

**Study on long non-coding RNAs expressed  
during *Helopeltis theivora* infestation in tea  
plant [*Camellia sinensis* (L.) O. Kuntze]**

**A THESIS SUBMITTED FOR THE AWARD OF THE DEGREE OF  
DOCTOR OF PHILOSOPHY (Ph.D.) IN BOTANY OF GAUHATI  
UNIVERSITY**

**By**

**KUNTALA SARMA BORDOLOI**

**Ph.D. Enrolment No. Bot-16/18**

**GU Registration No. 045755 of 2011-12**



**DEPARTMENT OF BOTANY**

**FACULTY OF SCIENCE**

**GAUHATI UNIVERSITY**

**ASSAM :: INDIA**

**AUGUST, 2023**

I would love to dedicate this thesis to my parents and my brother...

## Declaration

I hereby declare that this thesis is the result of my own research work which has been carried out under the guidance of Dr. Niraj Agarwala, Assistant Professor, Department of Botany of Gauhati University. I further declare that this thesis as a whole or any part thereof has not been submitted to any university (or institute) for the award of any degree or diploma.

This thesis contains less than 90,000 (ninety thousand) words excluding bibliography and captions.

Date: 28/08/2023

Place: Guwahati

*Kuntala Sarma Bordoloi*  
Kuntala Sarma Bordoloi



**DEPARTMENT OF BOTANY  
GAUHATI UNIVERSITY**

**Gopinath Bordoloi Nagar, Guwahati – 781014, Assam, India**

**Dr. Niraj Agarwala, M.Sc., Ph.D.**  
Assistant Professor

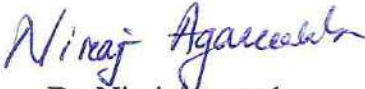
☎ : +918486027249(M), +917002284437(M)  
E-mail: niru.niraj.s@gmail.com  
niraj\_botany@gauhati.ac.in

Date: 25/08/2023

**Certificate**

This is to certify that the thesis titled "**Study on long non-coding RNAs expressed during *Helopeltis theivora* infestation in tea plant [*Camellia sinensis* (L.) O. Kuntze]**" is the result of research work of Kuntala Sarma Bordoloi, carried under my supervision, submitted to Gauhati University for the award of the degree of Doctor of Philosophy in Botany.

This thesis conforms to the standard of PhD thesis under Gauhati University including the standard related to plagiarism and has a similarity index not more than 10% (ten percent), excluding the bibliography.

  
Dr. Niraj Agarwala  
Supervisor

## Acknowledgement

First and foremost, with an immense sense of gratitude, I would like to offer my sincere thanks to my supervisor, Dr. Niraj Agarwala, for his constant guidance, motivation and constructive criticism during my research. He has been an integral source of my commitment and knowledge towards this field of interest and has intellectually contributed to my research work and its findings.

I would like to extend my deepest gratefulness to the Head of the Department of Botany, Gauhati University, for providing all the necessary facilities and for ensuring smooth conduct of my research work.

With immense pleasure I offer my sincere thanks to my RAC members and all the faculties and staff members of the Department of Botany, GU. I would also like to thank my fellow research scholars for their constant love and support.

My heartfelt thanks goes to my loving and supporting lab members Debasish B Krishnatreya, Bhaskar Dowarah and Pooja Moni Baruah for their constant help, support and criticism.

My sincere gratitude to all the authorities of Darjeeling Tea Research and Development Centre, Kurseong, West Bengal, authorities of Sungma and Turzum Tea Estate, West Bengal and Mrinmoy Purkayastha, Sonapur Tea Estate, for extending their helping hands for smooth conduct of my research. My utmost thanks to Genotypic, Bangalore, for carrying out the high-throughput Illumina RNA-sequencing. I would like to thank Preetom Regon and Mehzabin Rehman for extending their support during setting up of qPCR experiment.

And last but not the least, I offer my sincere gratitude to the Almighty, my beloved parents and my brother, without whose love and blessings, this research work wouldn't have been possible. I express my heartfelt thanks to them for helping me through thick and thin.

Date: 28/08/2023

*Kuntala Sarma Bordoloi*  
Kuntala Sarma Bordoloi

Place: Guwahati

## Abstract

*Helopeltis theivora* or the tea mosquito bug (TMB) is reportedly one of the most devastating pests of tea plant (*Camellia sinensis*) causing threat to the beverage crop. This study revealed a significant alternation of tea plant's biochemistry in terms of phenolics and flavonoids content, and increase in antioxidative enzymatic activities in response to TMB herbivory. Long non-coding RNAs (lncRNAs) constitute a group of endogenous RNAs that play gene regulatory roles in eukaryotes. In the present study, 9502 lncRNAs were identified from healthy and TMB-infested *C. sinensis* tissues using high-throughput RNA sequencing data, out of which 80 lncRNAs got differentially expressed in response to TMB infestation. Determination of genes that could act as potential targets of lncRNAs revealed that the identified lncRNAs could possibly target as many as 5804 genes. Differential gene expression (DGE) analysis led to the identification of 3665 differentially expressed genes (DEGs), of which, expression of 1767 genes got upregulated and 1898 genes got downregulated during tea plant's response to TMB. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis of DEGs and lncRNA-target genes have shown that TMB infestation might have triggered transcriptomic reprogramming leading to altered primary and secondary metabolism in *C. sinensis*. Identification of TMB-responsive circRNAs (another class of non-coding RNA) in tea plant revealed differential expression of 34 circRNAs during TMB feeding. These circRNAs are found to be associated with biological pathways like "tryptophan metabolism", "biosynthesis of secondary metabolites", "porphyrin metabolism", "nucleocytoplasmic transport", "N-glycan biosynthesis" etc. LncRNAs and circRNAs can act as competing endogenous RNAs (ceRNAs) to bind with common microRNA (miRNA) response elements (MREs) involving a competition between mRNAs, lncRNAs and circRNAs. Assessment of this "sponging" activity of lncRNAs and circRNAs led to the construction of the ceRNA network consisting of 5 lncRNAs, 17 circRNAs, 33 mRNAs competing against each other to bind with 37 miRNAs. The expression of 6 DEGs and 5 differentially expressed lncRNAs (DELs) have been validated by qRT-PCR.

# Table of contents

<b>List of figures</b>	<b>xii</b>
<b>List of tables</b>	<b>xix</b>
<b>1 INTRODUCTION</b>	<b>1</b>
<b>2 REVIEW OF LITERATURE</b>	<b>9</b>
2.1 History, taxonomy, cultivation of tea and its challenges	9
2.2 Plant-insect interaction – an overview	11
2.3 Transcription factors involved in plant-insect interaction	16
2.4 Taxonomy, life cycle of <i>H. theivora</i> and its infestation on tea	17
2.5 Non-coding RNAs – their potential in plant developmental processes and stress response	20
2.5.1 miRNA	21
2.5.2 siRNA	23
2.5.3 circRNA	24
2.5.4 Long non-coding RNAs – undeniable mediators of plant biotic stress response	25
<b>3 MATERIALS AND METHODS</b>	<b>29</b>
3.1 Sample harvesting for biochemical analysis	29
3.2 Total phenolics estimation	29
3.3 Totals flavonoids estimation	30
3.4 Antioxidative enzymatic assays	30
3.4.1 POX	30

3.4.2 APX	31
3.4.3 PAL	31
3.4.4 PPO	32
3.4.5 CAT	32
3.5 Plant materials and stress treatments	32
3.6 RNA isolation, library construction and Illumina sequencing	33
3.7 LncRNAs	34
3.7.1 Bioinformatics pipeline for identification of lncRNAs and differentially expressed lncRNAs (DELs)	34
3.7.2 LncRNA target gene prediction	35
3.7.3 Functional annotation and gene set enrichment analysis (GSEA) of lncRNA targets	35
3.7.4 Quantitative real-time PCR of selected DELs	36
3.8 Genes	39
3.8.1 Identification of differentially expressed genes (DEGs)	39
3.8.2 Functional annotation and gene set enrichment analysis (GSEA) of DEGs	39
3.8.3 Quantitative real-time PCR of selected DEGs	39
3.9 CircRNAs	41
3.9.1 Pipeline for identification of circRNAs and DECs	41
3.9.2 Expression correlation analysis of DECs and DEGs	42
3.9.3 Functional annotation of DECs	42
3.10 Collection of miRNA data	44
3.11 miRNA target prediction	44



3.12 eTM prediction . . . . .	44
3.13 Network construction and visualization . . . . .	44
<b>4 RESULTS</b>	<b>45</b>
4.1 Biochemical changes in the tea plant in response to TMB-herbivory . . . . .	45
4.1.1 Estimation of total phenolics . . . . .	45
4.1.2 Estimation of total flavonoids . . . . .	46
4.1.3 Enzymatic antioxidants . . . . .	46
4.1.3.1 POX . . . . .	47
4.1.3.2 APX . . . . .	47
4.1.3.3 PAL . . . . .	48
4.1.3.4 PPO . . . . .	49
4.1.3.5 CAT . . . . .	50
4.2 Sequencing reads summary . . . . .	51
4.3 LncRNAs . . . . .	52
4.3.1 Identification of lncRNAs . . . . .	52
4.3.2 Characterization of lncRNAs . . . . .	54
4.3.3 Expression of lncRNAs . . . . .	56
4.3.4 Differential expression of lncRNAs . . . . .	57
4.3.5 Identification of lncRNA-target genes . . . . .	61
4.3.6 Functional annotation and enrichment analysis of lncRNAs . . . . .	63
4.3.7 Quantitative real-time PCR of selected DELs . . . . .	70
4.4 Genes . . . . .	72
4.4.1 Expression of genes . . . . .	72
4.4.2 Differential expression of genes . . . . .	74

4.4.3	Functional annotation and enrichment analysis of DEGs . . . . .	75
4.4.4	Quantitative real-time PCR of selected DEGs . . . . .	80
4.5	Comparison between lncRNAs and mRNAs identified in this study . . . . .	82
4.6	CircRNAs . . . . .	84
4.6.1	Identification and characterization of circRNAs . . . . .	84
4.6.2	The differential expression pattern of identified circRNAs . . . . .	86
4.6.3	Functional characterization of DECs . . . . .	88
4.7	miRNA targets, eTMs . . . . .	91
4.8	ceRNA network . . . . .	118
<b>5</b>	<b>DISCUSSION</b>	<b>120</b>
5.1	Biochemical changes in the tea plant as a result of TMB feeding . . . . .	120
5.2	Change in expression pattern of genes/lncRNAs involved in primary metabolism . . . . .	121
5.2.1	DEGs related to primary metabolism of <i>C. sinensis</i> in response to TMB . . . . .	121
5.2.2	lncRNA mediated regulation of genes related to primary metabolism of <i>C. sinensis</i> in response to TMB . . . . .	124
5.3	Change in expression pattern of genes/lncRNAs involved in secondary metabolism . . . . .	126
5.3.1	DEGs related to secondary metabolism of <i>C. sinensis</i> in response to TMB . . . . .	126
5.3.2	lncRNA mediated regulation of genes related to secondary metabolism of <i>C. sinensis</i> in response to TMB . . . . .	128
5.4	Interaction between DEGs and their involvement in some important	

defense responsive biological pathways . . . . .	129
5.5 Involvement of circRNAs in tea-TMB interaction . . . . .	133
<b>SUMMARY AND CONCLUSION</b>	<b>136</b>
<b>REFERENCES</b>	<b>141</b>
<b>APPENDIX</b>	<b>177</b>

## List of Figures

Figure No.	Title	Page No.
1.1	Photographs of <i>H. theivora</i> infested tea plants	4
3.1	Healthy and <i>Helopeltis theivora</i> infested tea leaves. (a) Healthy tea leaf and bud (b) Development of symptoms after feeding (c, d) <i>H. theivora</i> feeding on leaves	33
3.2	A flowchart depicting pipeline employed for identification and characterization of lncRNAs in healthy and TMB-infested tea plants. The peach ellipses denote softwares/tools used in each corresponding step.	38
3.3	A flowchart depicting pipeline employed for identification of differentially expressed genes (DEGs) in healthy and TMB-infested tea plants. The peach ellipses denote softwares/tools used in each corresponding step.	41
3.4	A flowchart depicting pipeline employed for identification of differentially expressed circRNAs (DECs) in healthy and TMB-infested tea plants.	43
4.1	Total phenolic content in 11 clones of tea plant. (Data is represented as mean $\pm$ SE, n = 3). Statistical significance was calculated through a one-way ANOVA followed by a post-hoc t-test with Bonferroni's correction method. P-value $\leq$ 0.05 was used as the significance value. Statistically significant data are represented with *.	45
4.2	Total flavonoid content in 11 clones of tea plant. (Data is represented as mean $\pm$ SE, n = 3). Statistical significance was calculated through a one-way ANOVA followed by a post-hoc t-test with Bonferroni's correction method. P-value $\leq$ 0.05 was used as the	46

	significance value. Statistically significant data are represented with *.	
4.3	POX activity in 11 clones of tea plant. (Data is represented as mean $\pm$ SE, n = 3). Statistical significance was calculated through a one-way ANOVA followed by a post-hoc t-test with Bonferroni's correction method. P-value $\leq$ 0.05 was used as the significance value. Statistically significant data are represented with *.	47
4.4	APX activity in 11 clones of tea plant. (Data is represented as mean $\pm$ SE, n = 3). Statistical significance was calculated through a one-way ANOVA followed by a post-hoc t-test with Bonferroni's correction method. P-value $\leq$ 0.05 was used as the significance value. Statistically significant data are represented with *.	48
4.5	PAL activity in 11 clones of tea plant. (Data is represented as mean $\pm$ SE, n = 3). Statistical significance was calculated through a one-way ANOVA followed by a post-hoc t-test with Bonferroni's correction method. P-value $\leq$ 0.05 was used as the significance value. Statistically significant data are represented with *.	49
4.6	PPO activity in 11 clones of tea plant. (Data is represented as mean $\pm$ SE, n = 3). Statistical significance was calculated through a one-way ANOVA followed by a post-hoc t-test with Bonferroni's correction method. P-value $\leq$ 0.05 was used as the significance value. Statistically significant data are represented with *.	50
4.7	CAT activity in 11 clones of tea plant. (Data is represented as mean $\pm$ SE, n = 3). Statistical significance was calculated through a one-way ANOVA followed by a post-hoc t-test with Bonferroni's correction method. P-value $\leq$ 0.05 was used as the significance value. Statistically significant data are represented with *.	51

4.8	Filtering of transcripts for identification of putative lncRNAs	53
4.9	Coding potential analysis of the assembled transcripts through three tools	53
4.10	Chromosomal distribution of identified lncRNAs	54
4.11	Classification of identified lncRNAs based on their genomic position	55
4.12	Length distribution and number of exons found in the identified lncRNAs	56
4.13	A venn diagram showing number and proportion of lncRNAs expressed in the healthy/control and TMB-infested samples	56
4.14	Boxplot showing log <sub>2</sub> values of FPKM+1 of lncRNAs in the healthy/control and TMB-infested samples	57
4.15	Volcano plot showing the differential expression of lncRNAs in control v/s TMB-infested samples	60
4.16	Heatmap showing the differential expression pattern of lncRNAs in the six RNA-seq libraries	61
4.17	GO enrichment of lncRNA-target genes. BP; Biological processes, CC; Cellular component, MF; Molecular function. The size of the bubble represents number of lncRNA-target genes assigned to the particular GO term and the color of the bubble represents adjusted p-value (q value).	64
4.18	Gene set enrichment analysis based on GO enrichment of lncRNA-target genes. The left halve represents terms upregulated in response to TMB and the right halve represents terms downregulated in response to TMB. The size of the bubble represents number of lncRNA-target genes assigned to the particular GO term and the color of the bubble represents adjusted p-value (q value).	65

4.19	Gene set enrichment analysis based on KEGG pathway enrichment of lncRNA-target genes. The left halve represents pathways upregulated in response to TMB and the right halve represents pathways downregulated in response to TMB. The size of the bubble represents number of lncRNA-target genes assigned to the particular KEGG pathway and the color of the bubble represents adjusted p-value (q value).	67
4.20	Heatmaps showing expression patterns of lncRNA-target genes in different biological pathways (a) Terpenoids biosynthesis (b) Flavonoids biosynthesis (c) Zeatin biosynthesis (d) Plant hormone signal transduction (e) MAPK signaling pathway (f) Linoleic-acid metabolism (g) Brassinosteroids biosynthesis	69
4.21	Graphs showing positive/negative expression correlation between lncRNA and lncRNA-target genes. Blue line signifies FPKM values (x-axis) of lncRNAs and red line signifies FPKM values (x-axis) of lncRNA-target genes. Figures a and b denote negative correlation between lncRNA-mRNA pairs, figures c and d denote positive correlation between the pairs.	70
4.22	Melt curve plots of DELs TCONS_00040585 (a and b), TCONS_00083891 (c and d), TCONS_00096174 (e and f), TCONS_00032903 (g and h) and TCONS_00099260 (i and j) in control and TMB-infested samples	71
4.23	Bar diagrams showing result of qRT-PCR analysis of 5 DELs in control v/s TMB-infested samples. Values in the y-axis determine the relative expression of DELs normalized to UBC1 gene. Error bars indicate $\pm$ SEM of relative expression of triplicates.	72
4.24	Boxplot showing log <sub>2</sub> values of FPKM+1 of genes in the healthy/control and TMB-infested samples	73

4.25	A venn diagram showing number and proportion of genes expressed in the healthy/control and TMB-infested samples	74
4.26	Volcano plot and expression heatmap depicting DEGs	75
4.27	Gene set enrichment analysis based on GO enrichment of DEGs. The left halve represents terms upregulated in response to TMB and the right halve represents terms downregulated in response to TMB. The size of the bubble represents number of DEGs assigned to the particular GO term and the color of the bubble represents adjusted p-value (q value).	76
4.28	Gene set enrichment analysis based on KEGG pathway enrichment of DEGs. The left halve represents pathways upregulated in response to TMB and the right halve represents pathways downregulated in response to TMB. The size of the bubble represents number of DEGs assigned to the particular KEGG pathway and the color of the bubble represents adjusted p-value (q value).	79
4.29	Melt curve plots of DEGs CSS0024393.1 (a and b), CSS0006785.2 (c and d), CSS0016212.1 (e and f), CSS0018684.1 (g and h), CSS0023703.1 (i and j) and CSS0046901.1 (k and l) in control and TMB-infested samples	81
4.30	Bar diagrams showing result of qRT-PCR analysis of 6 DEGs in control v/s TMB-infested samples. Values in the y-axis determine the relative expression of DEGs normalized to UBC1 gene. Error bars indicate $\pm$ SEM of relative expression of triplicates.	82
4.31	(a) Bar diagram showing distribution of transcript length of both identified lncRNAs and expressed mRNAs (b) Comparison of exon numbers between identified	83



	lncRNAs and mRNAs (c) Boxplot showing distribution of FPKM values of lncRNAs and mRNAs	
4.32	Chromosomal distribution of identified circRNAs	84
4.33	Characterization of identified circRNAs in tea plant. (a) Pie chart showing classification of circRNAs based on their genomic positions (b) Venn diagram showing number of unique and common circRNAs identified in control and TMB-infested samples (c) Bar diagram showing length distribution of identified circRNAs (d) Percentage of circRNAs showing homology with already reported circRNAs of different plant species deposited in PlantcircBase	86
4.34	Differential expression pattern of identified circRNAs. (a) A heatmap showing the expression pattern of DECs across all six samples (b) Volcano plot showing differential expression of circRNAs	87
4.35	Annotation of the DEC-target genes based on GO. The x-axis depicts the enrichment ratio between number of DEC-target genes and all UniGenes enriched in a particular GO term. The size of the bubble represents number of DEC-target genes assigned to the particular GO term and the color of the bubble represents adjusted p-value (q value). BP; Biological processes, CC; Cellular component, MF; Molecular function	89
4.36	GSEA based on KEGG pathway of DEC-target genes. The x-axis depicts the enrichment ratio between number of lncRNA-target genes and all UniGenes enriched in a particular KEGG pathway. The size of the bubble represents number of DEC-target genes assigned to the particular KEGG pathway and the color of the bubble represents adjusted p-value (q value). The left half of	90

	the figure represents pathways upregulated in response to TMB and the right halve represents pathways downregulated in response to TMB.	
4.37	(a) Percentage of novel and conserved <i>C. sinensis</i> miRNAs putatively targeting DELs (b) Percentage of miRNAs targeting on DELs (c) Interactive networks of <i>C. sinensis</i> miRNAs and DELs represented by pink and purple ellipse nodes respectively. The connection between miRNAs and lncRNAs are shown as blue edges.	105
4.38	Result of the lncRNA mimic analysis. Query signifies miRNA and Subject signifies lncRNAs. Score determines the expectation threshold; Dash (-) determines bulge formation.	106
4.39	miRNA targeting lncRNA analysis through psRobot_tar; Score depicts the expectation threshold; Query determines the miRNA and subject determines the lncRNA; Asterisks (*) determine unpaired regions.	107
4.40	TMB-responsive miRNA-lncRNA-mRNA network in <i>C. sinensis</i> . miRNAs, mRNAs and lncRNAs are represented by pink prismatic shape, yellow ellipse and blue round rectangles respectively. The interactions between the RNA molecules are depicted by blue edges.	109
4.41	miRNA targeting mRNA analysis through psRobot_tar; Score depicts the expectation threshold; Query determines the miRNA and subject determines the mRNA; Asterisks (*) determine unpaired regions.	110
4.42	<i>C. sinensis</i> miRNA-DEC-DEG interaction network. miRNAs, genes and circRNAs are represented by red diamonds, green octagons and blue rectangles respectively. The connections between the RNA molecules are depicted by black edges.	111

4.43	<p>TMB-responsive ceRNA network in <i>C. sinensis</i>. miRNAs, mRNAs, lncRNAs and circRNAs are represented by purple prismatic shape, blue round rectangles, green octagons and red diamonds respectively. The interactions between the RNA molecules are depicted by black edges.</p>	119
5.1	<p>A hypothetical schematic regulatory network representing involvement of genes in defense responsive biological pathways during TMB infestation in <i>C. sinensis</i>. Peach colored ellipses represent DEGs and yellow ellipses represent genes that are not differentially expressed during TMB infestation. The heatmaps represent expression pattern of DEGs in six RNA-seq libraries (Control-1, Control-2, Control-3, Infested-1, Infested-2, Infested-3). The green rectangles represent plant hormones and blue rectangles represent the plant's response to TMB feeding downstream of the pathway involved.</p>	131

## List of Tables

<b>Table No.</b>	<b>Title</b>	<b>Page No.</b>
3.1	List of primers designed for validation of selected DELs	36
3.2	List of primers designed for validation of selected DEGs	40
4.1	Summary of reads generated from control and TMB-infested tea samples	52
4.2	List of DELs with their log <sub>2</sub> fold change and p-values	58
4.3	Results of lncRNA- <i>cis</i> target hybrid formation analysis using RIBlast algorithm	62
4.4	List of DECs with their log <sub>2</sub> fold change and p-values	86
4.5	List of DECs annotated with their parental genes	88
4.6	Table showing lncRNAs acting as eTMs and the corresponding mRNAs with their available information	91
4.7	Table showing <i>C. sinensis</i> miRNAs targeting DELs	92
4.8	List of DECs being targeted by <i>C. sinensis</i> miRNAs	111