Indonesian Journal of Biotechnology

VOLUME 28(1), 2023, 56-65 | RESEARCH ARTICLE

# Revealing disease-specific endogenous target mimic of microRNA from long non-coding RNA identification and characterization in *Musa* spp.

**Audie Masola Putra[1], Husna Nugrahapraja[1,3,*], Fenny Martha Dwivany[1,2,3]**

[1]School of Life Sciences and Technology, Institut Teknologi Bandung, Jalan Ganesha No. 10, Bandung, West Java 40132, Indonesia
[2]Bali International Research Center for Banana, Badung, Bali, 80361, Indonesia
[3]Bioscience and Biotechnology Research Center, Institut Teknologi Bandung, Jalan Ganesha No. 10, Bandung, West Java 40132, Indonesia
*Corresponding author: nugrahapraja@sith.itb.ac.id

**SUBMITTED** 6 September 2019 **REVISED** 20 April 2020 **ACCEPTED** 13 December 2022

**ABSTRACT** Banana (*Musa* spp.) is one of the most widely consumed fruits in the world. Unfortunately, the plants are at risk from many disease problems, which mainly derive from microorganism. It is a little known about the relationship between disease-inducing microorganisms and plants, particularly at the molecular level. This research aimed to characterize long non-coding RNA (lncRNA) from bananas that may have roles in regulating gene expression related to the disease response mechanism in banana derived from transcriptomic libraries. Furthermore, the detected transcripts were analyzed to identify the endogenous target mimics (eTMs) interaction between lncRNA and microRNA (miRNA) using computational approaches. Data from Cavendish banana (AAA group), Berangan (AAA group), Yunnan Banana (Itinerans), Dajiao (ABB group), and Klutuk (BB group) were used in this research. We found that lncRNA tends to be unsustainable, and most sizes are below 1000 bp (≥ 75%). Based on this result, we investigated the eTMs to determine lncRNA transcripts and miRNA, such as miR397 in Cavendish and miR444 in Klutuk. This transcript would be regulated following exposure to extreme temperatures and disease, indicating the possibility of disease-specific interaction between bananas and their environment at the molecular level.

**KEYWORDS** Banana; Characterization; Endogenous target mimics (eTMs); Long non-coding RNA; Transcriptomic

## 1. Introduction

Bananas (*Musa* spp.) are one of the widely consumed commodities in Indonesia. However, these plants are prone to disease caused by microorganisms attacks, such as viruses, bacteria, and fungi, resulting in a reduction of fruit production yield. In banana, blood disease bacteria and fusarium affect a massive drop in the fruit production. Several preventive methods have been applied, but the mechanisms of infection caused by the disease-causing agent were still not yet well understood at molecular level. Fortunately, study on transcriptomic in plant system is gradually becoming more extensive, particularly in understanding the role of non-coding RNA in gene regulation mechanisms in plants. It has been reported that the expression of lncRNAs have positive correlation between pathogen and plant cultivar in transcription expression level (Muthusamy et al. 2019). Therefore, we are interested to initiate our study to identify lncRNA from banana transcriptome data and determine their potential role in gene regulation mechanism.

In general, banana cultivated in different ploidy numbers, which is mostly triploids AAA, AAB, and ABB genome. This troploid bananas came from the progenitors of Musa acuminata (AA genome) and Musa balbisiana (BB genome). Currently, the cultivated bananas are also found in completely different groups, such as the S or T genome (Cheesman 1948). These complex genomic backgrounds in cultivated banana showed phenotypic variants ranged from susceptible to having a certain degree of resistance against abiotic and biotic stresses. For example, banana containing B genome is more tolerance to stress abiotic than A genome (Ravishankar et al. 2015)—Banana cv. Saba (ABB group) is more tolerant of the drought stress than other banana cultivars. While for banana cv. Grand Naine (AAA group) is susceptible compared to others (Muthusamy et al. 2016). Therefore, understanding the reason behind advantages of certain genomes thorough analysis using the omics approach would help to understand gene regulation networks and also the metabolic pathway. To date, there are extensive studies in banana in order to explore the molecular basis of abiotic and biotic stresses tolerant using omics approaches, particularly in functional genomics analysis (Ravishankar et al. 2015; Muthusamy et al. 2016). Furthermore, the study has also extended to the exploration of non-coding RNA, such as

miRNA and lncRNA, to have deeper understanding of molecular basis of the gene regulation mechanisms. The Long lncRNA have been found to have a role in gene regulation mechanisms when exposed to biotic and abiotic stresses (Yu et al. 2019). The miRNA studies have been explored in banana, while for lncRNA itself is still limited (Muthusamy et al. 2019; Sampangi-Ramaiah et al. 2019). In this study, the molecular basis underlying the banana-disease interaction will be conducted to discover lncRNA that could bring new perspectives of gene regulation mechanisms (Kim and Sung 2012).

lncRNAs are transcribed during the transcriptional process from DNA and part of the non-coding genome. lncRNA can be distinguished from the other type of non-coding RNAs due to its long base pair size (minimum length of 200 bp or above) and their lack of open reading frame (ORF). Also, lncRNA is known to have an essential function in living things, both animals and plants, since it has role in regulation of gene expression, both transcriptional and post-transcriptional regulation (Liu et al. 2015). lncRNAs act as regulator that can do cis-acting mechanism, such as chromatin modification, signals for gene transcription, interaction with effector molecules or repressors (Ballantyne et al. 2016).

Several studies on lncRNA (Kang and Liu 2015; Zhao et al. 2018; Yu et al. 2019) described some of the functions of lncRNA from the results of identification and characterization. In diploid strawberry (*Fragaria vesca*), regulatory interactions between lncRNAs and protein-coding (PC) gene targets showed positive and negative correlation on its expression, which can act in cis- and trans-actions (Kang and Liu 2015). Moreover, the natural antisense transcript-lncRNA (NAT-lncRNA) is required for the expression of cognate sense genes in *Arabidopsis* during stress and normal conditions (Zhao et al. 2018). On the other hand, there are extensive interactions between lncRNAs, MADS-box transcription factor RIPENING FACTOR (*RIN*), and miRNA (Yu et al. 2019). Borah et al. (2018) demonstrated the role of lncRNA as endogenous target mimics (eTMs) in soybean seeds, while other studies in maize found the lncRNA which binds explicitly to miR399 miRNA and represses the expression of the *PHO2* gene (Franco-Zorrilla et al. 2007). These target mimicries form a nonproductive interaction between lncRNA and miRNA to inhibit miRNA activity towards gene expression. Thus, association between miRNA and lncRNA might play an essential role in genes expression modulation, and deeper study of these interactions between two molecules representing the post-genomic era.

In this study, the lncRNA database is essential to elucidate the molecular basis of lncRNA interaction with miRNA. To accommodate the large quantity of data, several studies regarding lncRNA need to be stored in integrative molecular databases (Lesk 2005). The existence of this database is crucial for the availability of information, efficiency of time and cost of analysis since large genome data stored in database can be used as future references and can also be used as comparative analysis. These data can be used for applications in the field of technology, such as molecular plant breeding in agriculture system (Sampangi-Ramaiah et al. 2019). Moreover, the exploration of endogenous target mimic (eTMs) in banana also might open new perspectives of our understanding about gene regulation mechanism in bananas beyond identification and characterization. The aims of our research are to identify and characterize lncRNA coming from the banana transcriptomics datasets available; Berangan and Cavendish – AAA group, Yunnan – wild type, Dajiao – ABB group, and Klutuk – BB group. This study might provide a possible interaction of eTMs between lncRNA and miRNA on A and B genomes based on bioinformatics analysis. The description of lncRNA from different genomic background hopefully can give an insight regarding the molecular basis lncRNA as a respond to biotic stress to elucidate the disease-specific interaction in banana.

## 2. Materials and Methods

### *2.1. Transcriptomics datasets*

The transcriptomic datasets were obtained in the form of raw reads (SRA/FASTQ) from the RNA sequencing of five banana cultivars, namely Berangan – AAA group (SRR2132798; University of Malaya), Cavendish – AAA group (SRR924324; Chinese Academy of Tropical Agriculture Sciences), Dajiao – ABB group (SRR516083; Institute of Fruit Tree Research, Guangdong Academy of Agricultural Sciences), Klutuk – BB group (Backiyarani et al. 2015; Dwivany et al. 2021), and Yunnan Bananas – wild type banana (SRR6894782; Fujian Agriculture and Forestry University). The banana from each banana cultivar was obtained from the different experimental design and the source of data was from the various parts of organs and tissues combined in pooled transcript. We obtained all transcriptomic datasets from the NCBI database (Benson et al. 2017), EMBL-EBI (Leinonen et al. 2011), then we analyzed using in-house pipeline bioinformatics (Figure 1) in the Bioinformatics Laboratory, Biosciences and Biotechnology Research Center, Institut Teknologi Bandung.

### *2.2. Identification of lncRNA*

We started to identify lncRNA in all transcriptomics datasets using the Tuxedo pipeline (Trapnell et al. 2014), which include multiple software programs: Tophat2, Cufflinks, Cuffcompare, Cuffmerge, and Cuffdiff. First, we have checked the quality of the RNA-Seq libraries using FastQC, removed the adapter using Cutadapt version 1.2.1, and filtered low quality reads with ERNE-FILTER version 1.3 for quality control assessment (Figure 1).

The trimmed and clean reads from all libraries mapped against the genome reference DH Pahang version 2 (Genome A) for Cavendish, Dajiao, and Berangan (D'hont et al. 2012; Setiabudi et al. 2021) and the PKW Klutuk Wulung (Genome B) for Klutuk libraries (Davey et al. 2013), and Musa itinerans genome (Wu et al. 2016) for Yunnan

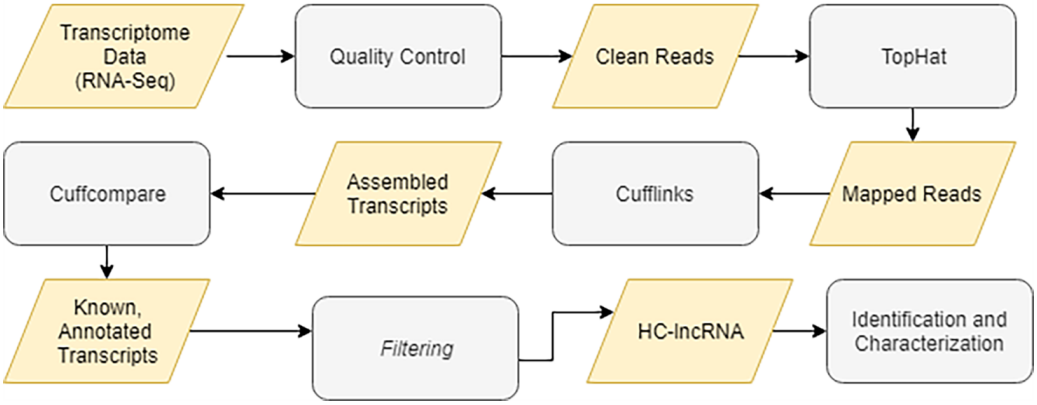


**FIGURE 1** Long non-coding (lncRNA) acquisition pipeline (Modified from (Wang et al. 2015)).

libraries using the TopHat2 version 2.0.9 program (Trapnell et al. 2014). We improved the mappability of the reads using the second iteration of mapping (Peri et al. 2020). The mapped reads then re-assembled independently using the Cufflinks version 2.0.9 program (Trapnell et al. 2014), producing output in the form of transcripts which compared with reference genome transcripts using the Cuffcompare program to identify known and annotated transcripts. The program will classify transcripts that have passed the Cufflinks stage. Then, we used Cuffmerge to obtain non-redundant transcripts, and the transcripts sequences retrieved using the gffreadtool. The compared transcript is filtered to get high confidence of lncRNA (Trapnell et al. 2014). Furthermore, the obtained lncRNA will be characterized using Cuffdiff for size, classification based on class code and possible interactions with miRNA.

The transcripts output from the Cuffcompare program (Trapnell et al. 2014) filtered with the parameters as following: 1) Size selection to eliminate the transcript size below 200 bp. 2) Eliminating the transcript with ORF and translated for more than 100 amino acids. 3) Eliminating the transcripts that have protein-coding regions based on a list in a protein database (Swiss Protein Database). 4) Transcript elimination using the Coding Potential Calculator (CPC) program to remove potentially coding transcripts. The CPC will analyze the quality, complexity, and sequence similarity with the ORF of the protein contained in the database, displayed in the form of a score. 5) Eliminating transcripts that have similarities with the sequence housekeeping genes. 6) Removing transcripts that have probabilities as small RNA (sRNA) precursors (Li et al. 2014). The output of the filtering results is that the remaining transcript is a high confidence level of lncRNA (HC-lncRNA), which is the result of the identification of lncRNA in the sample used. We obtained HC-lncRNA in the FASTA format.

### 2.3. *LncRNA characterisation based on class code*

Based on the position in the genome relative to the coding-gene, we classified the lncRNA transcript into four class codes: 1) intergenic lncRNA (class code: ”u”), located between two genes and not overlapping between the two genes. 2) intronic (Class code: ”i”) lncRNA, located in the intron of genes. 3) Sense lncRNA (class code: ”o”), located in the sense part of genes in DNA, and 4) antisense (class code: ”x”), located in the antisense part of genes in DNA (Ma et al. 2013).

### 2.4. *Prediction of Interaction of lncRNA with other molecules*

In this study, we sampled the obtained lncRNAs with the known banana miRNA. The lncRNA sequence is analyzed, then the psRNATarget program (https://plantgrn.noble.org/psRNATarget/analysis) (Dai et al. 2018) is used to predict the bond interaction between lncRNA and miRNA. The role in metabolic processes in banana plants was investigated by comparing them to miRNA databases such as miRBase and TAIR miRNA database (Dai et al. 2018).

### 2.5. *Validation of the existence of lncRNA*

We validated the lncRNA using Reverse Transcriptase PCR (RT-PCR) from total RNA in Cavendish – AAA group and Klutuk – BB group as a representative from A and B genomes and are available in the field collections. We determined the quality of RNA using a Biospectrophotometer (Eppendorf Inc.) and the integrity

**TABLE 1** LncRNA transcript primer design.

| Primer | Oligonucleotide Size (bp) | Fragment Size (bp) |
|---|---|---|
| MalncRNA_00002261-F | 5’-ACAAAG GAAAGGAGGGATGC-3’ | 125 |
| MalncRNA_00002261-R | 5’- ACAGTGTTGGCT-CATGAAGG -3’ | |
| MblncRNA_00057826-F | 5’- GGGAACAC-CATCGTTTCAGT -3’ | 100 |
| MblncRNA_00057826-R | 5’- CATCGTGTCATCCT-GCTCAA -3’ | |

using agarose gel electrophoresis (1.0%). We treated the RNA using DNase treatment as described in the DNase I RNase-free kit (Thermo Scientific, #EN0525) following manufacture protocol. RNA was reverse transcribed into cDNA using iScript cDNA Synthesis Kit (Biorad) following the manufacturer's instructions. We selected two lncRNA sequences from Cavendish and Klutuk. The primers used to conducted RT-PCR analysis (Table 1) were designed using Primer 3 Plus software (Rozen and Skaletsky 2000). The PCR amplification reaction (20 µl) contained GoTaq PCR Mastermix (Promega Inc.), five ρmole of the forward and five ρmole reverse primers, 2 µl of cDNA (1:5 diluted) and 0.5 units of Taq DNA polymerase (Promega Inc.). We performed the PCR reactions in a thermal cycler using the following conditions: the initial denaturation at 95 °C for 5 min, then followed by 40 cycles of 95 °C (30 s) / 59 °C (30 s)/72 °C (60 s) and a final extension step of 10 min at 72 °C, then the amplified PCR products were separated on an agarose gel (2%) for 25 min 100 V and visualized under UV-transilluminator. We used positive control (*Actin*), negative control (water), *MalncRNA_00002261*, and *MblncRNA_00057826*. *MalncRNA_00002261* refers to TCONS_00002261 lncRNA sequence from Cavendish while *MblncRNA_00057826* primers for TCONS_00057826 lncRNA sequence from Klutuk.

## 3. Results and Discussion

### 3.1. Identification of lncRNA

Based on the reads mapping result, Berangan, Cavendish – AAA group and Klutuk – BB group have a mapping rate of > 75% onto genome reference. Berangan and Cavendish were mapped onto DH Pahang reference – A genome and Klutuk was mapped onto Klutuk Wulung reference – B genome, while for Yunnan (*Musa itinerans*) and Dajiao (ABB group) bananas have a mapping rate of <50% due to different genomic background with both of the genome references (Table 2). As a wild type banana (Wu et al. 2016; Liu et al. 2018), Yunnan showed conserved collinearity with the estimation of divergence about 5.8 Mya with *M. acuminata*, but the mapping rate of Yunnan banana onto A genome is quite low. In this case, the genomics evolution between Yunnan and *M. acuminata* is interesting to be explored. On the other hand, our mapping strategy showed an increasing number of mapped reads compared with other study (Deng et al. 2018).

**TABLE 2** Sample mapping to genome reference result.

| Cultivar | *Overall Mapping Rate* |
|---|---|
| Berangan | 75.7% |
| Dajiao | 40.8% |
| Yunnan | 49.7% |
| Cavendish | 80.1% |
| Klutuk | 86.8% |

On the contrary, Dajiao banana showed different mapping rate ( 40% vs 60%) compared with another study (Yang et al. 2015). However, it is important to note that the bioinformatics pipeline used in this study is different from the previous research, which still use the *M. acuminata* reference version 1 and TopHat v1.4.0. The genomic background difference between Dajiao (ABB Group) with the A genome is quite interesting as this might be caused by the more dominant influence of the B genome in the plant. In summary, the low percentage value of mapping rate onto the genome reference affected the number of transcripts and types of transcripts (Trapnell et al. 2014).

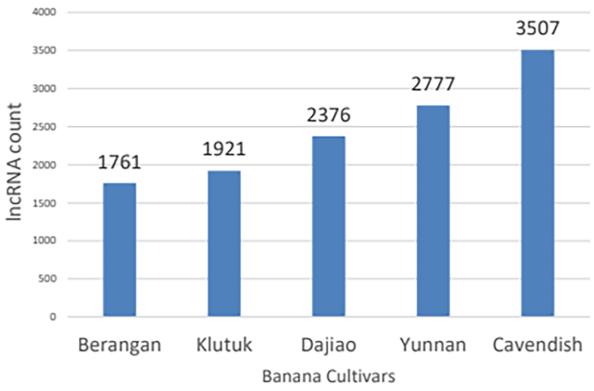


**FIGURE 2** The amount of lncRNA identified for Berangan, Klutuk, Dajiao, Yunnan, and Cavendish banana cultivar. Different cultivars were compared based on amount of lncRNA identified in this study.

The transcripts from the results of the mapped reads were reconstructed using Cufflinks and generated a unique transcriptome library using Cuffmerge. The transcripts datasets were then subjected to lncRNAs identification pipeline. The transcripts were filtered based on length, ORF length, protein domain, and high-similarity transcripts with housekeeping genes as previously described in methods. In final selection steps, lncRNA transcripts identified in this research were found to range from 3,509 to 1,761 of in Cavendish, Yunnan, Klutuk, Dajiao, and Berangan, respectively (Figure 2).

### 3.2. Characterization of lncRNA

The lncRNA transcripts were characterized based on class code in the genome into four classes (Table 3), then defined the distribution of lncRNA based on the genomic coordinate. The genic class code states the location of the lncRNA transcript relative to the surrounding genes, while the intergenic provide information on the existence of the transcript between two specific genes (Trapnell et al. 2010). In general, the intergenic composition in each sample has a lower amount compared to the genic region. Cavendish showed the highest abundance of lncRNAs within a reference intron, Klutuk showed the highest abundance for intergenic lncRNAs and generic exonic overlap, Dajiao showed the highest abundance for novel isoform lncRNAs and Yunnan revealed the most abundance lncRNAs at antisense region of reference. lncR-

**TABLE 3** Transcript distribution character based on class code.

| Class Code | Transcript Amount | | | | |
|---|---|---|---|---|---|
| | Berangan | Dajiao | Yunnan | Cavendish | Klutuk |
| i | 341 | 627 | 865 | 1364 | 23 |
| o | 31 | 21 | 13 | 14 | 45 |
| x | 178 | 185 | 349 | 278 | 70 |
| e | 258 | 734 | 1003 | 689 | 144 |
| j | 52 | 58 | 29 | 71 | 19 |
| u | 573 | 523 | 326 | 622 | 1687 |

NAs found in in this research, particularly in Cavendish bananas were quite impressive because usually lncRNAs are detected within intergenic regions and not many detected in the intronic region (Ma et al. 2013). The intergenic lncRNAs regulation is quite similar to mRNA and more conserved than others.

From another study, the genetic transcript has a specific role in overlapping genes region, so the lncRNA transcript quite similar sequence to mRNA. In this case, the transcript can work direct or indirect role for modification after the transcription, both attaching complementary to mRNA as well as Antisense lncRNAs, or can act as sponge miRNAs for endogenous competition. Intergenic and intronic lncRNAs are thought to be most likely to be regulated by diverse activation mechanisms of transcription, resulting in different activities depending on cell location. Even so, the molecular mechanism of intronic lncRNA remained questions because not yet explored so much (Wang and Chang 2011).

Unlike intronic lncRNA, the intergenic lncRNA widely explored, both the mechanism of cis-acting - the regulatory mechanism by lncRNA that will modulate the transcription of adjacent genes, and the trans-acting mechanism that will modulate genes that are located relatively far from lncRNA. All of the lncRNA transcripts detected in the banana become an exciting thing to explore for its function, because only based on the genomic coordinate can determine the possible mechanism of action, but not yet determine the role of each lncRNA in biological systems, especially in banana plants (Sampangi-Ramaiah et al. 2019). The presence of lncRNA (Table 3) showed the possibility of various gene regulation in banana plants, which then needs one-by-one analysis to determine its specific role in the biological system.

### *3.3. Distribution of lncRNA on chromosomes*

The distribution of lncRNA in chromosomes were defined and it was revealed that the distribution of lncRNAs among banana cultivars is uneven. In the Klutuk Banana (BB group), lncRNAs were similarly distributed on several chromosomes (four, eight, and nine), while in Cavendish, Dajiao, Berangan were concentrated on chromosome six, and in Yunnan (*M. itinerans*) is concentrated on chromosome four (Figure 3). Cavendish and Berangan (AAA

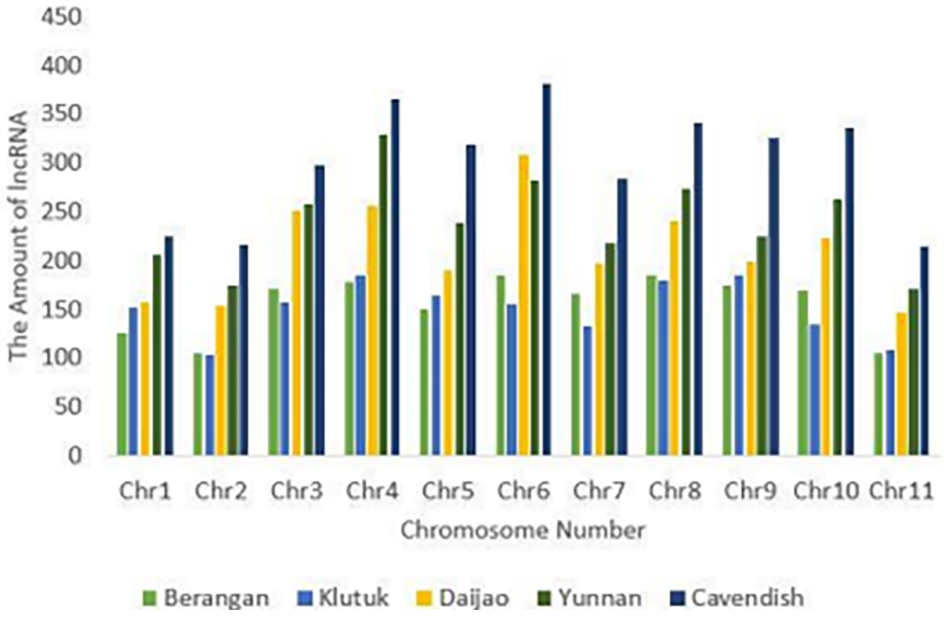


**FIGURE 3** The number of lncRNA Distribution Based on Chromosome Number. We used 11 chromosomes as baseline reference to highlight the distribution of lncRNA. Different cultivars were compared based on amount of lncRNA identified in this study.

group) showed a similar trend with Dajiao (ABB group). Here, the lncRNAs of each sample were not evenly distributed along banana chromosomes. The distribution of the lncRNA locus is per its mechanism of action, where it affects cis-acting gene expression by modifying the structure of chromosomes (Clark and Mattick 2011). In general, the distribution of lncRNA along chromosomes has a similar trend to the number of lncRNA transcripts detected (Figure 2).

### 3.4. lncRNA distribution based on size (bp)

Berangan, Dajiao, Yunnan, and Cavendish bananas observed as a whole has long dominated lncRNA types with the smallest range (200-400 bp) and 90% of all lncRNA consisting of transcripts with sizes below 1000 bp (Figure 4). Dajiao (ABB group) showed the highest amount of lncRNAs on the smallest range compared with others. It was suspected that lncRNAs with small size are more diverse compared to lncRNAs with a longer length. A smaller size of lncRNA transcript (<1000 bp) ensure the higher speed of transcription rate compared to lncRNAs with a length more than 1000 bp and minimize the possibility to be translated into protein. In agreement with Zhu et al. (2015), lncRNA tomatoes mostly have a small size of lncRNA where about 78% of the total identified lncRNA has a size of less than 1000 bp and only has 1 or 2 exons, which differs from the transcript which is encoded into protein translation (Zhu et al. 2015). In the lncRNAs on fruit, the composition of lncRNA of each cultivar can affect the process of fruit ripening to produce fruit with variations of taste, texture, biotic, and abiotic resistance, such as disease attack and extreme temperature changes (Hapsari and Lestari 2016).

### 3.5. Conservative sequence of lncRNAs in banana transcriptomic datasets (Musa spp.)

The lncRNA transcript does not have a conserved sequence relative to mRNA. Most lncRNAs are susceptible to mutations, but some showed to maintain specific lncRNAs, and the polymorphisms on lncRNA affecting the binding affinity to miRNA. In gastric cancer (Duan et al. 2018), SNPs in lncRNA can contribute to the development of gastric cancer. Like mRNA, conserved lncRNA thought to have an essential function in living things. Generally, four types of lncRNA conserved among humans and mice out of a total of 351 lncRNAs that are only sustainable in humans and 626 lncRNAs that are sustainable in mice, all of which interact with 56 genes that affect brain development (Li and Yang 2017).

Similar to the study, only a small portion of the banana crop was sustainable, but none of the lncRNAs conserved among the five banana cultivars, there was only a few specific lncRNAs maximum of four out of five samples (Figure 5). Most of the lncRNAs are specifically expressed in these cultivars and not found in other cultivars so that in banana lncRNA, so only a small portion are conserved. We

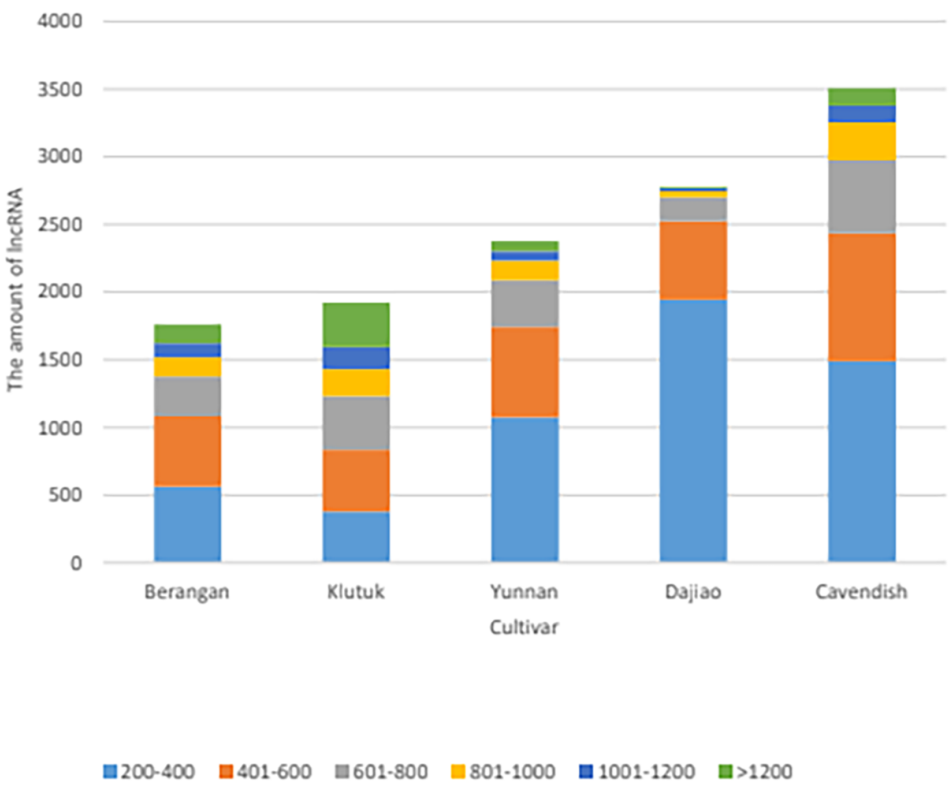


**FIGURE 4** lncRNA distribution in banana cultivars based on the length size (bp). The classification divided into different length classes: 1) 200-400 bp; 2) 401-600 bp; 3) 601-800 bp; 4) 801-1000 bp; 5) 1001-1200 bp; 6) > 1200 bp. The amount of lncRNA counted by numbers that have been identified.

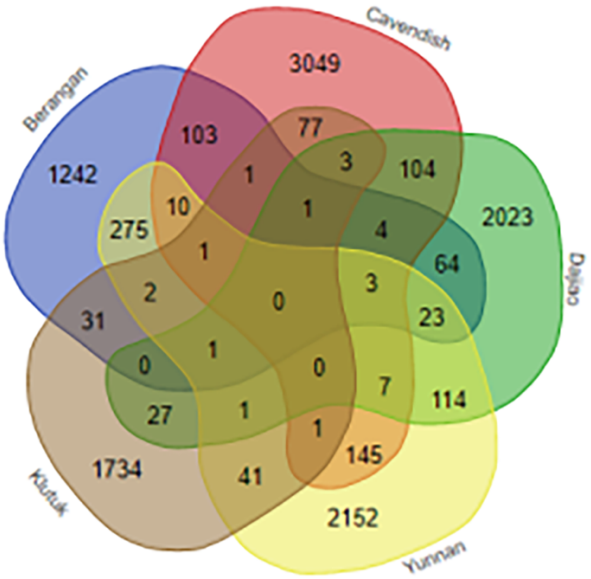


**FIGURE 5** Expressed lncRNA in different banana cultivars: Berangan, Cavendish, Dajiao, Klutuk, and Yunnan. Venn diagram showed specific and intersected lncRNA that expressed.

hypothesized the transcribed lncRNA was particular for each cultivar due to genomic background where it contains different genome groups (AAA, ABB, BB, Wild type).

### *3.6. Interaction of lncRNA with miRNA*

After the lncRNA expression profiling was done, we focused our observation to interaction between lncRNA and miRNA. Because of limited availability of resources, we decided to focus our study on analyzing lncRNA from the representative of A and B genome, which was the Cavendish and Klutuk banana, respectively. Furthermore, one of the lncRNA transcripts from both representatives were picked, TCONS_00057826 (Figure 6) was picked from klutuk banana, and it can be predicted from computational methods that it potentially has 45 interactions with miRNA.

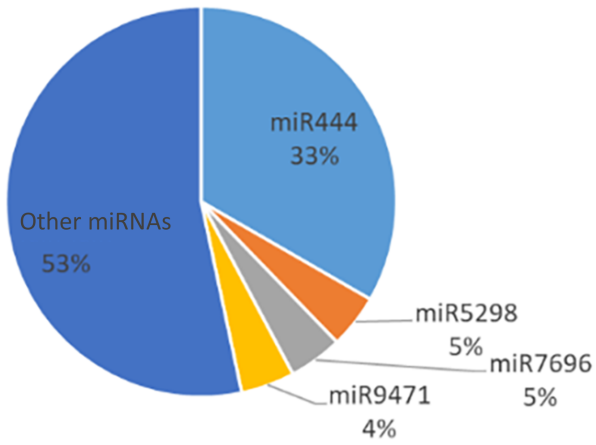


**FIGURE 6** Interaction probability percentage of TCONS_00057826 from Klutuk Banana and the respective miRNAs.

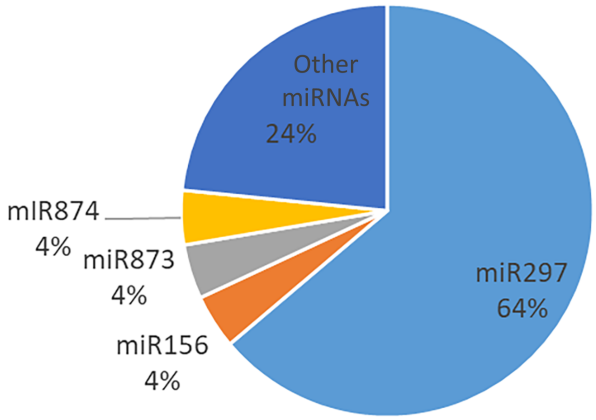


**FIGURE 7** Interaction probability percentage of TCONS_00002261 from Cavendish Banana and the respective miRNAs.

About 33% of the interactions showed the binding of lncRNA with the miR444 miRNA family, which hypothesized to have a role in the stress response to *Fusarium* wilt by FocTR4 in banana plants and responses to low temperatures (Fei et al. 2019). miR444 in rice plants can increase defence against RSV virus attacks by increasing the expression of the *OsRDR1* (antiviral RNA silencing) gene (Wang et al. 2016).

For lncRNA representative of Cavendish Banana - TCONS_00002261 (Figure 7) was picked. Using computational method, it was predicted to have interaction with miR397. The miR397 has a role in repressing the laccase gene (*LAC*), a gene that plays a role in temperature tolerance to protect the formation of reactive oxygen species (ROS) (Jones-Rhoades and Bartel 2004). The binding interactions percentage with miR397 about 64% and showed 30 interactions out of 47 possible binding events. Both lncRNAs are interactive and have a specific specificity to miRNA, so it is most likely that transcription of lncRNA will affect the expression of particular genes. The intronic position indicates that lncRNA will be expressed together with genes as introns, and inhibit the action of miRNA.

### *3.7. lncRNA Validation*

The two predicted lncRNAs, TCONS_000022261 and TCONS_00057826 were validated using the Reverse Transcription PCR (RT-PCR) method (Figure 8). A- acts as a negative control of Cavendish bananas and B- is a

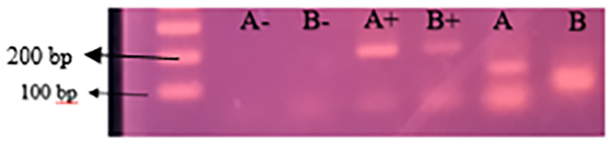


**FIGURE 8** Validation of lncRNA (TCONS_00057826 and TCONS_00002261) with Reverse Transcription PCR method using Cavendish and Klutuk Banana as an Example. A for Cavendish Banana and B for Klutuk Banana. Symbol of (-) for Negative and (+) for Positive Control in PCR.

negative control of Klutuk bananas, A + as a positive control on Cavendish bananas and B + as a positive control on Klutuk bananas, with negative and positive controls according to experimental design. A is the PCR result on Cavendish banana with MalncRNA_00002261 primer from TCONS lncRNA _00002261, B is PCR result on Klutuk banana using MblncRNA_00057826 primer from TCONS_00057826.

Further analysis is needed to see the overall significance of lncRNA, both for those that are validated and those that have not been verified to understand better the significance of these transcripts which affects expression of genes in banana plants, especially since all analysis were still considered putative. In further research to find out the function of lncRNA, it can be carried out an analysis of lncRNA inhibition with miRNA by analyzing expressions, one of which is transient agroinfiltration assay, using *Agrobacterium tumefaciens* which carries the expression factor inserted by the target lncRNA. Then, comparative study of gene expression can be done between expressions in controls and those treated with lncRNA, with both up-regulated group and down-regulated group (Wang et al. 2015).

## 4. Conclusions

The lncRNAs were successfully identified and characterized from different varieties of bananas taken from several transcriptomics libraries. The number of lncRNAs identified are different between varieties. Furthermore, these lncRNAs belonging to their respective plants have an uneven distribution among chromosomes and were found to have unsustainable and less conserved characteristics. Among all the interactions as eTMs, only two predicted interactions were analyzed, which are between TCONS_000022261, which has high affinity to bind with miR397 from Cavendish and might have potential role in regulating gene expressions related to environmental stresses. Secondly, the interaction between TCONS_00057826, which has high affinity to bind with miR444 from Klutuk that might play an important role as a response mechanism to react to diseases, particularly against fusarium. These transcripts expressions were also able to be validated with RT-PCR method. Although more research might be needed to have more conclusive results, this study could be used as an initial reference for similar studies in the future with more refined designs and methods.

## Acknowledgments

Authors thank “Riset Unggulan ITB 2019” from Biosciences and Biotechnology Research Center, Institut Teknologi Bandung for their support during this study.

## Authors’ contributions

AMP, FMD and HN designed the study. AMP carried out the laboratory work. AMP, FMD, HN analyzed the data. AMP, FMD, HN wrote the manuscript. All authors read and approved the final version of the document.

## Competing interests

Authors declared no competing interests, such as any financial, professional, or personal relationships that are relevant to the submitted work.

## References

Backiyarani S, Uma S, Saraswathi MS, Saravanakumar AS, Chandrasekar A. 2015. Transcriptome analysis of banana (*Musa balbisiana*) based on next-generation sequencing technology. Turkish J. Agric. For. 39(5):1–13. doi:10.3906/tar-1406-171.

Ballantyne MD, McDonald RA, Baker AH. 2016. lncRNA/MicroRNA interactions in the vasculature. Clin. Pharmacol. Ther. 99(5):494–501. doi:10.1002/cpt.355.

Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2017. GenBank. Nucleic Acids Res. 45(D1):D37–D42. doi:10.1093/nar/gkw1070.

Borah P, Das A, Milner MJ, Ali A, Bentley AR, Pandey R. 2018. Long non-coding rnas as endogenous target mimics and exploration of their role in low nutrient stress tolerance in plants. Genes (Basel). 9(9):459. doi:10.3390/genes9090459.

Cheesman E. 1948. Classification of the bananas II. The genus *Musa* L. Kew Bulletin 2:106–117.

Clark MB, Mattick JS. 2011. Long noncoding RNAs in cell biology. Semin. Cell Dev. Biol. 22(4):366–376. doi:10.1016/j.semcdb.2011.01.001.

Dai X, Zhuang Z, Zhao PX. 2018. PsRNATarget: A plant small RNA target analysis server (2017 release). Nucleic Acids Res. 46(W1):W49–W54. doi:10.1093/nar/gky316.

Davey MW, Gudimella R, Harikrishna JA, Sin LW, Khalid N, Keulemans J. 2013. ”A draft *Musa balbisiana* genome sequence for molecular genetics in polyploid, inter- and intra-specific *Musa* hybrids”. BMC Genomics 14(1):683. doi:10.1186/1471-2164-14-683.

Deng S, Cheng C, Liu Z, Chen Y, Zhang Z, Huang Y, Lin Y, Wang T, Lai Z. 2018. Comparative transcriptome analysis reveals a role for anthocyanin biosynthesis genes in the formation of purple peel in Minhou wild banana (*Musa itinerans* Cheesman). J. Hortic. Sci. Biotechnol. 94(2):1–17. doi:10.1080/14620316.2018.1473055.

D’hont A, Denoeud F, Aury JM, Baurens FC, Carreel F, Garsmeur O, Noel B, Bocs S, Droc G, Rouard M, Da Silva C, Jabbari K, Cardi C, Poulain J, Sou-

quet M, Labadie K, Jourda C, Lengellé J, Rodier-Goud M, Alberti A, Bernard M, Correa M, Ayyampalayam S, Mckain MR, Leebens-Mack J, Burgess D, Freeling M, Mbéguié-A-Mbéguié D, Chabannes M, Wicker T, Panaud O, Barbosa J, Hribova E, Heslop-Harrison P, Habas R, Rivallan R, Francois P, Poiron C, Kilian A, Burthia D, Jenny C, Bakry F, Brown S, Guignon V, Kema G, Dita M, Waalwijk C, Joseph S, Dievart A, Jaillon O, Leclercq J, Argout X, Lyons E, Almeida A, Jeridi M, Dolezel J, Roux N, Risterucci AM, Weissenbach J, Ruiz M, Glaszmann JC, Quétier F, Yahiaoui N, Wincker P. 2012. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. Nature 488(7410):213–217. doi:10.1038/nature11241.

Duan F, Jiang J, Song C, Wang P, Ye H, Dai L. 2018. Functional long non-coding RNAs associated with gastric cancer susceptibility and evaluation of the epidemiological e ffi cacy in a central Chinese population. Gene 646(January):227–233. doi:10.1016/j.gene.2017.12.063.

Dwivany FM, Nugrahapraja H, Sari LDN, Putri RR, Novianti C. 2021. Transcriptome dataset of ethylene-treated Klutuk Wulung banana. Data in Brief 38:107376. doi:10.1016/j.dib.2021.107376.

Fei S, Czislowski E, Fletcher S, Peters J, Batley J, Aitken E, Mitter N. 2019. Small RNA profiling of Cavendish banana roots inoculated with *Fusarium oxysporum* f. sp. *cubense* race 1 and tropical race 4. Phytopathol. Res. 1:22. doi:10.1186/s42483-019-0029-3.

Franco-Zorrilla JM, Valli A, Todesco M, Mateos I, Puga MI, Rubio-somoza I, Leyva A, Weigel D, García JA, Paz-ares J, Puga I, Franco-zorrilla M, Todesco M, Mateos I, Paz-ares J, Rubio-somoza I, Leyva A, Weigel D, Garcı JA. 2007. Target mimicry provides a new mechanism for regulation of microRNA activity. Nat. Genet. 39(8):1033–1037. doi:10.1038/ng2079.

Hapsari L, Lestari DA. 2016. Fruit characteristic and nutrient values of four Indonesian banana cultivars (*Musa* spp.) at different genomic groups. Agrivita 38(3):303–311. doi:10.17503/agrivita.v38i3.696.

Jones-Rhoades MW, Bartel DP. 2004. Computational Identification of Plant MicroRNAs and Their Targets, Including a Stress-Induced miRNA. Molecular Cell 14(6):787–799. doi:10.1016/j.molcel.2004.05.027.

Kang C, Liu Z. 2015. Global identification and analysis of long non-coding RNAs in diploid strawberry *Fragaria vesca* during flower and fruit development. BMC Genomics 16(1):1–15. doi:10.1186/s12864-015-2014-2.

Kim ED, Sung S. 2012. Long noncoding RNA: Unveiling hidden layer of gene regulatory networks. Trends in Plant Science 17(1):16–21. doi:10.1016/j.tplants.2011.10.008.

Leinonen R, Akhtar R, Birney E, Bower L, Cerdeno-Tárraga A, Cheng Y, Cleland I, Faruque N, Goodgame N, Gibson R, Hoad G, Jang M, Pakseresht N, Plaister S, Radhakrishnan R, Reddy K, Sobhany S, Hoopen PT, Vaughan R, Zalunin V, Cochrane G. 2011. The European nucleotide archive. Nucleic Acids Res. 39(SUPPL. 1):44–47. doi:10.1093/nar/gkq967.

Lesk AM. 2005. Database Annotation in Molecular Biology: Principles and Practice. John Wiley & Sons, Ltd. URL https://onlinelibrary.wiley.com/doi/book/10.1002/0470012420.

Li D, Yang MQ. 2017. Identification and characterization of conserved lncRNAs in human and rat brain. BMC Bioinformatics 18:489. doi:10.1186/s12859-017-1890-7.

Li L, Eichten SR, Shimizu R, Petsch K, Yeh CT, Wu W, Chettoor AM, Givan SA, Cole RA, Fowler JE, Evans MM, Scanlon MJ, Yu J, Schnable PS, Timmermans MC, Springer NM, Muehlbauer GJ. 2014. Genome-wide discovery and characterization of maize long non-coding RNAs. Genome Biol. 15(2):R40. doi:10.1186/gb-2014-15-2-r40.

Liu J, Wang H, Chua NH. 2015. Long non-coding RNA transcriptome of plants. Plant Biotechnol. J. 13(3):319–328. doi:https://doi.org/10.1111/pbi.12336.

Liu W, Cheng C, Lin Y, XuHan X, Lai Z. 2018. Genome-wide identification and characterization of mRNAs and lncRNAs involved in cold stress in the wild banana (*Musa itinerans*). PLoS One 13(7):1–27. doi:10.1371/journal.pone.0200002.

Ma L, Bajic VB, Zhang Z. 2013. On the classification of long non-coding RNAs. RNA Biol. 10(6):924–933. doi:10.4161/rna.24604.

Muthusamy M, Uma S, Backiyarani S, Saraswathi MS, Chandrasekar A. 2016. Transcriptomic Changes of Drought-Tolerant and Sensitive Banana Cultivars Exposed to Drought Stress. Frontiers Plant. Sci. 7(November):1–14. doi:10.3389/fpls.2016.01609.

Muthusamy M, Uma S, Suthanthiram B, Saraswathi MS, Chandrasekar A. 2019. Genome-wide identification of novel, long non-coding RNAs responsive to *Mycosphaerella eumusae* and *Pratylenchus coffeae* infections and their differential expression patterns in disease-resistant and sensitive banana cultivars. Plant Biotechnol. Rep. 13(1):73–83. doi:10.1007/s11816-018-00514-z.

Peri S, Roberts S, Kreko IR, McHan LB, Naron A, Ram A, Murphy RL, Lyons E, Gregory BD, Devisetty UK, Nelson AD. 2020. Read Mapping and Transcript Assembly: A Scalable and High-Throughput Workflow for the Processing and Analysis of Ribonucleic Acid Sequencing Data. Front. Genet. 10:1915–1927. doi:10.3389/fgene.2019.01361.

Ravishankar KV, Sampangi-ramaiah MH, Ajitha R, Khadke GN, Chellam V. 2015. Insights into *Musa balbisiana* and *Musa acuminata* species divergence and development of genic microsatellites by transcriptomics approach. Plant Gene 4:78–82. doi:10.1016/j.plgene.2015.09.007.

Rozen S, Skaletsky H. 2000. Primer3 on the WWW for

general users and for biologist programmers. Methods Mol. Biol. 2000(132):365–386. doi:10.1385/1-59259-192-2:365.

Sampangi-Ramaiah MH, Ravishankar KV, Rekha A, Hastantram M, Ramaiah S, Venkataramana K, Ajitha R. 2019. Long non-coding RNAs in banana : prediction , mapping and their comparative studies using *Musa balbisiana* and *Musa acuminata* transcriptome. Trees 33(2):359–369. doi:10.1007/s00468-018-1781-1.

Setiabudi E, Meitha K, Dwivany FM. 2021. *In silico* characterization and comparison of the fruit ripening related betaamylase (BAM) gene family in banana genome A and B. Indones. J. Biotechnol. 26(4):175–182. doi:10.22146/ijbiotech.65142.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2014. Erratum: Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc. 9(10):2513. doi:10.1038/nprot.2012.016.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28(5):511–515. doi:10.1038/nbt.1621.

Wang H, Jiao X, Kong X, Hamera S, Wu Y, Chen X, Fang R, Yan Y. 2016. A Signaling Cascade from miR444 to RDR1 in Rice Antiviral RNA Silencing Pathway. Plant Physiol. 170(4):2365–2377. doi:10.1104/pp.15.01283.

Wang J, Yu W, Yang Y, Li X, Chen T, Liu T, Ma N, Yang X, Liu R, Zhang B. 2015. Genome-wide analysis of tomato long non-coding RNAs and identification as endogenous target mimic for microRNA in response to TYLCV infection. Sci. Rep. 5:16946. doi:10.1038/srep16946.

Wang KC, Chang HY. 2011. Molecular Mechanisms of Long Noncoding RNAs. Molecular Cell 43(6):904–914. doi:10.1016/j.molcel.2011.08.018.

Wu W, Yang Yl, He Wm, Rouard M, Li Wm, Xu M, Roux N, Ge XJ. 2016. Whole genome sequencing of a banana wild relative *Musa itinerans* provides insights into lineage-specific diversification of the Musa genus. Sci. Rep. 6(31586):1–11. doi:10.1038/srep31586.

Yang Qs, Gao J, He Wd, Dou Tx, Ding Lj, Wu Jh, Li CY, Peng Xx, Zhang S, Yi GJ. 2015. Comparative transcriptomics analysis reveals difference of key gene expression between banana and plantain in response to cold stress. BMC Genomics 16:1–18. doi:10.1186/s12864-015-1551-z.

Yu T, Tzeng DT, Li R, Chen J, Zhong S, Fu D, Zhu B, Luo Y, Zhu H. 2019. Genome-wide identification of long non-coding RNA targets of the tomato mads box transcription factor rin and function analysis. Ann. Bot. 123(3):469–482. doi:10.1093/aob/mcy178.

Zhao X, Li J, Lian B, Gu H, Li Y, Qi Y. 2018. Global identification of *Arabidopsis* lncRNAs reveals the regulation of MAF4 by a natural antisense RNA. Nat. Commun. 9(1):1–12. doi:10.1038/s41467-018-07500-7.

Zhu B, Yang Y, Li R, Fu D, Wen L, Luo Y, Zhu H. 2015. RNA sequencing and functional analysis implicate the regulatory role of long non-coding RNAs in tomato fruit ripening. Journal of Experimental Botany 66(15):4483–4495. doi:10.1093/jxb/erv203.