# 3

# Materials and methods

## 3.1 Sample harvesting for biochemical analysis

Healthy and TMB-infested two leaves and a bud of tea plants belonging to nine clones viz. TV1, TV20, TV21, TV22, TV23, TV26, Tenali, TV9 and S3A3 were collected from commercial tea gardens of various parts of Assam like Biswanath, Behali, Sonapur. Tissues of two more clones P312 and AV2 were collected from Darjeeling Tea Research and Developmental Centre (DTRDC), West Bengal and Sungma and Turzum tea estate, Darjeeling, West Bengal. Samples were plucked and put in zip bags and then were immediately put inside boxes containing dry ice for transporting the samples to laboratory. After reaching the laboratory, the samples were transferred to -80ºC till further use.

## 3.2 Total phenolics estimation

For estimation of total phenolic content in the control and TMB-infested tea samples, the protocol suggested by Ainsworth and Gillepsie (2007) was followed. For extract preparation, powdered tissue was mixed well in 100% methanol in 10 mg/mL concentration and put in tubes. After shaking the mixture well, the tubes were kept in the dark for incubation upto 48 h. The extract was centrifuged in 13000 rpm for 5 mins and then 500 µL of the supernatant was collected for further processing. To the extract, 1 mL of freshly prepared 10% Folin-Ciocalteu reagent (F-C reagent) was added and vortexed briefly. The mixture was incubated for 5 mins followed by addition of 4 mL of freshly prepared 700 mM $Na_2CO_3$ solution. After a brief vortex, the samples were incubated for 2 h at room temperature in the dark. Finally, the absorbance of the samples was recorded at 765 nm against a blank. Gallic acid was used as a standard and the total phenolic

content was expressed as micrograms per milliliter of gallic acid equivalents (µg/mL of GAE).

### 3.3 Totals flavonoids estimation

The estimation of total flavonoid content was performed using a slightly modified protocol (Ahmed *et al*. 2015). Briefly, crushed control and TMB-infested tea samples were was dissolved in 100% ethanol in 10 mg/mL concentration and kept in the dark for incubation to upto 48 h. After incubation, the mixture was centrifuged at 13000 rpm for 5 mins and the supernatant was collected. To 1 mL of the extract, 300 µL of freshly prepared 5% sodium nitrite and 300 µL of freshly prepared 5% aluminium chloride solution were added. A brief vortex followed by incubation for 5 mins was done, after which 2 mL of 1M NaOH solution and 6.4 mL of $H_2O$ was added. The mixture was properly mixed and absorbance was taken at 510 nm. Quercetin was used as a standard and the total flavonoid content was expressed as micrograms per milliliter of quercetin equivalents (µg/mL of QE).

### 3.4 Antioxidative enzymatic assays

#### 3.4.1 Assay of peroxidase (POX)

The protocol provided by Chance and Maehly (1995) was followed with slight modifications. Briefly, 250 mg of powdered tissue was taken in 1 mL of freshly prepared chilled 0.1 M sodium phosphate buffer (pH 7). The mixture was centrifuged at 10000 rpm for 10 mins at 4°C. To 100 µL of the supernatant, 500 µL of 1% (v/v) $H_2O_2$ was added, followed by addition of 0.5 M pyrogallol to make the final volume to 3 mL. The absorbance was recorded at 430 nm at time interval of 1 min. POX activity was expressed

as units per mg protein. One unit of POX will form 1mg of purpurogallin from pyrogallol in 20 secs.

$$POX = \frac{\frac{\Delta Abs}{20secs} \times \ total\ assay\ volume}{12 \times volume\ of\ enzyme\ extract}$$

### 3.4.2 Assay of ascorbate peroxidase (APX)

For APX extraction, a slightly modified protocol of Nakano and Asada (1981) was followed. Briefly, 250 mg of powdered tissue was dissolved in 1 mL of 0.1 M sodium phosphate buffer (pH 7.2) and centrifuged at 2°C for 20 mins at 10000 rpm. The supernatant was collected as the enzyme source. The assay was done by mixing 10 µL of the enzyme extract, 10 µL of 0.5 mM L-ascorbic acid, 10 µL of $H_2O_2$ (30%) and 2.97 mL of 0.05 M sodium phosphate buffer (pH 7.2) and measuring reduction in absorbance per minute at 290 nm. The APX activity was expressed µmol of ascorbate oxidized per min per mg protein. One unit of APX activity is defined as the amount of enzyme that can oxidize 1.0 µM of ascorbic acid per minute.

$$APX = \frac{Abs}{2.8 \times g.f.w}$$

### 3.4.3 Assay of phenylalanine ammonia lyase (PAL)

For PAL estimation, protocol provided by Sadasivam and Manickam (1996) was followed with slight modifications. In 1 mL of sodium borate buffer, 250 mg of powdered tissue was dissolved and vortexed briefly. The mixture was centrifuged at 4°C at 10000 rpm for 15 mins. The reaction mixture constituted of 500 µL of enzyme extract, 300 µL of sodium borate buffer, 300 µL of 0.05 M L-phenylalanine and 4.9 mL distilled water.

The mixture was incubated at 40°C for 1 h. Absorbance was recorded at 290 nm. The increase in absorbance was recorded at 1 min interval for 3 mins. The enzyme activity in the sample was expressed as µmol of t-cinnamic acid produced from L-phenylalanine per min per mg protein.

$$PAL = \frac{\Delta Abs \times 6}{19.73 \times 0.5}$$

### 3.4.4 Assay of polyphenol oxidase (PPO)

Protocol from Singh and Ravindranath (1994) was followed. 250 mg of crushed tissue was dissolved in 1 mL of chilled 0.1 M sodium phosphate buffer (pH 7). Centrifugation was performed for 15 mins at 10000 rpm at 4°C. The reaction mixture consisted of 100 µL of enzyme extract, 1.9 mL of 0.1 M sodium phosphate buffer (pH 7) and 1 mL of 0.01 M catechol. Absorbance was measured at 380 nm. The activity was determined spectrophotometrically at 380 nm and expressed as units per mg protein. One unit of enzyme activity is defined as that amount which caused a rate of change of 0.0001 absorption units per min at 380 nm.

### 3.4.5 Assay of catalase (CAT)

Catalase activity was estimated by the UV method of Aebi (1983). Briefly, 100 mg of tissue was ground to powder by liquid nitrogen and the powdered tissue was suspended in 2 mL of 0.1 M potassium phosphate buffer (pH 6.8) containing 1mM EDTA and 1% (w/v) polyvinyl pyrrolidone (PVP) in cold condition. The mixture was centrifuged at 13000 g for 20 mins. About 100 µL of 10x diluted enzyme extract is added to a reaction mixture containing 1.8 mL of 50 mM potassium phosphate buffer (pH 6.8) and 100 µL of 100 mM $H_2O_2$. The absorbance was recorded at 240 nm at an interval of 15 secs for 2

mins. The enzyme activity was expressed as μmol of $H_2O_2$ reduced per min per mg protein.

$$CAT = \frac{Abs \times 100}{43.3 \times g.f.w./reaction\ mixture}$$

**3.5 Plant materials and stress treatments**

Phoobsering 312 (P312) is a widely grown tea clone of Darjeeling district of West Bengal producing high quality tea with distinctive aroma and flavour. This cultivar is also moderately susceptible to TMB infestation and therefore, we selected P312 clone for the current study. Adult TMBs were hand-collected from tea plantations of Darjeeling Tea Research and Development Centre (DTRDC, Kurseong) and Sungma-Turzum tea estate of Darjeeling district of West Bengal. Two-year old tea plants were planted in plastic pots (14 cm diameter and 15 cm height) and irrigated once every day. Plants were placed inside chambers well-caged by nets. Two sets of plants were maintained for control and treatment conditions. The health of the plants was carefully monitored. On the day of the experiment, 5-6 adult TMBs were placed on tender two leaves and a bud of each tea plant (healthy with no disease symptoms) in the treatment chamber. Harvesting of samples was done after 2h, 12h and 24h of TMB-feeding. Tissue harvesting was done by collecting two leaves and a bud from infested and healthy plants and immediately frozen in liquid nitrogen for RNA extraction.


Fig. 3.1 Healthy and *Helopeltis theivora* infested tea leaves. (a) Healthy tea leaf and bud (b) Development of symptoms after feeding (c, d) *H. theivora* feeding on leaves

### 3.6 RNA isolation, library construction and Illumina sequencing

TRIzol® reagent (Invitrogen) was used to extract the total RNA from each sample. The quality and purity of RNA was analyzed by using a 2100 Bioanalyzer and the RNA 6000 Nano LabChip Kit (Agilent, CA, USA), with a RIN number > 7.0. Removal of rRNA was performed using Ribo-Zero rRNA removal kit (Illumina, San Diego, CA, USA). Strand-specific cDNA library was then constructed for control and TMB-infested samples (three replicates each) through the TruSeq Stranded Kit (Illumina) and 150 bp paired-end reads were generated using Illumina HiSeq 2000 system.

### 3.7 LncRNAs

#### 3.7.1 Bioinformatics pipeline for identification of lncRNAs and differentially expressed lncRNAs (DELs)

The raw reads generated through Illumina sequencing were quality trimmed to remove adapters and low-quality sequences using Trimmomatic v0.39. (Bolger *et al*. 2014). Assessment of the quality of the data was done by FastQC followed by mapping of clean reads to the reference genome of *Camellia sinensis* var. sinensis (Xia *et al*. 2020) deposited in the Tea Plant Information Archive (TPIA) (Xia *et al*. 2019) using HISAT2

(Kim *et al*. 2015) software. Assembly of the reads and their abundance estimation was carried out by StringTie v2.0.5 (Pertea *et al*. 2016). Finally all assembled reads were merged into a final transcriptome using gffcompare (Pertea and Pertea, 2020) and the transcripts were further screened for identification of putative lncRNAs.

For basic filtering, the assembled transcripts that were less than 200 nucleotides and with exon numbers < 2 were filtered out (Wan *et al*. 2020) using a Perl script. Transcripts belonging to any of the class codes viz. u, i, o and x were only retained for further screening. For protein coding potential assessment, we used three programs viz. CPC (Kong *et al*. 2007), CNCI (Sun *et al*. 2013) and PLEK (Li *et al*. 2014a). Transcripts that were identified as "noncoding" by each of them were considered for further analysis. Those transcripts that possessed ORF length of ≥ 100 amino acids were discarded after analysing through TransDecoder (https://transdecoder.github.io) (Haas *et al*. 2013). Furthermore, the transcripts were scanned for the presence of housekeeping RNAs that were subsequently removed by performing cmscan against the Rfam database (Kalvari *et al*. 2018). We did a BLAST search against the protein sequences of *C. sinensis* with an e-value cut-off of 0.01 and the transcripts that showed homology with *C. sinensis* protein sequences were removed from further analysis. Finally BLASTX for the remaining transcripts against the Pfam (Finn *et al*. 2016) database was performed to discard potential protein coding transcripts. The remaining transcripts that fulfilled all requirements were considered as candidate lncRNAs. The raw read counts of the identified lncRNAs were generated using featureCounts (Liao *et al*. 2014) and finally DELs were identified by DESeq2 package (Love *et al*. 2014) setting log2 fold change cut-off of ≥1 (for upregulation) and ≤-1 (for downregulation) and adjusted p-value of ≤ 0.05.

### 3.7.2 LncRNA target gene prediction

LncRNAs can act on neighbouring genes through *cis*-regulation and it is a well-known phenomenon. To identify genes that may serve as *cis*-targets of lncRNAs, we performed a search for closest coding genes lying upstream and downstream of lncRNAs within a 10kb window through BEDTools v 2.25 program (Quinlan and Hall, 2010). To identify candidate genes as *trans*-targets of lncRNAs, we analysed the ability of lncRNAs to form RNA-RNA hybrid through an interaction ability search using RIblast algorithm v 1.2.0 (Fukunaga and Hamada, 2017). Only those interactions having hybridization energy threshold of less than -30 kcal $mol^{-1}$ were considered as significant targets. Simultaneously, an expression correlation analysis was also performed between DELs and DEGs to find out co-expressing pairs. For this, we calculated the Pearson's correlation coefficient (PCC) for DEL-DEG pairs and considered only those pairs as significantly correlated that showed PCC of $\leq -0.9$ (for negative correlation) or $\geq 0.9$ (for positive correlation) and with p-values $\leq 0.01$.

### 3.7.3 Functional annotation and gene set enrichment analysis (GSEA) of lncRNA targets

To understand the probable functions of genes targeted by lncRNAs, we performed a Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways annotation/enrichment analysis. The classification of the lncRNA target genes based on GO terms and KEGG pathways with which they are associated, was performed using clusterProfiler R package (Yu *et al*. 2012). GSEA in the GO annotation and KEGG pathways for lncRNA target genes was performed using clusterProfiler assuming a minimum of 3 genes and a maximum of 800 for each GO term/KEGG pathway to have an unbiased overview of the annotation and enrichment analysis. gseGO and gseKEGG

functions of clusterProfiler were used for gene set enrichment of GO and KEGG respectively.

### 3.7.4 Quantitative real-time PCR of selected DELs

Total RNA extraction was performed for all six samples including replicates (3 control and 3 TMB-infested samples harvested as previous) through TRIzol® reagent (Invitrogen). After RNA quantification, first strand cDNA synthesis for 1μg RNA was prepared using PrimeScript RT reagent kit with gDNA eraser (TaKaRa). Primers for five DELs were designed using Primer3 software with the following parameters- product size range was kept from 75-200 bp, optimum primer melting temperature at 55°C, optimum primer GC percentage at 55% and self-complementarity was set to zero.

Table 3.1 List of primers designed for validation of selected DELs

| lncRNA id | Forward primer (5′ to 3′) | Reverse primer (3′ to 5′) | Product size (bp) |
|---|---|---|---|
| TCONS_00040585 | AGCATGCTGAGTTTGGGCTT | AGCTCCCCAATACCAGTTCCA | 168 |
| TCONS_00083891 | TCTCGACAAGGCAAGGCAAA | GCGTCCACCTTCAATCTCCA | 176 |
| TCONS_00096174 | GGGAGTGGGTGAAGAAGATGG | GGAGGAGGAGGAGAAGAAGACT | 120 |
| TCONS_00032903 | AAGGGTTTGGGTGACTTGCTT | ATGCCTGCACCTGTGTAATGT | 162 |
| TCONS_00099260 | TCTTCGGGCCAGAACACATC | GTGCAGTTACAACAGTCCCCT | 188 |

The quantification of DELs was performed in an Applied Biosystems QuantStudio 5 Real-time PCR system (ThermoFisher Scientific). The quantitative real-time PCR (qRT-PCR) was performed using the PowerUp SYBR Green Master Mix (Applied Biosystems) according to the manufacturer's instructions. The reaction cycle was as follows: 50°C for 2 mins, 95°C for 2 mins followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. For melt curve analysis, the temperature was increased to 95°C for 15 s followed by 60°C for 1 min (1.6°C/s). The final dissociation was obtained at 95°C for 1 s (0.15°C/s). All

reactions were performed in triplicates. The UBC1 gene was used as an internal control for normalization of quantitative expression (Xu *et al*. 2020). The $2^{-\Delta\Delta Ct}$ method was used to calculate the fold change for the DELs in control v/s treatment condition and $\Delta C_T$ method was used to calculate the relative expression of DELs normalized to the reference/housekeeping gene (UBC1).

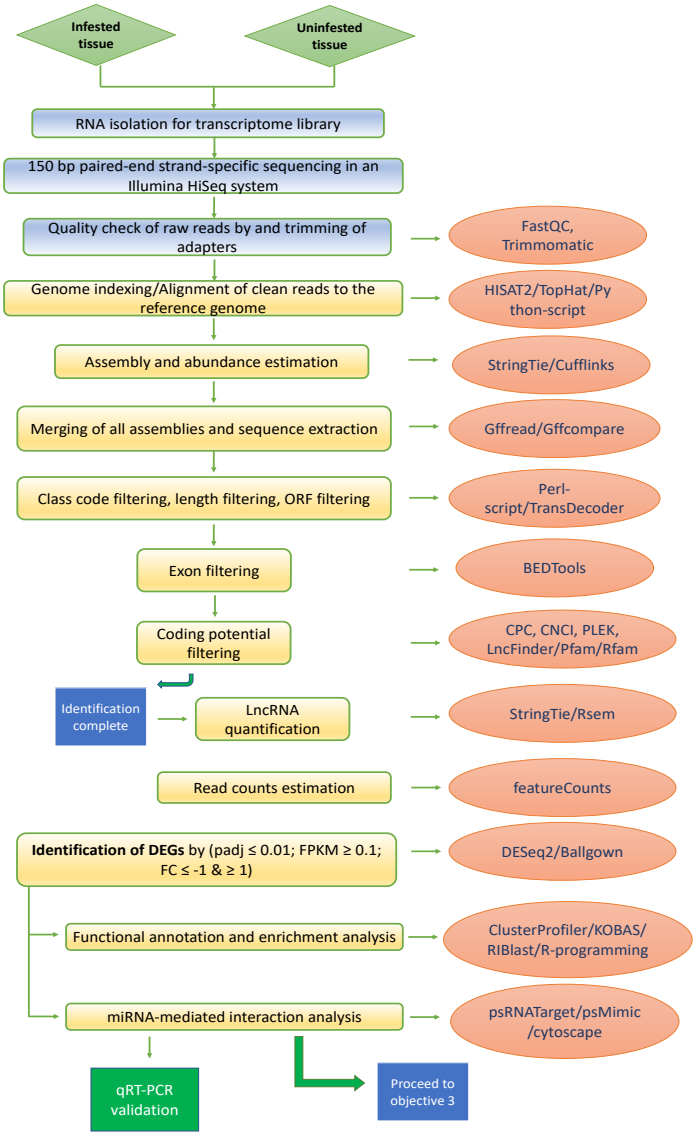


Fig. 3.2 A flowchart depicting pipeline employed for identification and characterization of lncRNAs in healthy and TMB-infested tea plants. The peach ellipses denote softwares/tools used in each corresponding step.

### 3.8 Genes

#### 3.8.1 Identification of differentially expressed genes (DEGs)

The expression of 57429 genes of *C. sinensis* were analyzed by StringTie v2.0.5 software. The expressed genes (FPKM ≥ 0.1) were subjected to differential expression analysis using featureCounts and DESeq2 package and lowly expressed genes (FPKM < 0.1) were discarded. Genes with an absolute value of log2 fold change ≥1 and ≤-1 and with adjusted p-values ≤ 0.05 found by DESeq2 software were considered as DEGs.

#### 3.8.2 Functional annotation and gene set enrichment analysis (GSEA) of DEGs

To understand the probable functions of DEGs, we performed a Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways annotation/enrichment analysis. The classification of the DEGs based on GO terms and KEGG pathways with which they are associated, was performed using clusterProfiler R package (Yu *et al*. 2012). GSEA in the GO annotation and KEGG pathways for DEGs was performed using clusterProfiler assuming a minimum of 3 genes and a maximum of 800 for each GO term/KEGG pathway to have an unbiased overview of the annotation and enrichment analysis. gseGO and gseKEGG functions of clusterProfiler were used for gene set enrichment of GO and KEGG respectively.

#### 3.8.3 Quantitative real-time PCR of selected DEGs

Total RNA extraction was performed for all six samples including replicates through TRIzol® reagent (Invitrogen). After RNA quantification, first strand cDNA synthesis for 1μg RNA was prepared using PrimeScript RT reagent kit with gDNA eraser (TaKaRa). Primers for six DEGs were designed using Primer3 software with the following parameters- product size range was kept from 100-200 bp, optimum primer melting

temperature at 55°C, optimum primer GC percentage at 55% and self-complementarity was set to zero.

Table 3.2 List of primers designed for validation of selected DEGs

| mRNA id | Forward primer (5′ to 3′) | Reverse primer (3′ to 5′) | Product size (bp) |
|---|---|---|---|
| CSS0024393.1 | CCTCTCGTCTCCAAGCTGTG | ACTTGTGTCTGTGGTGGCAA | 161 |
| CSS0006785.2 | GGCGGACTGGTCATGTAACA | TGACGTCGGTGATACTCCCT | 177 |
| CSS0016212.1 | CGAGCCGGTTTGAGTTCTCA | TCGTCAAATTCAGCCAGCGA | 176 |
| CSS0018684.1 | CCCAACTAGCCGGAACCAAT | CCGAACTTGGAGCCTGCATA | 132 |
| CSS0023703.1 | CGAGCTGGGCATCTTGAAGA | TCAATGACACGTAGGGCTGC | 146 |
| CSS0046901.1 | CGAGCCGGTTTGAGTTCTCA | TCGTCAAATTCAGCCAGCGA | 197 |

The quantification of DEGs was performed in an Applied Biosystems QuantStudio 5 Real-time PCR system (ThermoFisher Scientific). The qRT-PCR was performed using the PowerUp SYBR Green Master Mix (Applied Biosystems) according to the manufacturer's instructions. The reaction cycle was as follows: 50°C for 2 mins, 95°C for 2 mins followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. For melt curve analysis, the temperature was increased to 95°C for 15 s followed by 60°C for 1 min (1.6°C/s). The final dissociation was obtained at 95°C for 1 s (0.15°C/s). All reactions were performed in triplicates. The UBC1 gene was used as an internal control for normalization of quantitative expression. The $2^{-\Delta\Delta Ct}$ method was used to calculate the fold change for the DEGs in control v/s treatment condition and $\Delta C_T$ method was used to calculate the relative expression of DEGs normalized to the reference/housekeeping gene (UBC1).

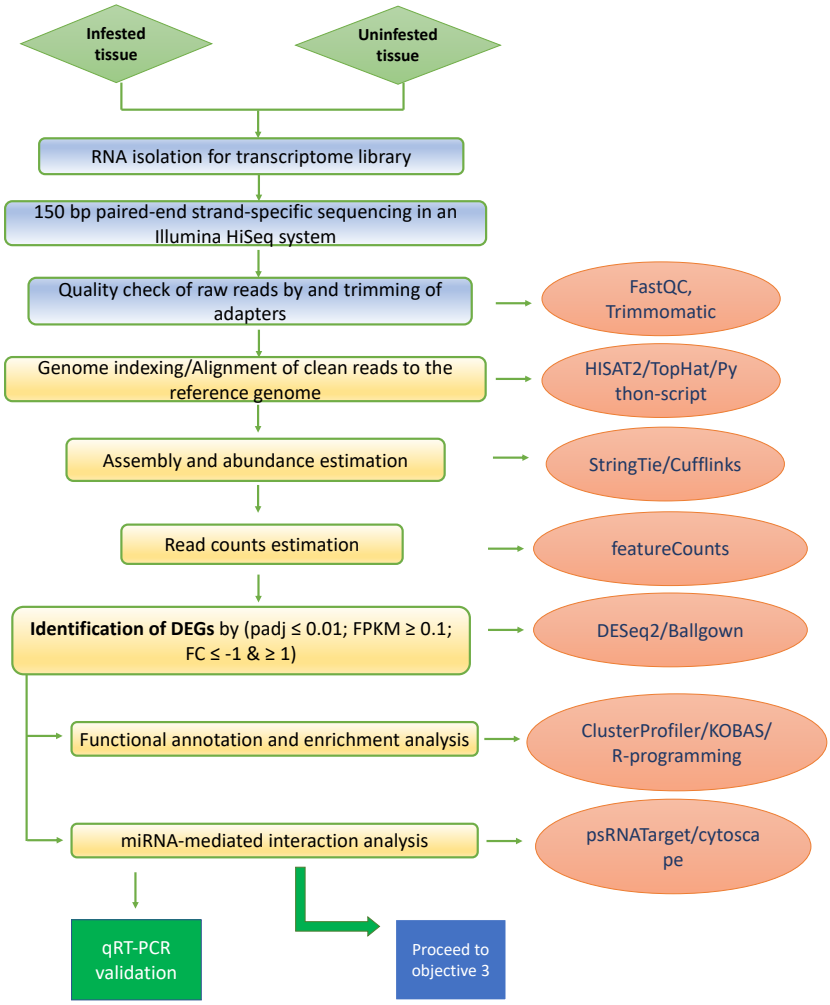


Fig. 3.3 A flowchart depicting pipeline employed for identification of differentially expressed genes (DEGs) in healthy and TMB-infested tea plants. The peach ellipses denote softwares/tools used in each corresponding step.

## 3.9 CircRNAs

### 3.9.1 Pipeline for identification of circRNAs and DECs

The high-quality clean reads were mapped to the reference genome using BWA-MEM (version 0.7.17) (Li and Durbin, 2009) using the parameters recommended by CIRI2 (version 2.0.6). Detection and identification of circRNAs was carried out through CIRI2

algorithm (Gao *et al*. 2018). CIRI2 scans the Sequence Alignment Map (SAM) files twice to detect circRNAs supported by at least two back-spliced reads. The CIRI2 runs a preliminary filtering by the paired end mapping (PEM) and GT-AG splicing signals and detects junction reads with paired chiastic clipping (PCC) signals in SAM files during the first scan. This is followed by a second scan to detect additional junction reads and simultaneously filters out false positives produced due to incorrect mapping. The sequences for the final candidate circRNAs were extracted using BEDTools v 2.25 (Quinlan and Hall, 2010). Raw counts for the identified circRNAs were generated using featureCounts (Liao *et al*. 2014) and the reads obtained for the candidate circRNAs were normalized to transcripts per million (TPM) values (Cooper *et al*. 2018). Differential expression (DE) analysis was performed using DESeq2 package (Love *et al*. 2014). Candidate circRNAs showing log2 fold change values of 1 or -1 were considered as DECs. For assessing whether the identified circRNAs of this study have any sequence similarity with known circRNAs deposited in PlantcircBase, we performed a sequence homology search for the identified circRNAs against those of other plant species of PlantcircBase (Chu *et al*. 2017).

### 3.9.2 Expression correlation analysis of DECs and DEGs

NcRNAs coexpressing with genes are known to regulate their expression (Bhatia *et al*. 2017, 2020). To decode the probable functions of DECs, an expression correlation analysis was performed for DECs and DEGs. The Pearson's correlation coefficient (PCC) for DEC-DEG pairs was estimated using a customized R-script and pairs exhibiting PCC of either 0.9 or -0.9 with p-values less than 0.01 were selected for further analysis. (Bordoloi *et al*. 2021).

### 3.9.3 Functional annotation of DECs

To functionally annotate the DECs, GO and KEGG analysis for DEC parental genes and genes coexpressing with DECs i.e., the DEC-target genes was carried out using the ClusterProfiler package of R (Yu *et al*. 2012).

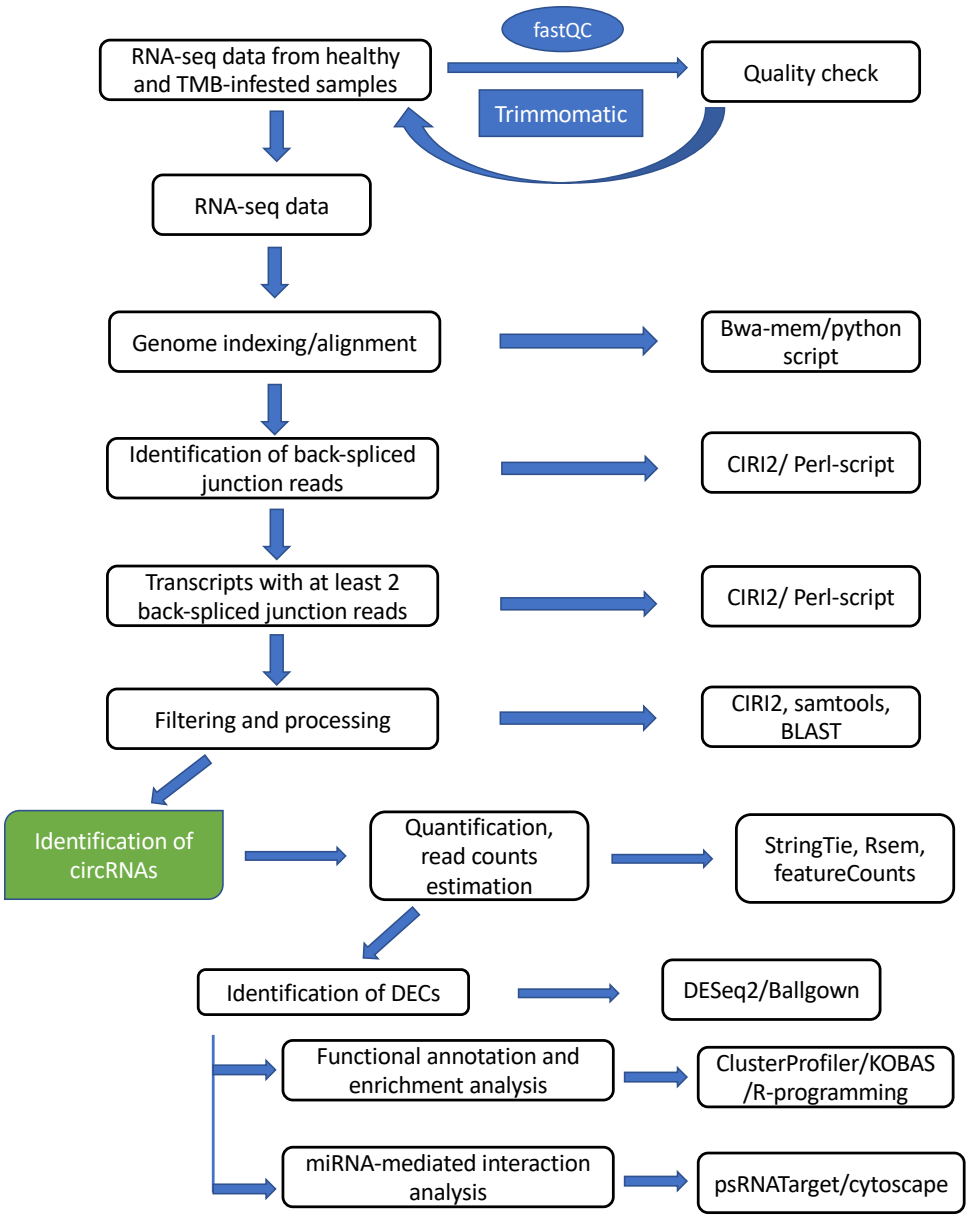


Fig. 3.4 A flowchart depicting pipeline employed for identification of differentially expressed circRNAs (DECs) in healthy and TMB-infested tea plants.

### 3.10 Collection of miRNA data

The ceRNA relationship was predicted based on the ceRNA hypothesis i.e., mRNAs, lncRNAs, circRNAs can regulate expression of each other using miRNA response elements (MREs) (Salmena *et al*. 2011). To evaluate the potential of DELs, DECs, DEGs as ceRNAs, we collected previously reported conserved and novel *C. sinensis* miRNAs (Jeyaraj *et al*. 2017a, 2017b).

### 3.11 miRNA target prediction

The DELs, DECs and DEGs that could potentially bind to miRNAs were identified by psRNATarget by setting target accessibility to 25 and other parameters to default (Dai and Zhao, 2011) and psRobot_tar with expectation threshold of 2.5 (Wu *et al*. 2012).

### 3.12 eTM prediction

The psMimic software (Wu *et al.* 2012) was used to find out the lncRNAs and circRNAs that can act as potential decoy of mRNAs and as miRNA sponges. Previously reported *C. sinensis* miRNAs were used to predict eTMs with the following parameters: (a) only three nucleotides were allowed to form the bulge (b) not more than three mismatches were allowed in the miRNA-lncRNA and miRNA-circRNA pairing region except the bulge.

### 3.13 Network construction and visualization

The ceRNA network was constructed by using the commonly identified miRNA-mRNA, miRNA-lncRNA and miRNA-circRNA pairs from both miRNA target prediction softwares. The ceRNA network was visualized using Cytoscape version 3.8.2 software (Shannon *et al*. 2003) to display the potential relationship between DELs, DECs, DEGs, and *C. sinensis* miRNAs.