Gene Expression Analysis For Lychee Under The Influence Of Peronophythora Litchii

Eric Zhuang

1 **SUMMARY**

2   Lychee (Litchi chinensis Sonn.) is a delicious and nutritious fruit that can be infected by a
3 plant pathogen called Peronophythora litchii. In this study, our main **hypothesis** is that Lychee
4 with or without the inoculation of P. Litchii should exhibit a set of differentially expressed genes
5 that can be further analyzed.

6   Although a whole genome sequence was recently completed, research on the functionality
7 of individual genes under the stress of P. Litchii is scarce. In this work, we acquired a genome-
8 wide transcriptomic data for two cultivars of Litchi with and without P. Litchii treatments from the
9 NCBI Gene Expression Omnibus (GEO) database. We then used Deseq2 library to identify and
10 analyze differentially expressed genes. In addition, we studied the top 18 differentially
11 expressed genes by locating them on the chromosome, describing their gene structure and
12 finding out a number of conserved motifs. Furthermore, we investigated the promoter area of
13 those genes (2000 bp upstream) and identified 7 cis-regulatory elements such as ABRE, ARE,
14 MeJA, TCA that are related to disease response and defense. This will aid in developing new
15 lychee cultivars that can better resist the infection of P. Litchii.

16

17 **INTRODUCTION**

18   Lychee (Litchi Chinensis Sonn.), also known as the alligator strawberry, is a delicious and
19 nutritious tropical fruit native to South China, Malaysia and northern Vietnam (1). It has been
20 widely cultivated in more than 20 countries worldwide. However, many diseases can affect the
21 production of this fruit. It is known that the most destructive disease called *Litchi downy blight* is
22 caused by Peronophythora litchi (P.Litchi), which leads to water brown spots (2)(3).

23   A recent effort of investigating the nature history of Lychee domestication produced the
24 whole genome wide sequencing for one of the cultivars (4)(5). However, most of the lychee
25 genes related to the response to P.Litchii infection are still not well understood. Hailun Liu et. al.
26 studied a group of LcCDPK genes and show they can express differently when exposed to

27 P.Litchii infection (6). Another study done by Peng Li, et. al. showed that a RXLR effector called
28 PlAvh202 is the main contributor to the virulence of P.Litchii (7).

29   J. Sun et. el. did a study on Lychee's response mechanism when infected by P. Litchii for
30 two Lychee cultivars. They showed how the two cultivars respond differently during the early
31 stage of infection, and arrived at the conclusion that quickly recognition and early responses to
32 pathogen is key to achieve better resistance to P. Litchii infection (8).

33   In this work, we aim to identify biomarkers that differentiate expressed genes for Lychee with
34 or without the inoculation of P. Litchii. Furthermore, we analyze the functionality of those genes
35 in several different ways. The outcome of this work might lead to the finding of more disease
36 resistant cultivars for Lychee.

37   A summary of contributions from our study:

38   1. We conducted differential gene expression analysis for two Lychee cultivars and
39   identified top differentially expressed genes with regard to P.Litchii inoculation.

40   2. We found that a lot more genes are highly expressed for the Yurong cultivar.

41   3. We analyzed top differentially expressed genes, found out their gene structure and
42   location on the Lychee chromosome.

43   4. We identified a number of conserved motifs out of those top genes.

44   5. Lastly, we examined cis-acting regulatory elements in the promoter area of the top
45   differentially expressed genes (2000 bp upstream) and identified 7 elements that are
46   related to disease and stress responses.

47

48 **RESULTS**

49   In order to study how Lychee responds to the Influence of Peronophythora Litchii in terms of
50 gene expression, it took several steps to identify top differentially expressed genes and study
51 their characteristics. During the first step, we run Deseq2 (9) to process dataset GSE201243
52 from NCBI's Gene Expression Omnibus (10). The dataset contains expression counts for 34K
53 genes with comparison between those treated with P.Litchii and those that are not. There are
54 two cultivars under study: 'Guiwei' and 'Yurong'. With each experiment repeated three times.
55 The output of the Deseq2 analysis is a data frame with each row being a gene and columns are
56 adjusted Pvalue, Pvalue and Log of Folding Change.

57   -log(Pvalue) and Log2 of fold changes are used to drawn as a volcano graph as shown in
58 **Figure 1**. The cutoff of the adjusted Pvalue is set to 0.05, so genes that are expressed
59 significantly differently are on the upper side of the graph, either marked as blue for
60 downregulated genes or red for upregulated genes. The top figure is for 'Guiwei' and the bottom

figure is for 'Yurong'. Based on the Pvalue threshold, we found 15 upregulated genes and 2 downregulated genes for 'Guiwei' and 39 upregulated genes and 78 downregulated genes for 'Yurong'. Most genes with high folding change are in the middle of that log range though. The 236 upregulated genes sit on the top of the figure while the 14 downregulated ones sit at the bottom. Most genes with significant folding changes are upregulated genes.
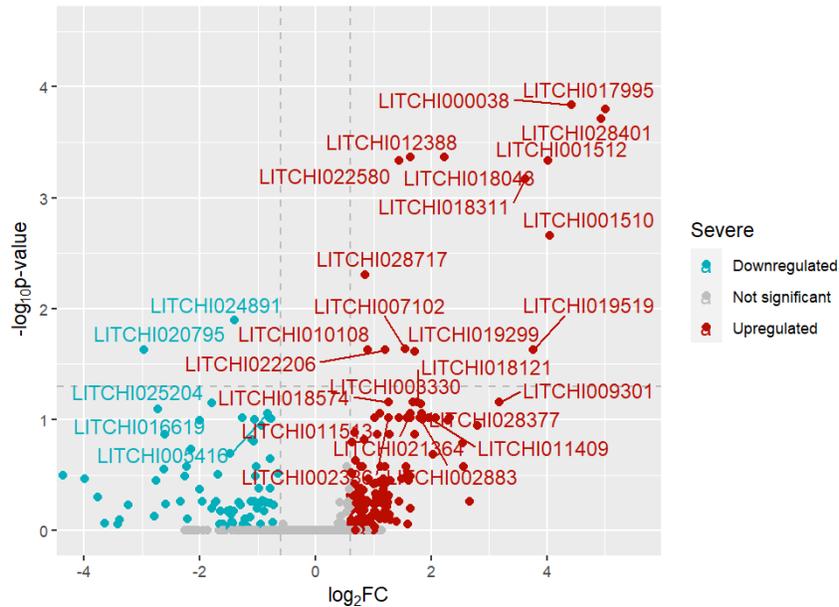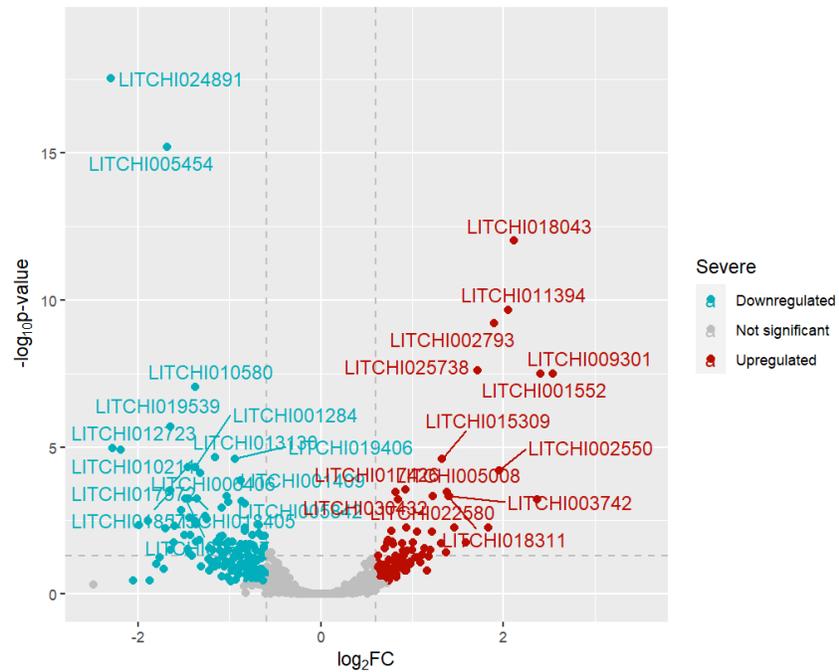

Figure 1. Volcano plot indicating top differentially expressed genes for GSE28674. Top: Guiwei, Bottom: Yurong, Red = upregulated, blue = downregulated.

70

71   We picked 18 genes with high folding changes from the two cultivars for further study. Next,

72 we compared the 18 genes in terms of folding changes across the two cultivars. **Figure 2** shows

73 a heatmap comparing those genes. The clustering shows that they can be categorized into

74 three types: high logFC for both cultivars, high logFC for Guiwei only and relatively low logFC

75 for both cultivars.
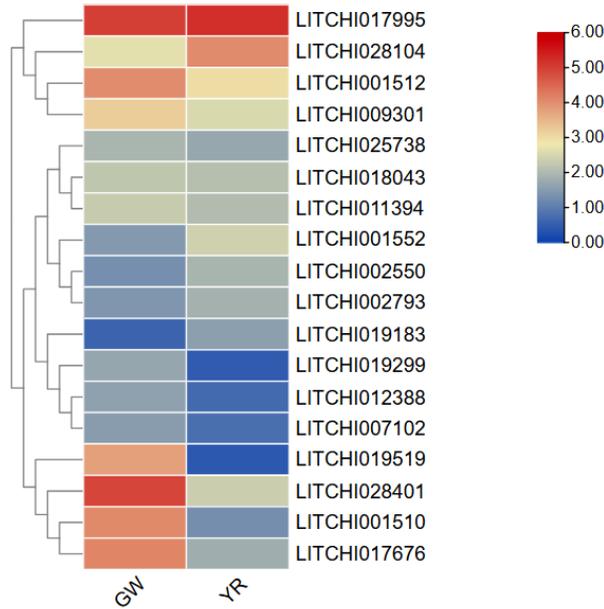
76
77 **Figure 2. Heatmap comparing top differentially expressed genes for Guiwei and Yurong**.
78 Color indicates different Log2FC values.

79
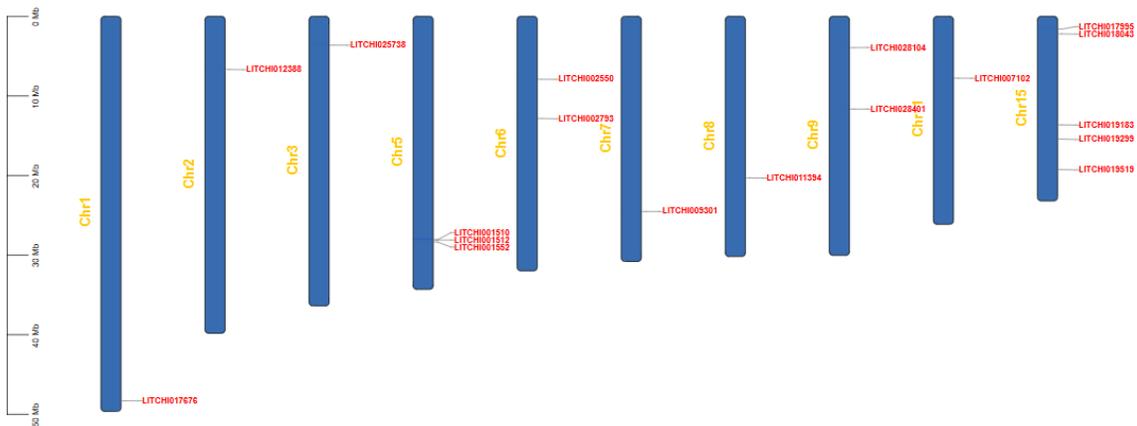
80
81 **Figure 3. Location of the top differentially expressed genes on the chromosome**.

4

82   Next, we show the position of the top differentially expressed genes on the chromosomes. It

83   can be seen from **Figure 3** that the 18 genes spread across 10 chromosomes (out of a total of

84   15 chromosomes for Lychee). Most of the chromosomes get at most 1 gene, but there are 5

85   genes on chromosome 15 and 3 on chromosome 5.

86

87
88
89   **Figure 4. Gene structure of top differentially expressed genes**.

90   Since most genes of Lychee are not well understood, we intend to further analyze their

91   internal structure. On **Figure 4**, we illustrate the structure of the selected genes. The yellow

92   segments are exons and the sold lines are introns. The number of exons varies greatly across

93   different genes. For example, LITCHI025738 contains 10 exons, whereas LITCHI007102 only

94   contains 2 exons. Another interesting gene is LITCHI001510, it is the longest among the 18

95   genes, but most part of the gene are introns. On the other hand, LITCH025738, LITCH012388

96   and LITCH028401 contain many small exons.

97   To further investigate the structure of these genes, we did an analysis on conserved motifs.

98   As shown in **Figure 5**, we were able to identify 20 motifs. Their distribution is also shown on the

99   figure. Most genes contain only a small number of motifs, except for LITCHI009301,

100 LITCHI002793 and LITCHI001510. Among all the motifs, Motif2, Motif5 and Motif7 are part of
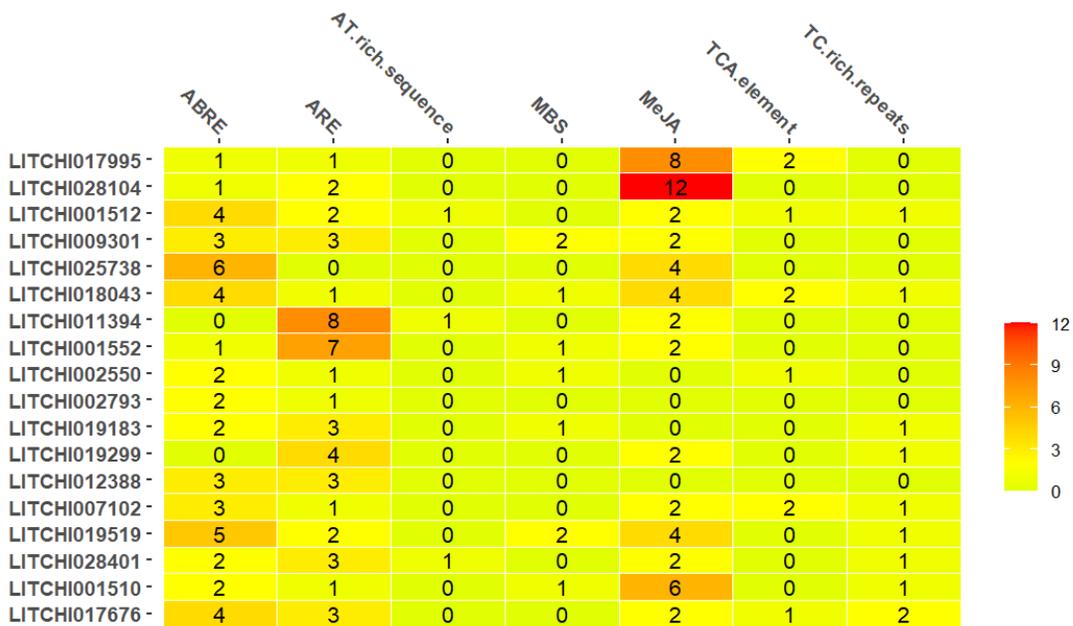101 the Thaumatin family.
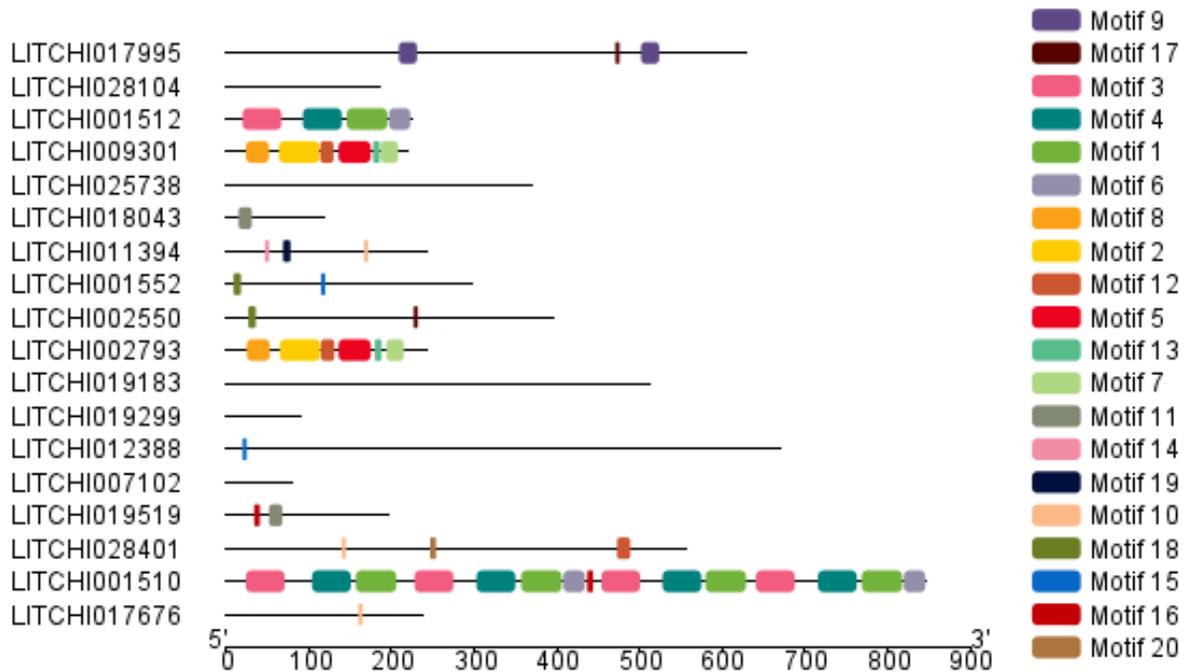
102
103 **Figure 5. Distribution of conserved motifs on top differentially expressed genes**.

104   Cis-acting regulatory elements encode the genomic functionality that leads to the proper
105 spatiotemporal patterning of expressed genes necessary for appropriate responses to the
106 environment (11).

107   Functionalities such as abscisic acid response, elicitor-mediated activation, MeJA responses
108 and salicylic acid responsiveness are essentially for plant development and stress response.
109 For the top differentially expressed genes identified earlier, we performed an analysis to search
110 for Cis-acting regulatory elements in their promoter area. Many Cis-acting regulatory elements
111 were found. In **Figure 6**, we looked at elements that might be related to Lychee's response to
112 P.Litchii. A total of seven cis-acting regulatory elements were investigated: TCA-element, ABRE,
113 ARE, MeJA, TC-rich repeats, AT-rich sequence and MBS. On the figure, we can see that those
114 elements spread across most of the genes except for LITCHI 002793, which is a relatively short
115 gene. Further on **Figure 7**, we tabularize the number of occurrences of each cis-acting
116 regulatory element on each top gene.

6
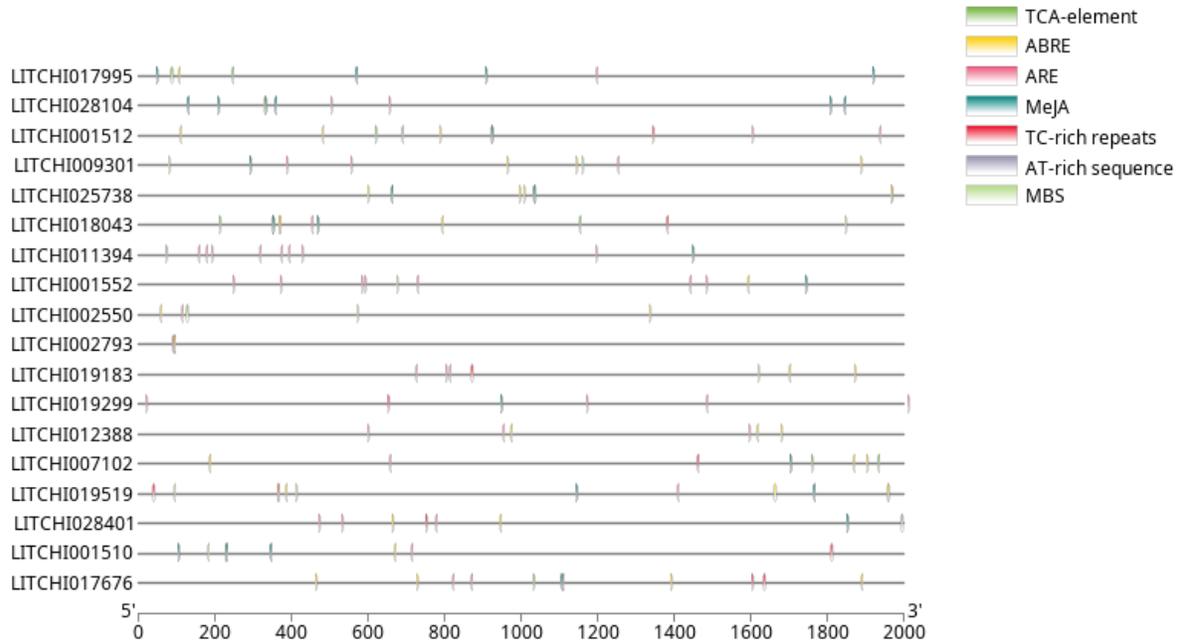
117
118 **Figure 6. The locations of cis-acting regulatory elements in promoters of top differentially**
119 **expressed genes**.

	ABRE	ARE	AT.rich.sequence	MBS	MeJA	TCA.element	TC.rich.repeats
LITCHI017995	1	1	0	0	8	2	0
LITCHI028104	1	2	0	0	12	0	0
LITCHI001512	4	2	1	0	2	1	1
LITCHI009301	3	3	0	2	2	0	0
LITCHI025738	6	0	0	0	4	0	0
LITCHI018043	4	1	0	1	4	2	1
LITCHI011394	0	8	1	0	2	0	0
LITCHI001552	1	7	0	1	2	0	0
LITCHI002550	2	1	0	1	0	1	0
LITCHI002793	2	1	0	0	0	0	0
LITCHI019183	2	3	0	1	0	0	1
LITCHI019299	0	4	0	0	2	0	1
LITCHI012388	3	3	0	0	0	0	0
LITCHI007102	3	1	0	0	2	2	1
LITCHI019519	5	2	0	2	4	0	1
LITCHI028401	2	3	1	0	2	0	1
LITCHI001510	2	1	0	1	6	0	1
LITCHI017676	4	3	0	0	2	1	2

120
121
122 **Figure 7. Types and numbers of cis-acting regulatory elements in promoters of top**
123 **differentially expressed genes.**

124

125   It is noteworthy that ABRE occurs frequently for LITCHI025738 and LITCHI019519; ARE

126   occurs frequently for LITCHI011394 and LITCHI001552; MeJA occurs a lot for LITHI017995,

127   LITCHI028104 and LITCHI001510. For the rest cis-acting regulatory elements (AT-rich

128   sequence, MBS, TCA-element and TC-rich repeats), they only show up sporadically on the top

129   differentially expressed genes.

130

131   **DISCUSSION**

132   **Figure 1** shows that there are a lot of more differentially expressed genes for Yurong, which

133   indicates that this cultivar responds quite differently from Guiwei. Hailun et. el. also found out

134   that several CDPK family genes were highly expressed after inoculation with P.Litchii and noted

135   that Yurong is a cultivar with high resistance to P.Litchii (6). This study provides more insights

136   into how Yurong responds differently to Lychee downy blight. A much larger number of

137   differentially expressed genes can indicate more complicated response mechanism for the

138   Yurong cultivar.

139   **Figure 3** and **Figure 4** show the structure and location of top differentially expressed genes.

140   Furthermore, we list their functionalities in **Table 1**. Proteins are categorized into "stress

141   response", calcium binding, enzyme and other.

142   LITCHI028104 is determined to be a Jacalin like Lectin, which is potentially an antiviral

143   protein (12). In the meantime, LITCHI011394 is considered part of a pathogen-related family.

144   Next, we have LITCHI012388, which belongs to the heat-shock protein 70 (HSP70) family,

145   some of which are only expressed under stress conditions (13). Furthermore, both

146   LITCHI001552 and LITCHI025738 are stress-related. They are cyclase-like plant proteins,

147   which are involved in responses to abiotic stress (14).

148   Both LITCHI009301 and LITCHI002793 are Osmotin/Thaumatin like proteins. Thaumatin is

149   an extremely sweet-tasting protein found in berries from Thaumatococcus daniellii, a tropical

150   flowering plant (15). If attacked by viroids, a single-stranded unencapsulated RNA molecule,

151   Thaumatin gets induced. Other the other hand, Osmotin is part of the PR-5 protein family whose

152   members are homologous to Thaumatin. Osmotin and other PR-5 proteins were known to have

153   antifungal activity against a broad range of fungi, including several plant pathogens (16).

154   The next two are: LITCHI017995, a Metalloendoproteinase, that is zinc-dependent, calcium-

155   activated protease; and LITCHI019519, which involved in EF-hand Calcium binding. Both genes

156   could be part of a pathway related to calcium binding.

157   We also observe two enzymes among the top differentially expressed genes. LITCHI002550,

158   belongs to the cytochrome P450 family. It is involved in many biosynthetic pathways and

159　catalyzes a range of reactions. LITCHI019299, LITCHI017676 are S-adenosylmethionine
160　related. It is a key regulatory enzyme in the biosynthesis of polyamines.

161　　For the remaining ones, LITCHI007102 is a plant specific phytosulfokine precursor proteins.
162　Phytosulfokines (PSK) belongs to the group of plant peptide growth factors, while
163　LITCHI019183 and LITCHI018043 are plant lipid transfer proteins.

164　　Note that, not all top expressed genes are well-understood. For example, LITCHI001510
165　and LITCHI001512 are not characterized, pending future study.

166

Protein	Functionality	Category
LITCHI028104	Jacalin-like lectin	stress response
LITCHI011394	pathogen-related protein family	stress response
LITCHI012388	Heat shock protein	stress response
LITCHI001552, LITCHI025738	cyclase-like protein	stress response
LITCHI009301, LITCHI002793	Osmotin, Thaumatin like	stress response
LITCHI019519	EF-hand Calcium binding	calcium binding
LITCHI017995	Metalloendoproteinase	calcium binding
LITCHI002550	cytochrome P450	enzyme
LITCHI019299, LITCHI017676	S-adenosylmethionine related	enzyme
LITCHI007102	phytosulfokines	other
LITCHI019183, LITCHI018043	plant lipid transfer protein	other

167　**Table 1. Functionality of top differentially expressed genes.**

168　　For the motifs show in **Figure 5**, Thaumatin family Motifs are also present. This family of
169　genes are known as pathogenesis-related group 5 (PR5). Such proteins are involved in
170　systemically acquired resistance and stress response. Many Thaumatin-like proteins
171　accumulate in plant when responding to infection by a pathogen and performing antifungal
172　activity (17).

173　　**Figure 6** and **Figure 7** show the seven cis-acting regulatory elements related to plant's
174　disease fighting. To better understand the functionality of such elements, we list their
175　functionalities in **Table 2**.

176　　First, ABRE element are ABA-responsive elements involved in abscisic acid responsiveness
177　(18)(19). Abscisic acid (ABA) plays a crucial role in plant's reaction to environmental stress
178　during vegetative growth. ABA-mediated responses induce expression of a large number of
179　genes, which is mediated by ABREs (20). In this study, we are mostly seeing ABREs in the form
180　of ACGTG.

181   ARE and MBS are two other cis-regulatory elements. Both respond to environmental
182 changes under various conditions, although they are not directly linked to Lychee downy blight
183 pathogen. We are seeing AREs in the form of AAACCA sequence and MBS in the form of
184 CAACTG sequence.

185   In the meantime, AT-rich sequence TAAAATACT were found in a few top differentially
186 expressed genes. They are elements for maximal elicitor-mediated activation. Elicitors are
187 compounds that can activate chemical defense in plants (21).

188   Methyl Jasmonate (MeJA) is a type of signaling molecule that helps plant development and
189 stress response. It can activate plant defense mechanism for a wide range of pathogens. For
190 example, Y. Zhang et. el. found out that MeJA-responsive proteins played a critical role in
191 Maize's response to pests and insects (22). For several top differentially expressed genes in
192 this study, we observed high concentration of MeJA cis-regulatory elements in their promoter
193 area, typically in the form of CGTCA or TGACG.

194   TCA-elements are involved in salicylic acid responsiveness. Note that hypersensitive
195 reaction is known to be away to prevent pathogen spreading by programmed cell death.
196 Hypersensitive reaction cab triggers systemic acquired resistance (SAR) by salicylic acid
197 mediated defense and provide broad-spectrum immunity to infection [from the early response
198 paper]

199   TC-rich repeats are known to participate in plant defense and stress responsiveness (23). In
200 this study, we are seeing TC-rich repeats in the form of GTTTTCTTAC.

201

cis-regulatory element	functionality
ABRE	cis-acting element involved in the abscisic acid responsiveness
ARE	cis-acting regulatory element essential for the anaerobic induction
AT-rich sequence	element for maximal elicitor-mediated activation
MBS	MYB binding site involved in drought-inducibility
MeJA	cis-acting regulatory element involved in the MeJA-responsiveness
TCA-element	cis-acting element involved in salicylic acid responsiveness
TC-rich repeats	cis-acting element involved in defense and stress responsiveness

202 **Table 2. Functionality of identified cis-regulatory elements in the promoters of top**
203 **differentially expressed genes.**

204   Note that, the cis-acting regulatory elements found in this study are commonly observed to
205 expressed as a mechanism for stress responses. For example, Y. Qian et. el. discovered a
206 similar set of cis-regulatory elements in maize (24).

207   In conclusion, we found many differentially expressed genes due to P.Litchii inoculation. We
208 then analyzed those genes, and found that many of them are related to stress responses. In

209 addition, we identified Thaumatin family Motifs out of those top genes, which is pathogenesis-
210 related. Lastly, we examined cis-acting regulatory elements in the promoter area of the top
211 differentially expressed genes (2000 bp upstream) and identified 7 elements that are related to
212 disease and stress responses.

213

214 **MATERIALS AND METHODS**

215   Gene expression data for Litchi Chinensis is scarce. We searched the NCBI Gene
216 Expression Omnibus (GEO) and found two sources of data: whole genome sequence from (5
217 and genome-wide transcriptomic data was obtained by high throughput sequencing for two
218 cultivars 'Guiwei' and 'Yurong' either inoculated with P.Litchii or not. This is from dataset
219 GSE201243 (10). The treated leaves were collected at 24 hpi.

220   Index of the reference genome was built using Bowtie v2.2.3 and paired-end clean reads
221 were aligned to the reference genome using TopHat v2.0.12. HTSeq was used to count reads
222 numbers mapped to each gene. And then FPKM of each gene was calculated based on the
223 length of the gene and reads count mapped to this gene.

224   Next, differential expression analysis of two conditions/groups was performed using the
225 DESeq2 R package and produce a list of differentially expressed genes. With a Pvalue
226 threshold of 0.05, there are 15 upregulated genes and 2 downregulated genes for Guiwei and
227 39 upregulated genes and 78 downregulated genes for Yurong. We then take the top 18 in
228 terms of folding change out of the two lists for further investigation.

229   TBtools (25) were used to draw the heatmap, extract gene coordinates from GFF3 file and
230 show their position on the chromosomes and their gene structure. In order to identify conserved
231 motifs, we first convert nucleotide sequences to protein sequences using TBtools. The resulting
232 fasta file is then fed to the MEME Motif-based sequence analysis tool (26), which produces an
233 XML file that can be visualized using TBtools.

234   To classify protein and motifs, we used PFAM (27) by passing in a list of motifs and top
235 differentially expressed genes as a fasta file.

236   In order to analyze Cis-acting regulatory elements, we used TBtools to extract 2000 bp
237 before each selected gene. This sequence is then input into PlantCare (28) to identify cis-acting
238 regulatory elements. The next step is to take the tab files from PlantCare, which is a fasta
239 format file with all identified cis-acting regulatory elements, filter out irrelevant elements and and
240 feed the rest into TBtools to draw the location of those elements. Lastly, relevant elements are
241 counted for each top differentially expressed gene and we used R's dplyr and ggplot2 library to
242 draw the graph with color gradient.

243

REFERENCES

1. Morton, J.F. "Lychee." *Fruits of Warm Climates.* Oct 2013, pp. 249–259.

2. Sun, J. et. al. "Transcriptome analysis of Phytophthora litchii reveals pathogenicity arsenals and confirms taxonomic status." *PLoS One.* 2017.

3. Yi, C. et al. "ATP-regulation of antioxidant properties and phenolics in litchi fruit during browning and pathogen infection process." *Food Chemistry.* Vol. 118, no. 1 Jan. 2010, pp. 42–47.

4. Hu, G. et. al. "Two divergent haplotypes from a highly heterozygous lychee genome suggest independent domestication events for early and late-maturing cultivars." *Nature Genetics.* Vol. 54, no. 1, Jan. 2022，pp. 3-83.

5. "Feizixiao whole genome." *National Center for Biotechnology Information, U.S. National Library of Medicine*, www.ncbi.nlm.nih.gov/bioproject/PRJNA747875/. Accessed Dec. 2023.

6. Liu, H. et. al. "Identification of LcCDPKs and analysis of their expression patterns in response to downy mildew stresses in lychee." *Journal of Fruit Science.* Vol. 3, Mar. 2023, pp. 442-456.

7. Li, P., et. al. "Peronophythora litchii RXLR effector P. litchii avirulence homolog 202 destabilizes a host ethylene biosynthesis enzyme." *Plant Physiology.* Vol. 193, no. 1, September 2023, pp 756–774.

8. Sun, J., et. al. "Early responses given distinct tactics to infection of Peronophythora litchii in susceptible and resistant litchi cultivar." *Scientific Reports.* No. 2810, Feb. 2019.

9. Love, M., et. al. "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." *Genome Biology.* Vol. 15, no. 12, 2014.

10. "Geo Accession Viewer." National Center for Biotechnology Information, U.S. National Library of Medicine, www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE201243

11. Marand, A., et. al. "cis-Regulatory Elements in Plant Development, Adaptation, and Evolution." *Annual Review of Plant Biology.* Vol. 74, May 2023, pp.111-137.

275   12. Ziolkowska, N. "Domain-swapped structure of the potent antiviral protein griffithsin and its
276   mode of carbohydrate binding." *Structure*. Vol.14, no. 7, Aug. 2006, pp. 1127-1135.

277   13. Snutch, T., et. al. "The Caenorhabditis elegans hsp70 gene family: a molecular genetic
278   characterization." *Gene*. Vol. 64, No. 2, Apr. 1988, pp. 241-255.

279   14. Qin, Y. et. al. "Characterization of a novel cyclase-like gene family involved in controlling
280   stress tolerance in rice." *Plant Physiology*. Vol 181, 2015, pp. 30-41.

281   15. Edens, L. et. al. "Cloning of cDNA encoding the sweet-tasting plant protein thaumatin and its
282   expression in Escherichia coli." *Gene*. Vol. 18, No. 1, Apr. 1982, pp. 1-12.

283   16. Min, K. et. al. "Crystal structure of osmotin, a plant antifungal protein." Proteins 54, 170-3,
284   (2004)

285   17. Ruiz-Medrano, R. et. al. "Nucleotide sequence of an osmotin-like cDNA induced in tomato
286   during viroid infection." *Plant Molecular Biology*. Vol. 20, no. 6, Dec. 1992, pp.1199-1202.

287   18. Hattori, T., et. al. "Experimentally determined sequence requirement of ACGT-containing
288   abscisic acid response element." *Plant & Cell Physiology*. Vol. 43, no. 1, Jan. 2002 pp.136-
289   140.

290   19. Fujita, Y., et. al. "ABA-mediated transcriptional regulation in response to osmotic stress in
291   plants." *Journal of Plant Research*, Vol. 124, No. 4, Mar. 2011, pp. 509-522.

292   20. Ingram, J. and Bartels, D. "The molecular basis of dehydration tolerance in plants." *Annual
293   review of plant biology*. Vol. 47, Jun. 1996, pp. 377-403.

294   21. Seki, H., et. al. "Combined Effects of Multiple m-Acting Elements in Elicitor-Mediated
295   Activation of PSCHS1 Gene", *Plant and Cell Physiology*. Vol. 38, No. 1, Jan. 1997, pp. 96-
296   100.

297   22. Zhang, Y., et. al. "Maize MeJA-responsive proteins identified by high-resolution 2-DE
298   PAGE." *Data Brief*. Vol. 5, Sep. 2015, pp.129-133.

299   23. Banerjee, J., et. al. "An intergenic region shared by At4g35985 and At4g35987
300   in Arabidopsis thaliana is a tissue specific and stress inducible bidirectional promoter
301   analyzed in transgenic Arabidopsis and tobacco plants." *PLoS One*. Vol. 8, No. 11, Nov.
302   2013.

303   24. Qian, Y., et. al. "Genome-wide identification, classification and expression analysis of the
304   JmjC domain-containing histone demethylase gene family in maize." *BMC Genomics*. Vol.
305   256, Apr. 2019.

306   25. Chen, C., et. al. "TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big
307   Biological Data." *Molecular Plant*. Vol. 13, No. 8, Mar. 2020, pp. 1194-1202.

308 26. Bailey, T. and Elkan, C. "Fitting a mixture model by expectation maximization to discover

309   motifs in biopolymers." *Proceedings of the Second International Conference on Intelligent*

310   *Systems for Molecular Biology*. Vol. 2, 1994, pp. 28-36. meme-suite.org/tools/meme.

311 27. "Pfam: database of a large number of proteins." pfam.xfam.org/interpro. Accessed Nov.

312   2023.

313 28. "Plantcare: cis-acting regulatory element." bioinformatics.psb.ugent.be/webtools/plantcare/

314   html. Accessed Nov. 2023.