

Landscape and regulation of alternative splicing and alternative polyadenylation in a plant pathogenic fungus

Ping Lu¹, Daipeng Chen^{1,2}, Zhaomei Qi¹, Haoming Wang¹, Yitong Chen¹, Qinhu Wang¹ , Cong Jiang¹ , Jin-Rong Xu²  and Huiquan Liu¹ 

¹State Key Laboratory of Crop Stress Biology for Arid Areas, College of Plant Protection, Northwest A&F University, Yangling, Shaanxi, 712100, China; ²Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN, 47907, USA

Author for correspondence:
Huiquan Liu
Email: liuhuiquan@nwsuaf.edu.cn

Received: 28 November 2021
Accepted: 30 March 2022

New Phytologist (2022) 235: 674–689
doi: 10.1111/nph.18164

Key words: alternative polyadenylation, alternative splicing, *Fusarium*, Iso-Seq, nonsense-mediated mRNA decay (NMD), RNA15, RNA-Seq.

Summary

- Alternative splicing (AS) and alternative polyadenylation (APA) contribute significantly to the regulation of gene expression in higher eukaryotes. Their biological impact in filamentous fungi, however, is largely unknown.
- Here we combine PacBio Isoform-Sequencing and strand-specific RNA-sequencing of multiple tissues and mutant characterization to reveal the landscape and regulation of AS and APA in *Fusarium graminearum*.
- We generated a transcript annotation comprising 51 617 isoforms from 17 189 genes. In total, 4997 and 11 133 genes are alternatively spliced and polyadenylated, respectively. Majority of the AS events alter coding sequences. Unexpectedly, the AS transcripts containing premature-termination codons are not sensitive to nonsense-mediated messenger RNA decay. Unlike in yeasts and animals, distal APA sites have strong signals, but proximal APA isoforms are highly expressed in *F. graminearum*. The 3'-end processing factors FgRNA15, FgHRP1, and FgFIP1 play roles in promoting proximal APA site usage and intron splicing. A genome-wide increase in intron inclusion and distal APA site usage and downregulation of the spliceosomal and 3'-end processing factors were observed in older and quiescent tissues, indicating intron inclusion and 3'-untranslated region lengthening as novel mechanisms in regulating aging and dormancy in fungi.
- This study provides new insights into the complexity and regulation of AS and APA in filamentous fungi.

Introduction

Alternative splicing (AS) and alternative polyadenylation (APA) are ubiquitous post-transcriptional regulatory processes in eukaryotes, enabling different messenger RNA (mRNA) isoforms to be generated from the same gene. Recent studies have revealed that AS and APA contribute significantly to enhance functional diversity and regulate gene expression in multiple ways in plants and animals (Tian & Manley, 2017; Chaudhary *et al.*, 2019). In fungi, however, the biological impact of AS and APA has not been well studied except in the budding yeast *Saccharomyces cerevisiae*, which has only *c.* 300 intron-containing genes and a simple saprophytic life cycle without elaborate asexual or sexual structures/fruitlet bodies. Filamentous fungi encompass an enormous diversity of species with varied morphologies, ecologies, and life cycle strategies. AS and APA appear to contribute to phenotypic complexity, pathogenicity, and ecological adaptation of these fungi (Rodriguez-Romero *et al.*, 2019; Muzafar *et al.*, 2021).

Although AS events have been analyzed by Illumina short-read RNA-sequencing (RNA-Seq) in *Trichoderma longibrachiatum*,

Verticillium dahliae, and a few other species (Zhao *et al.*, 2013; Xie *et al.*, 2015; Gehrmann *et al.*, 2016; Dong *et al.*, 2017; Jin *et al.*, 2017; Ibrahim *et al.*, 2021), a comprehensive analysis on splice isoforms in filamentous fungi is lacking. Besides short-read RNA-Seq data being unsuitable for accurately reconstructing full-length splice isoforms, most of the published fungal RNA-Seq data are not strand-specific, and overlapping transcripts between adjacent genes is widespread in fungi due to high gene density, which makes transcript assembly problematic. Knowledge of full-length transcript isoforms is necessary to deduce encoded proteins and assess the roles of splice isoforms in gene regulation. Isoform-sequencing (Iso-Seq) with PacBio single molecular real-time (SMRT) long-read technology offers a considerable advantage in characterizing transcriptome-wide full-length splice isoforms and post-transcriptional regulatory events without assembly. This approach has been widely used in plant and animal studies but rarely used for filamentous fungi (www.pacb.com/applications/rna-sequencing/rna-sequencing-for-plant-and-animal-sciences/).

The filamentous ascomycete fungus *Fusarium graminearum* is the predominant causal agent of Fusarium head blight (FHB),

one of the most devastating diseases on cereal crops worldwide. Given its scientific/economic importance, *F. graminearum* has been listed as fourth out of the top 10 most important fungal pathogens (Dean *et al.*, 2012). The ascospores discharged from perithecia (sexual fruiting bodies) are the primary inoculum of FHB. *Fusarium graminearum* is homothallic and produces abundant perithecia relatively synchronously under laboratory conditions, making it an ideal system for studying sexual development. Recently, A-to-I mRNA editing, a novel fungal epigenetic phenomenon, was discovered to specifically occur during its sexual stage (Liu *et al.*, 2016; Bian *et al.*, 2019). The relatively high rate of homologous recombination and ability to easily obtain homokaryotic transformants have enabled efficient and productive molecular studies in *F. graminearum*. To date, > 2000 genes have been functionally characterized by targeted deletion or disruption in *F. graminearum*, including large-scale gene knockout projects (Son *et al.*, 2011; Wang *et al.*, 2011; Yun *et al.*, 2015; Shin *et al.*, 2017; Jiang *et al.*, 2019, 2020).

Comparative and functional genomic studies rely on accurate genome assembly and annotation. *Fusarium graminearum* is one of the earliest fungal plant pathogens with a sequenced genome (Cuomo *et al.*, 2007). Owing to the small genome size and low repetitive sequence content, the genome assembly of *F. graminearum* strain PH-1 is of extremely high quality (King *et al.*, 2015). Nevertheless, errors in the reference genome and annotation were observed (Liu *et al.*, 2016). Moreover, despite improvements in the annotation by multiple revisions (King *et al.*, 2017), currently available gene models of *F. graminearum* are derived primarily from computational prediction, which is incomplete or sometimes inaccurate. Most gene models contain only coding sequences (CDSs), missing 5'- and/or 3'-untranslated regions (UTRs). No splice isoform information is available. Furthermore, although transcriptomics data have been accumulated to determine gene expression profiles during different developmental stages of *F. graminearum* (Dash *et al.*, 2012; Kazan & Gardiner, 2018; Brauer *et al.*, 2020), our understanding of the landscape and regulation of its transcriptome, especially in terms of splice isoforms and APA, is still lacking.

Here, we improved the reference genome and annotation of *F. graminearum* by using PacBio long-read sequencing and presented the first comprehensive analysis of a full-length transcriptome in a plant pathogenic fungus. We characterized the landscape of AS and APA and revealed their regulation in different tissues based on Iso-Seq and strand-specific RNA-Seq data together with mutant characterization.

Materials and Methods

Fungal materials and growth conditions

The *F. graminearum* strain PH-1 (Cuomo *et al.*, 2007) maintained in our laboratory (PH-1_YL) and its mutants generated in this study were routinely cultured on potato dextrose agar (PDA) plates at 25°C. The PH-1 laboratory stocks PH-1_MSU and PH-1_ZJU were kindly provided by Dr Frances Trail at Michigan State University (East Lansing, MI, USA) and Dr Yun Chen

at Zhejiang University (Hangzhou, China). Fungal tissues at different developmental stages or conditions were harvested as described (Liu *et al.*, 2015, 2016). Briefly, perithecia were harvested from carrot agar (CA) plates at 3-, 6-, 7- or 8-d post-fertilization (dpf). Conidiation samples were collected from 5-d-old liquid carboxymethyl cellulose (CMC) cultures. Conidia were collected by filtering 5-d-old CMC cultures through sterile glass wool. Germlings and vegetative hyphae were collected from conidia incubated for 12, 24, 36, and 48 h in liquid YEPD (1% yeast extract, 2% peptone, 2% glucose). Aerial mycelia were collected from 5-d-old PDA and CA plates. DON-producing hyphae were collected from 3-d-old liquid trichothecene biosynthesis induction (TBI) cultures supplied with 5 mM arginine or ammonium nitrate (NH₄NO₃). The inoculated spikelets of flowering wheat heads of cultivar Xiaoyan22 were collected 3 d after inoculation (dpi) with strain PH-1.

Generation of gene deletion mutants

The split-marker approach was used to generate the gene replacement constructs. The flanking sequences of individual genes were amplified and connected to the hygromycin phosphotransferase cassette by overlapping PCR. Protoplasts of the PH-1_YL1 strain were prepared and transformed with each gene replacement construct as described previously (Liu *et al.*, 2015). For transformant selection, hygromycin B (Calbiochem, La Jolla, CA, USA) was added to the final concentration of 250 µg ml⁻¹. Gene deletion mutants were confirmed by PCR assays. At least two independent deletion mutants were obtained for each gene.

PacBio and Illumina DNA library preparation and sequencing

Genomic DNA was isolated from 24 h vegetative hyphae by the CTAB method. The quality of DNA was evaluated by agarose gel electrophoresis and BioAnalyzer-2100 (Agilent, Santa Clara, CA, USA). The PacBio library was constructed with the SMRTbell™ Template Prep Kit according to manufacturer's instructions, and sequenced on the PacBio Sequel System. The Illumina library was prepared with the NEBNext® Ultra™ DNA Library Prep Kit for Illumina® (San Diego, CA, USA) following the manufacturer's instructions and sequenced on the Illumina HiSeq® 2500 System, with a 2 × 150 bp paired-end read mode.

PacBio and Illumina RNA library preparation and sequencing

Total RNA was extracted with the RNeasy Pure Plant Kit (Qiagen Biotech, Beijing, China). Poly(A)⁺ mRNA was enriched with Iso-Seq Express Oligo Kit (PN 101-737-500). The quality of DNA was evaluated by agarose gel electrophoresis and BioAnalyzer-2100 (Agilent). The PacBio Iso-Seq library was prepared according to the Iso-Seq™ Template Preparation for Sequel™ Systems protocol. The Iso-Seq libraries were sequenced on the PacBio RS II or Sequel System. Illumina strand-specific RNA-Seq libraries were prepared with the NEBNext® Ultra™

Directional RNA Library Prep Kit following the manufacturer's instructions, and sequenced on the Illumina HiSeq[®] 2500 system with the 2 × 150 bp paired-end read mode. For size-selected Iso-Seq, the amplified complementary DNA (cDNA) product was size selected with the BluePippin before library construction. For ribosomal RNA-depleted RNA-Seq, the total RNA of each sample was subjected to ribosomal RNA depletion as described (Liu *et al.*, 2016).

Reference genome correction

The public Rothamsted Research genome assembly (RR1, release-35) of strain PH-1 was downloaded from Ensembl Fungi. The *de novo* assembly of PacBio subreads was generated with CANU v.1.7 (Koren *et al.*, 2017) using default parameters and polished with ARROW from SMRT link. The PacBio genome assembly was aligned to the RR1 assembly using MUMMER4 (Marcais *et al.*, 2018), and different sites/regions between them were identified. The PacBio subreads and the Illumina reads of three PH-1 laboratory stocks were mapped onto the two assemblies by GMAP (Wu & Watanabe, 2005) and BOWTIE2 (Langmead & Salzberg, 2012), respectively. Each site/region that differed between the two assemblies were checked manually by using IGV (<https://software.broadinstitute.org/software/igv/>). Only the errors evidenced by at least two sequenced PH-1 laboratory stocks were corrected in the RR1 assembly.

Iso-Seq analysis pipeline

Our Iso-Seq analysis pipeline consisted of six steps (Supporting Information Fig. S1).

Step 1: Generating polished consensus transcripts The *pbtranscript* tool from SMRT link was used to extract the reads of insert from raw Iso-Seq datasets, classify them into full-length, nonartificial-concatemer (FLNC) and non-FLNC reads based on the location of primers and poly(A) tails, and then the FLNC reads were clustered and polished into consensus transcripts.

Step 2: Correcting indels, mismatches and splice junction The RNA-Seq data from the same RNA as Iso-Seq was used to generate the splicing junction annotation using STAR (Dobin *et al.*, 2013). The consensus transcripts were repaired using TRANSCRIPTCLEAN (Wyman & Mortazavi, 2019) according to the genome sequence and splicing junction annotation.

Step 3: Filtering fusion and polycistronic transcripts Because the presence of fusion or polycistronic transcripts can disturb the downstream analysis, they were detected and filtered out from the corrected consensus transcripts. Putative fusion transcripts were identified using the Cupcake script *fusion_finder.py*. Polycistronic transcripts were detected as transcripts containing two or more nonoverlapping open reading frames (ORFs) (≥ 100 aa) that overlap over 50% ORFs of each adjacent genes in same strand.

Step 4: Filtering artificial-concatemer transcripts and correcting wrong-stranded transcripts The consensus transcripts were aligned to the reference genome by BLAT (Kent, 2002), and the artificial-concatemer transcripts were identified and filtered out according to their mapping characteristics. The consensus transcripts were aligned to the reference genome and converted to GFF format by GMAP. Transcripts that shared the same splicing junction with the annotated genes but aligned to the opposite strand were identified using GFFCOMPARE (Pertea *et al.*, 2016) and corrected for their strand information.

Step 5: generating unique transcript isoforms The filtered and corrected consensus transcripts were aligned to the reference genome using GMAP. Redundant transcripts were collapsed into unique transcript isoforms by the Cupcake script *collapse_isoforms_by_sam.py*. To minimize inclusion of possible 5' truncated transcripts, transcript isoforms differing only in the 5' start of their first exon were collapsed to keep only the longest ones. Genes in the transcript isoform annotation that have only one single-exon transcript with a short length (≤ 200 nt) and low expression (≤ 1 TPM (transcripts per kilobase million)) were removed.

Step 6: Correcting unique transcript isoforms Overlapping transcripts from adjacent genes in the same strand were incorrectly grouped as one gene by Cupcake script. To correct these issues, the genes from transcript isoform annotation that overlap with two or more RR1 genes were detected by GFFCOMPARE and corrected manually with the aid of IGV.

RNA-Seq analysis

RNA-Seq data were aligned to the reference genome using HISAT2 (Pertea *et al.*, 2016). Transcript assembly and quantification of gene/transcript expression (TPM) were performed using STRINGTIE (Pertea *et al.*, 2016). Differential expression was analyzed with DESEQ2 (Love *et al.*, 2014) based on transcript abundance calculated by Salmon (Patro *et al.*, 2017). Transcripts with an adjusted *P*-value < 0.05 and $|\log_2 \text{fold change}| \geq 1$ were considered to be differentially expressed.

Identification and analysis of AS events

AS landscape in the transcript annotation were extracted using ASTALAVISTA (Foissac & Sammeth, 2007). According to the AS code assigned, the AS events were categorized as four basic types with '(n)^(n+1)~' for intron retention, '(n)^(n+1)~' for alternative 5'-donor, '(n)~(n+1)~' for alternative 3'-acceptor, and '(n)~(n+1)~' for exon skipping. Differential AS events among samples were identified by CASH v.2.2.1 (Wu *et al.*, 2018) based on the YL1 annotation. AS events with a false discovery rate (FDR) < 0.05 were regarded as differential AS events.

Nonsense-mediated mRNA decay (NMD) target prediction

The position and distance of the stop codon relative to the terminal exon-exon junction were determined for each AS transcript.

AS transcripts with a stop codon > 50-bp upstream of the terminal exon-exon junction were considered as putative nonsense-mediated mRNA decay (NMD) targets (Garcia-Moreno & Romao, 2020). The premature-termination codon (PTC)-containing NMD targets were identified by comparing the stop codon position of the predicted NMD target with that of the corresponding canonical transcript.

APA identification and analysis

The FLNC reads were aligned to the reference genome and used to identify the unique polyadenylation sites (PASs). To address the internal priming issue, the PASs with AAAAAA in the upstream 10 nt and downstream 20 nt or with more than 7 As in 10 nt sliding windows were discarded. For each of the PASs, the number of supporting FLNC reads were counted. Separating APA from heterogeneous cleavage was performed with a similar method as described (Liu *et al.*, 2017). SIGNALSLEUTH2 (Zhao *et al.*, 2014) and MEME-CHIP (Machanick & Bailey, 2011) were used to identify enriched 4-mer to 6-mer motifs in flanking sequences (−100 nt to 100 nt) around the PASs. Detecting APA and identifying differential APA usage from RNA-Seq alignments were performed with the *roar* (Grassi *et al.*, 2016). PAS with a *P*-value < 0.05 was considered to be differential APA usage.

Long noncoding RNA (lncRNA) identification

Novel transcripts from Iso-Seq transcript isoform sets that did not share any splice junctions with RR1 genes were used to predict long noncoding RNAs (lncRNAs). We used PLEK (Li *et al.*, 2014) to distinguish lncRNAs from protein-coding RNAs. The transcripts with a low coding potential were further scanned against the Pfam and Rfam databases to filter out transcripts encoding protein domains and/or harboring any known structural RNA motifs (*E* value < 10¹⁰).

Reverse transcription polymerase chain reaction (RT-PCR)

Primers used for PCR validation of AS events were designed to span the splicing events using PRIMER PREMIER 5. RNA samples were isolated with the TRIzol reagent (Invitrogen). The ReverAid First cDNA synthesis kit (Thermo Fisher Scientific, Waltham, MA, USA) was used for cDNA synthesis. PCR products were purified and sequenced by the Sanger method.

Statistical analysis

All statistical tests were performed using R. Hierarchical clustering was performed using R with average linkage cluster method and Pearson correlation coefficient distance. Gene ontology (GO) enrichment analysis was performed using BLAST2GO (www.blast2go.com) with the Fisher's Exact Test and Benjamini–Hochberg correction (FDR < 0.05). Only the most specific GO terms were retained.

Results

Revising the reference genome of PH-1

To improve the reference genome of *F. graminearum*, we generated > 300× long-reads for PH-1_YL1 with PacBio sequencing and > 90× short-reads for each of the three PH-1 laboratory stocks (YL, ZJU, and MSU) with Illumina sequencing (Tables S1, S2). The PacBio long-reads were *de novo* assembled into 10 contigs (Fig. 1a). By comparing the PacBio assembly with the most recent public assembly RR1 of PH-1, a total of 315 different regions/sites were identified. The long-reads and short-reads were aligned to the two assemblies, respectively. Based on the alignments, seven mis-assemblies, 200 base errors, and 42 InDel errors in the RR1 assembly (Fig. 1a–c; Table S3), which impacts the annotation of 74 protein-coding genes, were evidenced by at least two PH-1 laboratory stocks. We corrected all the errors in the RR1 assembly and generated an updated version of PH-1 assembly (named YL1).

Generating a comprehensive transcript annotation

To identify as many transcripts as possible, three size-selected (1–2 kb, 2–3 kb, and 3–6 kb) and one nonsize selected Iso-Seq libraries were constructed from combined equal amounts of poly (A)⁺ RNA purified from six tissues (Fig. 1d; Table S1). We also performed nonsize selected Iso-Seq for 24-h hyphae and 6-dpf perithecia with independent biological replicates. In total, 5423 051 reads of insert were obtained. We developed a computational pipeline (Fig. S1) and used it to obtain 47 589 high-quality unique transcript isoforms representing 13 712 genes.

Over 60% (8391) of the genes had two or more transcript isoforms (Fig. 1e). Remarkably, FG1G24420 had > 200 transcript isoforms. The Iso-Seq transcript isoforms (median 2341 nt) were 160% longer than the predicted RR1 transcripts (median 1462 nt) (Fig. 1f), which was mainly due to increases in the length of UTRs. In total, 63.4% (30 183) of the Iso-Seq transcript isoforms were not present in or different from RR1 transcripts. Among them, 5786 isoforms (12.2%) represented novel genes either from novel loci (2689) or overlapping with existing gene loci on the opposite strand (3097), and 16 022 isoforms (33.6%) were possible novel isoforms of existing genes that share at least one splice site with RR1 transcripts but differ at other splice sites (13 863) or overlapping with the RR1 transcripts but with different splice sites (2159).

There were 3187 genes without corresponding Iso-Seq transcripts. Those genes were generally expressed at lower levels (Fig. S2). To obtain their transcript annotations, we performed strand-specific RNA-Seq for each of the six tissues (Table S2), and obtained 1003 assembled transcripts for 882 RR1 genes without Iso-Seq transcripts and 300 assembled transcripts from novel gene loci. By combining the transcripts from both Iso-Seq and strand-specific RNA-Seq with the remaining 2305 RR1 genes without transcripts available, we generated a comprehensive reference annotation (named YL1) for PH-1 that has 51 617 transcript isoforms from 17 189 genes, including 5481 lncRNAs and 498

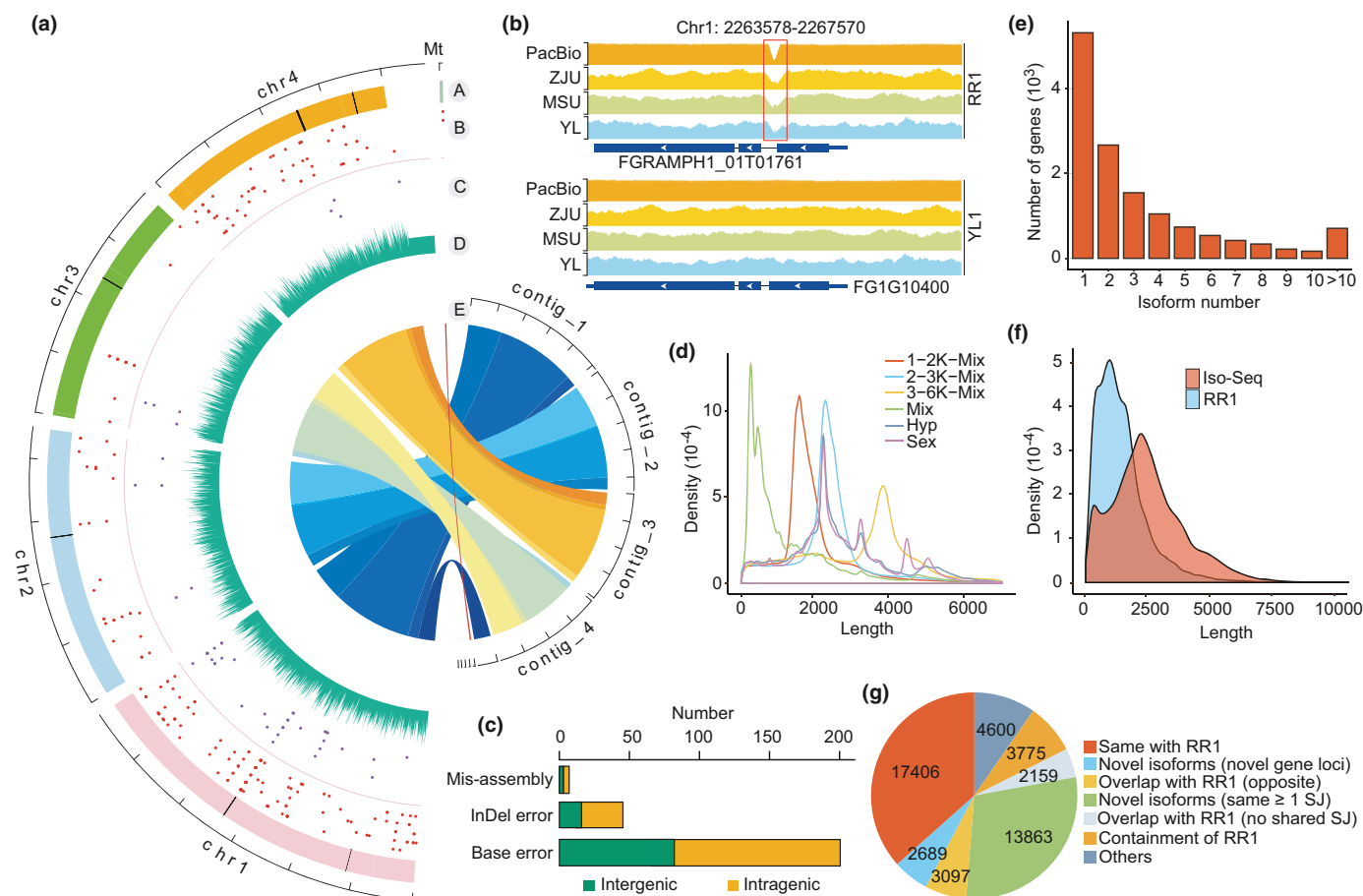
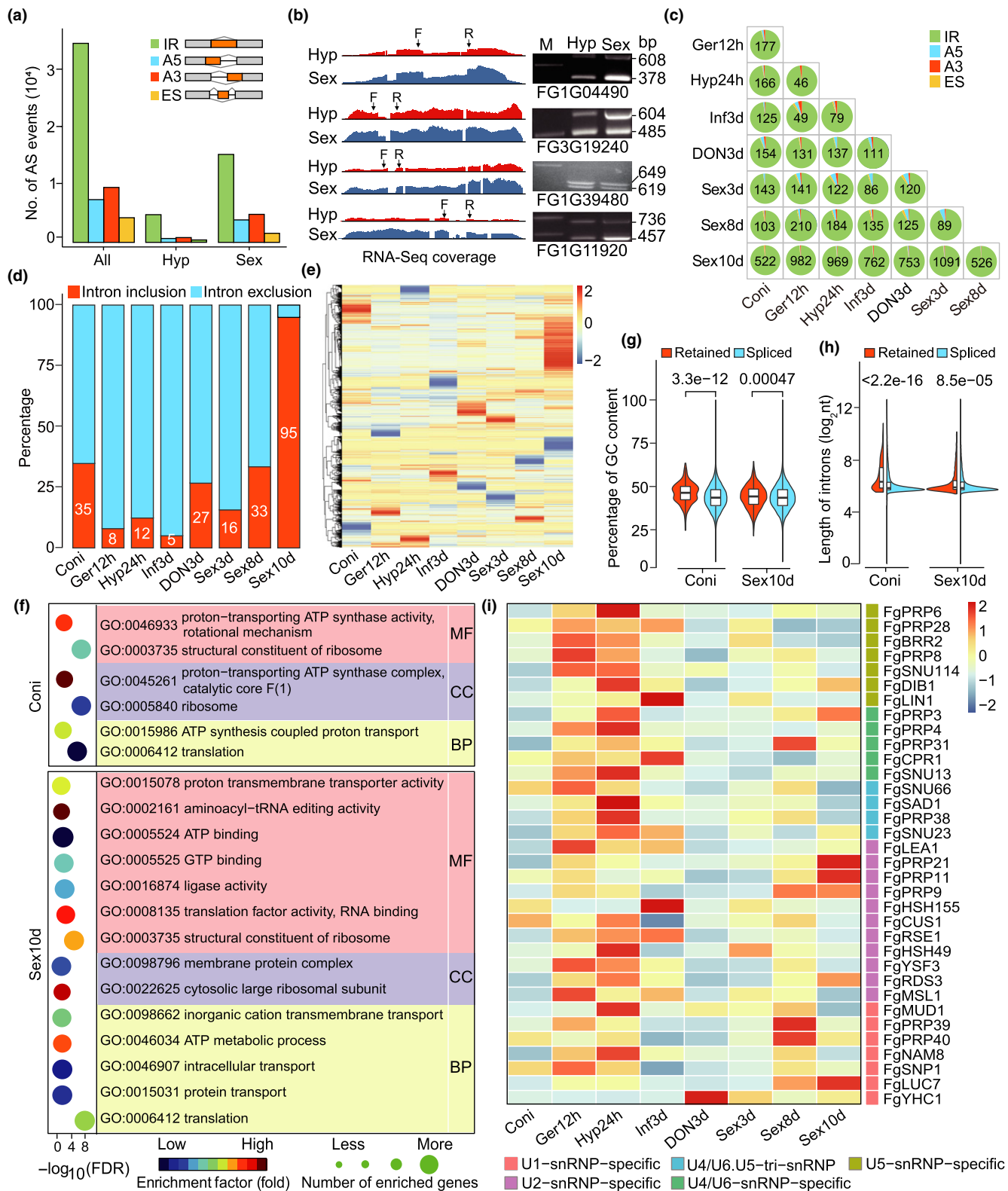


Fig. 1 Improvement of the genome sequence of *Fusarium graminearum* and its transcript annotation by PacBio single molecular real-time sequencing. (a) Comparison of PacBio and Rothamsted Research (RR1, Ensembl Fungi) assemblies of *F. graminearum* PH-1 strain. a: Ideograms of RR1 chromosomes (chr1–chr4) with mis-assembly regions indicated by black bars; b: Base error positions (red dots) in RR1 assembly; c: InDel error positions (purple dots) in RR1 assembly; d: PacBio read coverage in 10 kb bin; e: Syntenic blocks between PacBio and RR1 assemblies connected with ribbons. (b) An example of mis-assembly errors in the RR1 assembly. Coverage of Illumina reads from three PH-1 laboratory stocks (Zhejiang University (ZJU), Michigan State University (MSU), and Yangling (YL)) and PacBio reads from YL stock at the marked locus were mapped to the RR1 and YL1 assemblies. Red box highlights the mis-assembly region. (c) Number of corrected genomic errors located in the intragenic or intergenic regions. (d) Subread length distribution of different Isoform-sequencing (Iso-Seq) libraries. (e) Distribution of the number of transcript isoforms per gene. (f) Comparison of transcript length distributions between the RR1 annotations and Iso-Seq transcript isoforms. (g) Classification of Iso-Seq transcript isoforms.

novel protein-coding genes (Table S4). The YL1 annotation was adopted in a new gene naming scheme (e.g. FG1G00010), and associated with legacy annotation gene identifiers (IDs). We

created a *F. graminearum* genome database (FgBase) (fgbase.wheatcab.com) that allows users to browse, search, and download the updated genome and transcript annotation.

Fig. 2 Landscape and regulation of alternative splicing (AS) in *Fusarium graminearum*. (a) Distribution of four types of AS events, including intron retention (IR), alternative 5'-donor (A5), alternative 3'-acceptor (A3), and exon skipping (ES) in transcripts of all samples (All) as well as vegetative hyphae (Hyp) and sexual perithecia (Sex) samples. (b) Reverse transcription polymerase chain reaction (RT-PCR) validation of the four types of AS events in the marked loci with RNA isolated from hyphae (Hyp) and perithecia (Sex). The expected size for each band is indicated. M, DNA markers. RNA sequencing (RNA-Seq) read coverage of the four associated genes is shown in left panel. PCR primers (F, forward and R, reverse) are designed to flank the splicing events. (c) Number of significantly differential AS events between pairs of marked samples. Pie charts show the proportion of differential IR, A5, A3, and ES events. (d) Proportion of intron inclusion and intron exclusion in the nonredundant differential IR events relative to each of the marked samples. (e) Heatmap of Percent Spliced-In (PSI) values across different samples. High (orange to red) and low (yellow to blue) PSI values are depicted as Z-scores for each AS event. (f) Enriched gene ontology (GO) terms in genes with differentially retained introns in conidia (Coni) or 10-d post-fertilization (dpf) perithecia (Sex10d). MF, molecular function; CC, cell component; BP, biological process. (g, h) Violin plot comparison of GC contents (g) and intron lengths (h) between the spliced introns and differentially retained introns in conidia and 10-dpf perithecia. Horizontal line within the rectangle represents the median of values and whisker indicates the largest/lowest point inside the range defined by 1.5 times interquartile range (IQR). The *P* values are from two-tailed Wilcoxon rank sum test. (i) Heatmap of the expression of putative genes encoding spliceosomal components in different samples. High (orange to red) and low (yellow to blue) expression levels are depicted as Z-scores for each gene. Subunits of the five canonical snRNP subcomplexes are indicated with different colors on the right. For (c–i), see Supporting Information Table S2 for details of sample information.



Landscape of AS

Based on the new YL1 annotation, we performed a systematic characterization of AS in *F. graminearum*. A total of 54 613 AS events were identified in the YL1 annotation, which affected

42.7% (4997) of intron-containing genes and resulted in 17 229 splice isoforms. The AS genes were highly enriched for diverse biological processes (Fig. S3). There were 25 069 AS events in perithecia but nearly four times fewer (6552) in hyphae (Fig. 2a). Intron retention (IR) comprised the majority (61.9%) of AS

events. The relative proportion of non-IR events was increased in the perithecia. Additionally, 1960 genes had isoforms resulting from multiple combinatory AS events, suggesting that more complex AS events also occurred frequently in *F. graminearum*. We randomly selected four examples that showed each type of AS events for reverse transcription polymerase chain reaction (RT-PCR) (Table S5). For all the examples, we observed bands of expected sizes on gels (Fig. 2b).

A total of 14 080 transcript isoforms (52.8%) associated with 3291 genes were identified only in perithecia or hyphae (Fig. S4). Perithecia had the higher proportion of stage-specific transcript isoforms (40.7%). Remarkably, FG1G24420 generated 197 transcript isoforms in perithecia but only one in hyphae. Genes with perithecia-specific transcript isoforms were significantly enriched for regulation of transcription, oxidation-reduction, and transmembrane transport processes, while genes with hyphae-specific transcript isoforms were significantly enriched for RNA processing, methylation, and RNA modification processes (Fig. S4).

Differential AS events among different tissues

To examine the dynamics of AS events across different tissues, we calculated a Percent Spliced-In (PSI) index using SUPPA2 (Trincado *et al.*, 2018) for each AS event. PSI index is calculated as the fraction of the inclusion reads to the total reads (both inclusion and exclusion reads) to measure the inclusion level of a given splicing event. Hierarchical clustering revealed that the PSI values were variable in different samples (Fig. S4), suggesting that the AS events are subject to tissue-specific controls.

To accurately identify differential AS events among different tissues, we generated replicate strand-specific RNA-Seq data from conidia (Coni), 12-h conidial germlings (Ger12h), 24-h vegetative hyphae (Hyp24h), 3-dpi infection hyphae (Inf3d), 3-day-old DON production hyphae in TBI culture supplemented with NH_4NO_3 (DON3d), and three sexual stages (Sex3d, Sex8d, and Sex10d) (Table S2). Pairwise comparisons revealed that the number of differential AS events ranged from 46 to 1091 (Fig. 2c). Totally, the nonredundant differential AS events relative to each sample ranged from 844 to 1996 (Fig. S4). Sex10d had the greatest number of differential AS events, suggesting a distinct AS landscape in the later stage of sexual reproduction.

A global increase in intron inclusion in aging or dormant tissues

Over 95% of the detected differential AS events were IR (Figs 2c, S4), indicating that IR is more likely to be regulated across different tissues. The differential IR events were further subdivided into two opposite types: intron inclusion and intron exclusion. Interestingly, in Ger12h, Hyp24h and Inf3d, < 12% of the differential IR events were intron inclusion, whereas the fraction of intron inclusion was over 33% in Coni, Sex8d, and Sex10d (Fig. 2d). During sexual reproduction, intron inclusion continuously increased from Sex3d to Sex10d. In Sex10d, 95% of the differential IR events were intron inclusion (Fig. 2d). Consistent with these observations, the PSI values of most AS events were

higher in Sex10d but lower in Ger12h, Hyp24h and Inf3d (Fig. 2e). In comparison with Ger12h, Hyp24h and Inf3d that have active growth, Sex10d and Coni were representative of aging and dormant states, suggesting that increasing abundance of intron-retained isoforms is likely associated with those states.

Genes with increased intron inclusion in Coni were enriched for functions associated with ATP synthesis coupled proton transport and translation, while those in Sex10d were enriched for functions associated with ATP metabolic process, translation, inorganic cation transmembrane transport, intracellular transport, and protein transport (Fig. 2f). The retained introns that increased in Coni and Sex10d were generally longer and had higher GC contents compared to the spliced introns (Fig. 2g,h). For most of the genes encoding spliceosomal components, their expression levels were generally higher in the active tissues but lower in the old or quiescent tissues (Fig. 2i), indicating that depressed expression of spliceosomal genes may be partially responsible for increased intron inclusion in those tissues.

Majority of the AS transcript isoforms encoded altered ORFs

We next assayed the impact of AS on the protein level with respect to the canonical transcript isoform that was defined as the transcript isoform with the highest abundance in most samples for each gene. Overall, 78.5% of the AS transcript isoforms encoded altered ORFs. The most frequent category of ORF alteration is 5'-ORF shortening (26.4%) (Fig. 3a). Whereas only 4.4% have lengthened ORFs, 44.6% of AS transcript isoforms contain shortened ORFs at the 5'- and/or 3'-end. Moreover, the differential IR events among tissues also mainly resulted in 5'- or 3'-ORF shortening (72.6%) (Fig. S5). Therefore, AS may largely contribute to proteome complexity in *F. graminearum* if the resulting isoforms are translated.

AS may be not coupled to NMD generally

The AS transcript isoforms with shortened ORFs at 5'- and/or 3'-end generally had a lower expression level in different tissues (Fig. 3b). Since the ORF shortening is likely caused by introduction of a frameshift or nonsense codon via alternative splicing at the 5'- or 3'-region of the original ORFs, these ORF-shortened transcripts may be subjected to degradation by NMD, a mechanism that selectively degrades PTC-containing transcripts to prevent production of truncated proteins (Kishor *et al.*, 2019). Transcripts with a PTC or even a normal stop codon > 50 nt upstream of the last exon-exon junction can efficiently trigger NMD (Garcia-Moreno & Romao, 2020). We identified 2057 candidate NMD targets from the AS transcript isoforms. Among them, 764 contained a PTC. Surprisingly, the expression levels of both PTC-containing and non-PTC-containing NMD candidates were not significantly lower than that of non-NMD candidates (Fig. 3c), indicating that the AS transcripts containing a classical NMD-eliciting feature may be not sensitive to NMD in *F. graminearum*.

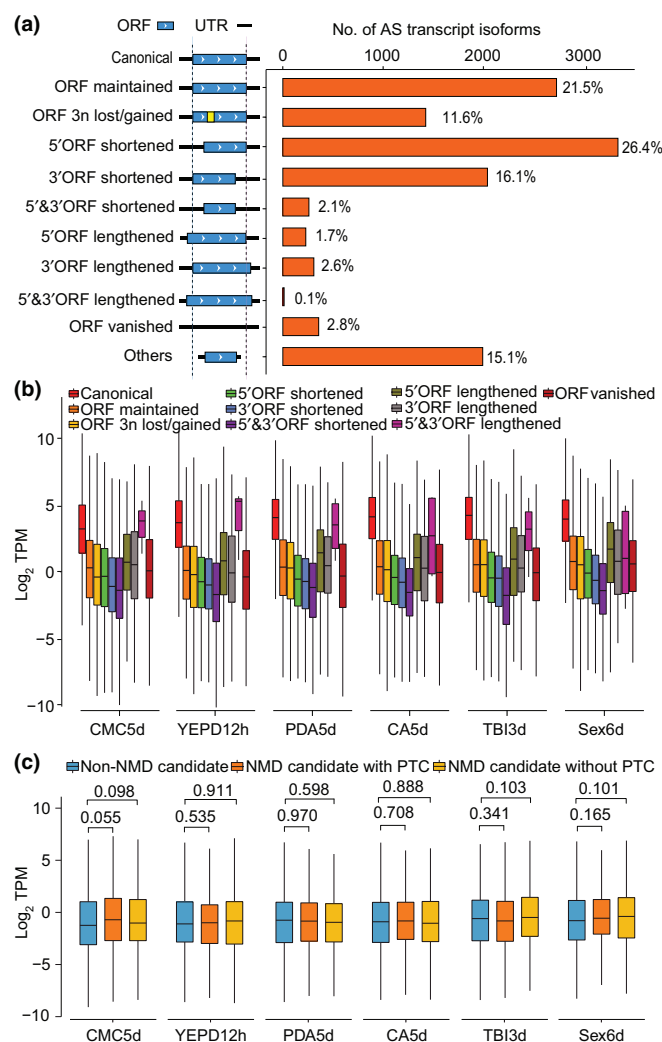


Fig. 3 Alteration of encoding proteins and expression of alternative splicing (AS) transcript isoforms in *Fusarium graminearum*. (a) The number and percentage of AS transcript isoforms for each marked category. The diagrams in the middle show the effects of each category of AS transcript isoforms on protein coding relative to the canonical transcript isoform. Yellow box represents in-frame open reading frame (ORF) alteration with the number of nucleotides lost or gained that can be divided by three. UTR, untranslated region. (b) Box plots comparing the expression levels ($\log_2$ transcripts per kilobase million, TPM) of different categories of AS transcript isoforms. (c) Box plots comparing the expression levels ($\log_2$ TPM) of the predicted premature-termination codon (PTC)-containing and non-PTC-containing nonsense-mediated mRNA decay (NMD) candidates and the non-NMD candidates of the 3'-ORF-shortened AS transcript isoforms. The *P* values are from two-tailed Kruskal–Wallis test. For (b) and (c), horizontal line within the rectangle represents the median of values and whisker indicates the largest/lowest point inside the range defined by 1.5 times interquartile range (IQR). See Supporting Information Table S2 for details of sample information.

To further characterize the relation of AS transcripts with NMD in *F. graminearum*, we identified and deleted *FgUPF1* (FG1G42730) (Fig. 4a; Tables S6, S7), an ortholog of the yeast *UPF1* gene encoding the central factor of the NMD pathway (Kishor *et al.*, 2019). The $\Delta Fgupf1$ mutant grew much slower than the wild-type on PDA plates (Fig. 4b), demonstrating the importance of *FgUPF1* in *F. graminearum*. Strand-specific

RNA-Seq analysis revealed that the abundance of the PTC-containing NMD candidates and ORF-shortened transcripts were not significantly increased or even slightly decreased in the $\Delta Fgupf1$ mutant (Fig. 4c; Table S2), confirming that these transcripts are not degraded by NMD. Additionally, although transcripts with upstream ORF (uORF) or long 3'-UTR were also reported to induce NMD (Garcia-Moreno & Romao, 2020), the upregulated transcripts in the $\Delta Fgupf1$ mutant were not enriched for uORF and their 3'-UTRs were shorter than those of nondifferentially expressed transcripts (Fig. 4d–f). Therefore, AS may be not coupled to NMD generally in *F. graminearum*.

The majority of upregulated and downregulated transcripts in the $\Delta Fgupf1$ mutant were ORF-maintained or canonical transcript isoforms (Fig. 4g). Interestingly, the downregulated transcripts were mostly enriched for genes associated with ribosome biogenesis (Fig. 4h). The repression of ribosome biogenesis may be directly related to the transcriptional alteration and severe growth defect of the $\Delta Fgupf1$ mutant.

Landscape of APA

APA can enhance transcriptome complexity by generating RNA isoforms that differ in their 3' end. Since the PASs of mRNAs are well represented in the FLNC reads, we characterized the global polyadenylation events in the *F. graminearum* genome. A total of 364 513 unique PASs were identified from FLNC reads (Fig. 5a). Among these, 83.9% were in the 3'-UTR regions. Upon APA, heterogeneous cleavage by the polyadenylation machinery can also leave multiple PAS located close to one another. To distinguish APA from heterogeneous cleavage, we modeled two types of distances between adjacent PASs and identified nine nt as the cutoff value (Fig. 5b). Nearby PASs within nine nt of each other in the same gene were clustered and defined as a PAS cluster (heterogeneous cleavage). In total, we identified 104 813 PAS clusters. In each cluster, the PAS with most supporting FLNC reads was selected for representative. In total, 11 133 genes (64.8%) displayed APA with an average of 9.2 sites per gene, including 6123 in perithecia and 5530 in hyphae (Fig. 5c).

For the APA genes in *F. graminearum*, the median length of the shortest and longest 3'-UTRs were 58 and 395 nt, respectively (Fig. 5d,e). Compared to hyphae, the shortest 3'-UTR isoforms were longer in perithecia but the longest ones had no significant difference. Furthermore, the alternative 3'-UTR (aUTR) in perithecia (peak value, 238 nt) were 127% longer than those in hyphae (peak value, 188 nt) (Fig. 5f).

A global increase in distal PAS usage in aging or dormant tissues

To examine the changes of PAS usage among different tissues, the expression ratio of the distal PAS isoform (*M*) to the proximal PAS isoform (*m*) for each PAS except the last ones was analyzed. The larger the *M/m* value, the higher the abundance for the distal PAS isoform. For the majority of PASs, the *M/m* value was < 1 (Fig. S6), suggesting that the distal PAS isoforms are generally less abundant in *F. graminearum*.

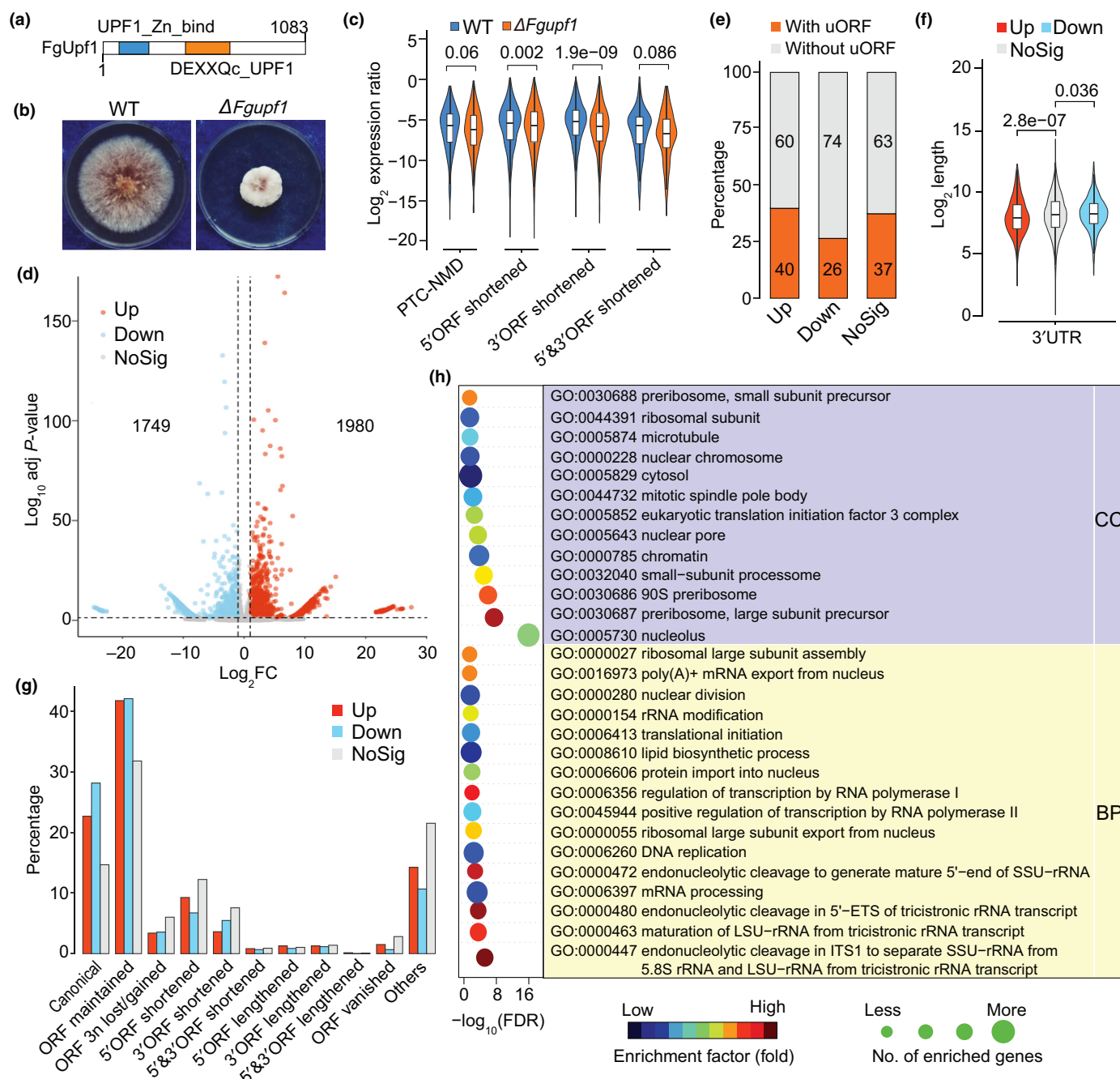


Fig. 4 Characterization of the $\Delta Fgupf1$ mutant of *Fusarium graminearum*. (a) Conserved domains of FgUpf1. DEXXQc_UPF1, DEXXQ-box helicase domain of Upf1; UPF1_Zn_bind, RNA helicase (Upf2 interacting domain). (b) Three-day-old cultures of the wild-type strain PH-1 (WT) and $\Delta Fgupf1$ mutant grown on potato dextrose agar (PDA) plates. (c) Box plots comparing the $\log_2$ expression ratios (transcripts per kilobase million (TPM) of transcript/total TPMs of all transcripts per gene) of the premature-termination codon (PTC)-containing nonsense-mediated mRNA decay (NMD) candidates (PTC-NMD) and the AS transcript isoforms with shortened open reading frames (ORFs) at 3'- and/or 5'-end in WT and $\Delta Fgupf1$ mutant. (d) Volcano plot of the significantly upregulated and downregulated transcripts in the $\Delta Fgupf1$ mutant ($|\log_2$ Fold Change| ≥ 1 , adjusted P -value < 0.05). (e, f) Percentage of transcripts with upstream ORF (uORF) (e) and box plots comparing the 3'-untranslated region (UTR) lengths (f) in the same three categories of transcripts (Up, Down, and NoSig). For (c–f), horizontal line within the rectangle represents the median of values and whisker indicates the largest/lowest point inside the range defined by 1.5 times interquartile range (IQR). The P values are from two-tailed Kruskal–Wallis test. (g) Distribution of the upregulated, downregulated, and nondifferentially expressed (NoSig) transcripts in the $\Delta Fgupf1$ mutant in each marked category. (h) Enriched gene ontology (GO) terms in genes with downregulated transcripts in the $\Delta Fgupf1$ mutant. CC, cell component; BP, biological process.

Notably, the M/m values of most PASs were largest in Sex10d (Fig. 5g), suggesting that the usage of distal PASs is generally increased in Sex10d. Furthermore, genes with increased distal

PAS usage in the older or dormant tissues as compared to younger tissues outnumbered those with decreased distal PAS usage (Fig. 5h). In comparison with germlings, the usage of distal

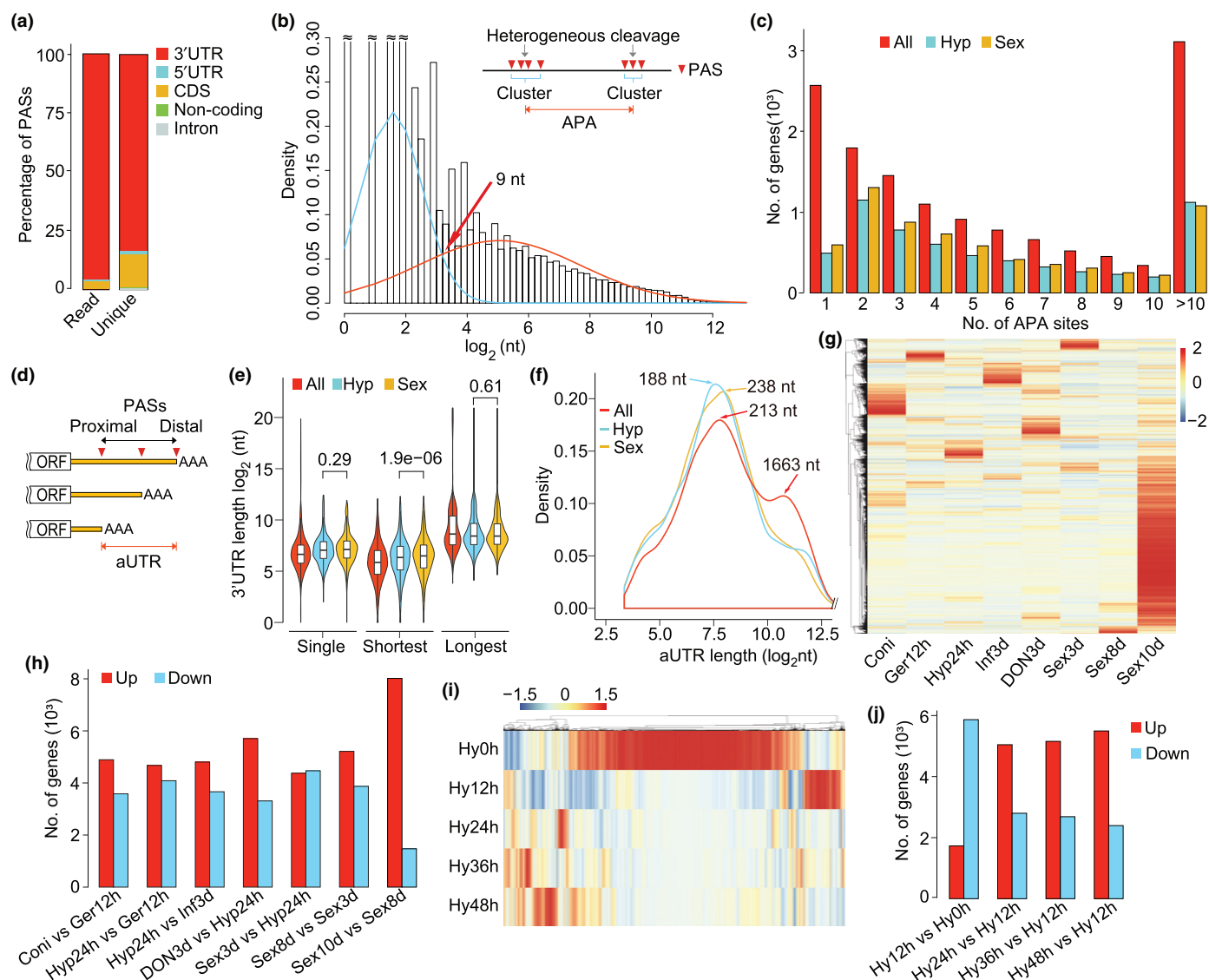


Fig. 5 Landscape and regulation of alternative polyadenylation (APA) in *Fusarium graminearum*. (a) Statistics of full-length, nonartificial-concatemer (FLNC) reads mapped to different regions of the *F. graminearum* genome and unique genomic polyadenylation sites (PASs). (b) Distribution of distances between adjacent PASs, with the blue line showing the distance between PASs within a PAS cluster and the red line showing the distance between PASs from different PAS clusters. The red arrow marks the crossover point used to group PASs into a PAS cluster. The PASs (arrowheads) within a PAS cluster were considered to be derived from heterogeneous cleavage of polyadenylation machinery while the PASs between different PAS clusters were considered to be derived from APA. (c) Distribution of the number of APA sites per gene detected in all samples (All) as well as in vegetative hyphae (Hyp) and sexual perithecia (Sex). (d) Schematic drawing of three APA isoforms derived from different PASs (arrowheads) in the 3'-untranslated region (UTR). The region between the first (most proximal) PAS and last (most distal) PAS is named alternative 3'-UTR (aUTR). (e) Violin plots comparing the 3'-UTR length of genes without APA (Single) and genes with APA (only the shortest and longest 3'-UTR isoforms were plotted). Horizontal line within the rectangle represents the median of values and whisker indicates the largest/lowest point inside the range defined by 1.5 times interquartile range (IQR). The *P* values are from two-tailed Kruskal–Wallis test. (f) Distribution of the length of aUTRs identified in all samples (All) as well as in vegetative hyphae (Hyp) and sexual perithecia (Sex). The peak values are indicated. (g) Heatmap of the *M/m* values for each PAS across marked samples. High (orange to red) and low (yellow to blue) *M/m* values are depicted as Z-scores for each PAS. (h) Number of genes with significantly increased (Up) or decreased (Down) distal PAS usage in marked pairwise comparisons (*P* < 0.05). (i) Heatmap of the *M/m* values for each PAS in conidia (Hy0h) and vegetative growth from germlings (Hy12h) to older hyphae (Hy48h). High (orange to red) and low (yellow to blue) *M/m* values are depicted as Z-scores for each PAS. (j) Number of genes with significantly increased (Up) or decreased (Down) distal PAS usage in marked pairwise comparisons (*P* < 0.05). For (g–j), see Supporting Information Table S2 for details of sample information.

PASs was remarkably increased in dormant conidia and also continuously increased in aging hyphae from 24 to 48 h (Fig. 5i,j). These results suggest that a global increase of distal PAS usage to generate long 3'-UTR isoforms is associated with aging or the dormant state.

A great number of GO terms related to autophagy, signal transduction, RNA binding, RNA processing, RNA splicing, RNA metabolism, translation, protein folding, protein localization, protein transport, and ubiquitin-mediated proteolysis were found to be significantly enriched in the genes with increased

distal PAS usage in Sex10d, Coni, and/or 48-h hyphae (Table S8). Many of these functions or pathways have been reported to be associated with cellular senescence or aging in animals (Deschenes & Chabot, 2017; Chen *et al.*, 2018; Angarola & Anczukow, 2021). Therefore, it is possible that 3'-UTR lengthening acts as a novel mechanism in regulating aging and dormancy in *F. graminearum*.

Polyadenylation signals surrounding PASs

To search for *cis*-acting elements that may guide cleavage and polyadenylation, we extracted upstream 100 nt and downstream 100 nt sequences surrounding each PAS to examine nucleotide distributions and enriched motifs. The single-nucleotide frequencies in *F. graminearum* are similar to that observed in yeasts (Liu *et al.*, 2017) (Fig. 6a). Four enriched hexameric motifs, AATWVA (W = A or T, V = A, C or G), TAKMTA (K = G or T, M = A or C), TTTTIT, and HGTGAH (H = A, C or T) were identified (Fig. 6b). The AATWVA and TTTTIT motifs occurred in a highly position-specific manner, suggesting that they are mechanistically important for cleavage and polyadenylation in *F. graminearum*. All four motifs were more abundant at distal PASs (Fig. 6c–f), indicating that distal PASs are stronger than proximal ones in *F. graminearum*.

Roles of 3'-end processing factors in PAS selection and intron splicing

Besides *cis*-acting elements surrounding the PASs, 3'-end processing machinery also plays important roles in APA regulation by influencing PAS selections. This machinery includes three major complexes that are necessary for cleavage and polyadenylation in yeast: Cleavage Factor IA (CFIA), Cleavage Factor IB (CFIB), and Cleavage and Polyadenylation Factor (CPF) (Vavasseur & Shi, 2014). The expression of the core 3'-end processing factors in *F. graminearum* was generally higher in the younger tissues but lower in the older or dormant tissues (Fig. 6g), suggesting that increased distal PAS usage in aging and dormant tissues may be due to global downregulation of these core 3'-end processing factors. To further characterize the function of the 3'-end processing factors in APA regulation, we selected five of them to delete, including *FgRNA15*, *FgCLP1*, *FgHRP1*, *FgYTH1*, and *FgFIP1* in *F. graminearum* (Fig. S7), which are orthologs of *RNA15*, *CLP1*, *HRP1*, *YTH1*, and *FIP1* in *Saccharomyces cerevisiae* (www.yeastgenome.org/),

respectively. Although all the five genes are essential for viability in *Saccharomyces cerevisiae*, we obtained deletion mutants for the *FgRNA15*, *FgHRP1*, and *FgFIP1* genes (Table S7). The *FgCLP1* and *FgYTH1* genes are also likely essential for viability in *F. graminearum* because we failed to identify deletion mutants after repeated attempts. The $\Delta Fghrp1$ and $\Delta Fgrna15$ mutants had severe growth defects and rarely produced aerial hyphae on V8 juice plates (Fig. 6h). They also failed to produce perithecia on mating plates, confirming their important roles in mRNA 3'-end processing. Unexpectedly, the $\Delta Fgfip1$ mutant had no obvious defects in growth (Fig. 6h). Perithecia formed by that mutant were normal in size and morphology but defective in ascospore release.

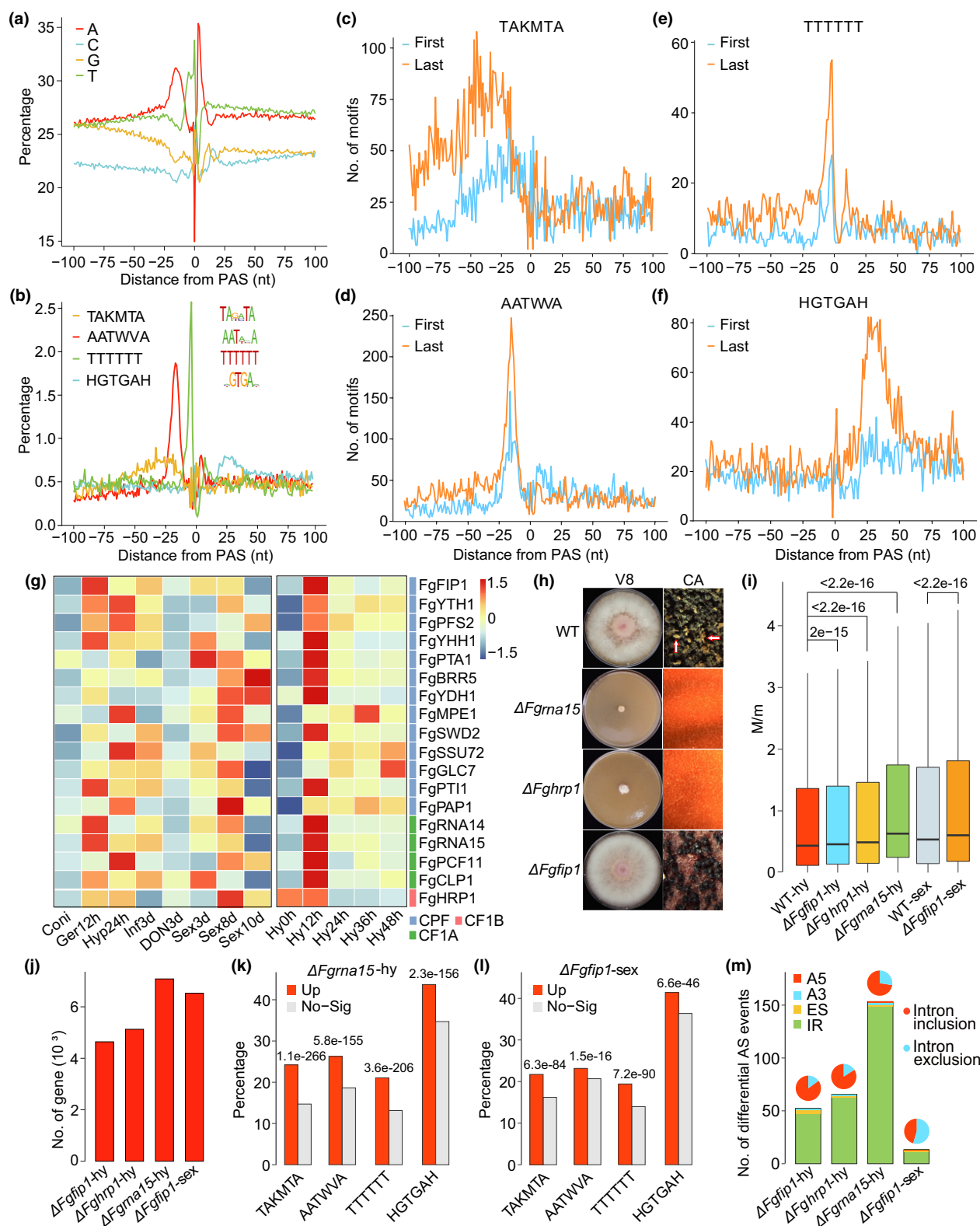
Strand-specific RNA-Seq analyses revealed that the *M/m* values were significantly increased in all three mutants compared to the wild-type (Fig. 6i). More than 4000 genes with significantly increased distal PAS usage were detected in each mutant (Fig. 6j), suggesting that the *FgRNA15*, *FgHRP1*, and *FgFIP1* genes are all functional in promoting proximal PAS usage. Especially, the *M/m* values of most PASs were substantially increased in the $\Delta Fgrna15$ mutant (Fig. S8), indicating the importance of *FgRNA15* in the regulation of the proximal PAS usage. The four motifs were significantly enriched in the vicinity of the PASs with increased distal PAS usage in all mutants except for the AATWVA motif, which was significantly depleted in the hyphal sample of the $\Delta Fgfip1$ and $\Delta Fghrp1$ mutants (Figs 6k, l, S9). Therefore, all three genes are required for the recognition of these motifs in *F. graminearum*.

Unexpectedly, we observed instances in which intron splicing changed in the $\Delta Fgrna15$ mutant (Fig. S10). We detected 153, 65, and 52 differential AS events in the hyphal sample of the $\Delta Fgrna15$, $\Delta Fghrp1$, and $\Delta Fgfip1$ mutants in comparison with wild-type (Fig. 6m). The majority of differential AS events were intron inclusion. These differentially spliced introns had a similar distribution throughout transcripts as the total introns (Fig. S11), implying that they are not likely caused by APA within introns. These results suggest that the *FgRNA15*, *FgHRP1*, and *FgFIP1* genes promote intron splicing in *F. graminearum*. Especially, *FgRNA15* plays a more important role.

Discussion

Although previous gene model annotation was generated by integrating multiple gene prediction algorithms, EST and RNA-Seq

Fig. 6 Roles of *cis*-acting elements and 3'-end processing factors in polyadenylation site (PAS) selection in *Fusarium graminearum*. (a) Nucleotide frequencies around the PASs. (b) Distribution of four enriched hexameric motifs in the flanking region of PASs. (c–f) Comparison of the distribution of each hexameric motif in the vicinity of the first and last PASs in the same terminal exon. (g) Heatmap of the expression of genes encoding core messenger RNA (mRNA) 3'-end processing factors in different samples. High (orange to red) and low (yellow to blue) expression levels are depicted as Z-scores for each gene. Subunits of three complexes (CPF, CP1B, and CF1A) are indicated with different colors as marked. (h) Defects of the $\Delta Fgfip1$, $\Delta Fghrp1$, and $\Delta Fgrna15$ mutants in growth and sexual reproduction. Three-day-old V8 cultures and perithecia formed on carrot agar (CA) plates at 8 d post-fertilization (dpf). Arrows point to ascospore cirrhi. (i) Box plots of the *M/m* values of PASs for the marked samples. Horizontal line within the rectangle represents the median of values and whisker indicates the largest/lowest point inside the range defined by 1.5 times interquartile range (IQR). The *P* values are from two-tailed Kruskal–Wallis test. (j) Number of genes with significantly increased distal PAS usage in marked mutants relative to the wild-type (WT) (*P* < 0.05). (k, l) Percentage of the four hexameric motifs detected in the vicinity of the PASs with increased (Up) or nondifferential (No-Sig) distal PAS usage in $\Delta Fgrna15$ mutant (k) and the sexual sample of the $\Delta Fgfip1$ mutant (l). The *P* values are from Fisher's exact test. (m) Number of significantly differential AS events in marked mutants relative to the WT. Pie charts show the proportions of differential intron inclusion and intron exclusion events. For (g, i, j and m), see Supporting Information Table S2 for details of sample information.



data were merely used for gene model support. In this study, we generated a comprehensive annotation (YL1) comprising 51 617 transcript isoforms from 17 189 genes for *F. graminearum* strain PH-1. Compared to the previous annotation (RR1) consisting of only 14 145 transcripts from 14 145 genes, the updated reference

annotation YL1 is much more comprehensive and complete. To our knowledge, this is the most comprehensive full-length transcriptome documented in filamentous fungi to date, which is a rich resource of transcript isoforms for comparative and functional genomic studies in filamentous ascomycetes. Our

methodology for Iso-Seq data analysis will also be a useful reference for analyzing the full-length transcriptome in other fungal species.

In *F. graminearum*, only 231 genes with AS events were previously reported (Zhao *et al.*, 2013). In this study, we detected 54 613 AS events in 4997 genes. One advantage of Iso-Seq is its suitability for identifying not only the AS events but also the splice isoforms. We detected a total of 17 229 splice isoforms and found that the splice isoforms of 1960 genes were derived from multiple combinatory AS events. These results suggest that AS significantly increases fungal transcriptome complexity, expanding our view of the regulatory potential of RNA splicing in fungi. Furthermore, detection and quantification of differential AS between sample groups from short-read RNA-Seq data commonly rely on pre-annotation of known spliced transcripts, the massive differential AS events detected from the RNA-Seq data of different tissues demonstrate the tremendous advantages of our YL1 annotation.

The regulatory function of AS in higher eukaryotes has been well established (Chaudhary *et al.*, 2019). However, the contribution of AS to fungal biology is still elusive. Recently, the AS event of *MoPTEN* was reported to be important for growth and pathogenesis in *Magnaporthe oryzae* (Wang *et al.*, 2021). In *Sclerotinia sclerotiorum*, a number of AS isoforms were differentially expressed on diverse host plants, which may contribute to its broad host spectrum (Ibrahim *et al.*, 2021). In *F. graminearum*, we found that a variety of biological processes were enriched in the AS genes, implying that AS may play important regulatory roles. Furthermore, a large number of differential AS events were identified among different tissues, indicating that these AS events are developmentally regulated. More importantly, we found a global increase in intron inclusion in aging or dormant tissues, suggesting that IR regulation is likely associated with aging or dormancy in *F. graminearum*. In animals, a global increase in intron inclusion has also been observed in aging tissues (Bhadra *et al.*, 2020; Angarola & Anczukow, 2021) and dysfunctional AS events contribute to the aging/senescence phenotype across multiple species (Deschenes & Chabot, 2017; Bhadra *et al.*, 2020; Angarola & Anczukow, 2021). As in animals (Adusumalli *et al.*, 2019), the differentially retained introns in the aging or dormant tissues have a higher GC content, and the expression levels of spliceosomal genes were reduced in the aging or dormant tissues in *F. graminearum*. Therefore, a global increase in intron inclusion may be a transcriptional signature of aging that is conserved between animals and fungi.

In *F. graminearum*, the majority of AS transcript isoforms potentially encode proteins with altered sequences relative to canonical transcript isoforms. The differential AS events among tissues also mainly caused ORF shortening. Are these transcript isoforms truly functional to produce proteins? Coupling AS to NMD is a ubiquitous regulatory mode of gene expression in eukaryotes (Garcia-Moreno & Romao, 2020). In *Saccharomyces cerevisiae*, AS is mainly used to control transcript levels rather than generate proteome diversity (Kawashima *et al.*, 2014). In this study, we found no evidence that the AS transcript isoforms with shortened ORFs or known NMD-eliciting features were

subjected to degradation by NMD, although the possibility that a small subset of them is degraded by NMD is not excluded. It is likely that the AS mainly contributes to increasing proteomic diversity in *F. graminearum*. However, since the 5'- and/or 3'-ORF shortened transcripts generally had a lower expression level, NMD-independent mechanism may act on them in *F. graminearum*. In fact, the existence of an intron retention-dependent mechanism of gene expression regulation that is not dependent on NMD has been suggested in the yeast *Cryptococcus neoformans* (Gonzalez-Hilarion *et al.*, 2016).

Distinct from yeasts, in which depletion of the Upf1 proteins does not affect growth (Gonzalez-Hilarion *et al.*, 2016), the $\Delta Fgupf1$ mutant of *F. graminearum* had severe growth defects. Interestingly, the downregulated transcripts in $\Delta Fgupf1$ mutant were mostly enriched for GO categories related to ribosome biogenesis. The downregulation of genes involved in translation and ribosome biogenesis was also found in the *upf1* mutant of *Arabidopsis* (Raxwal *et al.*, 2020). Therefore, the FgUpf1 likely plays an important role in translational gene regulation in *F. graminearum* as suggested in *Arabidopsis*.

Using a short-read sequencing-based 3'-T-fill method, 14 593 PASs were identified and 52% (4283) of *M. oryzae* genes were found to be alternatively polyadenylated (Rodriguez-Romero *et al.*, 2019). PASs can be accurately identified from the Iso-Seq FLNC reads at single-nucleotide resolution. We identified 364 513 unique PASs and showed that 64.8% (11 133) of the genes were alternatively polyadenylated in *F. graminearum*. These results suggest that APA is pervasive in filamentous fungi as in mammals and plants (Wu *et al.*, 2011; Tian & Manley, 2017). However, unlike in yeasts and mammals (Hoque *et al.*, 2013; Liu *et al.*, 2017; Gruber & Zavolan, 2019), distal APA sites have strong polyadenylation signals but proximal APA isoforms are highly expressed in *F. graminearum*. We showed that the core 3'-end processing factors FgRNA15, FgHRP1, and FgFIP1 all play important roles in recognizing the sequence motifs and promoting proximal PAS usage in *F. graminearum*. Deletion of them resulted in global increase in distal PAS usage. In humans, CSTF2/CSTF64, FIP1L1, and CFIm that are homologous to yeast RNA15, FIP1, and HRP1, respectively, were reported to promote cleavage and polyadenylation at proximal PASs (Gruber & Zavolan, 2019; Pereira-Castro & Moreira, 2021). However, unlike CFIm in humans, FgRNA15 has a particularly strong impact on PAS selection in *F. graminearum*. Furthermore, we found increased intron inclusion in these three mutants, especially in the $\Delta Fgrna15$ mutant, suggesting that the three 3'-end processing factors may also promote intron splicing in *F. graminearum*. A role for 3'-end processing factors in promoting AS was reported only recently in humans (Misra *et al.*, 2015), where the CPSF (homologous to the fungal CPF:CFII subcomplex) and SYMPK, but not other 3'-end processing factors, have a global role in promoting both inclusion and exclusion of internal exons. Our study indicates that the subunits of different 3'-end processing complex/subcomplex could be involved in promoting AS in fungi, including CFIA (FgRNA15), CFIB (FgHRP1), and CPF:PFI (FgFIP1).

Since 3'-UTR affects the stability, translation rate, and subcellular localization of mRNAs (Pereira-Castro & Moreira, 2021), changes in the ratio of 3'-UTR isoforms in different tissues may

affect cellular functions of corresponding proteins. We found a global increase in distal PAS usage in dormant and old tissues in *F. graminearum*. Moreover, genes with increased distal PAS usage in the dormant and old tissues were enriched for the ones functionally related to many senescence/age-related pathways known in mammals (Deschenes & Chabot, 2017; Chen *et al.*, 2018; Angarola & Anczukow, 2021). These results suggest that APA-mediated 3'-UTR lengthening may play a role in regulating aging and dormancy in *F. graminearum*. In mammals, global lengthening of 3'-UTRs in senescent cells was described only recently (Chen *et al.*, 2018). In yeasts, the expression of long 3'-UTR isoforms is favored in nutrient-restricted conditions, in which cells reach the quiescent state (Liu *et al.*, 2017). In mammals, it has been proposed that the generation of short 3'-UTR in proliferating cells is due to global upregulation of the 3'-end processing machinery. In *F. graminearum*, the increase of distal PAS usage in generating long 3'-UTR isoforms in the older or dormant tissues may be due to global downregulation of the 3'-end processing machinery. Therefore, APA regulating gene expression during cell senescence/aging and dormancy may be a common mechanism in mammals and fungi.

Retained introns located in the 5'- or 3'-UTR may increase the lengths of affected UTRs. Since both long 5'- and 3'-UTR isoforms tended to translate less efficiently (Wang *et al.*, 2016; Pereira-Castro & Moreira, 2021), our findings of the global increase in both intron inclusion and distal PAS usage in aging and dormant tissues demonstrate that AS and APA are an integral part of cell reprogramming processes to regulate the overall protein production in *F. graminearum*.

Overall, this study represents the first comprehensive analysis of a full-length transcriptome in a plant pathogenic fungus, providing new insights into the complexity and regulation of AS and APA in filamentous fungi. The updated genome sequences and comprehensive reference set of transcript isoforms generated in this study will be beneficial to the *Fusarium* community and fungal community in general for comparative and functional genomic studies.

Acknowledgements

The authors thank Mengchun Wu for preparing the fungal samples; Drs Xue Zhang, Guanghui Wang, and Chenfang Wang for helpful discussions; and Drs Frances Trail and Yun Chen for kindly providing the PH-1 laboratory stocks. The authors also thank Dr Larry Dunkle at Purdue University (West Lafayette, IN, USA) for critical reading of this manuscript. This study was supported by funding for HL from the National Key R&D Program of China (2019YFD1000605), National Natural Science Foundation of China (no. 31872918), and National Youth Talent Support Program (Z111021802), and funding for J-RX from USWBSI.

Author contributions

HL and PL designed the experiments and analysis pipelines. DC, ZQ, and YC performed the experiments, PL, HW, and QW

performed the analyses. HL, J-RX and CJ contributed advice and reagents. The figures were prepared by PL, HL and HW. The manuscript was written by HL, PL and J-RX. All authors read and approved the final manuscript.

ORCID

Cong Jiang  <https://orcid.org/0000-0002-5634-3200>
Huiquan Liu  <https://orcid.org/0000-0002-4723-845X>
Qinhu Wang  <https://orcid.org/0000-0003-1251-073X>
Jin-Rong Xu  <https://orcid.org/0000-0001-5999-5004>

Data availability

PacBio and Illumina Sequencing data have been submitted to the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under Bio-Project: PRJNA664227, PRJNA664506, PRJNA293594, and PRJNA699410. The updated version of PH-1 assembly (YL1) has been submitted to the NCBI GenBank under Bio-Project: PRJNA782099. The YL1 transcript annotation is freely available at FgBase (<http://fgbase.wheatcab.com/>). The PYTHON scripts developed in this study were submitted to GitHub (<https://github.com/xulab-nwafu/Iso-seq-tools>).

References

- Adusumalli S, Ngian ZK, Lin WQ, Benoukraf T, Ong CT. 2019. Increased intron retention is a post-transcriptional signature associated with progressive aging and Alzheimer's disease. *Aging Cell* 18: e12928.
- Angarola BL, Anczukow O. 2021. Splicing alterations in healthy aging and disease. *Wiley Interdisciplinary Reviews: RNA* 12: e1643.
- Bhadra M, Howell P, Dutta S, Heintz C, Mair WB. 2020. Alternative splicing in aging and longevity. *Human Genetics* 139: 357–369.
- Bian Z, Ni Y, Xu JR, Liu H. 2019. A-to-I mRNA editing in fungi: occurrence, function, and evolution. *Cellular and Molecular Life Sciences* 76: 329–340.
- Brauer EK, Subramaniam R, Harris LJ. 2020. Regulation and dynamics of gene expression during the life cycle of *Fusarium graminearum*. *Phytopathology* 110: 1368–1374.
- Chaudhary S, Khokhar W, Jabre I, Reddy ASN, Byrne LJ, Wilson CM, Syed NH. 2019. alternative splicing and protein diversity: plants versus animals. *Frontiers in Plant Science* 10: 708.
- Chen M, Lyu G, Han M, Nie H, Shen T, Chen W, Niu Y, Song Y, Li X, Li H *et al.* 2018. 3' UTR lengthening as a novel mechanism in regulating cellular senescence. *Genome Research* 28: 285–294.
- Cuomo CA, Guldener U, Xu JR, Trail F, Turgeon BG, Di Pietro A, Walton JD, Ma LJ, Baker SE, Rep M *et al.* 2007. The *Fusarium graminearum* genome reveals a link between localized polymorphism and pathogen specialization. *Science* 317: 1400–1402.
- Dash S, Van Hemert J, Hong L, Wise RP, Dickerson JA. 2012. PLEXDB: gene expression resources for plants and plant pathogens. *Nucleic Acids Research* 40: D1194–D1201.
- Dean R, Van kan JAL, Pretorius ZA, Hammond-kosack KE, Di pietro A, Spanu PD, Rudd JJ, Dickman M, Kahmann R, Ellis J *et al.* 2012. The top 10 fungal pathogens in molecular plant pathology. *Molecular Plant Pathology* 13: 414–430.
- Deschenes M, Chabot B. 2017. The emerging role of alternative splicing in senescence and aging. *Aging Cell* 16: 918–933.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15–21.
- Dong WX, Ding JL, Gao Y, Peng YJ, Feng MG, Ying SH. 2017. Transcriptomic insights into the alternative splicing-mediated adaptation of

- the entomopathogenic fungus *Beauveria bassiana* to host niches: autophagy-related gene 8 as an example. *Environmental Microbiology* 19: 4126–4139.
- Foissac S, Sammeth M. 2007. ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets. *Nucleic Acids Research* 35: W297–W299.
- Garcia-Moreno JF, Romao L. 2020. Perspective in alternative splicing coupled to nonsense-mediated mRNA decay. *International Journal of Molecular Sciences* 21: 9424.
- Gehrmann T, Pelkmans JF, Lugones LG, Wosten HA, Abeel T, Reinders MJ. 2016. *Schizophyllum commune* has an extensive and functional alternative splicing repertoire. *Scientific Reports* 6: 33640.
- Gonzalez-Hilarion S, Paulet D, Lee K-T, Hon C-C, Lechat P, Mogensen E, Moyrand F, Proux C, Barboux R, Bussotti G *et al.* 2016. Intron retention-dependent gene regulation in *Cryptococcus neoformans*. *Scientific Reports* 6: 32252.
- Grassi E, Mariella E, Lembo A, Molineris I, Provero P. 2016. ROAR: detecting alternative polyadenylation with standard mRNA sequencing libraries. *BMC Bioinformatics* 17: 423.
- Gruber AJ, Zavolan M. 2019. Alternative cleavage and polyadenylation in health and disease. *Nature Reviews Genetics* 20: 599–614.
- Hoque M, Ji Z, Zheng D, Luo W, Li W, You B, Park JY, Yehia G, Tian B. 2013. Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nature Methods* 10: 133–139.
- Ibrahim HMM, Kusch S, Didelon M, Raffaele S. 2021. Genome-wide alternative splicing profiling in the fungal plant pathogen *Sclerotinia sclerotiorum* during the colonization of diverse host families. *Molecular Plant Pathology* 22: 31–47.
- Jiang C, Cao S, Wang Z, Xu H, Liang J, Liu H, Wang G, Ding M, Wang Q, Gong C *et al.* 2019. An expanded subfamily of G-protein-coupled receptor genes in *Fusarium graminearum* required for wheat infection. *Nature Microbiology* 4: 1582–1591.
- Jiang C, Hei R, Yang Y, Zhang S, Wang Q, Wang W, Zhang Q, Yan M, Zhu G, Huang P *et al.* 2020. An orphan protein of *Fusarium graminearum* modulates host immunity by mediating proteasomal degradation of TaSnRK1alpha. *Nature Communications* 11: 4382.
- Jin L, Li G, Yu D, Huang W, Cheng C, Liao S, Wu Q, Zhang Y. 2017. Transcriptome analysis reveals the complexity of alternative splicing regulation in the fungus *Verticillium dahliae*. *BMC Genomics* 18: 130.
- Kawashima T, Douglass S, Gabunilas J, Pellegrini M, Chanfreau GF. 2014. Widespread use of non-productive alternative splice sites in *Saccharomyces cerevisiae*. *PLoS Genetics* 10: e1004249.
- Kazan K, Gardiner DM. 2018. Transcriptomics of cereal-*Fusarium graminearum* interactions: what we have learned so far. *Molecular Plant Pathology* 19: 764–778.
- Kent WJ. 2002. BLAT—the BLAST-like alignment tool. *Genome Research* 12: 656–664.
- King R, Urban M, Hammond-Kosack MC, Hassani-Pak K, Hammond-Kosack KE. 2015. The completed genome sequence of the pathogenic ascomycete fungus *Fusarium graminearum*. *BMC Genomics* 16: 544.
- King R, Urban M, Hammond-Kosack KE. 2017. Annotation of *Fusarium graminearum* (PH-1) v.5.0. *Genome Announcements* 5: e01479–e1516.
- Kishor A, Fritz SE, Hogg JR. 2019. Nonsense-mediated mRNA decay: the challenge of telling right from wrong in a complex transcriptome. *Wiley Interdisciplinary Reviews: RNA* 10: e1548.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. CANU: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research* 27: 722–736.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with BOWTIE 2. *Nature Methods* 9: 357–359.
- Li A, Zhang J, Zhou Z. 2014. PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC Bioinformatics* 15: 311.
- Liu H, Wang Q, He Y, Chen L, Hao C, Jiang C, Li Y, Dai Y, Kang Z, Xu JR. 2016. Genome-wide A-to-I RNA editing in fungi independent of ADAR enzymes. *Genome Research* 26: 499–509.
- Liu H, Zhang S, Ma J, Dai Y, Li C, Lyu X, Wang C, Xu JR. 2015. Two Cdc2 kinase genes with distinct functions in vegetative and infectious hyphae in *Fusarium graminearum*. *PLoS Pathogens* 11: e1004913.
- Liu X, Hoque M, Larochelle M, Lemay JF, Yurko N, Manley JL, Bachand F, Tian B. 2017. Comparative analysis of alternative polyadenylation in *S. cerevisiae* and *S. pombe*. *Genome Research* 27: 1685–1695.
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15: 550.
- Machanick P, Bailey TL. 2011. MEME-CHIP: motif analysis of large DNA datasets. *Bioinformatics* 27: 1696–1697.
- Marcais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. 2018. MUMMER4: a fast and versatile genome alignment system. *PLoS Computational Biology* 14: e1005944.
- Misra A, Ou J, Zhu LJ, Green MR. 2015. Global promotion of alternative internal exon usage by mRNA 3' end formation factors. *Molecular Cell* 58: 819–831.
- Muzafar S, Sharma RD, Chauhan N, Prasad R. 2021. Intron distribution and emerging role of alternative splicing in fungi. *FEMS Microbiology Letters* 368: fnab135.
- Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* 14: 417–419.
- Pereira-Castro I, Moreira A. 2021. On the function and relevance of alternative 3'-UTRs in gene expression regulation. *Wiley Interdisciplinary Reviews: RNA* 12: e1653.
- Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols* 11: 1650–1667.
- Raxwal VK, Simpson CG, Gloggnitzer J, Entinze JC, Guo W, Zhang R, Brown JWS, Riha K. 2020. Nonsense-mediated RNA decay factor UPF1 is critical for posttranscriptional and translational gene regulation in Arabidopsis. *Plant Cell* 32: 2725–2741.
- Rodriguez-Romero J, Marconi M, Ortega-Campayo V, Demuez M, Wilkinson MD, Sesma A. 2019. Virulence- and signaling-associated genes display a preference for long 3'UTRs during rice infection and metabolic stress in the rice blast fungus. *New Phytologist* 221: 399–414.
- Shin JY, Bui DC, Lee Y, Nam H, Jung S, Fang M, Kim JC, Lee T, Kim H, Choi GJ *et al.* 2017. Functional characterization of cytochrome P450 monooxygenases in the cereal head blight fungus *Fusarium graminearum*. *Environmental Microbiology* 19: 2053–2067.
- Son H, Seo Y-S, Min K, Park AR, Lee J, Jin J-M, Lin Y, Cao P, Hong S-Y, Kim E-K *et al.* 2011. A phenome-based functional analysis of transcription factors in the cereal head blight fungus, *Fusarium graminearum*. *PLoS Pathogens* 7: e1002310.
- Tian B, Manley JL. 2017. Alternative polyadenylation of mRNA precursors. *Nature Reviews Molecular Cell Biology* 18: 18–30.
- Trincado JL, Entinze JC, Hysenaj G, Singh B, Skalic M, Elliott DJ, Eyraes E. 2018. SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biology* 19: 40.
- Vavasseur A, Shi Y. 2014. Fungal pre-mRNA 3'-end processing. In: Sesma A, Haar T, eds. *Fungal RNA biology*. Cham, Switzerland: Springer International, 59–88.
- Wang C, Zhang S, Hou R, Zhao Z, Zheng Q, Xu Q, Zheng D, Wang G, Liu H, Gao X *et al.* 2011. Functional analysis of the kinome of the wheat scab fungus *Fusarium graminearum*. *PLoS Pathogens* 7: e1002460.
- Wang S, Liang H, Wei Y, Zhang P, Dang Y, Li G, Zhang SH. 2021. Alternative splicing of MoPTEN is important for growth and pathogenesis in *Magnaporthe oryzae*. *Frontiers in Microbiology* 12: 715773.
- Wang X, Hou J, Quedenau C, Chen W. 2016. Pervasive isoform-specific translational regulation via alternative transcription start sites in mammals. *Molecular Systems Biology* 12: 875.
- Wu TD, Watanabe CK. 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* 21: 1859–1875.
- Wu W, Zong J, Wei N, Cheng J, Zhou X, Cheng Y, Chen D, Guo Q, Zhang B, Feng Y. 2018. CASH: a constructing comprehensive splice site method for detecting alternative splicing events. *Briefings in Bioinformatics* 19: 905–917.

- Wu X, Liu M, Downie B, Liang C, Ji G, Li QQ, Hunt AG. 2011. Genome-wide landscape of polyadenylation in Arabidopsis provides evidence for extensive alternative polyadenylation. *Proceedings of the National Academy of Sciences, USA* 108: 12533–12538.
- Wyman D, Mortazavi A. 2019. TRANSCRIPTCLEAN: variant-aware correction of indels, mismatches and splice junctions in long-read transcripts. *Bioinformatics* 35: 340–342.
- Xie B-B, Li D, Shi W-L, Qin Q-L, Wang X-W, Rong J-C, Sun C-Y, Huang F, Zhang X-Y, Dong X-W *et al.* 2015. Deep RNA sequencing reveals a high frequency of alternative splicing events in the fungus *Trichoderma longibrachiatum*. *BMC Genomics* 16: 54.
- Yun Y, Liu Z, Yin Y, Jiang J, Chen Y, Xu JR, Ma Z. 2015. Functional analysis of the *Fusarium graminearum* phosphatome. *New Phytologist* 207: 119–134.
- Zhao C, Waalwijk C, de Wit PJ, Tang D, van der Lee T. 2013. RNA-Seq analysis reveals new gene models and alternative splicing in the fungal pathogen *Fusarium graminearum*. *BMC Genomics* 14: 21.
- Zhao Z, Wu X, Kumar PK, Dong M, Ji G, Li QQ, Liang C. 2014. Bioinformatics analysis of alternative polyadenylation in green alga *Chlamydomonas reinhardtii* using transcriptome sequences from three different sequencing platforms. *G3 (Bethesda)* 4: 871–883.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Isoform sequencing (Iso-Seq) analysis pipeline used in this study.

Fig. S2 Comparing expression levels of the predicted genes of strain PH-1 reference annotation (RR1, Ensembl Fungi) with and without Isoform sequencing (Iso-Seq) transcript isoforms in different samples of *Fusarium graminearum*.

Fig. S3 Enriched gene ontology (GO) terms in genes with alternative splicing events in *Fusarium graminearum*.

Fig. S4 Characterization of tissue-specific transcript isoforms and alternative splicing events in *Fusarium graminearum*.

Fig. S5 Alteration of encoding proteins by differential intron retention (IR) events among tissues of *Fusarium graminearum*.

Fig. S6 Relative abundance of distal and proximal polyadenylation site isoforms in *Fusarium graminearum*.

Fig. S7 Conserved domains of five selected 3'-end processing factors for functional characterization in *Fusarium graminearum*.

Fig. S8 Heatmap of the *M/m* values across marked samples of *Fusarium graminearum*.

Fig. S9 Percentage of the four hexameric motifs detected in the vicinity of the polyadenylation sites with increased (Up) or non-significantly differential (No-Sig) distal PAS usage in hyphal samples of the $\Delta Fgfp1$ and $\Delta Fghrp1$ mutants of *Fusarium graminearum*.

Fig. S10 Examples of genes with differential intron retention and alternative 3'-acceptor events in $\Delta Fgrna15$ mutant of *Fusarium graminearum*.

Fig. S11 Distribution of differentially spliced introns in $\Delta Fgrna15$, $\Delta Fghrp1$, and $\Delta Fgfp1$ mutants of *Fusarium graminearum* throughout transcripts.

Table S1 Information of PacBio single molecular real-time (SMRT) DNA and RNA sequencing data used in this study.

Table S2 Information of Illumina DNA- and RNA-sequencing (RNA-Seq) data used in this study.

Table S3 List of corrected errors in the RR1 genome assembly of *Fusarium graminearum* strain PH-1.

Table S4 Potentially novel protein-coding genes in *Fusarium graminearum*.

Table S5 Summary of primers used in this study for reverse transcription polymerase chain reaction (RT-PCR).

Table S6 Summary of primers used in this study for gene knock-out.

Table S7 The knockout strains of *Fusarium graminearum* generated in this study.

Table S8 Enriched gene ontology (GO) terms in genes with increased distal polyadenylation site usage in 10-d post-fertilization (dpf) perithecia, conidia, and/or 48-h hyphae of *Fusarium graminearum*.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.