Int J Adv Life Sci Res. Volume 7(2) 143-162
https://doi.org/10.31632/ijalsr.2024.v07i02.012

International Journal of Advancement in Life Sciences Research
Online ISSN: 2581-4877
Journal homepage http://ijalsr.org

Original Article

Transcriptome Analysis from leaf during non-flowering vegetative growth phase of mango (*Mangifera indica* L. *var* Fazli) from Murshidabad district of West Bengal, India.

Santi Ranjan Dey[1], Sayak Ganguli[2] and Mitu De[3*]

[1]*Department of Zoology, Rammohan College, Kolkata 700009, West Bengal, India.*
[2]*Postgraduate and Research Department of Biotechnology, St. Xavier's College (Autonomous), Kolkata 700016, West Bengal, India*
[3]*Department of Botany, Gurudas College, Kolkata 700054, West Bengal, India.*

**Correspondence E-mail: mitubotany@gmail.com*

Abstract

Mango (*Mangifera indica* L.) is a major fruit crop in numerous tropical and subtropical countries, facing various problems such as susceptibility to diseases, bi-annual production, low yield and a brief shelf life of the fruit, in cultivation. Traditional breeding methods have limited success in improving the quality of this fruit crop due to prolonged time of maturity, self-incompatibility and high degree of heterozygosity in breeding lines. However, recent studies utilizing genomic analysis have identified key genes responsible for economically important characters, suggesting the possibility of Marker Assisted Selection (MAS) in breeding lines through transgenic or specific genotype selection from seedlings. With the rapid advancements in genome sequencing and bioinformatics, it is now feasible to identify, label, clone, and manipulate numerous genes related to economically important characters. While research on the flowering and fruiting of mango has been extensive, research studies to preliminarily identify the functions of mango genes during the vegetative growth phase is very few. The major findings include few highly active genes viz. WRKY, NAM, MYB, GoGID1 which is channelizing the energy towards vegetative growth and inhibit flowering genes.

Keywords: *Fazli, Mango, MYB, Transcriptomics, West Bengal, WRKY*

Introduction

Mango, (*Mangifera indica* L.) is a fruit with succulent drupe that belongs to the Anacardiaceae family. It is most commonly grown fruit in the tropics, after banana and pineapple (according to FAOSTAT). This fruit is found in hillsides, river valleys and forests with elevations ranging from 200 to 1350 meters (Tharanathan *et al*., 2006). It is a popular fruit because of its' flavor, taste and significant nutritional value. The pulp is eaten, which is full of sugars, amino acids, aromatic as well as functional compounds such as pectin, vitamins, anthocyanins, polyphenols etc. (Subramanyam *et al.,* 1975).

Mango pulp can contain up to 200 mg/100 g of β-carotene, which is significantly higher (10 to 50 times) than the levels found in bananas and apples (Ledesma and Campbell, 2019). The primary active component in mango leaves, mangiferin, has been found to have antioxidant properties and contribute to antibacterial and immune-modulating effects (Sagar *et al.,* 1999). About 1000 cultivars of Mango are found globally. Depending on embryo, there are two kinds of varieties: mono-embryonic and poly-embryonic. Mono-embryonic (India) mango seeds contain only one zygotic embryo and are propagated through sexual reproduction, resulting in a single seedling (Iyer and Subramanyam, 1991; Mukherjee and Litz,2009; Iyer and Dinesh,1997; Viruel *et al.,*2005)). However, whether mono-embryonic or poly-embryonic, these seedlings may vary in their characteristics and do not necessarily maintain the desirable traits of the parent plant. The poly embryonic type of mango is typically

Received on :31st January 2024; Revised version received on :26th March 2024; Accepted: 5th April 2024

restricted in subtropical regions and has a red flesh and green or yellow fruit cover. *Mangifera siamensis* Warb. *ex* Craib (The Thai Mango) are of this kind (Mahato *et al.,* 2016; Shudo *et al.,* 2013, Tsai *et al.,* 2013; Yamanaka *et al.,* 2019). On the other hand, *Mangifera indica,* themono-embryonic mango is widely distributed throughout the world (De, M. *et al.,* 2017). In comparison to traditional breeding methods, molecular markers viz. SSR (Simple Sequence Repeats), CAPS (Cleavage Amplification Polymorphic Sequence) and RTIP (Reverse Transposon Insertion Polymorphic) markers are found to be more efficient and effective in mango research (Nashima *et al.*, 2017; Peng *et al.*, 2022;Tafolla-Arellano *et al.*, 2017). Now a days, there has been considerable and important improvement in mango breeding, with the use of next generation sequencing for identification of genes responsible for important traits (Pal, A. *et al.,* 2019). This has led to introduction of new mango varieties through direct transgene or genotype selection. As a result, there have been availability oftranscriptome data and genetic map data for mango (Arumuganathan, and Earle1991; Azim *et al.*, 2014; Kaur *et al.*, 2022, Kuhn *et al.*, 2017, Mukherjee, 1953; Sivankalyani *et al.*, 2016; Wang *et al.*, 2020).

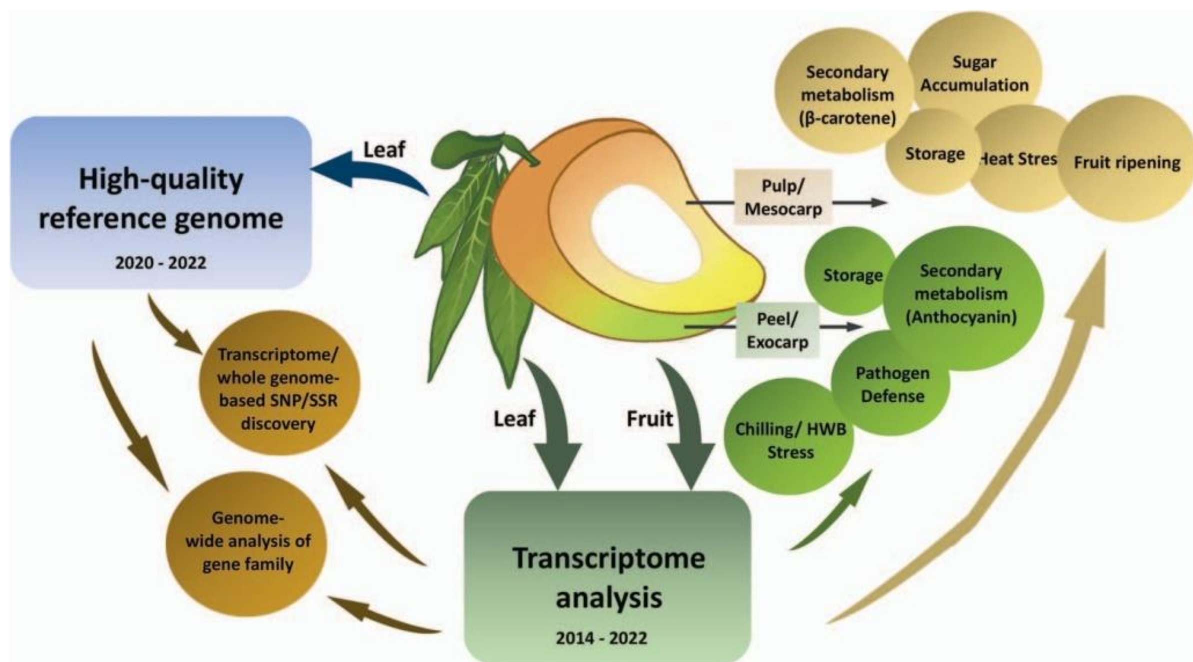


Figure1. Available Mango Genomics (Song *et al.,* 2023).

Transcriptome sequencing analysis of mango

Fig. 1 illustrates that most of the current findings on the mango transcriptome research, that has been focused on developmental process of fruit (Kaur *et al.*, 2022). Ripening of fruit is a complex process, that involves the development of flavor, development of pulp and peel color, changes in cell wall components, degradation of pulp starch and the development of volatile compound responsible for fragence. These factors ultimately determine the uniqueness of the variety (Sivankalyani *et al.*, 2016). In 2014, Azim *et al.* utilized Trinity, a short-read assembly program, to identify the transcriptome network of Langra mango (Azim *et al.*, 2014). Over 13,500 unigenes, were identified, which were assigned by 293 KEGG pathways. This provided a perspective of exploring regulatory genes involved in mango growth and development using transcriptome technology. (Gómez-Ollé *et al.*, 2023; Tafolla-Arellano *et al.*, 2017).

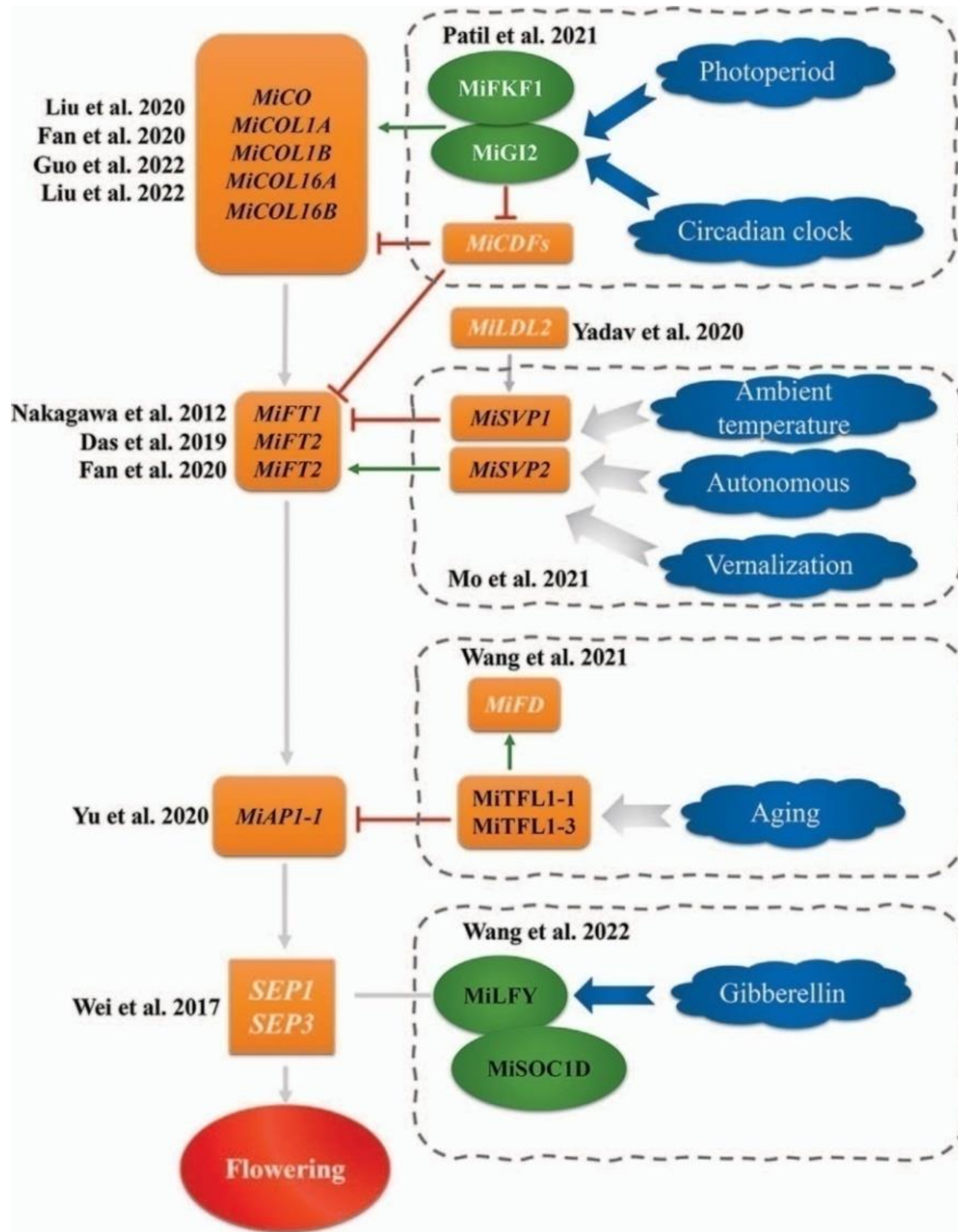


Figure2. Floweringregulatorygenesinmango. (Song *et al.,* 2023)

AP1: APETALA1;
CDF1: CyclingDOFFactor1;
CO: CONSTANS;
COL1A/B: CONSTANS-like1A/B;
COL16A/B: CONSTANS-like16A/B;
FD: FLOWERINGLOCUSD;
FKF: ,Flavin-Binding Kelch Repeat Fbox Protein;
FT1/2/3: FLOWERINGLOCUST1/2/3;
GI: Gigantea 2;
LDL2: Lysine Specific De-methylase Like 1;
LFY: LEAFY;
SEP 1/3: SEPALAATA 1/3;
SOC1: SUPPRESSOR OF OVER-EXPRESSION OFCONSTANS1
SVP1/2: ShortVegetative Phase1/2;
TFL1: TERMINAL FLOWER1.
Genes in black colour font have previously been validated by transgenic Arabidopsis or Tobacco (Gómez-Ollé *et al.*, 2023).

Materials and Methods

The Transcriptome Analysis of the Mango Genome (*Mangifera indica* L. var. Fazli) the methods that were used were cited in many papers viz. (Andrews, 2010; Chow *et al.*, 2016; Grabherr, 2011, Haas and Papanicolaou, 2016; Johnson *et al.*, 2008; Thiel, 2003;).

RNA extraction and QC

RNA was extracted from fresh, disinfected, fresh, mature mango leaf samples of variety Fazli, collected in RNA buffer, using standarized protocol. The quality of extracted RNA was checked on Bioanalyzer (Figure 3a & 3b) and quantified using QUBIT dsRNA HS kit.

Ladder

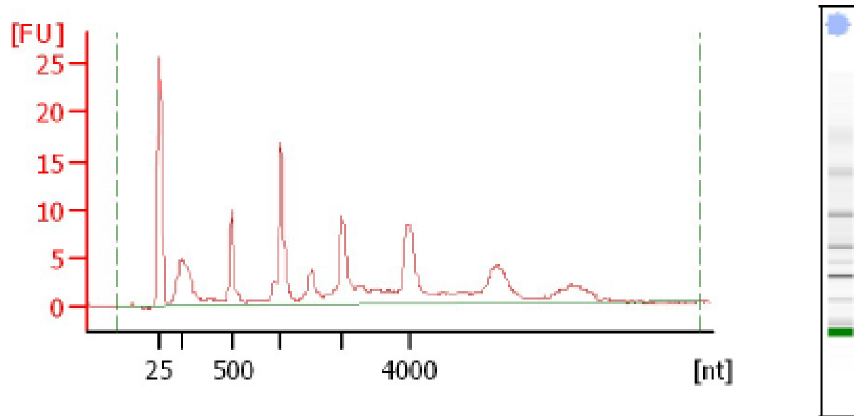
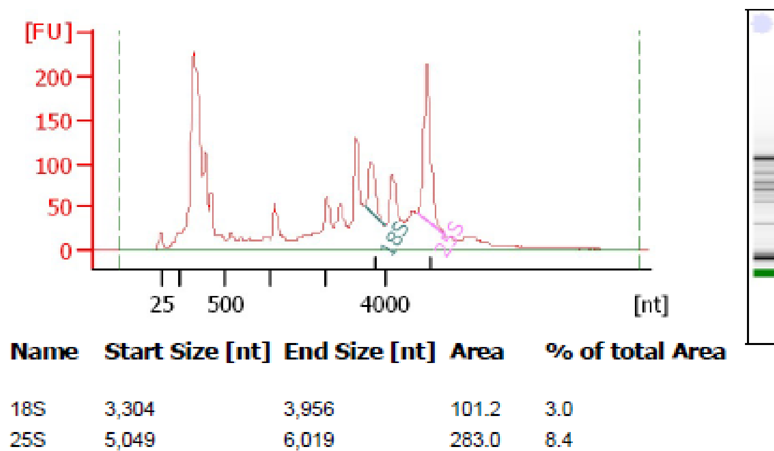


Figure 3a: Ladder for RNA quality check on Bioanalyzer

Mango leaf



Name	Start Size [nt]	End Size [nt]	Area	% of total Area
18S	3,304	3,956	101.2	3.0
25S	5,049	6,019	283.0	8.4

Figure 3b: Extracted Mango leaf RNA quality check using Bioanalyzer

The quantification of the extracted RNA sample was carried by QUBIT (Table 1). The extracted RNA has met the Illumina standards and has been taken further for sequencing.

Table 1: Quantification of extracted RNA by QUBIT

Sl. No.	Sample name	Concentration (ng/µl)	Volume (µl)
1	Mango leaf (Var. Fazli)	112.5	25

Library preparation

The library was prepared using “NEBNext® Ultra™ RNA Library Prep Kit for Illumina®” with Illumina standardized protocol. The following were the steps involved in the library preparation.

1. Ribosomal RNA depletion using r-RNA Depletion Kit
2. Purify and Fragment mRNA
3. First strand cDNA synthesis
4. Second strand cDNA synthesis
5. End Repair and adenylation of 3’ ends
6. Ligation of indexed Paired-End adapters

Library validation

The final enriched library was further validated for quality on Agilent Bioanalyser using DNA High Sensitivity chip (Figure 4a & 4b) and quantification on real time PCR (KAPA Library Quantification kit). The following graphs represent the prepared library validation using Bioanalyzer.

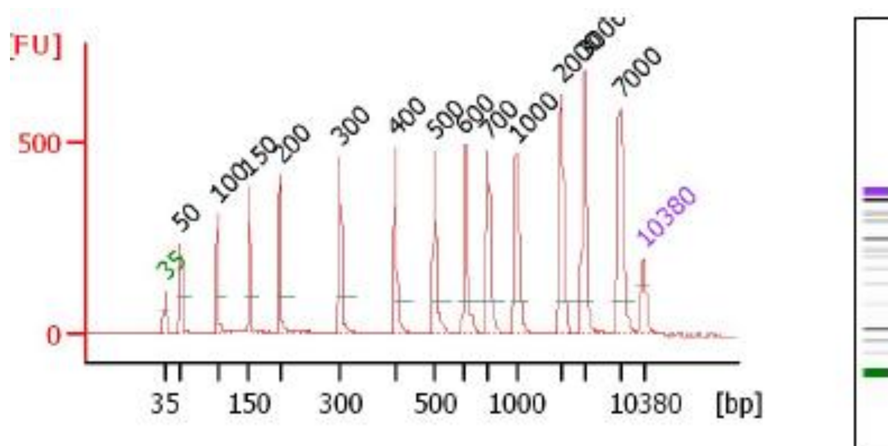
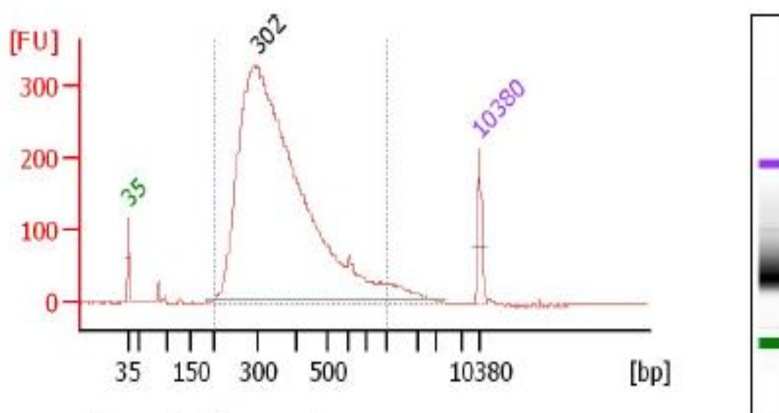


Figure 4a: Ladder for library quality check on Bioanalyzer

Sample Mango (*Mangifera indica* L. var. Fazli)



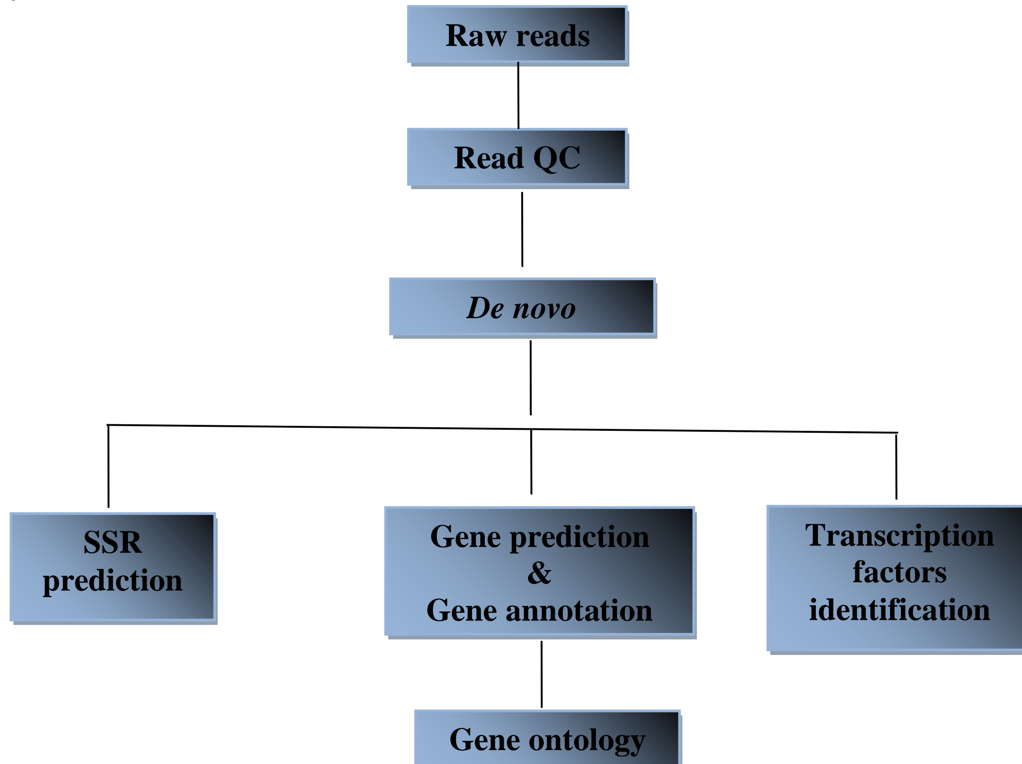
Peak		Size [bp]	Conc. [pg/µl]	Molarity [pmol/l]	Observations
1	◀	35	125.00	5,411.3	Lower Marker
2		302	4,460.77	22,348.8	
3	▶	10,380	75.00	10.9	Upper Marker

Figure 4b: Library quality check on Bioanalyzer of Mango leaf

Sequencing

The DNA library was normalized since it satisfies the Illumina standards for both quantity and quality, which are necessary for additional sequencing. The library was denatured with NaOH, and then 0.2N Tris, pH 7, was added to neutralize the pH levels. The cluster creation and sequencing process involved loading the completed library onto a NextSeq 500 reagent cartridge.

Data analysis



Quality control analysis of raw data

Read quality check

One of the most crucial phases in the pre-processing of sequenced data is the quality inspection of the sequenced raw reads. The goal is to evaluate the quality of the data and exclude low-quality reads by comprehending some pertinent aspects of an ensemble of next-generation sequencing reads, such as length, quality scores, and base distribution.

FASTQC was used to assess the quality of the raw reads from Illumina sequencing for ambiguous bases, Phred score >Q30, read length, nucleotide base content, and other factors. When evaluating the raw reads for quality, the following factors were taken into account.

The HTML report of the FASTQC is given in deliverables under folder FASTQC (Mango_Read1.html, Mango_Read2.html).

Results

Basic statistics

Basic statistics provides the information of file name, file type, encoding which is ASCII encoding of quality values, total number of sequences processed, sequence length which provides the length of the shortest and longest reads, and percentage of GC content (**Table 2**).

Table 2: Basic statistics summary of the raw data

Measure	Read 1	Read 2
Filename	Mango_Read1_25.fastq	Mango_Read2_25.fastq
File type	Conventional base calls	Conventional base calls
Encoding	Sanger / Illumina 1.9	Sanger / Illumina 1.9
Total sequence	18474575	18472256
Sequence length	26-151	26-151
%GC	44	44

The following parameters were considered for the quality check of raw reads

Per base sequence quality
The average and range of the quality scores across all bases are displayed for each base in the sequence quality. The y-axis of the graph is split into three categories: calls with extremely good quality (green), calls with reasonable quality (orange), and calls with bad quality (red). The location in reads is shown by the x-axis, while the per-base Phred scores are represented by the y-axis.

Figure 5 and 6 depict that per base quality reads is good.

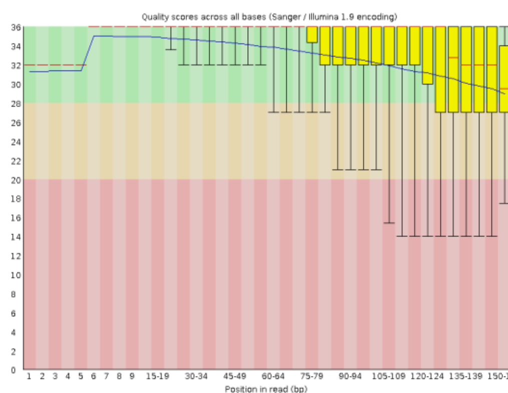


Figure 5: Per base sequence quality of R1

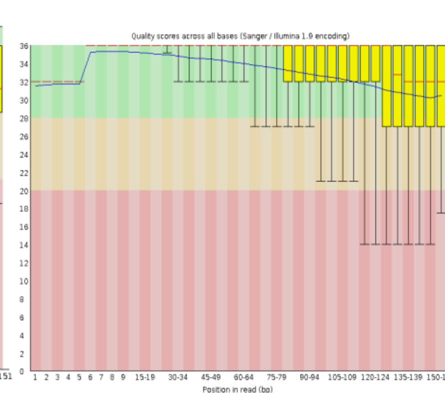
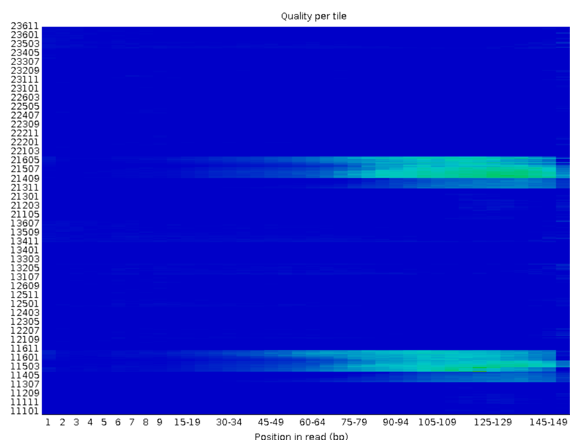


Figure 6: Per base sequence quality of R2

Explanation of Figure:
1. The central red line is the median value
2. The yellow box represents the inter-quartile range (25-75%)
3. The upper and lower whiskers represent the 10% and 90% points
4. The blue line represents the mean quality

Per tile sequence quality

A heat map showing the flow cell quality by tile is called a Per Tile sequence quality. The solid bright blue colour of Figures 7 and 8 indicates that the flow cell tile quality was continuously excellent.



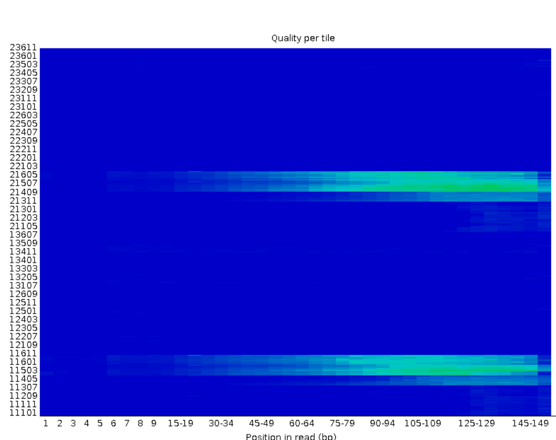
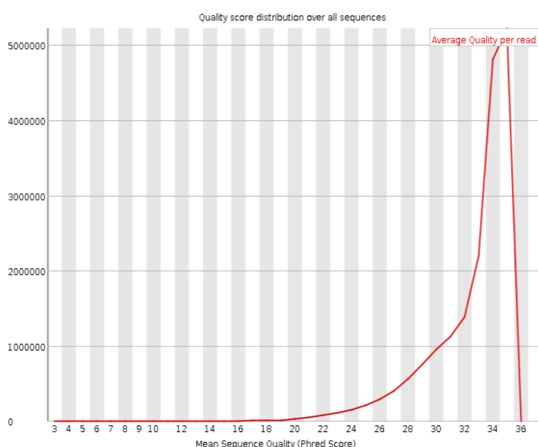


Figure 7: Per tile sequence quality of R1

Figure 8: Per tile sequence quality of R2

Per sequence quality scores

Per sequence quality scores (Phred score) represent the quality of each read and a good sample will have qualities all above 28. In Figure 9 and 10, reads are equally distributed with high quality Phred score between 30-35.



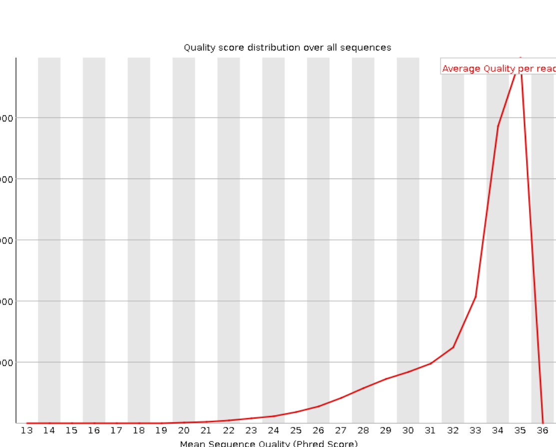
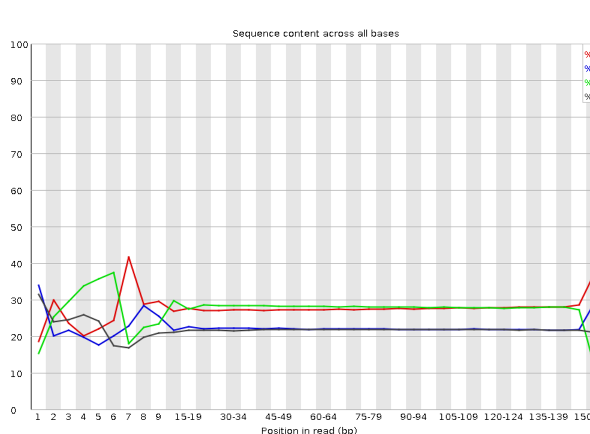


Figure 9: Per sequence quality score of R1

Figure 10: Per sequence quality score of R2

Per base sequence content

Per base sequence content describes the proportion of each bases position in the data for which each of the four normal DNA bases has been called. In Figure 11 and 12, 'A' &'T' and 'G' &'C' bases are being equal with a diverse library and even distribution.



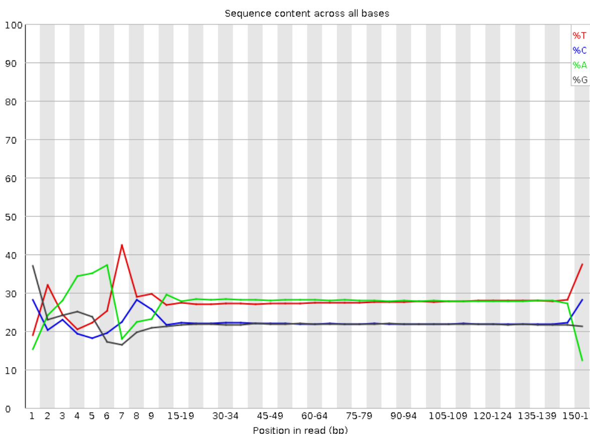


Figure 11: Per base sequence content of R1

Figure 12: Per base sequence content of R2

Per sequence GC content

Per sequence GC content measures the GC content across the whole length of each sequence in a data and compares it to the normal distribution of GC content. In Figure 13 and 14, reads GC count are equally distributed.

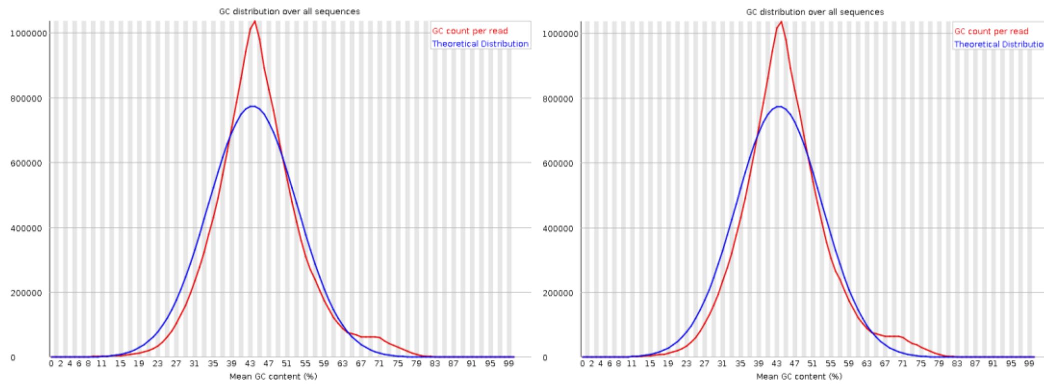


Figure 13: Per sequence GC content R1 **Figure 14:** Per sequence GC content R2

1.1.1 Per base N content

Per base N content plot, describe number of 'Ns' in the sequence. In Figure 15 and 16, the graph represents very low number of Ns at the end of sequence, which does not affect the analysis.

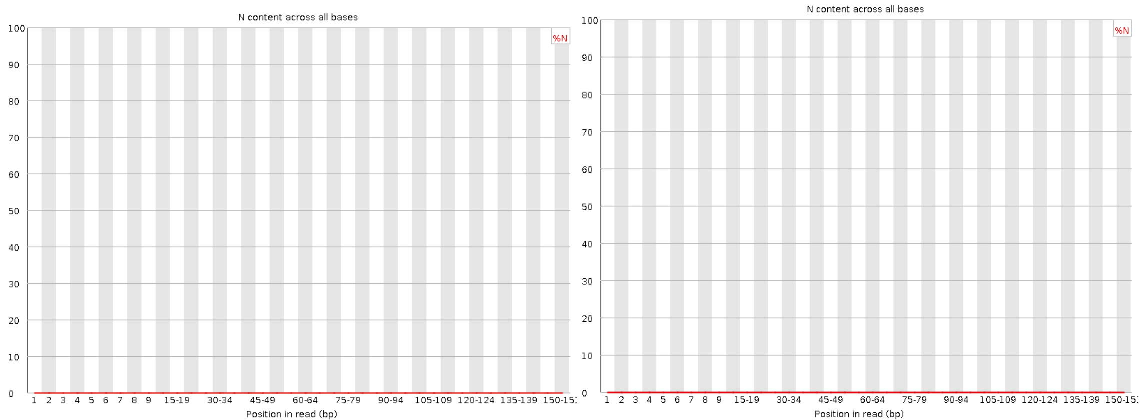


Figure 15: Per base N content R1 **Figure 16:** Per base N content R2

Sequence length distribution

This graph shows the distribution of fragment sizes in the data. In Figure 17 and 18, the graph shows the normal sequence length distribution.

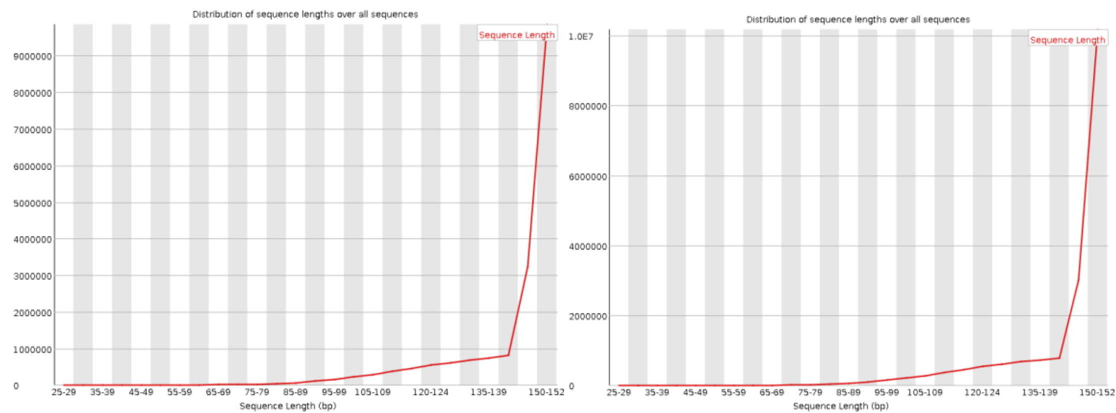


Figure 17: Sequence Length Distribution R1 **Figure 18:** Sequence Length Distribution R2

Sequence duplication levels

A figure illustrating the proportion of sequences with varying levels of duplication is produced by the sequence duplication levels graph, which counts the degree of duplication for each sequence in the set. Transcriptome sequences frequently have high read duplicate rates due to the possibility of a high gene density in a small number of genes. (Figures 19 and 20)

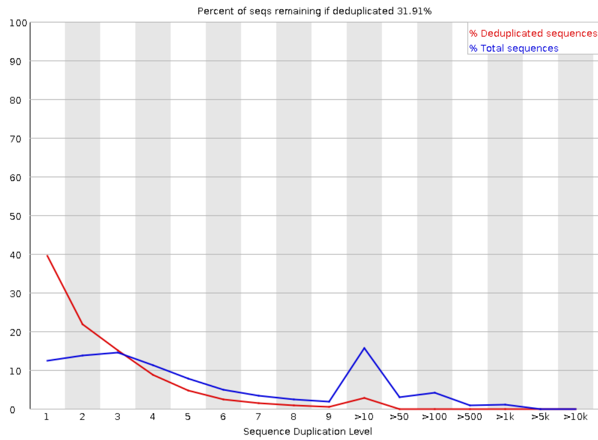


Figure 19: Sequence duplication levels R1

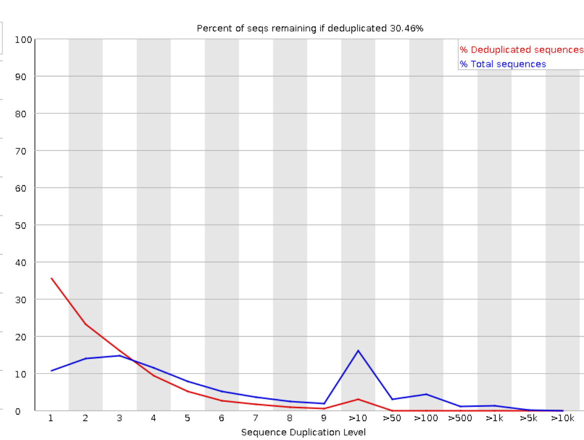


Figure 20: Sequence duplication levels R2

Over represented sequence/Adapter content

There was no adapter contamination in the sequence data output as represented by Figures 21 and 22.

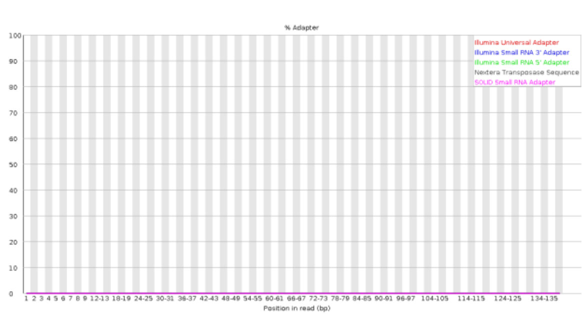


Figure 21`: Adapter content R1

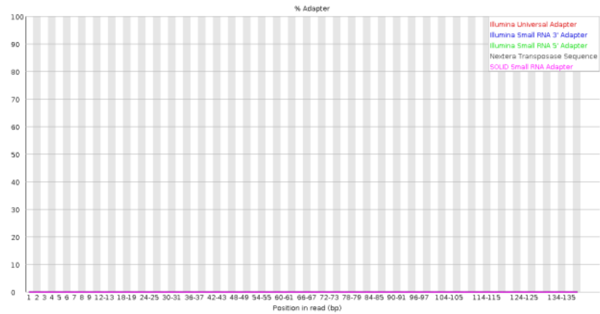


Figure 22: Adapter content R2

De novo Assembly

Using Trinity (version: 2.3.2) with default parameters and a K-mer size of 25 on mango (Mangifera indica) leaves, de novo assembly of high quality reads (Phred score>30) resulted in 93,300 transcripts clustered into 60,874 unigene with a GC percentage of 42.74. In order to eliminate any unimportant shorter transcripts, transcripts longer than 300 bp were kept for additional analysis (empirically calculated using our read length). The assembly statistics of the transcripts from mango leaves are compiled in Table 3. The Mango_transcripts.fasta deliverable contains the transcript sequence files.

Table 3: Assembly statistics of Mango leaf

Parameters	**Statistics**
Counts of transcripts, etc.	
Total trinity unigenes	60,874
Total trinity transcripts	93,300
Percent GC	42.74
Stats based on ALL transcript contigs	

Contig N10	2,382
Contig N20	1,887
Contig N30	1,583
Contig N40	1,326
Contig N50	1,096
Median contig length	577
Average contig	819.05
Total assembled bases	76,417,406

Gene prediction and Functional annotation

Gene prediction is the process of determining the DNA sequence that will encode for a protein, or a gene, as well as the gene's beginning and ending positions. Transdecoder (version: 3.0.1) was utilized in conjunction with internal tools to find potential coding areas in the transcripts that were produced. To reduce the likelihood of false positives during ORF predictions, an open reading frame that was at least 100 amino acids long was sought after.

BLASTP, Pfam, KEGG, COG, and BLASTX (version 2.2.31+) were used to annotate the transcripts containing ORFs. For the purpose of gene ontology, only transcripts relevant to plant species were retrieved. Table 4 summarizes the gene annotation findings for the assembled transcripts from Mango leaves using various techniques. The Mango_Annotation.xls delivery contains the annotation files.

Table 4: Annotation summary of Mango leaf sample

Parameters	
No. of transcripts	93,300
No. of ORFs	97,342
No. of transcripts annotated using BLASTX	80,770
No. of transcripts annotated using BLASTX only plant species	54,011
No. of plant annotated transcripts with Gene Ontology terms	43,958
No. of transcripts with ORFs annotated using BLASTP only for plant species	36,839
No. of transcripts with ORFs annotated using PFAM only for plant species	29,640
No. of transcripts with ORFs annotated using KEGG only for plant species	39,996
No. of transcripts with ORFs annotated using COG only for plant species	39,388

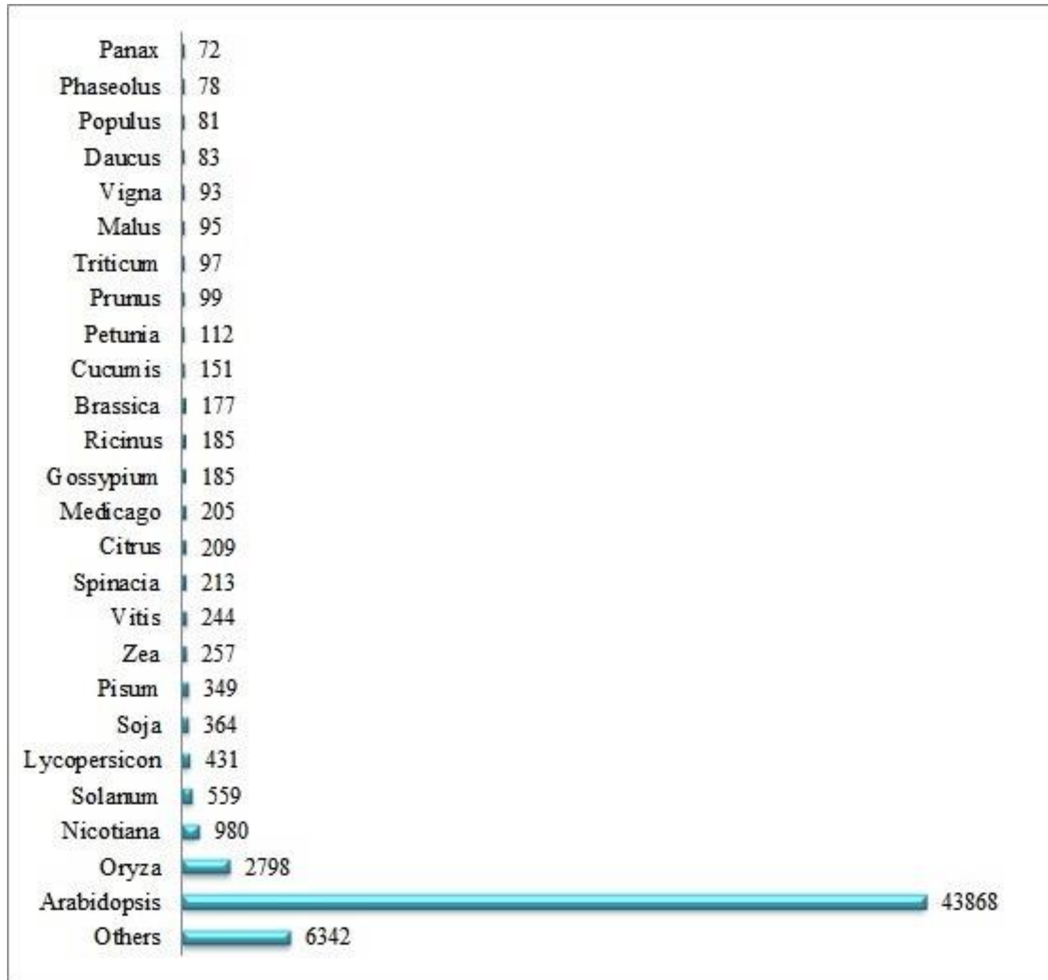


Figure 23: Plant species contributing for the functional annotation of Mango leaf assembled transcripts

Gene Ontology

According to Figure 22, the Mango leaf gene ontology study identified 86,873, 1,13,448 and 92,930 genes that are correspondingly linked to biological processes, cellular components, and recognized molecular functions. Figures 23, 24, and 25 show, respectively, the breakdown of proteins related to different cellular components, biological processes, and molecular functions.

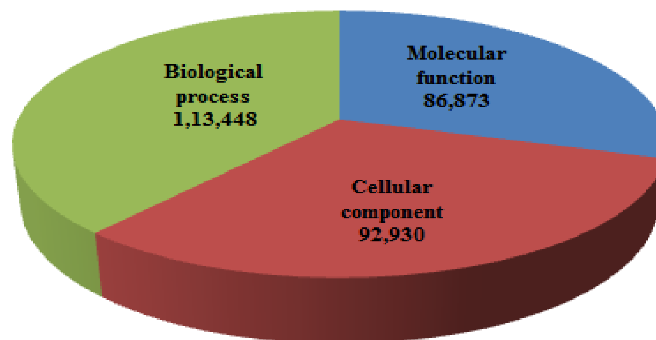


Figure 24: Genes associated with the molecular functions, biological process and cellular components

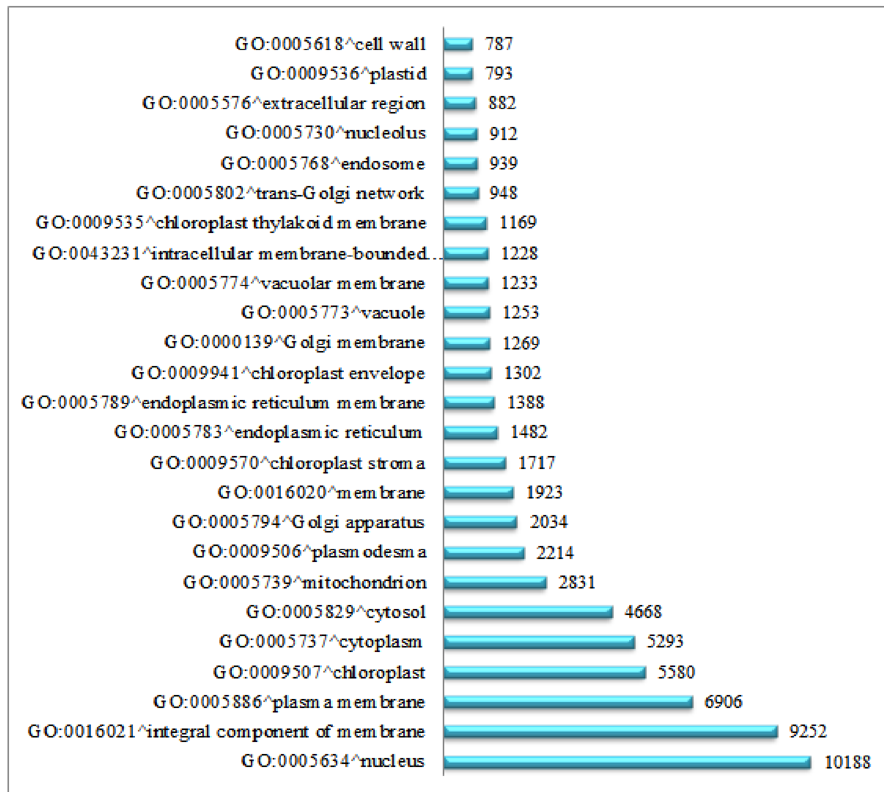


Figure 25: Break down of proteins associated with various cellular components

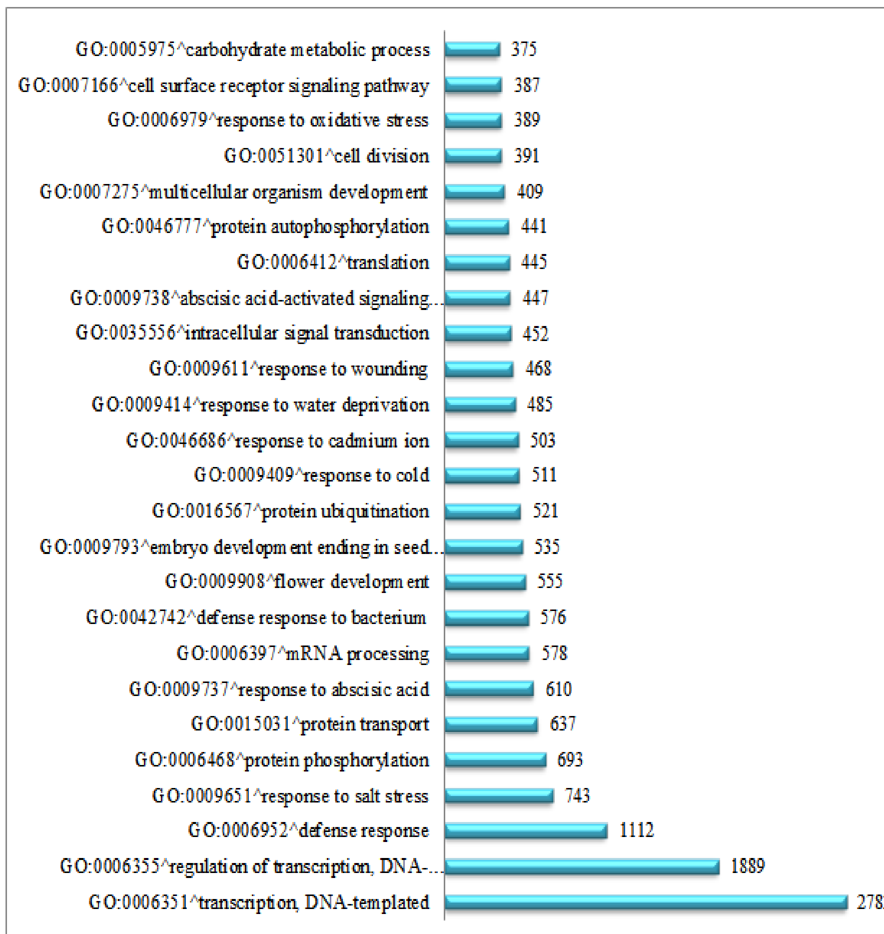


Figure 26: Break down of proteins associated with various biological processes

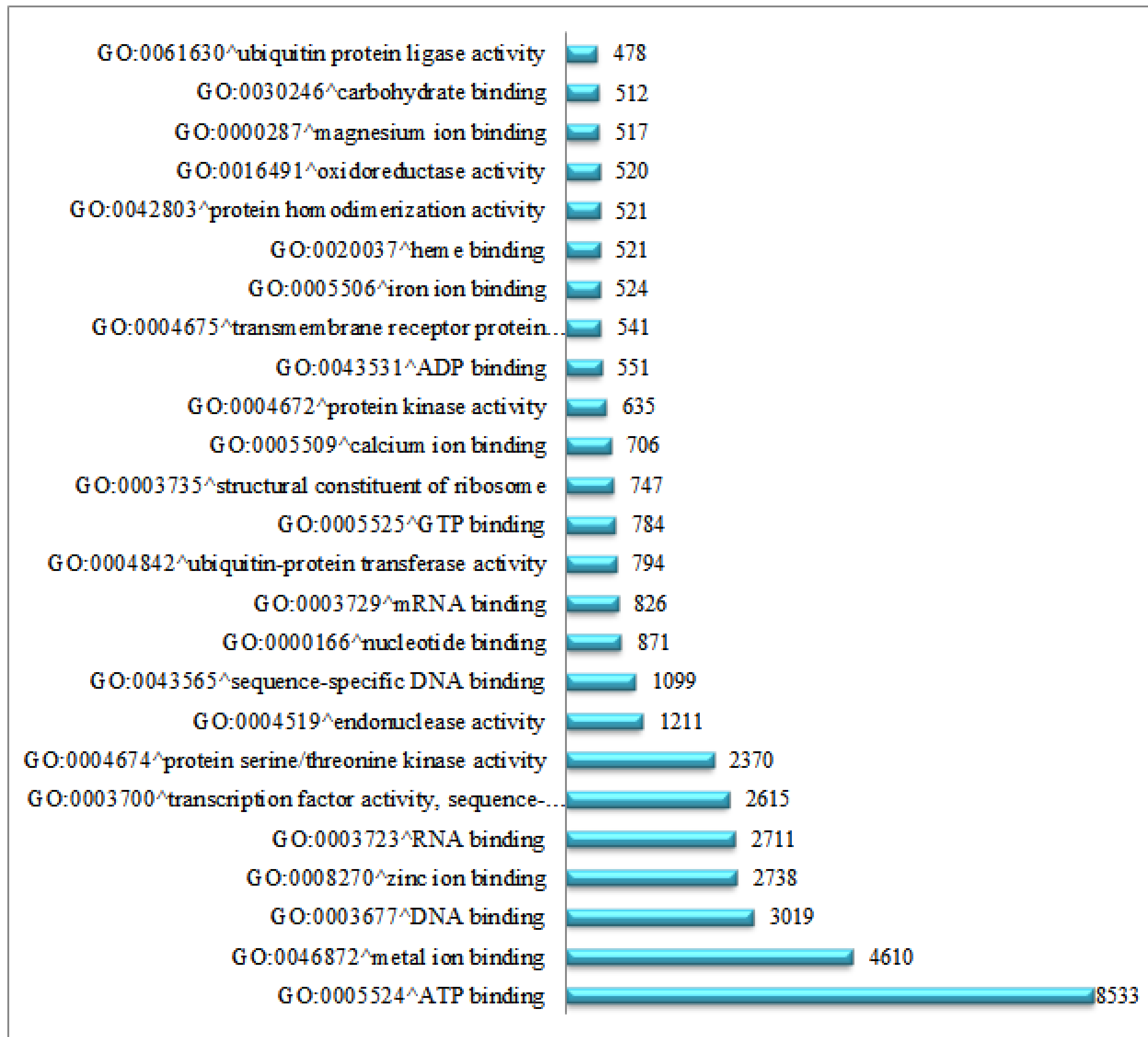


Figure 27: Break down of proteins associated with various molecular functions

Simple Sequence Repeats (SSRs) Prediction

Short repeat motifs with high levels of length variability resulting from insertion or deletion mutations of one or more repeat types are known as simple sequence repeats (SSRs) or short tandem repeats. Using MISA and internal tools, we examined the abundance of microsatellites, or SSRs, in annotated plant transcripts for mango leaf samples. The expected SSRs statistics are summarised in Tables 5 and 6. The transcript files for the various repetition type classes are distributed in detail and are included in the deliverables under the filename "Mango_SSR_results."

Table 5: Prediction of MICROSATELLITE for annotated transcripts

Description	
Total number of annotated transcripts	54,011
Total size of annotated transcripts(bp)	63,262,416
Number of SSRs identified	10,120
Number of annotated transcripts containing SSR	7,988
Number of annotated transcripts containing more than 1 SSR	1,612
Number of SSRs present in compound formation	1,171

Table6: Distribution of different repeat type classes

Repeat Type	
mono repeats	6,684
di repeats	1,626
tri repeats	1,655
tetra repeats	105
penta repeats	35
hexa repeats	15

Transcriptional factors identification

Transcription factors control the patterns of gene expression, which in turn dictate various biological processes. With PlantTFDB, transcription factors were found (version v4.0). The distribution of the transcription factors that have been found is summarized in Table 7, and the transcription factor family category in the mango leaf transcript is summarized in Table 8.

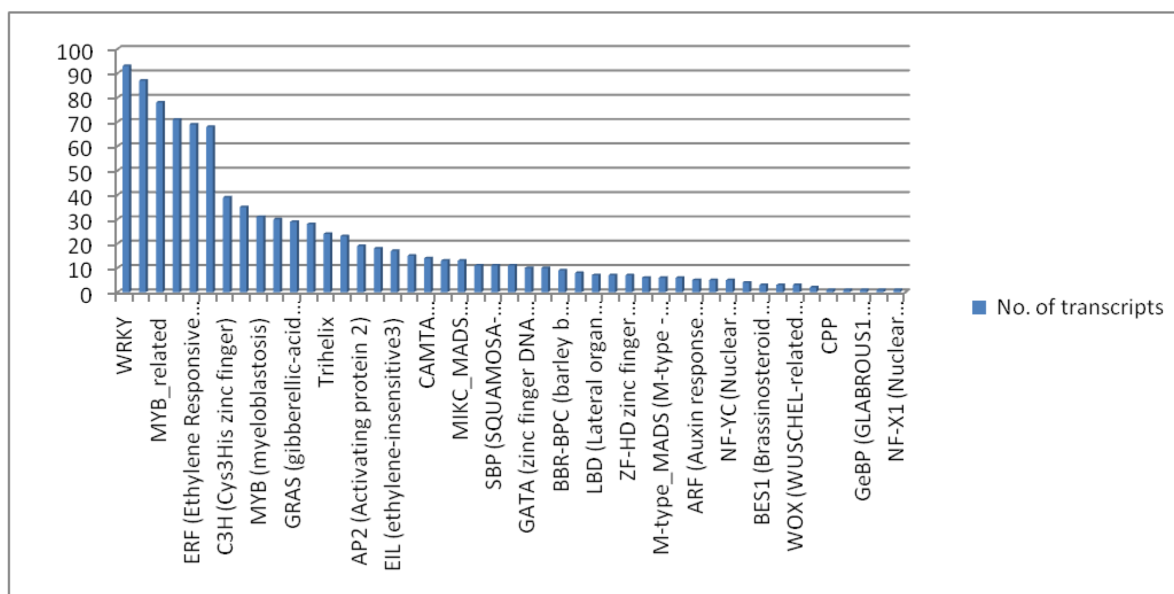
Table 7: Summary of transcription factors identified in Mango leaf

Description	
No. of assembled transcripts	54,011
No. of transcripts with transcription factor	976
No. of unique transcription factor in the leaf sample	44

Table 8: Transcriptional factors family identified in Mango leaf

Transcriptional factors family	No. of transcripts
AP2 (Activating protein 2)	19
ARF (Auxin response factors)	5
B3 (B3 DNA binding domain)	23
BBR-BPC (barley b recombinant/ basic pentacysteine)	9
BES1 (Brassinosteroid insensitive1-ethyl methanesulfonate-suppressor1)	3
bHLH (Basic helix-loop-helix proteins)	87
bZIP (Basic leucine-zipper superfamily)	71
C2H2 (Cys2–His2 zinc fingers)	35
C3H (Cys3His zinc finger)	39
CAMTA (atcamta5, calmodulin-binding transcription activator)	14
CO-like	6
CPP	1
DBB (double B-box)	5
Dof (DNA-binding One Zinc Finger)	8
E2F/DP (E2 factor /Dimerization Partner)	1

EIL (ethylene-insensitive3)	17
ERF (Ethylene Responsive Factor)	69
FAR1 (Fatty Acyl-CoA Reductase 1)	11
G2-like (Golden2-like)	30
GATA (zinc finger DNA binding proteins)	10
GeBP (GLABROUS1 enhancer-binding protein)	1
GRAS (gibberellic-acid insensitive (GAI), Repressor of GAI (RGA) and Scarecrow (SCR))	29
HB-other (homeodomain)	13
HD-ZIP (Homeodomain-leucine zipper)	28
HRT-like	1
HSF (heat shock factors)	18
LBD (Lateral organ boundaries domain)	7
LSD (lysergic acid diethylamide)	4
MIKC_MADS (MADS, Intervening (I), a Keratin-like (K) and a C-terminal domain)	13
M-type_MADS (M-type - MCM1, AGAMOUS, DEFICIENS, SRF)	6
MYB (myeloblastosis)	31
MYB_related	78
NAC (no apical meristem)	68
NF-X1 (Nuclear Transcription Factor, X-Box Binding 1)	1
NF-YA (Nuclear transcription factor Y subunit Alpha)	18
NF-YB (Nuclear transcription factor Y subunit Beta)	10
NF-YC (Nuclear transcription factor Y subunit Gamma)	5
Nin-like	2
SBP (SQUAMOSA-pROMOTER BINDING PROTEIN)	11
TALE (Transcription Activator-like Effectors)	15
TCP TEOSINTE BRANCHED 1, CYCLOIDEA, PCF1)	11
Trihelix	24
VOZ (Vascular plant One-Zinc finger)	7
Whirly	3
WOX (WUSCHEL-related homeobox)	3
WRKY	93
YABBY	6
ZF-HD zinc finger homeodomain)	7



Gene Expression Analysis: Considering Major Genes

Song, M. *et al.* (2023) in a review discussed detail findings on transcriptomics of mango, which are mostly oriented around the flower and fruit development. Little is known about the transcriptome analysis of vegetative stage.

Many studies have been conducted on the function of WRKY proteins in plant signalling, especially in regard to their participation in biotic stressors such pathogen assault. Recent discoveries, however, have also highlighted the critical function that these proteins play in initiating reactions to abiotic stressors. WRKY proteins are important transcription factors that have the ability to control gene expression either positively or negatively. Researchers have genetically altered some WRKY factors to increase plant resistance to abiotic stress after identifying their role in the production of stress-responsive genes. To properly target these genes, one must have a thorough understanding of the signalling pathways that result in the activation of WRKY proteins, their interactions with other signalling proteins, and the downstream genes they activate. Additionally, WRKY proteins have been found to confer tolerance to multiple abiotic stresses and potentially play a role in facilitating communication between abiotic and biotic stress responses.

For correct leaf and flower patterning, border development and lateral organ separation are controlled by the CUP-SHAPED COTYLEDON (CUC)/NO APICAL MERISTEM (NAM) gene family. Nevertheless, little is known about these genes' downstream targets. We detected a weak allele of the no-apical-meristem mutant mtnam-2 by a forward screen of the tobacco retrotransposon1 (Tnt1) insertion population in Medicago truncatula. We were also able to successfully regenerate a mature plant from the null allele mtnam-1. These resources offered a framework for in-depth analysis of MtNAM's and its downstream genes' functions. According to our research, MtNAM is strongly expressed in the apices of vegetative shoots and inflorescences, particularly where the shoot apical meristem and leaf/flower primordia meet.

Additionally, both the weak allele and the regenerated null allele showed clear floral abnormalities, including decreased floral whorls and organ counts as well as degraded floral organ identification. We discovered that all classes of floral homeotic genes are down-regulated in mtnam mutants using microarray and quantitative RT-PCR investigations. Mutations in MtNAM also caused a faulty shoot apical meristem and merged cotyledons and leaflets of the complex leaf. These results show that MtNAM is necessary for correct floral organ identification and development, and that it functions similarly to other CUC/NAM family genes in lateral organ separation and compound leaf development. Myeloblastosis (MYB)-related proteins, a broad subfamily of transcription factors that are involved in plant development and stress responses, comprise the MYB family. The promoter regions of BnMYB-related genes showed an enrichment of several cis-acting regulatory elements linked to stress response, phytohormone response, and development when analysed in silico.

Gibberellic acid (GA) signals are transduced by the phytohormone receptor GoGID1, and these signals are critical for many aspects of plant development. It has been demonstrated that ectopic

expression in tobacco promotes elongation and boosts biomass production. It was also found that in the presence of GA, GoGID1 can interact with the DELLA proteins in Arabidopsis. Additionally, it was discovered that GoGID1 overexpression can improve GA sensitivity in transgenic plants and restore the phenotypes of the Arabidopsis double-mutants atgid1a/atgid1c and atgid1a/atgid1b. Furthermore, GoGID1's diverse role in plant growth and development was demonstrated by the enhanced plant height, biomass, and leaf size of transgenic Arabidopsis and lucerne plants over expressing it.

Discussion

Mangos are a tropical fruit with a rich genetic resource, and their use in breeding programmes and related research has become essential. Mango research is making great strides in the fields of breeding and genetic enhancement, which could boost the industry and public acceptance of mangoes. New genetics that are more suited to production settings, more aesthetically pleasing to consumers, and protected by plant variety rights, patents, and trademarks are the main forces behind these advancements. Mangos have extensive intra-specific variability, with about 1600 cultivars cultivated worldwide. India has the greatest collection of mango cultivars, with over 1600 varieties farmed in all states of the nation. Every cultivar has distinct flavour, texture, taste, and size features.

However, when it comes to the core needs of crop improvement—variability, identification, collection, and conservation of germplasm—the entire potential of India's mango gene pool is still virtually untapped. On their farms, custodian farmers have been instrumental in maintaining, advancing, and modifying a variety of native mango cultivars. These farmers look after and preserve India's unique mango heritage, acting as protectors. A varied and well-maintained variety of variations and thorough and well-documented information on the genetic material that is now available are essential for breeding efforts (Dey *et al.*, 2020; Dutta *et al.*, 2020; De *et al.*, 2014; 2019, 2020a; 2020b; 2021a; 2021b; 2021c). By identifying particular genes, transcriptome analysis can be a useful technique for Marker Assisted Selection (MAS).

Conclusion

Technological developments in sequencing have made it possible to decipher the genome of the mango and analyze transcriptomes during the vegetative stage. This has yielded important insights into genetic diversity, variety identification, and the control of biological processes. Key regulatory genes for significant qualities like disease resistance will be identified in the future with the use of pan-genome construction, genetic and natural population analysis, and the utilization of multiomics data. This will have a significant positive impact on the speed and accuracy with which new, premium mango varieties can be developed, with an emphasis on enhancing tissue culture regeneration and gene editing technologies. Mango breeding, which has substantial economic significance in tropical and subtropical areas, could be revolutionized by these initiatives.

Acknowledgement

The authors thank the Principals of Rammohon College, St. Xavier's College (Autonomous) and Gurudas College for their support. The authors give thanks to the departmental colleagues who encouraged the authors to carry on this investigation.

Conflict of Interest

The authors declare that there are no conflicts of interest regarding the publication of this work.

References

Andrews, Simon.(2010). "FastQC: a quality control tool for high throughput sequence data.":175-176.*(Available from https://www.bioinformatics.babraham.ac.uk/projects/fastqc/)*

Arumuganathan, K., & Earle, E. D. (1991). Nuclear DNA content of some important plant species. *Plant molecular biology reporter*, *9*, 208-218. https://doi.org/10.1007/BF02672069

Azim, M. K., Khan, I. A., & Zhang, Y. (2014). Characterization of mango (*Mangifera indica* L.) transcriptome and chloroplast genome. *Plant molecular biology*, *85*, 193-208. https://doi.org/10.1007/s11103-014-0179-8

Chow, C. N., Zheng, H. Q., Wu, N. Y., Chien, C. H., Huang, H. D., Lee, T. Y., ... & Chang, W. C. (2016). PlantPAN 2.0: an update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. *Nucleic acids research*, *44*(D1), D1154-D1160. https://doi.org/10.1093/nar/gkv1035

De, M., Das, J., Saha, M., Pal, A., & Dey, S. R. (2014). Documentation and Characterization of the Indigenous Mango (Mangifera indica L.) varieties of Malda and Murshidabad districts of West Bengal. *J. Environ. & Sociobiol*, *11*(2), 187-198.

De, M., Dutta, S., Pahari, G. K., Ganguli, S., Ray, S., & Dey, S. R. (2019). Variation in flowering and fruit set among indigenous mango varieties from Malda and Murshidabad conserved as Field Gene Bank in Gurudas College, Kolkata, India. *International Journal of Advancement in Life Sciences Research*, *2*(4)48-61. https://doi.org/10.31632/ijalsr.2019v02i04.006

De, M., Dutta, S., Ray, S., & Dey, S. R. (2021b). Visual Clustering Analysis of some traditional Mango (*Mangifera indica* L.) varieties of Murshidabad District, West Bengal using Clust Vis web tool. *International Journal of Advancement in Life Sciences Research*, *4* (3), 42-43. https://doi.org/10.31632/ijalsr.2019v02i04.006

De, M., Pahari, G., Pal, A., Ganguli, S., Ray, S., Mandal, M., Dhar, S., Dutta, S., & Dey, S. (2019). Mango (Mangifera indica L.) field gene bank (FGB) in Gurudas College, Kolkata-54. *Int. J. Exp. Res. Rev.*, 18, 6-14.

De, M., Ray, S., Dutta, S., & Dey, S. R. (2021). Eugenol Found by GC-MS Analysis of The Methanolic Extract of The Fruit Pulp of Indigenous Mango (*Mangifera indica* L.) Variety, Champa from Murshidabad District of West Bengal, India. *International Journal of Advancement in Life Sciences Research*, 4(1), 22-30. https://doi.org/10.31632/ijalsr.20.v04i01.004

De, Mitu and Dey, S. R., **(**2021c). Assessment of the pulp quality of some rare indigenous mango (Mangifera indica L.) varieties of West Bengal, India. *International Journal of Experimental Research and Review, 26*, 70-78. https://doi.org/10.52756/ijerr.2021.v26.005

Dey, S. R., Dutta, S., & De, M. (2020). Neuro-Steroid Compound Found by GC-MS Analysis of The Methanolic Extract of The Leaves of *Aegle marmelos* (L) Corr. *International Journal of Advancement in Life Sciences Research*, 3(4), 51-56. https://doi.org/10.31632/ijalsr.20.v03i04.006

Dutta, S., De, S. R., & De, M. (2020). Characterisation of the mango (*Mangifera indica* L.) fruit and fruit pulp of some indigenous mango varieties of Murshidabad district of West Bengal. *International Journal of Advancement in Life Sciences Research*, 3(2), 23-28. https://doi.org/10.31632/ijalsr.20.v03i02.003

Gómez-Ollé, A., Bullones, A., Hormaza, J.I., Mueller, L.A. and Fernandez-Pozo, N. (2023). MangoBase: A Genomics Portal and Gene Expression Atlas for Mangifera indica. Plants, *12*(6), 1-12. https://doi.org/10.3390/plants12061273

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... & Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology*, *29*(7), 644-652. https://doi.org/10.1038/nbt.1883

Haas, B., & Papanicolaou, A. J. G. S. (2016). TransDecoder (find coding regions within transcripts). *https://transdecoder.github.io*

Iyer, C. P. A., & Dinesh, M. R. (1997). Advances in classical breeding and genetics in mango. *Acta Hortic*, *455*, 252-267. https://doi.org/10.17660/ActaHortic.1997.455.34

Iyer, C. P. A., & Subramanyam, M. D. (1991). Breeding mango for developing new varieties. *Acta Hortica*, *291,* 151-153. https://doi.org/10.17660/ActaHortic.1991.291.17

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic acids research*, *36*(2), W5-W9. https://doi.org/10.1093/nar/gkn201

Kaur, H., Sidhu, G. S., Mittal, A., Yadav, I. S., Mittal, M., Singla, D., ... & Chhuneja, P. (2023). Comparative transcriptomics in alternate bearing cultivar Dashehari reveals the genetic model of flowering in mango. *Frontiers in Genetics*, *13*, 1-13. https://doi.org/10.3389/fgene.2022.1061168

Kuhn, D. N., Bally, I. S., Dillon, N. L., Innes, D., Groh, A. M., Rahaman, J., ... & Sherman, A. (2017). Genetic map of mango: a tool for mango breeding. *Frontiers in plant science*, *8*, 1-11. https://doi.org/10.3389/fpls.2017.00577

Ledesma, N., & Campbell, R. J. (2019). The status of mango cultivars, market perspectives and mango cultivar improvement for the future. *Acta horticulturae*, (1244), 23-27. https://doi.org/10.17660/ActaHortic.2019.1244.3

Mahato, A. K., Sharma, N., Singh, A., Srivastav, M., Jaiprakash, Singh, S. K., ... & Singh, N. K. (2016). Leaf transcriptome sequencing for identifying genic-SSR markers and SNP heterozygosity in crossbred mango variety 'Amrapali'(*Mangifera indica* L.). *PloS one*, *11*(10), 1-16. https://doi.org/10.1371/journal.pone.0164325

Mukherjee, S. K. (1953). The mango—its botany, cultivation, uses and future improvement, especially as observed in India. *Economic botany*, *7*(2), 130-162. https://doi.org/10.1007/BF02863059

Mukherjee, S. K., & Litz, R. E. (2009). Introduction: botany and importance. In *The mango: Botany, production and uses*, *2*,.1-18.

Nashima, K., Terakami, S., Kunihisa, M., Nishitani, C., Shoda, M., Matsumura, M., ... & Yamamoto, T. (2017). Retrotransposon-based insertion polymorphism markers in mango. *Tree genetics & genomes*, *13*, 1-10. https://doi.org/10.1007/s11295-017-1192-2

Peng, L., Gao, W., Song, M., Li, M., He, D., & Wang, Z. (2022). Integrated metabolome and transcriptome analysis of fruit flavor and carotenoids biosynthesis differences between mature-green and tree-ripe of cv."Golden Phoenix" mangoes (Mangifera indica L.). *Frontiers in plant science*, *13*, 1-15. https://doi.org/10.3389/fpls.2022.816492

Sagar, V. R., Khurdiya, D. S., & Balakrishnan, K. A. (1999). Quality of dehydrated ripe mango slices as affected by packaging material and mode of packaging. *Journal of food science and technology (Mysore)*, *36*(1), 67-70.

Shudo, A., Tarora, K., Makishi, Y., Ichi, R., Takahashi, K., Matsumura, M., ... & Urasaki, N. (2013). Development of CAPS markers and their application in breeding for mango, Mangifera indica L. *Euphytica*, *190*(3), 345-355. https://doi.org/10.1007/s10681-012-0774-1

Sivankalyani, V., Sela, N., Feygenberg, O., Zemach, H., Maurer, D., & Alkan, N. (2016). Transcriptome dynamics in mango fruit peel reveals mechanisms of chilling stress. *Frontiers in plant science*, *7*, 1-17. https://doi.org/10.3389/fpls.2016.01579

Song, M., Wang, H., Fan, Z., Huang, H., & Ma, H. (2023). Advances in sequencing and key character analysis of mango (*Mangifera indica* L.). *Horticulture Research*, *10*(2), 1-16. https://doi.org/10.1093/hr/uhac259

Subramanyam, H., Krishnamurthy, S., & Parpia, H. A. B. (1975). Physiology and biochemistry of mango fruit. *Advances in Food Research*, *21*, 223-305. https://doi.org/10.1016/S0065-2628(08)60092-0

Tafolla-Arellano, J. C., Zheng, Y., Sun, H., Jiao, C., Ruiz-May, E., Hernández-Oñate, M. A., ... & Tiznado-Hernández, M. E. (2017). Transcriptome analysis of mango (*Mangifera indica* L.) fruit epidermal peel to identify putative cuticle-associated genes. *Scientific reports*, *7*(1), 1-13. https://doi.org/10.1038/srep46163

Tharanathan, R. N., Yashoda, H. M., & Prabha, T. N. (2006). Mango (*Mangifera indica* L.),"The king of fruits"—An overview. *Food Reviews International*, *22*(2), 95-123. https://doi.org/10.1080/87559120600574493

Thiel, T. (2003). MISA—Microsatellite identification tool. *Website http://pgrc.ipk-gatersleben.de/misa/ [accessed 17 June 2016]*.

Tsai, C. C., Chen, Y. K. H., Chen, C. H., Weng, I. S., Tsai, C. M., Lee, S. R., ... & Chiang, Y. C. (2013). Cultivar identification and genetic relationship of mango (*Mangifera indica*) in Taiwan using 37 SSR markers. *Scientia Horticulturae*, *164*, 196-201. https://doi.org/10.1016/j.scienta.2013.09.037

Viruel, M. A., Escribano, P., Barbieri, M., Ferri, M., & Hormaza, J. I. (2005). Fingerprinting, embryo type and geographic differentiation in mango (*Mangifera indica* L., Anacardiaceae) with microsatellites. *Molecular breeding*, *15*, 383-393. https://doi.org/10.1007/s11032-004-7982-x

Wang, P., Luo, Y., Huang, J., Gao, S., Zhu, G., Dang, Z., ... & Chen, Y. (2020). The genome evolution and domestication of tropical fruit mango. *Genome biology*, *21*, 1-17. https://doi.org/10.1186/s13059-020-01959-8

Yamanaka, S., Hosaka, F., Matsumura, M., Onoue-Makishi, Y., Nashima, K., Urasaki, N., ... & Yamamoto, T. (2019). Genetic diversity and relatedness of mango cultivars assessed by SSR markers. *Breeding science*, *69*(2), 332-344. https://doi.org/10.1270/jsbbs.18204