

Genome-wide Profiling of Long Non-coding RNA of the Rice Blast Fungus *Magnaporthe oryzae* During Infection

Gobong Choi

Seoul National University

Jongbum Jeon

Seoul National University

Hyunjun Lee

Seoul National University

Shenxian Zhou

Seoul National University

Yong-Hwan Lee (✉ yonglee@snu.ac.kr)

Seoul National University

Research Article

Keywords: host infection, long non-coding RNA, *Magnaporthe oryzae*, RNA-seq, transcriptional regulation

Posted Date: August 20th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-677726/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at BMC Genomics on February 15th, 2022.

See the published version at <https://doi.org/10.1186/s12864-022-08380-4>.

Abstract

Background

Long non-coding RNAs (lncRNAs) play essential roles in developmental processes and disease development at the transcriptional and post-transcriptional levels across diverse taxa. However, only few studies have profiled fungal lncRNAs in a genome-wide manner during host infection.

Results

Infection-associated lncRNAs were identified using lncRNA profiling over six stages of host infection (e.g., vegetative growth, pre-penetration, biotrophic, and necrotrophic stages) in the model pathogenic fungus, *Magnaporthe oryzae*. We identified 2,601 novel lncRNAs, including 1,286 antisense lncRNAs and 980 intergenic lncRNAs. Among the identified lncRNAs, 755 were expressed in a stage-specific manner and 560 were infection-specifically expressed lncRNAs (ISELs). To decipher the potential roles of lncRNAs during infection, we identified 365 protein-coding genes that were associated with 214 ISELs. Analysis of the predicted functions of these associated genes suggested that lncRNAs regulate pathogenesis-related genes, including xylanases and effectors.

Conclusions

The ISELs and their associated genes provide a comprehensive view of lncRNAs during fungal pathogen-plant interactions. This study expands new insights into the role of lncRNAs in the rice blast fungus, as well as other plant pathogenic fungi.

Background

Large portions of many genomes are composed of considerable numbers of dark matter non-coding transcripts, which function in gene regulation [1-3]. Non-coding RNAs longer than 200 nucleotides are considered long non-coding RNAs (lncRNAs), in contrast to small non-coding RNAs, such as microRNAs and small interfering RNAs [4, 5]. Based on their genomic positions and contexts within protein-coding genes, lncRNAs are categorized as intergenic lncRNAs, antisense lncRNAs, sense lncRNAs, and intronic lncRNAs [5-7]. lncRNAs can also be classified as cis-acting lncRNAs, which regulate target genes at adjacent regions, and trans-acting lncRNAs, which function at independent chromosomal loci [8]. lncRNAs modulate the transcriptome through multiple dimensions, including epigenetic, transcriptional, post-transcriptional, translational, and post-translational levels [9].

Following the discovery of *H19* in humans and *Xist* in mice, many more lncRNAs have been functionally characterized [10, 11]. Several studies have reported that mammalian lncRNAs are associated with cell differentiation and disease process; they also serve as biomarkers for cancer diagnoses [12-14]. Plant lncRNAs, such as *COLDAIR* and *GhIncNAT-ANX2*, have roles in development and in defense against pathogens [15, 16].

Functional analysis of lncRNAs in fungi has mainly been carried out in the yeast species, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. Yeast lncRNAs modulate vegetative growth, sexual reproduction, cell-cell adhesion, and phosphate regulation [17, 18]. lncRNAs also regulate the circadian clock (*qrf*) and cellulase genes (*HAX1*) in the saprotrophic fungi *Neurospora crassa* and *Trichoderma reesei*, respectively [19-21]. lncRNA *RZE1* regulates zinc finger transcription factor *ZNF2* and affects the yeast-to-hypha transition in the human pathogenic fungus *Cryptococcus neoformans* [22]. lncRNAs have also been reported to play roles in vegetative growth (*ncRNA1*), metabolic processes (*carP*), asexual/sexual reproduction (*GzmetE-AS*), and pathogenicity (*as-um02151*) in plant pathogenic fungi [23-26]. While genome-wide profiling of lncRNAs has been performed in some fungi during vegetative growth and sexual development, the profiling of lncRNAs associated with the infection process of plant pathogenic fungi is generally incomplete and has only been studied in the rice smut fungus *Ustilagoideia virens* [27-30].

Rice blast disease is caused by the filamentous fungus *Magnaporthe oryzae*, which is responsible for an annual yield loss of 10–30% [31]. In addition to its economic importance, this fungus has served as a model of host-pathogen interactions [32]. *M. oryzae* undergoes morphological and functional transitions during vegetative growth, appressorium formation, the biotrophic stage, and the necrotrophic stage during the infection process [33]. Following the completion of whole genome sequencing of this fungus, transcriptome profiling was performed to understand gene regulation during the infection process [34-37]. However, functional and genome-wide lncRNA investigations have not been performed in *M. oryzae*.

Here, we report the genome-wide identification of lncRNAs during specific stages of infection, including vegetative growth, pre-penetration, the biotrophic stage, and the necrotrophic stage. We identified infection-specifically expressed lncRNAs (ISELs), predicted the target genes using two different methods, and predicted the functions of ISEL-associated genes. This study expands the transcriptome-level knowledge of *M. oryzae*, from protein-coding genes to long non-coding transcripts; it also provides a novel foundation for understanding the role of non-coding RNAs in host-pathogen interactions.

Results

Genome-wide identification of lncRNAs in *M. oryzae*

RNA-seq data sets from vegetative mycelia, pre-penetration, biotrophic, and necrotrophic stages were used to identify lncRNAs during mycelial growth and disease development in *M. oryzae* [37]. Previously established pipelines were used to detect lncRNAs with some modifications (Fig. 1A) [38]. In total, 436.6 million reads were mapped to the *M. oryzae* genome with 27,480 predicted transcripts originating from 16,093 genomic loci (Fig. 1B). Among these transcripts, 23,586 transcripts were detected with an FPKM > 1 in at least one developmental or infection stage and were retained for further analysis. Novel transcripts (13,978) were identified using Gffcompare categorization [41]; known mRNAs from the Ensembl database and non-coding RNAs from the Rfam database were removed [42]. Coding transcripts were filtered out by removing coding potentials of < 0.54 and the remaining transcripts were scanned by InterProScan to

remove transcripts carrying known protein domains. The resulting 2,601 lncRNA candidates were identified with a majority of antisense lncRNAs (1,286; 49.4%), intergenic lncRNAs (980; 37.7%), sense lncRNAs (322; 12.4%), and intronic lncRNAs (13; 0.5%) (Table 1, Fig. 1C). Of the identified 2,601 lncRNAs, 1,599 (61.5%) lncRNAs were expressed at all stages; 2,199, 2,183, 2,025, 2,075, 2,170, and 2,352 lncRNAs were expressed at the vegetative mycelia, 18 h post-inoculation (hpi), 27 hpi, 36 hpi, 45 hpi, and 72 hpi stages, respectively.

Genomic features of *M. oryzae* lncRNAs

Properties such as genomic distribution, exon number, length, and GC ratio of lncRNAs were investigated by mRNA comparisons. lncRNAs and mRNAs were differentially distributed across chromosomes (chi-squared test: $p = 0.01413$) (Fig. 2A); lncRNAs (mean length = 1,584 nt) had shorter full-length transcripts than did mRNAs (mean length = 2,108 nt) (Fig. 2B). lncRNAs had fewer exons than did mRNAs (Fig. 2C); a greater proportion of lncRNAs possessed one or two exons, and lncRNAs exhibited a narrower range of exon numbers. The GC ratio of lncRNA (50.1%) was lower than the GC ratio of mRNA (55.5%) (Wilcoxon–Mann–Whitney test: $p = 1.51153e-106$) (Fig. 2D).

Conservation of *M. oryzae* lncRNAs was assessed by search the results for known lncRNAs from RNAcentral [46]. Only one *M. oryzae* lncRNA (MSTRG.14004.1) was matched to human pre-miRNA (HSALNT0126175). We also compared lncRNA and mRNA sequences with genomic sequences from eight Magnaporthales species, along with *N. crassa* as an outgroup. *M. oryzae* lncRNAs were less conserved than mRNAs in all species; fewer than 10% of *M. oryzae* lncRNAs were conserved in most species, with the exception of *M. grisea* (Additional file 1: Fig. S1).

Expression of lncRNA transcripts during infection

The expression dynamics of lncRNAs were assessed by generating heatmaps based on FPKM values from the 9,410 detected mRNAs and 2,601 lncRNAs (Fig. 3A, 3B). Clustered, stage-specific expression patterns were identified for both mRNAs and lncRNAs. Mean FPKM values indicated that expression levels of lncRNAs (4.3–7.3) were much lower than expression levels of mRNAs (35.3–47.1) at the vegetative stage and all infection stages (Fig. 3C). lncRNAs showed the highest mean expression level at 45 hpi (7.3), whereas mRNAs showed the highest mean expression level at 18 hpi (47.1). We found that lncRNAs had higher expression levels in the infection stages, compared with the vegetative growth stage, suggesting that lncRNAs have a role in disease development. The evaluation of specific transcripts involved assessment of the tissue specificity index τ (Tau) [54]. The larger mean tau value for lncRNAs indicated that the expression of lncRNAs (0.69) was more stage-specific than the expression of mRNAs (0.56) (Wilcoxon–Mann–Whitney test: $p = 1.872375e-14$) (Fig. 3D).

The specificity of lncRNA expression was assessed by categorizing 518 constitutive lncRNAs ($\tau \leq 0.5$), 1,328 intermediate lncRNAs ($0.5 < \tau \leq 0.8$), and 755 specific lncRNAs ($\tau > 0.8$) based on the stage specificity index. Of the specific lncRNAs, 195 mycelia-specifically expressed lncRNAs and 560 ISEs were detected. lncRNAs identified during infection included 72 lncRNAs at the pre-penetration stage (18

hpi), 243 lncRNAs at the biotrophic stage (27–36 hpi), and 245 lncRNAs at the necrotrophic stage (45–72 hpi) (Fig. 4A, Additional file 2: Table S1).

Prediction of stage-specifically expressed lncRNA

The functional roles of lncRNAs were predicted by investigating target genes using two distinct methods. ISELs were the focus of analysis because of their biological importance during infection. In total, 157 protein-coding genes from 143 ISELs were predicted to be cis-targeted genes based on genomic proximity. Trans-targeted genes (242) were predicted from 127 ISELs based on sequence complementarity. Fifty-six ISELs and 34 target genes were found using both methods, resulting in 214 predicted ISELs and 365 predicted target genes. Biological functions were inferred by conducting GO term enrichment analysis. The most enriched GO terms of the target genes groups included “carbohydrate metabolic process” and “interaction with host” terms (Additional file 3: Table S2). The terms “binding” and “mycelium development” were enriched for the target gene set for mycelia-specific lncRNA expression (Additional file 4: Table S3).

Forty-eight of the ISEL-target pairs belonged to carbohydrate-active enzyme (CAZyme) gene families involved in carbohydrate metabolic processes. A positive correlation was found for the majority of pairs (43 of 48), which had the highest expression in the necrotrophic stage (Fig. 4B). ISEL target genes were queried against PHI-base to identify pathogenesis-related genes [53]. As a result, 23 target genes were matched to the gene set from PHI-base (Table 2). The majority of these genes were targeted by trans-acting lncRNAs, with one pair acting through both cis- and trans-regulation. The ISEL-associated genes included 5 catabolic metabolism-related genes (4 xylanases and *MoSNF1*), 2 plant avirulence determinants (*MoCDIP4*, *ACE1*), and 1 hydrophobin gene (*MPG1*).

Verification of lncRNA production

lncRNA production was verified using RNA samples from vegetative mycelia and infected rice leaves (Fig. 5, Additional file 5: Fig. S2). The infection process was covered by collecting rice leaves at 24, 48, and 72 hpi for RNA extraction. Five antisense lncRNAs and 8 intergenic lncRNAs were selected for transcript-specific RT-PCR, which can distinguish the exact transcript of interest from overlapping transcripts, including antisense transcripts and alternatively spliced transcripts. All tested lncRNAs were confirmed to be expressed in either the mycelia or during infection.

Discussion

lncRNAs modulate gene expression at the transcriptional and post-transcriptional levels; they have important roles in various metabolic pathways throughout eukaryotic species [39]. Most lncRNA studies have been performed in model yeasts, while the functional characterization and profiling of plant pathogen lncRNAs have been rarely studied [17, 18]. Genome-wide profiling of plant pathogen lncRNAs in the disease process has been performed in the rice smut fungus *U. virens* [30]. The lack of lncRNA studies during disease development limits the understanding of the role of pathogen lncRNAs during infection. In

this study, we performed comprehensive profiling of lncRNAs over several infection stages and validated their production (Fig. 5). High-throughput sequencing data yielded 437 million mapped reads, which enabled us to capture non-coding transcripts with low expression levels, as well as transcripts that were actively expressed. While some lncRNAs without a poly(A) tail may have been missed because of poly(A)-capturing library preparation, the impact was presumably minimal because of the large number of lncRNAs transcribed by RNA polymerase II [40].

M. oryzae lncRNAs had shorter transcript length, fewer exons, lower GC ratios, and temporal-specific expression patterns, suggesting that functional lncRNAs exist in *M. oryzae*, because these features were observed in multiple eukaryotic organisms (Fig. 2, Fig. 3) [40]. The roles of lncRNAs are presumed to depend on the protein-coding genes with which they interact. Therefore, the prediction of lncRNA function depends on target gene prediction. Functional characterization of lncRNAs has revealed that both cis- and trans-acting lncRNAs have roles in gene regulation [21, 25]. However, previous fungal lncRNA profiling studies considered only cis-acting lncRNAs [29, 30]. Here, we performed target gene prediction for both cis- and trans-acting lncRNAs; we found more trans-acting lncRNA target genes than cis-acting lncRNA target genes. This extended prediction of target genes enabled us to identify a pool of unbiased lncRNA-associated genes that await further functional characterization of infection-related lncRNAs.

The mean level of lncRNA expression increased for all infection stages, compared with the vegetative growth stage, and a stage-specific pattern was observed. In this study, tau value was used to identify lncRNAs highly expressed only in particular infection stages, providing an accurate stage-specifically expressed lncRNAs. As expected, we identified more ISEs than mycelia-specifically expressed lncRNAs. Increased expression levels of lncRNAs during the developmental process were also observed in *F. graminearum* sexual reproduction and *U. virens* disease development [29, 30]. Our findings and other observations suggest that lncRNAs have roles in the pathogenesis of plant pathogenic fungi.

GO term enrichment analysis revealed that terms related to carbohydrate metabolism were enriched in ISEL-associated genes in *M. oryzae* (Additional file 3: Table S2). In *U. virens*, transport-related GO terms were enriched during all stages [30]. This difference may be relevant to the distinct lifestyles of biotrophs (*U. virens*) and hemibiotrophs (*M. oryzae*), although both species infect the same host. PHI-based analysis showed that *M. oryzae* lncRNAs may target genes encoding CAZymes, including plant cell wall-degrading enzymes (PCWDEs) (Fig. 4, Table 2). Notably, PCWDEs play important roles in rice blast disease progression by helping to overcome the physical barrier complex composed of cellulose, hemicellulose, pectin, lignin, and xylan [41]. A cellulase-regulating lncRNA was reported in the saprophyte *T. reesei*, where cellulases are essential for trophism [20]. Effectors such as *ACE1* and *MoCDIP4* were also found in *M. oryzae* lncRNA-associated genes. Effectors secreted from the pathogen act as major virulence determinants [42]. Taken together, the findings thus far suggest that lncRNAs function in the pathogenesis of *M. oryzae* by regulating associated genes.

Conclusions

In summary, this study reports the first genome-wide lncRNA profile in the model fungal pathogen, *M. oryzae*. The profiling of infection-specific lncRNAs and their associated genes suggest that lncRNA may regulate the infection process. Overall, this study provides extensive profiling of lncRNAs and the associated gene repertoire; it also demonstrates the potential roles of lncRNAs involved in rice blast disease development.

Methods

RNA extraction and strand-specific sequencing

M. oryzae strain KJ201 was obtained from the Center for Fungal Genetic Resources at Seoul National University (Seoul, Korea). Fungal mycelia were cultured with shaking (150 rpm) in a liquid complete medium (0.6% yeast extract, 0.6% tryptone, and 1% sucrose [w/v]) at 25°C for 3 days. Total RNA was extracted using an Easy-spin total RNA extraction kit (iNtRON Biotechnology, Seoul, Korea), in accordance with the manufacturer's instructions. Strand-specific cDNA synthesis and sequencing were performed at the National Instrumentation Center for Environmental Management at Seoul National University (Seoul, Korea). Shotgun sequencing was used to generate 75.3 million paired-end 151-bp reads using an Illumina HiSeq 2500.

Collection of *in planta* RNA-seq data

Six *M. oryzae* KJ201 RNA-seq libraries, including different infection stages of rice sheath, were used to identify lncRNA during mycelial growth and disease development (Jeon et al., 2020) (SRA accession no. SRX5076910- SRX5076915). The RNA-seq data contained paired-end 101-bp reads and included the following stages: vegetative mycelia, pre-penetration stage (18 hpi), biotrophic stage (27 and 36 hpi), and necrotrophic stage (45 and 72 hpi). These stages included appressorium formation (pre-penetration, 18 hpi), penetration and development of primary invasive hyphae (biotrophic stage, 27 hpi), development and growth of invasive hyphae (biotrophic stage, 36 hpi), active growth of invasive hyphae into neighboring host cells (necrotrophic stage, 45 hpi), and extensive proliferation and killing of host cells (necrotrophic stage, 72 hpi).

Transcriptome assembly

Raw reads were processed to remove low-quality reads and trim adapter sequences using NGS QC Toolkit v2.3.3 [43]. The resulting reads were mapped against the *M. oryzae* reference genome (MG8, Ensembl annotation 29) using HISAT2 v2.0.4 [32, 44]. The transcriptome was assembled using the genome-guided method of StringTie v1.3.3 with *de novo* annotation [45]. We used fragments per kilobase of transcript per million mapped read pairs (FPKM) as the expression value. If the expression value for a transcript was < 1 FPKM at all stages, the transcript was considered to be predicted, but not detected. Detected transcripts were used for subsequent analysis.

Pipeline for lncRNA identification

We used an established computational pipeline to identify lncRNAs. Transcripts shorter than 200 nucleotides were first filtered out. The assembled transcripts were then compared with protein-coding genes and categorized using Gffcompare [46]. We regarded antisense transcripts (class code “x”), sense transcripts (class codes “j” and “o”), intronic transcripts (class code “i”), and intergenic transcripts (class codes “u” and “p”) as novel transcripts. Known non-coding RNAs (tRNAs, rRNAs, snRNAs, and snoRNAs) were removed using Infernal v1.1.1 based on Rfam database release 14.0 [47, 48]. The coding potentials of transcripts were assessed using CPAT v.1.2.2 [49]. To maximize lncRNA detection, training was performed and the coding potential cutoff was set to 0.54 (Additional file 6: Fig. S3). Transcripts with coding potential below the cutoff were included; transcripts containing any known Pfam domain were removed using InterProScan version 5.29-68.0 [50].

lncRNA conservation analysis

The 2,601 *M. oryzae* lncRNAs identified in this study were BLAST searched against known lncRNAs downloaded from RNACentral with an E-value cutoff of 1e-5 [51]. The level of conservation between *M. oryzae* lncRNAs and other Magnaporthales species was assessed by BLAST searching predicted *M. oryzae* lncRNAs and annotated mRNAs against the genomes of eight Magnaporthales species (*Magnaporthe grisea*, *Gaeumannomyces graminis*, *Magnaporthe poae*, *Magnaportheopsis rhizophila*, *Magnaportheopsis incrustans*, *Magnaporthe salvinii*, *Ophioceras dolichostomum*, *Pseudohalonestria lignicola*), as well as *Neurospora crassa* as an outgroup, with an E-value cutoff of 1e-5. The genomes of *M. grisea*, *G. graminis*, *M. poae*, and *N. crassa* were obtained from the Comparative Fungal Genomics Platform (<http://cfgp.riceblast.snu.ac.kr>) [52]. The genomes of *M. rhizophila*, *M. incrustans*, *M. salvinii*, *O. dolichostomum*, and *P. lignicola* were downloaded from the National Center for Biotechnology Information [53].

Assessment of stage specificity and prediction of stage-specific lncRNAs

The stage specificities of transcripts were determined using the tissue specificity index as described previously [54].

$$\tau = \frac{\sum_{i=1}^n (1 - \hat{x}_i)}{n - 1}; \hat{x}_i = \frac{x_i}{\max_{1 \leq i \leq n} x_i}$$

where n is the number of stages and x_i is the expression level at stage i. The index varies from 0 (consistently expressed transcripts) to 1 (perfectly stage-specific transcripts).

Stage-specific lncRNAs were selected based on the following criteria: $\tau > 0.8$ and (FPKM of stage with the highest expression)/(FPKM of stage with the second highest expression) > 2 . lncRNAs with expression during the first and second peaks of the biotrophic stages were considered biotrophic stage-

specific lncRNAs; lncRNAs with expression during both the first and second peaks of the necrotrophic stages were considered necrotrophic stage-specific lncRNAs.

Target gene prediction

Protein-coding genes co-expressed with lncRNAs were identified using Pearson correlation coefficients, which were calculated between each mRNA–lncRNA pair based on expression values. Genes with an absolute value of coefficient > 0.9 were considered to be co-expressed. For these genes, possible target genes for cis- or trans-regulation were predicted using two independent criteria. For cis-target gene prediction, genes within a 10-kb window upstream or downstream of the lncRNAs were considered. For trans-target gene prediction, transcript sequence complementarity and RNA duplex energy were used to assess the impact of lncRNA binding on mRNA molecules using RNAplex (parameter: $1e-60$) [55]. Target genes were then subjected to Gene Ontology (GO) term enrichment analysis at a 5% false discovery rate using Blast2GO and AgriGO v2.0 [56, 57]. Pathogenesis-related genes were identified by querying target genes against a pathogen-host interactions database (PHI-base) [58].

Validation of lncRNA transcript production

The validation of lncRNA production was measured on the basis of lncRNA expression during vegetative mycelia and infection stages using strand-specific reverse transcription PCR (RT-PCR). Rice cultivar Nakdong was grown in a growth chamber at 28°C and 80% humidity with a 16/8-h light/dark photoperiod. Four-week-old rice seedlings were inoculated with *M. oryzae* KJ201 conidial suspension with 20×10^4 conidia/mL in 250 ppm Tween 20 using a sprayer. The inoculated plants were incubated for 24 hpi, 48 hpi, and 72 hpi. cDNA was synthesized using ImProm-II™ Reverse Transcription System (Promega, Madison, WI, USA), in accordance with the manufacturer's instructions. For strand-specific reverse transcription, transcript-specific primers were designed as previously reported [59]. Reverse transcription reactions were carried out with 200 ng of total RNA, 1 µl of 4 pmol/µl of transcript-specific primers, 2 µl of synthesized cDNA, and 1 µl of 10 pmol/µl nested primers, which were designed to amplify only the synthesized cDNA. I-star-max II PCR master mix was added for a total volume per reaction of 10 µl. Primers used in all RT-PCR experiments are listed in Table S4 (Additional file 7).

Abbreviations

lncRNA: Long non-coding RNA

ISEL: Infection-specifically expressed lncRNA

FPKM: Fragments per kilobase of transcript per million mapped read pairs

GO: Gene Ontology

CAZyme: Carbohydrate-active enzyme

PCWDE: Plant cell wall-degrading enzyme

Declarations

Acknowledgements

We thank Hyeunjeong Song for figure editing

Author's contribution

GC and YHL designed the study. GC, and JJ performed bioinformatics analysis. HL, and SZ performed associated biological experiments. GC, JJ, HL, and YHL wrote the manuscript. All authors read and approved the final manuscript

Funding

This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Korea government (MSIT) (2020R1A2B5B03096402 and 2018R1A5A1023599) and Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, and Forestry through Agricultural Microbiome Program (918017-04-1-CG000). SZ is grateful for a graduate fellowship through the Brain Korea 21 Plus Program. These funding bodies had no role in the study design, data collection, analysis, and preparation of the manuscript.

Availability of data and materials

All the data supporting our findings are contained within the manuscript. All raw transcriptome data reported in this article have been deposited in the Sequence Read Archive (SRA) under accession number SRP332970.

The used datasets include:

The genome of *M. oryzae* in Ensembl (MG8)

(https://fungi.ensembl.org/Magnaporthe_oryzae/Info/Annotation/);

The genomes of *M. grisea*, *G. graminis*, *M. poae*, and *N. crassa* in CFGP (BR29, R3-111a-1, ATCC 64411, OR74A)

(<http://cfgp.riceblast.snu.ac.kr>);

The genomes of *M. rhizophila*, *M. incrustans*, *M. salvinii*, *O. dolichostomum*, and *P. lignicola* in NCBI Assembly (GCA_003049465.1, GCA_003049425.1, GCA_003049435.1, GCA_003049485.1, GCA_003049395.1)

(<https://www.ncbi.nlm.nih.gov/assembly/>);

Rfam 14.0 in Rfam database (<https://rfam.xfam.org/>);

Pfam 31.0 in InterPro (<https://www.ebi.ac.uk/interpro/>);

Raw transcriptome data in NCBI SRA (SRX5076910-SRX5076915)

(<https://www.ncbi.nlm.nih.gov/sra/>);

Public access to all databases is open.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Kapranov P, Willingham AT, Gingeras TR. Genome-wide transcription and the implications for genomic organization. *Nat Rev Genet.* 2007;8:413-23.
2. Clark MB, Amaral PP, Schlesinger FJ, Dinger ME, Taft RJ, Rinn JL, et al. The reality of pervasive transcription. *PLoS Biol.* 2011;9:e1000625.
3. Wade JT, Grainger DC. Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. *Nat Rev Microbiol.* 2014;12:647-53.
4. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet.* 2009;10:155-9.
5. Ma L, Bajic VB, Zhang Z. On the classification of long non-coding RNAs. *RNA Biol.* 2013;10:924-33.
6. Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. *Cell.* 2009;136:629-41.
7. Laurent GS, Wahlestedt C, Kapranov P. The Landscape of long noncoding RNA classification. *Trends Genet.* 2015;31:239-51.
8. Gil N, Ulitsky I. Regulation of gene expression by cis-acting long non-coding RNAs. *Nat Rev Genet.* 2020;21:102-17.

9. Zhang X, Wang W, Zhu W, Dong J, Cheng Y, Yin Z, et al. Mechanisms and functions of long non-coding RNAs at multiple regulatory levels. *Int J Mol Sci.* 2019;20:5573.
10. Brannan CI, Dees EC, Ingram RS, Tilghman SM. The product of the H19 gene may function as an RNA. *Mol Cell Biol.* 1990;10:28-36.
11. Brockdorff N, Ashworth A, Kay GF, McCabe VM, Norris DP, Cooper PJ, et al. The product of the mouse *Xist* gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell.* 1992;71:515-26.
12. Li L, Wang M, Wu X, Geng L, Xue Y, Wei X, et al. A long non-coding RNA interacts with Gfra1 and maintains survival of mouse spermatogonial stem cells. *Cell Death Dis.* 2016;7:e2140-e.
13. Jalali S, Jayaraj GG, Scaria V. Integrative transcriptome analysis suggest processing of a subset of long non-coding RNAs to small RNAs. *Biol Direct.* 2012;7:1-13.
14. Wang W, Zhuang Q, Ji K, Wen B, Lin P, Zhao Y, et al. Identification of miRNA, lncRNA and mRNA-associated ceRNA networks and potential biomarker for MELAS with mitochondrial DNA A3243G mutation. *Sci Rep.* 2017;7:1-13.
15. Zhang YC, Chen YQ. Long noncoding RNAs: new regulators in plant development. *Biochem Biophys Res Commun.* 2013;436:111-4.
16. Zaynab M, Fatima M, Abbas S, Umair M, Sharif Y, Raza MA. Long non-coding RNAs as molecular players in plant defense against pathogens. *Microb Pathog.* 2018;121:277-82.
17. Till P, Mach RL, Mach-Aigner AR. A current view on long noncoding RNAs in yeast and filamentous fungi. *Appl Microbiol Biotechnol.* 2018;102:7319-31.
18. Li J, Liu X, Yin Z, Hu Z, Zhang KQ. An overview on identification and regulatory mechanisms of long non-coding RNAs in fungi. *Front Microbiol.* 2021;12:638617.
19. Kramer C, Loros JJ, Dunlap JC, Crosthwaite SK. Role for antisense RNA in regulating circadian clock function in *Neurospora crassa*. *Nature.* 2003;421:948-52.
20. Till P, Pucher ME, Mach RL, Mach-Aigner AR. A long noncoding RNA promotes cellulase expression in *Trichoderma reesei*. *Biotechnol Biofuels.* 2018;11:1-16.
21. Till P, Derntl C, Kiesenhofer DP, Mach RL, Yaver D, Mach-Aigner AR. Regulation of gene expression by the action of a fungal lncRNA on a transactivator. *RNA Biol.* 2020;17:47-61.
22. Chacko N, Zhao Y, Yang E, Wang L, Cai JJ, Lin X. The lncRNA *RZE1* controls cryptococcal morphological transition. *PLoS Genet.* 2015;11:e1005692.

23. Morrison EN, Donaldson ME, Saville BJ. Identification and analysis of genes expressed in the *Ustilago maydis* dikaryon: uncovering a novel class of pathogenesis genes. *Can J Plant Pathol*. 2012;34:417-35.
24. Parra-Rivero O, Pardo-Medina J, Gutiérrez G, Limón MC, Avalos J. A novel lncRNA as a positive regulator of carotenoid biosynthesis in *Fusarium*. *Sci Rep*. 2020;10:1-14.
25. Wang J, Zeng W, Xie J, Fu Y, Jiang D, Lin Y, et al. A novel antisense long noncoding RNA participates in asexual and sexual reproduction by regulating the expression of *GzmetE* in *Fusarium graminearum*. *Environ Microbiol*. 2021; doi:10.1111/1462-2920.15399.
26. Donaldson ME, Saville BJ. *Ustilago maydis* natural antisense transcript expression alters mRNA stability and pathogenesis. *Mol Microbiol*. 2013;89:29-51.
27. Arthanari Y, Heintzen C, Griffiths-Jones S, Crosthwaite SK. Natural antisense transcripts and long non-coding RNA in *Neurospora crassa*. *PLoS ONE*. 2014;9:e91353.
28. Donaldson ME, Ostrowski LA, Goulet KM, Saville BJ. Transcriptome analysis of smut fungi reveals widespread intergenic transcription and conserved antisense transcript expression. *BMC Genomics*. 2017;18:1-14.
29. Kim W, Miguel-Rojas C, Wang J, Townsend JP, Trail F. Developmental dynamics of long noncoding RNA expression during sexual fruiting body formation in *Fusarium graminearum*. *mBio*. 2018;9:e01292-18.
30. Tang J, Chen X, Yan Y, Huang J, Luo C, Tom H, et al. Comprehensive transcriptome profiling reveals abundant long non-coding RNAs associated with development of the rice false smut fungus, *Ustilaginoidea virens*. *Environ Microbiol*. 2021; doi:10.1111/1462-2920.15432.
31. Skamnioti P, Gurr SJ. Against the grain: safeguarding rice from rice blast disease. *Trends Biotechnol*. 2009;27:141-50.
32. Dean R, Van Kan JA, Pretorius ZA, Hammond-Kosack KE, Di Pietro A, Spanu PD, et al. The Top 10 fungal pathogens in molecular plant pathology. *Mol Plant Pathol*. 2012;13:414-30.
33. Fernandez J, Orth K. Rise of a cereal killer: the biology of *Magnaporthe oryzae* biotrophic growth. *Trends Microbiol*. 2018;26:582-97.
34. Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, et al. The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature*. 2005;434:980-6.
35. Kawahara Y, Oono Y, Kanamori H, Matsumoto T, Itoh T, Minami E. Simultaneous RNA-seq analysis of a mixed transcriptome of rice and blast fungus interaction. *PLoS ONE*. 2012;7:e49423.

36. Dong Y, Li Y, Zhao M, Jing M, Liu X, Liu M, et al. Global genome and transcriptome analyses of *Magnaporthe oryzae* epidemic isolate 98-06 uncover novel effectors and pathogenicity-related genes, revealing gene gain and lose dynamics in genome evolution. PLoS Pathog. 2015;11:e1004801.
37. Jeon J, Lee GW, Kim KT, Park SY, Kim S, Kwon S, et al. Transcriptome profiling of the rice blast fungus *Magnaporthe oryzae* and its host *Oryza sativa* during infection. Mol Plant Microbe Interact. 2020;33:141-4.
38. Weirick T, Militello G, Müller R, John D, Dimmeler S, Uchida S. The identification and characterization of novel transcripts from RNA-seq data. Brief Bioinform. 2016;17:678-85.
39. Marchese FP, Raimondi I, Huarte M. The multidimensional mechanisms of long noncoding RNA function. Genome Biol. 2017;18:1-13.
40. Quinn JJ, Chang HY. Unique features of long non-coding RNA biogenesis and function. Nat Rev Genet. 2016;17:47.
41. Quoc NB, Bao Chau NN. The role of cell wall degrading enzymes in pathogenesis of *Magnaporthe oryzae*. Curr Protein Pept Sci. 2017;18:1019-34.
42. Dodds PN, Rathjen JP. Plant immunity: towards an integrated view of plant–pathogen interactions. Nat Rev Genet. 2010;11:539-48.
43. Patel RK, Jain M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLoS One. 2012;7:e30619.
44. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12:357-60.
45. Perteau M, Perteau GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33:290-5.
46. Perteau M, Kim D, Perteau GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc. 2016;11:1650.
47. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 2013;29:2933-5.
48. Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, et al. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. Nucleic Acids Res. 2018;46:D335-42.
49. Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res. 2013;41:e74.

50. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 2014;30:1236-40.
51. Consortium R. RNAcentral 2021: secondary structure integration, improved sequence search and new member databases. *Nucleic Acids Res*. 2021;49:D212-20.
52. Choi J, Cheong K, Jung K, Jeon J, Lee GW, Kang S, et al. CFGP 2.0: a versatile web-based platform for supporting comparative and evolutionary genomics of fungi and Oomycetes. *Nucleic Acids Res*. 2013;41:D714-9.
53. Zhang N, Cai G, Price DC, Crouch JA, Gladieux P, Hillman B, et al. Genome wide analysis of the transition to pathogenic lifestyles in Magnaporthales fungi. *Sci Rep*. 2018;8:1-13.
54. Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. *Bioinformatics*. 2005;21:650-9.
55. Tafer H, Hofacker IL. RNAplex: a fast tool for RNA–RNA interaction search. *Bioinformatics*. 2008;24:2657-63.
56. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res*. 2008;36:3420-35.
57. Tian T, Liu Y, Yan H, You Q, Yi X, Du Z, et al. agriGO v2. 0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res*. 2017;45:W122-9.
58. Urban M, Cuzick A, Seager J, Wood V, Rutherford K, Venkatesh SY, et al. PHI-base: the pathogen–host interactions database. *Nucleic Acids Res*. 2020;48:D613-20.
59. Ho EC, Donaldson ME, Saville BJ. Detection of antisense RNA transcripts by strand-specific RT-PCR. *RT-PCR Protocols*: Springer; 2010. p. 125-38.

Tables

Table 1. Classification of lncRNAs in *M. oryzae*

Class of transcripts	Number of novel transcripts	Number of lncRNAs
Sense transcript	8,444	322
Antisense transcript	2,636	1,286
Intergenic transcript	2,876	980
Intronic transcript	22	13

Table 2. Target genes of infection specifically-expressed lncRNAs matched to genes from PHI-base

ISEL	Mode of action	Target gene	Description
MSTRG.14853.1	Trans	MoCDIP4	Plant cell death inducer
MSTRG.8963.1	Trans		
MSTRG.14634.1	Trans	ACE1	Polyketide synthase
MSTRG.10882.1	Trans	MoSNF1	AMP-activated protein kinase
MSTRG.5151.2	Cis	MET12	Methylenetetrahydrofolate reductase
MSTRG.12783.1	Trans	MoSOM1	Transcriptional regulator
MSTRG.14270.3	Cis/trans	MoSSK1	Response regulator
MSTRG.1779.3	Cis	MoCOD1	Zn ₂ Cys ₆ transcription factor
MSTRG.8913.3	Trans	MoPER1	GPI anchored-related gene
MSTRG.14270.1	Trans	MGG_08331T0	Endo-1,4-beta-xylanase
MSTRG.14270.2	Trans		
MSTRG.14853.1	Trans		
MSTRG.4487.2	Trans		
MSTRG.4487.3	Trans		
MSTRG.8648.1	Trans		
MSTRG.8648.2	Trans		
MSTRG.8648.1	Trans	MGG_08424T0	Endo-1,4-beta-xylanase
MSTRG.8648.2	Trans		
MSTRG.8655.2	Trans		
MSTRG.14853.1	Trans	MPG1	Hydrophobin
MSTRG.8407.2	Trans		
MSTRG.8648.1	Trans		
MSTRG.8648.2	Trans		
MSTRG.14853.1	Trans	MGG_10730T0	Na ⁺ -ATPase
MSTRG.13745.1	Trans	MoLDS1	Animal peroxidase
MSTRG.2819.1	Trans	SSM2	Non-ribosomal peptide synthetase
MSTRG.12783.1	Trans	MGG_15019T0	Peroxisomal copper amine oxidase
MSTRG.10882.1	Cis	MoRGS4	G-protein signaling regulator

MSTRG.8407.2	Cis	Pmc1	Vacuolar membrane-located Ca ²⁺ pump
MSTRG.13998.6	Trans		
MSTRG.1930.1	Trans	MST12	STE-like transcription factor
MSTRG.1930.3	Trans		
MSTRG.8389.1	Cis	XYL1	Endo-1,4-beta-xylanase
MSTRG.13915.1	Trans	XYL-6	Endo-1,4-beta-xylanase
MSTRG.10882.1	Trans	FZC87	Zn ₂ Cys ₆ transcription factor
MSTRG.14853.1	Trans	FZC12	Zn ₂ Cys ₆ transcription factor
MSTRG.1930.1	Trans	FZC42	Zn ₂ Cys ₆ transcription factor
MSTRG.1930.3	Trans		

Figures

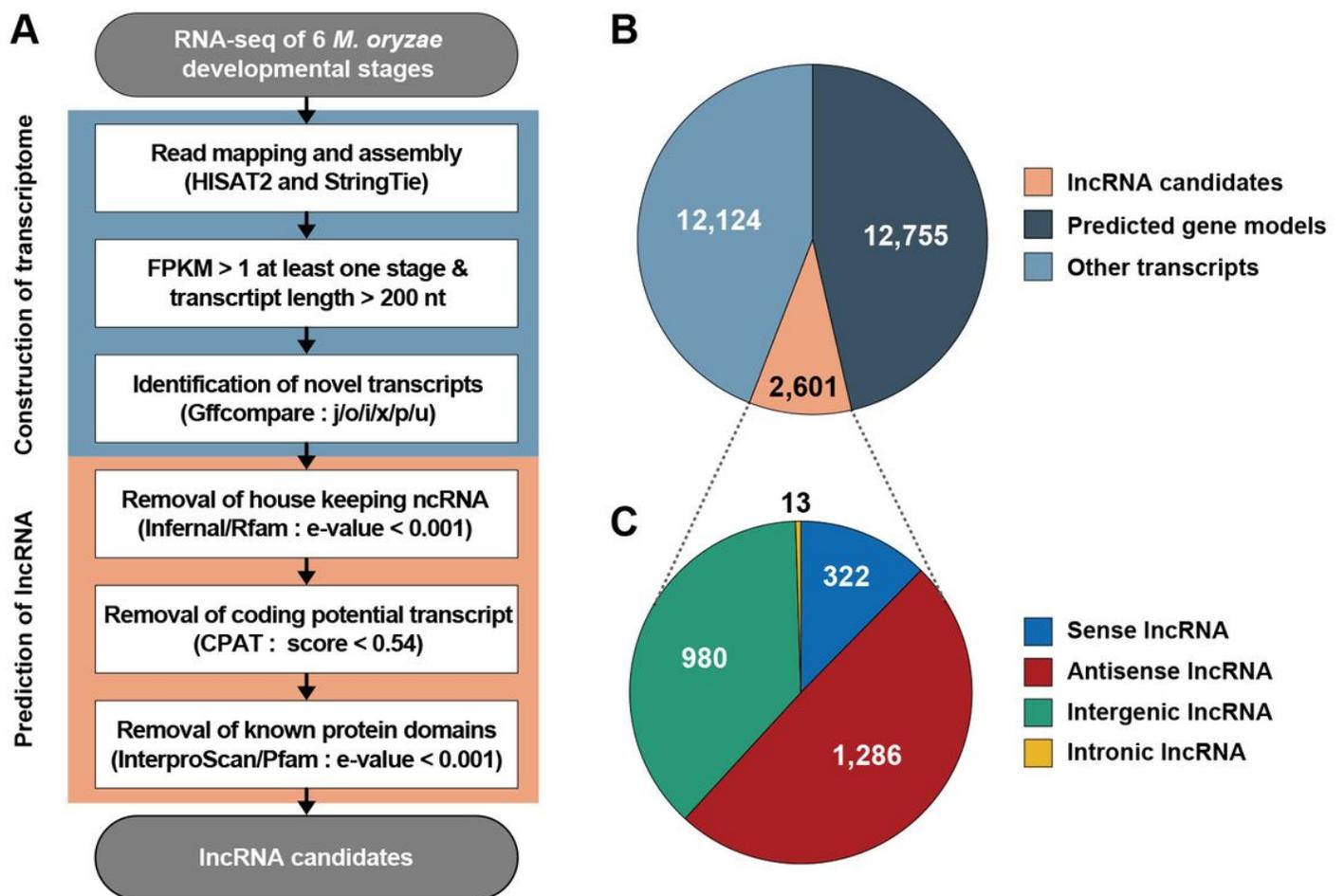


Figure 1

Schematic pipeline for identification of lncRNAs in *M. oryzae*. (A) Bioinformatic pipeline for lncRNA identification using RNA-seq data. CPAT, Coding Potential Assessment Tool. (B) Number of predicted transcripts. (C) Number of lncRNAs by different classes.

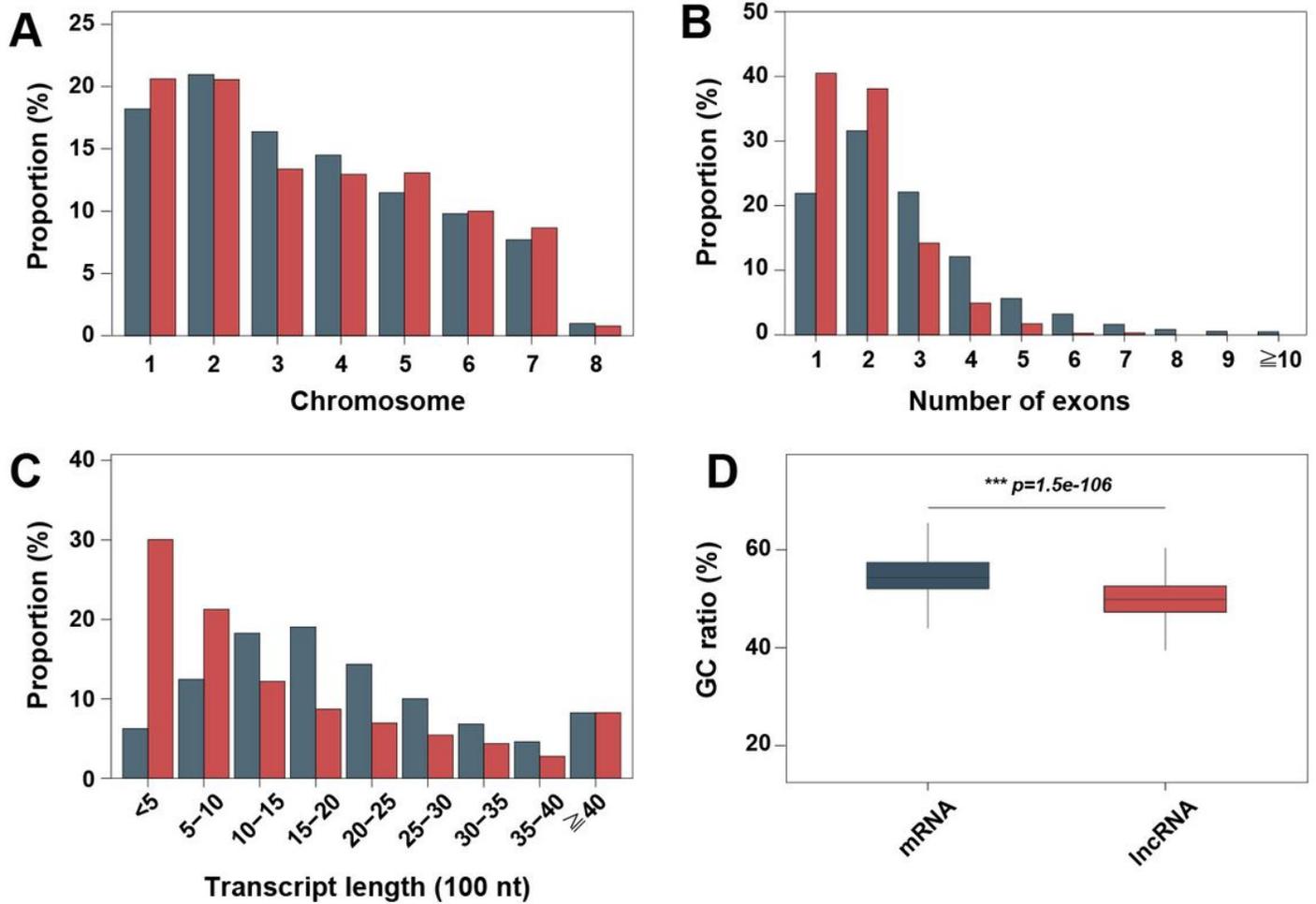


Figure 2

Genomic features of *M. oryzae* lncRNAs. (A) Distributions of mRNAs (bluish green) and lncRNAs (red) across chromosomes. (B) Distribution of transcript lengths. (C) Distribution of exon numbers per transcript. (D) GC ratio (%). The Wilcoxon–Mann–Whitney test confirmed a significant difference in GC ratio between the two groups. *** $p < 0.001$

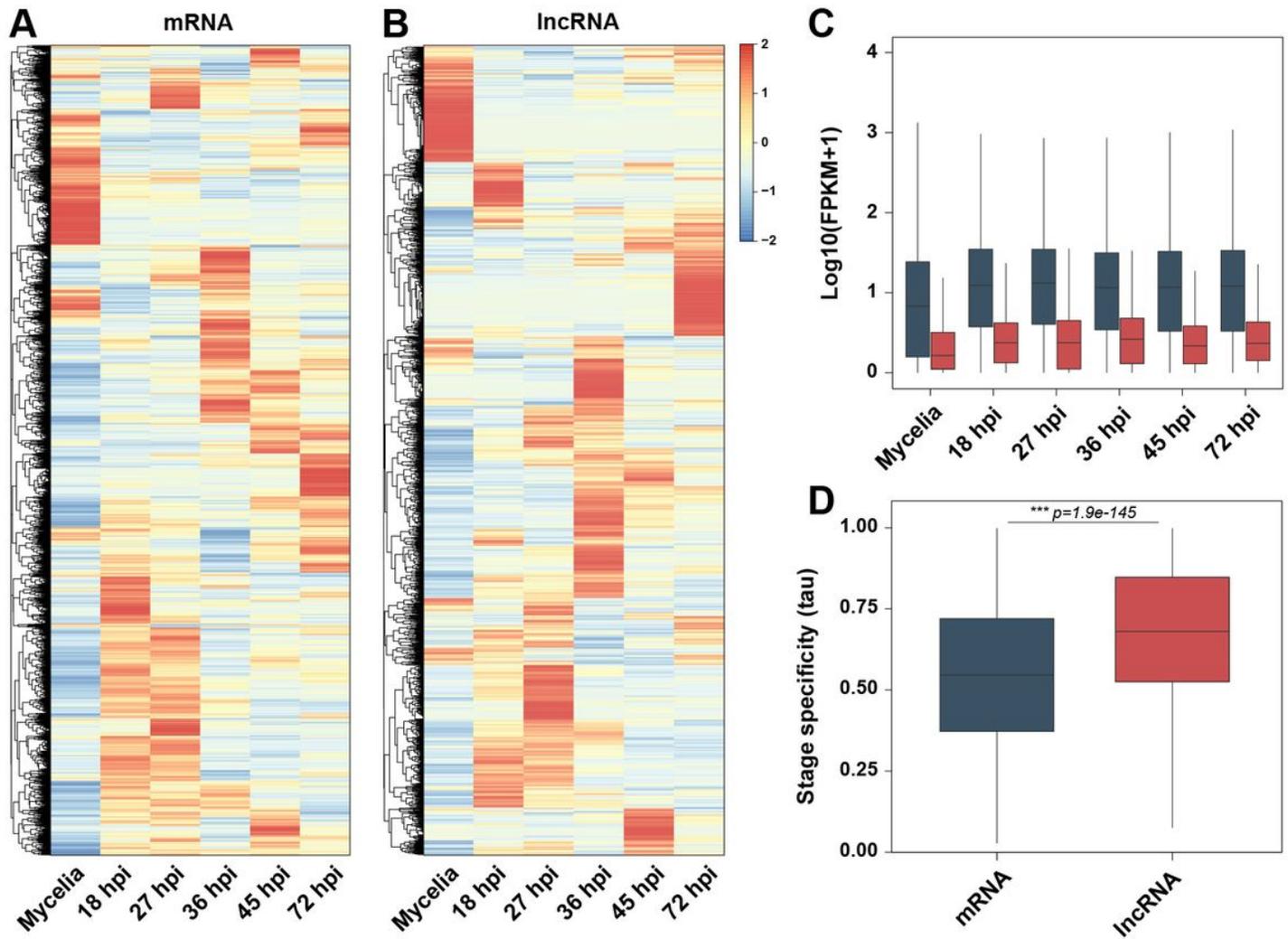


Figure 3

LncRNA expression level and pattern. (A), (B) Expression heatmaps of 9,410 mRNAs and 2,601 lncRNAs, respectively. Z-score normalization was applied to FPKM values across stages. (C) Boxplot of mRNA (bluish green) and lncRNA (red) expression patterns across developmental and infection stages. (D) Density plot of transcript stage specificity over six stages. τ (Tau) is used as a stage specificity index. The index varies from 0 (consistently expressed transcripts) to 1 (perfectly stage-specific transcripts). The Wilcoxon–Mann–Whitney test confirmed a significant difference in tau distribution between the two groups. *** $p < 0.001$

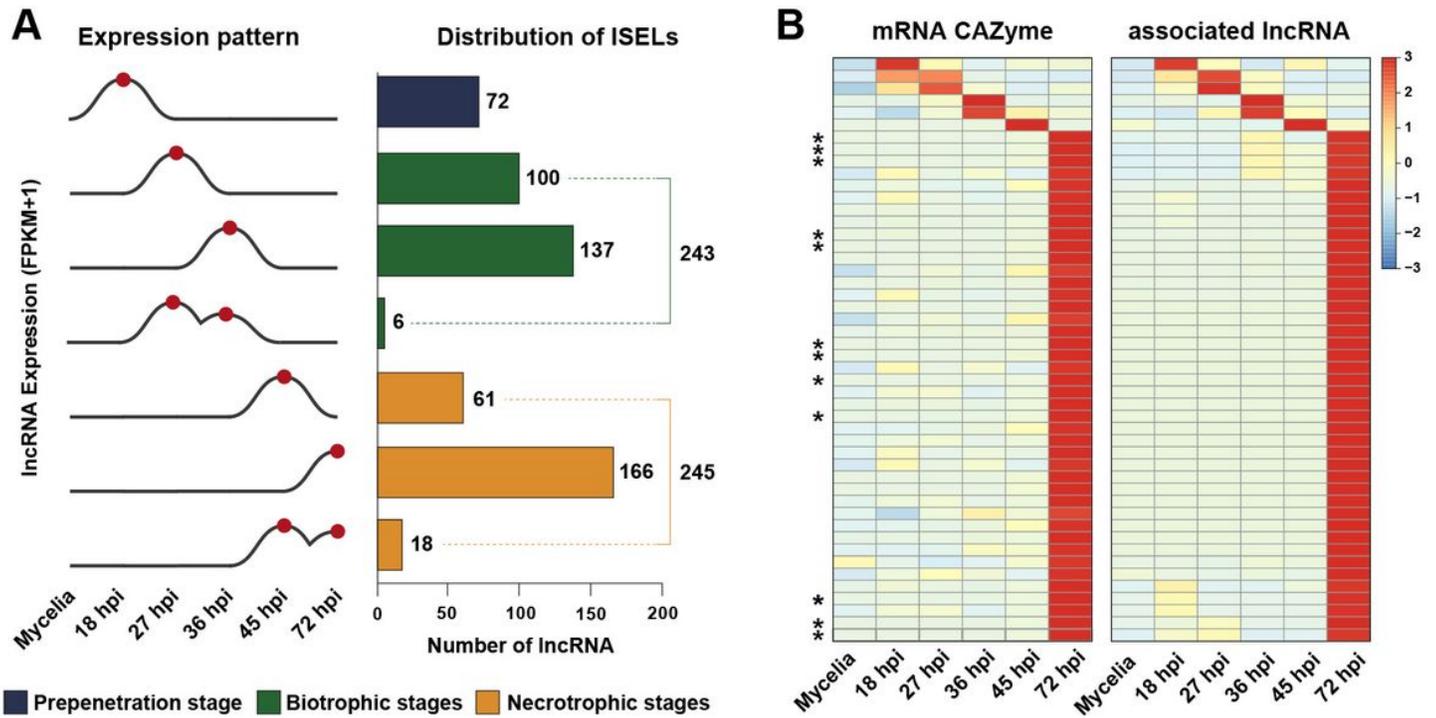


Figure 4

Infection stage-specific lncRNAs and their target genes. (A) Numbers of lncRNAs according to expression pattern. Red spots indicate an expression peak at each stage. (B) Expression heatmap of infection-specifically expressed lncRNAs and their CAZyme target genes. FPKM values were normalized across conditions based on Z-scores. Normalizations of mRNAs and lncRNAs were performed separately. * indicates a functionally characterized xylanase. CAZyme, Carbohydrate-active enzyme.

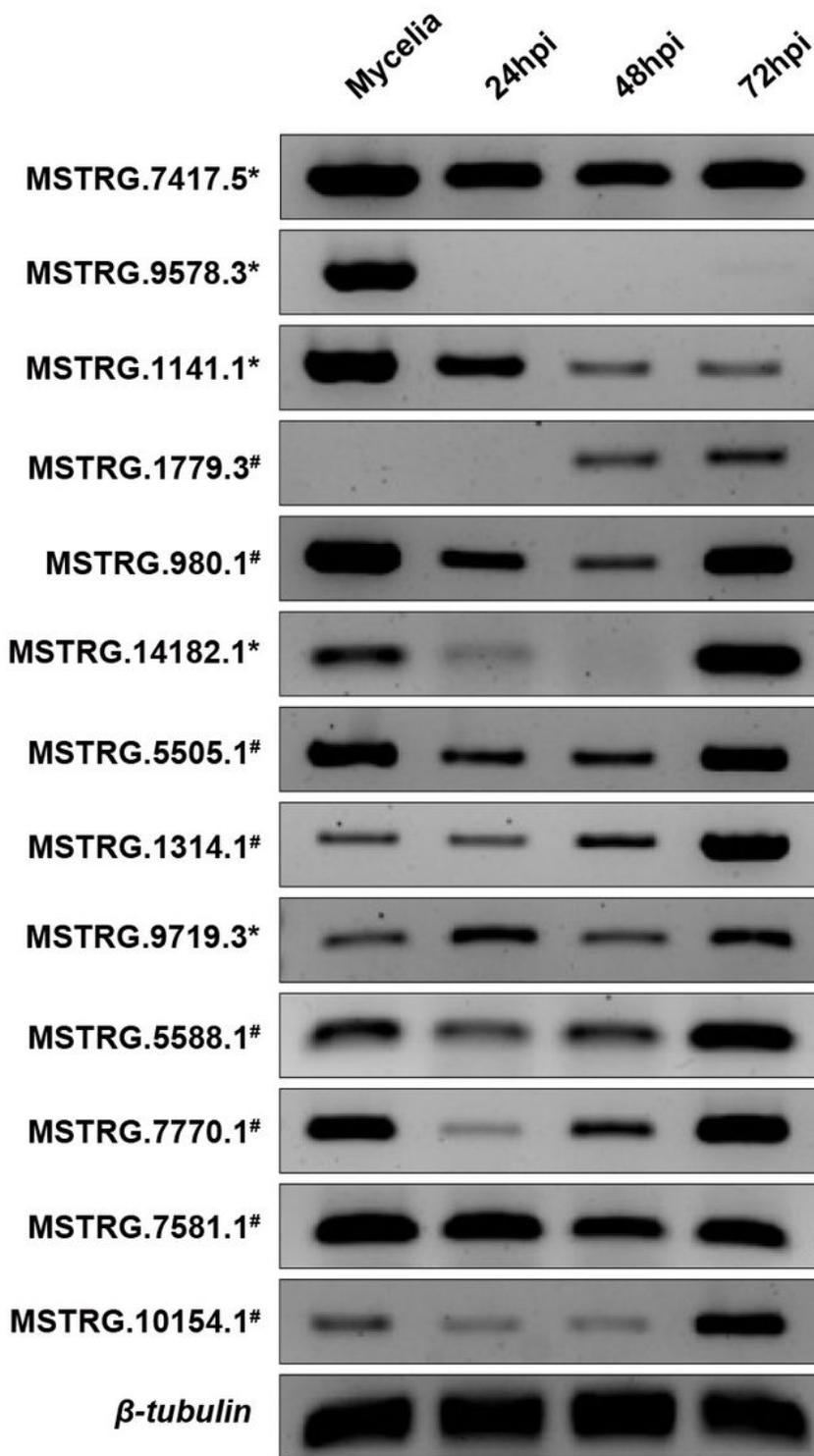


Figure 5

Validation of lncRNA production. Validation of lncRNAs was performed with strand-specific RT-PCR. Templates were cDNAs synthesized from the RNA of mycelia, 24 h post-inoculation (hpi), 48 hpi, and 72 hpi on infected rice leaves. * indicates an antisense lncRNA. # indicates an intergenic lncRNA. β -tubulin gene was used as a control. Full-length gels are presented in Fig. S2 (Additional file 5).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.docx](#)
- [Additionalfile2.xlsx](#)
- [Additionalfile3.xlsx](#)
- [Additionalfile4.xlsx](#)
- [Additionalfile5.docx](#)
- [Additionalfile6.docx](#)
- [Additionalfile7.xlsx](#)