, 2011), corroborating our results. Several studies have demonstrated the essential role of CFI_m in non-canonical PAS selection and APA. The CFI_m complex enhances the efficiency of cleavage by selecting a canonical or non-canonical PAS and helps recruit other components of the polyadenylation machinery (Venkataraman *et al.*, 2005). The RRM of CFI_m68 facilitates RNA looping and enhances the RNA binding by CFI_m25 (Yang *et al.*, 2011).

Rice blast represents a major threat to rice crops, and affects severely global rice production (Fisher *et al.*, 2012). The disease is caused by the filamentous ascomycete fungus *Magnaporthe oryzae,* which can colonise different plant organs at all stages of development (Illana *et al.*, 2013). In addition to rice, blast disease has severely affected wheat in the last decades, especially in America, and more recently in Asia (Cruz & Valent, 2017). *M. oryzae* shares many characteristics associated with other important plant pathogens, such as appressorium formation and intracellular tissue invasion (Ryder & Talbot, 2015). Because of its economic importance and genetic tractability, *M. oryzae* is being extensively studied and represents an important model for the investigation of fungal diseases in crops (Yan & Talbot, 2016).

Rbp35 is not essential for *M. oryzae* viability but is required for fungal development and plant infection (Franceschetti *et al.*, 2011), which has prompted us to investigate in more detail how these cellular processes are regulated by APA in the rice blast fungus. For this purpose, we have identified genome-wide the PAS usage of the *M. oryzae* rice isolate Guy11 and the Δ*rbp35* mutant under different nutrient conditions and during infection. More than half of *M. oryzae* expressed genes are alternatively polyadenylated and their PAS regions contain novel motifs not found in yeast. Deletion of Rbp35 does not affect mRNA poly(A) tail lengths *in vivo*. However, Rbp35 is required for distal PAS selection through the recognition of the UGUAH motif. We have confirmed that Rbp35 binds the UGUAH *in vitro*, and suggest a mode of non-canonical PAS selection mechanism for Rbp35. Importantly, APA is active during infection: distal PAS selection prevails on a subset of genes with signaling and virulence functions during *M. oryzae* plant colonisation, and it is significant the participation of the long 3'UTR isoform of *14-3-3B* in *M. oryzae* pathogenicity. Particularly, we have

observed that Rbp35 regulates the expression of different types of transposons, which points the potential role of Rbp35 and APA in *M. oryzae* genome evolution.

MATERIALS AND METHODS

Fungal media, growth conditions and infection assays

Fungal strains were grown on CM (complete medium), MM (minimal medium) and DCM (defined complex medium) (Foster *et al.*, 2003). Starvation conditions were carried out on MM depleted of nitrogen or carbon sources. Leaf and root infections were conducted as previously described (Tucker *et al.*, 2010).

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 (Galhano *et al.*, 2017). Three biological replicates were independently harvested and extracted for each library preparation. Library preparation and RNA sequencing on illumina HiSeq 2000 platform was carried out as previously described by EMBL (Heidelberg) sequencing service (Wilkening *et al.*, 2013).

An additional standard RNA sequencing experiment was carried out using infected plant tissues. Samples were collected from infected roots or leaves at 8h, 24h, 48h and 72h post-inoculation. Uninfected rice and *M. oryzae* spores were included as control samples. Two biological replicates were sequenced in parallel. Total RNA was extracted as previously described. Twenty-two libraries were sequenced in ten channels on Illumina HiSeq™ 2000 (2x75 pb) at Centro de Análisis Genómico (CNAG), Spain.

Additional analyses and experiments including mRNA half-life studies, polysomal profiling, RNA-binding assays and generation of 14-3-3B-derived strains and plasmids are described in Supporting Information Methods S1.

ePAT, qPCR and Western blotting

The ePAT reactions were carried out as previously described (Illana *et al.*, 2017), using 1 µg of total RNA mixed with 1µl ePAT primer (100 µM) or TVN-ePAT primer (100 µM) for size control reaction (Jänicke *et al.*, 2012) (Table S1). For qPCR reactions and Western blotting we followed previously described protocols (Franceschetti *et al.*, 2011; Rodríguez-Romero *et al.*, 2015). qPCR reactions were performed using primers detailed in Table S1. The 14-3-3 (pan) polyclonal antibody (Cell Signaling

#8312) recognizes total levels 14-3-3 proteins. This antibody detects all known isoforms of mammalian 14-3-3 proteins and is predicted to detect 14-3-3 homologs in other species.

Availability of data and materials

The data that support the findings of this study have been deposited at the National Center for Biotechnology Information (NCBI) with accession number SRX3390558. A database and publicly-available SPARQL Endpoint is available at [https://dydra.com/markw/polyadenylation-sites-in-magnaporthe-oryza/sparql]. This contains all PAS data, represented using the Resource Description Framework [https://www.w3.org/RDF/] Semantic Web standard following the FAIR Data Principles (Wilkinson *et al.*, 2016). A user-friendly interface for this database, providing search based on *Magnaporthe oryzae* genome locus ID, is available on the Web at http://linkeddata.systems/Magnaporthe/polyA_Sites/.

RESULTS

Polyadenylation sites identified for *M. oryzae* 8,327 annotated gene transcripts

Using a mapping protocol termed 3'T-fill based on deep sequencing of nucleotides located before the polyadenosine tail (Wilkening *et al.*, 2013), we completed a genome-wide analysis of *M. oryzae* polyadenylation sites (PASs; Figure S1A). A total of 24 libraries derived from two different *M. oryzae* strains, the wild-type Guy11 and the Δ*rbp35* mutant; four different growth conditions, i.e. complete medium (CM), minimal medium (MM), nitrogen starvation (MM-N) and carbon starvation (MM-C); and three biological replicates were built for this purpose. The data files ranged from 4,751,592 to 11,517,077 total reads with an average distance between pair-end reads of ~100 bp. The three replicates correlation varied between 92% and 98% (Figure S1B). Reads were mapped against the complete genome of *M. oryzae* 70-15 (Dean *et al.*, 2005), which contains 12,593 protein-coding genes and 13,218 annotated gene transcripts (MG8/version 27, Ensembl Fungi; http://fungi.ensembl.org/index.html).

Taking all wild-type data sets, a total of 14,593 PASs were reliably assigned to annotated features. The majority of PASs (>90%) localized to 3´UTRs, with the rest appearing in 5´UTRs, coding sequences (CDSs), and introns (Figure S1C). In the wild-type strain, and combining all growth conditions, 8,327 annotated transcripts that represent ~71% of the total expressed transcripts, were found with at least one high-confidence PAS. In CM, *M. oryzae* expressed 7,105 genes while the number of expressed genes in carbon-starved cells reached 7,607 (Figure S1D).

Transcription not associated with annotated genes is widespread in *M. oryzae*

We found 261 PASs with significant expression (>100 supporting reads) in the wild-type combining all data that could not be assigned to any annotated genomic feature, and were classified as "orphan" PASs (Figure S2A-B). Of these, 81 (31%) overlapped with gene regions in the antisense orientation (Figure S2B-C and Table S2). The remaining 180 (69%) orphan PASs mapped to intergenic regions that usually showed homology to transposable elements like MAGGY (Farman *et al.*, 1996a), Inago2 (Sanchez *et al.*, 2011), Pot3 (Dioh *et al.*, 2000), MINE (Couch *et al.*, 2005) and MGL (Dioh *et al.*, 2000). Many of these orphan PASs were found to be differentially expressed between different growth media (Figure S2C). 3'UTRs that matched antisense orphans PASs were significantly longer than those not containing any orphan PAS (Figure S2D). Decreasing the stringency of PAS calling to only 10 supporting reads, we found 3,165 orphan PASs, suggesting that transcription unlinked from genomic features is widespread. Of those, 1,097 (35%) overlapped antisense with annotated genes and 698 were differentially expressed (Figure S2C).

~52% of *M. oryzae* genes are alternatively polyadenylated

Combining all wild-type data sets, 4,283 genes had two or more PASs (Figure 1A). This is consistent with the reported high level of alternative polyadenylation (APA) in many species (Mayr, 2017; Tian & Manley, 2017). To determine if APA was more frequently observed in specific functional gene groups, we performed a statistical enrichment analysis using FungiFun2 (Priebe *et al.*, 2015), and found several functional categories (FunCat) to be enriched considering all alternatively polyadenylated genes (Figure 1B). Notably, APA occurred more frequently than expected in genes associated with nucleotide-, DNA-, RNA- and protein-binding, cell cycle control, protein translocation and cellular polarity. Remarkably, global modulators of gene expression showed alternatively polyadenylated mRNA isoforms (Table S3), such as transcriptional regulators (79 genes), β-importins (10 genes), RNA-binding proteins (41 genes), proteins involved in Ca2+-signaling and homeostasis (13 genes), and other signaling proteins (42 genes) including Pmk1, Mst7, Mst11 and Mst50. Overall, this analysis provided an indication of the potential of this post-transcriptional mechanism to regulate gene networks and fungal pathogenesis in *M. oryzae*.

Nucleotide context of *M. oryzae* PASs differs from yeast

We analyzed the nucleotide context of the 14,593 PASs and found several differences from that of *S. cerevisiae* (Figure 1C). The A-rich motif was less uniform than in yeast (Figure 1D); the canonical AAUAAA motif, ubiquitous in metazoans, was still the most abundant but accounted for only 7% of

the *M. oryzae* PASs. Other A-rich motifs identified were AUAAAA, AAUACA and AAGAAA. Although the yeast upstream U-rich element (UUE), also present in all studied eukaryotes, was clearly visible, the downstream U-rich element (DUE) was less conspicuous in *M. oryzae*, and the polyadenylation efficiency element (PEE) recognized by yeast Hrp1 and represented by the UAYRUA consensus sequence (Wilkening *et al.*, 2013), was entirely absent (Figure 1E). Inspecting more closely the *M. oryzae* nucleotide profile, we identified three novel motifs; the UGUAH and UAGNH, located upstream of the A-rich region at approximately -50 bp and -35 bp, respectively, and the HGUGA motif located downstream of the A-rich region at +20 (Figure 1F). In addition, *M. oryzae* PASs showed a preference for SA dinucleotides at the cut-site position, rather than YA dinucleotides observed in *S. cerevisiae* (Figure 1C).

Carbon starvation strongly affects gene expression and increases the use of distal PASs

The highest number of up/down-regulated genes in the wild-type strain arose when comparing CM and MM-C, involving 4,649 genes (Figure 2A, Table S4). This indicated that the depletion of carbon had a stronger impact on *M. oryzae* gene expression than the lack of nitrogen. Enrichment analysis of FunCat terms in these gene sets showed that the fungal adaptation to carbon free medium consists mostly of the down-regulation of transcripts associated with translation, oxidation-reduction, and biosynthesis of organic compounds. Up-regulated genes were linked to cellular sensing, transport, and fatty acid/carbohydrate metabolisms (Figure S2E).

Testing if nutritional adaptation can influence polyadenylation, we found that carbon starvation induces the highest number of significant changes in PAS usage. The change from CM to MM-C induced differential PAS selection (DPS) in 455 genes (Figure 2B, Table S5). Of these, 346 (76%) showed a preference for distal PASs under carbon starvation (Figure 2C). Similar patterns were observed in the change from MM to MM-C (Figure 2C).

Of the 455 genes, we observed that 31 had PASs in both their 5'UTR and 3'UTR. These were differentially selected, with the preference shifting from the 5'UTR PAS to the 3'UTR PAS, or vice-versa, in carbon-starved cells (Figure S2F and Table S6), which suggested the existence of distinct post-transcriptional mechanisms regulating their expression. Examining the nucleotide context of the preferred PASs in MM-C, the UUE was frequently depleted (Figure 2D), suggesting that proteins involved in the recognition of this motif are under differential regulation. In fact, the yeast UUE interacts with Yth1, a protein component of the Polyadenylation Factor I complex (Barabino *et al.*, 1997), and the *M. oryzae* ortholog of Yth1, MGG_00680, is down-regulated in carbon starvation.

Finally, we analysed any potential link between gene expression and use of alternative PASs by looking at genes differentially expressed in the change from CM to MM-C (Figure 2A; 4,649 genes) and genes with DPS under the same condition (455 genes; Figure 2C). We found only 227 differentially expressed genes (49% of the total 455 genes) that showed DPS (Figure 2E). The total number of non-redundant genes expressed in CM and MM-C in the wild-type is 8,100 genes. This means that the percentage of genes expected to change expression from CM to MM-C (4,649 over 8,100; 57%), is higher than the percentage of genes with differential expression and DPS (49%; *p*-value/lower tail of 0.0005). Thus, genes with DPS have generally reduced tendency to change expression in the change from CM to MM-C, suggesting that DPS and differential expression are two events not strictly correlated in *M. oryzae*.

The Δ*rbp35* mutant exhibits fewer changes in response to differing nutrient conditions

We analyzed how the lack of Rbp35 affected gene expression under different nutritional conditions. With the exception of the change from CM to MM, the Δ*rbp35* mutant was dramatically altered in its ability to adapt to nutritional changes as revealed by a 48% reduction in the number of differentially expressed genes (3,091) compared to the wild-type (5,935; Figure 2A). The strongest gene expression changes were observed between CM/MM and MM-C. When compared to the wild-type, combining all growth media, the mutant showed 1,236 differentially expressed genes (Figure 2F). A FunCat analysis found terms significantly enriched only in Δ*rbp35* down-regulated genes (732 genes), with functions associated with secondary metabolism, cellular transport and protein processing (Figure 3A). Particularly, we found several genes already characterized in *M. oryzae* encoding transcriptional regulators (8 genes), G protein-coupled receptors (4 genes) and proteins involved in synthesis of secondary metabolites (16 genes) such as melanin (Table S7). Our results are consistent with the defects displayed by Δ*rbp35* in the target of rapamycin (TOR) pathway (Franceschetti *et al.*, 2011; Marroquin-Guzman & Wilson, 2015), one of the major signaling pathways involved in the response to nutritional stimuli (Gonzalez & Hall, 2017).

Rbp35 is required for distal PAS selection and recognition of the UGUAH motif

We compared all identified PASs between wild-type and mutant. Overall, ~39% (3,218) of all expressed genes displayed DPS in Δ*rbp35* compared to wild-type (Figure 3B). Approximately ~75% of these showed a preference for proximal PASs, which resulted in a global shortening of 3'UTRs in Δ*rbp35* (Figure 3C, Figure S3A). The 3'UTR length variation ranged from a few nucleotides to more than one thousand. The majority of 3'UTRs changed no more than 200 nt. Further analysis showed

that 61% (2,595) of alternatively polyadenylated genes in the wild-type were also regulated by Rbp35 (*p*-value< 2.2e-308, Figure S3B), which supported the relevant role of Rbp35 in the regulation of APA in *M. oryzae*.

We analyzed the nucleotide context of new PASs in Δ*rbp35* that were not previously identified in the wild-type. These PASs appeared to represent genuine PASs because both nucleotide profile and predicted RNA secondary structure resembled "canonical" PASs (Figure S3C-D), with slight variations described below. However, they were never selected in the wild-type under the conditions tested, which suggested that their selection in Δ*rbp35* was not the result of random cleavage but rather governed by similar signals as in non-Rbp35-regulated PASs.

A motif analysis of PASs that newly appeared, or became preferentially selected in Δ*rbp35* found strongly depleted the UGUAH element at approximately -55 nt (Figure 3D; purple line), compared to transcripts with PASs not affected by the lack of Rbp35 (Figure 3D; grey line). The overall structure of these PASs were slightly different; for those preferentially selected in Δ*rbp35* the UGUAH motif was depleted overall, however when present, appeared closer to the cut site (-30 nt; Figure 3D purple line) than in those preferred in WT (-55nt; Figure 3D blue line). This indicated that Rbp35 binding is involved in selection of PASs with a UGUAH element at -55 nt; however when the UGUAH element lies closer to a potential cut site, binding of Rbp35 inhibits the cleavage reaction possibly by occlusion (Figure 3E).

We confirmed that Rbp35 was sufficient to bind the UGUAH motif *in vitro* as opposed to CFI25 carrying out RNA-binding assays on native polyacrylamide gels (Figure 3F and Figure S3E). In these assays, the probe disappeared when we increased the amount of Rbp35. However, the corresponding band shift of the probe bound to Rbp35 was not detected, possibly due to its concealment by either Rbp35 or Rbp35/CFI25 complex. The presence of the proteins was confirmed probing the membranes with the anti-His antibody. CFI25 was not enough to sequester the probe, although it increased Rbp35 binding efficiency. Double amount of Rbp35 was needed to bind the 4xUGUAH probe at similar rates as the 2xUGUAH probe, which supported further the binding site specificity of Rbp35 (Figure 3F).

We also looked at the conservation of the UGUAH motif within the fungal kingdom (Figure S3F). Only fungal species that did not contain clear homologues of the Rbp35/CFI25 complex, e.g. *S. cerevisiae*, *Schizosaccharomyces pombe* and *Candida albicans* showed alterations in the conservation of the UGUAH motif, which was present although in different positions.

The A-rich element AAUAHA was also significantly reduced in PASs preferentially selected in Δ*rbp35* and in wild-type carbon-starved cells (Figure 3G), possibly due to changes in the polyadenylation machinery protein composition.

Rbp35 regulates expression of non-coding RNAs and transposable elements

Among all genes with DPS in the mutant, only 385 (12%) showed differential expression compared to wild-type (*p*-value: 1.8e-09, Figure S3G and Table S8), which revealed that mRNAs with altered PASs in Δ*rbp35* were not usually affected in their expression levels. By contrast, tRNAs and some transposable elements (TEs) showed significant changes in their expression levels in Δ*rbp35* (Figure S3H). We know that the observed expression changes in tRNAs and rRNAs have little effect in global translation in the mutant (Rodríguez-Romero *et al.*, 2015). The up-regulation of the expression of one of the most active TEs in *M. oryzae* such as MAGGY (Farman *et al.*, 1996b; Ikeda *et al.*, 2001), can explain the high tendency of the Δ*rbp35* mutant to generate colony sectors (Figure S3I).

RNA pol III lacks the CTD domain present in RNA pol II required for recruiting CPSF into the polyadenylation complex. Therefore, RNA pol III-dependent transcripts do not undergo canonical polyadenylation (Ustyantsev *et al.*, 2017). Non-canonical polyadenylation plays an important role in RNA quality control pathways in eukaryotes, and sends many ncRNAs into degradation pathways, including tRNAs, rRNAs, snRNAs, and snoRNAs (Reinisch & Wolin, 2007). In the nucleus, one of these degradation pathways is regulated by the yeast TRAMP complex, which adds a short tail of adenosines in the 3' end of the ncRNA to trigger its processing or degradation by the nuclear exosome (LaCava *et al.*, 2005). Poly(A) stretches have also been found on human 18S and 28S rRNA fragments combined with other nucleotides, suggesting the existence of similar degradation mechanisms for rRNAs (Slomovic *et al.*, 2006). Therefore, it is likely that these modified ncRNAs have been captured during library preparation in our experiments. Consequently, the observed changes in expression for tRNAs and rRNAs and in DPS for tRNAs in the Δ*rbp35* mutant are possibly related with deficiencies in ncRNA degradation pathways.

Rbp35 regulates alternative polyadenylation of genes with signaling and virulence functions

We looked at the 315 characterised *M. oryzae* genes with virulence-related functions in the host-pathogen interaction database (http://www.phi-base.org/) (Urban *et al.*, 2017). Among these, 270 were expressed in one or more of the growth conditions in our experiment and 127 (~47%) showed DPS in Δ*rbp35* (Table S9). The correlation between genes with DPS and genes with virulence

functions in Δ*rbp35* surpassed statistical significance (Fisher's exact test *p*-value:1.6e-05), suggesting that Rbp35 is an important regulator of PASs in virulence-associated genes.

Among the 127 virulence genes, 58 encoded transcription factors and components of signalling cascades such as the two mitogen-activated protein kinases (MAPKs) *PMK1* and *MPS1*, and their upstream components *MST7,* a MAPK kinase (Zhao & Xu, 2007) (MEK), and *MST11*, a MEK kinase (Park *et al.*, 2006) (MEKK) (Table S8). We confirmed by extension poly(A) test (ePAT) the involvement of Rbp35 in the regulation of PAS selection in *MST7* (Figure 4A), and two novel signalling components, the *M. oryzae 14-3-3* genes, which participate in signalling pathways such as TOR (Figure S4A and Table S5).

Rbp35 does not regulate poly(A) tail length

To investigate if poly(A) tails were contributing to defects exhibited by the Δ*rbp35* mutant, we analyzed and compared poly(A) tail lengths from wild-type and Δ*rbp35* using bulk mRNA. No obvious alteration in poly(A) tail lengths was detected in Δ*rbp35* transcripts (Figure 4B). *M. oryzae* poly(A) tails were >100 nt in length, whereas the maximum poly(A) tail length of *S. cerevisiae* was ~70 nt. This result indicates that Δ*rbp35*-associated defects are not caused by changes in mRNAs poly(A) tail length.

mRNAs with short 3'UTRs are more stable

Since 3'UTRs can control mRNAs half-life (Barreau *et al.*, 2005; Bartel, 2009; Hogg & Goff, 2010), we measured global mRNA stability to see if it was affected in Δ*rbp35*. We carried out global transcriptional shut-off experiments applying the transcriptional inhibitor thiolutin to actively growing mycelia (Pelechano & Perez-Ortin, 2008), followed by dot-blotting of extracted mRNAs and subsequent probing with biotin-labeled oligo(dT) (Rouble *et al.*, 2014) (Figure 4C). The intensity of the signal found in Δ*rbp35* mRNAs was stronger, suggesting that the shortening of mRNAs in Δ*rbp35* increased their stability, at least for several of them. We then analysed the involvement of long 3'UTRs in transcript stability, particularly in three Rbp35-dependent transcripts associated with signalling, *MST7*, *14-3-3A* and *14-3-3B*. In all cases, mRNAs with short 3'UTRs were more stable than those with long 3'UTRs (Figure 4D).

mRNA isoforms can accumulate differentially in polysomal fractions

Long 3'UTRs can regulate the translation efficiency of a given transcript (Floor & Doudna, 2016), and we analysed this by looking at the abundance of mRNA isoforms in polysomal fractions under

different nutrient conditions. Similar polysomal profiles were obtained for wild-type and Δ*rbp35* strains (Figure 4E), which supported that the lack of Rbp35 does not affect overall translation. In carbon-starved cells, we observe an overall reduction of the 80S complex, which is responsible of translation initiation (Sonenberg & Hinnebusch, 2009), followed by a reduction in polysomal mRNAs in both strains. This suggested that the lack of carbon reduced canonical translation in the fungal cell. We then looked at the presence of mRNA isoforms regulated by Rbp35 in polysomal fractions. The location of *MST7* mRNA variants in polysomes was not affected by nutritional conditions (Figure 4F). Long mRNAs were observed less frequently in *MST7* than in *14-3-3A* and *14-3-3B*, both of which had predominantly long mRNA isoforms loaded into polysomes. An increase of mRNAs in Δ*rbp35* polysomal fractions was only observed for *14-3-3A*. Overall, our results indicated that *14-3-3A* and *14-3-3B* mRNAs with long 3'UTRs accumulated better than those with short 3'UTRs in polysomal fractions, suggesting they were being translated more efficiently. By contrast, *MST7* mRNA variants showed no clear preference for their accumulation in polysomes.

Distal PAS selection prevails during leaf infection

In order to identify alternatively polyadenylated genes and their mRNA isoforms abundance during plant infection we carried out an additional transcriptome analysis, a standard RNA-seq using RNAs extracted from a time course infection experiment at 8h, 24h, 48h and 72h. Fungal spores were collected from CM media and sprayed on either three week-old rice plants or one week-old roots that have grown on water-agar petri dishes. In our experiment, the percentage of detected fungal reads varied between 0.02-0.07% on leaves to 0.02-4% on roots (Figure S4B).

We compared the PASs of *M. oryzae* transcripts found in spores and inside the plant tissues with *in vitro* fungal PASs (Figure S4C). Relative expression levels of PASs extracted from spores and leaves resembled more to those found in carbon- and nitrogen-starved cells, respectively, whereas relative expression levels of PASs from infected roots differed only from MM-C (Figure 5A). From a biological perspective, this suggested that carbon sources are scarce in spores in contrast to roots, and that nitrogen availability for *M. oryzae* is challenging on leaves.

When looking at different time points of infection (Figure 5B), a relatively high proportion of fungal mRNAs displayed DPS, with long 3'UTRs observed on infected leaves at all time points (13-22%) and on roots only at 8h (35%; Table S10). *MST7*, *14-3-3A* and *14-3-3B* pre-mRNAs also varied their PAS selection during infection. We confirmed their 3'UTR length change *in planta* by qPCR (Figure 5C). The *14-3-3B* gene differed significantly in the abundance of mRNAs with long 3'UTR in infected plants and was selected for further studies.

The 3'UTR of *14-3-3B* regulates *M. oryzae* plant infection

We found three PASs in *14-3-3B* mRNA, with the distance between the first and last PAS being ~550 nt (Figure S4A, left panel). We confirmed the presence of these PASs by ePAT and the abundance of short *14-3-3B* transcripts in Δ*rbp35*, supporting the Rbp35 preference for distal PASs (Figure S4A, right panel). To investigate any functional role of the long 3'UTR of *14-3-3B* we generated the deletion mutant Δ*14-3-3b* (Figure S4D-F). We complemented this mutant with a *14-3-3B:mCherry* fusion construct that either contained its native 3'UTR (*14-3-3B-S/L* for short and long 3'UTR) or included only the first PAS (*14-3-3B-S* for short 3'UTR). The Δ*14-3-3b* mutant was slightly affected in growth on CM, MM and under different stresses including osmotic stress (Figure S5A). The *14-3-3B-S* and *14-3-3B-S/L* constructs complemented Δ*14-3-3b* growth defects in MM supplemented with NaCl (Figure S5B), indicating that the C-terminal mCherry tag and the 3'UTR length were not affecting protein function.

Infection tests revealed that Δ*14-3-3b* was strongly impaired in leaf and root infection (Figure 6A). A strong symptom reduction was observed in Δ*14-3-3b* carrying the short 3'UTR construct whereas the strain with *14-3-3B-S/L* showed wild-type disease symptoms, suggesting a relevant function for the full 3'UTR length of *14-3-3B* during infection. In addition, the *14-3-3B-S/L* construct retained the ability to reduce its expression levels in MM-C; however, this preference was lost in *14-3-3B-S* (Figure 6B), suggesting the existence of a region within the long 3'UTR responsible of this effect. Supporting this, a strong preference for the use of one of the two distal PASs of *14-3-3B* mRNA was observed in MM-C (Figure S4A).

We found less *14-3-3B-S* mRNAs compared to *14-3-3B-S/L* despite the short isoform is more stable, and the protein was barely detectable in the Δ*14-3-3b/14-3-3B-S* strain (Figure 6C-D). The 45% reduction of mRNA levels and the failure of short *14-3-3B* transcripts to translate efficiently could explain this drastic reduction in protein expression in Δ*14-3-3b/14-3-3B-S*. Confocal imaging confirmed 14-3-3b protein abundance in the Δ*14-3-3b*/*14-3-3B-S/L* strain and located 14-3-3b in the cytoplasm within aggregates near the spore cell wall (Figure 6D).

Discussion

APA can generate multiple mRNA isoforms that can be susceptible to further post-transcriptional regulatory events, including mRNA stability, localisation or translation (Mayr, 2017; Tian & Manley, 2017). Therefore, APA must be tightly controlled in eukaryotic cells.

The *Magnaporthe oryzae* Δ*rbp35* mutant is impaired in fungal development and plant colonisation, indicating that Rbp35 regulates APA of mRNAs associated with these processes. To identify these mRNAs we conducted a genome-wide sequencing of PASs using 3'T-filled approach in *M. oryzae* wild-type and Δ*rbp35* strains under different nutrient conditions This sequencing approach has allowed us to obtain a detailed map of RNA 3' ends. The cDNA libraries have been generated from fragmented RNAs, previously purified from oligo(dT) columns and reverse transcribed with an oligo(dT) primer coupled to an adapter and biotin (Wilkening et al., 2013). Consequently, the status of their 5' ends is undefined. Mature eukaryotic mRNAs contain a cap structure at their 5'end that is removed before they undergo 5′–3′ exonucleolytic decay, a process that occurs co-translationally while they are still associated with ribosomes (Hu et al., 2009). We have assumed that the vast majority of polyadenylated RNAs in our experiment are stable and do not derived from degradation products, meaning that the majority of sequenced molecules are capped and contain intact 5'ends.

APA is observed in ~52% of the *M. oryzae* transcripts. More of 90% of the PASs localize in the 3'UTR although PASs are also found in upstream regions such as CDSs, introns and 5'UTRs. In yeast and mammals, RNAs polyadenylated in the CDS and lacking a stop codon are unstable and rapidly degraded through the non-stop decay mechanism (NSD) (Isken & Maquat, 2007). Polyadenylation within introns can either generate an alternative protein product synthesis or affect the stability of the mRNA via the nonsense-mediated mRNA decay (NMD) pathway (Lykke-Andersen & Jensen, 2015). When APA occurs within 5'UTRs they often play regulatory roles (Hinnebusch *et al.*, 2016). It can result in the transcription of an upstream open reading frame (uORF), which normally precludes the transcription of the downstream protein-coding gene as in the case of the yeast transcriptional activator Gcn4p (Hinnebusch, 2005). uORFs are also frequent in plants, and can play an important roles during stress adaptation (Hellens *et al.*, 2016). In *Arabidopsis thaliana*, hypoxia leads to an increase in the number of uORFs that are overrepresented in polysomes, suggesting that the generation of alternative mRNA isoforms is an important feature of the abiotic stress response (de Lorenzo *et al.*, 2017). In fungi, several genes, including *S. cerevisiae GCN4*, *S. cerevisiae CPA1*, and *Neurospora crassa arg-2*, are regulated by uORFs, which control their expression in response to specific physiological signals (reviewed in Hood *et al.*, 2009). In the case of *M. oryzae*, the *RBP35* gene contains two polyadenylated transcripts within its 5'UTR intron, which are expressed during carbon starvation, and one of them, uORF1, is required for restoring fungal growth in the presence of rapamycin (Rodríguez-Romero *et al.*, 2015).

Pervasive transcription, i.e. transcription not associated with annotated genomic features, is widespread and is seen in both yeast and humans (Jensen *et al.*, 2013). A plethora of non-coding RNAs (ncRNAs), of which most are by-products of transcription, occur as a result of 'leaky' transcription initiation. The *M. oryzae* orphan PASs are possibly associated with these transcription events. However, there is a potential for regulatory roles for several of them since a significant number of antisense orphan PASs match to UTRs in *M. oryzae*. They overlap with longer 3' UTRs more often than with 5' UTRs, similar to yeast (David *et al.*, 2006), and some of them are differentially expressed. We note that they often locate to intergenic regions with homology to transposable elements (TEs). In humans, it is frequent to find non-annotated transcription start sites (TSS) in long terminal repeats (LTRs) modulated by epigenetic mechanisms (Brocks *et al.*, 2017). In our experiments, this type of epigenetic regulation can be responsible for the distinct expression of orphan PASs observed in *M. oryzae* TEs under different nutritional conditions.

A small RNA profiling study (<200 nt) identified capped and polyadenylated sRNAs (CPA-sRNAs) in *M. oryzae* (Gowda *et al.*, 2010). A total of 14,547 CPA-sRNAs were found ranging between 16 and 218 nt in length, and mapped primarily to transcription initiation and termination sites of protein-coding genes. Of these, 14% (1,994) matched with our sequencing data and located mainly in 3' UTRs (1,107 / 55%), CDSs (244 / 12%), 5' UTRs (148 / 7%) and intergenic regions (395 / 20%), similarly to previously described CPA-sRNAs (Gowda *et al.*, 2010). The orphan PASs of our study did not match with CPA-sRNAs.

The *M. oryzae* polyadenylation signal is not well conserved compared to yeast (Ozsolak *et al.*, 2010; Mata, 2013), and many variations of the canonical AAUAAA are allowed. The *M. oryzae* PAS also contains three novel auxiliary elements compared to the yeast PAS: UGUAH and UAGNH located upstream of the A-rich region, and HGUGA located downstream. In addition, the *M. oryzae* PAS has lost the yeast polyadenylation efficiency element (PEE), which interacts with the CFIB/Hrp1 protein (Perez-Canadillas, 2006). Rbp35 is not found in *S. cerevisiae*, which may explain the absence of the UGUAH element in yeast PASs. The significance of the UAGNH motif is unknown, but considering its occurrence and location it is plausible that it is being recognized by a protein only present in *M. oryzae* polyadenylation machinery. The depletion of yeast motifs in *M. oryzae* PASs like the upstream U-rich element (UEE), and appearance of novel ones also point to differences between the components of *M. oryzae* and yeast polyadenylation machineries, either in protein composition or modular structure. Remarkably, Rbp35 recognizes UGUAH without the need of CFI25 as opposed to humans, whose CFIm59/68 only bind to RNA in the presence of CFIm25 (Dettwiler *et al.*, 2004).

We used carbon and nitrogen starvation to expand the coverage of PASs identified in *M. oryzae* to genes potentially required for plant colonization. Nitrogen and carbon metabolisms play a significant role during plant infection (Fernandez *et al.*, 2014; Marroquin-Guzman *et al.*, 2017), and their starvation can trigger expression of genes associated with infection (Donofrio *et al.*, 2006). There is a strong preference for distal PAS usage during carbon starvation, regulating up to 553 alternatively polyadenylated genes (Table S5). This differential PAS selection (DPS) normally affects to 3'UTRs distal PASs. A significant number of genes associated with signaling undergo DPS in MM-C, including at least 29 transcription factors, and 19 protein kinases and phosphatases, several of which have been shown to be necessary for infection such as Mst12 (Park *et al.*, 2002) and Mps1 (Xu *et al.*, 1998).

The Δ*rbp35* mutant shows a higher number of PASs and 39% of all expressed genes are affected in PAS selection in Δ*rbp35*. The majority of them are alternatively polyadenylated in the wild-type, indicating that Rbp35 is an important regulator of APA in *M. oryzae*. The vast majority of 3'UTRs are shorter in Δ*rbp35* compared to the wild-type, which indicates that Rbp35 is required for distal PAS selection as with the human CFI_m68 (Martin *et al.*, 2012), despite of showing little homology.

Generally, transcripts with altered PASs in Δ*rbp35* show no changes in expression. However, some repetitive elements of endogenous (rDNA and tRNAs) or exogenous (transposon elements, TEs) origin show both differential expression as well as differential PAS selection in Δ*rbp35.* The observed tendency of the Δ*rbp35* mutant to revert its developmental defects can be due to dysregulation of several TEs, which are known to be a major tool for genetic rearrangement and pathogenic evolution in the rice blast fungus (Chadha & Sharma, 2014).

Transcripts are generally more stable in Δ*rbp35* than in wild-type cells, and this correlates with the long half-lives of short *MST7*, *14-3-3A* and *14-3-3B* mRNA isoforms. Long 3'UTRs are more frequent targets for non-mediated decay (Hogg & Goff, 2010), a process that drives specific mRNAs for degradation and can control their translation rates (Merchante *et al.*, 2015).

Carbon starvation consistently reduced the formation of the 80S complex and the amount of mRNAs loaded into high molecular weight polysomes. This suggests that overall translation is reduced, in accordance with what it is known in mammalian systems, where cells under severe stress can re-programme their translational status inhibiting canonical or cap-dependent translation to maintain the expression of specific proteins involved in growth and survival (Spriggs *et al.*, 2010). We also observed that *14-3-3A* and *14-3-3B* mRNAs with long 3'UTRs accumulate in polysomal fractions, which points that at least for these two genes, a positive correlation is found between 3'UTR

length and translation efficiency. In transcripts of murine and human naive and activated T cells, the shortening of 3'UTRs does not affect mRNA and protein levels, but influences the RNA-binding protein interactome (Gruber *et al.*, 2014), suggesting that it may well be happening in *M. oryzae*.

The analyses of *in planta* fungal transcripts reveal differences between leaf and root infection pathways (Figure 7). The mRNAs found in spores tend to have long 3'UTRs, as in carbon-starved cells, and suggest that this is required for fungal adaptation at early stages of infection (0 – 8h). After 24h the preference for long 3'UTRs is reduced but still maintained on leaves, as opposed to roots where the number of short mRNAs rises significantly. A high percentage of 3'UTR transcripts remain unaltered during infection, which points the existence of regulatory mechanisms for those transcripts that show different PAS usage. Rbp35 regulates a significant number of infection-associated transcripts, and particularly those involved in signaling cascades. These genes are alternatively polyadenylated and subjected to DPS in carbon starved cells, which supports the importance of the regulation of 3'UTR lengths in signaling components.

To our knowledge, this is the first time that a genome-wide approach has been carried out to identify PASs in a filamentous fungus. We show differences in the mode of recognition of fungal CFI complex compared to humans, and propose a mechanism for Rbp35 non-canonical PAS selection. We provide significant evidence that differential PAS selection is active *in planta*, and that the long 3'UTR of *14-3-3B* is required for full disease symptoms production. Our results suggest that APA allows fine tuning of proteins required to adapt fungal growth by producing long mRNAs with higher turnover and lower translation rates. The involvement of Rbp35 in the regulation of PAS selection in TEs points its potential capacity to drive genomic rearrangements. This study will help improving fungal genome annotations, and offers new avenues for understanding pathogen evolution and regulatory mechanisms controlling fungal host invasion.

ACKNOWLEDGMENTS

A.S. thanks Cristina Arribas (C.A.) for her technical support. A.S. and M.W. are funded by institutional support from the Universidad Politécnica de Madrid, the Spanish Research council (MINECO, ref. BIO2014-53211-R and TIN2014-55993-RM), and the Community of Madrid (grant ref. S2013/ABI-2734 and PEJD-2016/BIO-2857). J.R.R. and V.O.R. are funded by MINECO (refs. BIO2014-54233-JIN and BES-2015-072234, respectively). M.M. is funded by the Isaac Peral Programme of the Universidad Politécnica de Madrid through a staff award to MW. C.A. contract was co-funded by the Community of Madrid and the European Commission (Youth Employment Initiative, YEI) ref. PEJ15/BIO/TL-0086.

Author contributions

A.S. conceived the study; J.R.R., M.M., V.O.C., M.D., M.W. and A.S. conceived, designed and performed all experiments. M.M. analyzed all the RNA-seq data. J.R.R., M.M., V.O.C., M.D., M.W. and A.S. contributed to data analysis and writing of the manuscript.

REFERENCES

Barabino SM, Hubner W, Jenny A, Minvielle-Sebastia L, Keller W. 1997. The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins. *Genes & Development* **11**(13): 1703-1716.

Barreau C, Paillard L, Osborne HB. 2005. AU-rich elements and associated factors: are there unifying principles? *Nucleic Acids Research* **33**(22): 7138-7150.

Bartel DP. 2009. MicroRNAs: Target Recognition and Regulatory Functions. *Cell* **136**(2): 215-233.

Berkovits BD, Mayr C. 2015. Alternative 3 ' UTRs act as scaffolds to regulate membrane protein localization. *Nature* **522**(7556): 363-+.

Brocks D, Schmidt CR, Daskalakis M, Jang HS, Shah NM, Li DF, Li J, Zhang B, Hou YR, Laudato S, et al. 2017. DNMT and HDAC inhibitors induce cryptic transcription start sites encoded in long terminal repeats. *Nature Genetics* **49**(7): 1052-1060.

Chadha S, Sharma M. 2014. Transposable Elements as Stress Adaptive Capacitors Induce Genomic Instability in Fungal Pathogen *Magnaporthe oryzae*. *PLoS One* **9**(4): ARTN e94415.

Chan S, Choi EA, Shi YS. 2011. Pre-mRNA 3 '-end processing complex assembly and function. *Wiley Interdisciplinary Reviews RNA* **2**(3): 321-335.

Couch BC, Fudal I, Lebrun MH, Tharreau D, Valent B, van Kim P, Notteghem JL, Kohn LM. 2005. Origins of host-specific populations of the blast pathogen *Magnaporthe oryzae* in crop domestication with subsequent expansion of pandemic clones on rice and weeds of rice. *Genetics* **170**: 613-630.

Cruz CD, Valent B. 2017. Wheat blast disease: danger on the move. *Tropical Plant Pathology* **42**(3): 210-222.

David L, Huber W, Granovskaia M, Toedling J, Palm CJ, Bofkin L, Jones T, Davis RW, Steinmetz LM. 2006. A high-resolution map of transcription in the yeast genome. *Proceedings of the National Academy of Sciences of the United States of America* **103**(14): 5320-5325.

Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, Thon M, Kulkarni R, Xu JR, Pan H. 2005. The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* **434**: 980-986.

de Lorenzo L, Sorenson R, Bailey-Serres J, Hunt AG. 2017. Noncanonical Alternative Polyadenylation Contributes to Gene Regulation in Response to Hypoxia. *Plant Cell* **29**(6): 1262-1277.

Dettwiler S, Aringhieri C, Cardinale S, Keller W, Barabino SML. 2004. Distinct sequence motifs within the 68-kDa subunit of cleavage factor I-m mediate RNA binding, protein-protein interactions, and subcellular localization. *Journal of Biological Chemistry* **279**(34): 35788-35797.

Dioh W, Tharreau D, Notteghem JL, Orbach M, Marc-Henri L. 2000. Mapping of Avirulence Genes in the Rice Blast Fungus, *Magnaporthe grisea*, with RFLP and RAPD Markers. *Molecular Plant-Microbe Interactions* **13**(2): 217-227.

Donofrio NM, Oh Y, Lundy R, Pan H, Brown DE, Jeong JS, Coughlan S, Mitchell TK, Dean RA. 2006. Global gene expression during nitrogen starvation in the rice blast fungus, *Magnaporthe grisea*. *Fungal Genetics and Biology* **43**(9): 605-617.

Eckmann CR, Rammelt C, Wahle E. 2011. Control of poly(A) tail length. *Wiley Interdisciplinary Reviews-Rna* **2**(3): 348-361.

Farman ML, Taura S, Leong SA. 1996a. The *Magnaporthe grisea* DNA fingerprint probe MGR586 contains the 3' end of an inverted repeat transposon. *Molecular and General Genetics* **251**(6): 675-681.

Farman ML, Tosa Y, Nitta N, Leong SA. 1996b. MAGGY, a retrotransposon in the genome of the rice blast fungus Magnaporthe grisea. *Molecular and General Genetics* **251**(6): 665-674.

Fernandez J, Marroquin-Guzman M, Wilson RA. 2014. Mechanisms of Nutrient Acquisition and Utilization During Fungal Infections of Leaves. *Annual Review of Phytopathology* **52**: 155-174.

Fisher MC, Henk DA, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, Gurr SJ. 2012. Emerging fungal threats to animal, plant and ecosystem health. *Nature* **484**(7393): 186-194.

Floor SN, Doudna JA. 2016. Tunable protein synthesis by transcript isoforms in human cells. *Elife* **5**.

Foster AJ, Jenkinson JM, Talbot NJ. 2003. Trehalose synthesis and metabolism are required at different stages of plant infection by *Magnaporthe grisea*. *Embo Journal* **22**(2): 225-235.

Franceschetti M, Bueno E, Wilson RA, Tucker SL, Gómez-Mena C, Calder G, Sesma A. 2011. Fungal Virulence and Development Is Regulated by Alternative Pre-mRNA 3' End Processing in *Magnaporthe oryzae*. *Plos Pathogens* **7**(12): e1002441.

Galhano R, Illana A, Ryder LS, Rodriguez-Romero J, Demuez M, Badaruddin M, Martinez-Rocha AL, Soanes DM, Studholme DJ, Talbot NJ, et al. 2017. Tpc1 is an important $Zn(II)_2Cys_6$ transcriptional regulator required for polarized growth and virulence in the rice blast fungus. *PLoS Pathog* **13**(7): e1006516.

Gonzalez A, Hall MN. 2017. Nutrient sensing and TOR signaling in yeast and mammals. *Embo Journal* **36**(4): 397–408.

Gowda M, Nunes CC, Sailsbery J, Xue MF, Chen F, Nelson CA, Brown DE, Oh Y, Meng SW, Mitchell T, et al. 2010. Genome-wide characterization of methylguanosine-capped and polyadenylated small RNAs in the rice blast fungus *Magnaporthe oryzae*. *Nucleic Acids Research* **38**(21): 7558-7569.

Gruber AR, Martin G, Muller P, Schmidt A, Gruber AJ, Gumienny R, Mittal N, Jayachandran R, Pieters J, Keller W, et al. 2014. Global 3' UTR shortening has a limited effect on protein abundance in proliferating T cells. *Nat Commun* **5**: 5465.

Hellens RP, Brown CM, Chisnal MAW, Waterhouse PM, Macknight RC. 2016. The Emerging World of Small ORFs. *Trends in Plant Science* **21**(4): 317-328.

Hinnebusch AG. 2005. Translational regulation of GCN4 and the general amino acid control of yeast. *Annual Review of Microbiology* **59**: 407-450.

Hinnebusch AG, Ivanov IP, Sonenberg N. 2016. Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science* **352**(6292): 1413-1416.

Hogg JR, Goff SP. 2010. Upf1 senses 3'UTR length to potentiate mRNA decay. *Cell* **143**(3): 379-389.

Hood HM, Neafsey DE, Galagan J, Sachs MS. 2009. Evolutionary Roles of Upstream Open Reading Frames in Mediating Gene Regulation in Fungi. *Annual Review of Microbiology* **63**: 385-409.

Hu WQ, Sweet TJ, Chamnongpol S, Baker KE, Coller J. 2009. Co-translational mRNA decay in *Saccharomyces cerevisiae*. *Nature* **461**(7261): 225-229.

Ikeda K, Nakayashiki H, Takagi M, Tosa Y, Mayama S. 2001. Heat shock, copper sulfate and oxidative stress activate the retrotransposon MAGGY resident in the plant pathogenic fungus *Magnaporthe grisea*. *Molecular Genetics and Genomics* **266**(2): 318-325.

Illana A, Marconi M, Rodriguez-Romero J, Xu P, Dalmay T, Wilkinson MD, Ayllon MA, Sesma A. 2017. Molecular characterization of a novel ssRNA ourmia-like virus from the rice blast fungus *Magnaporthe oryzae*. *Archives of Virology* **162**(3): 891-895.

Illana A, Rodriguez-Romero J, Sesma A 2013. Major Plant Pathogens of the Magnaporthaceae family. In: Horwitz BA ed. *Genomics of Soil- and Plant-associated Fungi*. Berlin Heidelberg: Springer-Verlag Ltd., 45-88.

Isken O, Maquat LE. 2007. Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function. *Genes Dev* **21**(15): 1833-1856.

Jänicke A, Vancuylenberg J, Boag PR, Traven A, Beilharz TH. 2012. ePAT: A simple method to tag adenylated RNA to measure poly(A)-tail length and other 3 ' RACE applications. *RNA* **18**(6): 1289-1295.

Jensen TH, Jacquier A, Libri D. 2013. Dealing with Pervasive Transcription. *Molecular Cell* **52**(4): 473-484.

LaCava J, Houseley J, Saveanu C, Petfalski E, Thompson E, Jacquier A, Tollervey D. 2005. RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell* **121**(5): 713-724.

Lykke-Andersen S, Jensen TH. 2015. Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes. *Nature Reviews Molecular Cell Biology* **16**: 665.

Marroquin-Guzman M, Hartline D, Wright JD, Elowsky C, Bourret TJ, Wilson RA. 2017. The *Magnaporthe oryzae* nitrooxidative stress response suppresses rice innate immunity during blast disease. *Nature Microbiology* **2**: 17054.

Marroquin-Guzman M, Wilson RA. 2015. GATA-Dependent Glutaminolysis Drives Appressorium Formation in *Magnaporthe oryzae* by Suppressing TOR Inhibition of cAMP/PKA Signaling. *PLoS Pathog* **11**(4): e1004851.

Martin G, Gruber AR, Keller W, Zavolan M. 2012. 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 Reports* **1**(6): 753-763.

Mata J. 2013. Genome-wide mapping of polyadenylation sites in fission yeast reveals widespread alternative polyadenylation. *RNA Biol* **10**(8): 1407-1414.

Mayr C. 2017. Regulation by 3'-Untranslated Regions. *Annual Review of Genetics* **51**(1): 171-194.

Merchante C, Brumos J, Yun J, Hu Q, Spencer KR, Enriquez P, Binder BM, Heber S, Stepanova AN, Alonso JM. 2015. Gene-specific translation regulation mediated by the hormone-signaling molecule EIN2. *Cell* **163**(3): 684-697.

Ozsolak F, Kapranov P, Foissac S, Kim SW, Fishilevich E, Monaghan AP, John B, Milos PM. 2010. Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. *Cell* **143**(6): 1018-1029.

Park G, Xue C, Zhao X, Kim Y, Orbach M, Xu J-R. 2006. Multiple Upstream Signals Converge on the Adaptor Protein Mst50 in *Magnaporthe grisea*. *Plant Cell* **18**(10): 2822-2835.

Park G, Xue C, Zheng L, Lam S, Xu JR. 2002. MST12 Regulates Infectious Growth but not Appressorium Formation in the Rice Blast Fungus *Magnaporthe grisea*. *Molecular Plant-Microbe Interactions* **15**(3): 183-192.

Pelechano V, Perez-Ortin JE. 2008. The transcriptional inhibitor thiolutin blocks mRNA degradation in yeast. *Yeast* **25**(2): 85-92.

Perez-Canadillas JM. 2006. Grabbing the message: structural basis of mRNA 3 ' UTR recognition by Hrp1. *Embo Journal* **25**(13): 3167-3178.

Priebe S, Kreisel C, Horn F, Guthke R, Linde J. 2015. FungiFun2: a comprehensive online resource for systematic analysis of gene lists from fungal species. *Bioinformatics* **31**(3): 445-446.

Proudfoot NJ. 2016. Transcriptional termination in mammals: Stopping the RNA polymerase II juggernaut. *Science* **352**(6291): aad9926.

Reinisch KM, Wolin SL. 2007. Emerging themes in non-coding RNA quality control. *Current Opinion in Structural Biology* **17**(2): 209-214.

Rodríguez-Romero J, Franceschetti M, Bueno E, Sesma A. 2015. Multilayer Regulatory Mechanisms Control Cleavage Factor I Proteins in Filamentous Fungi. *Nucleic Acids Research* **43**: 179-195.

Rouble AN, Biggar KK, Storey KB. 2014. A high-throughput protocol for message RNA quantification using RNA dot-blots. *Analytical Biochemistry* **452**: 31-33.

Ryder LS, Talbot NJ. 2015. Regulation of appressorium development in pathogenic fungi. *Current Opinion in Plant Biology* **26**: 8-13.

Sanchez E, Asano K, Sone T. 2011. Characterization of Inago1 and Inago2 retrotransposons in *Magnaporthe oryzae*. *Journal of General Plant Pathology* **77**(4): 239-242.

Sesma A. 2016. RNA metabolism and regulation of virulence programs in fungi. *Seminars in Cell & Developmental Biology* **57**: 120–127.

Shalem O, Carey L, Zeevi D, Sharon E, Keren L, Weinberger A, Dahan O, Pilpel Y, Segal E. 2013. Measurements of the Impact of 3 ' End Sequences on Gene Expression Reveal Wide Range and Sequence Dependent Effects. *Plos Computational Biology* **9**(3).

Shi Y, Manley JL. 2015. The end of the message: multiple protein-RNA interactions define the mRNA polyadenylation site. *Genes & Development* **29**(9): 889-897.

Shi YS, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR, Frank J, Manley JL. 2009. Molecular Architecture of the Human Pre-mRNA 3 ' Processing Complex. *Molecular Cell* **33**(3): 365-376.

Slomovic S, Laufer D, Geiger D, Schuster G. 2006. Polyadenylation of ribosomal RNA in human cells. *Nucleic Acids Research* **34**(10): 2966-2975.

Sonenberg N, Hinnebusch AG. 2009. Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. *Cell* **136**(4): 731-745.

Spriggs KA, Bushell M, Willis AE. 2010. Translational regulation of gene expression during conditions of cell stress. *Molecular Cell* **40**(2): 228-237.

Tian B, Manley JL. 2017. Alternative polyadenylation of mRNA precursors. *Nature Reviews Molecular Cell Biology* **18**(1): 18-30.

Tucker SL, Besi MI, Galhano R, Franceschetti M, Goetz S, Lenhert S, Osbourn A, Sesma A. 2010. Common Genetic Pathways Regulate Organ-Specific Infection-Related Development in the Rice Blast Fungus. *Plant Cell* **22**(3): 953-972.

Urban M, Cuzick A, Rutherford K, Irvine A, Pedro H, Pant R, Sadanadan V, Khamari L, Billal S, Mohanty S, et al. 2017. PHI-base: a new interface and further additions for the multi-species pathogen-host interactions database. *Nucleic Acids Research* **45**(D1): D604-D610.

Ustyantsev IG, Golubchikova JS, Borodulina OR, Kramerov DA. 2017. Canonical and noncanonical RNA polyadenylation. *Molecular Biology* **51**(2): 226-236.

Venkataraman K, Brown KM, Gilmartin GM. 2005. Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. *Genes & Development* **19**(11): 1315-1327.

Weill L, Belloc E, Bava FA, Mendez R. 2012. Translational control by changes in poly(A) tail length: recycling mRNAs. *Nature Structural & Molecular Biology* **19**(6): 577-585.

Wilkening S, Pelechano V, Jarvelin AI, Tekkedil MM, Anders S, Benes V, Steinmetz LM. 2013. An efficient method for genome-wide polyadenylation site mapping and RNA quantification. *Nucleic Acids Research*: e65.

Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, Santos LBD, Bourne PE, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **3**.

Xu JR, Staiger CJ, Hamer JE. 1998. Inactivation of the mitogen-activated protein kinase Mps1 from the rice blast fungus prevents penetration of host cells but allows activation of plant defense responses. *Proceedings of the National Academy of Sciences, USA* **95**(21): 12713-12718.

Yan X, Talbot NJ. 2016. Investigating the cell biology of plant infection by the rice blast fungus *Magnaporthe oryzae*. *Current Opinion in Microbiology* **34**: 147-153.

Yang Q, Coseno M, Gilmartin GM, Doublie S. 2011. Crystal Structure of a Human Cleavage Factor CFI(m)25/CFI(m)68/RNA Complex Provides an Insight into Poly(A) Site Recognition and RNA Looping. *Structure* **19**(3): 368-377.

Yang Q, Doublie S. 2011. Structural biology of poly(A) site definition. *Wiley Interdisciplinary Reviews RNA* **2**(5): 732-747.

Zhao X, Xu JR. 2007. A highly conserved MAPK-docking site in Mst7 is essential for Pmk1 activation in *Magnaporthe grisea*. *Molecular Microbiology* **63**(3): 881-894.

FIGURE LEGENDS

Figure 1. Analysis of polyadenylation sites (PASs) in *M. oryzae* wild-type strain. A) ~52% of expressed genes undergo alternative polyadenylation (APA). Most APA events involve two PASs. **B)** FungiFun2 enrichment analysis of functional category (FunCat) terms. 1,654 genes exhibit APA and possess known functional domains. Only the top ten significant FunCat terms are listed. **C)** Nucleotide profile of *S. cerevisiae* and *M. oryzae* PAS regions. **D)** Top ten AU-rich element forms present in *M. oryzae* PASs. **E)** The AU-rich Polyadenylation Enhancer Element (PEE) recognized by yeast Hrp1 is not found in *M. oryzae*. **F**) Distribution of three novel elements in *M. oryzae* PAS region.

Figure 2. Carbon starvation dramatically affects *M. oryzae* gene expression and PAS selection. **A)** Number of genes with differential expression in wild-type (WT) and Δ*rbp35*. Carbon starvation (MM-C) presents the highest number of differentially expressed genes. **B)** Number of genes with differential PAS selection (DPS) between different growth conditions in WT. MM-C provides the highest number of genes with DPS. **C)** Distal PASs are generally preferred in MM-C compared to complete medium (CM) or minimal medium (MM). A positive fold-change means preference for distal PASs; a negative fold-change means preference for proximal PASs. **D)** The U-rich Upstream Element (UUE) is depleted in PASs preferred in MM-C, suggesting that RNA-binding proteins involved in the recognition of this element can be under differential regulation. **E)** The percentage of genes with differential expression and DPS (49%; p-value < 0.0005) is lower than the percentage of

genes expected to change expression from CM to MM-C (4,649 over 8,100; 57%). **F)** Number of genes differentially expressed between wild-type (WT) and Δ*rbp35*.

Figure 3. Rbp35 regulates PASs selection in ~39% of expressed genes. A) FungiFun2 enrichment analysis of functional category (FunCat) terms in the Δ*rbp35* mutant. Only the top ten significant FunCat terms are listed. **B)** Number of genes with DPS between WT and Δ*rbp35*. The mutant shows high levels of DPS. **C)** Proximal PASs are more frequent in Δ*rbp35* compared to WT. A positive fold-change means preference for distal PASs, a negative fold-change means preference for proximal PASs. **D)** The UGUAH motif at -55 nt is depleted in PASs preferred in Δ*rbp35*, suggesting that this element is the possible binding site for the Rbp35 protein. PASs preferred in the mutant also show an enrichment of the UGUAH motif, but closer to the cut site. **E)** Working model of Rbp35 in the selection of *M. oryzae* PASs. Three different modes of PAS selection are represented. 1) When Rbp35 is not involved in PAS selection, Rbp35 may or may not be a component of the polyadenylation machinery, and the UGUAH element is not present. In transcripts with multiple PASs, the WT shows no clear preference for which is selected. 2) In distal PASs with a UGUAH motif at -55nt, Rbp35 guides the polyadenylation machinery to a proper cleavage site. Proximal PASs usually have a UGUAH box that lies at approx. -25nt, and its recognition by Rbp35 inhibits the cleavage reaction by the polyadenylation machinery (in grey), possibly through occlusion of other binding sites. 3) When Rbp35 is not present, proximal PASs previously blocked by Rbp35 are recognised more efficiently, masking the cut of distal PASs by the polyadenylation machinery (in grey). Red columns show exemplar RNA-seq read-frequencies found for each mode of PAS selection. **F)** RNA-binding assays with Rbp35 and CFI25. Bottom panels: detection of biotinylated probes with streptavidin-horseradish peroxidase (strep-HRP). Top panels: same membranes probed with anti-His antibody, confirming the presence of proteins used in these assays (see also Figure S3E).Rbp35 recognises strongly 2x and 4x UGUAH probes while CFI25 enhances Rbp35 binding to UGUAH. A poly(C)$_{42}$ RNA homopolymer and BSA (bovine serum albumin) resuspended in binding buffer were used as negative controls. **G)** The AAUAHA motif, which includes the canonical polyadenylation signal AAUAAA is depleted in PASs preferred in Δ*rbp35* and in WT grown on MM-C.

Figure 4. Effect of 3'UTRs length in *M. oryzae*. A) Lack of carbon and Rbp35 alter PAS selection in *MST7*. (Left) Schematic representation of *MST7* mRNA. (Right) Extension poly(A) test (ePAT) reaction run in a 2% of agarose gel. The ePAT analyzes both poly(A) tail length and alternative PAS selection. Lack of carbon increases *MST7* mRNAs with long 3'UTRs in the wild-type (WT) but not

in Δ*rbp35*. The TVN-PAT reaction amplifies the 3'UTR without poly(A) tail and is used as a control. **B)** Poly(A) tail length profiles in WT, Δ*rbp35* and *S. cerevisiae* (*S.c.*). *M. oryzae* poly(A) tails are not regulated by Rbp35. **C)** (Top) Schematic representation of the experiment. (Bottom) Dot-blot of total RNA probed with biotinylated oligo(dT). Δ*rbp35* mRNAs remained stable at longer thiolutin exposition times compared to the WT. **D)** Estimation of mRNA half-lives of *MST7*, *14-3-3A* and *14-3-3B* after transcriptional inhibition with thiolutin. mRNA half-life is represented by the equation t½ =0.693/k, where k is the constant rate for mRNA decay (i.e., percentage change over time). After addition of thiolutin (6 µg/ml), CDS and longest 3'UTRs, which contain short and long (S/L) or only long (L) transcripts, respectively, were amplified by qPCR (n=3; SD). **E)** Polysomal profiling of WT and Δ*rbp35* in CM and MM-C. MW: molecular weight. **F)** qPCRs of *MST7*, *14-3-3A* and *14-3-3B* using mRNAs from polysomal fractions 1-8 (n=3; error bars represent SD; letters indicate significance at $p < 0.05$ using Fisher's ANOVA, degrees of freedom=5, F=4.88).

Figure 5. Characterization of fungal 3' UTRs during infection. A) Box-plot analysis. The relative expression of PASs at the spore stage is more similar to that observed in MM-C, while PASs expression during leaf infection most resembles that of MM-N. **B)** 3'UTR lengths vary during plant infection. (Left) 3'UTR change is showed by displaying the differential posterior distribution of the estimated binomial proportion between the considered infection point and the spore stage. Each dot indicates the mean of the posterior distribution while the vertical lines display 89% credible intervals. At the beginning of root infection longer 3'UTRs prevail whereas a global 3'UTR shortening occurs at later stages. On leaves, a relatively high number of transcripts with long 3'UTRs are observed during all stages of the infection process. (Right) Graph representing the percentages derived from this analysis. **C)** qPCRs from infected rice leaves and roots using primers that amplified CDS or longest 3'UTR, which contain short and long (S/L) or only long (L) transcripts, respectively. The abundance of long 3'UTR isoforms varies during rice infection (hpi, hours post-inoculation).

Figure 6. Relevance of long *14-3-3B* mRNA isoform for plant colonization. A) Infection experiments using wild-type (WT) strain and the Δ*14-3-3b* mutant complemented either with the full length sequence of *14-3-3B*, which contains short and long 3'UTRs (short and long, *S/L*) or with a short 3'UTR (*S*). *14-3-3B-S* is unable to restore Δ*14-3-3b* disease symptoms. **B)** qPCR analysis of *14-3-3B* mRNAs (n=3; SD), and **C)** Western blotting of proteins, extracted from mycelia grown on CM or MM-C. Comassie stain and rpS6 antibody (Ab) were used as loading controls. *14-3-3B-S* has

lost its regulation in MM-C. **D)** Confocal analysis of 14-3-3b:mCherry. These images were taking with identical confocal settings.

Figure 7. Regulation of APA during infection in the rice blast fungus. Summary diagram of the results derived from our genome-wide studies. Infection initiates with spores landing on plant surfaces. This stage leads to a preference for long 3' UTRs of mRNAs, possibly due to a low availability of carbon sources. Leaf penetration requires well-established sequential steps (adhesion, germination and appressorium formation). Throughout this process carbon-rich molecules are being released from leaf cell walls and a change in APA preference is observed from long to short 3' UTRs. After penetration occurs (4-8 h post-inoculation, hpi), the fungus stays in the first infected cell up to 36-48 hpi. During this time, the fungus reprograms its metabolism to initiate plant invasive growth, and a highly regulated molecular switch occurs, including the modification of 3'UTR lengths. Between 48 and 120 hpi, the fungus invades rice cells where it founds a significant amount of nutrients, which leads to a global preference for short 3'UTRs.