O-ORT-04

Exploration of cassava lincRNAs relevant to defense response against cassava brown streak virus (CBSV) infection

Rungaroon Suksamran[1], Treenut Saithong[2,3], Chinae Thammarongtham[4], Saowalak Kalapanulak[2,3]*

[1]*Biotechnology Program, School of Bioresources and Technology, King Mongkut's University of Technology Thonburi (Bang Khun Thian), Bangkok, 10150, Thailand*
[2]*Bioinformatics and Systems Biology Program, School of Bioresources and Technology, King Mongkut's University of Technology Thonburi (Bang Khun Thian), Bangkok, 10150, Thailand*
[3]*Systems Biology and Bioinformatics Research Group, Pilot Plant Development and Training Institute, King Mongkut's University of Technology Thonburi (Bang Khun Thian), 10150, Thailand*
[4]*Biochemical Engineering and Systems Biology Research group, National Center for Genetic Engineering and Biotechnology at King Mongkut's University of Technology Thonburi (Bang Khun Thian), 10150, Thailand*

*Corresponding author. E-mail: saowalak.kal@kmutt.ac.th
DOI:

ABSTRACT

Cassava brown streak virus (CBSV) poses a huge threat to cassava production and can lead to substantial yield and quality losses. Understanding of plant gene regulation against CBSV infection is necessary for developing resistant cultivars. Transcriptome analysis offers insights into the underlying gene regulation and defense mechanisms against pathogens. However, most of the CBSV-related transcriptome analysis in cassava has mainly focused on protein coding genes, and comprehensive knowledge of the roles of long non-coding RNAs (lncRNAs), which have been linked to the defense mechanisms of some plants, remains elusive for cassava. Here, we performed a transcriptome-wide screening for long intergenic non-coding RNAs (lincRNAs) in a CBSV-resistant cultivar under CBSV infection. A total of 415 expressed lincRNAs were found to be distributed across the cassava genome. Twenty-nine of them were significantly differentially expressed relative to control (absolute $\log_2$ fold change ≥ 2 and corrected p-value < 0.05). Among these, 4 were potentially involved in regulating transcripts of disease resistance proteins,

heat shock proteins, chaperones, late embryogenesis abundant proteins, major latex proteins, stress- responsive GRAS transcription factors, and hormone- regulated proteins at the early stage of infection. To our knowledge, this work provides the first evidence of lincRNAs involved in cassava defense response against CBSV. It provides insightful information towards a comprehensive understanding of post-transcriptional regulation in plants against viral infection, for which further wet lab experimentation is crucial.

Keywords: CBSV, Disease resistance gene, LincRNA, Transcriptome

INTRODUCTION

Cassava is a starchy root crop, upon which over 800 million people rely annually. It is widely utilized in the food industry and as an alternative resource for energy production. However, pathogens have threatened its production, and although some resistant cultivars have been bred, their consistency in different environments has been questioned. One of the most important plant pathogens is the cassava brown streak virus (CBSV), which is transmitted by the whitefly (*Bemisia tabaci*) or through infected planting materials, and symptoms include chlorosis of leaves and necrosis of storage roots. Most cassava cultivars are susceptible to CBSV. Some have been reported to be resistant, remaining free of symptoms even after exposure to the virus, while some others can survive up to 12 months with small chlorotic tissue (Amuge et al., 2017).

A high number of differentially expressed genes, relative to control, have also been reported for a resistant cultivar ('Namikonga') under CBSV infection (Amuge et al., 2017). Some of the differentially expressed genes were matched to defense response genes — leucine-rich repeats (LRRs) and nucleotide binding domain (NBARC), pathogenesis-related (PR) proteins, late embryogenesis abundant (LEA) proteins, transcription factors (NAM, ATAF, NAC, GATA, and GRAS), chaperones, and heat shock proteins (HSP), reported in earlier studies. Overall, CBSV defense mechanisms in cassava, including important regulators, are poorly understood as most of the previous works focused on protein coding genes (Maruthi et al., 2014; Amuge et al., 2017; Anjanappa et al., 2018; Sheat et al., 2019). Moreover, not much is known about the roles of long non-coding RNAs (lncRNAs) in cassava, although studies in some other species have demonstrated that lncRNAs are important regulators and are involved in plant defense mechanisms.

Long non-coding RNAs are transcripts that are longer than 200 nt without protein coding potential. They can be classified as lncNAT (long antisense

transcript), intronic/sense lncRNA, and lincRNA (long intergenic non-coding RNA) based on their genomic location. Recent works have demonstrated the importance of lncRNAs for plant disease tolerance. For example, silencing of two lncRNAs, GhlncNAT-ANX2 and GhlncNAT-RLP7, has been reported to confer resistance against wilt disease in cotton, caused by *Verticillium dahliae* (Zhang et al., 2018). Also, lncRNA has been used to enable tomato resistance to *Phytophthora infestans* through the miRNA target mimicry mechanism (Jiang et al., 2019). These and other similar findings emphasize the need for in-depth research into the importance of lncRNAs and ways they could be exploited for the development of resistant cultivars.

Nowadays, the availability of RNAseq data and bioinformatics tools facilitate transcriptome-wide lncRNA detection and analysis to characterize their functions. Here, we explored of lincRNAs that are relevant to cassava brown streak virus resistance in the 'Namikongka' cultivar, using RNAseq datasets from Amuge et al. (2107). We identified interesting differentially expressed lincRNAs, which could be involved in post-transcriptional regulation of defense-response proteins.

MATERIALS AND METHODS

Data source and RNAseq data processing

RNAseq data from Amuge et al. (2017) with BioProject ID PRJNA360340 were retrieved from NCBI. Expression in leaves of a CBSV- resistant cultivar, 'Namikonga,' was studied in triplicate at 3 time points, namely 0 h — before graft-inoculation with axillary buds from CBSV-infected (infection) and healthy (control) plants, early infection (5 days after grafting), and late infection (54 days after grafting). High-quality reads (Phred-quality score > 20) without primer or adapter contamination in each sample were aligned with the cassava reference genome (AM560v6.1) using STAR aligner (Dobin et al., 2013). Sequences and locations of assembled transcripts were obtained using Cufflinks (Trapnell et al., 2010).

Bioinformatics pipeline for lincRNA identification

Assembled transcripts in intergenic regions with class code "u" and not overlapped with protein coding genes were processed using several filtering criteria, as shown in Figure 1. Transcripts longer than 200 nt with more than 500 bp distance from nearest coding genes were analyzed for coding potential using computational tools, CPC (Kong et al., 2007) and CPAT (Wang et al., 2013). Then, transcripts without similarity to protein sequences or domains were identified as putative lincRNAs (Figure 1A).

Prediction of lincRNA target

Protein coding transcripts of cassava were predicted as targets for the lincRNAs based on sequence complementary and free energy of molecular interaction using LncTar (Li et al., 2015) with a normalized free energy (ndG) cutoff of less than -0.1. The functions of the lincRNAs were then inferred from those of their targets.

Expression analysis and visualization

Count read table for both the lincRNAs and corresponding protein coding transcripts was generated using a computational tool, Verse (Zhu et al., 2016). The expression of each transcript was normalized by relative log expression (RLE) in DESeq2 (Love et al., 2014). Transcripts with an absolute $\log_2$ fold change of ≥ 2, relative to control, at a corrected p-value of < 0.05 were considered differentially expressed transcripts. Charts and plots were constructed by ggplot2 in R, and Cytoscape (Shannon et al., 2003) was used for network visualization.

RESULTS AND DISCUSSION

LincRNAs expression and characteristics in a CBSV-resistant cultivar

After aligning the high-quality reads to the reference genome, two replicate samples in control, IDs SRR5154680 and SRR5154730 from 5 and 54 DAG, respectively, that varied highly from their sister-replicates in the intergenic region were excluded, remaining two biological replicates each. In total, we found 415 expressed lincRNAs using the bioinformatics pipeline for lincRNA identification (Figure 1A). This relatively small number of lincRNAs could be because poly A tailed RNA libraries were used for the identification, which limited the scope of the discovery. Overall, the number of expressed lincRNAs at the various infection stages differed (Figure 1B), which is indicative of the expression dynamism, and 89 lincRNAs were consistently expressed (Figure 1C).
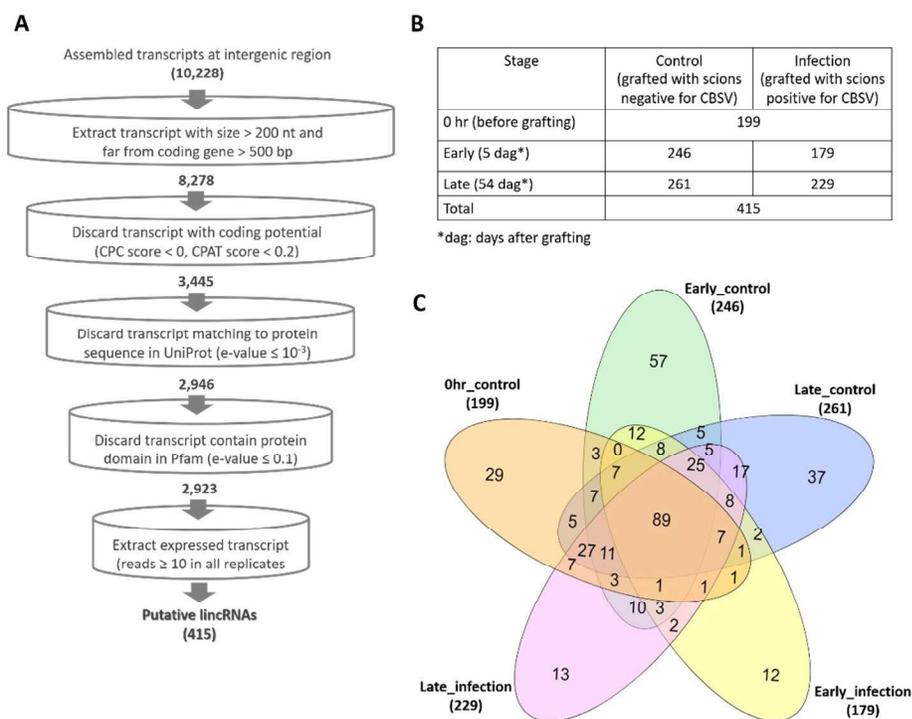
Figure 1. Bioinformatics pipeline for lincRNA identification (A). The number of expressed lincRNAs under control and infection conditions at three time points: before grafting (0 h), early stage (5 dag) and late stage (54 dag) (B) and their prevalence in each condition are represented by a Venn diagram (C).
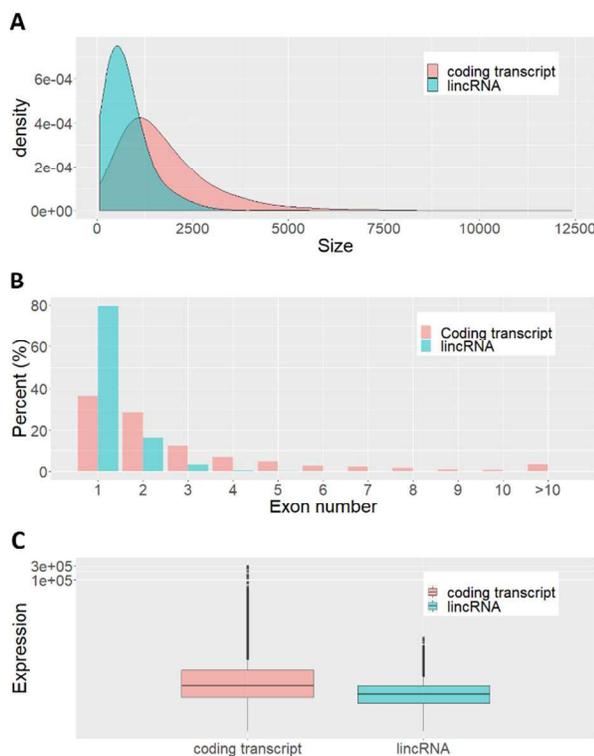
Figure 2.   A comparison of the 415 putative lincRNAs and protein coding
transcripts based on transcript size (A), exon number per transcript (B),
and expression level after normalization by DESeq2 (C).

The characteristics of the 415 lincRNAs were compared to those of the 8,047 expressed (mapped reads $\geq$ 10 in all replicates) protein coding transcripts (Figure 2). The lincRNAs showed a size of $204 - 5,169$ nt, an exon number of $1 - 5$ and an expression level of $0.80 - 1,093.82$; whereas the protein coding transcripts showed a size of $72 - 12,406$ nt, an exon number of $1 - 48$, and an expression level of $0.80 - 303,604.61$. These lower values observed in lincRNAs compared to their targets agree with previously reported findings in other plants, e.g. soybean (Golicz et al., 2018) and tomato (Wang et al., 2018).

Differential expression of lincRNAs at the early CBSV infection and implications for defense-response gene regulation

The relevance of lincRNAs in CBSV resistance was investigated by analyzing their expression in control and infection conditions using DESeq2. Twenty-seven differentially-expressed (DE) lincRNAs were found at early CBSV infection (23 up-regulation and 4 down-regulation), whereas 7 were found at the late stage (4 up-regulation and 3 down-regulation). These corroborate Amuge et al. (2017), who reported a higher number of DE transcripts at early CBSV infection stage, and could be indicative of its importance for CBSV resistance. Functions of the lincRNAs that were differentially expressed at early CBSV infection were inferred from their predicted targets using LncTar.

In total, 9,692 cassava protein coding transcripts were predicted as targets of the lincRNAs, and 69 of them were differentially expressed at the early infection stage. Among the DE targets, 7 have been reported as defense-response genes (Figure 3) encoding heat shock proteins, disease resistant proteins (CC-NBS-LRR class), chaperones, GRAS transcription factors, late embryogenesis abundant (LEA) proteins, a major latex protein, and a gibberellin-regulated protein (Amuge et al., 2017;Yang et al., 2015; De Bruyne et al., 2014).

Furthermore, the up- and down-regulation patterns found in defense response transcripts were consistent with Amuge et al. (2017) notwithstanding the stricter criteria employed. Interestingly, we found a subnetwork of lincRNAs, TCONS_00040811 that performs as a master regulator, regulating a disease resistance protein, a heat shock protein, a chaperone and a GRAS transcription factor (Figure 3). Master regulators might offer insights into the complexity of gene regulation and could be explored for improving tolerance to CBSV.
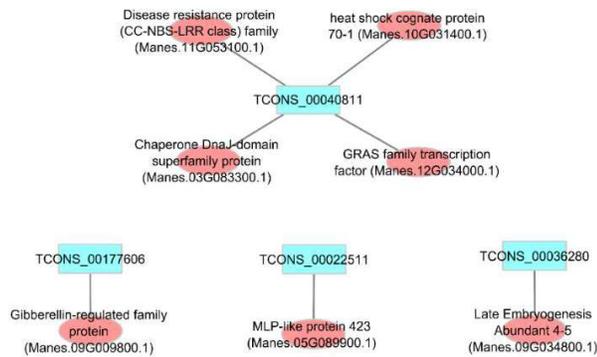
Figure 3. Interaction of DE lincRNAs and their DE predicted targets that were matched to previously reported defense response genes at early CBSV infection. Blue rectangular and red oval nodes represent lincRNAs and their targets, respectively.

The co-expression analysis of the seven defense response-related lincRNAs-target pairs (Figure 3) revealed similar expression profiles (fold change) in two of the pairs (Figure 4A-B), which was indicative of positive regulation in response to CBSV. Four pairs had dissimilar profiles (Figure 4C-F), probably indicating negative regulation. The remaining pair showed no evidence of a relationship at the three time points (Data not shown).
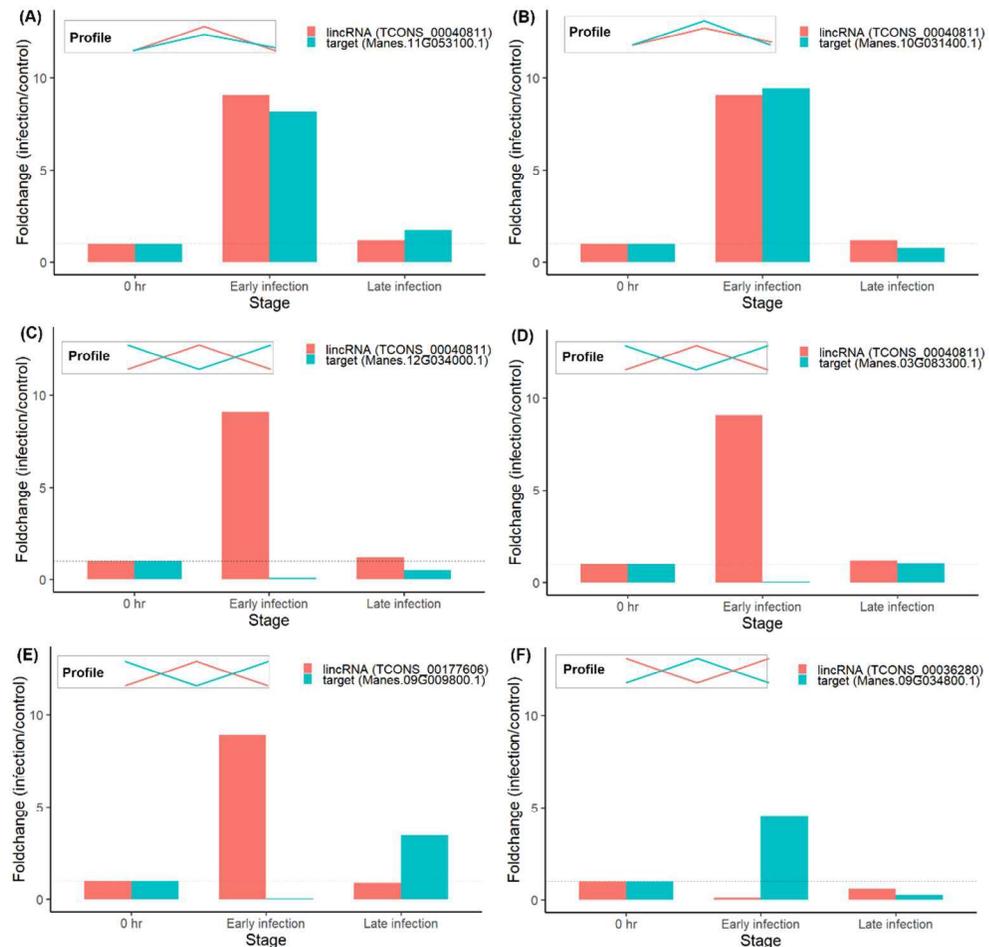
Figure 4. Co-expression profiles (Fold change) of identified CBSV-responsive lincRNAs-target pairs at three time points. A similarity in profiles of fold change along the time points suggests a positive regulation, whereas dissimilar trends suggest a negative regulation. Red and blue colors represent lincRNAs and their targets, respectively.

The positive regulations identified between lincRNA TCONS_00040811 and a disease resistance transcript (Manes. 11G053100.1) and between lincRNA TCONS_00040811 and a heat shock cognate transcript (Manes. 10G031400.1) could be due to target regulation by lncRNAs through miRNA target mimicry or competitive endogenous RNA mechanism (Yoon et al., 2014; Wang et al., 2016),

which inhibit miRNA –mediated target silencing resulting in a concordant profile between lincRNAs and their targets. On the other hand, the negative regulation could be explained by the suppression of targets by lncRNA-generated miRNAs (lncRNA acts as miRNA precursor), leading to discordant lncRNA and target profiles (Keniry et al., 2012).

CONCLUSIONS

This study identified a collection of 415 lincRNAs in a CBSV-resistant cassava cultivar, based on transcriptome-wide screening using a bioinformatics pipeline. Twenty-nine of them were differentially expressed, relative to control, at the early stage of CBSV infection. Among them, 4 could bind to at least one of the CBSV-responsive (five) and plant defense (two) transcripts, from previous studies. However, only 6 of the lincRNA-target pairs had profiles that showed evidence of a relationship at 0, early, and late infection stages. The knowledge gained from our findings could be of immense benefit for cassava improvement against pathogens. However, experimental validation of these lincRNAs and further investigation of their molecular mechanisms are necessary.

ACKNOWLEDGEMENTS

REFERENCES

Amuge, T., Berger, D.K., Katari, M.S., Myburg, A.A., Goldman, S.L., and Ferguson, M.E. 2017. A time series transcriptome analysis of cassava (*Manihot esculenta* Crantz) varieties challenged with Ugandan cassava brown streak virus. Scientific Reports. 7(1): 9747.

Anjanappa, R.B., Mehta, D., Okoniewski, M.J., Szabelska-Berȩsewicz, A., Gruissem, W., and Vanderschuren, H. 2018. Molecular insights into Cassava brown streak virus susceptibility and resistance by profiling of the early host response. Molecular Plant Pathology. 19(2): 476-489.

De Bruyne, L., Höfte, M., and De Vleesschauwer, D. 2014. Connecting growth and defense: the emerging roles of brassinosteroids and gibberellins in plant innate immunity. Molecular Plant. 7(6): 943-959.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., and Gingeras, T.R. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 29(1): 15–21.

Golicz, A.A., Singh, M.B., and Bhalla, P.L. 2018. The long intergenic noncoding RNA (lincRNA) landscape of the soybean genome. Plant physiology. 176(3): 2133–2147.

Jiang, N., Cui, J., Shi, Y., Yang, G., Zhou, X., Hou, X. … Luan, Y. 2019. Tomato lncRNA23468 functions as a competing endogenous RNA to modulate NBS-LRR genes by decoying miR482b in the tomato-*Phytophthora infestans* interaction. Horticulture Research. 6(28).

Keniry, A., Oxley, D., Monnier, P., Kyba, M., Dandolo, L., Smits, G., and Reik, W. 2012. The *H19* lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. Nature cell biology. 14(7): 659–665.

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 research. 35: 345–349.

Li, J., Ma, W., Zeng, P., Wang, J., Geng, B., Yang, J., and Cui, Q. 2015. LncTar: a tool for predicting the RNA targets of long noncoding RNAs. Briefings in Bioinformatics. 16(5): 806-812.

Love, M.I., Huber, W., and Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 15(12): 550.

Maruthi, M.N., Bouvaine, S., Tufan, H.A., Mohammed, I.U., and Hillocks, R.J. 2014. Transcriptional response of virus-infected cassava and identification of putative sources of resistance for cassava brown streak disease. PloS one. 9(5): e96642.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., … Ideker, T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research. 13(11): 2498-2504.

Sheat, S., Fuerholzner, B., Stein, B., and Winter, S. 2019. Resistance against cassava brown streak viruses from Africa in cassava germplasm from south America. Frontiers in plant science. 10: 567.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., … Pachter, L. 2010. Transcript assembly and abundance estimation from

RNA- Seq reveals thousands of new transcripts and switching among isoforms. Nature Biotechnology. 28(5): 511–515.

Wang, G.Q., Wang, Y., Xiong, Y., Chen, X.C., Ma, M.L., Cai, R., … Pang, W.J. 2016. Sirt1 AS lncRNA interacts with its mRNA to inhibit muscle formation by attenuating function of miR-34a. Scientific reports. 6: 21865.

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 research. 41(6): e74.

Wang, M., Zhao, W., Gao, L., and Zhao, L. 2018. Genome-wide profiling of long non-coding RNAs from tomato and a comparison with mRNAs associated with the regulation of fruit ripening. BMC plant biology. 18(1), 75. doi:10.

Yang, C.L., Liang, S., Wang, H.Y., Han, L.B., Wang, F.X., … Xia, G.X. 2015. Cotton major latex protein 28 functions as a positive regulator of the ethylene responsive factor 6 in defense against *Verticillium dahlia*. Molecular Plant. 8(3): 399-411.

Yoon, J.H., Abdelmohsen, K., and Gorospe, M. 2014. Functional interactions among microRNAs and long noncoding RNAs. Seminars in cell & developmental biology. 34: 9–14.

Zhang, L., Wang, M., Li, N., Wang, H., Qiu, P., Pei, L., … Zhang, X. 2018. Long noncoding RNAs involve in resistance to *Verticillium dahliae*, a fungal disease in cotton. Plant biotechnology journal. 16(6): 1172–1185.

Zhu Q, Fisher, S.A., Shallcross, J., and Kim, J. 2016. VERSE: a versatile and efficient RNA- Seq read counting tool [Online] . Available: https://pdfs.semanticscholar.org/22cc/eb0158c3805cf4c57fa54b02ef30e662e373.pdf [2018, November 30].