Sequencing and Functional Annotation of mRNAs and lncRNAs from Tea (*Camellia sinensis*) Leaves During Infection by the Fungal Pathogen *Lasiodiplodia theobromae*

Dongxue Li,¹ Silong Jiang,^{1,2} Xiaodong Wen,^{1,2} Xingchen Song,^{1,2} Yuanyou Yang,¹ Yu Wang,¹ Honglin Huang,¹ Delu Wang,^{3,†} and Zhuo Chen^{1,†}

¹ Key Laboratory of Green Pesticide and Agricultural Bioengineering, Ministry of Education, Guizhou University, Guiyang, Guizhou 550025, China

² Agricultural College, Guizhou University, Guiyang, Guizhou 550025, China

³ College of Forestry, Guizhou University, Guiyang, Guizhou 550025, China

Abstract

Due to the absence of effective control measures, tea leaf spot, caused by *Lasiodiplodia theobromae*, can seriously reduce the production and quality of tea leaves. Analysis of the relationship between messenger RNAs (mRNAs) and long noncoding RNAs (lncRNAs) of tea could help to prevent and control the disease. In this current study, high-performance sequencing of mRNA and lncRNA from tea leaves during infection by *L. theobromae* was conducted using the Illumina HiSeq 4000 platform. Of the differentially expressed mRNAs, 2,503 or 5,278 were significantly up- or down regulated, respectively, in challenged leaves relative to unchallenged leaves. Differentially expressed lncRNAs were also analyzed at the level of genes for the above two treatments. The numbers of significantly up- or downregulated lncRNAs were 3,403 or 1,477, respectively. Functional analysis of mRNA revealed that *L. theobromae* challenge can induce the upregulation of the genes of *seven transmembrane MLO family protein -3*, *-6*, *-11*, *major allergen Pru ar 1*, and so on. In total, 176 differentially expressed lncRNAs occurred in response to *L. theobromae* challenge as a result of *cis* regulation, with 146 differentially expressed mRNAs. At the biological process of defense response, the *glycosyl hydrolase 9A1* gene was *cis* regulated by the differentially expressed lncRNA. The expression profiling and association studies of the tea host infected by *L. theobromae* will provide a resource for future research into host–pathogen interactions and disease resistance.

The genus-rich fungal family Botryosphaeriaceae (phylum Ascomycota) consists of saprobes, endophytes, and phytopathogens, with largely woody plant hosts (Chethana et al. 2016; Morales-Cruz et al. 2015; Paolinelli-Alfonso et al. 2016; Phillips et al. 2013; Yan et al. 2013). *Lasiodiplodia* Ellis & Everh. is a clade of 25 species in this family and is highly pathogenic toward many plant species (Phillips et al. 2013; Yan et al. 2013). In this genus, *Lasiodiplodia theobromae* (Pat.) Griffon & Maubl. is the anamorph of *Botryosphaeria rhodina* (Berk. & M. A. Curtis) Arx, which can cause diseases in many unrelated plant species such as canker and dieback of grapevine (Rodríguez-Gálvez et al. 2015), branch canker of olive (Pérez et al. 2018), branch dieback of cashew (Cardoso et al. 2002), fruit spot of eggplant (Woodward et al. 2005), fruit rot of pomelo (Luo et al. 2011), postharvest fruit rot in avocado (Garibaldi et al. 2012), gummosis or inflorescence blight of mango trees (Li et al. 2013; Serrato-Díaz et al. 2013), inflorescence blight and fruit rot of longan (Serrato-Díaz et al. 2014), leaf spot of tea oil (*Camellia oleifera*) (Zhu et al. 2014), collar rot of peanut (Phipps and Porter 1998), strawberry dieback (Yildiz et al. 2014), and leaf spot of tea (*C. sinensis*) (Li et al. 2019). In a previous study, messenger RNAs (mRNAs) and

Funding

The project was supported by National Key Research Development Program of China (2017YFD0200308) and its post-subsidy project (2018-5262), National Natural Science Foundation of China (numbers 31860515 and 21977023), Program of Introducing Talents to Chinese Universities (111 Program, D20023), and China Agriculture Research System of MOF and MARA (CARS-23-D09).

Keywords

Lasiodiplodia theobromae, lncRNAs, mRNAs, sequencing and functional analysis, tea leaf spot

[†]Corresponding authors: D. L. Wang; dlwang@gzu.edu.cn and Z. Chen; gychenzhuo@aliyun.com

e-Xtra: Supplementary materials are available online.

The author(s) declare no conflict of interest.

Accepted for publication 24 May 2021.

Copyright © 2021 The Author(s). This is an open access article distributed under the CC BY-NC-ND 4.0 International license.

microRNAs (miRNAs) of *L. theobromae* and the tea plant host during infection of tea leaves were sequenced, and differentially expressed genes of the pathogen and plant host were analyzed. The target genes of miRNAs of the pathogen and plant host were also predicted (Jiang et al. in press).

Long noncoding RNAs (lncRNAs) exhibit various biological functions by regulating gene expression at the transcriptional, translational, or posttranslational levels (Chi et al. 2019). As a consequence, the analysis of lncRNAs is of marked significance in terms of understanding disease and resistance during the process of infection (Zaynab et al. 2018). In this study, 3-year-old potted tea plants of *C. sinensis* 'Fuding-dabaicha' were grown in the greenhouse and used as the host in the pathogenicity test. During the one-bud and fifth-leaf stages, the third or fourth leaves on each healthy twig were inoculated with the conidial suspensions of *L. theobromae* strain GZHS-2017-010 (10^6 conidia/ml, 20 to 30 μ l per inoculation aliquot) onto physically wounded tea leaves. Both sides of the main vein on tea leaves were punctured using sterilized syringe needles, making four neighboring holes within a range of 6 mm in diameter. For the uninfected treatment, sterile water was used to inoculate the tea leaves. For the infected and uninfected treatments, 15 healthy twigs were pooled to one treatment group. The experiment was independently conducted in three biological treatment groups. At 48 h postinoculation, a lesion formed on the tea leaf. The total RNA of the tea leaf was extracted at 48 h postinoculation, using a TRIzol Plus RNA Purification Kit (Invitrogen, Carlsbad, CA, U.S.A.), following the manufacturer's procedure. A cDNA library was prepared using a TruSeq stranded mRNA Sample Preparation kit (Illumina, San Diego, CA, U.S.A.), and sequenced on an Illumina HiSeq 4000 platform by Lc-Bio Technologies Co., Ltd. (Hangzhou, China), following the manufacturer's recommended protocol. Low-quality reads were removed using the Cutadapt program (command line parameters: -O 3, -m 100) (Martin 2011). For mRNA analysis, valid data were aligned to the reference genome of *C. sinensis* var. *sinensis* (<http://pcsb.ahau.edu.cn:8080/CSS/>), using HISAT software (version 2.0.4) (command line parameters: -l fr-firststrand, -min-intronlen 20, -max-intronlen 500000) (<http://www.ccb.jhu.edu/software/hisat/index.shtml>) (Kim et al. 2015). The mapped reads of each sample were assembled; then, the quantified expression abundance for mRNAs was calculated as the fragment per kilobase of transcript per million mapped reads value at the level of $\log_2$ (fold change) > 1 or $\log_2$ (fold change) $\leq$ 1 and with statistical significance of *P* value < 0.05 using StringTie software (version:stringtie-1.3.4d; default parameters) (<https://ccb.jhu.edu/software/stringtie/>) (Frazee et al. 2015; Pertea et al. 2015) and Ballgown (default parameters) in R (Frazee et al. 2015). Functional analysis of the differentially expressed mRNAs was conducted by Gene Ontology (GO) (version 2016.04) (<http://geneontology.org>) and Kyoto Encyclopedia of Genes and Genomes (version 2016.05) (<http://www.kegg.jp/kegg>) using the Blast2GO bioinformatics tool (<http://www.ncbi.nlm.nih.gov/blast>) (Conesa et al. 2005).

For lncRNA analysis, sequence quality was further verified using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Valid reads were used to map reads to the genome of *C. sinensis* var. *sinensis* (<http://pcsb.ahau.edu.cn:8080/CSS/>) using Bowtie2 (default parameters) in R (Langmead and Salzberg 2012) and TopHat2 (command line parameters: -mate-inner-dist 50, -mate-std-dev 20, -min-anchor 4, -min-intron-length 20, -max-intron-length 500000, -read-edit-dist 2, -min-segment-intron 70, -max-segment-intron 40000, -read-mismatches 2, -library-type fr-firststrand) (Kim et al. 2013). The mapped reads of each sample were assembled using StringTie software (version stringtie-1.3.4d; default parameters) (Pertea et al. 2015); then, the transcriptomes from the six samples involving infected and uninfected treatments (three biological replicates of each) were merged to construct a comprehensive transcriptome using the Perl program (<https://www.perl.org>). After the final transcriptome was generated, the abundance of all transcripts was estimated using StringTie software (default parameters) (Pertea et al. 2015) and Ballgown (default parameters) (Frazee et al. 2015). All transcripts were filtered as follows. First, transcripts that overlapped with known mRNAs and novel mRNAs with coding potential were filtered; second, transcripts that were shorter than 200 bp and the coverage of reads < 3 were filtered. Transcripts with coding potential were used for prediction using the Coding Potential Calculator (CPC) software (version 0.9-r2) (command line parameters: -cpc_score 0.5) (Kong et al. 2007) and Coding-Non-Coding Index (CNCI) software (version 2.0) (command line parameters: -novel_trans_length 200, -novel_trans_num_exon 1, -cnci_score 0) (Sun et al. 2013). All transcripts with CPC score $\leq$ 1 and CNCI score < 0 were removed. The

Table 1. Messenger RNA (mRNA) and long noncoding RNA (lncRNA) sequences from tea (*Camellia sinensis*) leaves during infection by the fungal pathogen *Lasiodiplodia theobromae*

Sequence	Sample source ^a	
	Uninfected treatment	Infected treatment
mRNAs		
Raw reads	46,851,992	49,801,319
Valid reads	46,241,852	48,419,248
Valid bases	6.82 G	7.12 G
Infected versus uninfected treatment	–	2,503 (upregulation)
	–	5,278 (downregulation)
lncRNAs		
Raw reads	141,647,262	145,772,279
Valid reads	139,330,285	143,076,549
Valid bases	20.90 G	21.46 G
Differentially expressed lncRNAs versus mRNAs ^b	–	176 lncRNAs
	–	146 mRNAs

^a Tea leaves were punctured with four holes using sterilized syringe needles, with the holes clustered within a range of 6 mm in diameter; then, sterile water (uninfected treatment) or the conidial suspensions (infected treatment) were used to inoculate the tea leaves.

^b *cis* Regulation between differentially expressed lncRNAs and differentially expressed mRNAs.

remaining transcripts were considered to be lncRNAs. The expression level of lncRNAs was calculated using StringTie software (default parameters) (<https://ccb.jhu.edu/software/stringtie/>) (Pertea et al. 2015; Trapnell et al. 2010) and Ballgown (default parameters) in R (Frazee et al. 2015). *Cis*-target genes of lncRNAs in 100,000 upstream and downstream were predicted by the Python program (<https://www.python.org>). Then, functional analysis of the target genes for lncRNAs was conducted by using Blast2GO (Conesa et al. 2005). Significance was expressed as a *P* value < 0.05.

Infection by *L. theobromae* significantly up- or downregulated 2,503 or 5,278 of the tea leaf mRNAs, respectively. When the GO database annotation of all genes was carried out, the total number of genes (TB gene numbers) and the number of genes with significantly differential expression between the infected treatments and the uninfected treatments (TS gene numbers) were 21,398 and 4,255, respectively. At the level of biological process (BP), there was enrichment with particular respect to the number of genes related to defense response, with the number of genes with significantly differential expression annotated in the specified database (S gene number) and number of genes annotated in the specified GO term (B gene number) being 18 and 63, respectively. Some genes of *seven transmembrane MLO family protein -3*, *-6*, *-11*, and *major allergen Pru ar 1* were upregulated in the infected treatment relative to the uninfected treatment.

The differentially expressed lncRNAs were analyzed at the levels of genes. Infection by *L. theobromae* significantly up- and downregulated 3,403 and 1,477 of genes, respectively. As a result of *cis* regulation, 176 differentially expressed lncRNAs occurred, with 146 differentially expressed mRNAs (Table 1).

GO enrichment analysis of differentially expressed mRNAs targeted by lncRNAs indicated that TB gene numbers and TS gene numbers were 1,454 and 220, respectively. At the level of BP, differentially expressed mRNAs were mainly enriched with respect to defense response, with S gene numbers and B gene numbers being 3 and 7, respectively. GO analysis indicated that the *glycosyl hydrolase 9A1* gene enriched at defense response was *cis* regulated by the differentially expressed lncRNA.

For tea leaf spot caused by *L. theobromae*, mRNAs and lncRNAs from infected and uninfected tea leaves were sequenced. The differential expression of mRNAs or lncRNAs was analyzed; then the differentially expressed lncRNAs were identified and their corresponding *cis* target genes were predicted. This dataset will provide an important resource for studying the mechanism underlying the interaction between *L. theobromae* and the tea plant. The sequences of mRNAs and lncRNAs of tea have been deposited in GenBank under Sequence Read Archive accession PRJNA637492 for mRNAs and PRJNA689677 for lncRNAs. In addition, bioinformatics analysis of mRNAs and lncRNAs are presented in in Supplementary Table S1. *L. theobromae* strain GZHS-2017-010 has been deposited in China General Microbiological Culture Collection Center, with the accession number CGMCC3.20151.

Literature Cited

- Cardoso, J. E., Vidal, J. C., Santos, A. A., Freire, F. C. O., and Viana, F. M. P. 2002. First report of black branch dieback of cashew caused by *Lasiodiplodia theobromae* in Brazil. *Plant Dis.* 86:558.
- Chethana, K. W. T., Li, X.-H., Zhang, W., Hyde, K. D., and Yan, J.-Y. 2016. Trail decryption of molecular research on *Botryosphaeriaceae* in woody plants. *Phytopathol. Mediterr.* 55:147-171.
- Chi, Y.-J., Wang, D., Wang, J.-P., Yu, W.-D., and Yang, J.-C. 2019. Long non-coding RNA in the pathogenesis of cancers. *Cells* 8:1015.
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. 2005. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674-3676.
- Frazee, A. C., Perte, G., Jaffe, A. E., Langmead, B., Salzberg, S. L., and Leek, J. T. 2015. Ballgown bridges the gap between transcriptome assembly and expression analysis. *Nat. Biotechnol.* 33:243-246.
- Garibaldi, A., Bertetti, D., Amati, M. T., Cardinale, J., and Gullino, M. L. 2012. First report of postharvest fruit rot in avocado (*Persea americana*) caused by *Lasiodiplodia theobromae* in Italy. *Plant Dis.* 96:460.
- Jiang, S.-L., Yin, Q.-X., Li, D.-X., Wu, X., Wang, Y., Wang, D.-L., and Chen, Z. Integrated mRNA and small RNA sequencing for analyzing tea leaf spot pathogen *Lasiodiplodia theobromae*, under in vitro conditions and the course of infection. *Phytopathology* In press.
- Kim, D., Langmead, B., and Salzberg, S. L. 2015. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* 12:357-360.
- Kim, D., Perte, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. 2013. TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14:R36.
- Kong, L., Zhang, Y., Ye, Z.-Q., Liu, X.-Q., Zhao, S.-Q., Wei, L., and Gao, G. 2007. CPC: Assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* 35:W345-W349.
- Langmead, B., and Salzberg, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9:357-359.
- Li, D.-X., Bao, X.-T., Ren, Y.-F., Wang, Y., Song, B.-A., and Chen, Z. 2019. First report of *Lasiodiplodia theobromae* causing leaf spot on tea plant in Guizhou Province of China. *Plant Dis.* 103:374.
- Li, Q.-L., Guo, T.-X., Pan, Z.-B., Huang, S.-P., Mo, J.-Y., Ning, P., and Hsiang, T. 2013. An outbreak of gummosis of mango trees caused by *Lasiodiplodia theobromae* in Guangxi, South China. *Plant Dis.* 97:690.
- Luo, M., Dong, Z.-Y., Bin, S.-Y., and Lin, J.-T. 2011. First report of fruit rot disease on pomelo caused by *Lasiodiplodia theobromae* in China. *Plant Dis.* 95:1190.
- Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* 17:10-12.
- Morales-Cruz, A., Amrine, K. C. H., Blanco-Ulate, B., Lawrence, D. P., Travadon, R., Rolshausen, P. E., Baumgartner, K., and Cantu, D. 2015. Distinctive expansion of gene families associated with plant cell wall degradation, secondary metabolism, and nutrient uptake in the genomes of grapevine trunk pathogens. *BMC Genomics* 16:469.
- Paolinelli-Alfonso, M., Villalobos-Escobedo, J. M., Rolshausen, P., Herrera-Estrella, A., Galindo, C., and López-Hernández, J. F. 2016. Global transcriptional analysis suggests *Lasiodiplodia theobromae* pathogenicity factors involved in modulation of grapevine defensive response. *BMC Genomics* 17:615.
- Pérez, B. A., Niz, J., Salvador, R., Salvador, R., Berretta, M., Roca, M. E. M., Pesce, M. V., Matías, A. C., González De Urreta, M., Paccioretti, M., and Otero, M. L. 2018. First report of *Lasiodiplodia theobromae* causing branch canker on 'Manzanilla' olive in Northwestern Argentina. *Plant Dis.* 102:677.
- Perte, M., Perte, 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.
- Phillips, A. J. L., Alves, A., Abdollahzadeh, J., Slippers, B., Wingfield, M. J., Groenewald, J. Z., and Crous, P. W. 2013. The *Botryosphaeriaceae*: Genera and species known from culture. *Stud. Mycol.* 76:51-167.
- Phipps, P. M., and Porter, D. M. 1998. Collar rot of peanut caused by *Lasiodiplodia theobromae*. *Plant Dis.* 82:1205-1209.
- Rodríguez-Gálvez, E., Maldonado, E., and Alves, A. 2015. Identification and pathogenicity of *Lasiodiplodia theobromae* causing dieback of table grapes in Peru. *Eur. J. Plant Pathol.* 141:477-489.
- Serrato-Díaz, L. M., Perez-Cuevas, M., Rivera-Vargas, L. I., and French-Monar, R. D. 2013. First report of *Lasiodiplodia theobromae* causing inflorescence blight of mango. *Plant Dis.* 97:1380.
- Serrato-Díaz, L. M., Rivera-Vargas, L. I., Goenaga, R., and French-Monar, R. D. 2014. First report of *Lasiodiplodia theobromae* causing inflorescence blight and fruit rot of longan (*Dimocarpus longan* L.) in Puerto Rico. *Plant Dis.* 98:279.
- Sun, L., Luo, H.-T., Bu, D.-C., Zhao, G.-G., Yu, K.-T., Zhang, C.-H., Liu, Y.-N., Chen, R.-S., and Zhao, Y. 2013. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* 41:e166.
- Trapnell, C., Williams, B. A., Perte, G., Mortazavi, A., Kwan, G., Van Baren, M. J., Salzberg, S. L., Wold, B. J., and Pachter, L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28:511-515.
- Woodward, J. E., Langston, D. B., Brock, J. H., Kemerait, C. J., Brenneman, T. B., and Beard, G. H. 2005. First demonstration of Koch's postulates for *Lasiodiplodia theobromae* fruit spot on eggplant (*Solanum melongena*). *Plant Dis.* 89:687.
- Yan, J.-Y., Xie, Y., Zhang, W., Wang, Y., Liu, J.-K., Hyde, K. D., Seem, R. C., Zhang, G.-Z., Wang, Z.-Y., Yao, S.-W., Bai, X.-J., Dissanayake, A. J., Peng, Y.-L., and Li, X.-H. 2013. Species of *Botryosphaeriaceae* involved in grapevine dieback in China. *Fungal Divers.* 61:221-236.
- Yildiz, A., Benlioglu, K., and Benlioglu, H. S. 2014. First report of strawberry dieback caused by *Lasiodiplodia theobromae*. *Plant Dis.* 98:1579.
- 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.
- Zhu, H., Niu, X.-Q., Song, W.-W., Yu, F.-Y., Tang, Q.-H., Qin, W.-Q., and Chen, L.-Q. 2014. First report of leaf spot of tea oil camellia (*Camellia oleifera*) caused by *Lasiodiplodia theobromae* in China. *Plant Dis.* 98:1427.