농학박사 학위논문

벼 도열병균의 긴 비암호화 리보핵산 분석 및 짧은 비암호화 리보핵산과의 상호작용

Genome-wide analysis of long non-coding RNAs and their interaction with small RNAs in *Magnaporthe oryzae*

2023년 2월

서울대학교 대학원

협동과정 농생명유전체학전공

최 고 봉

Genome-wide analysis of long non-coding RNAs and their interaction with small RNAs in *Magnaporthe oryzae*

A dissertation submitted in partial
fulfillment of the requirement for
the degree of

DOCTOR OF PHILOSOPHY

to the Faculty of
Interdisciplinary Program in Agricultural Genomics
at

SEOUL NATIONAL UNIVERSITY

by

Gobong Choi

FEBRUARY 2023

농학박사 학위논문

벼 도열병균의 긴 비암호화 리보핵산 분석 및 짧은 비암호화 리보핵산과의 상호작용

지도교수 이 용 환

이 논문을 농학박사 학위논문으로 제출함
2022년 12월

서울대학교 대학원
협동과정 농생명유전체학전공

최 고 봉

최고봉의 박사 학위논문을 인준함

2022년 12월

위 원 장 최 도 일 (인)

부 위 원 장 이 용 환 (인)

위 원 김 광 형 (인)

위 원 김 기 태 (인)

위 원 정 경 채 (인)

A THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Genome-wide analysis of long non-coding RNAs and their interaction with small RNAs in *Magnaporthe oryzae*

UNDER THE DIRECTION OF DR. YONG-HWAN LEE

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF SEOUL NATIONAL UNIVERSITY

BY

GOBONG CHOI

INTERDISCIPLINARY PROGRAM IN AGRICULTURAL GENOMICS

DECEMBER 2022

APPROVED AS A QUALIFIED THESIS OF GOBONG CHOI
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
BY THE COMMITTEE MEMBERS

CHAIRMAN	Doil Choi
VICE CHAIRMAN	Yong-Hwan Lee
MEMBER	Kwang-Hyung Kim
MEMBER	Ki-Tae Kim
MEMBER	Kyeongchae Cheong

ABSTRACT

Genome-wide long non-coding RNAs and their interaction with small RNAs in *Magnaporthe oryzae*

Gobong Choi
Interdisciplinary Program in Agricultural Genomics
The Graduate School
Seoul National University

Transcription occurs in the protein-coding regions as well as the regions where any protein-coding sequence is absent. Although these non-coding RNAs lack coding potential, they play roles in transcriptional, post-transcriptional, translational, and post-translational regulation by controlling protein-coding genes. Non-coding RNAs, which are longer than 200 nucleotides, are considered as long non-coding RNAs (lncRNAs). As the sequencing technology has advanced, a repertoire of lncRNA transcriptomes has been accumulated and the functional characterization of each lncRNA has been performed. LncRNAs have been reported to participate in the development, responses to abiotic stresses, and host-microbe interaction. However, their role in plant fungal pathogens was poorly understood due to the limited range of studied species. In this study, we profiled lncRNAs of the rice blast fungus, *Magnaporthe oryzae*, during disease development to decipher the role of lncRNAs in

response to the host. We identified lncRNAs and analyzed their genomic feature and expression pattern to understand their properties, which could be related to their functions. Moreover, specifically expressed lncRNAs in infection stages and their target genes were identified to investigate functional lncRNAs. The analysis of target gene functions suggests that these lncRNAs play roles in pathogenesis such as cell wall degradation and evasion of host immunity. LncRNAs could function solely or in cooperation with small RNAs (sRNAs). LncRNAs generally interact with sRNAs in three ways. LncRNAs could be precursors of sRNAs, be regulated by sRNAs, and regulate sRNA activity. However, their interaction is not well understood in fungi. We profiled lncRNAs and sRNAs in the defect of sRNA biogenesis machinery genes to unravel their interaction in *M. oryzae*. We selected sRNAs processed by RNA interference machinery to filter out the debris. The analysis of genes targeted by non-coding RNAs suggests that two classes of non-coding RNAs be involved in different biological processes depending on the type of interaction. This study provides a repertoire of non-coding RNAs and a foundation for functional studies to elucidate their biological roles. This comprehensive study helps to understand the crosstalk between two classes of non-coding RNAs and suggests that non-coding RNAs can be key regulators in biological processes including pathogenesis. Taken together, this work shed light on the complex regulatory network in plant pathogenic fungi.

Keywords: *Magnaporthe oryzae*, host infection, long non-coding RNA, small non-coding RNA, rice blast disease

Student number: 2015-21805

CONTENTS

LIST OF TABLES

CHAPTER I

CHAPTER II

CHAPTER III

LIST OF FIGURES

CHAPTER II

page

CHAPTER III

CHAPTER I

Long non-coding RNA in fungi

ABSTRACT

Even though non-coding RNAs are not translated into proteins, they participate in diverse biological processes by regulating protein-coding genes at the transcriptional and post-transcriptional levels. These non-coding RNAs comprise small non-coding RNAs (sRNAs) and long non-coding RNAs (lncRNAs). LncRNAs, which are longer than 200 nucleotides, and have been less well studied than sRNAs, which include microRNAs, small interfering RNAs, and piwi-interacting RNAs. This review focuses on fungal lncRNAs, specifically their role in development and stress responses, based on molecular and functional characterizations. Novel areas of research of the crosstalks between lncRNAs and sRNAs in regulating gene expression are also proposed.

INTRODUCTION

Non-canonical transcripts lack the structure of protein-coding genes but account for a considerable part of the genome in both eukaryotes and prokaryotes (Clark et al., 2011; Wade and Grainger, 2014). These pervasive transcripts are heterogeneous and include long non-coding RNAs (lncRNAs) and small non-coding RNAs (sRNAs) (Laurent et al., 2015; Ma et al., 2013). LncRNAs are longer than 200 nucleotides, often have a 5′ end cap and a poly(A) tail, and undergo splicing events similar to mRNAs (Nojima and Proudfoot, 2022). However, they differ from mRNAs by their smaller size, lower expression levels, and lower sequence conservation (Quinn and Chang, 2016). LncRNAs are classified as sense, antisense, intergenic, and intronic lncRNAs depending on their genomic context (Laurent et al., 2015; Ma et al., 2013; Ponting et al., 2009). They can also be categorized as cis-acting and trans-acting lncRNAs, based on their distance to regulated genes (Gil and Ulitsky, 2020).

Following the discovery of *H19*, the number of functionally characterized lncRNAs has steadily increased (Brannan et al., 1990). LncRNAs have been shown to modulate chromatin regulation, scaffolding, transcriptional regulation, and post-transcriptional regulation (Statello et al., 2021). In mammals, they are involved in cell differentiation, disease processes, and the immune response (Chen et al., 2017; Fatica and Bozzoni, 2014; Hobuß et al., 2019), and they have been explored as biomarkers and therapeutic targets (Dhuri et al., 2020; Wang et al., 2017). In plants, lncRNAs participate in development and pathogen resistance (Zaynab et al., 2018; Zhang and Chen, 2013).

Advances in high-throughput sequencing (HTS) have increased the number of available fungal transcriptomes, but studies of profiling and characterizing lncRNAs in fungi have thus far been limited to yeast. In this review, we examine the genome-wide profiling of lncRNAs in fungi, including yeasts, saprotrophs, animal pathogens, and plant pathogens, and summarize the findings from molecular studies that have led to functional characterizations of lncRNAs.

I. LncRNA profiling in fungi

As a first step in studying the role of lncRNAs, their identification is essential. Northern blotting and reverse-transcription polymerase chain reaction are used to detect both lncRNAs and mRNAs (Brannan et al., 1990; Furuno et al., 2006), but these methods have limitations in terms of quantitative transcriptome analyses. Earlier, lncRNAs were profiled on a genomic scale using expressed sequenced tags and microarrays (Gupta et al., 2010; Jia et al., 2010). However, neither method is appropriate for analyzing trends under different biological conditions, due to limitations in the number of identified lncRNAs. With the introduction of HTS technology, RNA sequencing has enabled the identification of novel lncRNAs and the evaluation of their expression (Wang et al., 2009).

LncRNAs are expressed at low levels because they are targeted by RNA surveillance (Nair et al., 2020). In yeast, the deletion of the exosome subunit Rrp6 enabled the detection of cryptic unstable transcripts. Cryptic unstable transcripts are degraded in the nucleus (Wyers et al., 2005), in contrast to Xrn1-sensitive unstable transcripts (XUTs) and stable unannotated transcripts (SUTs), which are processed in the cytoplasm (Garneau et al., 2007). XUTs accumulate in the absence of the 5′–3′ exonuclease Xrn1 (Van Dijk et al., 2011), whereas SUTs are less affected by the deletion of RNA surveillance components and are processed by cytoplasmic exosomes (Xu et al., 2009). Nrd1-unterminated transcripts have been observed in mutants lacking the RNA-binding factor Nrd1, which is related to early termination of transcription (Schulz et al., 2013). Recently, Dicer-sensitive unstable transcripts were detected under deficiency of RNase III Dicer (Atkinson et al., 2018).

Saprotrophic fungi

Many species of fungi obtain nutrients from non-living material and play essential roles in ecosystems as decomposers (Boddy and Hiscox, 2016). Genome-wide lncRNA profiling has been used to investigate the lncRNA transcriptome in these saprotrophic fungi. In *Neurospora crassa*, a model fungus for the study of circadian rhythms, light induces the expression of lncRNAs (Arthanari et al., 2014; Cemel et al., 2017). In the medicinal mushroom *Taiwanofungus camphoratus*, lncRNAs are differentially expressed during asexual and sexual development (Chen et al., 2022). Differentially expressed lncRNAs were also detected in the fruiting body stage of two mushroom-forming fungi, *Pterula gracilis* and *Pleurotus ostreatus* (Merényi et al., 2022). In the wood-decaying fungi *Coniophora puteana* and *Serpula lacrymans*, lncRNAs differentially expressed in response to organic matter extracts were shown to be related to proteolysis and carbohydrate metabolism (Borgognone et al., 2019).

Pathogenic fungi

Disease-causing fungi extract nutrients from living or dead hosts (Barelli et al., 2016; Rai and Agarkar, 2016; Rokas, 2022). Profiling studies suggest that lncRNAs reflect the fungal lifestyle. Studies of human pathogenic fungi have largely focused on members of the genus *Candida* (Table 1). LncRNAs include those specifically expressed and co-expressed with pathogenicity genes, but across the pathogenic species *Candida albicans*, *C. tropicalis*, *C. parapsilosis*, *C. auris*, and *C. glabrata*, they are poorly conserved (Hovhannisyan and Gabaldón, 2021). The dynamics of lncRNAs have been characterized under infection-relevant conditions in *Cryptococcus neoformans* (Kalem and Panepinto, 2022). LncRNAs differentially

expressed during sexual or asexual development were also detected in three insect pathogens (*Conidiobolus obscurus*, *Cordyceps militaris*, and *Metarhizium robertsii*) (Wang et al., 2019a; Wang et al., 2019b; Ye et al., 2021).

Fungi are the causal agents of economically important plant diseases (Doehlemann et al., 2017). In three smut fungi (*Ustilago maydis*, *Ustilago hordei*, and *Sporisorium reilianum*), the detection of conserved lncRNAs has been reported (Donaldson et al., 2017; Donaldson and Saville, 2013). Profiling of lncRNAs during the sexual development of *Fusarium graminearum* revealed XUTs (Kim et al., 2018). The modulated expression of lncRNAs as a regulator of protein-coding genes during disease development was described in *Magnaporthe oryzae* (Choi et al., 2022; Li et al., 2021). In *Ustilaginoidea virens*, a role for lncRNAs in transport-related regulation during infection was proposed (Tang et al., 2021).

Table 1. Genome-wide profiling of fungal lncRNA

Species	Life style	Host	Condition	Reference
N. crassa	Saprotroph	-	Vegetative growth, light, temperature stress	Arthanari et al., 2014; Cemel et al., 2017
T. Camphoratus	Saprotroph	-	Mycelia, arthrospore	Chen et al., 2022
P. gracilis	Saprotroph	-	Developmental stages	Merényi et al., 2021
P. ostreatus	Saprotroph	-	Developmental stages	Merényi et al., 2021
C. puteana	Saprotroph	-	Mycelia, nutrient stress	Borgognone et al., 2019
S. lacrymans	Saprotroph	-	Mycelia, nutrient stress	Borgognone et al., 2019
C. albicans	Pathogen	Human	Infection, heat, PH, nutrient, drug stress	Hovhannisyan et al., 2021
C. auris	Pathogen	Human	Infection, drug stress	Hovhannisyan et al., 2021
C. glabrata	Pathogen	Human	Infection, nutrient, PH stress	Hovhannisyan et al., 2021
C. parapsilosis	Pathogen	Human	Infection, nutrient, temperature, oxidative stress	Hovhannisyan et al., 2021
C. tropicalis	Pathogen	Human	Infection, nutrient, PH, drug, ultrasound stress	Hovhannisyan et al., 2021
C. neoformans	Pathogen	Human	Nutrient, temperature stress	Kalem et al., 2022
C. obscurus	Pathogen	Insect	Mycelia	Ye et al., 2021
C. militaris	Pathogen	Insect	Sexual and asexual development	Wang et al., 2019a
M. robertsii	Pathogen	Insect	Conidial germination, heat stress	Wang et al., 2019b
F. graminearum	Pathogen	Wheat, corn	Sexual development	Kim et al., 2018
M. oryzae	Pathogen	Rice	Infection, vegetative growth, conidia	Li et al., 2021; Choi et al., 2022
S. reilianum	Pathogen	Maize, sorghum	Haploid mycelia, dikaryon mycelia, teliospore	Donaldson et al., 2017
U. maydis	Pathogen	Maize	Haploid mycelia, dikaryon mycelia, teliospore	Donaldson et al., 2013; Donaldson et al., 2017
U. hordei	Pathogen	Barley, oat	Haploid mycelia, dikaryon mycelia, teliospore	Donaldson et al., 2017
U. virens	Pathogen	Rice	Infection	Tang et al., 2021

II. Biological functions of lncRNAs in fungi

Although thousands of fungal lncRNAs have been identified by genome-wide profiling, only a few have been functionally characterized (Table 2). Those studies showed that lncRNAs are involved in diverse biological processes, including cell differentiation, nutrient metabolism, and stress response. Specific biological processes such as host infection have been examined.

Fungal growth and differentiation

Cell–cell adhesion is a prerequisite for tissue formation (Alberts, 2017). In *Saccharomyces cerevisiae*, the *FLO11* gene, encoding a cell wall glycoprotein, is regulated by two lncRNAs: *ICR1* and *PWR1* (Bumgarner et al., 2009). The lncRNAs *IRT1*, *RME2*, and *RME3* control the expression of *IME1*, *IME4*, and *ZIP2*, respectively, which are involved in meiosis (Gelfand et al., 2011; Hongay et al., 2006; Van Werven et al., 2012). The pHO lncRNA induces nucleosome repositioning and influences the transcription of *HO* endonuclease, required for mating-type interconversion (Yu et al., 2016). The antisense lncRNA *SUT169* controls the relative expression of *SPS100* mRNA isoforms, which are responsible for sporulation (Bunina et al., 2017). The antisense lncRNA *ADF1* suppresses the expression of *MDF1* to influence vegetative growth (Li et al., 2010). In *Schizosaccharomyces pombe*, meiRNAs localize at the *sme2* locus and bind Mmi1 to promote meiosis (Shichino et al., 2014). In *C. neoformans*, the lncRNA *RZE1* controls the transcript level of *ZNF2*, which regulates yeast-to-hyphal transition.

Metabolisms and nutrition

In *S. cerevisiae*, two lncRNAs, *GAL4* and *GAL10*, participate in galactose metabolism, by regulating the transcription of the *GAL* gene cluster, which is induced in the absence of glucose and in the presence of galactose (Geisler et al., 2012; Houseley et al., 2008). In *S. pombe*, mlonRNAs are induced in the absence of glucose and remodel chromatin structure to promote the expression of the *fbp1* gene (Hirota et al., 2008). In both *S. cerevisiae* and *S. pombe*, lncRNAs play a role in controlling phosphate metabolism (Ard et al., 2014; Camblong et al., 2007; Chatterjee et al., 2016; Garg et al., 2018; Shah et al., 2014; Uhler et al., 2007). In *S. cerevisiae*, nitrogen starvation induces the expression of *ncASP3* and facilitates access to *ASP3* via histone modifications (Huang et al., 2010). The lncRNA *SRG1*, induced under serine-rich conditions, suppresses the expression of the *SER3* gene, involved in serine biosynthesis (Martens et al., 2005). Beyond yeast species, lncRNAs were shown to regulate carotenoid biosynthesis and cellulose metabolism in *Fusarium* species and *Trichoderma reesei*, respectively (Parra-Rivero et al., 2020; Till et al., 2020; Till et al., 2018).

Responses to abiotic stresses

In *S. cerevisiae*, antisense lncRNAs of the *CDC28* gene interact with the stress-responsive protein kinase Hog1 during osmotic stress to enable access of chromatin remodelers to *CDC28* and in turn activate its expression (Nadal-Ribelles et al., 2014). In *S. pombe*, the lncRNA *SPNCRNA.1164* regulates the expression of the transcription factor *atf1* in response to oxidative stress (Leong et al., 2014). In *N. crassa*, the antisense lncRNA *qrf* negatively regulates the circadian clock gene *frq*

via chromatin modifications (Kramer et al., 2003). In *U. virens*, the antisense lncRNAs *UvlncNAT-MFS* and *UvMFS* form an RNA duplex that regulates vegetative growth, conidiation, and the response to osmotic and oxidative stresses (Tang et al., 2021).

Pathogenicity

lncRsp1 and *GzmetE-AS*, two lncRNAs present in *F. graminearum*, which is the causal agent of fusarium head blight, participate in sexual reproduction and host infection (Wang et al., 2022; Wang et al., 2021). In the absence of *lncRsp1*, the virulence of the mutants is decreased due to their reduced ability to discharge ascospores. *lncRsp1* together with the adjacent sugar transporter gene *Fgsp1* negatively regulates the deoxynivalenol biosynthesis genes *TRI4*, *TRI5*, *TRI6*, and *TRI13*. Similarly, the antisense lncRNA *GzmetE-AS* negatively regulates the HOA-encoding gene *GzmetE*, required for vegetative growth, asexual and sexual reproduction, and pathogenesis. In the common corn smut fungus *U. maydis*, deletion of the lncRNA *ncRNA1* and of the antisense transcript of *um02151* affects virulence, although the underlying mode of action is unknown (Donaldson and Saville, 2013; Morrison et al., 2012).

Table 2. Molecular functional characterization of fungal lncRNA

Species	LncRNA	Target gene	Condition	Reference
S. cerevisiae	*ICR1*	*FLO11*	Cell-cell adhesion	Bumgarner et al., 2009
	PWR1	*ICR1* (*FLO11*)	Cell-cell adhesion	Bumgarner et al., 2009
	GAL10 lncRNA	*GAL10, GAL1*	Galactose utilization	Houseley et al., 2008
	GAL4 lncRNA	*GAL4*	Galactose utilization	Geisler et al., 2012
	pHO-lncRNA	*HO* genes	Mating type interconversion during cell cycle re-entry	Yu et al., 2016
	IRT1	*IME1*	Meiosis	van Werven et al., 2012
	RME2	*IME4*	Meiosis	Hongay et al., 2006
	RME3	*ZIP2*	Meiosis	Gelfand et al., 2011
	ncASP3	*ASP3*	Nitrogen starvation	Huang et al., 2010
	as-CDC28	*CDC28*	Osmostress	Nadal-Ribelles et al., 2014
	as-PHO5	*PHO5*	Phosphate metabolism	Uhler et al., 2007
	as-PHO84	*PHO84*	Phosphate metabolism	Camblong et al., 2007
	SRG1	*SER3*	Serine biosynthesis	Martens et al., 2004
	SUT169	*SPS100*	Sporulation	Bunina et al. 2017
	ADF1	*MDF1*	Vegetative growth	Li et al., 2010
S. pombe	mlonRNA	*fbp1*	Glucose starvation	Hirota et al., 2008
	meiRNA-S and L	*sme2*	Meiosis	Shichino et al., 2014
	SPNCRNA.1164	*atf1*	Oxidative stress	Leong et al., 2014
	nc-tgp1	*tgp1*	Phosphate metabolism	Ard et al., 2014
	prt / *nc-pho1*	*pho1*	Phosphate metabolism	Chatterjee et al., 2016; Shah et al., 2014
	prt2	*pho84*	Phosphate metabolism	Garg et al., 2018
C. neoformans	*RZE1*	*ZNF2*	Yeast-to-hyphal transition	Chacko et al., 2015
F. graminearum	*GzmetE-AS*	*GzmetE*	Asexual and sexual reproduction, and plant infection	Wang et al., 2021
F. graminearum	*lncRsp1*	Deoxynivalenol biosynthesis genes	Sexual reproduction and plant infection	Wang et al., 2022

F. oxysporum, *F. fujikuroi*	*carP*	*CarS*	Carotenoid biosynthesis	Parra-Rivero et al., 2020
N. crassa	*qrf*	*frq*	Circadian rhythm	Kramer et al., 2003
T. reesei	*HAX1*	Cellulase genes	Cellulose metabolism	Till et al., 2018; Till et al., 2020
U. maydis	*ncRNA1*	-	Pathogenic development	Morrison et al., 2012
U. maydis	*as-UMAG_02151*	-	Pathogenic development	Donaldson and Saville, 2013
U. virens	*UvlncNAT-MFS*	*UvMFS*	Conidiation, growth and various stress response	Tang et al., 2021

PERSPECTIVE

LncRNAs participate in biological processes by regulating protein-coding genes. However, functional characterizations by *in silico* comparative analyses based on comparisons of primary sequences have been limited, due to the low sequence conservation level and insufficient transcriptome databases. Advanced HTS technology has improved genome and transcriptome information both quantitatively and qualitatively, which has led to more accurate gene models, by the inclusion of all full-length transcript isoforms. *In silico* comparative analyses that incorporate secondary/tertiary structures and genomic contexts from various species will allow more accurate predictions of lncRNA function. However, this approach has been applied far less often in fungi than in animals and plants. Its broader application in fungi will provide a larger repertoire of fungal transcriptomes.

Crosstalks between lncRNAs and sRNAs have been investigated in animals and plants (Ulitsky, 2018; Zhou et al., 2020), but due to the absence of RNA interference (RNAi) machinery in budding yeast, few such studies have been conducted in fungi (Drinnenberg et al., 2009). The lncRNAs of fission yeasts were recently profiled under RNAi machinery deficiency (Atkinson et al., 2018; Szachnowski et al., 2019). Investigations of the crosstalks among ncRNAs in fungi will fill many of the knowledge gaps regarding gene regulatory networks and their functions under diverse conditions. A better understanding of the biological roles of lncRNAs in fungi will have many practical applications, including in industry, medicine, and agriculture.

LITERATURE CITED

Alberts, B. 2017. Molecular biology of the cell. WW Norton & Company.

Ard, R., Tong, P., and Allshire, R.C. 2014. Long non-coding RNA-mediated transcriptional interference of a permease gene confers drug tolerance in fission yeast. Nat. Commun. 5:1-9.

Arthanari, Y., Heintzen, C., Griffiths-Jones, S., and Crosthwaite, S.K. 2014. Natural antisense transcripts and long non-coding RNA in *Neurospora crassa*. PLoS ONE. 9:e91353.

Atkinson, S.R., Marguerat, S., Bitton, D.A., Rodríguez-López, M., Rallis, C., Lemay, J.-F., Cotobal, C., Malecki, M., Smialowski, P., and Mata, J. 2018. Long noncoding RNA repertoire and targeting by nuclear exosome, cytoplasmic exonuclease, and RNAi in fission yeast. RNA. 24:1195-1213.

Barelli, L., Moonjely, S., Behie, S.W., and Bidochka, M.J. 2016. Fungi with multifunctional lifestyles: endophytic insect pathogenic fungi. Plant Mol. Biol. 90:657-664.

Boddy, L., and Hiscox, J. 2016. Fungal ecology: principles and mechanisms of colonization and competition by saprotrophic fungi. Microbiol. Spectr. 4:4.6. 17.

Borgognone, A., Sanseverino, W., Aiese Cigliano, R., and Castanera, R. 2019. Distribution, characteristics, and regulatory potential of long noncoding RNAs in brown-rot fungi. Int. J. Genomics. 2019:9702342

Brannan, C.I., Dees, E.C., Ingram, R.S., and Tilghman, S.M. 1990. The product of

the *H19* gene may function as an RNA. Mol. Cell. Biol. 10:28-36.

Bumgarner, S.L., Dowell, R.D., Grisafi, P., Gifford, D.K., and Fink, G.R. 2009. Toggle involving cis-interfering noncoding RNAs controls variegated gene expression in yeast. Proc. Natl. Acad. Sci. USA. 106:18321-18326.

Bunina, D., Štefl, M., Huber, F., Khmelinskii, A., Meurer, M., Barry, J.D., Kats, I., Kirrmaier, D., Huber, W., and Knop, M. 2017. Upregulation of *SPS100* gene expression by an antisense RNA via a switch of mRNA isoforms with different stabilities. Nucleic Acids Res. 45:11144-11158.

Camblong, J., Iglesias, N., Fickentscher, C., Dieppois, G., and Stutz, F. 2007. Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in *S. cerevisiae*. Cell. 131:706-717.

Cemel, I.A., Ha, N., Schermann, G., Yonekawa, S., and Brunner, M. 2017. The coding and noncoding transcriptome of *Neurospora crassa*. BMC Genomics. 18:978.

Chatterjee, D., Sanchez, A.M., Goldgur, Y., Shuman, S., and Schwer, B. 2016. Transcription of lncRNA *prt*, clustered *prt* RNA sites for Mmi1 binding, and RNA polymerase II CTD phospho-sites govern the repression of *pho1* gene expression under phosphate-replete conditions in fission yeast. RNA. 22:1011-1025.

Chen, C.-L., Li, W.-C., Chuang, Y.-C., Liu, H.-C., Huang, C.-H., Lo, K.-Y., Chen, C.-Y., Chang, F.-M., Chang, G.-A., and Lin, Y.-L. 2022. Sexual Crossing, Chromosome-Level Genome Sequences, and Comparative Genomic Analyses for the Medicinal Mushroom *Taiwanofungus Camphoratus* (Syn. *Antrodia Cinnamomea*, *Antrodia Camphorata*). Microbiol. Spectr.

10:e02032-02021.

Chen, Y.G., Satpathy, A.T., and Chang, H.Y. 2017. Gene regulation in the immune system by long noncoding RNAs. Nature Immun. 18:962-972.

Choi, G., Jeon, J., Lee, H., Zhou, S., and Lee, Y.-H. 2022. Genome-wide profiling of long non-coding RNA of the rice blast fungus *Magnaporthe oryzae* during infection. BMC Genomics. 23:132.

Clark, M.B., Amaral, P.P., Schlesinger, F.J., Dinger, M.E., Taft, R.J., Rinn, J.L., Ponting, C.P., Stadler, P.F., Morris, K.V., and Morillon, A. 2011. The reality of pervasive transcription. PLoS Biol. 9:e1000625.

Dhuri, K., Bechtold, C., Quijano, E., Pham, H., Gupta, A., Vikram, A., and Bahal, R. 2020. Antisense oligonucleotides: an emerging area in drug discovery and development. J. Clin. Med. 9:2004.

Doehlemann, G., Ökmen, B., Zhu, W., and Sharon, A. 2017. Plant pathogenic fungi. Microbiol. Spectr. 5:5.1. 14.

Donaldson, M.E., Ostrowski, L.A., Goulet, K.M., and Saville, B.J. 2017. Transcriptome analysis of smut fungi reveals widespread intergenic transcription and conserved antisense transcript expression. BMC Genomics. 18:340.

Donaldson, M.E., and Saville, B.J. 2013. *Ustilago maydis* natural antisense transcript expression alters mRNA stability and pathogenesis. Mol. Microbiol. 89:29-51.

Drinnenberg, I.A., Weinberg, D.E., Xie, K.T., Mower, J.P., Wolfe, K.H., Fink, G.R., and Bartel, D.P. 2009. RNAi in budding yeast. Science. 326:544-550.

Fatica, A., and Bozzoni, I. 2014. Long non-coding RNAs: new players in cell

differentiation and development. Nat. Rev. Genet. 15:7-21.

Furuno, M., Pang, K.C., Ninomiya, N., Fukuda, S., Frith, M.C., Bult, C., Kai, C., Kawai, J., Carninci, P., and Hayashizaki, Y. 2006. Clusters of internally primed transcripts reveal novel long noncoding RNAs. PLoS Genet. 2:e37.

Garg, A., Sanchez, A.M., Shuman, S., and Schwer, B. 2018. A long noncoding (lnc) RNA governs expression of the phosphate transporter Pho84 in fission yeast and has cascading effects on the flanking *prt* lncRNA and *pho1* genes. J. Biol. Chem. 293:4456-4467.

Garneau, N.L., Wilusz, J., and Wilusz, C.J. 2007. The highways and byways of mRNA decay. Nat. Rev. Mol. Cell Biol. 8:113-126.

Geisler, S., Lojek, L., Khalil, A.M., Baker, K.E., and Coller, J. 2012. Decapping of long noncoding RNAs regulates inducible genes. Mol. Cell. 45:279-291.

Gelfand, B., Mead, J., Bruning, A., Apostolopoulos, N., Tadigotla, V., Nagaraj, V., Sengupta, A.M., and Vershon, A.K. 2011. Regulated antisense transcription controls expression of cell-type-specific genes in yeast. Mol. Cell. Biol. 31:1701-1709.

Gil, N., and Ulitsky, I. 2020. Regulation of gene expression by cis-acting long non-coding RNAs. Nat. Rev. Genet. 21:102-117.

Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.-C., Hung, T., Argani, P., and Rinn, J.L. 2010. Long non-coding RNA *HOTAIR* reprograms chromatin state to promote cancer metastasis. Nature. 464:1071-1076.

Hirota, K., Miyoshi, T., Kugou, K., Hoffman, C.S., Shibata, T., and Ohta, K. 2008. Stepwise chromatin remodelling by a cascade of transcription initiation of

non-coding RNAs. Nature. 456:130-134.

Hobuß, L., Bär, C., and Thum, T. 2019. Long non-coding RNAs: at the heart of cardiac dysfunction? Front. Physiol. 10:30.

Hongay, C.F., Grisafi, P.L., Galitski, T., and Fink, G.R. 2006. Antisense transcription controls cell fate in *Saccharomyces cerevisiae*. Cell. 127:735-745.

Houseley, J., Rubbi, L., Grunstein, M., Tollervey, D., and Vogelauer, M. 2008. A ncRNA modulates histone modification and mRNA induction in the yeast *GAL* gene cluster. Mol. Cell. 32:685-695.

Hovhannisyan, H., and Gabaldón, T. 2021. The long non-coding RNA landscape of *Candida* yeast pathogens. Nat. Commun. 12:1-13.

Huang, Y.-C., Chen, H.-T., and Teng, S.-C. 2010. Intragenic transcription of a noncoding RNA modulates expression of *ASP3* in budding yeast. RNA. 16:2085-2093.

Jia, H., Osak, M., Bogu, G.K., Stanton, L.W., Johnson, R., and Lipovich, L. 2010. Genome-wide computational identification and manual annotation of human long noncoding RNA genes. RNA. 16:1478-1487.

Kalem, M.C., and Panepinto, J.C. 2022. Long non-coding RNAs in *Cryptococcus neoformans*: insights into fungal pathogenesis. Front. Cell. Infect. Microbiol. 12:858317

Kim, W., Miguel-Rojas, C., Wang, J., Townsend, J.P., and Trail, F. 2018. Developmental dynamics of long noncoding RNA expression during sexual fruiting body formation in *Fusarium graminearum*. mBio. 9:e01292-01218.

Kramer, C., Loros, J.J., Dunlap, J.C., and Crosthwaite, S.K. 2003. Role for antisense RNA in regulating circadian clock function in *Neurospora crassa*. Nature.

421:948-952.

Laurent, G.S., Wahlestedt, C., and Kapranov, P. 2015. The Landscape of long noncoding RNA classification. Trends Genet. 31:239-251.

Leong, H.S., Dawson, K., Wirth, C., Li, Y., Connolly, Y., Smith, D.L., Wilkinson, C.R., and Miller, C.J. 2014. A global non-coding RNA system modulates fission yeast protein levels in response to stress. Nat. Commun. 5:1-10.

Li, D., Dong, Y., Jiang, Y., Jiang, H., Cai, J., and Wang, W. 2010. A *de novo* originated gene depresses budding yeast mating pathway and is repressed by the protein encoded by its antisense strand. Cell Res. 20:408-420.

Li, Z., Yang, J., Peng, J., Cheng, Z., Liu, X., Zhang, Z., Bhadauria, V., Zhao, W., and Peng, Y.-L. 2021. Transcriptional Landscapes of Long Non-coding RNAs and Alternative Splicing in *Pyricularia oryzae* Revealed by RNA-Seq. Front. Plant Sci. 12:723636.

Ma, L., Bajic, V.B., and Zhang, Z. 2013. On the classification of long non-coding RNAs. RNA Biol. 10:924-933.

Martens, J.A., Wu, P.-Y.J., and Winston, F. 2005. Regulation of an intergenic transcript controls adjacent gene transcription in *Saccharomyces cerevisiae*. Genes Dev. 19:2695-2704.

Merényi, Z., Virágh, M., Gluck-Thaler, E., Slot, J.C., Kiss, B., Varga, T., Geösel, A., Hegedüs, B., Bálint, B., and Nagy, L.G. 2022. Gene age shapes the transcriptional landscape of sexual morphogenesis in mushroom-forming fungi (Agaricomycetes). Elife. 11:e71348.

Morrison, E.N., Donaldson, M.E., and Saville, B.J. 2012. Identification and analysis of genes expressed in the *Ustilago maydis* dikaryon: uncovering a novel

class of pathogenesis genes. Can. J. Plant Pathol. 34:417-435.

Nadal-Ribelles, M., Solé, C., Xu, Z., Steinmetz, L.M., de Nadal, E., and Posas, F. 2014. Control of Cdc28 CDK1 by a stress-induced lncRNA. Mol. Cell. 53:549-561.

Nair, L., Chung, H., and Basu, U. 2020. Regulation of long non-coding RNAs and genome dynamics by the RNA surveillance machinery. Nat. Rev. Mol. Cell Biol. 21:123-136.

Nojima, T., and Proudfoot, N.J. 2022. Mechanisms of lncRNA biogenesis as revealed by nascent transcriptomics. Nat. Rev. Mol. Cell Biol. 23:389-406.

Parra-Rivero, O., Pardo-Medina, J., Gutiérrez, G., Limón, M.C., and Avalos, J. 2020. A novel lncRNA as a positive regulator of carotenoid biosynthesis in *Fusarium*. Sci. Rep. 10:1-14.

Ponting, C.P., Oliver, P.L., and Reik, W. 2009. Evolution and functions of long noncoding RNAs. Cell. 136:629-641.

Quinn, J.J., and Chang, H.Y. 2016. Unique features of long non-coding RNA biogenesis and function. Nat. Rev. Genet. 17:47-62.

Rai, M., and Agarkar, G. 2016. Plant–fungal interactions: what triggers the fungi to switch among lifestyles? Crit. Rev. Microbiol. 42:428-438.

Rokas, A. 2022. Evolution of the human pathogenic lifestyle in fungi. Nat. Microbiol. 7:607-619.

Schulz, D., Schwalb, B., Kiesel, A., Baejen, C., Torkler, P., Gagneur, J., Soeding, J., and Cramer, P. 2013. Transcriptome surveillance by selective termination of noncoding RNA synthesis. Cell. 155:1075-1087.

Shah, S., Wittmann, S., Kilchert, C., and Vasiljeva, L. 2014. lncRNA recruits RNAi

and the exosome to dynamically regulate pho1 expression in response to phosphate levels in fission yeast. Genes Dev. 28:231-244.

Shichino, Y., Yamashita, A., and Yamamoto, M. 2014. Meiotic long non-coding meiRNA accumulates as a dot at its genetic locus facilitated by Mmi1 and plays as a decoy to lure Mmi1. Open Biol. 4:140022.

Statello, L., Guo, C.-J., Chen, L.-L., and Huarte, M. 2021. Gene regulation by long non-coding RNAs and its biological functions. Nat. Rev. Mol. Cell Biol. 22:96-118.

Szachnowski, U., Andjus, S., Foretek, D., Morillon, A., and Wery, M. 2019. Endogenous RNAi pathway evolutionarily shapes the destiny of the antisense lncRNAs transcriptome. Life Sci. Alliance. 2: e201900407.

Tang, J., Chen, X., Yan, Y., Huang, J., Luo, C., Tom, H., and Zheng, L. 2021. Comprehensive transcriptome profiling reveals abundant long non-coding RNAs associated with development of the rice false smut fungus, *Ustilaginoidea virens*. Environ. Microbiol. 23:4998-5013.

Till, P., Derntl, C., Kiesenhofer, D.P., Mach, R.L., Yaver, D., and Mach-Aigner, A.R. 2020. Regulation of gene expression by the action of a fungal lncRNA on a transactivator. RNA Biol. 17:47-61.

Till, P., Mach, R.L., and Mach-Aigner, A.R. 2018. A current view on long noncoding RNAs in yeast and filamentous fungi. Appl. Microbiol. Biotechnol. 102:7319-7331.

Uhler, J.P., Hertel, C., and Svejstrup, J.Q. 2007. A role for noncoding transcription in activation of the yeast *PHO5* gene. Proc. Natl. Acad. Sci. USA. 104:8011-8016.

Ulitsky, I. 2018. Interactions between short and long noncoding RNAs. FEBS Lett. 592:2874-2883.

Van Dijk, E., Chen, C., d'Aubenton-Carafa, Y., Gourvennec, S., Kwapisz, M., Roche, V., Bertrand, C., Silvain, M., Legoix-Né, P., and Loeillet, S. 2011. XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. Nature. 475:114-117.

Van Werven, F.J., Neuert, G., Hendrick, N., Lardenois, A., Buratowski, S., van Oudenaarden, A., Primig, M., and Amon, A. 2012. Transcription of two long noncoding RNAs mediates mating-type control of gametogenesis in budding yeast. Cell. 150:1170-1181.

Wade, J.T., and Grainger, D.C. 2014. Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. Nat. Rev. Microbiol. 12:647-653.

Wang, J., Zeng, W., Cheng, J., Xie, J., Fu, Y., Jiang, D., and Lin, Y. 2022. *lncRsp1*, a long noncoding RNA, influences *Fgsp1* expression and sexual reproduction in *Fusarium graminearum*. Mol. Plant Pathol. 23:265-277.

Wang, J., Zeng, W., Xie, J., Fu, Y., Jiang, D., Lin, Y., Chen, W., and Cheng, J. 2021. A novel antisense long non-coding RNA participates in asexual and sexual reproduction by regulating the expression of *GzmetE* in *Fusarium graminearum*. Environ. Microbiol. 23:4939-4955.

Wang, W., Zhuang, Q., Ji, K., Wen, B., Lin, P., Zhao, Y., Li, W., and Yan, C. 2017. Identification of miRNA, lncRNA and mRNA-associated ceRNA networks and potential biomarker for MELAS with mitochondrial DNA A3243G mutation. Sci. Rep. 7:1-13.

Wang, Y., Shao, Y., Zhu, Y., Wang, K., Ma, B., Zhou, Q., Chen, A., and Chen, H.

2019a. XRN1-associated long non-coding RNAs may contribute to fungal virulence and sexual development in entomopathogenic fungus *Cordyceps militaris*. Pest Manag. Sci. 75:3302-3311.

Wang, Z., Gerstein, M., and Snyder, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10:57-63.

Wang, Z., Jiang, Y., Wu, H., Xie, X., and Huang, B. 2019b. Genome-wide identification and functional prediction of long non-coding RNAs involved in the heat stress response in *Metarhizium robertsii*. Front. Microbiol. 10:2336.

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.-C., Dufour, M.-E., Boulay, J., Régnault, B., Devaux, F., Namane, A., and Séraphin, B. 2005. Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly (A) polymerase. Cell. 121:725-737.

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L.M. 2009. Bidirectional promoters generate pervasive transcription in yeast. Nature. 457:1033-1037.

Ye, G., Zhang, L., and Zhou, X. 2021. Long noncoding RNAs are potentially involved in the degeneration of virulence in an aphid-obligate pathogen, *Conidiobolus obscurus* (Entomophthoromycotina). Virulence. 12:1705-1716.

Yu, Y., Yarrington, R.M., Chuong, E.B., Elde, N.C., and Stillman, D.J. 2016. Disruption of promoter memory by synthesis of a long noncoding RNA. Proc. Natl. Acad. Sci. USA. 113:9575-9580.

Zaynab, M., Fatima, M., Abbas, S., Umair, M., Sharif, Y., and Raza, M.A. 2018.

Long non-coding RNAs as molecular players in plant defense against pathogens. Microb. Pathog. 121:277-282.

Zhang, Y.-C., and Chen, Y.-Q. 2013. Long noncoding RNAs: new regulators in plant development. Biochem. Biophys. Res. Commun. 436:111-114.

Zhou, X., Cui, J., Meng, J., and Luan, Y. 2020. Interactions and links among the noncoding RNAs in plants under stresses. Theor. Appl. Genet. 133:3235-3248.

CHAPTER II

Genome-wide profiling of long non-coding RNA of the rice blast fungus *Magnaporthe oryzae* during infection

This chapter was published in BMC Genomics.

Choi, G. et al. (2022) *BMC Genomics*. 23:132.

ABSTRACT

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. 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. 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.

INTRODUCTION

Genomes encode large numbers of non-coding transcripts, which function in gene regulation (Clark et al., 2011; Kapranov et al., 2007; Wade and Grainger, 2014). 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 (Ma et al., 2013; Mercer et al., 2009). Based on their genomic positions and contexts within protein-coding genes, lncRNAs are categorized as intergenic lncRNAs, antisense lncRNAs, sense lncRNAs, and intronic lncRNAs (Laurent et al., 2015; Ma et al., 2013; Ponting et al., 2009). 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 (Gil and Ulitsky, 2020). LncRNAs modulate the transcriptome through multiple dimensions, including epigenetic, transcriptional, post-transcriptional, translational, and post-translational levels (Zhang et al., 2019).

Following the discovery of *H19* in humans and *Xist* in mice, many more lncRNAs have been functionally characterized (Brannan et al., 1990; Brockdorff et al., 1992). Several studies have reported that mammalian lncRNAs are associated with cell differentiation and disease process; they also serve as biomarkers for cancer diagnoses (Jalali et al., 2012; Li et al., 2016; Wang et al., 2017). Plant lncRNAs, such as *COLDAIR* and *GhlncNAT-ANX2*, have roles in development and in defense against pathogens (Zaynab et al., 2018; Zhang and Chen, 2013).

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 (Li et al., 2021; Till et al., 2018a). LncRNAs also regulate the circadian clock (*qrf*) and cellulase genes (*HAX1*) in the saprotrophic fungi *Neurospora crassa* and *Trichoderma reesei*, respectively (Kramer et al., 2003; Till et al., 2020; Till et al., 2018b). LncRNA *RZE1* regulates zinc finger transcription factor *ZNF2* and affects the yeast-to-hypha transition in the human pathogenic fungus *Cryptococcus neoformans* (Chacko et al., 2015). 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 (Donaldson and Saville, 2013; Morrison et al., 2012; Parra-Rivero et al., 2020; Wang et al., 2021). 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 *Ustilaginoidea virens* (Arthanari et al., 2014; Donaldson et al., 2017; Kim et al., 2018; Tang et al., 2021).

Rice blast disease is caused by the filamentous fungus *Magnaporthe oryzae*, which is responsible for an annual yield loss of 10 − 30% (Skamnioti and Gurr, 2009). In addition to its economic importance, this fungus has served as a model of host–pathogen interactions (Dean et al., 2012). *M. oryzae* undergoes morphological and functional transitions during vegetative growth, appressorium formation, the biotrophic stage, and the necrotrophic stage during the infection process (Fernandez and Orth, 2018). Following the completion of whole genome sequencing of this

fungus, transcriptome profiling was performed to understand gene regulation during the infection process (Dean et al., 2005; Dong et al., 2015; Jeon et al., 2020; Kawahara et al., 2012). 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.

MATERIALS AND METHODS

I. 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 with NEXTflex Rapid Directional mRNA-seq Kit (Bioo Scientific, Austin, TX, USA) 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.

II. 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 (SRA accession no. SRX5076910- SRX5076915) (Jeon et al., 2020). 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).

III. Transcriptome assembly

Raw reads were processed to remove low-quality reads and trim adapter sequences using NGS QC Toolkit v2.3.3 (Patel and Jain, 2012). The resulting reads were mapped against the M. oryzae reference genome (MG8, Ensembl annotation 29) using HISAT2 v2.0.4 (Dean et al., 2012; Kim et al., 2015). The transcriptome was assembled using the genome-guided method of StringTie v1.3.3 with de novo annotation (Pertea et al., 2015). Transcriptome assembly proceeded through two steps. In the first step, the strand-specific RNA-seq data was used. Then, *in planta* RNA-seq data and the updated transcriptome annotation from the first step were used in the second step. 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.

IV. LncRNA identification

We used an established computational pipeline to identify lncRNAs. Transcripts whose spliced sequences are shorter than 200 nucleotides were first filtered out. The

assembled transcripts were then compared with protein-coding genes and categorized using Gffcompare (Pertea et al., 2016). 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 (Kalvari et al., 2018; Nawrocki and Eddy, 2013). The coding potentials of transcripts were assessed using CPAT v.1.2.2 (Wang et al., 2013). To maximize lncRNA detection, training was performed using transcript sequences of *F. graminearum* and the coding potential cutoff was set to 0.54 (Figure 1). Transcripts with coding potential below the cutoff were included; transcripts containing any known Pfam domain were removed using InterProScan version 5.29–68.0 (Jones et al., 2014).

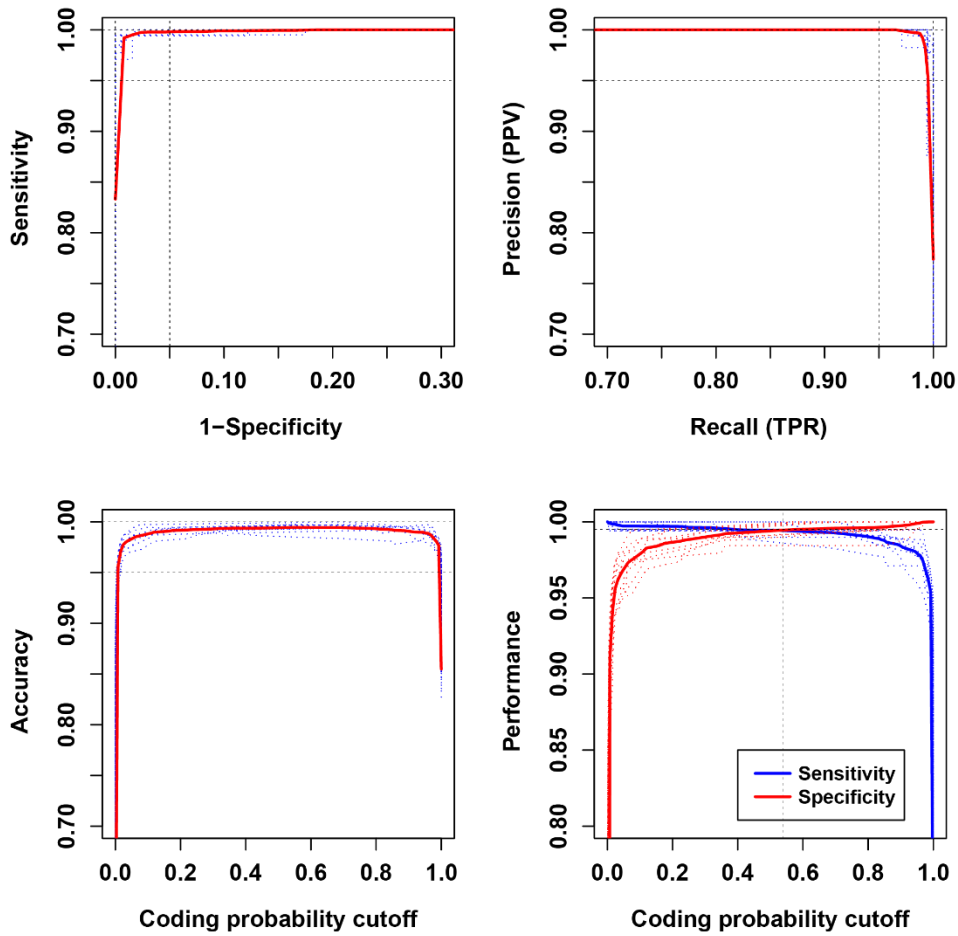

Figure 1. Coding potential model in filamentous fungi. Performance evaluation of non-coding transcript prediction. Two-graph receiver operating characteristic analysis was performed to determine an optimal CPAT cutoff value for non-coding transcript calls. The dashed curves represent 10-fold cross-validation, while solid curves represent the mean curve from 10 validation runs. Sensitivity measures the proportion of positives that are correctly identified. Specificity measures the proportion of negatives that are correctly identified.

V. 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 (Consortium, 2021). 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*, *Magnaporthiopsis rhizophila*, *Magnaporthiopsis incrustans*, *Magnaporthe salvinii*, *Ophioceras dolichostomum*, *Pseudohalonectria 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) (Choi et al., 2013). The genomes of *M. rhizophila*, *M. incrustans*, *M. salvinii*, *O. dolichostomum*, and *P. lignicola* were downloaded from the National Center for Biotechnology Information (Zhang et al., 2018).

VI. 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 (Yanai et al., 2005).

$$\tau = \frac{\sum_{i=1}^{n}(1-\hat{x_i})}{n-1}; \hat{x_i} = \frac{x_i}{\max\limits_{1\le i\le n} x_i}$$

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.

VII. 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) (Tafer and Hofacker, 2008). Target genes were then subjected to Gene Ontology (GO) term enrichment analysis at a 5% false discovery rate using Blast2GO and AgriGO v2.0 (Götz et al., 2008; Tian et al., 2017). Pathogenesis-related genes were identified by querying target genes against a pathogen-host interactions database (PHI-base)

(Urban et al., 2020).

VIII. 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 × 104 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 (Ho et al., 2010). 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 1.

Table 1. The primers used in this study

Primer	Strand	Sequence (5' → 3')
MSTRG.1141.1-F	forward	CAAGAGATGTGATCGGAGCCCG
MSTRG.1141.1-R	reverse	GAGAAACACCCCCTTCTTACCACAGAC
MSTRG.14182.1-F	forward	CGGTTAGATTGCGATTTCGAAGAGGG
MSTRG.14182.1-R	reverse	TTGCGACAAACTACGCAACAGACC
MSTRG.1314.1-F	forward	TGATGGATCTCGAATTCGGGGTGAC
MSTRG.1314.1-R	reverse	TCATTTCAACGCCCACTGGCTCTA
MSTRG.1779.3-F	forward	TAACCCACTGCCGGCAAATCAAG
MSTRG.1779.3-R	reverse	TAAATAGTCGGCGGGTAGTGTAGGG
MSTRG.7417.5-F	forward	CCTGGTAAGCATACACGTGCGATG
MSTRG.7417.5-R	reverse	TAAAACCTGCGGATTATCCCCCCAA
MSTRG.9578.3-F	forward	TATGTTGTCCCATCATGCCTAAGTGGC
MSTRG.9578.3-R	reverse	GAGCTGCCAACGTTGTAAACCAGATG
MSTRG.9719.3-F	forward	CTCCAGACCATAGATTGTCACAGGCA
MSTRG.9719.3-R	reverse	TTTCACTTACCTCGATAACCTCGCCC
MSTRG.7770.1-F	forward	CCGACGACTTCTTCGTGGTTTGAAC
MSTRG.7770.1-R	reverse	GTCGATTCTATTGCAAATCAACCGGTCC
MSTRG.10154.1-F	forward	GGTGTTGATGATGAACCACCAGGTTG
MSTRG.10154.1-R	reverse	CTATATCAGATCGCCATGCGAATTGCC
MSTRG.7581.1-F	forward	GAGCCGGGAATTGTATTGGGCAAAAG
MSTRG.7581.1-R	reverse	TACCGTAGGGACATCACCAATCCATC
MSTRG.5505.1-F	forward	CTCCAAGTTACCCAAACCACGAAATGG
MSTRG.5505.1-R	reverse	CTTGTTTCCATACGAGTTTCTGGCTGC
MSTRG.5588.1-F	forward	CAGAACGAGATACGAGCCAGTGAC
MSTRG.5588.1-R	reverse	GGCTTCTAATGTGGGAGGGTTAGATG
MSTRG.980.1-F	forward	CGACGACGGAACGAGAAACGATTG
MSTRG.980.1-R	reverse	GCGTGACATTGATGGAAATTTGCCG
Beta-tubulin-F	forward	ACAACTTCGTCTTCGGTCAG
Beta-tubulin-R	reverse	GTGATCTGGAAACCCTGGAG

RESULTS

I. 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* (Jeon et al., 2020). Previously established pipelines were used to detect lncRNAs with some modifications (Figure 2A) (Weirick et al., 2016). In total, 436.6 million reads were mapped to the *M. oryzae* genome with 27,480 predicted transcripts originating from 16,093 genomic loci (Figure 2B). 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 (Pertea and Pertea, 2020); known mRNAs from the Ensembl database and non-coding RNAs from the Rfam database were removed (Kalvari et al., 2018). 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 2, Figure 2C). 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 (Figure 3).

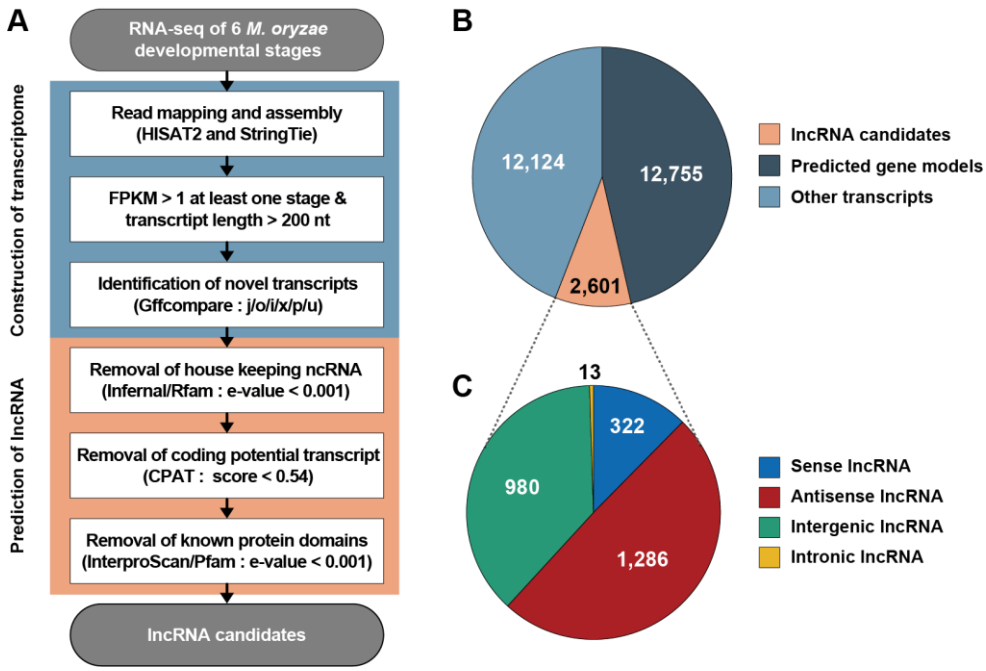

Figure 2. Schematic pipeline for identification of lncRNAs in *M. oryzae*. (A) Bioinformatic pipeline for lncRNA identification using RNA-seq data. CPAT, Coding Potential Assessment Tool. (B) Number of predicted transcripts. (C) Number of lncRNAs by different classes.

Table 2. 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

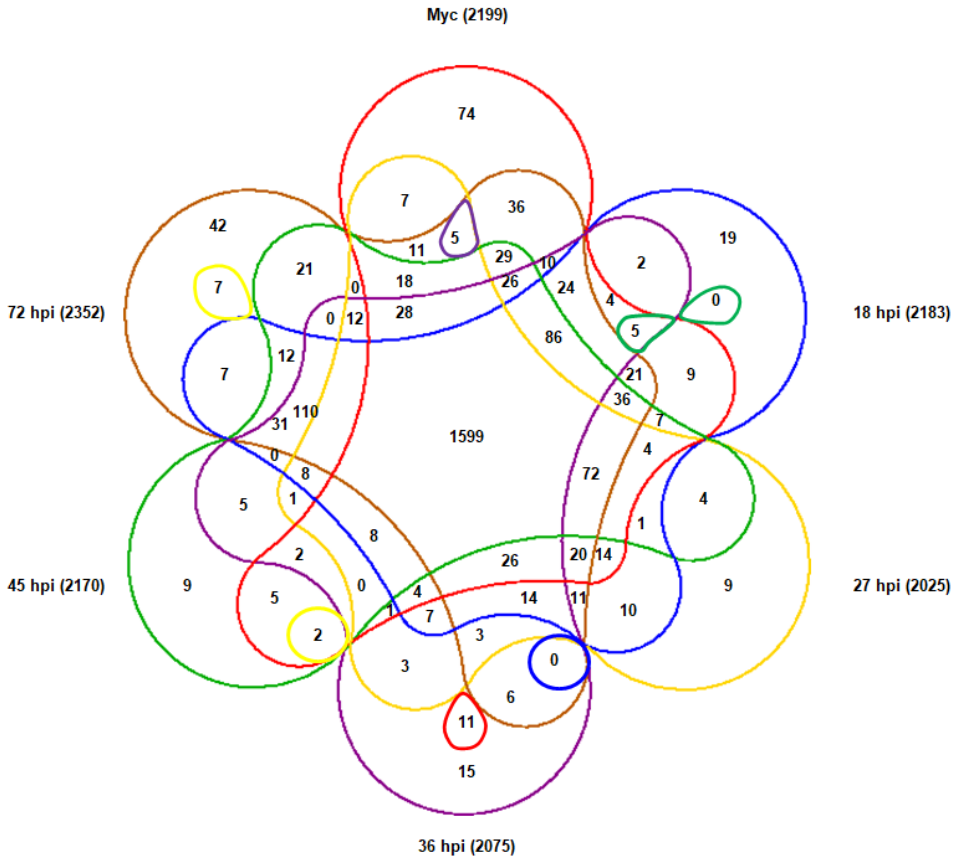

Figure 3. Venn diagram showing the number of lncRNAs expressed among stages.

II. 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, test for equality of proportions: p = 5.635e-09) (Figure 4A); lncRNAs (mean length = 1,584 nt) had shorter full-length transcripts than did mRNAs (mean length = 2,108 nt) (Figure 4B). LncRNAs had fewer exons than did mRNAs (Figure 4C); 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) (Figure 4D).

Conservation of *M. oryzae* lncRNAs was assessed by comparison to known lncRNAs from RNAcentral (Consortium, 2021). No significantly conserved lncRNA was discovered. 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* (Figure 5).

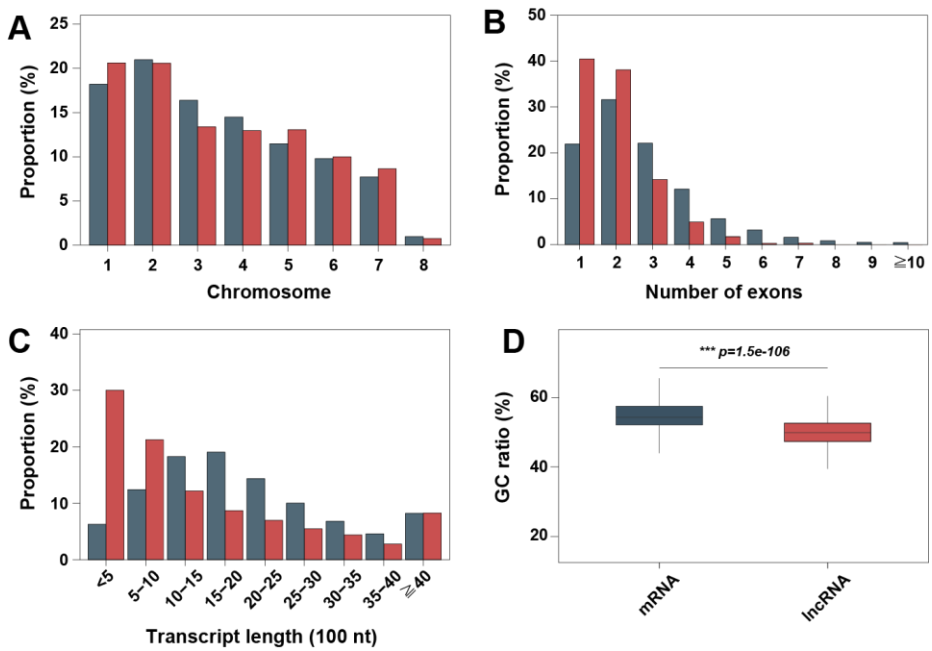

Figure 4. Genomic features of *M. oryzae* lncRNAs. (A) Distributions of mRNAs (bluish green) and lncRNAs (red) across chromosomes. (B) Distribution of transcript lengths. (C) Distribution of exon numbers per transcript. (D) GC ratio (%). The Wilcoxon–Mann–Whitney test confirmed a significant difference in GC ratio between the two groups. *** $p < 0.001$

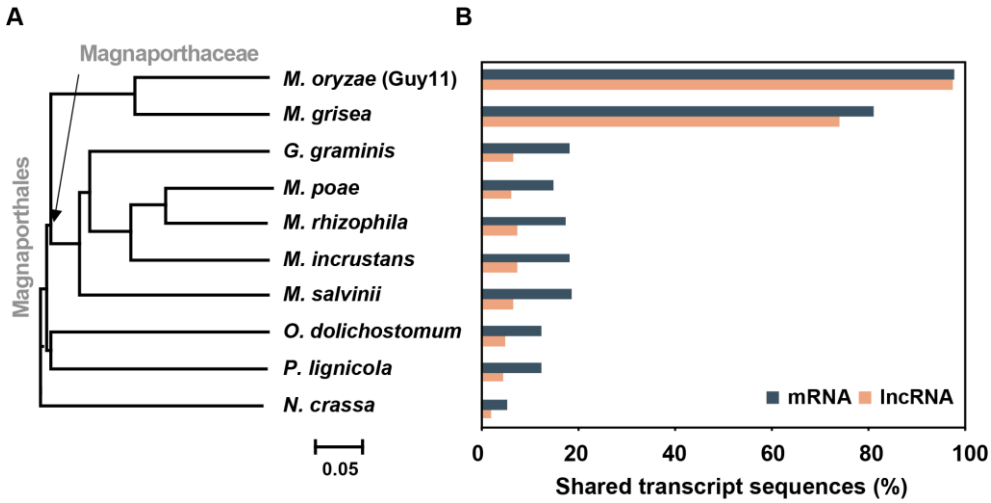

Figure 5. Conservation of lncRNAs among Magnaporthales species and *N. crassa*. Proportion of mRNAs (bluish green) and lncRNAs (brown) identified using BLASTn and *M. oryzae* lncRNA sequences against the genome with an e-value cutoff < 1e-5. *M. oryzae*, *Magnaporthe oryzae*; *M. grisea*, *Magnaporthe grisea*; *G. gramininis, Gaeumannomyces graminis*; *M. poae, Magnaporthe poae*; *M. rhizophila, Magnaporthiopsis rhizophila*; *M. incrustans, Magnaporthiopsis incrustans*; *M. salvinii, Magnaporthe salvinii*; *O. dolichostomum, Ophioceras dolichostomum*; *P. lignicola, Pseudohalonectria lignicola*; *N. crassa, Neurospora crassa*.

III. 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 (Figure 6A, 6B). 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 (Figure 6C). 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 the assessment of the tissue specificity index τ (Tau) (Yanai et al., 2005). 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) (Figure 6D).

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 ISELs 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) (Figure 7A, Table 3).

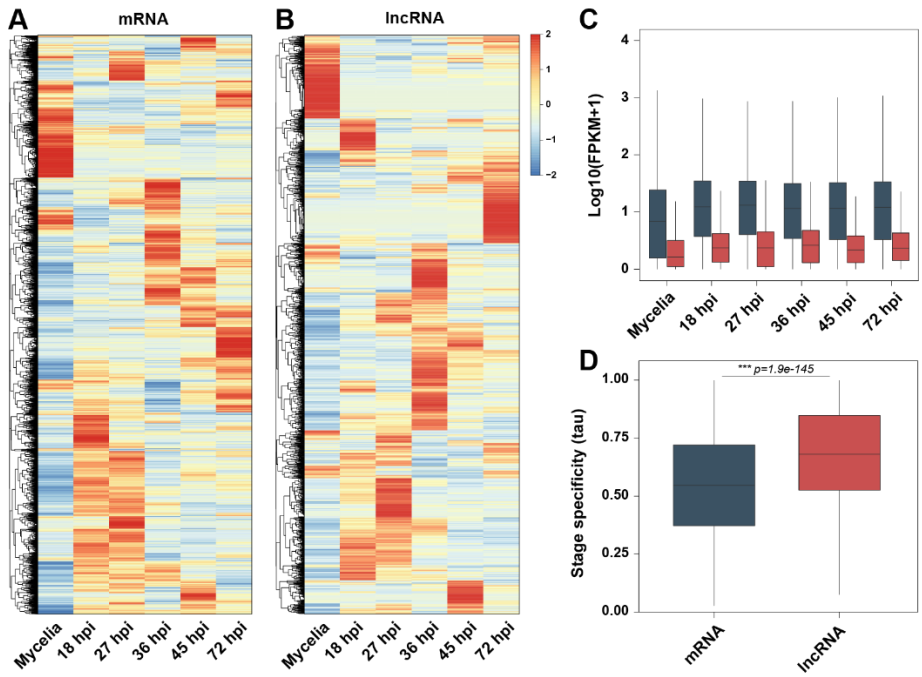

Figure 6. LncRNA expression level and pattern. (A), (B) Expression heatmaps of 9,410 mRNAs and 2,601 lncRNAs, respectively. Z-score normalization was applied to FPKM values across stages. (C) Boxplot of mRNA (bluish green) and lncRNA (red) expression patterns across developmental and infection stages. (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. *** $p < 0.001$

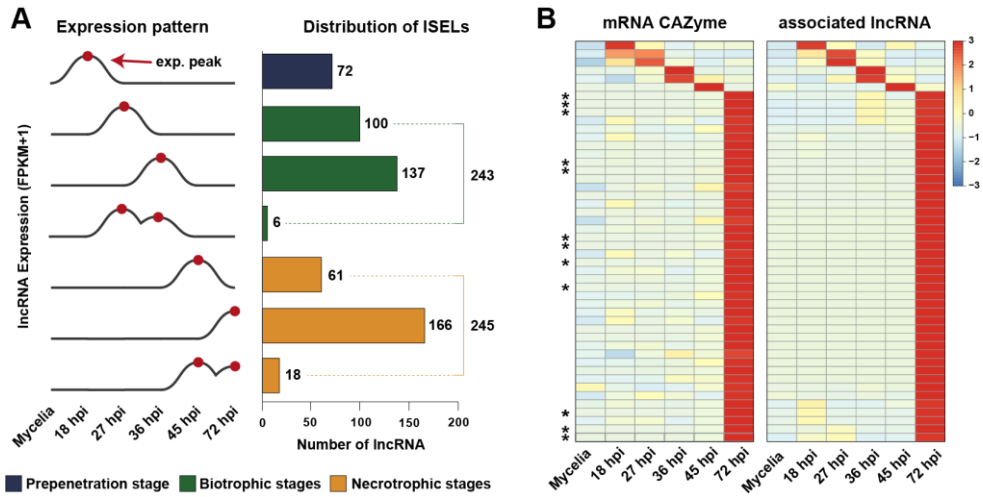

Figure 7. Infection stage-specific lncRNAs and their target genes. (A) Numbers of lncRNAs according to expression pattern. Red spots indicate an expression peak at each stage. (B) Expression heatmap of infection-specifically expressed lncRNAs and their CAZyme target genes. FPKM values were normalized across conditions based on Z-scores. Normalizations of mRNAs and lncRNAs were performed separately. * indicates a functionally characterized xylanase. CAZyme, Carbohydrate-active enzyme.

Table 3. Number of stage-specifically expressed lncRNAs

Stage	Number of lncRNAs
Mycelia	195
Pre-penetration	72
Biotrophic stage	243
Necrotrophic stage	245

IV. 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 (Table 4). The terms “binding" and “mycelium development" were enriched for the target gene set for mycelia-specific lncRNA expression (Table 5).

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 (Figure 7B). ISEL target genes were queried against PHI-base to identify pathogenesis-related genes (Urban et al., 2020). As a result, 23 target genes were matched to the gene set from PHI-base (Table 6). The proportion of the pathogenesis-related genes from PHI-base was higher in the target genes of ISELs than those of non-ISELs (two-proportions z-test: $p = 0.01085$) (Table 7). 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*).

Table 4. Enriched GO terms of infection stage-specifically expressed lncRNA target genes

GO ID	Category	Description	Gene number	FDR	Source
GO:0030248	MF	cellulose binding	9	0.0000123	Blast2GO
GO:0030247	MF	polysaccharide binding	9	0.0000123	Blast2GO
GO:0016798	MF	hydrolase activity, acting on glycosyl bonds	23	0.0000123	Blast2GO
GO:0004553	MF	hydrolase activity, hydrolyzing O-glycosyl compounds	23	0.0000123	Blast2GO
GO:0051701	BP	interaction with host	14	0.0001	AgriGO
GO:0052047	BP	interaction with other organism via secreted substance during symbiotic interaction	12	0.00016	AgriGO
GO:0044046	BP	interaction with host via substance released outside of symbiont	12	0.00016	AgriGO
GO:0052048	BP	interaction with host via secreted substance during symbiotic interaction	12	0.00016	AgriGO
GO:0052051	BP	interaction with host via protein secreted by type II secretion system	11	0.00052	AgriGO
GO:0052211	BP	interaction with other organism via protein secreted by type II secretion system during symbiotic interaction	11	0.00052	AgriGO
GO:0030246	MF	carbohydrate binding	11	0.00062	Blast2GO
GO:0005975	BP	carbohydrate metabolic process	24	0.0012	Blast2GO
GO:0051704	BP	multi-organism process	16	0.005	AgriGO
GO:0044419	BP	interspecies interaction between organisms	14	0.028	AgriGO
GO:0044403	BP	symbiosis, encompassing mutualism through parasitism	14	0.028	AgriGO

FDR: false discovery rate; MF: molecular function; BP: biological process

Table 5. Enriched GO terms of mycelia-specifically expressed lncRNA target genes

GO ID	Category	Description	Gene number	FDR	Source
GO:0005488	MF	binding	127	1.8E-09	Blast2GO
GO:0007275	BP	multicellular organismal development	67	2.8E-10	AgriGO
GO:0043581	BP	mycelium development	67	2.8E-10	AgriGO
GO:0032502	BP	developmental process	72	2.8E-10	AgriGO
GO:0032501	BP	multicellular organismal process	67	2.8E-10	AgriGO
GO:0048856	BP	anatomical structure development	70	2.8E-10	AgriGO

FDR: false discovery rate; MF: molecular function; BP: biological process

Table 6. 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_2Cys_6 transcription factor
MSTRG.8913.3	Trans	*MoPER1*	GPI anchored-related gene
MSTRG.14270.1	Trans		
MSTRG.14270.2	Trans		
MSTRG.14853.1	Trans		
MSTRG.4487.2	Trans	MGG_08331T0	Endo-1,4-beta-xylanase
MSTRG.4487.3	Trans		
MSTRG.8648.1	Trans		
MSTRG.8648.2	Trans		
MSTRG.8648.1	Trans		
MSTRG.8648.2	Trans	MGG_08424T0	Endo-1,4-beta-xylanase
MSTRG.8655.2	Trans		
MSTRG.14853.1	Trans		
MSTRG.8407.2	Trans	*MPG1*	Hydrophobin
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 MSTRG.13998.6	Cis Trans	*Pmc1*	Vacuolar membrane-located Ca^{2+} pump
MSTRG.1930.1 MSTRG.1930.3	Trans Trans	*MST12*	STE-like transcription factor
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_2Cys_6 transcription factor
MSTRG.14853.1	Trans	*FZC12*	Zn_2Cys_6 transcription factor
MSTRG.1930.1 MSTRG.1930.3	Trans	*FZC42*	Zn_2Cys_6 transcription factor

Table 7. Contingency table of pathogenesis-related genes matched to PHI-base

	Target genes of ISELs	Target genes of non-ISELs
Genes matched to PHI-base	23	18
Genes not matched to PHI-base	342	581

V. Verification of lncRNA production

LncRNA production was verified using RNA samples from vegetative mycelia and infected rice leaves (Figure 8). 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.

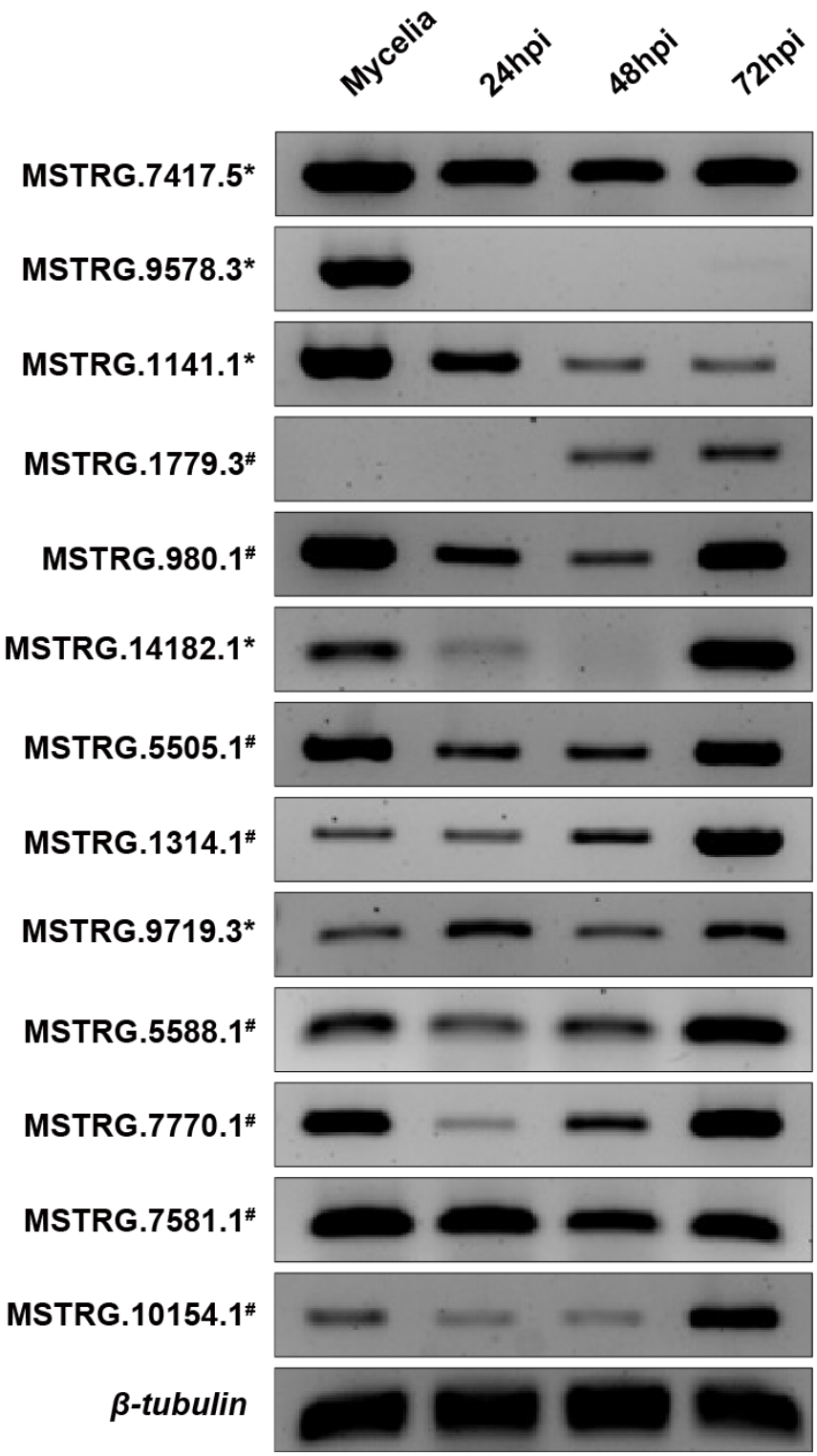
Mycelia
24hpi
48hpi
72hpi
MSTRG.7417.5*
MSTRG.9578.3*
MSTRG.1141.1*
MSTRG.1779.3#
MSTRG.980.1#
MSTRG.14182.1*
MSTRG.5505.1#
MSTRG.1314.1#
MSTRG.9719.3*
MSTRG.5588.1#
MSTRG.7770.1#
MSTRG.7581.1#
MSTRG.10154.1#
β-tubulin

Figure 8. 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.

DISCUSSION

LncRNAs modulate gene expression at the transcriptional and post-transcriptional levels; they have important roles in various metabolic pathways throughout eukaryotic species (Marchese et al., 2017). Most lncRNA studies have been performed in model yeasts, while the functional characterization and profiling of plant pathogen lncRNAs have been rarely studied (Li et al., 2021; Till et al., 2018a). Genome-wide profiling of plant pathogen lncRNAs in the disease process has been performed in the rice smut fungus *U. virens* (Tang et al., 2021). 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 (Figure 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 (Quinn and Chang, 2016). Specifically expressed transcripts at infection stages would also be underrepresented due to low sequencing depth and ambiguity of strand specificity (Zhao et al., 2015).

M. oryzae lncRNAs had shorter transcript lengths, 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

(Figure 2, Figure 3) (Quinn and Chang, 2016). Low GC content of lncRNAs would be related to their temporal-specific expression and low stability. The positive correlation between the GC content and stability of transcripts was also reported (Clark et al., 2012). 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 (Till et al., 2020; Wang et al., 2021). However, previous fungal lncRNA profiling studies considered only cis-acting lncRNAs (Kim et al., 2018; Tang et al., 2021). 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 a well-defined stage-specifically expressed lncRNAs. As expected, we identified more ISELs than mycelia-specifically expressed lncRNAs. Increased expression levels of lncRNAs during the developmental process were also observed in *Fusarium graminearum* sexual reproduction and *U. virens* disease development (Kim et al., 2018; Tang et al., 2021). 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 4: Table S2). In *U. virens*, transport-related GO terms were enriched during all stages (Tang et al., 2021). 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) (Figure 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 (Quoc and Bao Chau, 2017). A cellulase-regulating lncRNA was reported in the saprophyte *T. reesei*, where cellulases are essential for trophism (Till et al., 2018b). 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 (Dodds and Rathjen, 2010). Taken together, the findings thus far suggest that lncRNAs function in the pathogenesis of *M.* oryzae by regulating associated genes.

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 suggests 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.

LITERATURE CITED

Arthanari, Y., Heintzen, C., Griffiths-Jones, S., and Crosthwaite, S.K. 2014. Natural antisense transcripts and long non-coding RNA in *Neurospora crassa*. PLoS ONE. 9:e91353.

Berretta, J., Pinskaya, M., and Morillon, A. 2008. A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in *S. cerevisiae*. Genes Dev. 22:615-626.

Brannan, C.I., Dees, E.C., Ingram, R.S., and Tilghman, S.M. 1990. The product of the *H19* gene may function as an RNA. Mol. Cell. Biol. 10:28-36.

Brockdorff, N., Ashworth, A., Kay, G.F., McCabe, V.M., Norris, D.P., Cooper, P.J., Swift, S., and Rastan, S. 1992. 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. 71:515-526.

Chacko, N., Zhao, Y., Yang, E., Wang, L., Cai, J.J., and Lin, X. 2015. The lncRNA *RZE1* controls cryptococcal morphological transition. PLoS Genet. 11:e1005692.

Choi, J., Cheong, K., Jung, K., Jeon, J., Lee, G.-W., Kang, S., Kim, S., Lee, Y.-W., and Lee, Y.-H. 2013. CFGP 2.0: a versatile web-based platform for supporting comparative and evolutionary genomics of fungi and Oomycetes. Nucleic Acids Res. 41:D714-D719.

Clark, M.B., Amaral, P.P., Schlesinger, F.J., Dinger, M.E., Taft, R.J., Rinn, J.L., Ponting, C.P., Stadler, P.F., Morris, K.V., and Morillon, A. 2011. The reality

of pervasive transcription. PLoS Biol. 9:e1000625.

Clark, M.B., Johnston, R.L., Inostroza-Ponta, M., Fox, A.H., Fortini, E., Moscato, P., Dinger, M.E., and Mattick, J.S. 2012. Genome-wide analysis of long noncoding RNA stability. Genome Res. 22:885-898.

Consortium, R. 2021. RNAcentral 2021: secondary structure integration, improved sequence search and new member databases. Nucleic Acids Res. 49:D212-D220.

Dean, R., Van Kan, J.A., Pretorius, Z.A., Hammond-Kosack, K.E., Di Pietro, A., Spanu, P.D., Rudd, J.J., Dickman, M., Kahmann, R., and Ellis, J. 2012. The Top 10 fungal pathogens in Mol. Plant Pathol. Mol. Plant Pathol. 13:414-430.

Dean, R.A., Talbot, N.J., Ebbole, D.J., Farman, M.L., Mitchell, T.K., Orbach, M.J., Thon, M., Kulkarni, R., Xu, J.-R., and Pan, H. 2005. The genome sequence of the rice blast fungus *Magnaporthe grisea.* Nature. 434:980-986.

Dodds, P.N., and Rathjen, J.P. 2010. Plant immunity: towards an integrated view of plant–pathogen interactions. Nat. Rev. Genet. 11:539-548.

Donaldson, M.E., Ostrowski, L.A., Goulet, K.M., and Saville, B.J. 2017. Transcriptome analysis of smut fungi reveals widespread intergenic transcription and conserved antisense transcript expression. BMC Genomics. 18:340.

Donaldson, M.E., and Saville, B.J. 2013. *Ustilago maydis* natural antisense transcript expression alters mRNA stability and pathogenesis. Mol. Microbiol. 89:29-51.

Dong, Y., Li, Y., Zhao, M., Jing, M., Liu, X., Liu, M., Guo, X., Zhang, X., Chen, Y.,

and Liu, Y. 2015. 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. 11:e1004801.

Fernandez, J., and Orth, K. 2018. Rise of a cereal killer: the biology of *Magnaporthe oryzae* biotrophic growth. Trends Microbiol. 26:582-597.

Gil, N., and Ulitsky, I. 2020. Regulation of gene expression by cis-acting long non-coding RNAs. Nat. Rev. Genet. 21:102-117.

Götz, S., García-Gómez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talón, M., Dopazo, J., and Conesa, A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36:3420-3435.

Ho, E.C., Donaldson, M.E., and Saville, B.J. 2010. Detection of antisense RNA transcripts by strand-specific RT-PCR. RT-PCR Protocols. Springer. 125-138.

Jalali, S., Jayaraj, G.G., and Scaria, V. 2012. Integrative transcriptome analysis suggest processing of a subset of long non-coding RNAs to small RNAs. Biol. Direct. 7:1-13.

Jeon, J., Lee, G.-W., Kim, K.-T., Park, S.-Y., Kim, S., Kwon, S., Huh, A., Chung, H., Lee, D.-Y., and Kim, C.-Y. 2020. Transcriptome profiling of the rice blast fungus *Magnaporthe oryzae* and its host *Oryza sativa* during infection. Mol. Plant Microbe Interact. 33:141-144.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., and Nuka, G. 2014. InterProScan 5: genome-scale

protein function classification. Bioinformatics. 30:1236-1240.

Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E.P., Rivas, E., Eddy, S.R., Bateman, A., Finn, R.D., and Petrov, A.I. 2018. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. Nucleic Acids Res. 46:D335-D342.

Kapranov, P., Willingham, A.T., and Gingeras, T.R. 2007. Genome-wide transcription and the implications for genomic organization. Nat. Rev. Genet. 8:413-423.

Kawahara, Y., Oono, Y., Kanamori, H., Matsumoto, T., Itoh, T., and Minami, E. 2012. Simultaneous RNA-seq analysis of a mixed transcriptome of rice and blast fungus interaction. PLoS ONE. 7:e49423.

Kim, D., Langmead, B., and Salzberg, S.L. 2015. HISAT: a fast spliced aligner with low memory requirements. Nat. Methods. 12:357-360.

Kim, W., Miguel-Rojas, C., Wang, J., Townsend, J.P., and Trail, F. 2018. Developmental dynamics of long noncoding RNA expression during sexual fruiting body formation in *Fusarium graminearum*. mBio. 9:e01292-01218.

Kramer, C., Loros, J.J., Dunlap, J.C., and Crosthwaite, S.K. 2003. Role for antisense RNA in regulating circadian clock function in *Neurospora crassa*. Nature. 421:948-952.

Laurent, G.S., Wahlestedt, C., and Kapranov, P. 2015. The Landscape of long noncoding RNA classification. Trends Genet. 31:239-251.

Li, J., Liu, X., Yin, Z., Hu, Z., and Zhang, K.-Q. 2021. An overview on identification and regulatory mechanisms of long non-coding RNAs in fungi. Front. Microbiol. 12:638617.

Li, L., Wang, M., Wu, X., Geng, L., Xue, Y., Wei, X., and Jia, Y. 2016. A long non-coding RNA interacts with Gfra1 and maintains survival of mouse spermatogonial stem cells. Cell Death Dis. 7:e2140-e2140.

Ma, L., Bajic, V.B., and Zhang, Z. 2013. On the classification of long non-coding RNAs. RNA Biol. 10:924-933.

Marchese, F.P., Raimondi, I., and Huarte, M. 2017. The multidimensional mechanisms of long noncoding RNA function. Genome Biol. 18:1-13.

Mercer, T.R., Dinger, M.E., and Mattick, J.S. 2009. Long non-coding RNAs: insights into functions. Nat. Rev. Genet. 10:155-159.

Morrison, E.N., Donaldson, M.E., and Saville, B.J. 2012. Identification and analysis of genes expressed in the *Ustilago maydis* dikaryon: uncovering a novel class of pathogenesis genes. Can. J. Plant Pathol. 34:417-435.

Nawrocki, E.P., and Eddy, S.R. 2013. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 29:2933-2935.

Parra-Rivero, O., Pardo-Medina, J., Gutiérrez, G., Limón, M.C., and Avalos, J. 2020. A novel lncRNA as a positive regulator of carotenoid biosynthesis in *Fusarium*. Sci. Rep. 10:1-14.

Patel, R.K., and Jain, M. 2012. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLoS ONE. 7:e30619.

Pertea, G., and Pertea, M. 2020. GFF utilities: GffRead and GffCompare. F1000Res. 9:304.

Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., and Salzberg, S.L. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat. Protoc. 11:1650-1667.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. 33:290-295.

Ponting, C.P., Oliver, P.L., and Reik, W. 2009. Evolution and functions of long noncoding RNAs. Cell. 136:629-641.

Quinn, J.J., and Chang, H.Y. 2016. Unique features of long non-coding RNA biogenesis and function. Nat. Rev. Genet. 17:47-62.

Quoc, N.B., and Bao Chau, N.N. 2017. The role of cell wall degrading enzymes in pathogenesis of *Magnaporthe oryzae*. Curr. Protein Pept. Sci. 18:1019-1034.

Skamnioti, P., and Gurr, S.J. 2009. Against the grain: safeguarding rice from rice blast disease. Trends Biotechnol. 27:141-150.

Tafer, H., and Hofacker, I.L. 2008. RNAplex: a fast tool for RNA–RNA interaction search. Bioinformatics. 24:2657-2663.

Tang, J., Chen, X., Yan, Y., Huang, J., Luo, C., Tom, H., and Zheng, L. 2021. Comprehensive transcriptome profiling reveals abundant long non-coding RNAs associated with development of the rice false smut fungus, *Ustilaginoidea virens*. Environ. Microbiol. 23:4998-5013.

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., Xu, W., and Su, Z. 2017. agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. Nucleic Acids Res. 45:W122-W129.

Till, P., Derntl, C., Kiesenhofer, D.P., Mach, R.L., Yaver, D., and Mach-Aigner, A.R. 2020. Regulation of gene expression by the action of a fungal lncRNA on a transactivator. RNA Biol. 17:47-61.

Till, P., Mach, R.L., and Mach-Aigner, A.R. 2018a. A current view on long

noncoding RNAs in yeast and filamentous fungi. Appl. Microbiol. Biotechnol. 102:7319-7331.

Till, P., Pucher, M.E., Mach, R.L., and Mach-Aigner, A.R. 2018b. A long noncoding RNA promotes cellulase expression in *Trichoderma reesei*. Biotechnol. Biofuels. 11:1-16.

Urban, M., Cuzick, A., Seager, J., Wood, V., Rutherford, K., Venkatesh, S.Y., De Silva, N., Martinez, M.C., Pedro, H., and Yates, A.D. 2020. PHI-base: the pathogen–host interactions database. Nucleic Acids Res. 48:D613-D620.

Wade, J.T., and Grainger, D.C. 2014. Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. Nat. Rev. Microbiol. 12:647-653.

Wang, J., Zeng, W., Xie, J., Fu, Y., Jiang, D., Lin, Y., Chen, W., and Cheng, J. 2021. A novel antisense long non-coding RNA participates in asexual and sexual reproduction by regulating the expression of *GzmetE* in *Fusarium graminearum*. Environ. Microbiol. 23:4939-4955.

Wang, L., Park, H.J., Dasari, S., Wang, S., Kocher, J.-P., and Li, W. 2013. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res. 41:e74.

Wang, W., Zhuang, Q., Ji, K., Wen, B., Lin, P., Zhao, Y., Li, W., and Yan, C. 2017. Identification of miRNA, lncRNA and mRNA-associated ceRNA networks and potential biomarker for MELAS with mitochondrial DNA A3243G mutation. Sci. Rep. 7:1-13.

Weirick, T., Militello, G., Müller, R., John, D., Dimmeler, S., and Uchida, S. 2016. The identification and characterization of novel transcripts from RNA-seq data. Brief. Bioinform. 17:678-685.

Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., Bar-Even, A., Horn-Saban, S., Safran, M., and Domany, E. 2005. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics. 21:650-659.

Zaynab, M., Fatima, M., Abbas, S., Umair, M., Sharif, Y., and Raza, M.A. 2018. Long non-coding RNAs as molecular players in plant defense against pathogens. Microb. Pathog. 121:277-282.

Zhang, N., Cai, G., Price, D.C., Crouch, J.A., Gladieux, P., Hillman, B., Khang, C.H., LeBrun, M.-H., Lee, Y.-H., and Luo, J. 2018. Genome wide analysis of the transition to pathogenic lifestyles in Magnaporthales fungi. Sci. Rep. 8:1-13.

Zhang, X., Wang, W., Zhu, W., Dong, J., Cheng, Y., Yin, Z., and Shen, F. 2019. Mechanisms and functions of long non-coding RNAs at multiple regulatory levels. Int. J. Mol. Sci. 20:5573.

Zhang, Y.-C., and Chen, Y.-Q. 2013. Long noncoding RNAs: new regulators in plant development. Biochem. Biophys. Res. Commun. 436:111-114.

Zhao, S., Zhang, Y., Gordon, W., Quan, J., Xi, H., Du, S., von Schack, D., and Zhang, B. 2015. Comparison of stranded and non-stranded RNA-seq transcriptome profiling and investigation of gene overlap. BMC Genomics. 16:675.

CHAPTER III

Comprehensive genome-wide analysis of non-coding RNAs reveals functions of lncRNA-sRNA crosstalk in the rice blast fungus *Magnaporthe oryzae*

This chapter is in preparation for publication.

ABSTRACT

The crosstalks between two classes of non-coding RNAs, small non-coding RNAs (sRNAs) and long non-coding RNAs (lncRNAs), involve diverse biological pathways. However, these crosstalks have not been investigated in the context of plant pathogenic fungi. Therefore, we profiled sRNAs and lncRNAs under Dicer deficiency and detected 1,022 Dicer-dependent Argonaute (AGO)-enriched sRNAs from 348 loci in the model fungus *Magnaporthe oryzae*. We identified 29 sRNA loci originating from 38 lncRNAs and 100 lncRNAs regulated by sRNAs from 61 sRNA loci. Functional enrichment analysis showed that different biological pathways were involved in non-coding RNAs depending on the type of crosstalk between sRNAs and lncRNAs. In addition, we constructed a lncRNA-sRNA-mRNA network and identified pathogenesis-related subnetworks. This study provides a comprehensive view of non-coding RNAs in the rice blast fungus, as well as plant pathogenic fungi.

INTRODUCTION

As high-throughput sequencing technology has advanced, the repertoire of transcripts has expanded (Lowe et al., 2017). A large proportion of these transcripts have low protein-coding potential (Zhao et al., 2021). Non-coding RNAs are classified into small non-coding RNAs (sRNAs) and long non-coding RNAs (lncRNAs), with 200 nucleotides commonly being used as the cutoff length (Laurent et al., 2015; Ma et al., 2013).

sRNAs are involved in gene regulation in the RNA interference (RNAi) pathway at the transcriptional and post-transcriptional levels (Hannon, 2002; Wilson and Doudna, 2013). Dicer cleaves long double-stranded RNA (dsRNA) and hairpin-structured RNA into short single-stranded RNAs (ssRNAs). Argonaute (AGO) binds these ssRNAs for incoporation into the RNA-induced silencing complex (RISC). RISC functions as a key regulator via mRNA degradation, translation repression, or heterochromatin formation.

LncRNAs modulate the transcriptome through multiple pathways, including epigenetic, transcriptional, post-transcriptional, translational, and post-translational pathways (Mercer et al., 2009; Ponting et al., 2009). LncRNAs are transcribed by RNA polymerase II and are often 5′-end capped, spliced, and polyadenylated (Nojima and Proudfoot, 2022). Although they are transcribed similarly to mRNAs, they have different characteristics. LncRNAs have shorter lengths, lower expression, and a lower sequence conservation level than mRNAs (Quinn and Chang, 2016). They are classified as sense lncRNAs, antisense lncRNAs, intergenic lncRNAs, or

intronic lncRNAs depending on their genomic location and context (Laurent et al., 2015; Ma et al., 2013).

Crosstalks between sRNAs and lncRNAs have categorized into three types. First, lncRNAs can be precursors for sRNAs. For example, some Dicer-like-1-dependent sRNAs originate from lncRNAs in *Arabidopsis thaliana* to regulate the expression of mRNAs (Ma et al., 2014). Second, sRNAs regulate the expression of lncRNAs. *CDR1-AS* circular RNA is cleaved, and downregulated, by miR-671 (Hansen et al., 2011). Finally, lncRNAs regulate the activity of sRNAs. In the competing endogenous RNAs (ceRNAs) hypothesis, lncRNAs compete with other transcripts, including mRNAs, for sRNA targets (Salmena et al., 2011). For instance, *PTENP1* transcribed from a pseudogene of tumor suppressor *PTEN* has the potential to stabilize homologous mRNAs through competition (Tay et al., 2011). While several studies have been conducted on the genome-wide identification and functional characterization of lncRNAs or sRNAs in fungi, there is little research on crosstalks between these two classes of non-coding RNAs (Li et al., 2021a; Torres-Martínez and Ruiz-Vázquez, 2017). Genome-wide lncRNA profiling was performed of RNAi machinery gene deletion in two fission yeast species (*Schizosaccharomyces pombe*, *Naumovozyma castellii*), without sRNA analysis (Atkinson et al., 2018; Szachnowski et al., 2019). In *Fusarium graminearum*, sRNA-enriched loci overlapping with lncRNAs were identified and the expression levels of two classes of non-coding RNAs were correlated during sexual development (Kim et al., 2018).

Magnaporthe oryzae causes rice blast disease, which is responsible for severe yield losses in cultivated rice worldwide (Nalley et al., 2016; Skamnioti and Gurr, 2009). *M. oryzae* and rice have been studied as a model of the interactions between

fungal pathogens and plant hosts (Dean et al., 2012).. sRNA profiling has been performed under diverse conditions including starvation, infection, and RNAi machinery deficiency (Lee et al., 2022; Nguyen et al., 2018; Nunes et al., 2011; Raman et al., 2017; Raman et al., 2013). LncRNA profiling was also performed during disease development (Choi et al., 2022; Li et al., 2021b). However, an integrative analysis of the two non-coding RNA classes has not been performed.

In this study, we profiled lncRNAs and sRNAs in the context of RNAi machinery gene deficiency in *M. oryzae*. We identified Dicer-dependent sRNAs and their associated lncRNAs, performed functional enrichment analysis of their target genes, and found pathogenesis-related genes. We constructed a ceRNA network based on RNA-seq and sRNA-seq to reveal the function of lncRNA-sRNA-mRNA crosstalk. This study provides novel insight into the role of non-coding RNAs in host-pathogen interactions.

MATERIALS AND METHODS

I. Collection of RNA-seq and sRNA-seq data

M. oryzae RNA-seq and sRNA-seq data sets of wild-type, Dicer-deleted mutants, and MoERI-1-deleted mutants were used to identify lncRNAs and sRNAs in the absence of RNAi machinery genes in M.(BioProject accession no. PRJNA856435) (Lee et al., 2022). The AGO-associated sRNA-seq sets for AGO enrichment analysis were retrieved from DDBJ Sequence Read Archive (SRA accession no. DRA005932) (Nguyen et al., 2018).

II. RNA-seq data analysis

Low-quality reads were removed and adapter sequences were trimmed from raw reads using NGS QC Toolkit v2.3.3 (Patel and Jain, 2012). The resulting reads were mapped against the *M. oryzae* 70-15 reference genome (MG8, Ensembl annotation 29) using HISAT2 v2.0.4 (Dean et al., 2005; Kim et al., 2015). The transcriptome was assembled using the genome-guided method of StringTie v1.3.3 with *de novo* annotation (Pertea et al., 2015). We used fragments per kilobase of transcripts per million mapped reads (FPKM) as the expression index. If the FPKM was < 1 under all conditions, the transcript was considered to be predicted, but not detected. Detected transcripts were used in the subsequent analysis.

We used an established computational pipeline to identify lncRNAs (Choi et al., 2022). First, transcripts whose spliced sequences are shorter than 200 nucleotides

were filtered out. The assembled transcripts were then compared with protein-coding genes and categorized using Gffcompare (Pertea and Pertea, 2020). 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 (Kalvari et al., 2018; Nawrocki and Eddy, 2013). The coding potentials of transcripts were assessed using CPAT v1.2.2 (Wang et al., 2013). Transcripts with coding potential below the cutoff were included; transcripts containing any known Pfam domain were removed using InterProScan v5.29-68.0 (Jones et al., 2014).

III. sRNA-seq data analysis

Adapter sequences were removed by Cutadapt v1.8.1 (Martin, 2011). Reads were filtered by quality and size using Sickle v1.33 (Joshi and Fass, 2011). Processed reads were mapped to the reference genome of the *M. oryzae* strain 70-15 (MG8, Ensembl annotation 29) using bowtie v1.2.2 (Dean et al., 2005; Langmead et al., 2009). To compare the different libraries, the coordinates were fixed and sRNA loci were clustered by ShortStack3 (Johnson et al., 2016). The abundance of each locus was normalized to reads per kilobase per million reads (RPKM). Dicer-dependent sRNA loci were defined as follows: RPKM > 1 in at least one condition and fold-change > 2 in the wild-type compared to the Dicer-deleted mutant. AGO-enriched sRNAs were selected based on the following criteria: RPKM > 3 in at least one AGO-binding library and fold-change > 3 compared to the control set.

IV. Target gene prediction and analysis

Prediction of transcripts targeted by sRNA was performed using psRNATarget and TargetFinder (Dai et al., 2018; Fahlgren and Carrington, 2010). If a small RNA was predicted to target the same target transcript by both tools and the Pearson correlation coefficient was below −0.3, the sRNA-target transcript pair was used in further analysis. Genes within a 10-kb window upstream or downstream of lncRNAs and an absolute value of Pearson correlation coefficient > 0.9 were predicted to be targeted by lncRNAs. Target genes were then subjected to Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) term enrichment analysis (5% false discovery rate) using g:Profiler (Raudvere et al., 2019). Pathogenesis-related genes were identified by querying target genes against a pathogen-host interactions database (PHI-base) (Urban et al., 2020). A lncRNA-sRNA-mRNA network was constructed using Cytoscape v3.9.1 (Smoot et al., 2011).

RESULTS

I. Identification of lncRNAs and Dicer-dependent sRNAs

We used sRNA-seq and RNA-seq data from Dicer-deleted mutants, *MoERI-1*-deleted mutants, and wild-type as the control to identify lncRNA-associated sRNAs (Figure 1A). In total, we predicted 28,335 transcripts, 15,580 of which were identified as novel transcripts. A total of 2,171 lncRNAs were identified, with the largest proportion being antisense transcripts (n = 1,001; 46.1%), followed by intergenic lncRNAs (n = 816; 37.6%) and sense lncRNAs (n = 352; 16.2%) (Table 1). LncRNAs showed a tendency toward low overall expression compared with mRNAs in both the wild-type and Dicer-deleted mutant (Figure 2A). The average expression level of lncRNAs in the Dicer-deleted mutant was similar to that of the wild-type (Figure 2B).

We constructed 7,397 sRNA loci using Shortstack3 for comparison among conditions (Johnson et al., 2016). We identified Dicer-dependent sRNAs based on changes in expression level. We identified 1,695 Dicer-dependent sRNA loci with more than two-fold decrease in expression of the Dicer-deleted mutant. We then selected AGO-enriched sRNAs that were likely to regulate gene expression (because loading into AGO proteins is a prerequisite for the functioning of sRNAs). We identified 1,022 Dicer-dependent AGO-enriched sRNAs from among 348 loci. They showed U preferences at 5`-ends and peaks at 19~20 nt, regardless of the type of AGO protein (Figure 3A and 3B).

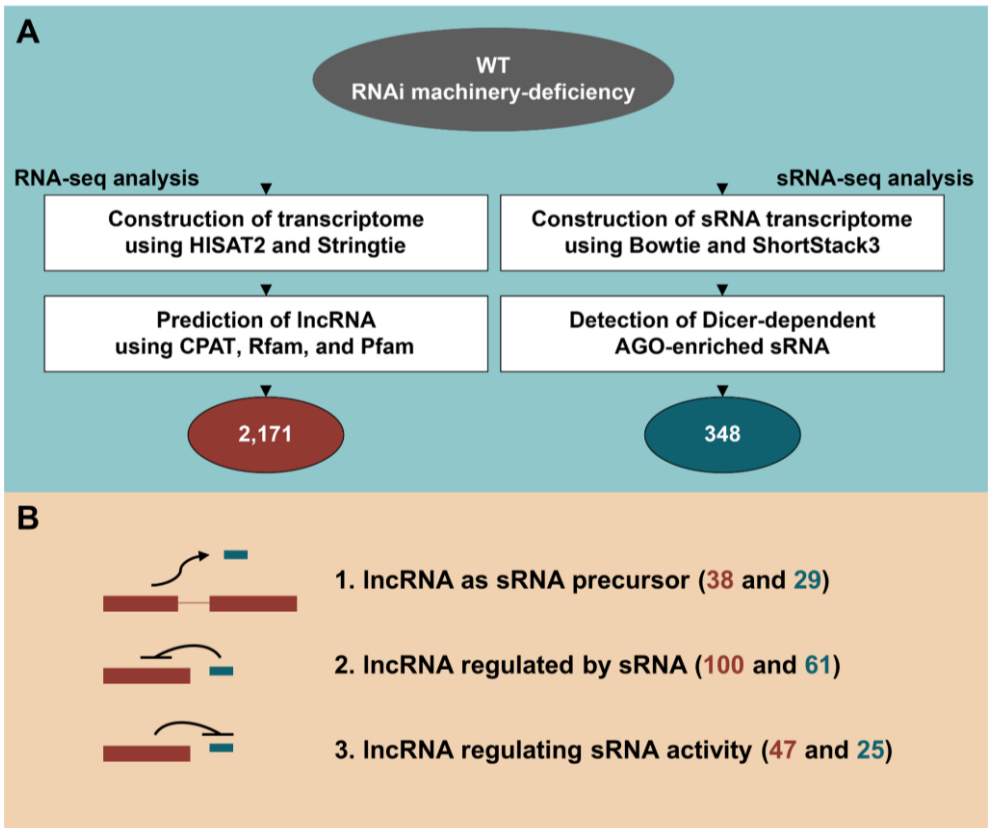

Figure 1. Schematic pipeline for identifying lncRNAs and associated sRNAs under RNAi machinery-deficient conditions in *M. oryzae*. (A) Bioinformatic pipeline for lncRNA identification using RNA-seq and sRNA-seq data. CPAT, Coding Potential Assessment Tool. (B) Type of crosstalk between lncRNAs and sRNAs. The numbers in parentheses mean the number of lncRNAs and that of sRNA loci, respectively.

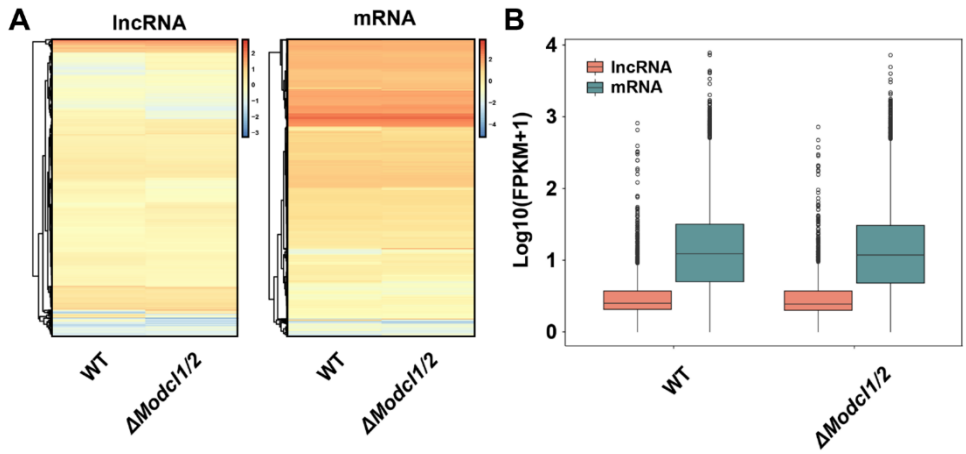

Figure 2. Pattern of lncRNA expression in the wild-type and Dicer-deleted mutant. (A) Heatmaps of lncRNA and mRNA. Log10 normalization was applied to FPKM values across libraries. (B) Boxplot of lncRNA (pink) and mRNA (blue).

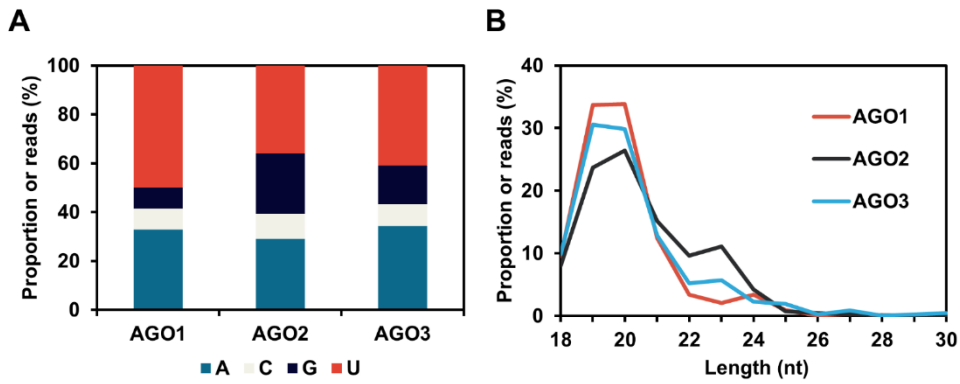

Figure 3. Characteristics of *M. oryzae* Dicer-dependent AGO-enriched sRNAs.

(A) Nucleotide composition at the 5'-ends of sRNAs. (B) Length distribution of sRNAs.

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

Class of lncRNA	Number of lncRNAs
Sense lncRNA	352
Antisense lncRNA	1,001
Intergenic lncRNA	816
Intronic lncRNA	2

II. Identification of small RNAs originating from lncRNAs

We aimed to identify lncRNAs serving as precursors of Dicer-dependent AGO-enriched sRNAs. We identified 72 lncRNAs overlapping with 59 sRNA loci, and selected pairs from among 38 lncRNAs and 29 sRNA loci with positive correlations (Pearson correlation coefficient > 0.3) (Figure 1B). LncRNAs serving as sRNA precursors were mainly antisense transcripts (19; 50.0%), followed by sense lncRNAs (15; 39.5%) and intergenic lncRNAs (3; 7.9%) (Table 2). The lncRNA classes differed compared to those for total lncRNAs. sRNAs originating from lncRNAs were predicted to target 35 protein-coding genes (Table 3). GO and KEGG term enrichment analysis revealed that terms related to glycosyltransferase were enriched in sRNA target genes (Figure 4). Target genes were queried against PHI-base to identify pathogenesis-related genes (Urban et al., 2020). Two target genes were matched to the gene set from PHI-base: the MoCKb regulatory subunit homolog and Som1 homolog (Table 4).

Table 2. Classification of lncRNAs which are sRNA precursors

Class of lncRNA	Number of lncRNAs
Sense lncRNA	15
Antisense lncRNA	19
Intergenic lncRNA	3
Intronic lncRNA	1

Table 3. Protein-coding genes targeted by sRNAs originating from lncRNAs

Gene ID	Description
MGG_17026T0	hypothetical protein
MGG_05654T0	hypothetical protein
MGG_04708T1	hypothetical protein
MGG_04773T0	oligosaccharyl transferase STT3 subunit
MGG_07572T0	high-affinity nickel transporter nic1
MGG_13483T0	glutamyl-tRNA (Gln) amidotransferase subunit A
MGG_05035T0	hypothetical protein
MGG_17648T0	hypothetical protein
MGG_01385T0	choline transporter
MGG_10751T0	peroxisomal copper amine oxidase
MGG_02351T0	tyrocidine synthetase 1
MGG_15622T0	hypothetical protein
MGG_17294T0	hypothetical protein
MGG_08548T0	proline-specific permease
MGG_16813T0	hypothetical protein
MGG_00602T0	cross-pathway control protein 1
MGG_02821T0	dolichyl-di-phosphooligosaccharide-protein glycotransferase
MGG_05008T0	aldehyde dehydrogenase
MGG_09985T0	hypothetical protein
MGG_01933T0	hypothetical protein
MGG_09907T0	ubiquitin carboxyl-terminal hydrolase 13
MGG_13020T0	glucose-repressible alcohol dehydrogenase transcriptional effector
MGG_16018T0	hypothetical protein
MGG_05651T0	*MoCKb2*, MoCKb regulatory subunit homolog
MGG_07481T0	AP-3 complex subunit delta
MGG_03041T0	glucokinase
MGG_09003T0	hypothetical protein
MGG_09991T0	phosphatidylinositol transfer protein SFH5
MGG_11888T0	hypothetical protein
MGG_06360T0	hypothetical protein
MGG_11518T0	G/U mismatch-specific uracil DNA glycosylase
MGG_01068T0	JmjC domain-containing protein
MGG_01815T0	spindle pole body component alp6
MGG_00374T0	hypothetical protein
MGG_04708T0	*MoSOM1*, Som1 homolog

Table 4. Target genes of sRNAs and lncRNAs in PHI-base

NcRNA type	Target gene	Description
lo-sRNA	MGG_05651	*MoCKb2*, MoCKb regulatory subunit homolog
lo-sRNA	MGG_04708	*MoSOM1*, Som1 homolog
st-lncRNA	MGG_01481	*MoPEX7*, peroxisomal targeting signal 2 receptor
st-lncRNA	MGG_02773	*MoMCM1*, MADS-box transcription factor
st-lncRNA	MGG_12655	*AVR-Pi9*, avirulence effector
ce-sRNA	MGG_00365	*MagB*, G protein α subunit
ce-sRNA	MGG_09523	*TRA1*, transcription factor
ce-sRNA	MGG_12349	*CONx1*, transcription factor
ce-sRNA	MGG_12814	*MoAP1*, transcription factor
ce-sRNA	MGG_14767	*SSM2*, non-ribosomal peptide synthetase related to siderophore synthesis
ce-sRNA	MGG_15972	*AVR-Pik*, avirulence effector

lo-sRNA: lncRNA-originating lncRNA; st-lncRNA: sRNA-targeted lncRNA; ce-sRNA: sRNA in ceRNA (competing endogenous RNA) regulatory cascades.

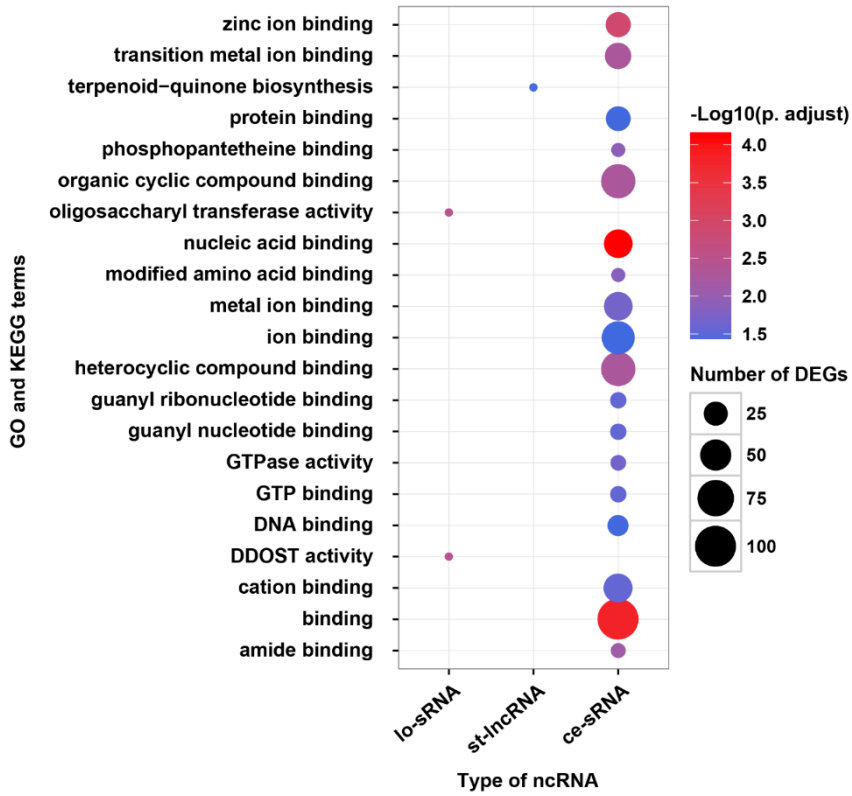

Figure 4. Functional enrichment analysis of protein-coding genes targeted by non-coding RNAs. LncRNA are categorized by type of crosstalk with sRNA; lo-sRNA, lncRNA-originating sRNA. st-lncRNA, sRNA-targeted lncRNA. ce-sRNA, sRNA in ceRNA (competing endogenous RNA) regulatory cascades.

III. Identification of sRNAs regulating lncRNA expression

We investigated lncRNAs that could be regulated by sRNAs. We identified 329 lncRNAs targeted by 275 Dicer-dependent sRNAs from among 102 loci using psRNATarget and TargetFinder (Dai et al., 2018; Fahlgren and Carrington, 2010). Among them, we selected pairs of ncRNAs with negative correlations (Pearson correlation coefficient < -0.3) (Figure 1B). We identified 100 possible target lncRNAs and 54 protein-coding genes targeted by these lncRNAs (Table 5). GO and KEGG term enrichment analysis revealed that the term "terpenoid-quinone biosynthesis" was enriched in lncRNA target genes (Figure 4). Three target genes were matched to the gene set from PHI-base, including a peroxisomal targeting signal 2 receptor, a MADS-box transcription factor, and an avirulence effector (Table 4).

Table 5. Target genes of lncRNAs regulated by sRNAs

Gene ID	Description
MGG_02287T0	O-methyltransferase
MGG_16152T0	hypothetical protein
MGG_14587T0	Ser/Thr protein phosphatase
MGG_06769T0	hypothetical protein
MGG_06858T0	hypothetical protein
MGG_16203T0	SMC1A protein
MGG_09918T0	hypothetical protein
MGG_09919T0	amino transferase
MGG_16432T0	hypothetical protein
MGG_05650T0	PAP2 domain-containing protein
MGG_04738T0	short chain dehydrogenase/oxidoreductase
MGG_04737T0	hypothetical protein
MGG_04736T0	hypothetical protein
MGG_04543T0	hypothetical protein
MGG_04567T0	hypothetical protein
MGG_04740T0	hypothetical protein
MGG_08888T0	hypothetical protein
MGG_11317T0	long chain fatty acid oxidase
MGG_08891T0	maltose permease
MGG_08892T0	hypothetical protein
MGG_01481T0	MoPEX7, peroxisomal targeting signal 2 receptor
MGG_01859T0	hypothetical protein
MGG_01863T0	aminopeptidase Y
MGG_01866T0	hypothetical protein
MGG_01867T0	hypothetical protein
MGG_08118T0	hypothetical protein
MGG_08121T0	hypothetical protein
MGG_08123T0	beta-glucosidase
MGG_08206T0	amidohydrolase
MGG_12316T0	hypothetical protein
MGG_08208T0	hypothetical protein
MGG_15211T0	hypothetical protein
MGG_05283T0	uricase
MGG_03290T0	N-(5-amino-5-carboxypentanoyl)-L-cysteinyl-D-valine synthase
MGG_01138T0	cytochrome P450 2C31
MGG_17312T0	hypothetical protein
MGG_01070T0	hypothetical protein
MGG_01068T0	JmjC domain-containing protein
MGG_01067T0	monothiol glutaredoxin-5
MGG_14701T0	hypothetical protein
MGG_00220T0	NADP-dependent alcohol dehydrogenase 6
MGG_11682T0	phenylacetone monooxygenase
MGG_00041T0	O-methyltransferase
MGG_17699T0	aminodeoxychorismate synthase
MGG_03914T1	hypothetical protein
MGG_17729T0	hypothetical protein
MGG_09417T0	hypothetical protein
MGG_17812T0	hypothetical protein

MGG_09795T0	hypothetical protein
MGG_17934T0	hypothetical protein
MGG_17967T0	hypothetical protein
MGG_02775T0	hypothetical protein
MGG_02773T0	*MoMcm1*, MADS-box transcription factor
MGG_12655T0	avirulence effector *Avr-Pi9*

IV. Construction of a lncRNA-sRNA-mRNA network

Based on the ceRNA hypothesis, we constructed a lncRNA-sRNA-mRNA network. We predicted mRNAs targeted by Dicer-dependent AGO-enriched sRNAs for network construction. The network consisted of 47 lncRNAs, 25 sRNA loci, and 208 mRNAs (Figure 1B). GO and KEGG enrichment analysis showed that binding-related terms were enriched in sRNA target genes (Figure 4). We identified five subnetworks including genes extracted from PHI-base within a pathogenesis-related network (Figure 5). This pathogenesis-related network contained 13 lncRNAs, 5 sRNA loci, and 61 mRNAs. Six genes matched to the PHI-base included a G protein α subunit, three transcription factors, a non-ribosomal peptide synthetase related to siderophore synthesis, and an avirulence effector (Table 4).

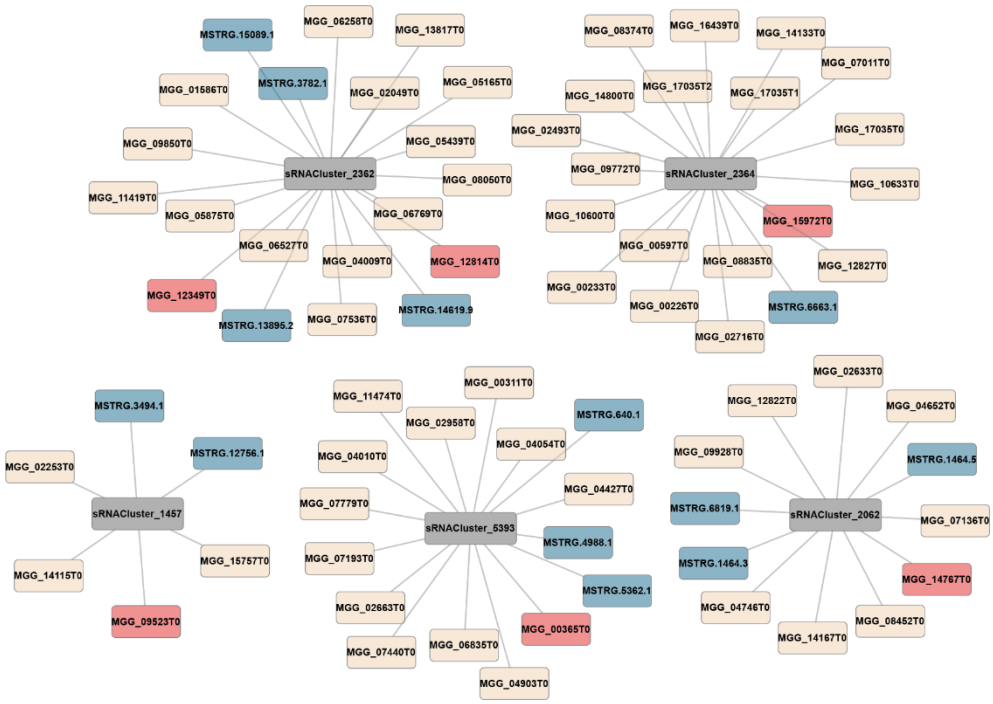

Figure 5. Pathogenesis-related lncRNA-sRNA-mRNA co-expression network. Subnetworks including genes in PHI-base were identified within the total network. LncRNAs, sRNA clusters, mRNAs, and mRNAs matched to genes in PHI-base are shown as blue, green, yellow, and red rectangles, respectively.

DISCUSSION

In plants and animals, lncRNAs serving as sRNA precursors, and sRNAs as transcription regulators of lncRNAs, have been studied using RNA-seq and sRNA-seq under Dicer deficiency conditions (Ma et al., 2014; Zheng et al., 2014). In fungi, genome-wide profiling and functional analysis of lncRNAs and sRNAs have recently been carried out (Li et al., 2021a; Torres-Martínez and Ruiz-Vázquez, 2017). However, integrative analyses of lncRNAs and sRNAs have not been reported under Dicer deficiency conditions in the context of plant pathogenic fungi. Because both MoDCL2 and MoDCL1 participate in sRNA biogenesis, we profiled the entire transcriptome using RNA-seq and sRNA-seq data from Δ*Modcl1/2* to detect Dicer-dependent sRNAs and associated lncRNAs (Lee et al., 2022; Raman et al., 2017). We may have missed non-coding RNAs expressed at particular disease stages, including lncRNAs expressed under conditions of infection (Choi et al., 2022). This profiling based on bioinformatic approaches could identify lncRNAs and sRNAs on a genomic scale. However, this study has limitations in functional characterization. Molecular biological techniques such as polysome profiling and degradome sequencing would clarify their functional characteristics. Nevertheless, our results shed light on the crosstalk between lncRNAs and sRNAs.

Dicer-dependent AGO-enriched sRNAs showed a 20-nt peak and 5`-end U preference in the wild-type but not Δ*Modcl1/2* (Figure 3) (Raman et al., 2017). Another study reported that a major 23-nt peak and minor 20-nt peak in the wild-type were not present in Δ*Modcl1/2*, and speculated that 23-nt sRNAs might be

processed into 20-nt sRNAs to function (Lee et al., 2022). According to this hypothesis, we identified Dicer-dependent AGO-enriched sRNAs likely to regulate transcripts.

It was reported that microRNA-like RNAs regulated genes including MGG_07848, MGG_06375, and MGG_17707 in *M. oryzae* (Li et al., 2020). These genes were not among the target genes identified in this study (Supplementary Table 2 and 3), which might be due to a different combination of tools being used for target gene prediction. Sequence complementarity between sRNAs and these genes was predicted when psRNATarget or TargetFinder was used alone. Target genes in this study predicted by both tools were filtered out. Sampling conditions, which differed between the present and previous studies, might also influence the expression of sRNAs and lncRNAs due to their tendency to be stage-specific (Choi et al., 2022; Mohorianu et al., 2011).

Our functional enrichment analysis of target genes showed that different GO and KEGG terms were related to different types of crosstalk between sRNAs and lncRNAs (Figure 4). LncRNAs acting as precursors of sRNAs, lncRNAs targeted by sRNAs, and sRNAs targeting protein-coding genes competing with lncRNAs were involved in glycosyl transfer, terpenoid-quinone biosynthesis and diverse binding, respectively. The network of the last group included pathogenesis-related transcription factors and an avirulence effector. A common target gene of the first and second groups, MGG_01068, was identified (Supplementary Table 2 and 3). Although functional analysis of this gene was not performed, the same histone methylase, *MoJMJ1*, is known to be related to vegetative growth, asexual

reproduction, appressorium formation, and invasive growth in *M. oryzae* (Huh et al., 2017). This common target gene may be regulated through a feedback loop in which lncRNAs interact with sRNAs.

We constructed a lncRNA-sRNA-mRNA network and identified a pathogenesis-related subnetwork based on PHI-base (Figure 5) (Urban et al., 2020). In this ceRNA network, lncRNAs could indirectly regulate the expression of genes by competing for sRNA targets (Salmena et al., 2011). Genes in this pathogenesis-related network showed enrichment in binding-related terms, including nucleic acid binding. Three transcription factor genes known to be involved in pathogenicity were also included in this network. This network may provide pathogenesis-related candidate genes. For example, *MoMAS5* (MGG_02253) and the transcription factor *TRA1* (MGG_09523) were observed in the same sub-network. It has recently been shown that *MoMas5* is required for the suppression of host innate immunity (Gong et al., 2022).

In summary, we performed genome-wide lncRNA and sRNA profiling under Dicer deficiency conditions and comprehensively analyzed their crosstalks in *M. oryzae*. Dicer-dependent AGO-enriched sRNAs and their associated lncRNAs were identified. Functional enrichment analysis and construction of the lncRNA-sRNA-mRNA network showed that different biological pathways were involved in different types of crosstalks between lncRNAs and sRNAs. This study provides a foundation for investigating the roles of sRNAs and lncRNAs in rice blast disease development.

LITERATURE CITED

Atkinson, S.R., Marguerat, S., Bitton, D.A., Rodríguez-López, M., Rallis, C., Lemay, J.-F., Cotobal, C., Malecki, M., Smialowski, P., and Mata, J. 2018. Long noncoding RNA repertoire and targeting by nuclear exosome, cytoplasmic exonuclease, and RNAi in fission yeast. RNA. 24:1195-1213.

Choi, G., Jeon, J., Lee, H., Zhou, S., and Lee, Y.-H. 2022. Genome-wide profiling of long non-coding RNA of the rice blast fungus *Magnaporthe oryzae* during infection. BMC Genomics. 23:132.

Dai, X., Zhuang, Z., and Zhao, P.X. 2018. psRNATarget: a plant small RNA target analysis server (2017 release). Nucleic Acids Res. 46:W49-W54.

Dean, R., Van Kan, J.A., Pretorius, Z.A., Hammond-Kosack, K.E., Di Pietro, A., Spanu, P.D., Rudd, J.J., Dickman, M., Kahmann, R., and Ellis, J. 2012. The Top 10 fungal pathogens in Mol. Plant Pathol. Mol. Plant Pathol. 13:414-430.

Dean, R.A., Talbot, N.J., Ebbole, D.J., Farman, M.L., Mitchell, T.K., Orbach, M.J., Thon, M., Kulkarni, R., Xu, J.-R., and Pan, H. 2005. The genome sequence of the rice blast fungus *Magnaporthe grisea*. Nature. 434:980-986.

Fahlgren, N., and Carrington, J.C. 2010. miRNA target prediction in plants. Plant MicroRNAs. Springer. 51-57.

Gong, Z., Ning, N., Li, Z., Xie, X., Wilson, R.A., and Liu, W. 2022. Two *Magnaporthe* appressoria-specific (MAS) proteins, MoMas3 and MoMas5, are required for suppressing host innate immunity and promoting biotrophic growth in rice cells. Mol. Plant Pathol. 23:1290-1302.

Hannon, G.J. 2002. RNA interference. Nature. 418:244-251.

Hansen, T.B., Wiklund, E.D., Bramsen, J.B., Villadsen, S.B., Statham, A.L., Clark, S.J., and Kjems, J. 2011. miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. EMBO J. 30:4414-4422.

Huh, A., Dubey, A., Kim, S., Jeon, J., and Lee, Y.-H. 2017. *MoJMJ1*, encoding a histone demethylase containing JmjC domain, is required for pathogenic development of the rice blast fungus, *Magnaporthe oryzae*. Plant Pathol. J. 33:193.

Johnson, N.R., Yeoh, J.M., Coruh, C., and Axtell, M.J. 2016. Improved placement of multi-mapping small RNAs. G3 (Bethesda). 6:2103-2111.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., and Nuka, G. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics. 30:1236-1240.

Joshi, N., and Fass, J. 2011. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files. https://github.com/najoshi/sickle.

Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E.P., Rivas, E., Eddy, S.R., Bateman, A., Finn, R.D., and Petrov, A.I. 2018. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. Nucleic Acids Res. 46:D335-D342.

Kim, D., Langmead, B., and Salzberg, S.L. 2015. HISAT: a fast spliced aligner with low memory requirements. Nat. Methods. 12:357-360.

Kim, W., Miguel-Rojas, C., Wang, J., Townsend, J.P., and Trail, F. 2018. Developmental dynamics of long noncoding RNA expression during sexual fruiting body formation in *Fusarium graminearum*. mBio. 9:e01292-01218.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. 2009. Ultrafast and

memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10:1-10.

Laurent, G.S., Wahlestedt, C., and Kapranov, P. 2015. The Landscape of long noncoding RNA classification. Trends Genet. 31:239-251.

Lee, H., Choi, G., Lim, Y.-J., and Lee, Y.-H. 2022. Comparative profiling of canonical and non-canonical small RNAs in the rice blast fungus, *Magnaporthe oryzae*. Front. Microbiol. 13:995334.

Li, J., Liu, X., Yin, Z., Hu, Z., and Zhang, K.-Q. 2021a. An overview on identification and regulatory mechanisms of long non-coding RNAs in fungi. Front. Microbiol. 12:638617.

Li, Y., Liu, X., Yin, Z., You, Y., Zou, Y., Liu, M., He, Y., Zhang, H., Zheng, X., and Zhang, Z. 2020. MicroRNA-like milR236, regulated by transcription factor MoMsn2, targets histone acetyltransferase MoHat1 to play a role in appressorium formation and virulence of the rice blast fungus *Magnaporthe oryzae*. Fungal Genet. Biol. 137:103349.

Li, Z., Yang, J., Peng, J., Cheng, Z., Liu, X., Zhang, Z., Bhadauria, V., Zhao, W., and Peng, Y.-L. 2021b. Transcriptional Landscapes of Long Non-coding RNAs and Alternative Splicing in *Pyricularia oryzae* Revealed by RNA-Seq. Front. Plant Sci. 12:723636.

Lowe, R., Shirley, N., Bleackley, M., Dolan, S., and Shafee, T. 2017. Transcriptomics technologies. PLoS Comput. Biol. 13:e1005457.

Ma, L., Bajic, V.B., and Zhang, Z. 2013. On the classification of long non-coding RNAs. RNA Biol. 10:924-933.

Ma, X., Shao, C., Jin, Y., Wang, H., and Meng, Y. 2014. Long non-coding RNAs: a

novel endogenous source for the generation of Dicer-like 1-dependent small RNAs in *Arabidopsis thaliana*. RNA Biol. 11:373-390.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17:10-12.

Mercer, T.R., Dinger, M.E., and Mattick, J.S. 2009. Long non-coding RNAs: insights into functions. Nat. Rev. Genet. 10:155-159.

Mohorianu, I., Schwach, F., Jing, R., Lopez-Gomollon, S., Moxon, S., Szittya, G., Sorefan, K., Moulton, V., and Dalmay, T. 2011. Profiling of short RNAs during fleshy fruit development reveals stage-specific sRNAome expression patterns. Plant J. 67:232-246.

Nalley, L., Tsiboe, F., Durand-Morat, A., Shew, A., and Thoma, G. 2016. Economic and environmental impact of rice blast pathogen (*Magnaporthe oryzae*) alleviation in the United States. PLoS ONE. 11:e0167295.

Nawrocki, E.P., and Eddy, S.R. 2013. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 29:2933-2935.

Nguyen, Q., Iritani, A., Ohkita, S., Vu, B.V., Yokoya, K., Matsubara, A., Ikeda, K.-i., Suzuki, N., and Nakayashiki, H. 2018. A fungal Argonaute interferes with RNA interference. Nucleic Acids Res. 46:2495-2508.

Nojima, T., and Proudfoot, N.J. 2022. Mechanisms of lncRNA biogenesis as revealed by nascent transcriptomics. Nat. Rev. Mol. Cell Biol. 23:389-406.

Nunes, C.C., Gowda, M., Sailsbery, J., Xue, M., Chen, F., Brown, D.E., Oh, Y., Mitchell, T.K., and Dean, R.A. 2011. Diverse and tissue-enriched small RNAs in the plant pathogenic fungus, *Magnaporthe oryzae*. BMC Genomics. 12:288.

Patel, R.K., and Jain, M. 2012. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLoS ONE. 7:e30619.

Pertea, G., and Pertea, M. 2020. GFF utilities: GffRead and GffCompare. F1000Res. 9:304.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. 33:290-295.

Ponting, C.P., Oliver, P.L., and Reik, W. 2009. Evolution and functions of long noncoding RNAs. Cell. 136:629-641.

Quinn, J.J., and Chang, H.Y. 2016. Unique features of long non-coding RNA biogenesis and function. Nat. Rev. Genet. 17:47-62.

Raman, V., Simon, S.A., Demirci, F., Nakano, M., Meyers, B.C., and Donofrio, N.M. 2017. Small RNA functions are required for growth and development of *Magnaporthe oryzae*. Mol. Plant Microbe Interact. 30:517-530.

Raman, V., Simon, S.A., Romag, A., Demirci, F., Mathioni, S.M., Zhai, J., Meyers, B.C., and Donofrio, N.M. 2013. Physiological stressors and invasive plant infections alter the small RNA transcriptome of the rice blast fungus, *Magnaporthe oryzae*. BMC Genomics. 14:326.

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., and Vilo, J. 2019. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. 47:W191-W198.

Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P.P. 2011. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell. 146:353-358.

Skamnioti, P., and Gurr, S.J. 2009. Against the grain: safeguarding rice from rice

blast disease. Trends Biotechnol. 27:141-150.

Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.-L., and Ideker, T. 2011. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics. 27:431-432.

Szachnowski, U., Andjus, S., Foretek, D., Morillon, A., and Wery, M. 2019. Endogenous RNAi pathway evolutionarily shapes the destiny of the antisense lncRNAs transcriptome. Life Sci. Alliance. 2: e201900407.

Tay, Y., Kats, L., Salmena, L., Weiss, D., Tan, S.M., Ala, U., Karreth, F., Poliseno, L., Provero, P., and Di Cunto, F. 2011. Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. Cell. 147:344-357.

Torres-Martínez, S., and Ruiz-Vázquez, R.M. 2017. The RNAi universe in fungi: a varied landscape of small RNAs and biological functions. Annu. Rev. Microbiol. 71:371-391.

Urban, M., Cuzick, A., Seager, J., Wood, V., Rutherford, K., Venkatesh, S.Y., De Silva, N., Martinez, M.C., Pedro, H., and Yates, A.D. 2020. PHI-base: the pathogen–host interactions database. Nucleic Acids Res. 48:D613-D620.

Wang, L., Park, H.J., Dasari, S., Wang, S., Kocher, J.-P., and Li, W. 2013. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res. 41:e74.

Wilson, R.C., and Doudna, J.A. 2013. Molecular mechanisms of RNA interference. Annu. Rev. Biophys. 42:217-239.

Zhao, L., Wang, J., Li, Y., Song, T., Wu, Y., Fang, S., Bu, D., Li, H., Sun, L., and Pei, D. 2021. NONCODEV6: an updated database dedicated to long non-

coding RNA annotation in both animals and plants. Nucleic Acids Res. 49:D165-D171.

Zheng, G.X., Do, B.T., Webster, D.E., Khavari, P.A., and Chang, H.Y. 2014. Dicer-microRNA-Myc circuit promotes transcription of hundreds of long noncoding RNAs. Nat. Struct. Mol. Biol. 21:585-590.

벼 도열병균의 긴 비암호화 리보핵산 분석 및 짧은 비암호화 리보핵산과의 상호작용

최 고 봉

초 록

단백질을 암호화하는 구역 및 암호화하는 서열이 없는 구역에서도 전사는 일어난다. 비암호화 리보핵산은 단백질을 만드는 정보가 없지만 유전자를 조절함으로써 전사 과정, 전사 후 과정, 번역 과정, 번역 후 과정에서 일어나는 조절 과정에 관여한다. 비암호화 리보핵산은 200개의 염기보다 긴 경우 긴 비암호화 리보핵산(lncRNA)으로 간주된다. 시퀀싱(sequencing) 분석 기술이 발전하면서 비암호화 리보핵산 전사체가 축적되고 기능 분석이 수행되고 있다. 긴 비암호화 리보핵산은 발달 과정, 비생물적 자극에 대한 반응, 기주와 미생물의 상호작용에 참여한다고 보고되었다. 그러나 제한된 종에서의 연구로 인해 식물병원성 곰팡이에서는 긴 비암호화 리보핵산에 대한 역할에 대한 이해가 부족하다.

해당 연구는 기주에 대한 반응에서 긴 비암호화 리보핵산의 역할을 이해하기 위해 병이 발생하는 동안 벼 도열병균(*Magnaporthe oryzae*)에서 프로파일링(profiling)을 수행했다. 긴 비암호화 리보핵산을 확인 후

기능과 관련이 있을 수 있는 유전체 서열 특징과 발현 경향을 분석했다. 추가적으로, 기능을 할 가능성이 큰 긴 비암호화 리보핵산을 조사하기 위해서 감염 단계에 특이적으로 발현한 경우의 대상 유전자를 탐색했다. 유전자 분석 결과는 긴 비암호화 리보핵산은 세포벽 분해와 기주의 면역 체계 회피 같은 역할을 수행하여 병원성에 관여한다고 제시해 준다.

긴 비암호화 리보핵산은 단독으로 또는 짧은 비암호화 리보핵산(sRNA)와 협력해서 기능한다. 상호작용 방식은 일반적으로 세 가지가 있다. 전자가 후자의 전구체가 되는 경우, 후자가 전자를 조절하는 경우, 전자가 후자의 활동을 조절하는 경우로 구분할 수 있다. 곰팡이에서는 이들의 상호작용에 대한 이해가 부족한 상황이다. 벼 도열병균에서 상호작용을 밝히기 위해 짧은 비암호화 리보핵산의 생합성 유전자가 없는 상황에서 두 비암호화 리보핵산의 프로파일링을 수행했다. 그 과정에서 짧은 비암호화 리보핵산 중 잔해를 배제하기 위해서 리보핵산 간섭 도구에 의해 처리되는 것들을 선별했다. 대상 유전자의 분석 결과 상호작용의 종류에 따라 다른 생물학적 과정과 연관되어 있음을 밝혔다.

해당 연구는 비암호화 리보핵산의 레퍼토리를 제공하여 생물학적 기능을 알아보기 위한 기능적 연구의 기반을 제공한다. 또한 종합적인 연구를 통해 두 종류의 비암호화 리보핵산의 상호작용에 대한 이해를 돕고, 병원성을 포함하는 생물학적 과정에서 이들이 핵심 요소라는 점을 제안한다. 따라서 본 연구는 식물 병원성 곰팡이에서 복잡한 조절망에 대한 연구 방향을 제시한다.

주요어 : 벼 도열병균, 기주 감염, 긴 비암호화 리보핵산, 짧은 비암호화 리보핵산, 벼 도열병

학번 : 2015-21805