# GENOMIC AND TRANSCRIPTOMIC ANALYSIS OF *FICUS RELIGIOSA* REVEALS INSIGHTS INTO CARBON FIXATION PATHWAY AND NEURAL DISEASE PATHWAYS

**A THESIS TO BE SUBMITTED TO**
**THE UNIVERSITY OF TRANS-DISCIPLINARY HEALTH SCIENCES AND TECHNOLOGY**

THE UNIVERSITY OF TRANS-DISCIPLINARY HEALTH SCIENCES & TECHNOLOGY

**FOR THE AWARD OF THE DEGREE OF DOCTOR OF PHILOSOPHY**

**BY**

**ASHALATHA K L**
**REG. NO. – 20318020203**

**UNDER THE GUIDANCE OF**

**DR. MALALI GOWDA**
**Professor**

**JANUARY 2024**

The University of Trans-Disciplinary Health Sciences and Technology (TDU),
Post Attur via Yelahanka, Jarakabande Kaval, Yelahanka Bengaluru - 560064

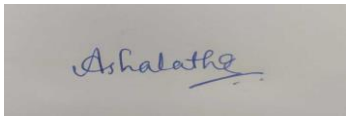
**THE UNIVERSITY OF TRANS-DISCIPLINARY HEALTH SCIENCES AND TECHNOLOGY**

**Private University Established in Karnataka by ACT 35 of 2013**
**BENGALURU - 560064**

## DECLARATION BY THE CANDIDATE

I declare that this thesis entitled **"Genomic and Transcriptomic Analysis of *Ficus religiosa* reveals insights into carbon fixation pathway and neural disease pathways"** submitted for the award of Doctor of Philosophy to THE UNIVERSITY OF TRANS-DISCIPLINARY HEALTH SCIENCES AND TECHNOLOGY (TDU), Yelahanka, Bengaluru, is my original work, conducted under the supervision of my guide Dr. Malali Gowda. I also wish to inform that no part of the research has been submitted for a degree or examination at any university. References, help and material obtained from other sources have been duly acknowledged.

I hereby confirm the originality of the work and that there is no plagiarism in any part of the dissertation.



**Place: Bengaluru** **Signature of the Candidate**

**Date: 09-01-2024**

**Name of candidate: Ashalatha K L**
**Reg. No.: 20318020203**
**Ph.D Student**
**The University of Trans-Disciplinary Health Sciences and Technology (TDU), Yelahanka**
**Bengaluru - 560064**

**THE UNIVERSITY OF TRANS-DISCIPLINARY HEALTH SCIENCES AND TECHNOLOGY**

**Private University Established in Karnataka by ACT 35 of 2013**
**BENGALURU - 560064**

# CERTIFICATE

This is to certify that the work incorporated in this thesis **"Genomic and Transcriptomic Analysis of *Ficus religiosa* reveals insights into carbon fixation pathway and neural disease pathways"** submitted by **Ashalatha K L** was carried out under my supervision. No part of this thesis has been submitted for a degree or examination at any university. References, help and material obtained from other sources have been duly acknowledged. I hereby confirm the originality of the work and that there is no plagiarism in any part of the dissertation.

**Research Supervisor**

**Date: 09-01-2024**

**Dr. Malali Gowda,**

**Professor**

**The University of Trans-Disciplinary Health Sciences and Technology (TDU), Yelahanka**

**Bengaluru - 560064**

# ACKNOWDLEGEMENT

Foremost, I would like to express my gratitude to my supervisor Dr. Malali Gowda, for the continuous support for my Ph.D study and research. I would like to thank my doctoral advisory committee members Dr. Arun Kumar K. P. Scientist at Central Muga Eri Research and Training Institute (CMER&TI), Assam and Dr. M. N. Balakrishnan Nair, Emeritus Professor at The University of Trans-Disciplinary Health Sciences and Technology (TDU), Bengaluru, for their encouragement, remarkable comments and suggestions.

My sincere thanks goes to Tata Education and Development Trust fellowship, who provided fund support to me for 2 years during my Ph.D study. I would like to acknowledge the Next Generation Genomics Facility at the Centre for Cellular and Molecular Platforms (C-CAMP) and Bengaluru Genomics Center Pvt. Ltd for support in sequencing the plant samples. I would like to thank Dr. Noorunnisa Begum, Associate Professor at TDU for providing the microscope lab facility for the stomatal work. I thank my fellow labmates Mr. Ravindra Raut, TDU lab member, who accompanied me for collecting the leaf samples. I thank Mr. Patturaju, TDU lab member, who helped me for stomatal work.

I thank my friends and other lab members in TDU University: Annapoorna Prabhu, Divya, Manjunath R, Dr. Niranjan, Dr. Lavanya, Dr. Prassanna Koti, Dr. Venu S, Dr. Sheetal Ambardar, Puja S, Rashmi, Bhanushree, Sachin, Rakshith, Hemanth, Santosh Hegde and Dr. Pavithra N for interesting discussions, working together and memories gathered in those working days. I extend my heartfelt thanks to my friends in life, teachers, family members, brothers and well-wishers who supported and guided me at every stage through motivation, encouragement and helped to grow in life and career.

Last but utmost, my parents for their love, prayers, caring and sacrifices for bringing me up with education and values. I am very much thankful to my husband for his love, understanding and supportive for my research work. Finally, I thank the God for his showers of blessings to complete this work successfully

# DEDICATION

I would like to dedicate this thesis to my beloved parents and my husband who have supported me throughout my education and career.

I dedicate to those committed researchers in the hope that this work will contribute in some way to the exploration of genomics by current and future researchers in biotechnology and life sciences.

## CONTENTS

## List of tables

### Chapter III



### Chapter IV



## List of text data

### Chapter III

# List of figures

## Chapter II



## Chapter III

**Chapter IV**

## Additional files

1. Additional file 1: Text file S1.1 - https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-023-09270-z/MediaObjects/12864_2023_9270_MOESM8_ESM.txt

2. Additional file 2: Text file S1.2 - https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-023-09270-z/MediaObjects/12864_2023_9270_MOESM9_ESM.txt

3. Additional file 3: Table S1 - https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-023-09270-z/MediaObjects/12864_2023_9270_MOESM11_ESM.txt

4. Additional file 4: Table S2 - https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-023-09270-z/MediaObjects/12864_2023_9270_MOESM17_ESM.txt

5. Additional file 5: Table S3 - https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-023-09270-z/MediaObjects/12864_2023_9270_MOESM18_ESM.xls

6. Additional file 6: Table S4 - https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-023-09270-z/MediaObjects/12864_2023_9270_MOESM21_ESM.txt

7. Additional file 7: Table S5.1 - https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-023-09270-z/MediaObjects/12864_2023_9270_MOESM22_ESM.txt

8. Additional file 8: Table S5.2 - https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-023-09270-z/MediaObjects/12864_2023_9270_MOESM23_ESM.txt

9. Additional file 9: Table S6 - https://static-content.springer.com/esm/art%3A10.1186%2Fs12864-023-09270-z/MediaObjects/12864_2023_9270_MOESM27_ESM.txt

## Abbreviations

WGS: Whole Genome Sequencing

BUSCO: Benchmarking Universal Single-Copy Orthologous

ESTs: Expressed Sequence Tags; PPR repeat: Pentatricopeptide repeat

FPKM: Fragments Per Kilobase of transcript per Million mapped reads

MSA: Multiple Sequence Alignment

BWA - Burrows-Wheeler Aligner

GO terms - Gene ontology terms

Pfam - protein family

KEGG - Kyoto Encyclopedia of Genes and Genomes

LTR - Long terminal repeats

GGAT - glutamate-glyoxylate aminotransferase enzyme

PEPC - PEP carboxylase enzyme

MDH - Malate dehydrogenase

CAM pathway - Crassulacean Acid Metabolism pathway

C3 pathway - Calvin-Benson cycle pathway

C4 - Dicarboxylic pathway

# SYNOPSIS

Indian subcontinent is rich in plant biodiversity and associated with traditional medicinal practices. *Ficus religiosa* (Peepal tree) is a key stone species, huge semi-evergreen and lives for thousands of years. It belongs to the Moraceace family and to the groups of fig trees. The various parts of *F. religiosa* like bark, leaves, fruits and latex are used in traditional medicine to treat different disease conditions. *F. religiosa* is known for the Buddha's meditation and enlightenment and popularly known as Bodhi Tree.

Peepal tree is generally believed to be unique in photosynthesis pathways in plant Kingdom, it believed to produce oxygen during night and day throughout its life cycle. In addition, its stomata are sunken and giant or hydathode, much larger than the normal stomata where it can store more water molecules. However, limited molecular studies have been carried out on *F. religiosa* to understand its molecular and physiological aspects.

In this study, we aimed to establish the molecular data of *F. religiosa*, using next generation sequencing technology. We have carried-out the whole genome sequencing and transcriptomic sequencing and bioinformatics analyses. We sequenced the entire genome of the fig species *F. religiosa* using two next-generation technology platforms that use shotgun sequencing. Based on ab initio, homology, and mRNA evidence utilized for annotation, the assembled genome resulted in a size of 406 Mb and 35,093 protein-coding genes. The repetitive sequences made up around 53% of the entire genome. The gene expression patterns of the carbon fixation pathway were revealed by RNA sequencing from leaf samples collected at day and night period (2 PM and 2 AM, respectively). As a result, it was shown that the genome of Peepal is closely related to the genus Ficus (*F. carica* and *F. macrocarpa*), and relatively close to the genus *M. notabilis* in the same family. This study has provided the annotated genes and pathways related to carbon fixation, neural disease pathways, molecular, metabolic and physiological pathways.

## List of publications

1. Ashalatha, K. L., Arunkumar, K. P., & Gowda, M. (2023). Genomic and transcriptomic analysis of sacred fig (Ficus religiosa). *BMC genomics*, *24*(1), 1-13.

2. Ashalatha, K. L., & Gowda, M. (2019). Heritage of Neem–Peepal Tree Resides a Profound Scientific Facts. *The Neem Genome*, 13-19.

# CHAPTER I

# GENERAL INTRODUCTION

## I.1 Peepal/Buddha tree (*Ficus religiosa*)

Ficus (Moraceae) is considered as one of the largest genera of angiosperms, comprising species of trees, hemi-epiphytes, shrubs, climbers and creepers in the tropical and subtropical regions throughout the world [1]. Some of the fig species are *F. religiosa, F. racemosa, F. carica, F. benjamina, F. glomerata, F. benghalensis, F. maxima, F. sycomorus,* etc. Ficus species have important value in several areas such as religious, ecological, evolutionary, and medicinal. The co-evolution of *Ficus* and their pollinator fig wasps is an example to show obligate mutualism [2]. Longevity is an attractive feature of exclusive plants that have a lifespan of hundreds to thousands of years. Interestingly, *F. benghalensis* and *F. religiosa* species possess longevity charaters and not any other members from this genus [3].

The Peepal tree (*Ficus religiosa* L.) is a sacred fig hemi-epiphyte belonging to the family Moraceae, with a diploid sporophyte chromosome count (2n = 26) [4]. It is known to be a keystone species, long-lived deciduous tree species related to the ~755 to 800 fig species found worldwide [5]. Generally, the Peepal tree is thought to produce oxygen day and night, which is why it holds a special significance in communities across India. They have a special type of stomata, called sunken, giant, or hydathode at the lower leaf epidermis. These are larger than the normal stomata and either overlie veins or are mixed with normal stomata. Such stomata have been shown to retain gas and water molecules for a longer [6]. As far as we know, there is no scientific evidence that Peepal trees produce oxygen at night.

### I.2 Cultural and Traditional practices associated with Peepal tree

Indian subcontinent is rich in plant biodiversity and associated with traditional medicinal practices. The Peepal tree is a cosmopolitan species, having value for cultural and spiritual practices in Buddhism, Hinduism, and Jainism. It is popularly called the Bodhi tree, where Buddha is believed to have meditated and attained spiritual enlightenment underneath this tree. Hence, the culture is spread across Asia and it has been worshipped. Peepal has several vernacular names, like Pippali, Ashwatha, Arali, and so on; it is frequently found together with the Neem tree near Indian temples [7].

### I.3 Peepal tree in Ayurveda

In Ayurveda, the Peepal tree has been classified as a Rasayana (a type of drug), whereby rejuvenators and antioxidants aid in relieving the body's stress [8]. Peepal tree alleviates Pitta and Kapha (Ayurvedic classifications), hence prescribed for treatment of the disorders like respiratory and inflammatory disorders, ulcers, stomatitis, hiccup, arthritis, gout, skin diseases, bone fracture, diabetes, etc., [8]. The various parts of *F. religiosa* like bark, leaves, fruits and latex are used in traditional medicine to treat different disease conditions like intestinal worms, neuro-disorders, dysentery, mumps, jaundice, heart diseases, diabetes, constipation, skin diseases, etc.

In Ayurveda, *F. religiosa* belong to a class of drug called rasayana, which are rejuvenators, antioxidants that relive body stress. The Ayurvedic properties of *F. religiosa* include Rasa: kashaya (astringent), Guna: guru (heavy), ruksha (dry), Veerya: shita (cold) and Vipaka: katu (pungent). It has coloring or pigmenting (varnya) action, ability to arrest pain (vedanasthaapana), remove edematous swellings (shothahara) and conserves blood (raktasamgrahaka) [30]. The different parts of the *F. religiosa* tree such as, leaves, latex, bud,

stem, bark, roots, flowers, seeds, fruits have been used extensively in the traditional medicines. Peepal tree is of great Ayurvedic medicine for diabetes (Vd.7.16), (Ss.Ci.11.9), rheumatoid arthritis (Cs.Ci.29.158), asthma, inflammatory Disorders (Ah.U.25.29), wound healing (Cs.Ci.25.95, 113), burnt healing (Vm.45.21), severe colic (Sb.4.506), splint in fractures (Ss.Ci.3.6), relieving pain (Ss.U.21.9-10), infectious and sexual disorders (Ss.Ci.26.27). The bark forms an important ingredient of many Ayurvedic formulations, like "PanchaValkaladiTailum" this oil contains *F. religiosa*, *F. benghalensis* L., *F. glomerata* Roxb., *F. infectoria* Willd., *Azadirachta indica* A. Juss., *Curcuma longa* L. and *Hemides musindicus* R. Br. and "PanchaValkalaKashaya" (decoction containing *F. religiosa*, *F. benghalensis, F. glomerata*, *F. infectoria* and *A. indica* used for skin diseases to balance Pitta and Kapha (Panda, 2005; Singh and Panda, 2005). *F. religiosa* mollifies Pitta and Kapha, hence prescribed as an Ayurvedic remedy for the disorders associated with their imbalance like, respiratory disorders, ulcers, stomatitis, hiccup, arthritis, gout, skin diseases, allergies, inflammatory disorders, bone fracture, diabetes, etc [8]. More scientific studies and evidences are still required to prove it detail.

**Some of the Ayurveda Shlokas describing the Peepal tree**

|| **कैयदेवनघिण्टु** ||

अश्वत्थः शीतलो रूक्षः कषायो दुर्जरो गुरुः ||४३२||
व्रणपित्तकफास्रघ्नो वर्ण्यो योनविशोधनः |४३३|

अश्वत्थः शीतलो रूक्षः - Ashvattha tree has coolant property
कषायो दुर्जरो गुरुः - Having an astringent taste and difficult to digest
व्रणपित्तकफास्रघ्नो - It is used to treat the ulcers and wounds. And also balances the pitta, kapha doshas [types of prakriti (nature) classified in ayurveda]
वर्ण्यो - It gives the good complexion to skin and improves the skin color.
योनविशोधनः - It also cures the female reproductive diseases

||**राजनघिण्टु**||

पप्पिलः सुमधुरस्तु कषायः शीतलश्च कफपित्तविनाशी |
रक्तदाहशमनः स हि सद्यो योनिदोषहरणः कलि पक्वः ||११४||
अश्वत्थवृक्षस्य फलानि पक्वान्यतीवहृद्यानि च शीतलानि |
कुर्वन्ति पित्तास्रविषार्तिदाहवच्छ्रर्दिशोषारुचिदोषनाशम् ||११५||

सुमधुरस्तु - sweet in taste
रक्तदाहशमनः - Useful for bleeding disorder
सद्यो योनिदोषहरणः - Quick relieve from vaginal and urinary tract problems
कलि पक्वः - Bark is used
फलानि पक्वान्यतीवहृद्यानि - Fruits are tasty and cold
पित्तास्रविषार्तिदाहवच्छ्रर्दिशोषारुचिदोषनाशम् - Removes toxic and poisonous condition. And it relieves burning, nausea, weakness anorexia and cures skin diseases

(Source: e-Nighantu (Collection of Āyurvedic Lexicons), Central Council for Research in Ayurvedic Sciences (CCRAS), New Delhi. http://niimh.nic.in/ebooks/e- Nighantu/ )

## I.4. Overview on studies of Peepal tree

In animal models such as rats, the Peepal tree has been tested for the treatment of neurodegenerative disorders such as Parkinson's disease and Huntington's disease [9] [10], as well as anti-ulcer activity in albino mice [11].

Some of the previous studies shows that *F. religiosa* extracts are effective in treating neurological disorders:

1. *F. religiosa* has been used traditionally for the treatment of neurodegenerative disorders including Parkinson's disease (PD) and also been reported to possess antioxidant

activity, this plant may prove to be effective in the remedy of PD. Hence, it was evaluated for anti-Parkinson's effect using neurotoxin induced Parkinson's model in rats. The anti-Parkinson's activity of petroleum ether extract of *F. religiosa* leaves was studied in haloperidol and 6 hydroxydopamine (6-OHDA) induced experimental animal models. The research showed the effects of *F. religiosa* using *in vivo* behavioral parameters like catalepsy, muscle rigidity, and locomotor activity and its effects on neurochemical parameters (MDA, CAT, SOD, and GSH) in rats [9].

2. Several plants have been reported in the literature that depicts AChE inhibitory activity. The most accepted strategy for the management of Alzheimer's disease is the use of acetyl cholinesterase (AChE) inhibitors. Here they have selected 37 traditional Indian medicinal plants (including the stem bark of *F. religiosa*) known to be useful in treating cognitive decline, improving memory or related CNS activities, and screened for their AChE inhibitory activity [12].
3. In Huntington's disease, the study showed that the 14 days administration of 3-nitropropoionic acid (NP) increased the oxidative stress in the brain. It significantly increased the level of lipid peroxidation and depletion of endogenous antioxidants enzyme such as catalase, superoxide dismutase and reduced glutathione. Treatment with higher dose of petroleum ether extract of *F. religiosa* (400 mg/kg) resulted in a decreased level of Malondialdehyde (MDA) and increased levels of Superoxide dismutase (SOD), Catalase and reduced Glutathione (GSH) that indicate antioxidant effect in the brain of 3 NP treated animals [10].
4. The study has shown the administration of the *Ficus racemosa* bark extract elevated acetyl choline (Ach) levels and improved memory in rats. The collective pharmalogical actions attributed by *F. racemosa* extract may serve as beneficial and supporting agent in the Alzheimer's Disease [13].

5. *F. religiosa* fig extract (FRFE) resulted in a significant improvement of memory, as its treatment attenuated the scopolamine-induced anterograde and retrograde amnesia dose-dependently. Further, cyproheptadine pre-treatment significantly reversed the anti-amnesic effect of FRFE. FRFE has anti-amnesic activity against scopolamine-induced amnesia, in a dose-dependent manner. Inhibition of the anti-amnesic effect of FRFE by cyproheptadine substantiates the involvement of serotonergic pathways for its activity [14].
6. The molecular mechanism of methanol extract of *Ficus religiosa* leaf (MFL)-mediated attenuation underlies the down-regulation of extracellular signal-regulated kinase (ERK), c-Jun N-terminal kinase (JNK) and p38 mitogen-activated protein kinase (MAPK) signaling pathway, and suppresses the nuclear factor kappaB (NF- κ B) activation. The results suggest that MFL exhibits anti-inflammatory properties in lipid polysaccharide (LPS)-induced activation of BV2 microglial cells, and that might have a therapeutic potential for various neurodegenerative diseases [15].

**I.5. Genomic resources of Ficus species and other non-model plant species.**

Next-generation sequencing (NGS) technologies have accelerated the generation of draft genome sequences of Moraceae plant species. The genome size of *Morus notabilis* is 330 Mb [16], 333 Mb in *Ficus carica* [17], 436 Mb in *F. microcarpa* and 370 Mb in *F. hispida* [2]. The genome sequencing of non-model plant species Peepal was first mentioned in The Neem Genome book chapter [7]. A recent study has generated the genomic resource of *F. religiosa* (332 Mb) and *F. benghalensis* (392 Mb). However, they generated a limited size of genome assembly and genes (23,929) for the Peepal tree when compared to the present study [18]. Recently, a few research groups have attempted sequencing of non-model plant species like pineapple (*Ananas comosus*) and *Kalanchoë* species revealing the gene expressions of the

Crassulacean acid metabolism (CAM) pathway [19] [20]. The whole-genome sequencing has shown common or crystalline ice plants (*Mesembryanthemum crystallinum*) to switch from Calvin- Benson Cycle (C3) to CAM photosynthesis under a salt stress [21]. The study described by comparing both species with and without the C4 trait and different tissues within a C4 plant using RNA-seq suggests ways of integration into the underlying C3 metabolism [22]. These findings and other physiological features of the Peepal tree enabled us to characterize the pathways in the present work.

### I.6 Research objectives of the present study

Despite its ecological, medicinal, cultural, and historic importance, the molecular biology and genomics studies on the Peepal tree are scanty. As Peepal is relevant to traditional medicinal practices and Buddha's meditation, we envisaged elucidating the genome sequence and studying the transcriptome of photosynthetic tissue (leaf tissue) in diurnal and nocturnal periods. The objective of the present study was to generate a genome sequence and annotate genes of the Peepal tree. The transcriptomic analysis has been undertaken to identify the expression of genes in the diurnal and nocturnal periods for photosynthetic activity using a molecular approach. In this study, we aimed to characterize the genes involved in various physiological, biochemical metabolic, and other pathways. Also, a comparative genomic analysis has been carried out to study the relationship of the Peepal tree with closely related species of its Moraceae family.

# CHAPTER II

# MATERIALS AND METHODS

## II.1 Collection of leaf samples and extraction of nucleic acids

The mature leaves were collected from a cultivated Peepal tree (15 years old) at a Private property, Anuganalu village, Hassan District, India (13.0647° N, 76.0363° E). We have followed a non-invasive method for collecting leaf samples. Genomic DNA was extracted from the leaves using the Qiagen DNeasy Plant Mini kit (Catalog #69106), and the quality and quantity of DNA were confirmed using the Nanodrop. From the same Peepal tree, the leaf samples were collected and immediately placed on dry ice during the day (2 PM) and night (2 AM) periods. Total RNA was isolated from the leaf samples using the QiagenRNeasy Plant Mini kit (Catalog #74904) method and was treated with RNase-free DNase I (Catalog #M0303S) from New England BioLabs for 30 min at 37 °C to remove residual DNA. RNA integrity and quantity were confirmed on Qubit and Tape station using a dsDNA HS (Catalog #32854) kit from Invitrogen and RNA screen tape from Agilent respectively. The sample collection for DNA seq and RNA seq is shown in the Figure II.1A, B.

**A)** **B)**

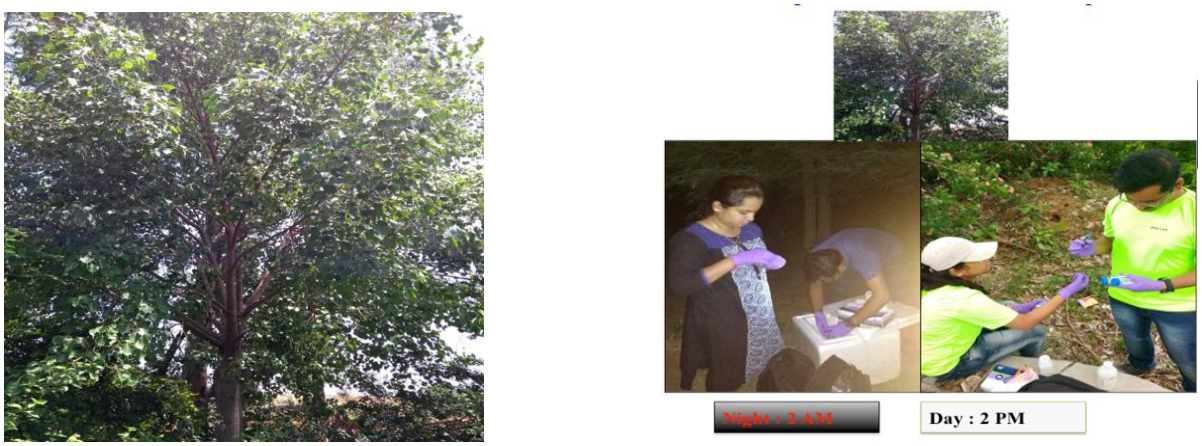
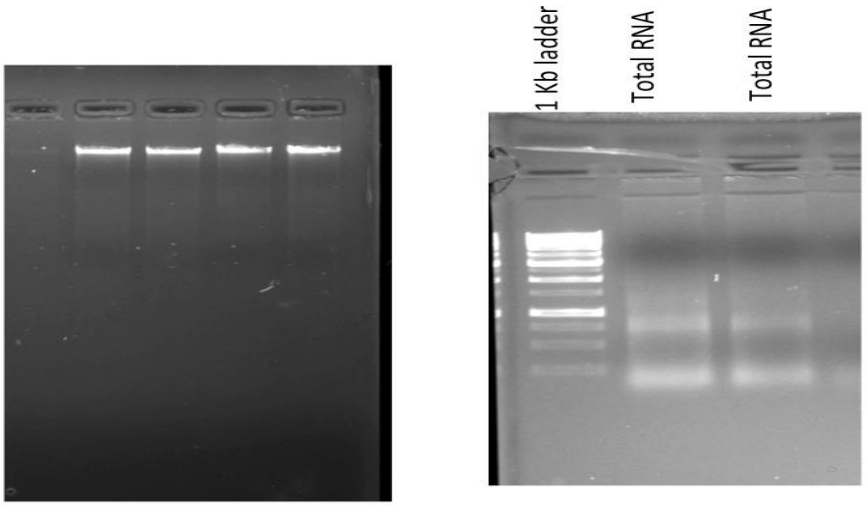



**Figure II.1:** The images shows A) Leaf sample collection for DNA Seq from the Peepal tree B) Leaf sample collection for RNA Seq from the Peepal tree at 2:00 AM and 2:00 PM.

The gel electrophoresis image of the isolated DNA and RNA from the Peepal tree leaf sample were shown in the Figure IV.2 A, B.




**A)** **B)**

**Figure II.2:** A) Agarose gel electrophoresis image of DNA of *F. religiosa* (B) Agarose gel electrophoresis image of RNA of *F. religiosa*.

## II.2 DNA and RNA library preparation and sequencing

We applied the two next-generation technology platforms having shotgun sequencing method, to sequence the whole-genome of fig species *F. religiosa*. Whole-genome shotgun DNA library preparation was performed using the Illumina TrueSeq DNA sample preparation kit (FC-121-2001). The paired-end (PE) (2 x 100 nts) sequencing was carried out using Illumina HiSeq-1000. Also, to increase the size of genome data, we sequenced the genome with paired-end (PE) (2 x 100 nts) using the MGISEQ-2000 platform.

The RNA libraries were prepared using “TruSeq RNA Library Prep Kit v2 from Illumina®” with Illumina standardized protocol. The RNA libraries were quantified on Qubit (dsDNA HS kit) and validated on the TapeStation instrument (D1000 screen tape). These RNA libraries were used for sequencing with the Illumina HiSeq-2500 platform.

### II.3 Genome size estimation and assembly

Each of the Illumina and MGISEQ-2000 raw reads were processed for a quality check using the FastQC v0.11.6 tool [8]. Then filtering and trimming of raw reads were done to remove the low complexity bases using the TrimGalore-0.4.5 (https://www.bioinformatics.babraham.ac.uk/projects/trimgalore/) and reads having quality value Q>20 and length above 20 bases were taken for constructing the assembly. To estimate the genome size, filtered reads were taken for the k-mer distribution (different k-mers from 21 to 77) and abundance analysis using Jellyfish v1.1.12 [23] and GenomeScope v2.0 [24]. The separate Illumina and MGI Seq generated raw reads were used to construct the assembly using the tools SPAdes-3.13.0 [25] and MaSuRCA-3.2.9 [26] respectively. The parameters were the default k-mer sizes of 21, 33, and 55 for Illumina assembly. The constructed assemblies were used to build the super scaffolds using the tool SSPACE standard v3.0 [27].

The combined Illumina HiSeq and MGISEQ raw reads were used to construct the hybrid assembly using the assembler SPAdes-v3.13.0 [25]. The parameters were the default k-mer sizes 21, 33, and 55, with a 77 mer also set. The gaps in the assembly were closed by GMcloser-1.6.2 [28]. The assembly statistics were obtained using the tool Quast v4.6.1 [29]. The completeness and evaluation of the assembly were done by BUSCOv3 tool [30] with the Embryophyte and Eukaryota database and by aligning the RNA-seq reads to the genome.

### II.4 Structural gene prediction and functional annotation

Peepal tree assembled scaffolds were processed for structural and functional gene annotation using the MAKER-P v.2.31.10 software [31]. The RNA-sequenced data of *Morus notabilis* [16] consists of expression sequence tags (ESTs) and the GFF (Gene finding format) file which

contains the gene features and structures of genes, protein data of *A. thaliana* and RNA-sequence data of Peepal tree were imported as evidence for annotation support. The structural and functional annotation of predicted genes and proteins was performed using BLASTP in the Uniprot database. The protein family, structures, and gene ontology (GO) terms were identified for protein-coding genes using InterProScan-V5.27-66.0 [32].

**II.5 Gene family construction, identification of homologous and orthologous genes**

Protein sequences of *A. thaliana, M. notabilis, P. persica, C. sativa, Z. jujuba,* and the protein sequences of the current study *F. religiosa* were taken for the homologous and orthologous gene identification. The homologous genes were identified in *F. religiosa* proteome sequences using the BLASTP analysis against the other 5 proteomes of *A. thaliana, M. notabilis, P. persica, C. sativa,* and *Z. jujuba*. OrthoVenn2 [33] was used to cluster orthologous genes and identify the single-copy orthologous genes in all six proteomes. Further, these single-copy orthologous genes were used for constructing the phylogenetic tree using the tool MAFFT-v7 [34].

**II.6 Comparative genome analysis**

We downloaded the genomes of *F. carica, F. microcarpa,* and *M. notabilis*. We aligned these genomes against the Peepal tree genome assembly to understand their relationships using the BWA-V0.7.17 (Burrows-Wheeler Aligner) [35] and Samtools v1.7 [36].

**II.7 Prediction of repetitive elements: TEs and SSR**

The RepeatModeller-open-1.0.11and RepeatMasker-4.0 tools were used for repeat library building and repeat identification in the assembly respectively. The MicroSAtellite identification tool (MISA) [37] was used for the identification of SSRs from assembled genome

sequences of *F. religiosa*. The parameters were set to identify perfect di-, tri-, tetra-, penta-, and hexa nucleotide motifs with a minimum threshold of 6, 5, 5, 5, and 5 repeats, respectively.

### II.8 Prediction of transcription factor families

The families of transcription factors (TFs) were predicted in genome annotations and differentially expressed transcripts of the Peepal tree using Plant Transcription Factor Database v5.0 [38].

### II.9 Non-coding RNA genes

The transfer RNAs in the Peepal tree genome were found using tRNAscan-SE (v2.0.3) [39] with the 'eukaryotes' option. tRNAscan-SE software deployed with the covariance models identifies the primary sequence and secondary structure information of tRNA and gives the complete tRNA genes for the query genome and transcriptome sequences. tRNAscan-SE software is integrated with Infernal v1.1 to enhance the tRNA search with better covariance and other updated models. Using the isotype-specific covariance model provides the functional classification of tRNAs and in the first pass search cutoff score, 10 is set. The miRbase database (http://www.mirbase.org) was used for the identification of putative miRNAs in the genome and unique identified transcripts sequence data based on the homology search. The long non-coding RNAs (lncRNAs) were identified with the Coding Potential Calculator tools [40].

### II.10 Transcriptome sequencing, assembly, and annotation

High-quality stranded RNA sequencing (ssRNA-seq) reads were assembled into putative transcripts using Trinity v2.9.0 [41]. Assembled transcripts were passed through Transdecoder v5.02 [42] to predict the coding sequences. The transcripts were clustered to find the unigenes by removing the redundant transcripts using the tool CD-HIT-est v.0.0.1 [43] with a 95%

sequence identity threshold. Transcripts assembled from Trinity and CD-HIT-v0.0.1 were used in downstream analyses for gene prediction. Unigenes were used to predict the putative genes using the NCBI non-redundant (nr) database using the BLASTX program and proteins were predicted from the Uniprot database using the BLASTP program. The Trinity assembled transcripts were annotated using Trinotate- V3.11. The raw reads were mapped to scaffold assembled genome using Cufflinks-v2.2.1 [44] and considered as reference assembly.

### II.11 Transcript quantification and differential gene expression analysis

The estimation of transcripts abundance was determined using RNA-Seq by Expectation-Maximization (RSEM) tool [45], which quantifies transcript level abundance from RNA-seq data. RSEM first generates and pre-processes a set of reference transcript sequences and then aligns reads to reference transcripts followed by an estimation of transcript abundances. Normalized transcripts obtained from the transcript quantification methods were used in the next step for the differential gene expression analysis. FPKM and Trimmed Mean of M-values (TMM) are calculated to understand the expression levels of genes in day and night samples of the Peepal tree. For further analysis, the gene expression was estimated using FPKM and TMM value minimum $\geq 1$. The TMM value was used to cluster the genes according to their expression pattern using the edgeR package in the R tool. The parameters used in the differential expression analysis were a probability value P-value of 0.001 and a fold change value of log2. The expression value was also determined for assembled transcripts to verify the expression of genes predicted from gene models. The differentially expressed genes were annotated using BLAST2GO Annotation software [46].

### II.12 Pathway Analysis

The annotated genes from the assembled genome and the differentially expressed genes from the Peepal tree leaf tissues collected during the day (2 PM) and night (2 AM) were used for pathway analysis in the KAAS (KEGG Automatic Annotation Server) (KEGG) server [47] using the BBH (bi-directional best hit) method and the search against a default set of 40 eukaryotic organisms. It provided the list of pathways where the candidate genes were mapped based on the orthologous homology alignment.

# CHAPTER III

# RESULTS

## III.1 *De novo* hybrid assembly using Illumina and MGI short reads

We used two next-generation technology platforms to sequence the whole genome of the fig species, the Peepal tree. A total of 266 and 645 million paired-end reads were generated from Illumina HiSeq1000 and MGISEQ-2000 platforms respectively. The data of 88.44 billion high-quality bases (Quality>20) was used for genome assembly.

**Table III.1:** Details on raw sequence data of *F. religiosa* genome and transcriptome

| **Sample (Period of collection)** | **Data type** | **Library type** | **Sequencer** | **Insert length (base)** | **Number of reads (paired-end)** | **Number of bases (paired-end)** |
|---|---|---|---|---|---|---|
| Leaf | DNA Seq | Paired-end | Illumina HiSeq | 387-390 | 266,593,648 | 23,943,191,355 |
| Leaf | DNA Seq | Paired-end | MGISEQ-2000 | 409-420 | 645,008,220 | 64,500,822,000 |
| | | | | **Total** | 911,601,868 | 88,444,013,355 |
| Leaf (Day -2 PM) | RNA Seq | Paired-end | Illumina HiSeq | 250-500 | 56,472,586 | 5,613,340,767 |
| Leaf (Night -2 AM) | RNA Seq | Paired-end | Illumina HiSeq | 250-500 | 54,997,562 | 5,469,938,094 |
| | | | | **Total** | 111,470,148 | 11,083,278,861 |

A hybrid assembly was performed using a sequencing depth of 65.5X Illumina reads and 158.86X MGI reads. The raw data details are given in the Table III.1. The evaluation of the distribution of k-mers in both Illumina and MGI reads to estimate the genome size provided

genome sizes of 319 Mb and 273 Mb respectively (Figures III.1A and III.1B). The combination of Illumina and MGI reads was used for assembling the genome. Hybrid genome assembly yielded a genome of 406 Mb. The contig N50 length is 5,817 bp and the largest contig length is 148 Kb. The GC content of the Peepal tree genome is 34.23%. The gap-closing step was performed for the hybrid assembly. There were 35,811 (5.5%) misassembled contigs and 604,807 (94.4%) truly assembled contig sequences in the final assembled genome.

**Table III.2:** Final assembly and annotation of Peepal genome

| **Features** | **Final assembly** |
|---|---|
| Total length of assembled sequence (Mb) | 406,103,086 |
| Number of scaffolds/contigs | 202,258 |
| Minimum scaffold length (bp) | 200 |
| Maximum scaffold length (bp) | 148,483 |
| GC % | 34.23 |
| N50 (bp) | 5,817 |
| L50 | 17,605 |
| Number of annotated genes | 35,093 |

The workflow of the genome assembly is presented in the supplementary material (Figure III.2). The final scaffold assembly of the genome is given in Table III.2 and the statistics of assembly contigs and scaffolds are shown in the Table III.3. The alignment of raw reads to the hybrid genome sequence was performed, which mapped 99.5% and 99.27% of Illumina and MGI reads respectively (Text data III.1A and III.1B).

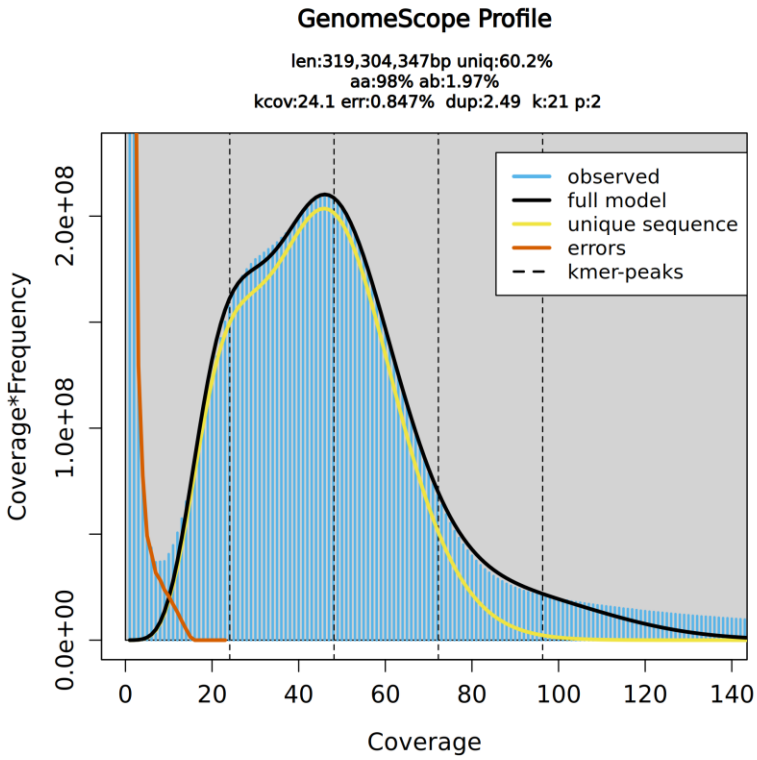



**Figure III.1A:** The kmer histogram distribution for the Illumina reads using the GenomeScope tool.

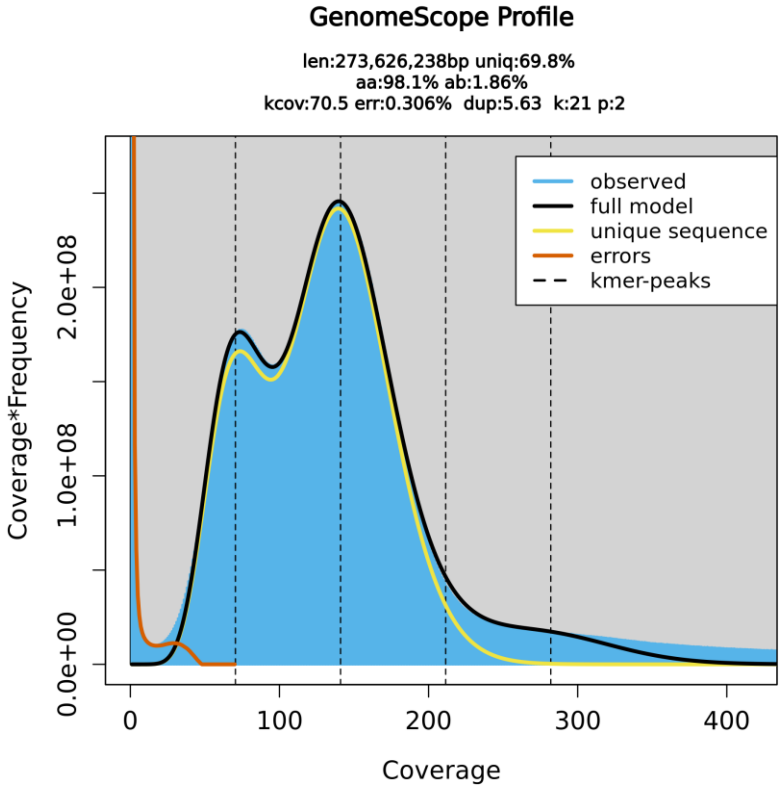



**Figure III.1B:** The k-mer histogram distribution for the MGI reads using the GenomeScope tool.

The completeness of the Peepal tree genome assembly was assessed with the BUSCO tool. The results showed that 76.5% (232 out of 303) and 84.1% (1,210 out of 1,440) of genes were conserved as single-copy orthologs in eukaryotic and plant universal data sets, respectively. Out of 232 complete genes in the Eukaryota database, 214 are single-copy orthologs, 18 are duplicates, 57 are fragmented and 14 are missing. Out of 1,210 complete genes in the Embryophyte database, 1,173 are single-copy orthologs, 37 are duplicates, 105 are fragmented and 125 are missing (Figures III.3A and III.3B). The transcriptome sequence reads aligned with the assembled genome showed that 99.46% of all reads were mapped and of these 88.25% of paired reads were mapped (Text data III.2).

**Figure III.2:** Flow chart of *De Novo* Whole Genome Analysis (WGA) of *Ficus religiosa*

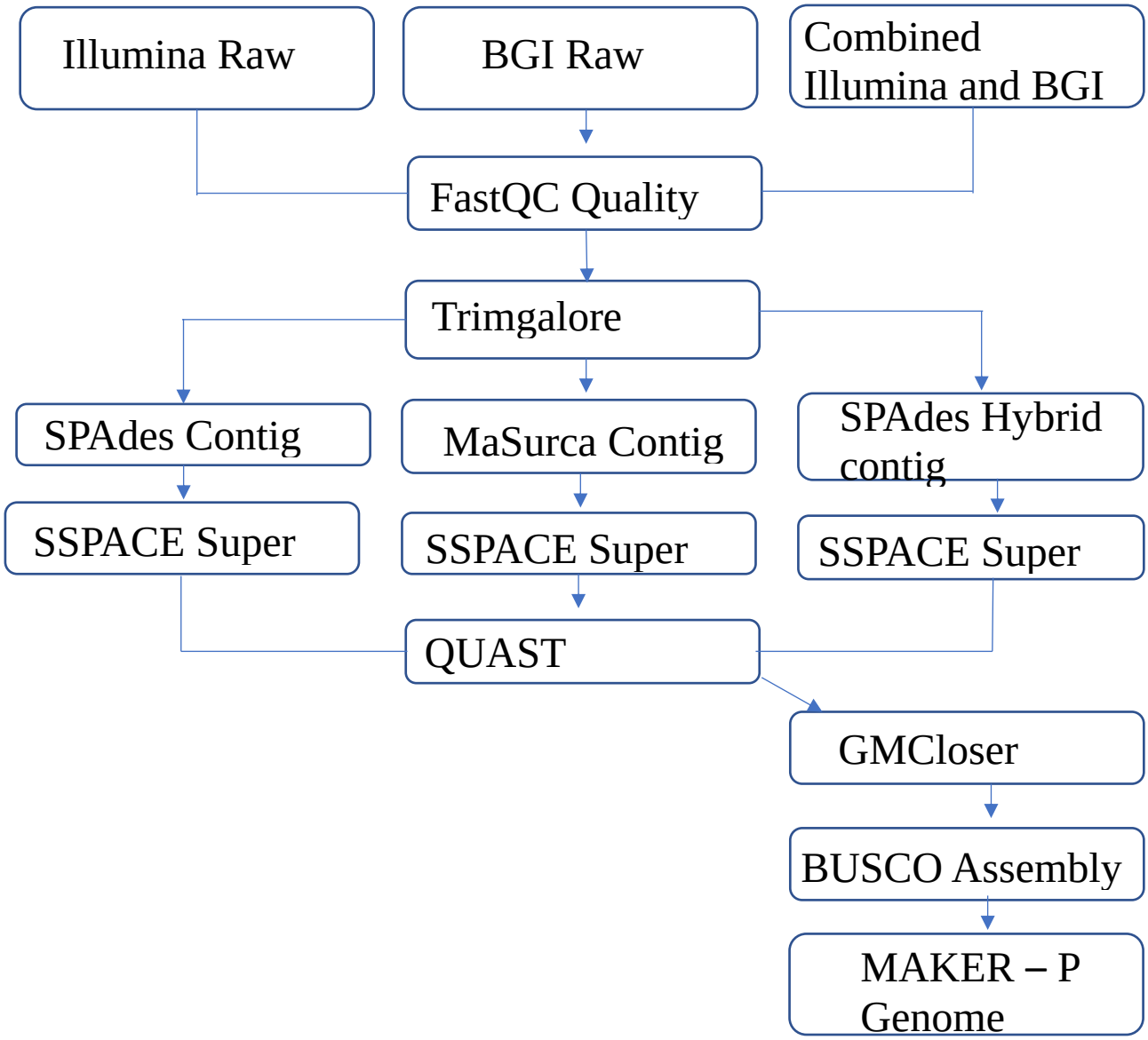


**Table III.3:** Contig and scaffold assembly statistics of *F. religiosa* genome

| **Contig assembly** | **Illumina assembly** | **MGI assembly** | **Hybrid assembly** | |
|---|---|---|---|---|
| No. of contigs | 92,794 | 234,282 | 136,974 | |
| Largest contig (bp) | 45,835 | 29,929 | 95,945 | |
| Total assembled bases | 248,292,704 | 300,267,781 | 307,589,416 | |
| Minimum scaffold length (bp) | 56 | 300 | 78 | |
| Maximum scaffold length (bp) | 39,390 | 29,929 | 95,945 | |
| GC % | 33.13 | 33.40 | 33.61 | |
| N50 (bp) | 4,862 | 1,481 | 3,916 | |
| L50 | 14,285 | 54,768 | 19,679 | |
| **Scaffold assembly** | **Illumina assembly** | **MGI assembly** | **Hybrid assembly** | **Gap closed assembly** |
| No. of scaffolds | 87,060 | 118,027 | 121,895 | 121,696 |
| Largest scaffold (bp) | 56,337 | 54,685 | 174,746 | 174,006 |
| Total assembled bases | 264,662,650 | 345,020,555 | 386,980,121 | 381,047,120 |
| Minimum scaffold length (bp) | 56 | 300 | 78 | 78 |
| Maximum scaffold length (bp) | 56,337 | 54,685 | 174,748 | 174,748 |
| GC % | 32.99 | 33.44 | 33.69 | 33.70 |
| N50 (bp) | 5,340 | 4,771 | 6,482 | 6,385 |
| L50 | 13,703 | 20,196 | 15,619 | 15,539 |

**Text data III.1A** The statistics of Illumina reads mapped to Peepal hybrid whole genome

281217169 + 0 in total (QC-passed reads + QC-failed reads)
266593648 + 0 primary
0 + 0 secondary
14623521 + 0 supplementary
0 + 0 duplicates
0 + 0 primary duplicates
279840572 + 0 mapped (99.51% : N/A)
265217051 + 0 primary mapped (99.48% : N/A)
266593648 + 0 paired in sequencing
133296824 + 0 read1
133296824 + 0 read2
146303940 + 0 properly paired (54.88% : N/A)
264362740 + 0 with itself and mate mapped
854311 + 0 singletons (0.32% : N/A)
105379514 + 0 with mate mapped to a different chr
80340131 + 0 with mate mapped to a different chr (mapQ>=5)

**Text data III.1B** The statistics of MGI reads mapped to Peepal hybrid whole genome

647258709 + 0 in total (QC-passed reads + QC-failed reads)
645008220 + 0 primary
0 + 0 secondary
2250489 + 0 supplementary
0 + 0 duplicates
0 + 0 primary duplicates
642546929 + 0 mapped (99.27% : N/A)
640296440 + 0 primary mapped (99.27% : N/A)
645008220 + 0 paired in sequencing
322504110 + 0 read1
322504110 + 0 read2
498335748 + 0 properly paired (77.26% : N/A)
639033956 + 0 with itself and mate mapped
1262484 + 0 singletons (0.20% : N/A)
135314868 + 0 with mate mapped to a different chr
100064168 + 0 with mate mapped to a different chr (mapQ>=5)

**Text data III.2** The statistics of transcriptome reads mapped to Peepal hybrid whole genome

127288304 + 0 in total (QC-passed reads + QC-failed reads)
0 + 0 secondary
15818156 + 0 supplementary
0 + 0 duplicates
126599752 + 0 mapped (99.46% : N/A)
111470148 + 0 paired in sequencing
55735074 + 0 read1
55735074 + 0 read2
98367496 + 0 properly paired (88.25% : N/A)
110544540 + 0 with itself and mate mapped
237056 + 0 singletons (0.21% : N/A)
5078802 + 0 with mate mapped to a different chr
4371548 + 0 with mate mapped to a different chr (mapQ>=5)

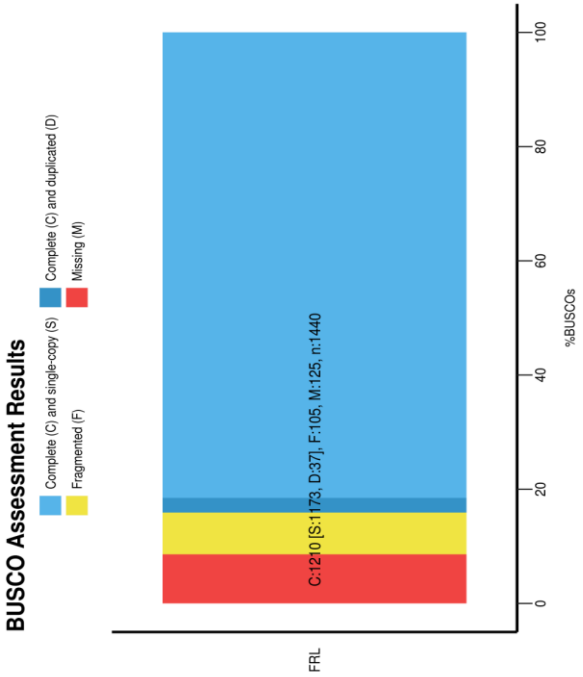



**Figure III.3A:** BUSCO Assessment results using the plant universal single-copy orthologs (embryophyta database)

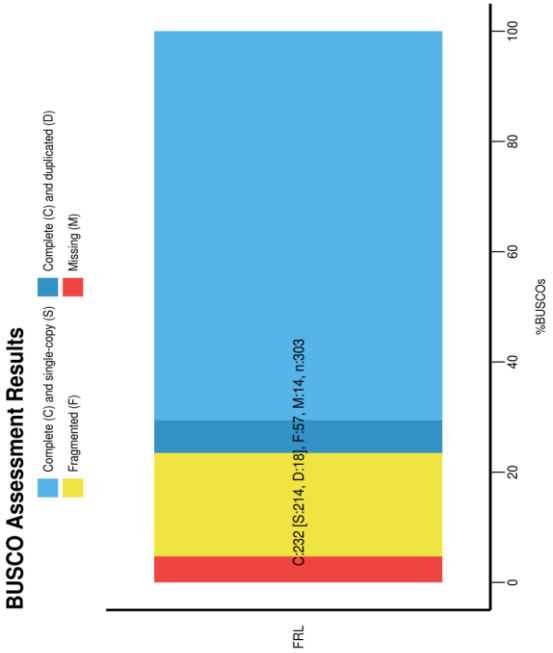



**Figure III.3B:** BUSCO Assessment results using the eukaryote universal single-copy orthologs (eukaryota database)

### III.2 Genome and pathways annotation

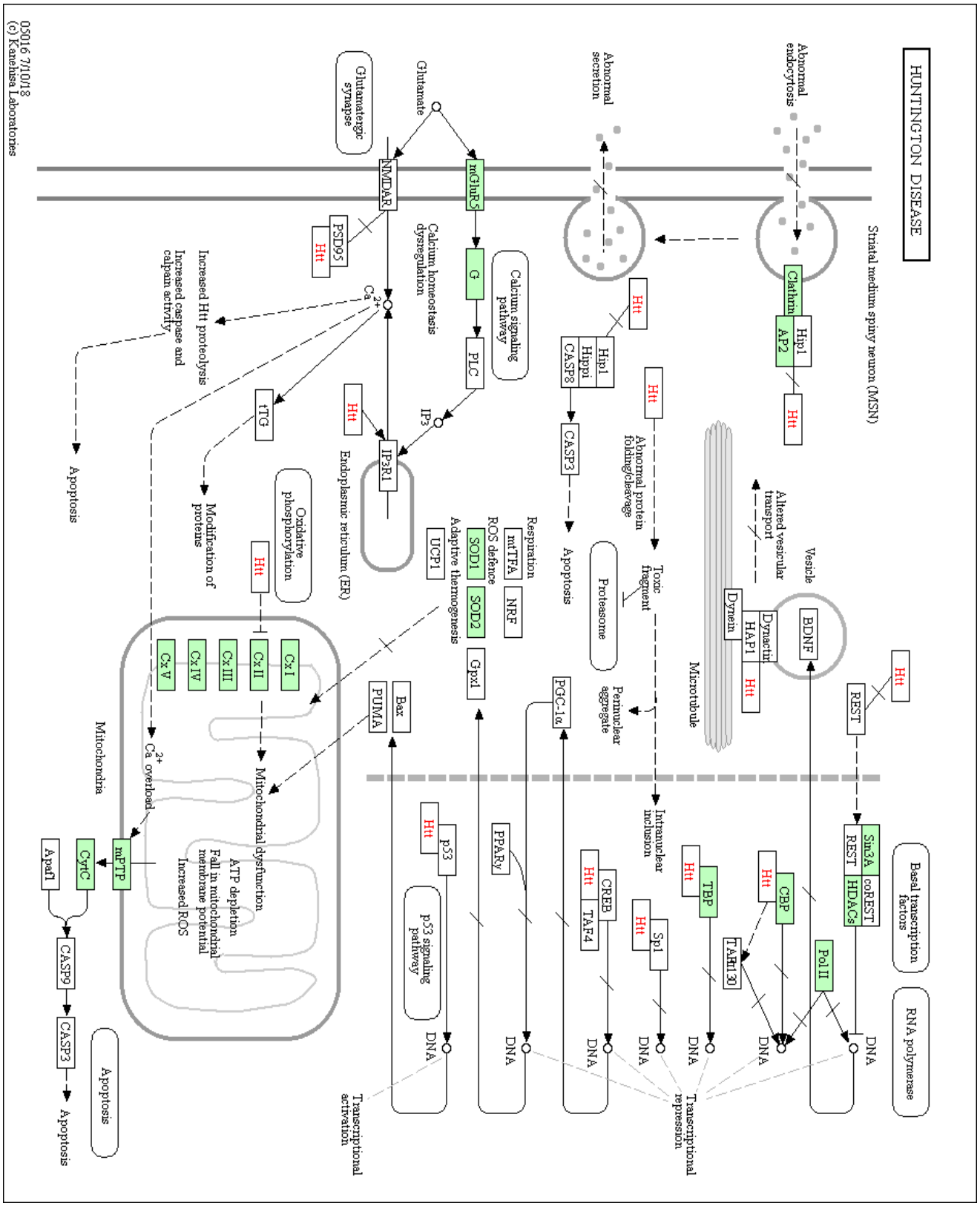
We identified 35,093 protein-coding genes with the complete structures in the Peepal tree genome (Additional file 1: Text file S1.1 and Additional file 2: Text file S1.2). RNA-seq data from two leaf tissue samples of the Peepal tree and alternative reference ESTs from *Morus notabilis* and *Arabidopsis thaliana* protein sequences were used as protein homology evidence during genome annotation. Based on a homology search using BLASTN, out of 35,093 genes predicted, 32,255 genes (91.9%) were having evidence from transcriptome assembly. About 76.3% of RNA-seq reads from the day and night leaf tissue samples were mapped to the annotated genes in the Peepal tree.

Based on the sequence similarities, the complete set of annotated genes and their amino acid sequences were used in the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis [47]. This result showed the pathways in metabolism, biosynthesis of secondary metabolites, genetic and environmental information processing, and signal transduction pathway were common as several others. The top 5 highest gene count for pathways like Ribosome (123 genes), Spliceosome (96 genes), Oxidative phosphorylation (86 genes), Thermogenesis (82 genes), and RNA transport (74 genes) was found. In addition, important candidate genes were also found for human disease pathways like Huntington's disease (68 genes), Parkinson’s disease (57 genes), Alzheimer's disease (55 genes), and others shown in Table III.4 and the neural pathways is shown in the Figure III.4 (A, B, C).

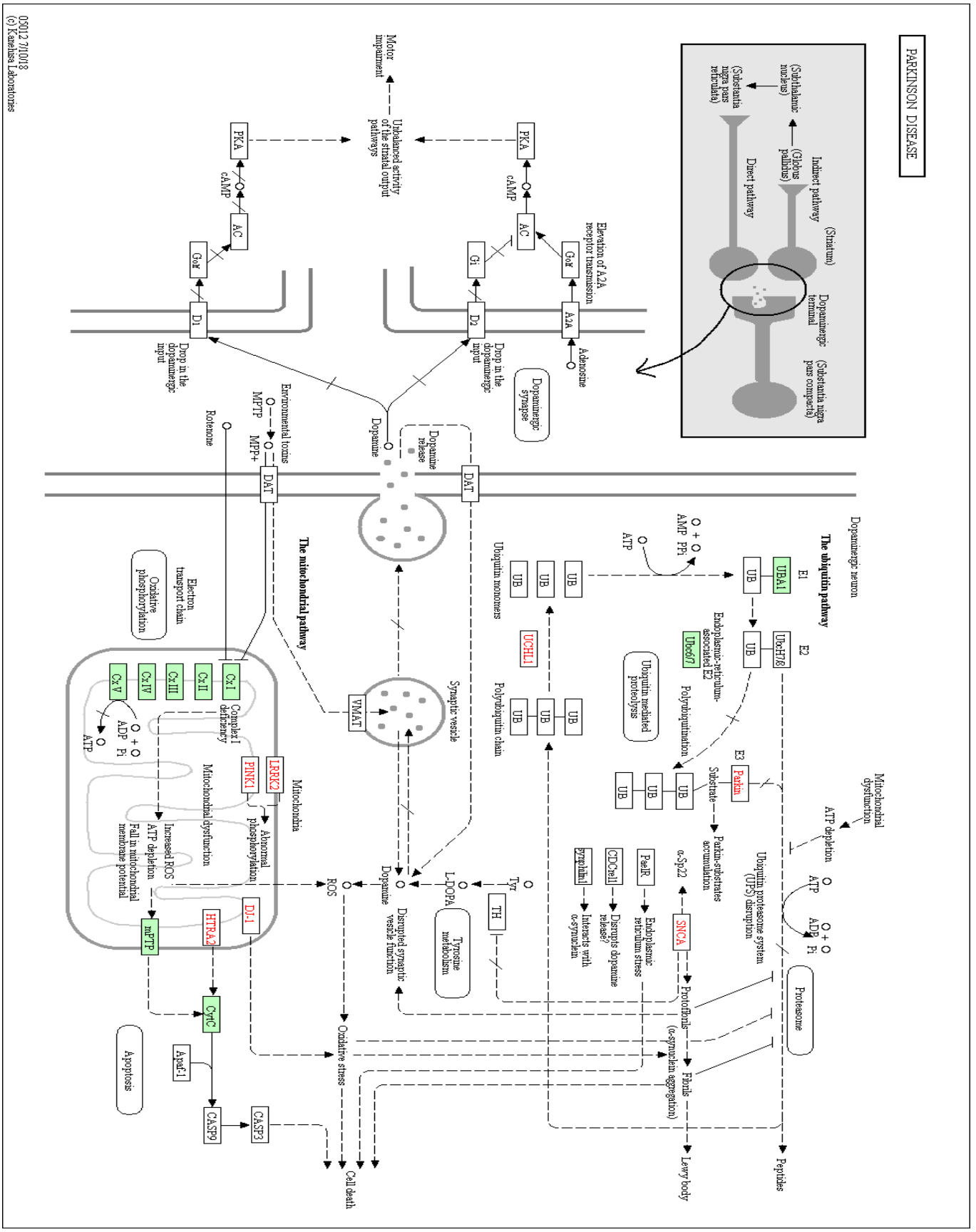
**Table III.4:** Top 10 pathways with highest gene counts in *F. religiosa* genome

| **Pathway names** | **Gene counts** |
|---|---|
| Ribosome | 123 |
| Spliceosome | 96 |
| Oxidative phosphorylation | 86 |
| Thermogenesis | 82 |
| RNA transport | 74 |
| Protein processing in endoplasmic reticulum | 73 |
| Huntington disease | 68 |
| Parkinson disease | 57 |
| Endocytosis | 55 |
| Alzheimer disease | 55 |

**A)**


HUNTINGTON DISEASE
05016 7/10/18
(c) Kanehisa Laboratories
Striatal medium spiny neuron (MSN)
Abnormal endocytosis
Clathrin
Hip1
AP2
Htt
Abnormal secretion
Htt
Hip1
Hippi
CASP8
CASP3
Apoptosis
Vesicle
BDNF
Altered vesicular transport
Dynactin
HAP1
Htt
Dynein
Microtubule
Htt
Abnormal protein folding/cleavage
Toxic fragment
Proteasome
Perinuclear aggregate
Intranuclear inclusion
Htt
REST
Sin3A
REST
coREST
HDACs
Pol II
DNA
Basal transcription factors
RNA polymerase
Htt
CBP
TAFII30
DNA
Htt
TBP
DNA
Htt
Sp1
DNA
Htt
CREB
TAF4
DNA
Transcriptional repression
Glutamate
Glutamatergic synapse
NMDAR
mGluR5
PSD95
Htt
Ca2+
Calcium homeostasis dysregulation
Increased caspase and calpain activity
Increased Htt proteolysis
Modification of proteins
G
PLC
IP3
IP3R1
Htt
Endoplasmic reticulum (ER)
Calcium signaling pathway
tTG
Apoptosis
Oxidative phosphorylation
Htt
Cx I
Cx II
Cx III
Cx IV
Cx V
Mitochondria
Ca2+ overload
mPTP
Mitochondrial dysfunction
ATP depletion
Fall in mitochondrial membrane potential
Increased ROS
Apaf1
CytC
CASP9
CASP3
Apoptosis
Respiration
mtTFA
NRF
ROS defence
SOD1
SOD2
Adaptive thermogenesis
UCP1
Gpx1
PGC-1α
PPARγ
DNA
Htt
p53
p53 signaling pathway
DNA
Bax
PUMA
DNA
Transcriptional activation

**B)**


PARKINSON DISEASE
(Striatum)
Indirect pathway
(Globus pallidus)
(Subthalamic nucleus)
Direct pathway
(Substantia nigra pars reticulata)
Dopaminergic terminal
(Substantia nigra pars compacta)
Motor impairment
Unbalanced activity of the striatal output pathways
PKA
cAMP
AC
Golf
D1
Drop in the dopaminergic input
Gi
D2
A2A
Elevation of A2A receptor transmission
Adenosine
Dopaminergic synapse
Dopamine
Dopamine release
DAT
Environmental toxins
MPTP
MPP+
Rotenone
Dopaminergic neuron
The ubiquitin pathway
E1
UBA1
E2
UbcH7/8
Endoplasmic-reticulum-associated E2
Ube2g2
Polyubiquitination
E3
Parkin
Substrate
Ubiquitin mediated proteolysis
Ubiquitin monomers
Polyubiquitin chain
UCHL1
UB
ATP
AMP PPi
Mitochondrial dysfunction
ATP depletion
ADP Pi
Ubiquitin proteasome system (UPS) disruption
Proteasome
Peptides
Parkin-substrates accumulation
α-Sp22
PaelR
CDCrel1
Pyrimidin1
SNCA
Endoplasmic reticulum stress
Disrupts dopamine release?
Interacts with α-synuclein
Protofibrils
(α-synuclein aggregation)
Fibrils
Lewy body
Cell death
Oxidative stress
Apoptosis
CASP3
CASP9
Apaf-1
CytC
mPTP
HTRA2
DJ-1
Abnormal phosphorylation
Increased ROS
ATP depletion
Fall in mitochondrial membrane potential
Mitochondrial dysfunction
Complex I deficiency
Cx I
Cx II
Cx III
Cx IV
Cx V
ADP Pi
ATP
PINK1
LRRK2
Mitochondria
The mitochondrial pathway
Oxidative phosphorylation
Electron transport chain
ROS
Dopamine
L-DOPA
Tyr
TH
Tyrosine metabolism
Disrupted synaptic vesicle function
Synaptic vesicle
VMAT
Dopamine release
DAT
06012 710118
(c) Kanehisa Laboratories

**C)**

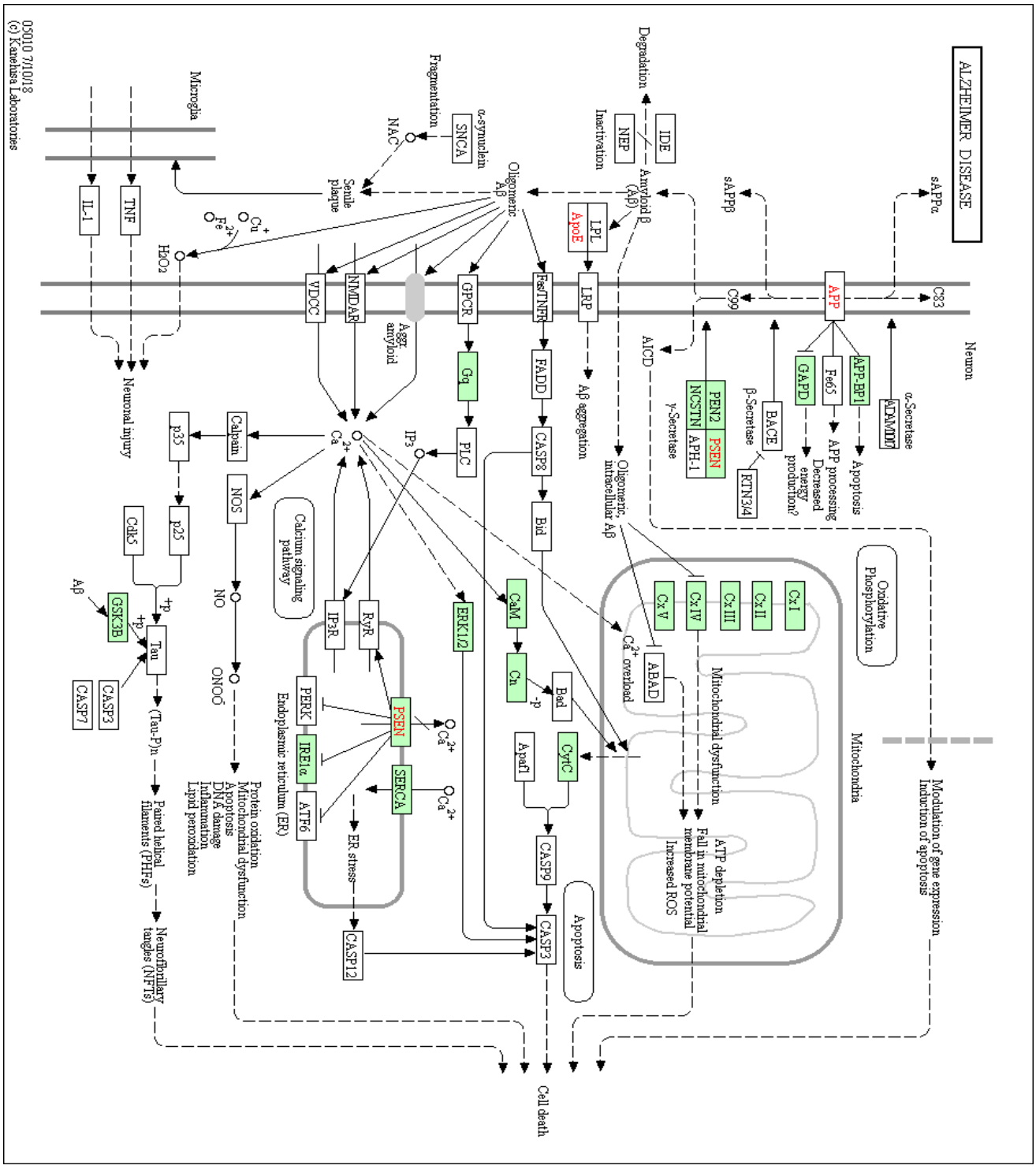



**Figure III.4:** Graphical representation of genome annotated genes of *Ficus religiosa* mapped to reference
A) Huntington disease pathway (map05016),
B) Parkinson disease pathway (map05012),
C) Alzheimer disease pathway (map05010) from KEGG.

The green color boxes indicate *F. religiosa* genes mapping on the reference and the white boxes indicate no mapping.

Source: The permission to use of KEGG pathway image is granted under the CC BY 4.0 open-access license from Kanehisa Laboratories and sources are listed in the references.

### III.3 Protein family and Gene Ontology analysis

The protein family (Pfam) ID and Gene Ontology (GO terms) were assigned to genes using an InterProScan module [32]. Out of 35,093 genes, 24,163 consisted of Pfam IDs that were distributed across 3,759 types of Pfam domains, and their gene ontology (GO) terms were also identified. The Pfam domain consisting of proteins that were large in the Peepal tree genome included 3-Deoxy-D-manno-octulosonic-acid transferase, Ring finger domain, PPR repeat family, Helix-loop-helix DNA-binding protein, DYW family of nucleic acid deaminases, Lysine methyltransferase, Putative GTPase activating protein for Arf, Ankyrin repeats and others (Additional file 3: Table S1).

Catalase is an antioxidant enzyme known to catalyze $H_2O_2$ into water and oxygen. We identified the gene sequences for the Catalase gene (FRLM_016351-RA) and its isozyme CAT1 Catalase isozyme 1 (FRLM_016350-RA), (FRLM_012250-RA) in the Peepal genome annotation. Two catalase genes were identified in the differential expression of transcriptome data: the KatE gene known as a monofunctional catalase, and the KatG gene known as a catalase-peroxidase [48]. KatE gene also known as CatB, is differentially expressed during the day and night period with the Fragments Per Kilobase of transcript per Million mapped reads (FPKM) values 937.49 and 1786.02 respectively. KatG gene is differentially expressed during the day and night, with the FPKM values 162.03 and 81.53 respectively. The KatE gene has been reported to be involved in physiological pathways such as glyoxylate and dicarboxylate metabolism, tryptophan metabolism, MAPK signaling pathway – plant, FoxO signaling pathway, and serine-pyruvate transaminase pathway. The KatG gene is involved in tryptophan metabolism, tyrosine metabolism, biosynthesis of secondary metabolism, and drug metabolism pathways.

### III.4 Identification of homologous, orthologous, and singleton genes

To understand the gene evolution and relationships among *F. religiosa* and other taxa, we performed homologous and orthologous gene detection analysis for the Peepal tree with an additional 5 species. Homologous gene identification and orthologous clustering of the proteomes of six species, including the model organism *A. thaliana*, *M. notabilis* (closest relative species of Ficus), and other closely related species of the Moraceae family were selected for the analysis. Based on proteome sequence homology analysis, 29,516 homologous genes were found in *Arabidopsis thaliana*, 29,924 in *Morus notabilis*, 29,750 in *Cannabis sativa*, 29,909 in *Prunus persica*, and 29,830 in *Ziziphus jujuba* with respect to *F. religiosa* proteome sequences (35,093). *F. religiosa, A. thaliana, M. notabilis, C. sativa, P. persica*, and *Z. jujuba* form a cluster of 24,310 orthologous genes and are conserved within the species. The number of specific orthologous gene clusters identified was 15,016 in *F. religiosa*, 16,170 in *A. thaliana*, 16,517 in *M. notabilis*, 15,921 in *C. sativa*, 16,655 in *P. persica*, and 17,235 in *Z. jujuba*. A total of 1,184 single-copy gene clusters were found across the six species and the number of specific singletons identified was 10,154 in *F. religiosa*, 4,469 in *A. thaliana*, 2,284 in *M. notabilis*, 1,912 in *C. sativa*, 1,802 in *P. persica*, and 4,209 in *Z. jujuba* (Figure III.5).

The identified single-copy clusters were used to illustrate the taxonomic and phylogenetic relationships among a group of species. Based on the similarity of proteomes and single-copy orthologous clustering, we deduced the phylogenetic tree for *F. religiosa* and the other five species. The multiple sequence alignment (MSA) and Neighbour-Joining (NJ) methods were used for constructing an evolutionary phylogenetic tree. It was found that *F. religiosa* is closely related to *M. notabilis* by having more similarities in their proteomes as they are evolving from the Moraceae family, followed by *C. sativa, Z. jujuba, P. persica,* and *A. thaliana* (Figure III.6).

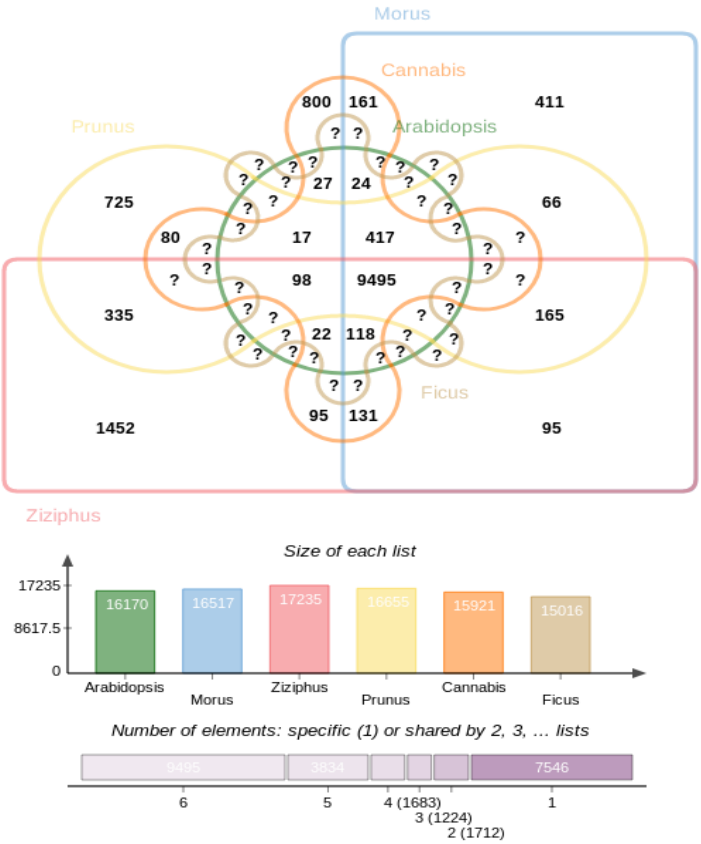



**Figure III.5:** Orthologous clustering of 6 species using proteome data deduced 24310 orthologous gene clusters and 1184 single-copy gene clusters across the above 6 species

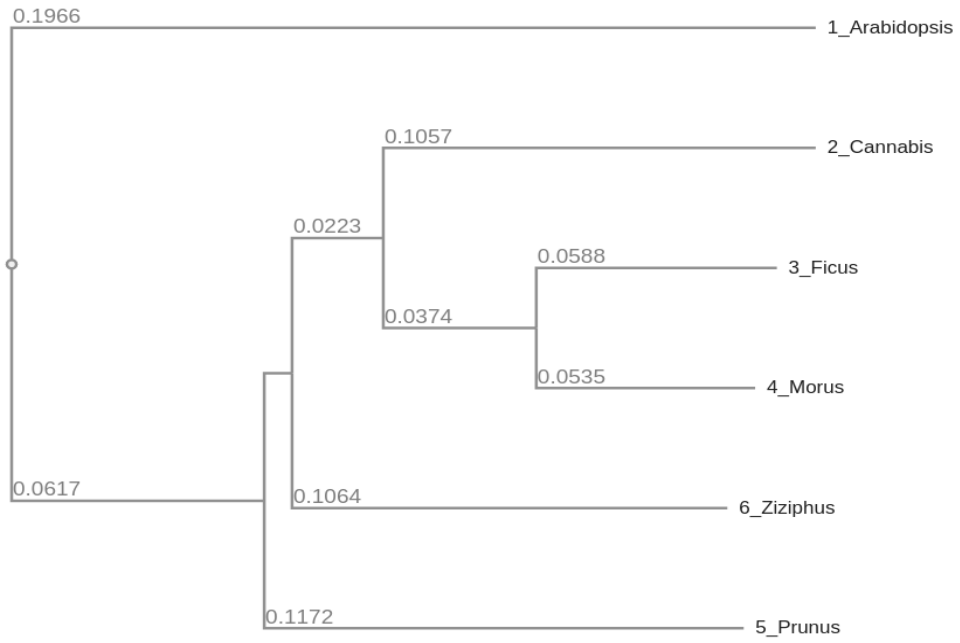



**Figure III.6:** Phylogenetic analysis of *Ficus religiosa* with other plant species like *A. thaliana, C. sativa*, *M. notabilis, Z. jujuba, P. persica*

### III.5 Comparative analysis of Peepal tree genome

We aligned the genome of the Peepal tree with those of the three Moraceae family members, *F. carica*, *F. microcarpa*, and *M. notabilis*. Comparison of our assembly against these genomes resulted in a mapping of 88.62% to *F. carica*, 89.6% to *F. microcarpa*, and 46.9% to *M. notabilis*. The results showed Peepal genome to be closer to the genus Ficus (*F. carica* and *F. microcarpa*) and also relatively closer to the genus Morus of the same family. The statistics of genome sequence alignment of *F. religiosa* against *F. carica*, *M. notabilis*, *F. microcarpa* genomes is presented in the supplementary material (Text data S3).

**Text data III.3:** The statistics of genome sequence alignment of *F. religiosa* against *F. carica*, *M. notabilis*, *F. microcarpa* genomes

**#*F. carica* assembly v/s *F. religiosa* assembly**
549342 + 0 in total (QC-passed reads + QC-failed reads)
0 + 0 secondary
347084 + 0 supplementary
0 + 0 duplicates
486801 + 0 mapped (88.62% : N/A)
0 + 0 paired in sequencing
0 + 0 read1
0 + 0 read2
0 + 0 properly paired (N/A : N/A)
0 + 0 with itself and mate mapped
0 + 0 singletons (N/A : N/A)
0 + 0 with mate mapped to a different chr
0 + 0 with mate mapped to a different chr (mapQ>=5)

***#M. notabilis assembly v/s F. religiosa assembly***
277314 + 0 in total (QC-passed reads + QC-failed reads)
0 + 0 secondary
75056 + 0 supplementary
0 + 0 duplicates
130283 + 0 mapped (46.98% : N/A)
0 + 0 paired in sequencing
0 + 0 read1
0 + 0 read2
0 + 0 properly paired (N/A : N/A)
0 + 0 with itself and mate mapped
0 + 0 singletons (N/A : N/A)
0 + 0 with mate mapped to a different chr
0 + 0 with mate mapped to a different chr (mapQ>=5)

**#*F. microcarpa* assembly v/s *F. religiosa* assembly**
536874 + 0 in total (QC-passed reads + QC-failed reads)
0 + 0 secondary
334616 + 0 supplementary
0 + 0 duplicates
481026 + 0 mapped (89.60% : N/A)
0 + 0 paired in sequencing
0 + 0 read1
0 + 0 read2
0 + 0 properly paired (N/A : N/A)
0 + 0 with itself and mate mapped
0 + 0 singletons (N/A : N/A)
0 + 0 with mate mapped to a different chr
0 + 0 with mate mapped to a different chr (mapQ>=5)

### III.6 Repeats in the genome of the Peepal tree

Repeat library building and repeat identification were performed using the ReapeatModeller and RepeatMasker tools (www.repeatmasker.org) respectively. *De novo* repeat identification resulted in 53.55% (269.62 Mb) repetitive sequences in the Peepal tree genome. The RNA elements, long terminal repeats (LTR) constitute about 5% of repeats and 43.71% of these repeats did not belong to any of the annotated repeats families. The 53.55% of repetitive sequences in the Peepal tree genome are closest to its Moraceae family species, 47% are found in the closest species mulberry (*M. notabilis*), 46.5% in *F. microcarpa*, and 48.9% in *F. hispida*. The repetitive sequences were classified into known categories, such as LINE1 (0.19%), long terminal repeat retrotransposon (5.09%), DNA transposons (1.09%), and simple repeats (3.25%) and unclassified (43.71%) shown in Table III.5.

**Table III.5:** Repeat content in the assembled *F. religiosa* genome

| **Classification** | **Copy number** | **DNA content (nt)** | **DNA content (%)** |
|---|---|---|---|
| **Non-LTR Retrotransposon** | | | |
| LINE1 | 2,723 | 939,296 | 0.19 |
| LTR-Retrotransposon | 134,518 | 25,612,270 | 5.09 |
| DNA transposons | 25,332 | 5,494,527 | 1.09 |
| Unclassified | 1,805,148 | 220,080,246 | 43.71 |
| Interspersed repeats | - | 252,126,339 | 50.07 |
| Simple Sequence repeats (SSR) | 368,277 | 16,386,007 | 3.25 |
| Low complexity | 45,966 | 2,238,964 | 0.44 |
| **Total repeat size in Mb (%)** | | **269.62 Mb** | **53.55 %** |

### III.7 Simple sequence repeats (SSRs)

We identified SSRs from the assembled Peepal tree genome. In total, 799,992 SSRs were identified on 267,593 sequences, which are composed of mono- (606,169), di- (143,113), tri- (34,327), tetra- (11,791), penta- (2,911), and hexa- (1,681) type repeats (Table S5.1). Among mono repeats, the 'A/T' (73.91%) type was the highest followed by 'C/G' (1.87%). Similarly, the 'AT /TA', 'AG/CT', 'AC/GT', and 'CG/CG' types of di repeats were in 9.8%, 2.76%, 1.41%, and 0.09% fractions, respectively. 'AAT/ATT', 'AAG/CTT' 'ATA/TAT', 'TTA/TTC', and 'GAA/TAA', were the most abundant tri repeats and 'AAAT' was predominant in tetra repeats. The detailed distribution of all types of repeats and statistics is shown in the Table III.6 and in the supplementary material (Additional file 4: Table S2).

**Table III.6:** Simple sequence repeats (SSR) prediction in the genome of *F. religiosa*

| **Features** | **Numbers** |
|---|---|
| No. of sequences | 1,189,520 |
| Total size of sequences (bp) | 503,523,744 |
| Total no. of identified SSRs | 799,992 |
| No. SSR containing sequences | 267,593 |
| Mono repeats | 606,169 |
| Di repeats | 143,113 |
| Tri repeats | 34,327 |
| Tetra repeats | 11,791 |
| Penta repeats | 2,911 |
| Hexa repeats | 1,681 |

### III.8 Transcription Factors (TFs)

Transcription factors act in regulating gene expression driven by several external and internal signals by activating or suppressing the downstream genes. The MAKER annotated protein sequences of Peepal tree genome assembly were used for BLAST analysis with the Plant

Transcription Factor Database v5.0 [38] using the *A. thaliana* protein sequence as a reference. A total of 1,264 protein sequences from 35,093 protein-coding genes with genome annotation shows evidence for 56 families of Transcription factors (Additional file 5: Table S3). The TFs families include the ERF, M-type MADS, ARF, DBB, MIKC MADS, WOX, C3H, G2-like, MYB, TALE, B3, HB-other, and MYB-related family proteins. The transcription factors play an important role in regulating growth, developmental processes, and environmental responses in the plant's [49].

### III.9 Transcriptome sequencing, assembly, and annotation

*De novo* transcriptome assembly was performed for the mature leaf samples of the Peepal tree collected during the day and night periods. The assembly was performed for each sample and also a combined assembly was performed for the reads of both samples.

The combined transcriptome assembly resulted in 152.8 Mb assembled bases with an N50 length of 2,076 bp and an average transcript length of 13.16 Kb and 42.17% GC content. The transcriptome assembly and annotation workflow are given in the supplementary material (Figure III.7). The statistics of assembly contigs and sequence assembled contigs are also provided in the Table III.7.

**Table III.7:** Assembly Statistics of *F. religiosa* Transcriptome

| **Counts of transcripts** | **Assembly (Day - 2PM)** | **Assembly (Night -2AM)** | **Combined Assembly** |
|---|---|---|---|
| Number of genes | 35,928 | 36,271 | 43,227 |
| Number of transcripts | 82,679 | 90,498 | 116,038 |
| Percent GC | 42.75 | 42.42 | 42.17 |
| Contig N50 (nts) | 1,914 | 2,053 | 2,076 |
| Median contig length | 929 | 1,011 | 969 |
| Average Contig (nts) | 1238.49 | 1332.73 | 1316.83 |
| Total assembled bases | 102,397,092 | 120,608,982 | 152,802,655 |

The *de novo* assembled transcript sequences (116,038) were processed for annotation. *De novo* assembled transcripts were clustered to exclude the redundant transcripts and identified 26,479 unique transcripts sequences. The statistics of unigenes are given in Table III.8. *De novo* assembled transcripts and unigenes were annotated to find the structural and functional genes.

**Table III.8:** Statistics of Uni-genes in Peepal Transcripts

| **Features** | **Sample (Day - 2 PM)** | **Sample (Night - 2 AM)** | **Combined Sample** |
|---|---|---|---|
| Number of genes | 22,597 | 23,360 | 26,479 |
| Number of transcripts | 16,912 | 16,780 | 18,173 |
| Percent GC | 46.37 | 46.24 | 46.25 |
| Contig N50 (nts) | 1,404 | 1,407 | 1,374 |
| Median contig length | 813 | 801 | 753 |
| Average Contig (nts) | 1060.06 | 1055.95 | 1017.44 |
| Total assembled bases | 23,954,145 | 24,667,005 | 27,055,237 |

The protein families were identified for the uniquely characterized transcripts of RNA data. Out of 26,479 transcripts, Pfam IDs for 19,175 were distributed across 3,977 types of protein family (Pfam) domains and their gene ontology (GO) terms were also identified (Additional file 6: Table S4). Pfam IDs and GO terms were assigned to predict the function of unique gene sequences and encoded translated proteins.

### III.10 Differential gene expression analysis of diurnal and nocturnal period transcriptome data

The differentially expressed genes (35,182) from the day and night periods of leaf tissue samples of the Peepal tree (Additional file 7: Table S5.1) were used for the pathway analysis. The top 272 highly up-regulated differentially expressed transcripts were identified for diurnal and nocturnal periods (Additional file 8: Table S5.2).

The TFs were identified from the differentially expressed transcripts. From the day sample, 2 transcripts coded for specific TFs like C3H family protein and nuclear factor Y, subunit A7 (NF-YA7), and in the night period sample, the 6 transcripts coded for specific TFs like ERF family protein, CONSTANS-like 2, MYB-related family protein shown in the Table III.9. In plants, the nuclear factor-YA has a role in drought stress responses. In rice, NF-YA7 is involved in the drought tolerance pathway which is independent of the Abscisic acid manner [50]. Expression of C3H and NF during the day could influence plant growth and development in the Peepal tree. *The Ethylene response factor* ERF105 showed as the cold-regulated transcription factor gene of Arabidopsis [51]. In the Peepal tree, ERF, MYB -related family proteins like REVEILLE 1 (RVE1) [52] and late elongated hypocotyl gene (LHY) are expressed during the night period. RVE1 functions in the circadian clock and auxin pathways and LHY maintains the circadian rhythm in Arabidopsis [53]. Both the RVE1 and LHY are found expressed in night-specific Peepal tree transcripts indicating the active circadian rhythms and pathways during the dark time.

**Table III.9** Families of Transcription factors (TFs) identified in transcripts from day and night leaf samples

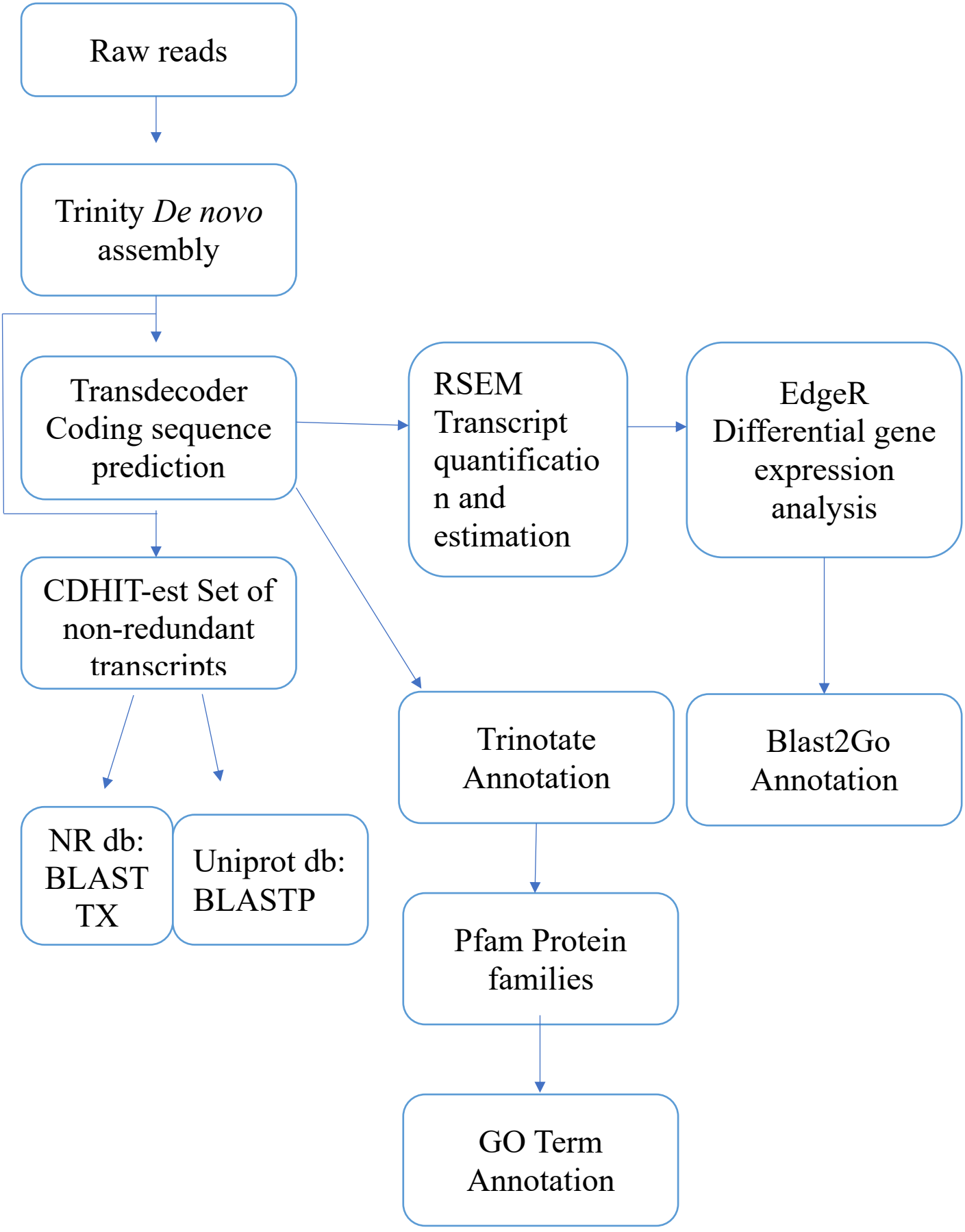
**Day leaf sample**

| **Transcript ID** | **Family** | **Best hit in A. thaliana** | **E-value** | **Description of Family** | **FPKM (day)** | **FPKM (night)** |
|---|---|---|---|---|---|---|
| TRINITY_DN8525_c0_g1_i14 | NF-YA | AT1G30500.2 | 6.00E-78 | nuclear factor Y, subunit A7 | 4.512 | 0.366 |
| TRINITY_DN7886_c1_g1_i1 | C3H | AT4G29190.1 | 8.00E-76 | C3H family protein | 54.377 | 4.366 |

**Night leaf sample**

| Transcript ID | Family | Best hit in A. thaliana | E-value | Description of Family | FPKM (day) | FPKM (night) |
|---|---|---|---|---|---|---|
| TRINITY_DN9099_c2_g8_i1 | ERF | AT5G51190.1 | 3.00E-37 | ERF family protein | 17.225 | 174.909 |
| TRINITY_DN9136_c3_g1_i3 | DBB | AT3G02380.1 | 2.00E-61 | CONSTANS-like 2 | 8.319 | 123.196 |
| TRINITY_DN10449_c1_g2_i6 | MYB_related | AT1G18330.1 | 2.00E-54 | MYB_related family protein | 4.3 | 47.347 |
| TRINITY_DN10449_c2_g1_i15 | MYB_related | AT5G17300.1 | 5.00E-67 | MYB_related family protein | 6.65 | 144.695 |
| TRINITY_DN9519_c3_g2_i11 | MYB_related | AT1G01060.3 | 1.00E-153 | MYB_related family protein | 0.329 | 53.994 |
| TRINITY_DN9519_c3_g2_i17 | MYB_related | AT1G01060.3 | 1.00E-147 | MYB_related family protein | 0.388 | 7.898 |

**Figure III.7:** Flow chart of *De novo* Transcriptome Analysis of *Ficus religiosa*



### III.11 Non-coding RNA genes in the Peepal tree genome

Based on a coding potential calculator (CPC), *de novo*-based assembled transcripts (26,691) were further categorized into protein coding (19,911) and non-coding (6,780). Based on BLASTN analysis, out of 6,780 non-coding transcripts, 4,219 transcripts targeted genome-annotated genes and 2,561 remained non-coding. A total of 30,973 Cufflinks assembled transcripts (reference-based alignment with genome assembly) were further categorized into protein-coding (7,163) and non-coding (23,810). Out of 23,810 non-coding transcripts, 14,605 were having alignment to genome-annotated genes using BLASTN. Further, categorization of specific day and night sample transcripts resulted in 6,628 (day) and 7,339 (night) protein-coding and 20,528 and 25,494 non-coding transcripts respectively. From these non-coding transcripts, 18,893 (day) and 22,232 (night) transcripts were aligned to MAKER-P predicted genes using BLASTN. The remaining transcripts were considered to be non-coding transcripts, as we did not find any match to predicted gene evidence to support them. Hence, the majority of RNA sequences are found to have protein-coding sequences, while the non-coding genes have been shown biologically relevant in recent years, and deepen studies are needed to understand their functions.

**miRNAs**: microRNAs are a major class of non-coding RNAs. Based on the homology search, we identified the microRNA precursors using the miRbase database (http://www.mirbase.org). These microRNAs belong to MIR396, MIR2916, MIR156, MIR164, MIR6236, MIR166, MIR168, and MIR395 families. Among the identified miRNAs, MIR408 was found to be specific to the night period transcripts of the Peepal tree. MIR 408 was identified on the genes like TPK5 Two-pore potassium channel 5, prfA peptide chain release factor 1, and also on proteins of unknown function in the Peepal genome. MIR408 is a highly conserved microRNA

in plants and is involved in enhancing photosynthesis by mitigating the efficiency of irradiation utilization and the capacity for carbon dioxide fixation [54].

The unigene transcripts were used to identify the microRNAs. MIR168 and MIR166 homologs were identified on the two transcripts. We identified the miRNAs on genomic scaffolds based on mapping the transcriptome data to the genome. This provides information on miRNAs specific to the day and night leaf tissue transcriptome. We identified 23 and 25 pre-miRNA expressions in the day and night period respectively shown in the Table III.10. The statistics of transfer RNAs (tRNA) were identified in the genome and their details are given in the Table III.11.

**Table III.10:** Details on long non-coding RNAs, miRNAs in *De novo* transcripts, reference-based transcripts and genome

| **Genes** | **Total transcript / contig** | **No. of coding transcripts** | **No. of non-coding transcripts** | **Non-coding transcripts v/s genome annotated genes (BLASTN)** | **miRNAs families (micro RNAs) predicted** | **Homologous miRNA in other organisms** | **No. of transcripts consists miRNAs** |
|---|---|---|---|---|---|---|---|
| *De novo* transcripts | 26,691 | 19,911 | 6,780 | 4,219 | MIR 168 and its subfamilies | mdm, ppe, tcc, mes, crt, csi, ccl, gma, mtr, fve, vun, aau, lja, cme, vvi, ptc, nta, lus, bra, aof, nta, rco, cln, pta, far | 2 |
| Reference based transcript | 30,973 | 7,163 | 23,810 | 14,605 | MIR396, MIR7534, MIR166, MIR6236, MIR399, MIR168, MIR4995, MIR2916, MIR171, | csi, fve, gma, ccl, ptc, sly, atr, tcc, mes, nta, pab, ppe, vvi, cme, mdm, lja, cln, pta, mmu, crt, | 30 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | MIR3629, MIR5523, MIR8576, MIR6173, MIR156, MIR393, MIR827, MIR172, MIR167, MIR529, MIR394, MIR398, MIR408, MIR5538, MIR162, | mtr, vun, aau, lus, bra, aof, aly, ath, rco, peu, bdi, bcy, bgy, hbr, amg, zma, sof, sbi, ghr, cas, osa, eun, ata, vca, ctr, bna, pab, egu, dpr, cca, far, cpa, hpe, htu | |
| Day sample transcript | 27,156 | 6,628 | 20,528 | 18,893 | MIR7534, MIR166, MIR6236, MIR168, MIR4995, MIR2916, MIR11602, MIR171, MIR396, MIR3629, MIR5523, MIR8576, MIR6173, MIR156, MIR393, MIR827, MIR399, MIR167, MIR529, MIR398, MIR5141, MIR5538, MIR162 | lja, cln, pta, mmu, mdm, ppe, tcc, mes, crt, csi, ccl, gma, mtr, fve, vun, aau, cme, vvi, ptc, nta, lus, bra, aof, aly, ath, rco, peu, pla, atr, bdi, bcy, bgy, hbr, amg, aau, lus, zma, sof, sbi, sly, ghr, cas, osa, eun, ata, vca, ppe, aof, dpr, cca, far, cas, ahy, gra, rgl, cpa, vun, hpe, htu | 26 |
| Night sample transcript | 32,833 | 7,339 | 25,494 | 22,232 | MIR396, MIR166, MIR6236, MIR399, MIR168, MIR4995, MIR11602, MIR5523, | csi, fve, gma, ccl, ptc, sly, atr, tcc, mes, nta, pab, ppe, vvi, cme, mdm, cln, pta, mmu, crt, | 31 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | MIR6173, MIR827, MIR167, MIR5538, MIR5141, MIR408, MIR7534, MIR2916, MIR171, MIR3629, MIR8576, MIR156, MIR393, MIR398, MIR162 and their subfamilies | mtr, vun, aau, lja, lus, bra, aof, aly, ath, rco, pla, bdi, bcy, bgy, hbr, amg, zma, sof, sbi, ghr, cas, osa, eun, ata, vca, rgl, peu, hbr, dpr, cca, far, cas, ahy, gra, cpa, hpe, htu | |
| Genome | 121,696 | - | - | - | MIR2916, MIR8175, MIR6236, MIR396, MIR7534, MIR166, MIR156, MIR157, MIR164, MIR395, MIR399, MIR319, MIR297, MIR529, MIR168, MIR4995, MIR5538, MIR11602, MIR171, MIR172, MIR160, MIR169, MIR396, MIR8576, MIR3629, MIR2111, MIR390, MIR6300, MIR6236, MIR5523, MIR167, MIR170, | peu, ath, mmu, csi, fve, gma, ccl, ptc, sly, atr, tcc, mes, nta, pab, ppe, vvi, cme, mdm, lja, cln, pta, hbr, aof, aqc, rco, vun, bcy, bgy, lus, vca, gra, mtr, stu, cpa, cas, bna, aly, ctr, rgl, ssl, eun, nta, dpr, bdi, pta, osa, mml, mdm, ahy, crt, aau, bra, hbr, atr, stu, pla, cca, pvu, ghb, amg, zma, sof, sbi, ghr, htu, hpa, ssl, egu, far, | 429 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | MIR6173, MIR5141, MIR5368, MIR159, MIR393, MIR6478, MIR8005, MIR827, MIR394, MIR398, MIR397, MIR408, MIR828, MIR477 and their subfamilies | | |

**Table III.11:** Statistics of Transfer RNAs predicted in the genome

| **Features** | **Numbers** |
|---|---|
| No. bases in tRNAs | 83,569 |
| Total tRNAs predicted | 1,104 |
| Average tRNA length | 75 |
| Infernal-confirmed tRNAs | 1,046 |
| tRNAs decoding standard 20 amino acids | 890 |
| Suppressor tRNAs | 2 |
| tRNAs with unknown isotypes | 4 |
| Predicted pseudogenes | 150 |
| tRNAs with introns | 59 |

### III.12 GO enrichment Analysis

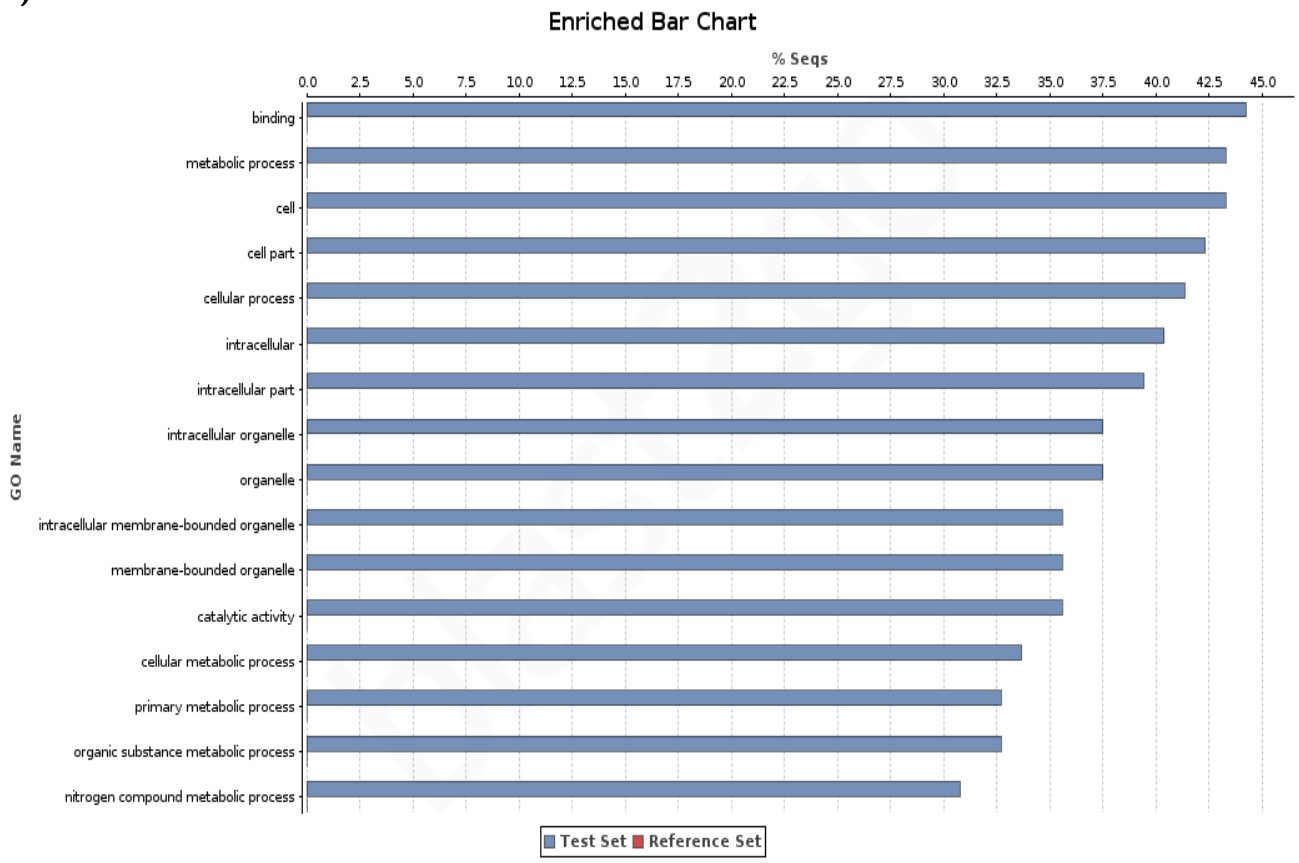
The Gene Ontology (GO) analysis was performed with the Blast2Go tool for the up-regulated genes obtained from the Differential gene expression analysis. GO enrichment is classified under three categories, genes which are having gene ontology terms and function for biological process, cellular process and molecular process. The GO terms for enriched genes of highly expressed genes from night sample on the basis of sequence percentage is given in the Table III.12 and for day sample is given in the Table III.13. The bar chart representing the GO terms for enriched genes of up regulated genes in night sample and day sample is given in the Figure

III.8 (A and B) respectively. and for the night sample in the Figure. The GO enrichment analysis of night sample resulted the GO terms for the genes involved in biological process are cellular component organization or biogenesis, signalling, localization, response to stimulus, regulation of biological process, biological regulation, cellular process and metabolic process; GO terms of genes involved in cellular process are protein containing complex, organelle part, membrane part, membrane, organelle, cell part and cell; GO terms of genes involved in molecular process are transporter activity, catalytic activity and binding. The GO enrichment analysis of day sample resulted the GO terms for the genes involved in biological process are response to stimulus, cellular process and metabolic process; GO terms of genes involved in cellular process are protein-containing complex, organelle part, organelle, cell, cell part, membrane and membrane part; GO terms of genes involved in molecular process are catalytic activity and binding. The pie chart for all the three GO term categories for night sample are shown below in the Figure III.9A, III.9B, III.9C and for the day sample are given in the Figure III.10A, III.10B, III.10C.

**Table III.12:** GO terms for enriched genes of highly expressed genes from night sample

| GO Name | Percent of Seqs (%) |
|---|---|
| binding | 44.23076923 |
| metabolic process | 43.26923077 |
| cell | 43.26923077 |
| cell part | 42.30769231 |
| cellular process | 41.34615385 |
| intracellular | 40.38461538 |
| intracellular part | 39.42307692 |
| intracellular organelle | 37.5 |
| organelle | 37.5 |
| intracellular membrane-bounded organelle | 35.57692308 |
| membrane-bounded organelle | 35.57692308 |
| catalytic activity | 35.57692308 |
| cellular metabolic process | 33.65384615 |
| primary metabolic process | 32.69230769 |
| organic substance metabolic process | 32.69230769 |
| nitrogen compound metabolic process | 30.76923077 |

**A)**




**B)**

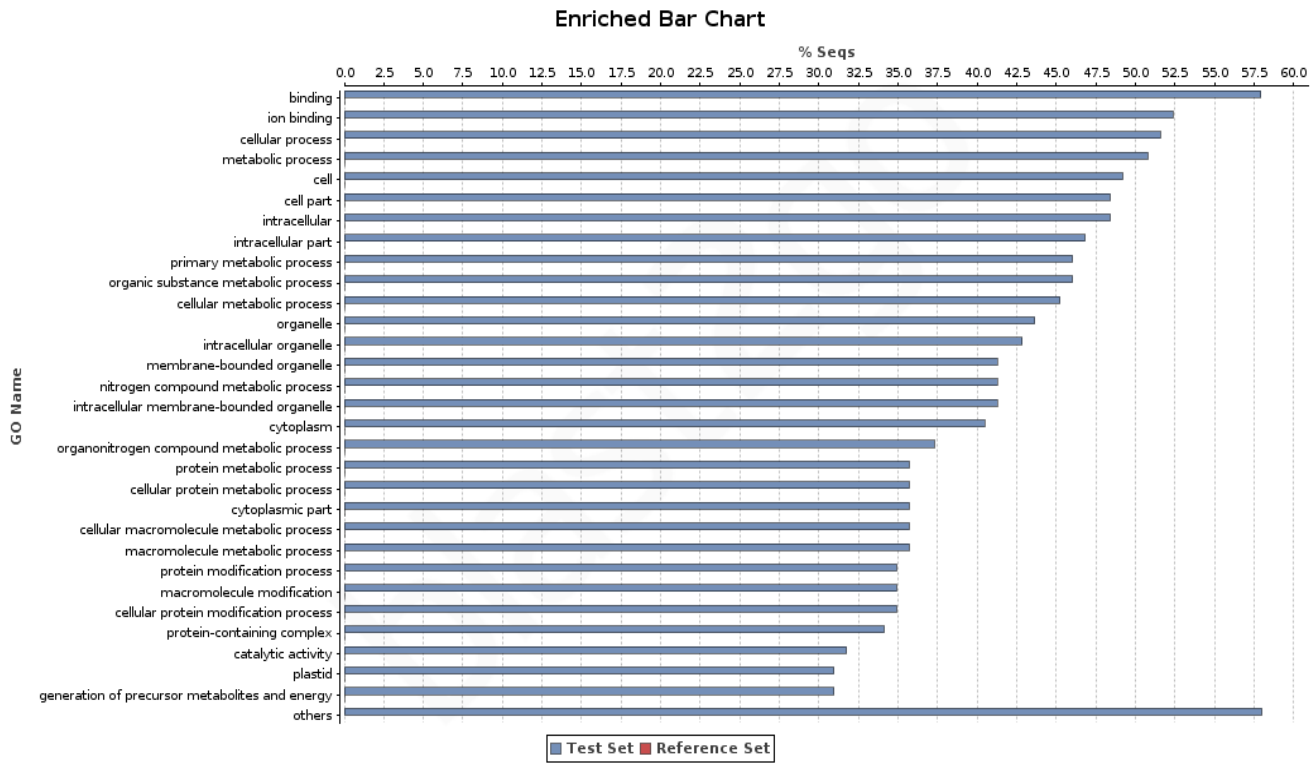



**Figure III.8:** Bar graph representing the GO terms for enriched genes of up regulated genes in A) Night sample B) Day sample

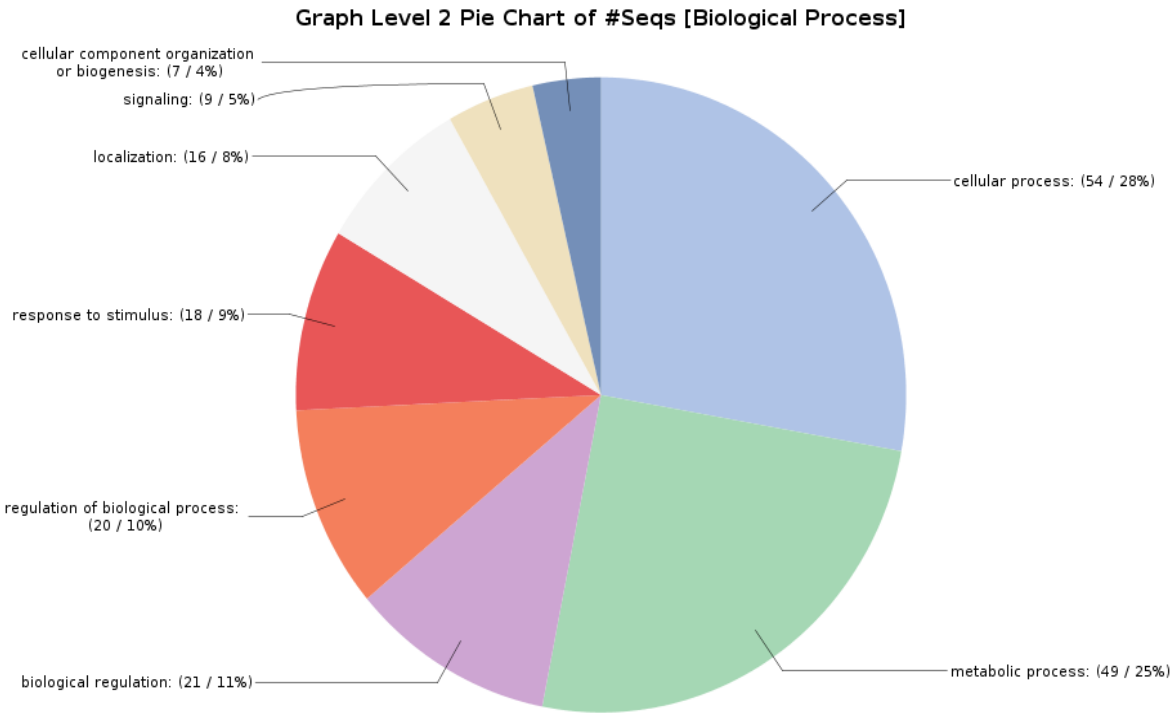



**Figure III.9A:** Genes involved in the biological process and their GO terms. in night sample

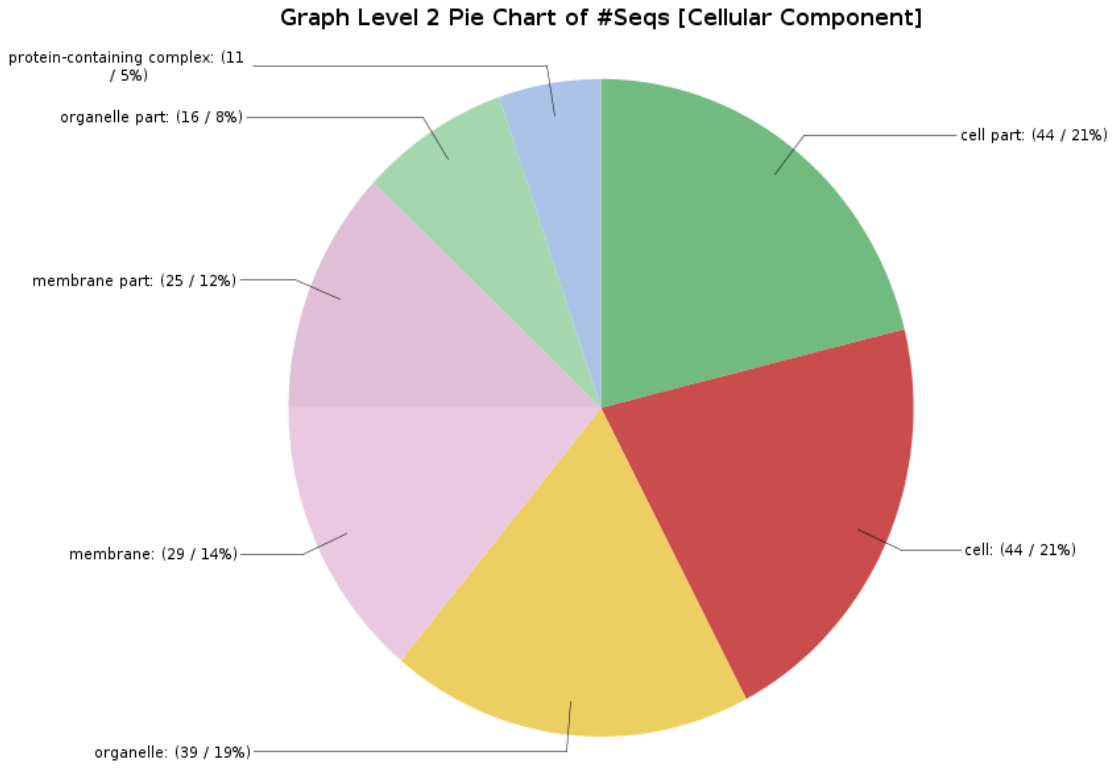



**Figure III.9B:** Genes involved in the cellular process and their GO terms in night sample

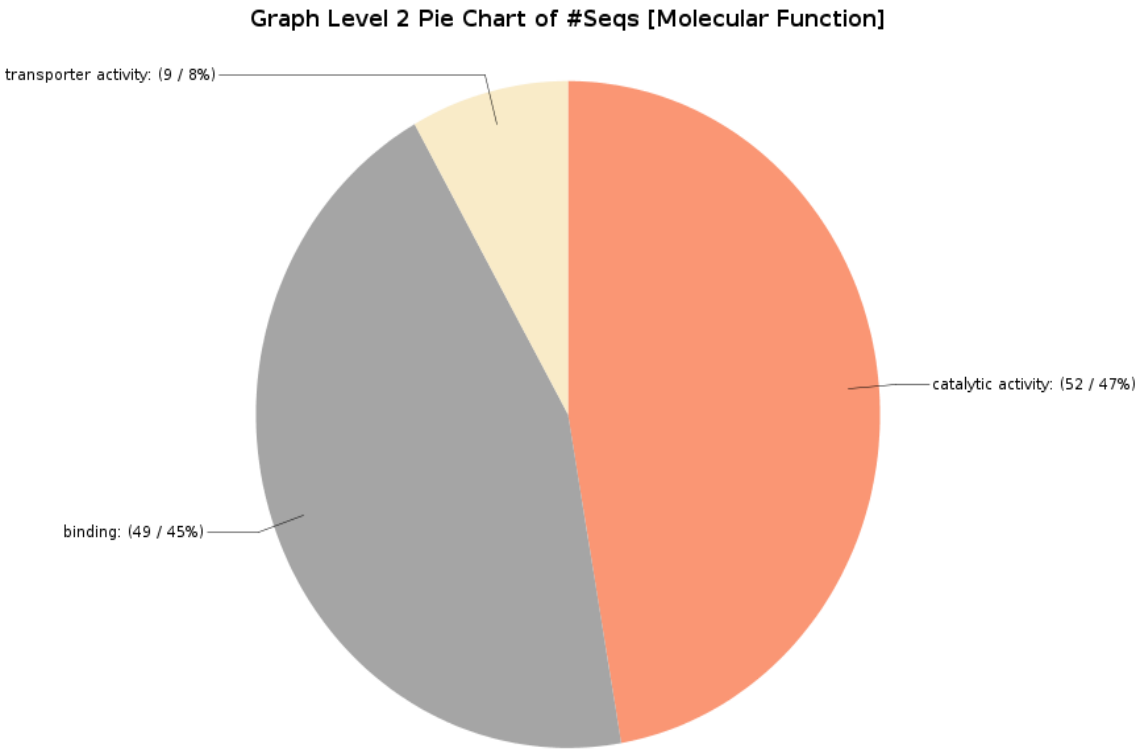



**Figure III.9C:** Genes involved in the molecular process and their GO terms in night sample

**Table III.13:** GO terms for enriched genes of highly expressed genes from day sample

| GO Name | % Seqs |
|---|---|
| binding | 58.73015873 |
| ion binding | 53.17460317 |
| cellular process | 51.58730159 |
| metabolic process | 50.79365079 |
| cell | 50 |
| cell part | 49.20634921 |
| intracellular | 49.20634921 |
| intracellular part | 47.61904762 |
| primary metabolic process | 46.03174603 |
| organic substance metabolic process | 46.03174603 |
| cellular metabolic process | 45.23809524 |
| organelle | 44.44444444 |
| intracellular organelle | 43.65079365 |
| membrane-bounded organelle | 42.06349206 |
| intracellular membrane-bounded organelle | 42.06349206 |
| cytoplasm | 41.26984127 |
| nitrogen compound metabolic process | 41.26984127 |
| organonitrogen compound metabolic process | 37.3015873 |
| cytoplasmic part | 36.50793651 |
| protein metabolic process | 35.71428571 |
| cellular protein metabolic process | 35.71428571 |

| | |
|---|---|
| cellular macromolecule metabolic process | 35.71428571 |
| macromolecule metabolic process | 35.71428571 |
| protein modification process | 34.92063492 |
| macromolecule modification | 34.92063492 |
| protein-containing complex | 34.92063492 |
| cellular protein modification process | 34.92063492 |
| plastid | 31.74603175 |
| catalytic activity | 31.74603175 |
| thylakoid | 30.95238095 |
| generation of precursor metabolites and energy | 30.95238095 |
| photosynthesis | 28.57142857 |

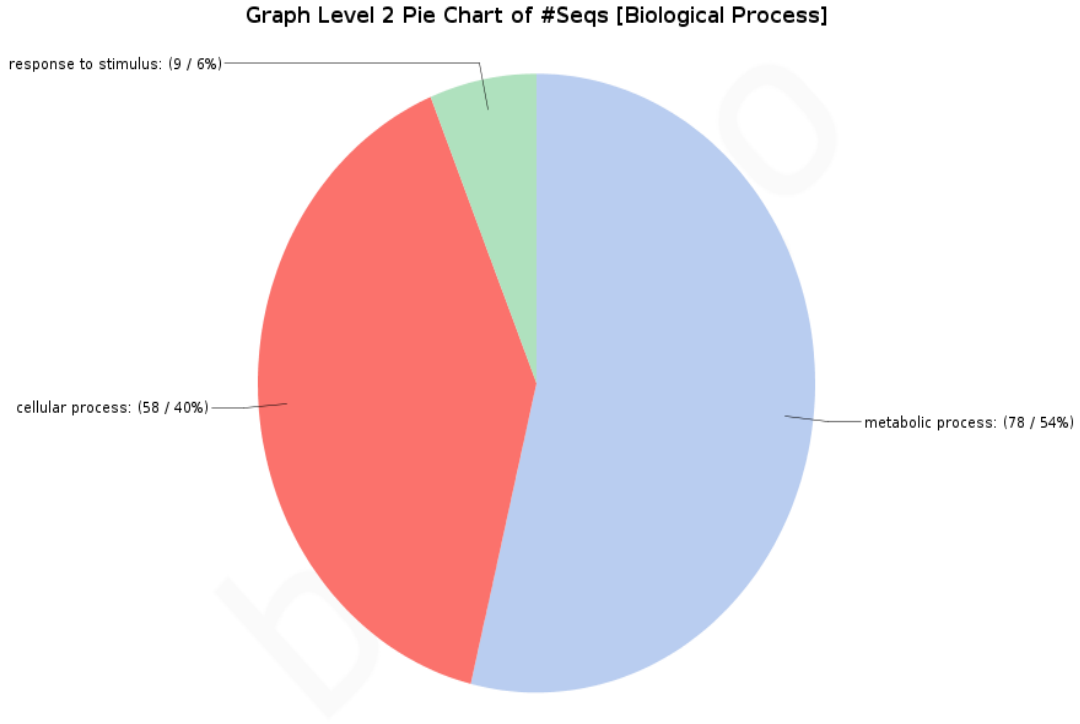



**Figure III.10A:** Genes involved in the biological process and their GO terms in day sample

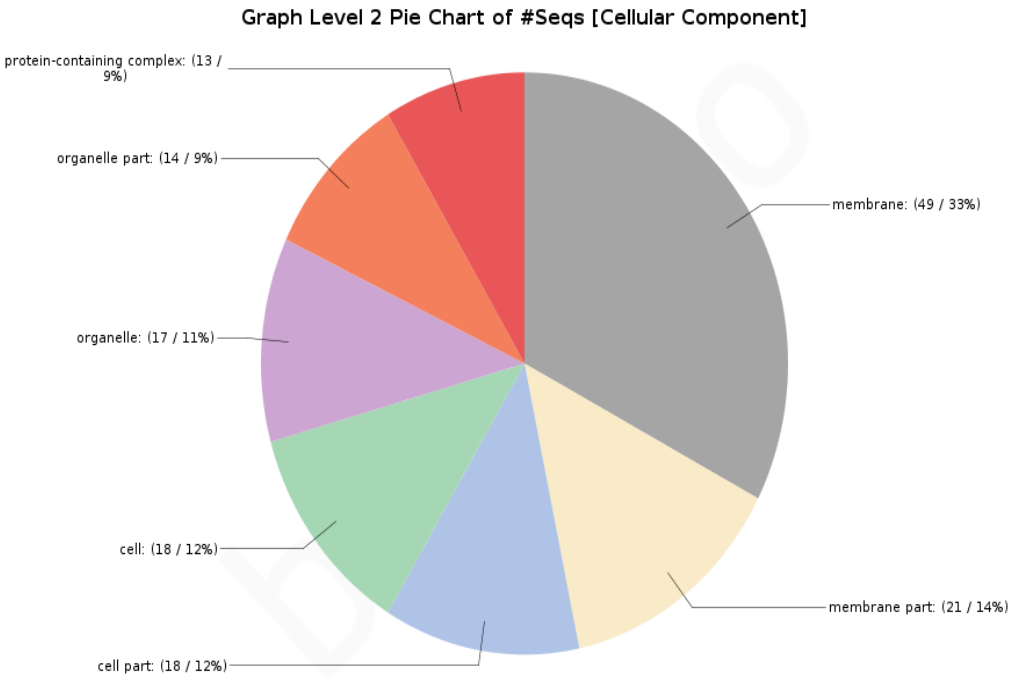



**Figure III.10B:** Genes involved in the cellular process and their GO terms in day sample

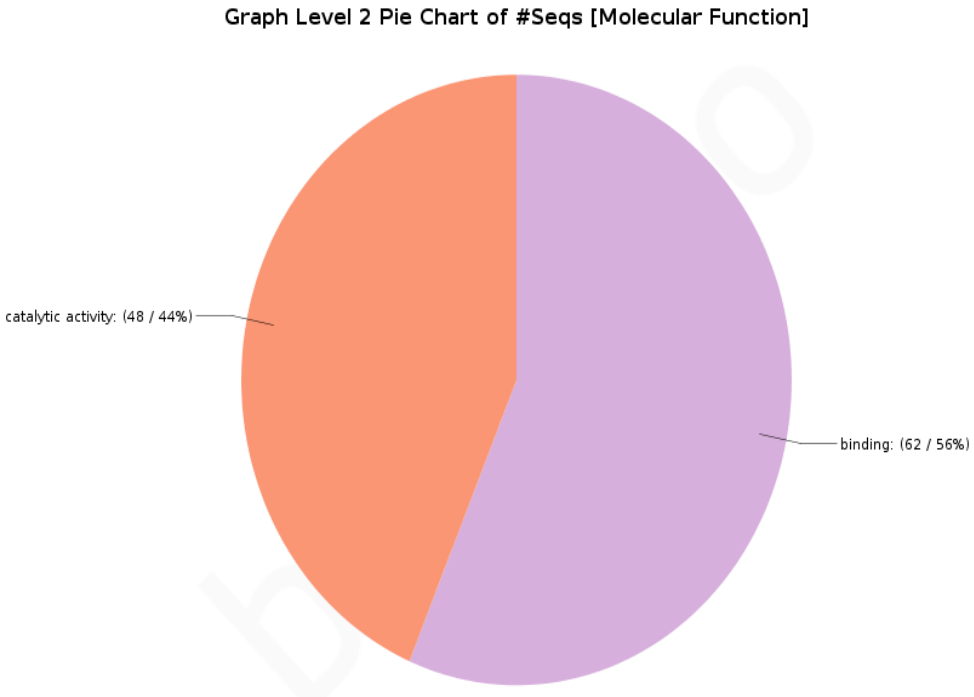



**Figure III.10C:** Genes involved in the molecular process and their GO terms in day sample

### III.13 Identification of carbon fixation pathway in Peepal tree

The study was conducted to analyse the gene expression patterns in the leaf tissues of the Peepal tree under the diurnal (2:00 PM) and the nocturnal period (2:00 AM). Through the pathway analysis, the candidate genes for carbon fixation pathways like the CAM pathway, Calvin-Benson cycle ($C_3$) pathway, and $C_4$ - Dicarboxylic pathway were identified and estimated based on their transcript abundance.

The transcriptome data contained 20 putative genes involved in the carbon fixation module of CAM, $C_3$, and $C_4$ including the key genes fructose-bisphosphate aldolase class I, fructose-1,6-bisphosphate, phosphoenolpyruvate carboxylase (PEPC/PPC), phosphoenolpyruvate carboxylase kinase (PPCK), NAD+ and NADP+, malate dehydrogenase (MDH) and pyruvate orthophosphate dikinase (PPDK) genes (Additional file 9: Table S6). Gene mapping was completed for CAM and $C_4$ cycle pathways and could not find a mapping for three genes in the $C_3$ cycle pathway. Those three genes, fructose-6-phosphate phosphoketolase (EC 4.1.2.22) and phosphoketolase (EC 4.1.2.9) were purified in *Acetobacter xylinum* [55] and sedoheptulokinase (EC 2.7.1.14) was shown in *Bacillus* species [56]. These three genes were not found in the Peepal tree for the $C_3$ cycle.

The differentially expressed genes from transcriptomic data were mapped to the reference carbon fixation in photosynthetic organisms pathway on the KEGG database (Figure III.11) [57], [58], and [59]. The diagrammatic representation of the genes involved in the carbon fixation pathway is shown in the Figure III.12.

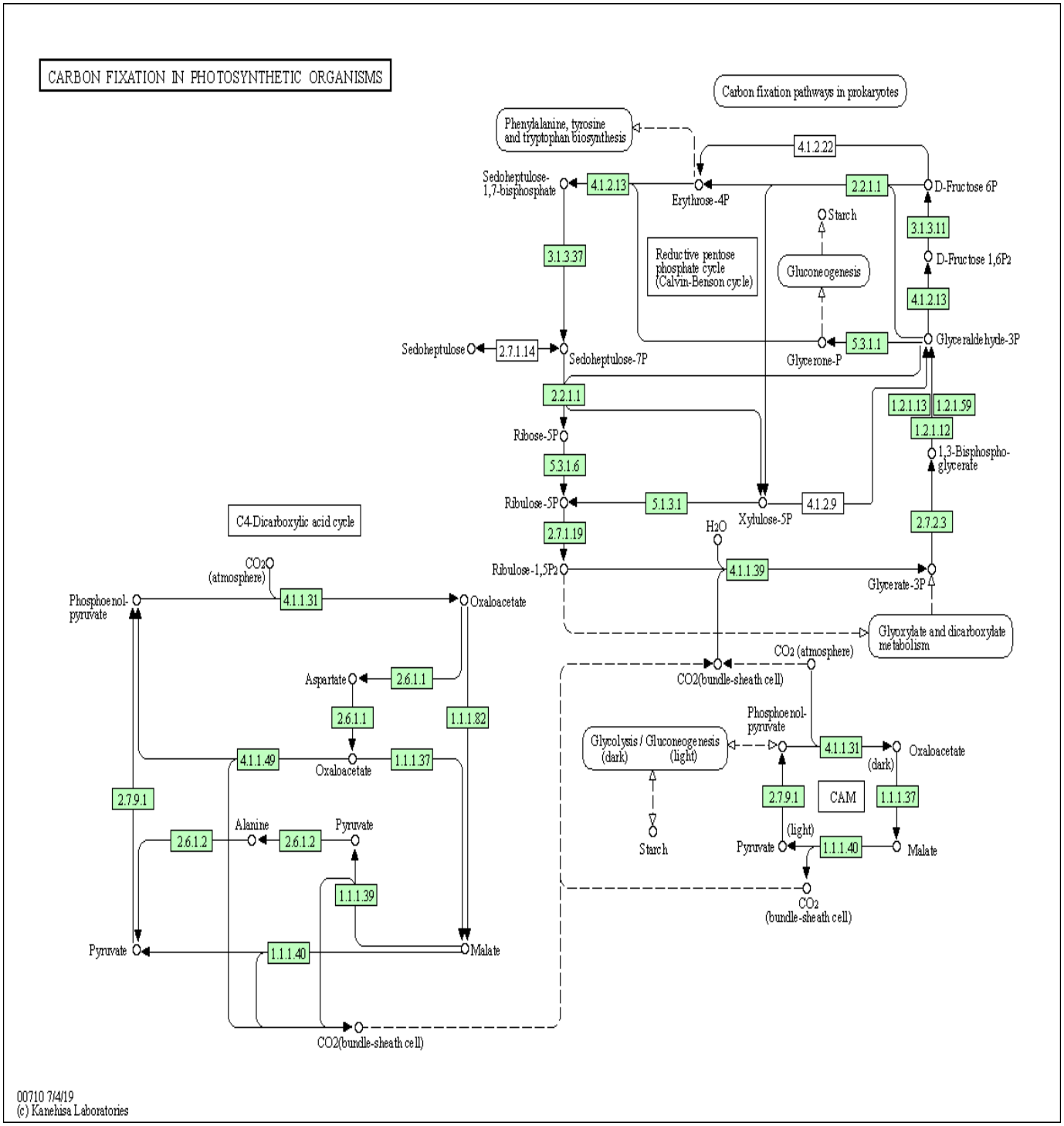



**Figure III.11:** Graphical representation of differentially expressed genes of *Ficus religiosa* mapped to reference Carbon fixation in photosynthetic organisms pathway (map00710) from KEGG. The green color boxes indicate *F. religiosa* genes mapping on the reference and the white boxes indicate no mapping.

Source: The permission to use of KEGG pathway image is granted under the CC BY 4.0 open-access license from Kanehisa Laboratories and sources are listed in the references.

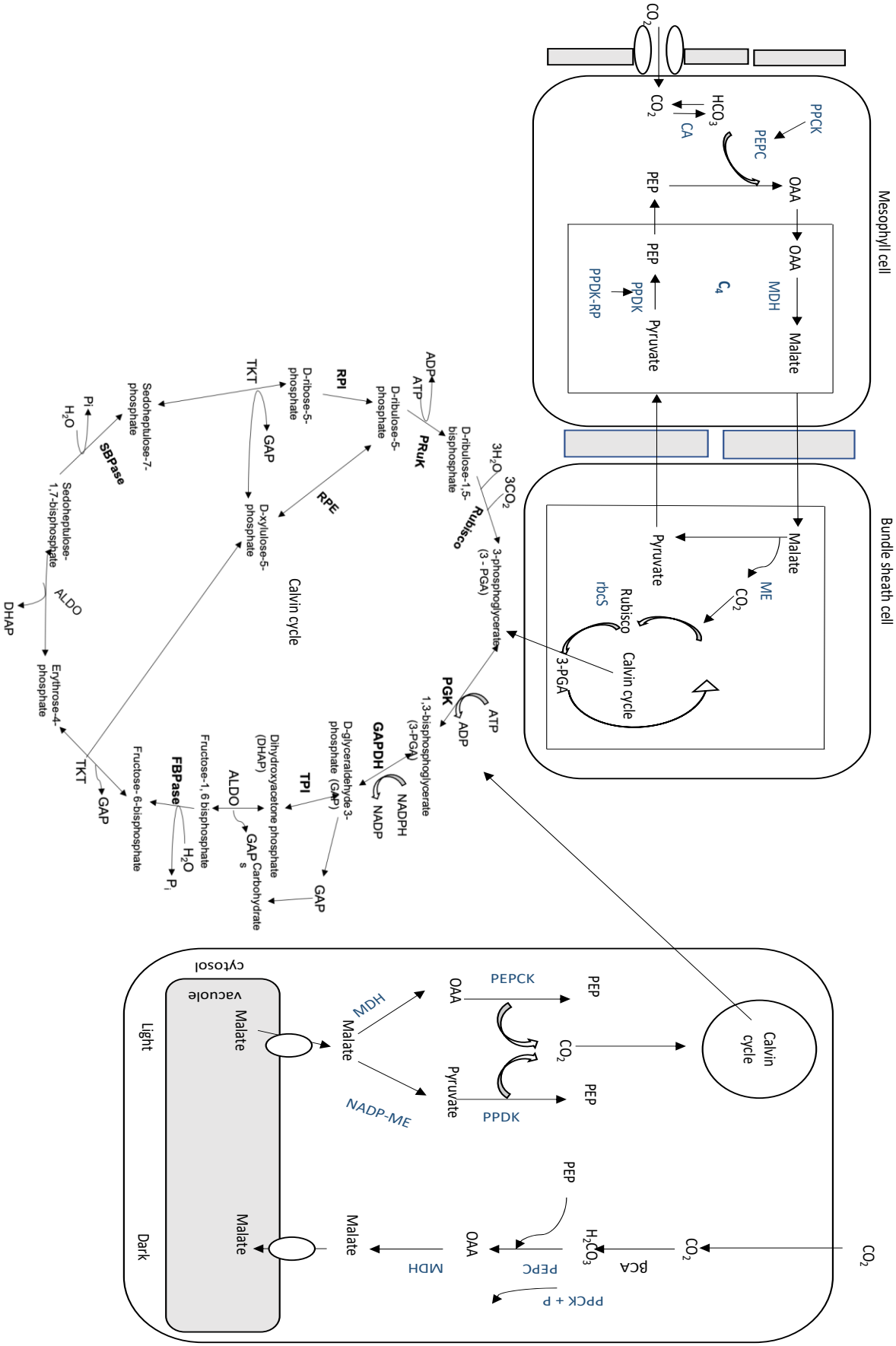



**Figure III.12:** Diagrammatic representation of candidate genes of Carbon fixation pathway ($C_3$, $C_4$, CAM cycle)

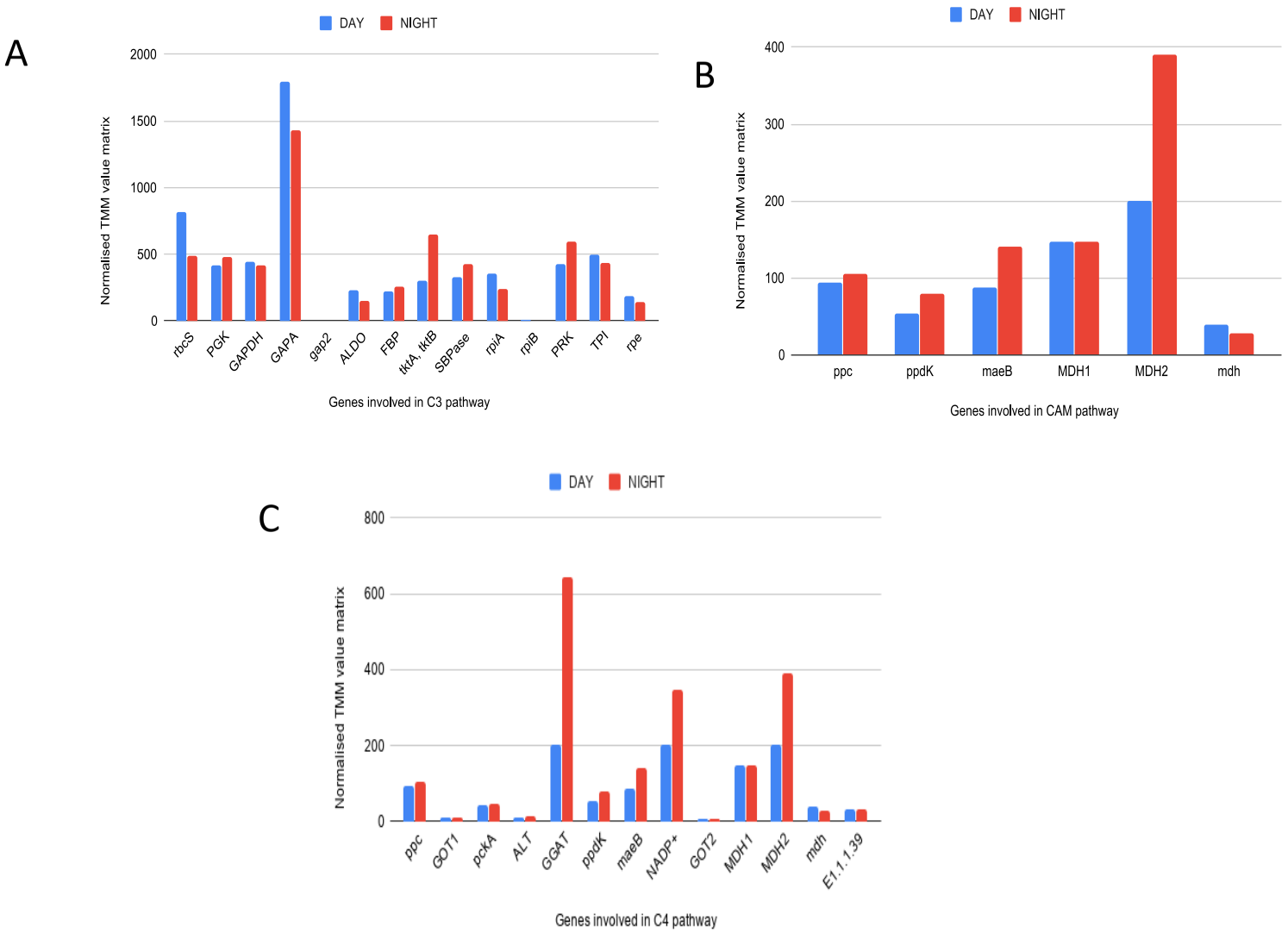



**Figure III.13:** The candidate genes involved in the $C_3$, CAM and $C_4$ cycle.

A) Calvin- Benson ($C_3$) cycle B) Crassulacean acid metabolism (CAM) cycle C) $C_4$ cycle. X – axis represents the genes involved pathways, Y – axis is the matrix of normalised expression trimmmed mean of M (TMM) values; Blue graph – leaf tissue collected during day period (2 PM), Red graph – leaf tissue collected during night period (2 AM)

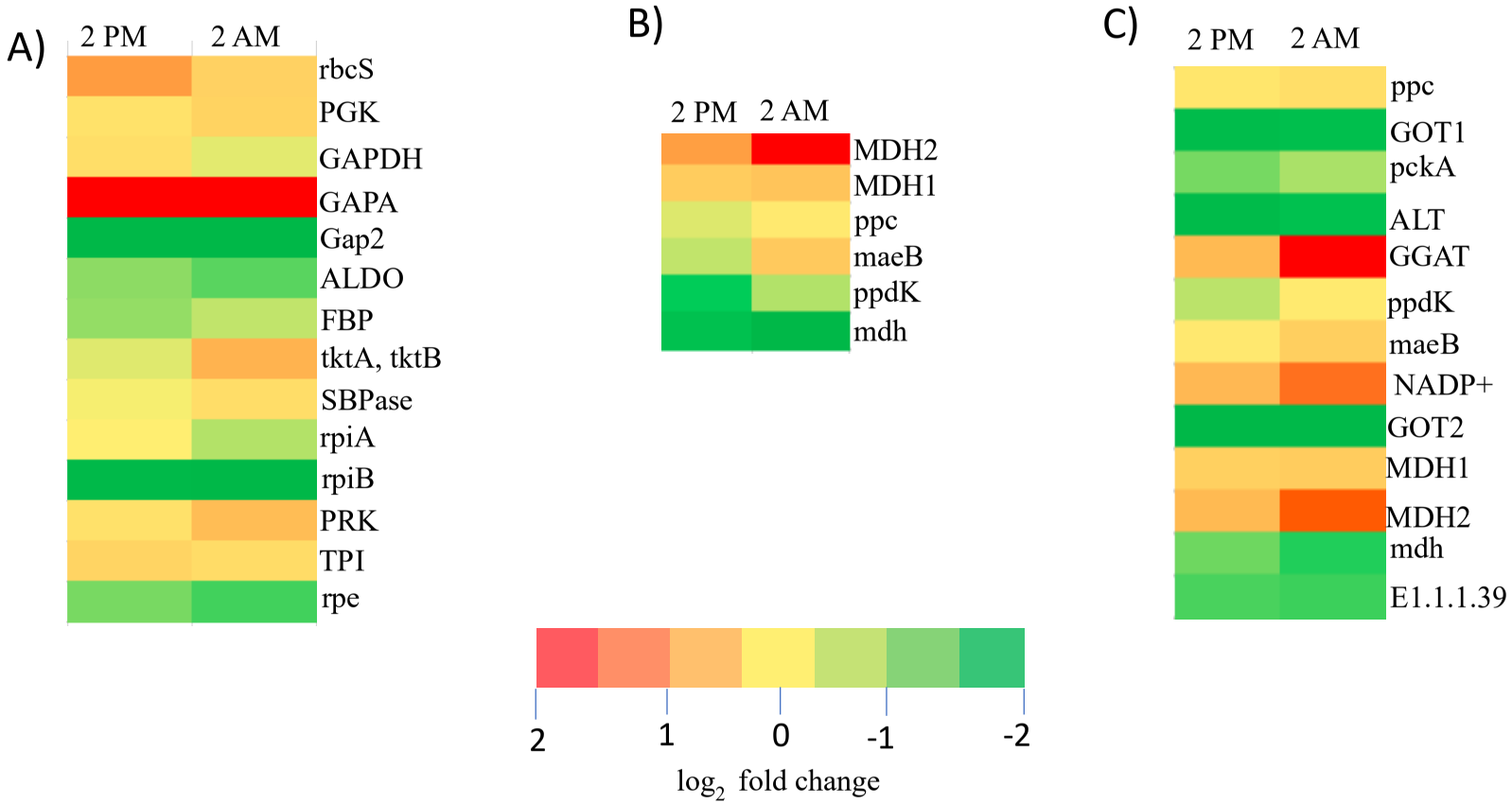



**Figure III.14:** Gene expression pattern of *F. religiosa* carbon fixation genes across the diurnal (2 PM) and nocturnal (2 AM) expression data. A) $C_3$ cycle, B) CAM cycle C) $C_4$ cycle with fold change log2-transformed Fragments Per Kilobase of transcript per Million mapped reads (FPKM) value based expression profiles are shown

The important genes expressed in the $C_3$ cycle are rubisco and glyceraldehyde-3-phosphate dehydrogenase. Ribulose-bisphosphate carboxylase (RuBP carboxylase or *rubisco)* small chain enzyme that is enriched in leaf tissue collected during the day (2 PM). Rubisco is the most abundant protein in chloroplasts. The glyceraldehyde-3-phosphate dehydrogenase (NADP+) is enriched in day sample leaf tissue, the enzyme responsible for the reversible conversion of glyceraldehyde 3-phosphate to ribulose bisphosphate using ATP, the acceptor for CO2. The trancriptomic genes have been mapped to the $C_3$ cycle except for the three genes mentioned above. [60].

The signature genes responsible for the CAM cycle were expressed in the Peepal tree during the night. The phosphoenolpyruvate carboxylase kinase (PPCK), NAD(P)-ME (maeB), and Malate dehydrogenase (MDH) transcripts were highly enriched in the photosynthetic leaf tissue collected during the night period than the day. It indicates that the Peepal tree adapts to the CAM pathway and can fix nocturnal carbon dioxide using the PEP carboxylase (PEPC) enzyme and accumulate malate by the enzyme malate dehydrogenase. The transcriptomic genes of the Peepal tree have been completely mapped to the KEGG pathway of the CAM cycle.

In the $C_4$ Dicarboxylic cycle, the high expression of glutamate-glyoxylate aminotransferase enzyme (GGAT) in the leaf tissue collected during the night period (2 AM) indicates the photorespiration in the Peepal tree. The carbon fixation begins in the mesophyll cells, where CO2 is converted into bicarbonate. It adds the 3-carbon acid phosphoenolpyruvate (PEP) by an enzyme called phosphoenolpyruvate carboxylase. The product of this reaction is the four-carbon acid oxaloacetate, which is reduced to malate another four-carbon acid [61]. The second highest expression is NADP-malate dehydrogenase (MDH), which converts the oxaloacetate generated by PEPC to malate. The differentially expressed genes from the Peepal tree had a complete mapping to the $C_4$ cycle. The gene expression pattern of the carbon-fixation pathway

in the Peepal tree suggests that the plant switches between the $C_3$, $C_4$, and CAM cycles during the diurnal and nocturnal periods. The FPKM and Trimmed Mean of M-values (TMM) values of the differentially expressed genes for the carbon fixation pathway are shown in Figures III.13A, B, C, and III.14A, B, and C.

# CHAPTER IV

# DISCUSSION

## IV.1 SUMMARY

This study generated and annotated the genomics and transcriptomics data for the keystone species Peepal tree (*F. religiosa*). This plant species is well known for Buddha's enlightenment while he was meditating underneath this tree. This tree is being worshipped as birth giving, regenerative and medicinally valued for many diseases and this culture is spread across Asia. Generally, Peepal tree is known for high production of oxygen throughout the day as well as during night time, but there is a lack of scientific evidence. We used two next-generation sequencing technologies to sequence the Peepal genome and characterized the hybrid whole genome. The assembled genome resulted in a size of 406 Mb with 35,093 protein-coding genes, based on *ab initio*, homology, and mRNA evidence used for annotation. A total of about 53% of the genome consisted of the repetitive sequences. Photosynthetic tissues at distinct conditions (diurnal and nocturnal) were used for RNA sequencing to understand the genes, proteins, and molecular pathways. The transcriptome analysis yielded 26,691 unique transcripts. The completeness of the Peepal tree genome was confirmed based on BUSCO analysis and comparative analysis of transcriptome data. Based on BUSCO analysis, 84.1% of genes had completeness of Peepal tree genome assembly for the conserved genes in plant universal single-copy orthologs data set (embryophyta database) and 81.5% had single-copy orthologs in the genome; and 76.5% which indicates the completeness of Peepal tree genome assembly for the conserved genes in eukaryotic universal single-copy orthologs data set (eukaryota database) and 70.6% had single- copy orthologs in the genome, it has provided the confidence for the downstream analyses. We compared the genomes of the three Moraceae family members, *F. carica*, *F. microcarpa*, and *M. notabilis*, with the Peepal tree. According

to the findings, the Peepal genome is considerably more similar to the genera Ficus (*F. carica* and *F. microcarpa*) than it is to the related genus Morus.

The downstream analysis has been carried out to understand the molecular functions in the photosynthesis of the Peepal tree in the diurnal and nocturnal periods. We performed the downstream analysis of genomic and transcriptomic data to understand the microRNAs, TFs, and molecular pathways of the Peepal tree. The miRNA MIR408 was identified to be specially expressed in the in Peepal leaf tissue during the night period. The miRNA MIR408 responds to copper deficiency and light in Arabidopsis [62]. In *O. sativa*, MiR408 plants were efficient at saving and converting light energy into sugars, suggesting that miR408 can promote photosynthesis by down-regulating the uclacyanin (*UCL8)* gene [63]. Thus, MIR408 found specific expression in transcripts of night period leaf tissue of the Peepal tree indicating the similar conversion of light energy and accumulation of sugars at night. It may also aid in photosynthesis by enhancing carbon fixation.

In the current study, Catalase gene expression was found to be high in the night period transcripts of Peepal tree leaf tissue. A previous study on the Peepal tree showed that leaf tissue collected at night time exhibited the scotoactive opening of stomata during the night, which indicates that through the stomatal opening molecular oxygen ($O_2$) is released by the action of catalase enzyme on hydrogen peroxide ($H_2O_2$) [64]. The physiological interaction between catalase and its substrate $H_2O_2$ in the plant was determined by quantifying $H_2O_2$ and assaying the catalase, in which catalase showed a 4-fold increase in activity, especially during the night. Peepal tree has a higher amount of $H_2O_2$ deposition during the night than day [64], which is an indication of pathway switching between carbon fixation pathways.

The RNA sequencing from diurnal (2 PM) and nocturnal (2 AM) leaf samples showed the gene expression patterns of the carbon fixation pathway. The gene expression of mRNA in the $C_3$,

$C_4$ and CAM cycles indicated that depending on the carbohydrate, amino acids biosynthesis and metabolism and environmental conditions the plant switches between these three cycles in a time-structured manner.

The day mRNA expression data suggested Peepal tree can carry out the diurnal carbon fixation by the $C_3$ cycle. GGAT plays an important role in the biosynthesis and metabolism of major amino acids. GGAT is involved in the photorespiratory process. It catalyzes the reaction of glutamate and glyoxylate in the 2-oxoglutarate and glycine .High expression of GGAT in the $C_4$ cycle indicates that there could be photorespiration in the Peepal tree during the night. Plants adapt to the CAM cycle to grow during water constraints and increase the level of carbon dioxide uptake than their $C_3$ and $C_4$ cycles [65]. The Peepal tree study provides information on plants using the CAM pathway to fix nocturnal carbon dioxide using the PEP carboxylase (PEPC) enzyme and the accumulation of malate by the enzyme malate dehydrogenase.

The Peepal tree gene expression analysis for the $C_3$, $C_4$, and CAM cycles suggested that plants could switch between these three cycles depending on the carbohydrate, amino acids biosynthesis, metabolism, and environmental conditions. In the *Kalanchoë fedtschenkoi* genome study, the convergence in protein sequence and re-scheduling of diel transcript expression of genes was reported to be involved in nocturnal $CO_2$ fixation, stomatal movement, heat tolerance, circadian clock, and carbohydrate metabolism with the other CAM species in comparison with non-CAM species [20]. Some of the previous studies in the pineapple genome revealed the gene lineage transition from C3 photosynthesis to CAM, and CAM-related genes exhibit a diel expression pattern in photosynthetic tissues [19]. The evolution of CAM in *Agave* from $C_3$ photosynthesis shows that the core metabolic components required for CAM have ancient genomic origins which could be traceable to non-vascular plants while regulatory

proteins required for diel re-programming of metabolism have shared among the recent origin of $C_3$, $C_4$, and CAM species [66].

The plant model Arabidopsis encodes several orthologues of human proteins that function in mechanisms similar to those in other eukaryotes [67] [68]. Previous findings showed that 70% of oncogenes involved in cancer have orthologs in Arabidopsis, 67% in *D. melanogaster*, 72% in *C. elegans*, and 41% in *S. cerevisiae*. This 'disease gene' similarity is comparable to that observed in other model organisms [69]. The research in Arabidopsis and many other model systems has led to the discovery or analysis of genes and processes important to human health. Previous studies showed that the Peepal tree has been tested for the treatment of neurodegenerative disorders such as Parkinson's disease and Huntington's disease [9], [10], as well as anti-ulcer activity in animal models. [11]. The studied showed the effects of Peepal tree using *in vivo* behavioral parameters like catalepsy, muscle rigidity, and locomotor activity and its effects on neurochemical parameters (MDA, CAT, SOD, and GSH) in rats [9]. Another study showed that methanol extract of *F. religiosa* has proven anti-inflammatory properties in LPS-induced activation of BV2 microglial cells, and it might have therapeutic potential for various neurodegenerative diseases [15]. In the present work, we found that Peepal genes show similarities with human disease pathways, which can be utilized to further understand the traditional medicinal practices. The gene expression data for carbon fixation pathway is an indication of physiological changes in Peepal tree. This is an indication to support the previous data of stomatal structure of Peepal tree and its one of the reason of storing more moisture, water and gaseous molecule in this plant [6]. Such as, generally believed to plant produces more oxygen and may have helped Buddha's meditation and enlightenment. There is no direct evidence how significantly Peepal tree has influenced Buddha's meditation or enlightenment. Thus, plant research opens up new frontiers in terms of drug development and treatment of

diseases of great importance to human health. Plants seem to be a part of this diverse portfolio of tools necessary to understand fundamental cellular processes.

In summary, the genome data and transcript abundance evidence indicate the molecular switch in the carbon fixation pathway of the Peepal tree (*F. religiosa*) during the day and night periods depending on its physiological and environmental conditions. Our study is a foundation for further experiments to determine the underlying mechanisms in $C_3$, $C_4$, and CAM metabolism.

## IV.2 CONCLUSION

In this study, we generated the genomic and transcriptomic data for Peepal/Bodhi tree. Genomic data pathway analyses identified the genes associated with several physiological, biochemical, metabolic, and disease pathways. Differential expression data from diurnal and nocturnal leaf tissue samples of Peepal revealed gene expression patterns in the carbon fixation pathway during light and dark. The transcript abundance indicates that plants could switch between the three $C_3$, $C_4$, and CAM pathways. Genome, transcriptome resources is provided with the prediction of genes, proteins and its related pathways from Peepal. Accordingly, the Peepal genome has been shown to be closely related to the genera Ficus, and relatively close to the Morus genus of the same family. We have also characterized the genes that are involved in different biological pathways. The well-annotated genome for the Peepal tree will have broader implications for studies regarding the physiology, evolution, conservation of species, and human neurological diseases.

## IV. Future investigations:

### Preliminary Study on Stomata of *Ficus religiosa*

There are few evidences from Peepal tree leaf tissues which have shown the sctoactive opening of the stomata during night [64]. Based on previous studies and the present study data, we hypothesized that the Peepal tree leaf tissues can exhibit opening and closing of stomata throughout day and night. It could be the reason either due to organic acid metabolism, enzymes, gaseous molecules, molecular, physiological and environmental changes in the plant. The objective of this study was to observe the stomatal movement (opening and closing) on an hourly basis for complete 24 hours in a day.

**Materials and Methods:**

We collected leaf samples from a Peepal tree near Patanjali block at The University of Trans-Disciplinary Health Sciences and Technology (TDU), Yelahanka, Bengaluru, every for 24 hours to analyze the stomatal movement. In the months of July 2018, June 2019 and November 2019, we collected the leaf samples. Leaf tissues were peeled from the lower epidermis of the leaf (abaxial side) of *F. religiosa* using the blade and transferred into the petridish containing the water. The specimen was transferred to the slide, a drop of 50% glycerol was added on the peel and covered with cover slip. The leaf peel was observed under the Compound light microscope (Olympus CX-33) under 10X and 40X magnification to confirm the presence of stomata. After the identification of stoma, the specimens were washed with the water and treated 1% saffranine. The specimens were washed again with the water and dehydrated through alcohol-xylol grades ( 25%, 50%, 75%, 100% series for 20 minutes at each grade) and mounted using DPX. The prestaining and staining procedure was performed according to [70]. All permanent slides were examined by Olympus CX-33 Compound light microscope and photographed using Olympus CX-33 magCam camera (at 40X magnification) and magnum software.

**Results:**

From this stomatal experiment we observed that there is an opening and closing of stomata during the diurnal and nocturnal periods. We have provided the information an open and close of the stomata for every hour in the *F. religiosa* leaf tissue. We observed that some stomata are open, slight open, close, few open and close types in the leaf tissues for every hour in 24hrs time format in different seasonal years, the details were given below in the Table 1 and Figure 1A, B, C, D, E, F, G, H, I, J, K and L. We observed that stomata were completely closed during 12 AM, 5 AM, 6 AM, 3 PM, 4 PM, 5 PM, 6 PM, 7 PM, 11 PM; slight opening of stomata during 7 AM, 8 AM, 9 AM, 8 PM, 9 PM, 10 PM, 1 PM (Few open and close), 2 PM (Few open and close); stomata are fully open during 10 AM, 11AM, 12 PM, 2AM, 3AM and 4 AM. This data on the stomatal opening and closing indicates that the stomata in Peepal tree leaf exhibits between the type of stomata such as of photoactive (open in day time) and scotoactive (open in night time) stomata based on its physiological mechanisms. Hence, we suggest further study and investigation is required to confirm this data using high-end microscopes. This preliminary data provides the indication of stomatal movement during day and night periods in *F. religiosa* leaves but further verification is required to confirm more evidently.

**Table IV. 1:** Stomatal movement observation in 24 hours for different seasonal years.

| **Time period** | **July 2018** | **June 2019** | **Nov 2019** | **Observation** |
|---|---|---|---|---|
| 12AM | Close | Close | Close | Close |
| 1AM | Slight open | Slight open | Slight open | Slight open |
| 2AM | Open | Open | Open | Open |
| 3AM | Open | Open | Open | Open |
| 4AM | Open | Open | Open | Open |

| | | | | |
|---|---|---|---|---|
| 5AM | Close | Close | Close | Close |
| 6AM | Close | Close | Not taken | Close |
| 7AM | Slight open | Slight open | Not taken | Slight open |
| 8AM | Slight open | Slight open | Not taken | Slight open |
| 9AM | Slight open | Slight open | Not taken | Slight open |
| 10AM | Open | Open | Open | Open |
| 11AM | Open | Open | Open | Open |
| 12PM | Open | Open | Open | Open |
| 1PM | Close | Open | Few open & close | Few open & close |
| 2PM | Few open & close | Few open & close | Few open & close | Few open & close |
| 3PM | Close | Close | Close | Close |
| 4PM | Close | Close | Close | Close |
| 5PM | Close | Close | Close | Close |
| 6PM | Close | Close | Close | Close |
| 7PM | Close | Close | Close | Close |
| 8PM | Slight open | Slight open | Slight open | Slight open |
| 9PM | Slight open | Slight open | Slight open | Slight open |
| 10PM | Slight open & close | Slight open & close | Slight open & close | Slight open & close |
| 11PM | Close | Close | Close | Close |

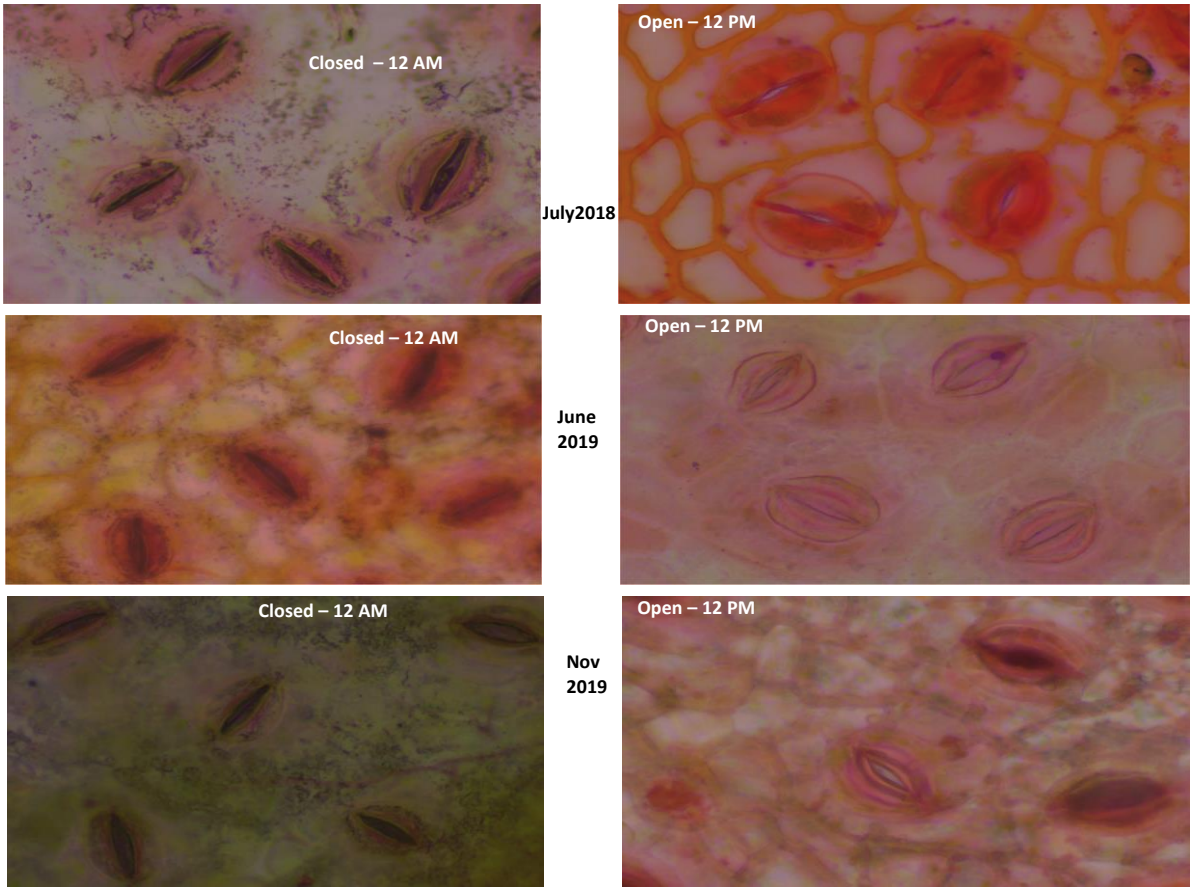



**Figure IV. 1A:** Distribution stomata from lower epidermis of Peepal tree leaf tissue collected at 12 AM and 12 PM(July 2018, June 2019 and November 2019)

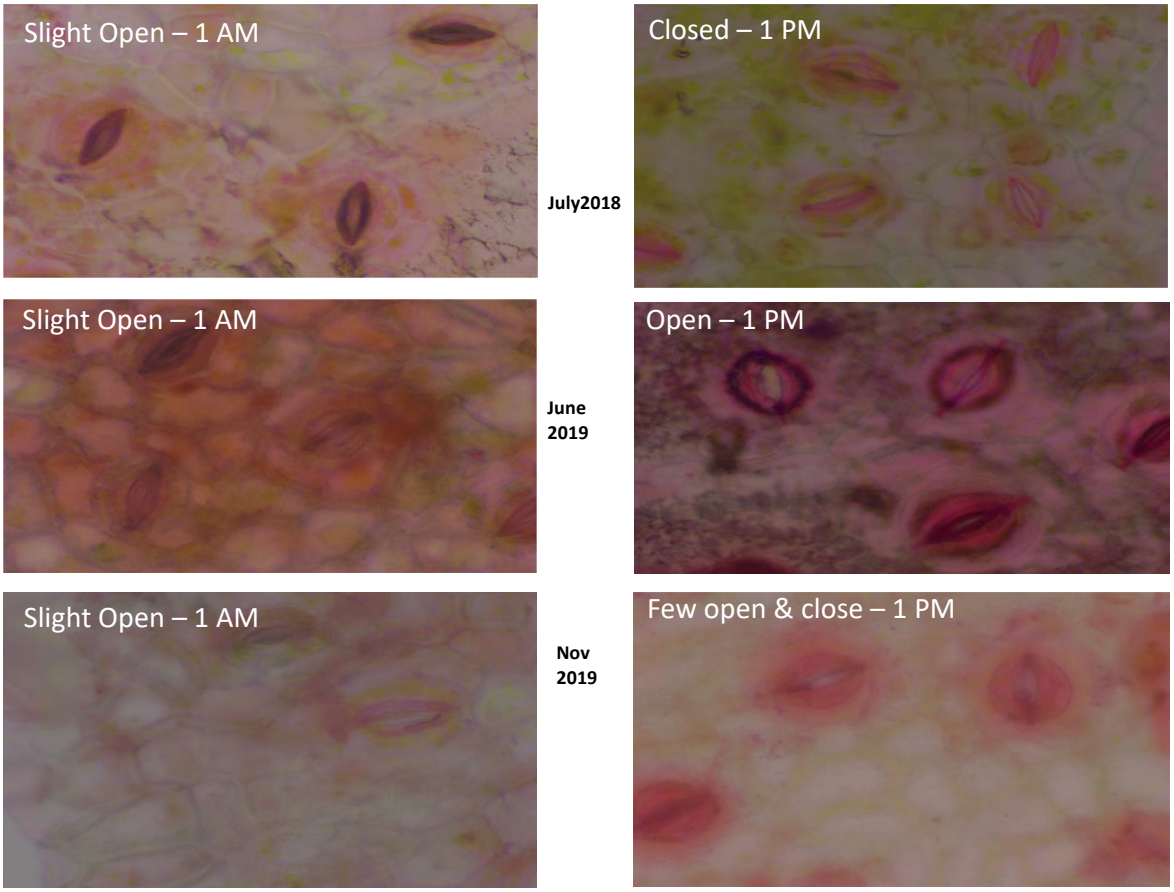



**Figure IV. 1B:** Distribution stomata from lower epidermis of Peepal tree leaf tissue collected at 1 AM and 1 PM (July 2018, June 2019 and November 2019)

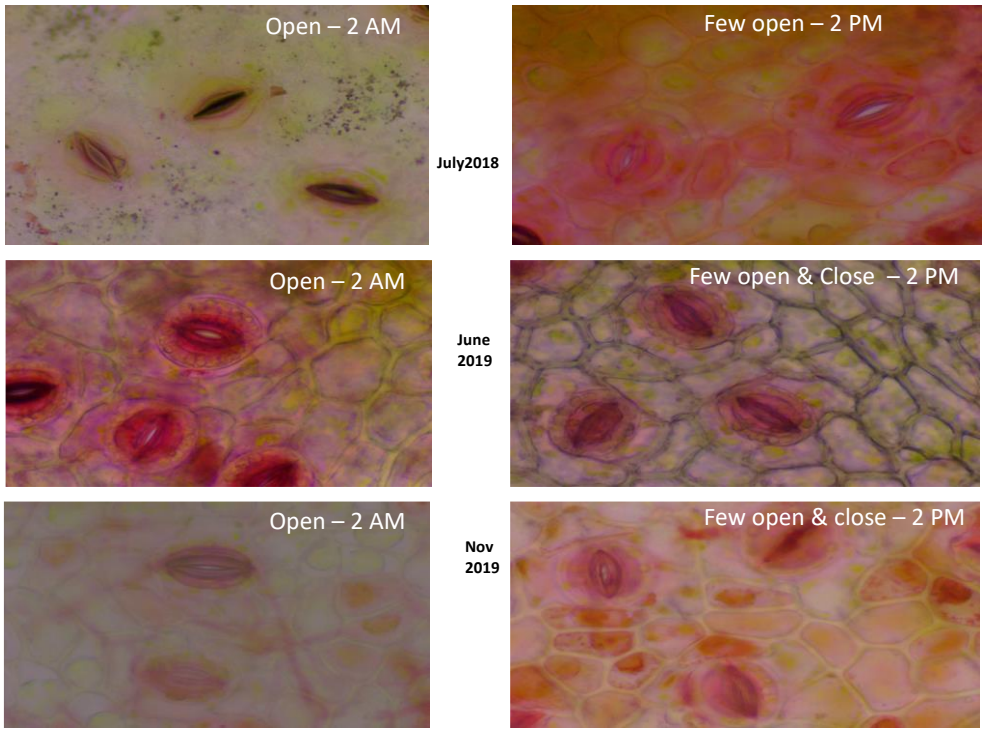



**Figure IV.1C:** Distribution stomata from lower epidermis of Peepal tree leaf tissue collected at 2 AM and 2 PM (July 2018, June 2019 and November 2019)

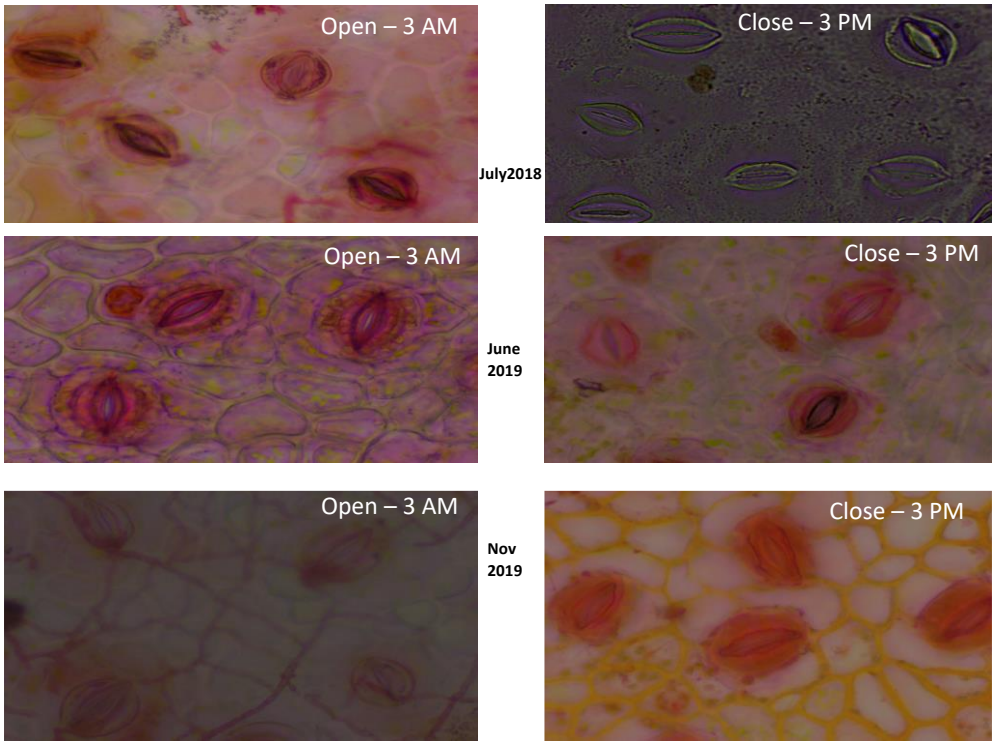



**Figure IV. 1D:** Distribution stomata from lower epidermis of Peepal tree leaf tissue collected at 3 AM and 3 PM (July 2018, June 2019 and November 2019)

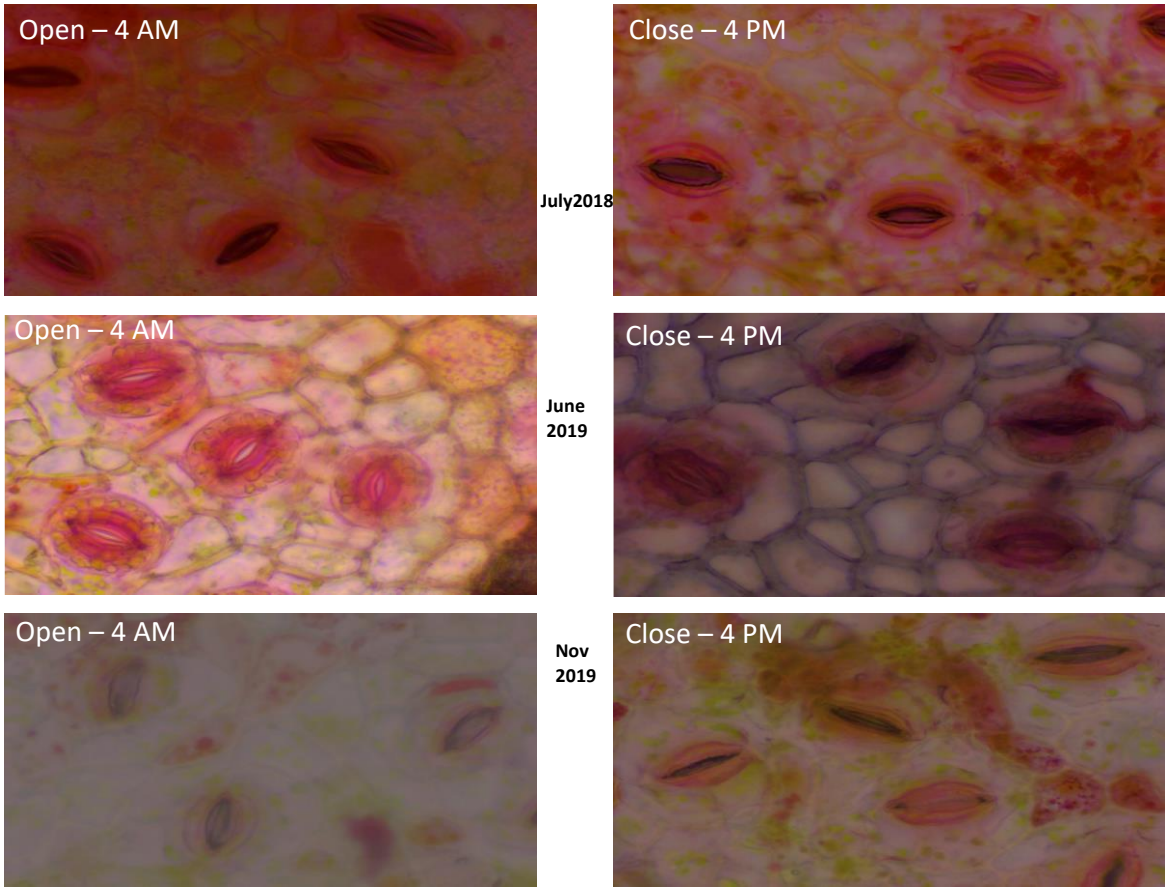



**Figure IV. 1E:** Distribution stomata from lower epidermis of Peepal tree leaf tissue collected at 4 AM and 4 PM(July 2018, June 2019 and November 2019)

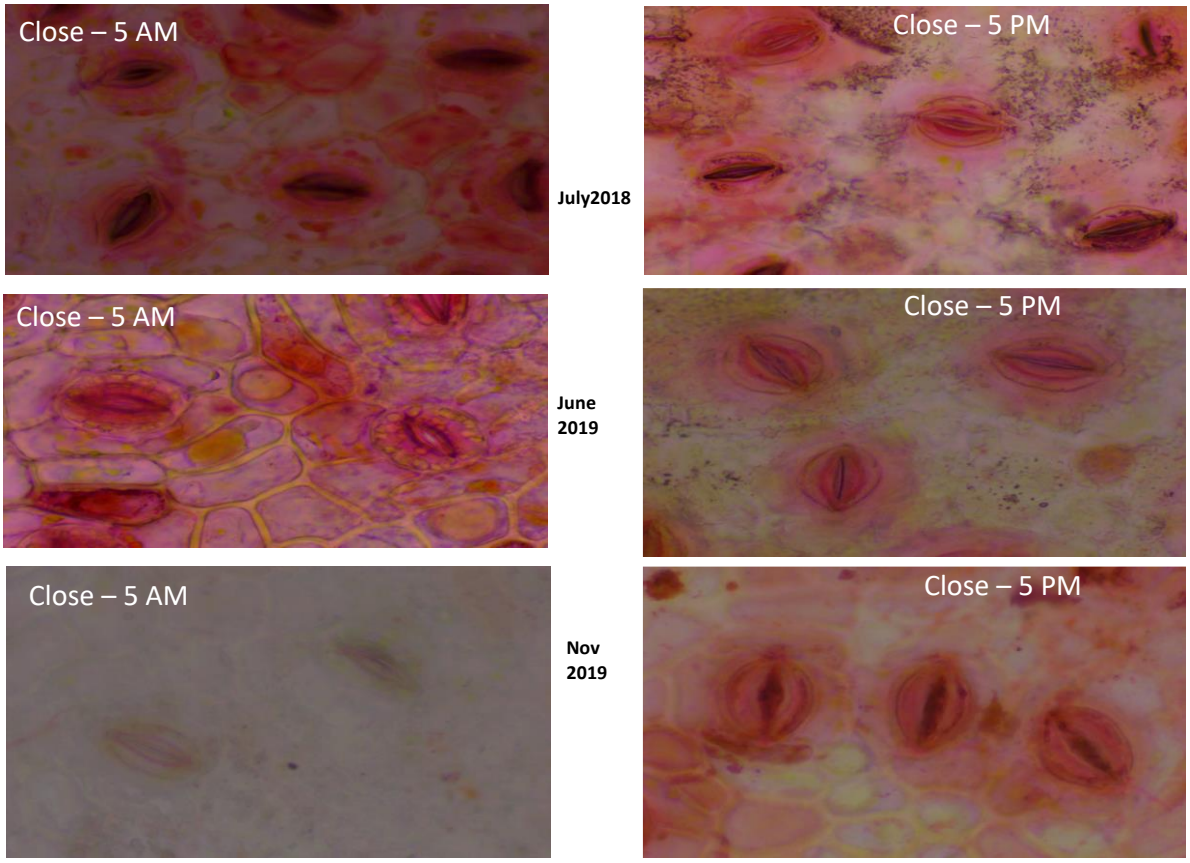



**Figure IV. 1F:** Distribution stomata from lower epidermis of Peepal tree leaf tissue collected at 5 AM and 5 PM(July 2018, June 2019 and November 2019)

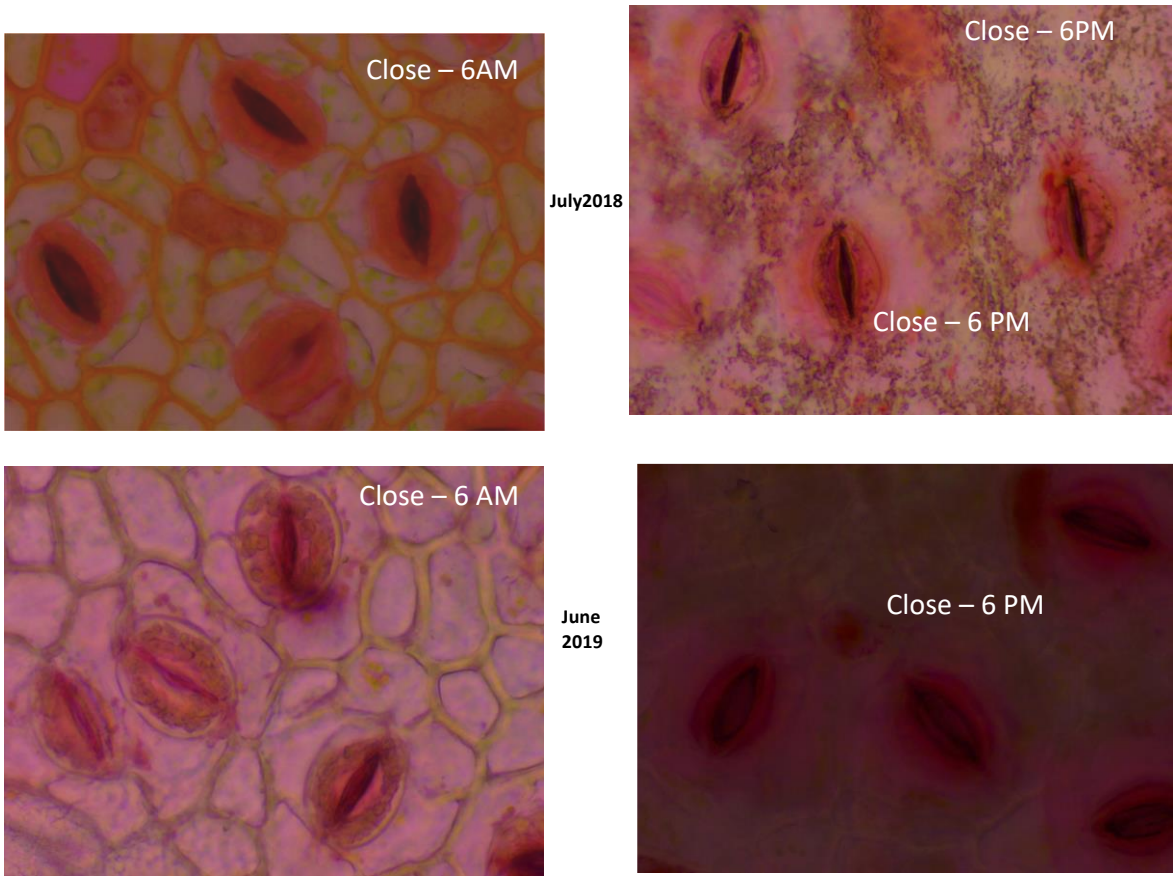



**Figure IV. 1G:** Distribution stomata from lower epidermis of Peepal tree leaf tissue collected at 6 AM and 6 PM(July 2018 and June 2019)

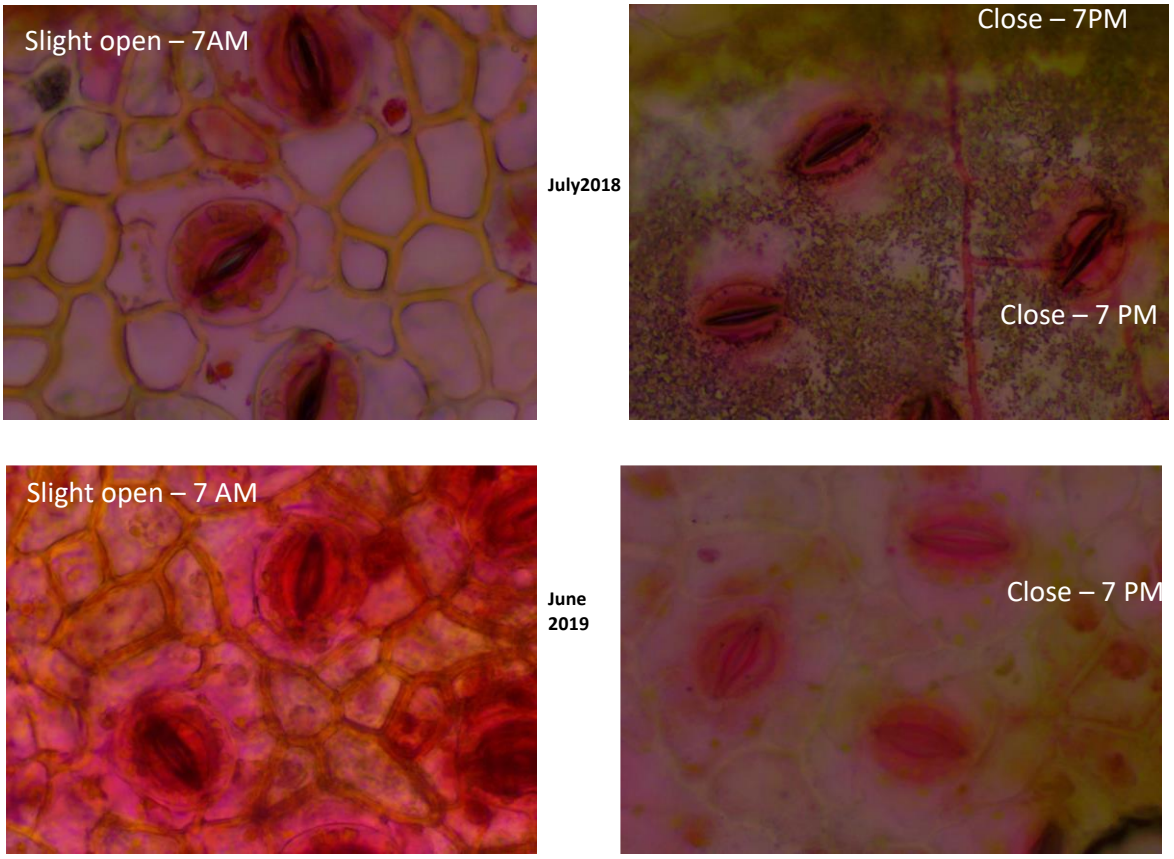



**Figure IV. 1H:** Distribution stomata from lower epidermis of Peepal tree leaf tissue collected at 7 AM and 7 PM(July 2018 and June 2019)

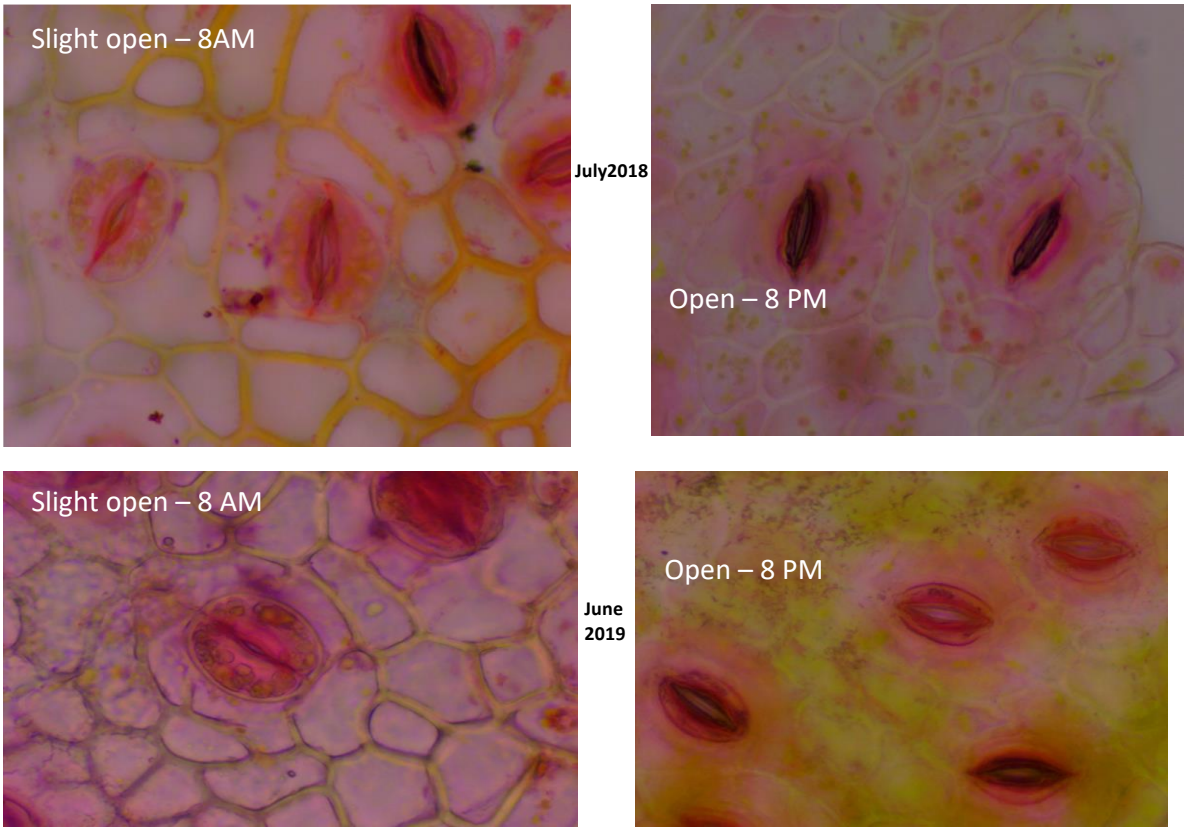



**Figure IV. 1I:** Distribution stomata from lower epidermis of Peepal tree leaf tissue collected at 8 AM and 8 PM(July 2018 and June 2019)

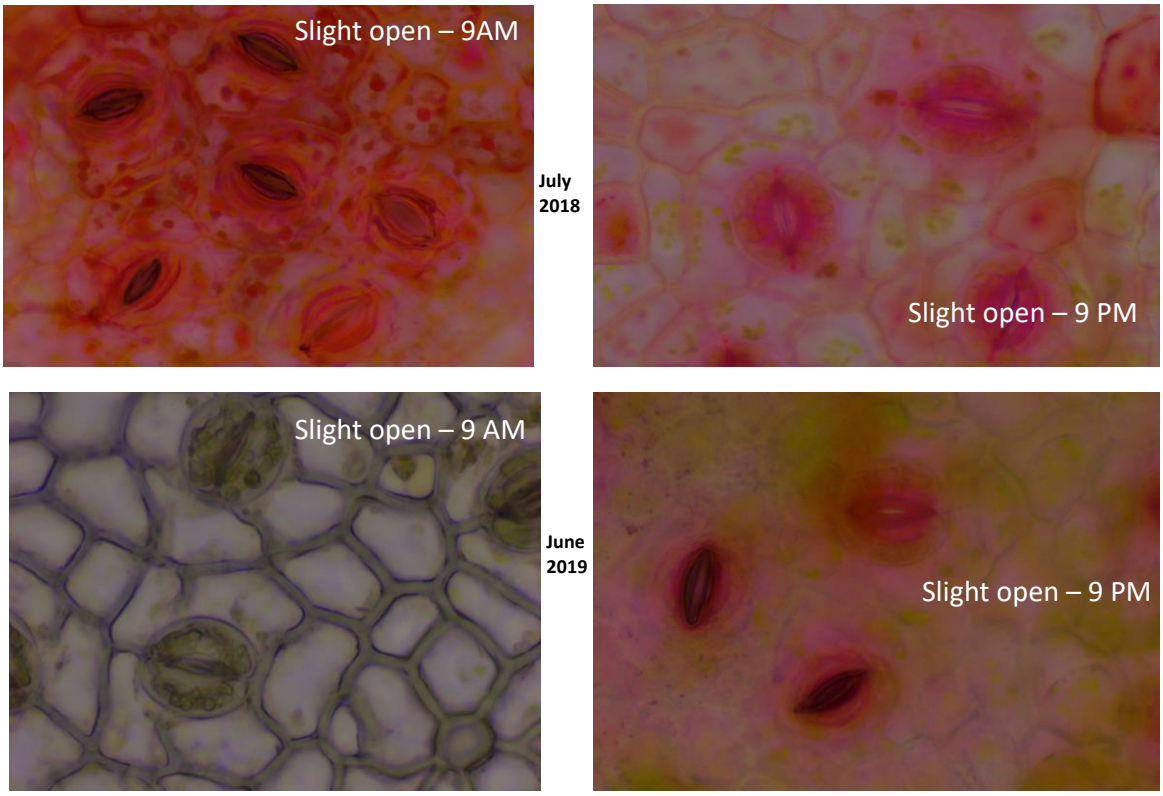



**Figure IV. 1J:** Distribution stomata from lower epidermis of Peepal tree leaf tissue collected at 9 AM and 9 PM(July 2018 and June 2019)

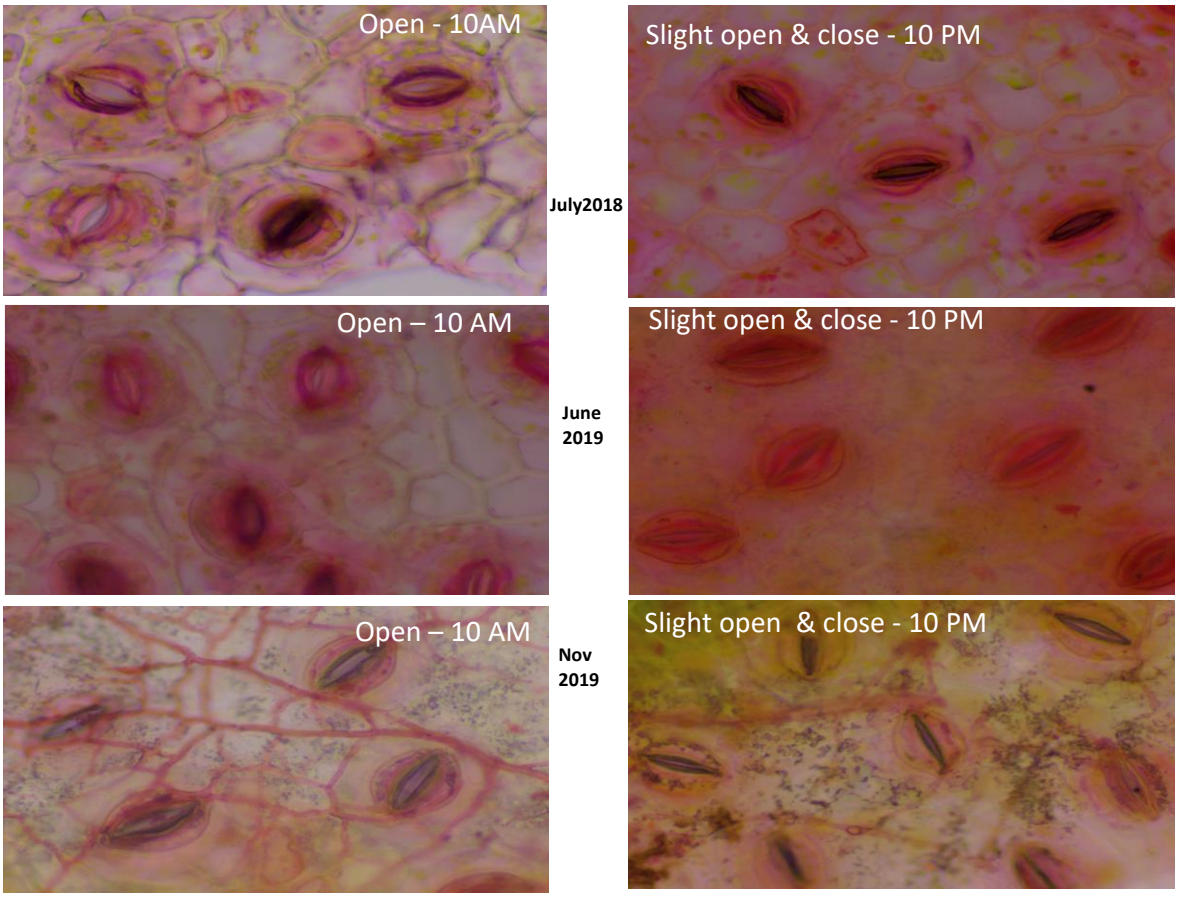



**Figure IV. 1K:** Distribution stomata from lower epidermis of Peepal tree leaf tissue collected at 10 AM and 10 PM(July 2018, June 2019 and November 2019)

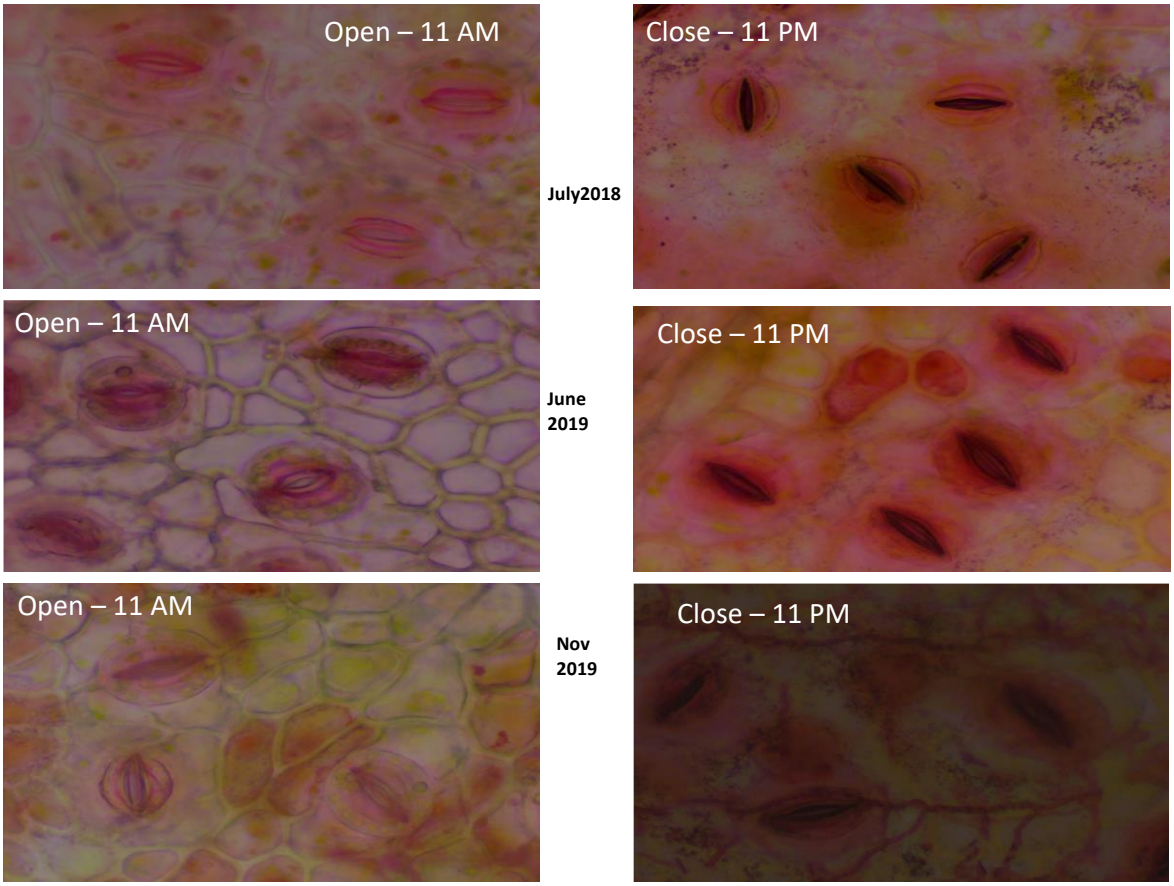



**Figure IV. 1L:** Distribution stomata from lower epidermis of Peepal tree leaf tissue collected at 11 AM and 11 PM(July 2018 and June 2019)

**Summary:**

From the Peepal leaf tissues study, we observed the stomatal opening and closing in diurnal and nocturnal periods. It indicates the physiological changes happening in the leaf tissues of Peepal tree during day and night. The stomatal opening and closing observations from this study indicates that the stomata in Peepal tree leaf might exhibits both photoactive (open in day time) and scotoactive (open in night time) stomata. This stomatal changes observed in the present study is an indicative of an oxygen synthesis during throughout the day and night. This data acts as an substantial support to the previous studies of Peepal tree stomata are sunken, giant, or hydathode at the lower leaf epidermis. Such stomata have been shown to

retain gas and water molecules for a longer [6]. It could be one of main reason that under the Peepal tree, there is much cooler condition and also people believe that this tree produces more oxygen throughout the day and night. We need further more investigations and high-end microscopical and physiological research on this work to confirm this data more evidently.

**References**

1. Mawa S, Husain K, Jantan I. Ficus carica L. (Moraceae): Phytochemistry, Traditional Uses and Biological Activities. Evidence-Based Complement Altern Med. 2013;2013:1–8.
2. Zhang X, Wang G, Zhang S, Chen S, Wang Y, Wen P, et al. Genomes of the Banyan Tree and Pollinator Wasp Provide Insights into Fig-Wasp Coevolution. Cell. 2020;183:875-889.e17.
3. Chakraborty A, Mahajan S, Bisht MS, Sharma VK. Genome sequencing and comparative analysis of Ficus benghalensis and Ficus religiosa species reveal evolutionary mechanisms of longevity. iScience. 2022;25:105100.
4. Ohri D, Khoshoo TN. Nuclear DNA contents in the genusFicus (Moraceae). Plant Syst Evol. 1987;156:1–4.
5. Van Noort S, Gardiner AJ, Tolley KA. New records of Ficus (Moraceae) species emphasize the conservation significance of inselbergs in Mozambique. South African J Bot. 2007;73:642–9.
6. Chantarasuwan B, Baas P, van Heuven B-J, Baider C, van Welzen PC. Leaf anatomy of Ficus subsection U rostigma (Moraceae). Bot J Linn Soc. 2014;175:259–81.
7. Ashalatha KL, Gowda M. Heritage of Neem–Peepal Tree Resides a Profound Scientific Facts. In: The Neem Genome. Springer, Cham; 2019. p. 13–9.
8. Singh D, Singh B, Goel RK. Traditional uses, phytochemistry and pharmacology of Ficus religiosa: A review. J Ethnopharmacol. 2011;134:565–83.
9. Bhangale JO, Acharya SR. Anti-Parkinson Activity of Petroleum Ether Extract of Ficus religiosa (L.) Leaves. Adv Pharmacol Sci. 2016;2016:1–9.
10. Bhangale JO, Acharya NS, Acharya SR. Protective effect of Ficus religiosa (L.) against 3-nitropropionic acid induced Huntington disease. Orient Pharm Exp Med. 2016;16:165–74.
11. Gregory M, Divya B, Mary RA, Viji MMH, Kalaichelvan VK, Palanivel V. Anti–ulcer activity of Ficus religiosa leaf ethanolic extract. Asian Pac J Trop Biomed. 2013;3:554–6.
12. Vinutha B, Prashanth D, Salma K, Sreeja SL, Pratiti D, Padmaja R, et al. Screening of selected Indian medicinal plants for acetylcholinesterase inhibitory activity. J Ethnopharmacol. 2007;109:359–63.
13. Ahmed F, Manjunath S, Narendra Sharath Chandra J. Acetylcholine and memory-enhancing activity of Ficus racemosa bark. Pharmacognosy Res. 2011;3:246.
14. Kaur H, Singh D, Singh B, Goel RK. Anti-amnesic effect of Ficus religiosa in scopolamine-induced anterograde and retrograde amnesia. Pharm Biol. 2010;48:234–40.
15. Jung HW, Son HY, Minh C Van, Kim YH, Park Y-K. Methanol extract of Ficus leaf inhibits the production of nitric oxide and proinflammatory cytokines in LPS-stimulated microglia via the MAPK pathway. Phyther Res. 2008;22:1064–9.
16. He N, Zhang C, Qi X, Zhao S, Tao Y, Yang G, et al. Draft genome sequence of the mulberry tree Morus notabilis. Nat Commun. 2013;4:2445.
17. Usai G, Mascagni F, Giordani T, Vangelisti A, Bosi E, Zuccolo A, et al. Epigenetic patterns within the haplotype phased fig ( Ficus carica L.) genome. Plant J. 2020;102:600–14.
18. Abhisek Chakraborty, Shruti Mahajan, Manohar Singh Bisht VKS. Genome sequencing and comparative analysis of Ficus benghalensis and Ficus religiosa trees reveal evolutionary mechanisms of longevity. https://doi.org/doi: https://doi.org/10.1101/2021.10.14.464369.
19. Ming R, VanBuren R, Wai CM, Tang H, Schatz MC, Bowers JE, et al. The pineapple genome and the evolution of CAM photosynthesis. Nat Genet. 2015;47:1435–42.
20. Yang X, Hu R, Yin H, Jenkins J, Shu S, Tang H, et al. The Kalanchoë genome provides insights into convergent evolution and building blocks of crassulacean acid metabolism. Nat Commun. 2017;8:1899.
21. Guan Q, Tan B, Kelley TM, Tian J, Chen S. Physiological Changes in

Mesembryanthemum crystallinum During the C3 to CAM Transition Induced by Salt Stress. Front Plant Sci. 2020;11:283.
22. Schlüter U, Denton AK, Bräutigam A. Understanding metabolite transport and metabolism in C4 plants through RNA-seq. Curr Opin Plant Biol. 2016;31:83–90.
23. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011;27:764–70.
24. Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, et al. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics. 2017;33:2202–4.
25. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol. 2012;19:455–77.
26. Zimin A V., Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. The MaSuRCA genome assembler. Bioinformatics. 2013;29:2669–77.
27. Boetzer M, Henkel C V., Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics. 2011;27:578–9.
28. Kosugi S, Hirakawa H, Tabata S. GMcloser: closing gaps in assemblies accurately with a likelihood-based selection of contig or long-read alignments. Bioinformatics. 2015;:btv465.
29. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29:1072–5.
30. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210–2.
31. Campbell MS, Law M, Holt C, Stein JC, Moghe GD, Hufnagel DE, et al. MAKER-P: A Tool Kit for the Rapid Creation, Management, and Quality Control of Plant Genome Annotations. Plant Physiol. 2014;164:513–24.
32. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30:1236–40.
33. Xu L, Dong Z, Fang L, Luo Y, Wei Z, Guo H, et al. OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res. 2019;47:W52–8.
34. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. 2019;20:1160–6.
35. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.
36. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10:giab008.
37. Thiel T, Michalek W, Varshney R, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). Theor Appl Genet. 2003;106:411–22.
38. Jin J, Tian F, Yang D-C, Meng Y-Q, Kong L, Luo J, et al. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res. 2017;45:D1040–5.
39. Lowe TM, Eddy SR. tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. Nucleic Acids Res. 1997;25:955–64.
40. Kong L, Zhang Y, Ye Z-Q, Liu X-Q, Zhao S-Q, Wei L, et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. 2007;35 suppl_2:W345–9.
41. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol.

2011;29:644–52.
42. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013;8:1494–512.
43. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22:1658–9.
44. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28:511–5.
45. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323.
46. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21:3674–6.
47. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 2007;35 Web Server:W182–5.
48. Italiani VCS, da Silva Neto JF, Braz VS, Marques M V. Regulation of Catalase-Peroxidase KatG Is OxyR Dependent and Fur Independent in Caulobacter crescentus. J Bacteriol. 2011;193:1734–44.
49. Liu C, Xu X, Kan J, Cheng Z ming, Chang Y, Lin J, et al. Genome-wide analysis of the C3H zinc finger family reveals its functions in salt stress responses of Pyrus betulaefolia. PeerJ. 2020;8:e9328.
50. Lee D-K, Kim H Il, Jang G, Chung PJ, Jeong JS, Kim YS, et al. The NF-YA transcription factor OsNF-YA7 confers drought stress tolerance of rice in an abscisic acid independent manner. Plant Sci. 2015;241:199–210.
51. Bolt S, Zuther E, Zintl S, Hincha DK, Schmülling T. ERF105 is a transcription factor gene of Arabidopsis thaliana required for freezing tolerance and cold acclimation. Plant Cell Environ. 2017;40:108–20.
52. Rawat R, Schwartz J, Jones MA, Sairanen I, Cheng Y, Andersson CR, et al. REVEILLE1, a Myb-like transcription factor, integrates the circadian clock and auxin pathways. Proc Natl Acad Sci. 2009;106:16883–8.
53. Mizoguchi T, Wheatley K, Hanzawa Y, Wright L, Mizoguchi M, Song H-R, et al. LHY and CCA1 Are Partially Redundant Genes Required to Maintain Circadian Rhythms in Arabidopsis. Dev Cell. 2002;2:629–41.
54. Pan J, Huang D, Guo Z, Kuang Z, Zhang H, Xie X, et al. Overexpression of microRNA408 enhances photosynthesis, growth, and seed yield in diverse plants. J Integr Plant Biol. 2018;60:323–40.
55. Schramm M, Klybas V, Racker E. Phosphorolytic cleavage of fructose-6-phosphate by fructose-6-phosphate phosphoketolase from Acetobacter xylinum. J Biol Chem. 1958;233:1283–8.
56. Ebata M, Sato R, Bak T. The enzymatic phosphorylation of sedoheptulose. J Biochem. 1955;42:715–25.
57. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000;28:27–30.
58. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. Protein Sci. 2019;28:1947–51.
59. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. 2023;51:D587–92.
60. McFarlane CR, Shah NR, Kabasakal B V., Echeverria B, Cotton CAR, Bubeck D, et al.

Structural basis of light-induced redox regulation in the Calvin–Benson cycle in cyanobacteria. Proc Natl Acad Sci. 2019;116:20984–90.
61. Cousins AB, Baroli I, Badger MR, Ivakov A, Lea PJ, Leegood RC, et al. The Role of Phospho enol pyruvate Carboxylase during C4 Photosynthetic Isotope Exchange and Stomatal Conductance. Plant Physiol. 2007;145:1006–17.
62. Abdel-Ghany SE, Pilon M. MicroRNA-mediated Systemic Down-regulation of Copper Protein Expression in Response to Low Copper Availability in Arabidopsis. J Biol Chem. 2008;283:15932–45.
63. Zhang J-P, Yu Y, Feng Y-Z, Zhou Y-F, Zhang F, Yang Y-W, et al. MiR408 Regulates Grain Yield and Photosynthesis via a Phytocyanin Protein. Plant Physiol. 2017;175:1175–85.
64. Smitha RB, Bennans T, Mohankumar C, Benjamin S. Oxidative stress enzymes in Ficus religiosa L.: Biochemical, histochemical and anatomical evidences. J Photochem Photobiol B Biol. 2009;95:17–25.
65. NOBEL PS. Achievable productivities of certain CAM plants: basis for high values compared with C 3 and C 4 plants. New Phytol. 1991;119:183–205.
66. Yin H, Guo H-B, Weston DJ, Borland AM, Ranjan P, Abraham PE, et al. Diel rewiring and positive selection of ancient plant proteins enabled evolution of CAM photosynthesis in Agave. BMC Genomics. 2018;19:588.
67. Hays JB. Arabidopsis thaliana, a versatile model system for study of eukaryotic genome-maintenance functions. DNA Repair (Amst). 2002;1:579–600.
68. Spampinato CP, Gomez-Casati DF. Research on Plants for the Understanding of Diseases of Nuclear and Mitochondrial Origin. J Biomed Biotechnol. 2012;2012:1–12.
69. Jones AM, Chory J, Dangl JL, Estelle M, Jacobsen SE, Meyerowitz EM, et al. The Impact of Arabidopsis on Human Health: Diversifying Our Portfolio. Cell. 2008;133:939–43.
70. Johansen DA. Plant microtechnique. London; McGraw-Hill Book Company, Inc:; 1940.

**APPENDIX I**

Research Article

# Genomic and transcriptomic analysis of sacred fig (*Ficus religiosa*)

Ashalatha *et al. BMC Genomics* (2023) 24:197
https://doi.org/10.1186/s12864-023-09270-z

BMC Genomics

RESEARCH Open Access

# Genomic and transcriptomic analysis of sacred fig (*Ficus religiosa*)

K. L. Ashalatha[1], Kallare P Arunkumar[2] and Malali Gowda[1,3*]

**Abstract**

**Background** Peepal/Bodhi tree (*Ficus religiosa* L.) is an important, long-lived keystone ecological species. Communities on the Indian subcontinent have extensively employed the plant in Ayurveda, traditional medicine, and spiritual practices. The Peepal tree is often thought to produce oxygen both during the day and at night by Indian folks. The goal of our research was to produce molecular resources using whole-genome and transcriptome sequencing techniques.

**Results** The complete genome of the Peepal tree was sequenced using two next-generation sequencers Illumina HiSeq1000 and MGISEQ-2000. We assembled the draft genome of 406 Mb, using a hybrid assembly workflow. The genome annotation resulted in 35,093 protein-coding genes; 53% of its genome consists of repetitive sequences. To understand the physiological pathways in leaf tissues, we analyzed photosynthetically distinct conditions: bright sunny days and nights. The RNA-seq analysis supported the expression of 26,479 unigenes. The leaf transcriptomic analysis of the diurnal and nocturnal periods revealed the expression of the significant number of genes involved in the carbon-fixation pathway.

**Conclusions** This study presents a draft hybrid genome assembly for *F. religiosa* and its functional annotated genes. The genomic and transcriptomic data-derived pathways have been analyzed for future studies on the Peepal tree.

**Keywords** *Ficus religiosa*, Peepal, Bodhi, Hybrid genome, Transcriptome, Carbon-fixation pathway

## Background

The Peepal tree (*Ficus religiosa* L.) is a sacred fig, hemi-epiphyte that belongs to the Moraceae family and has a diploid sporophytic chromosome count (2n=26) [1]. It is known to be a long-lived deciduous species related to the 755 fig species widespread worldwide [2]. The Peepal tree is a cosmopolitan species, having value for cultural and spiritual practices in Buddhism, Hinduism, and Jainism. It is popularly called the Bodhi tree, where Buddha is believed to have meditated and attained spiritual enlightenment underneath this tree. Hence, the culture is spread across Asia and it has been worshipped. Peepal has several vernacular names, like Pippali, Ashwatha, Arali, and so on; it is frequently found together with the Neem tree near Indian temples [3]. Generally, the Peepal tree has a special significance in communities across India as it is believed to produce oxygen day and night. They have a special type of stomata called sunken, giant, or hydathode at the lower leaf epidermis. These are larger than the normal stomata and occur over the veins or are mixed with normal stomata. It indicates that such stomata hold

*Correspondence:
Malali Gowda
malali.gene@gmail.com
[1]The University of Trans-Disciplinary Health Sciences and Technology (TDU), Yelahanka, Bengaluru 560064, India
[2]Central Silk Board, Central Muga Eri Research and Training Institute (CMER&TI), Ministry of Textiles Lahdoigarh, Jorhat, Assam 785700, India
[3]The University of Trans-Disciplinary Health Sciences and Technology (TDU), DNA Life Organization, Yelahanka, Bengaluru 560064, India

gaseous and water molecules for a longer time [4]. To our knowledge, there is no reported scientific evidence to claim oxygen production from the Peepal tree at night.

In Ayurveda, the Peepal tree has been classified as a Rasayana (a type of drug), whereby rejuvenators and antioxidants aid in relieving the body's stress [5]. Peepal tree alleviates Pitta and Kapha (Ayurvedic classifications), hence prescribed for treatment of the disorders like respiratory and inflammatory disorders, ulcers, stomatitis, hiccup, arthritis, gout, skin diseases, bone fracture, diabetes, etc., [5]. In animal models such as rats, the Peepal tree has been tested for the treatment of neurodegenerative disorders such as Parkinson's disease and Huntington's disease [6] [7], as well as anti-ulcer activity in albino mice [8].

Next-generation sequencing (NGS) technologies have accelerated the generation of draft genome sequences of Moraceae plant species. The genome size of *Morus notabilis* is 330 Mb [9], 333 Mb in *Ficus carica* [10], 436 Mb in *F. microcarpa* and 370 Mb in *F. hispida* [11]. The genome sequencing of non-model plant species *F. religiosa* was first mentioned in The Neem Genome book chapter [3]. A recent study has generated the genomic resource of *F. religiosa* (332 Mb) and *F. benghalensis* (392 Mb). However, they generated a limited size of genome assembly and genes (23,929) for the Peepal tree when compared to the present study [12]. Recently, a few research groups have attempted sequencing of non-model plant species like pineapple (*Ananas comosus*) and *Kalanchoë* species revealing the gene expressions of the Crassulacean acid metabolism (CAM) pathway [13] [14]. The whole-genome sequencing has shown common or crystalline ice plants (*Mesembryanthemum crystallinum*) to switch from Calvin- Benson Cycle (C3) to CAM photosynthesis under a salt stress [15]. The study described by comparing both species with and without the C4 trait and different tissues within a C4 plant using RNA-seq suggests ways of integration into the underlying C3 metabolism [16]. These findings and other physiological features of the Peepal tree enabled us to characterize the pathways in the present work.

Despite its ecological, medicinal, cultural, and historic importance, the molecular biology and genomics studies on the Peepal tree are scanty. As Peepal is relevant to traditional medicinal practices and Buddha's meditation, we envisaged elucidating the genome sequence and studying the transcriptome of photosynthetic tissue (leaf tissue) in diurnal and nocturnal periods. The objective of the present study was to generate a genome sequence and annotate genes of the Peepal tree. The transcriptomic analysis has been undertaken to identify the expression of genes in the diurnal and nocturnal periods for photosynthetic activity using a molecular approach. In this study, we aimed to characterize the genes involved in various physiological, biochemical metabolic, and other pathways. Also, a comparative genomic analysis has been carried out to study the relationship of the Peepal tree with closely related species of its Moraceae family.

**Table 1** Final assembly and annotation of Peepal genome

| Features | Final assembly |
|---|---|
| Total length of assembled sequence (Mb) | 406.103 |
| Number of scaffolds/contigs | 202,258 |
| Minimum scaffold length (bp) | 200 |
| Maximum scaffold length (bp) | 148,483 |
| GC content (%) | 34.23 |
| N50 (bp) | 5,817 |
| L50 | 17,605 |
| Number of annotated genes | 35,093 |

## Results

### De novo hybrid assembly using Illumina and MGI short reads

We used two next-generation technology platforms to sequence the whole genome of the fig species, the Peepal tree. A total of 266 and 645 million paired-end reads were generated from Illumina HiSeq1000 and MGISEQ-2000 platforms respectively. The data of 88.44 billion high-quality bases (Quality > 20) was used for genome assembly. A hybrid assembly was performed using a sequencing depth of 65.5X Illumina reads and 158.86X MGI reads. The raw data details are given in the supplementary material (Additional file 1: Table S1.1). The evaluation of the distribution of k-mers in both Illumina and MGI reads to estimate the genome size provided genome sizes of 319 Mb and 273 Mb respectively (Additional file 2: Figures S1A and S1B). The combination of Illumina and MGI reads was used for assembling the genome. Hybrid genome assembly yielded a genome of 406 Mb. The contig N50 length is 5,817 bp and the largest contig length is 148 Kb. The GC content of the Peepal tree genome is 34.23%. The gap-closing step was performed for the hybrid assembly. There were 35,811 (5.5%) misassembled contigs and 604,807 (94.4%) truly assembled contig sequences in the final assembled genome. The workflow of the genome assembly is presented in the supplementary material (Additional file 3: Figure S2). The statistics of assembly contigs and scaffolds are shown in the supplementary material (Additional file 4: Table S1.2) and the final scaffold assembly of the genome is given in Table 1. The alignment of raw reads to the hybrid genome sequence was performed, which mapped 99.5% and 99.27% of Illumina and MGI reads respectively (Additional file 5: Text file S1.1).

The completeness of the Peepal tree genome assembly was assessed with the BUSCO tool. The results showed that 76.5% (232 out of 303) and 84.1% (1,210 out of 1,440) of genes were conserved as single-copy orthologs

in eukaryotic and plant universal data sets, respectively. Out of 232 complete genes in the Eukaryota database, 214 are single-copy orthologs, 18 are duplicates, 57 are fragmented and 14 are missing. Out of 1,210 complete genes in the Embryophyte database, 1,173 are single-copy orthologs, 37 are duplicates, 105 are fragmented and 125 are missing (Additional file 6: Figures S3A and S3B). The transcriptome sequence reads aligned with the assembled genome showed that 99.46% of all reads were mapped and of these 88.25% of paired reads were mapped (Additional file 7: Text file S1.2).

#### Genome and pathways annotation

We identified 35,093 protein-coding genes with the complete structures in the Peepal tree genome (Additional file 8: Text file S2.1 and Additional file 9: Text file S2.2). RNA-seq data from two leaf tissue samples of the Peepal tree and alternative reference ESTs from *Morus notabilis* and *Arabidopsis thaliana* protein sequences were used as protein homology evidence during genome annotation. Based on a homology search using BLASTN, out of 35,093 genes predicted, 32,255 genes (91.9%) were having evidence from transcriptome assembly. About 76.3% of RNA-seq reads from the day and night leaf tissue samples were mapped to the annotated genes in the Peepal tree.

Based on the sequence similarities, the complete set of annotated genes and their amino acid sequences were used in the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis [17]. This result showed the pathways in metabolism, biosynthesis of secondary metabolites, genetic and environmental information processing, and signal transduction pathway were common as several others. The top 5 highest gene count for pathways like Ribosome (123 genes), Spliceosome (96 genes), Oxidative phosphorylation (86 genes), Thermogenesis (82 genes), and RNA transport (74 genes) was found. In addition, important candidate genes were also found for human disease pathways like Huntington's disease (68 genes), Parkinson's disease (57 genes), Alzheimer's disease (55 genes), and others (Additional file 10: Table S2).

#### Protein family and gene ontology analysis

The protein family (Pfam) ID and Gene Ontology (GO terms) were assigned to genes using an InterProScan module [18]. Out of 35,093 genes, 24,163 consisted of Pfam IDs that were distributed across 3,759 types of Pfam domains, and their gene ontology (GO) terms were also identified. The Pfam domain consisting of proteins that were large in the Peepal tree genome included 3-Deoxy-D-manno-octulosonic-acid transferase, Ring finger domain, PPR repeat family, Helix-loop-helix DNA-binding protein, DYW family of nucleic acid deaminases, Lysine methyltransferase, Putative GTPase activating protein for Arf, Ankyrin repeats and others (Additional file 11: Table S3).

Catalase is an antioxidant enzyme known to catalyze $H_2O_2$ into water and oxygen. We identified the gene sequences for the Catalase gene (FRLM_016351-RA) and its isozyme CAT1 Catalase isozyme 1 (FRLM_016350-RA), (FRLM_012250-RA) in the Peepal genome annotation. Two catalase genes were identified in the differential expression of transcriptome data: the KatE gene known as a monofunctional catalase, and the KatG gene known as a catalase-peroxidase [19]. KatE gene also known as CatB, is differentially expressed during the day and night period with the Fragments Per Kilobase of transcript per Million mapped reads (FPKM) values 937.49 and 1786.02 respectively. KatG gene is differentially expressed during the day and night, with the FPKM values 162.03 and 81.53 respectively. The KatE gene has been reported to be involved in physiological pathways such as glyoxylate and dicarboxylate metabolism, tryptophan metabolism, MAPK signaling pathway – plant, FoxO signaling pathway, and serine-pyruvate transaminase pathway. The KatG gene is involved in tryptophan metabolism, tyrosine metabolism, biosynthesis of secondary metabolism, and drug metabolism pathways.

#### Identification of homologous, orthologous, and singleton genes

To understand the gene evolution and relationships among *F. religiosa* and other taxa, we performed homologous and orthologous gene detection analysis for the Peepal tree with an additional 5 species. Homologous gene identification and orthologous clustering of the proteomes of six species, including the model organism *A. thaliana*, *M. notabilis* (closest relative species of Ficus), and other closely related species of the Moraceae family were selected for the analysis. Based on proteome sequence homology analysis, 29,516 homologous genes were found in *Arabidopsis thaliana*, 29,924 in *Morus notabilis*, 29,750 in *Cannabis sativa*, 29,909 in *Prunus persica*, and 29,830 in *Ziziphus jujuba* with respect to *F. religiosa* proteome sequences (35,093). *F. religiosa, A. thaliana*, *M. notabilis, C. sativa, P. persica*, and *Z. jujuba* form a cluster of 24,310 orthologous genes and are conserved within the species. The number of specific orthologous gene clusters identified was 15,016 in *F. religiosa*, 16,170 in *A. thaliana*, 16,517 in *M. notabilis*, 15,921 in *C. sativa*, 16,655 in *P. persica*, and 17,235 in *Z. jujuba*. A total of 1,184 single-copy gene clusters were found across the six species and the number of specific singletons identified was 10,154 in *F. religiosa*, 4,469 in *A. thaliana*, 2,284 in *M. notabilis*, 1,912 in *C. sativa*, 1,802 in *P. persica*, and 4,209 in *Z. jujuba* (Fig. 1).

The identified single-copy clusters were used to illustrate the taxonomic and phylogenetic relationships

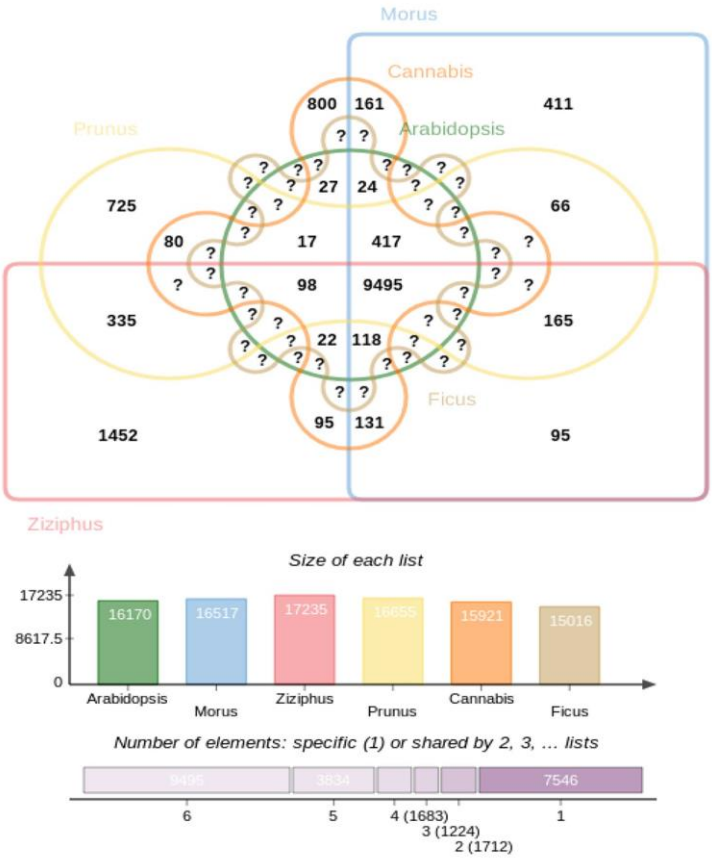



**Fig. 1** Orthologous clustering of 6 species using proteome data deduced 24,310 orthologous gene clusters and 1,184 single-copy gene clusters across the above 6 species

among a group of species. Based on the similarity of proteomes and single-copy orthologous clustering, we deduced the phylogenetic tree for *F. religiosa* and the other five species. The multiple sequence alignment (MSA) and Neighbour-Joining (NJ) methods were used for constructing an evolutionary phylogenetic tree. It was found that *F. religiosa* is closely related to *M. notabilis* by having more similarities in their proteomes as they are evolving from the Moraceae family, followed by *C. sativa, Z. jujuba, P. persica*, and *A. thaliana* (Additional file 12: Figure S4).

**Comparative analysis of Peepal tree genome**

We aligned the genome of the Peepal tree with those of the three Moraceae family members, *F. carica*, *F. microcarpa*, and *M. notabilis*. Comparison of our assembly against these genomes resulted in a mapping of 88.62% to *F. carica*, 89.6% to *F. microcarpa*, and 46.9% to *M. notabilis*. The results showed Peepal genome to be closer to the genus Ficus (*F. carica* and *F. microcarpa*) and also relatively closer to the genus Morus of the same family. The statistics of genome sequence alignment of *F. religiosa* against *F. carica*, *M. notabilis*, *F. microcarpa* genomes is presented in the supplementary material (Additional file 13: Text file S3).

### Repeats in the genome of the Peepal tree

Repeat library building and repeat identification were performed using the ReapeatModeller and RepeatMasker tools (www.repeatmasker.org) respectively. *De novo* repeat identification resulted in 53.55% (269.62 Mb) repetitive sequences in the Peepal tree genome. The RNA elements, long terminal repeats (LTR) constitute about 5% of repeats and 43.71% of these repeats did not belong to any of the annotated repeats families. The 53.55% of repetitive sequences in the Peepal tree genome are closest to its Moraceae family species, 47% are found in the closest species mulberry (*M. notabilis*), 46.5% in *F. microcarpa*, and 48.9% in *F. hispida*. The repetitive sequences were classified into known categories, such as LINE1 (0.19%), long terminal repeat retrotransposon (5.09%), DNA transposons (1.09%), and simple repeats (3.25%) and unclassified (43.71%) (Additional file 14: Table S4).

### Simple sequence repeats (SSRs)

We identified SSRs from the assembled Peepal tree genome. In total, 799,992 SSRs were identified on 267,593 sequences, which are composed of mono- (606,169), di- (143,113), tri- (34,327), tetra- (11,791), penta- (2,911), and hexa- (1,681) type repeats (Additional file 15: Table S5.1). Among mono repeats, the 'A/T' (73.91%) type was the highest followed by 'C/G' (1.87%). Similarly, the 'AT /TA', 'AG/CT', 'AC/GT', and 'CG/CG' types of di repeats were in 9.8%, 2.76%, 1.41%, and 0.09% fractions, respectively. 'AAT/ATT', 'AAG/CTT' 'ATA/TAT', 'TTA/TTC', and 'GAA/TAA', were the most abundant tri repeats and 'AAAT' was predominant in tetra repeats. The detailed distribution of all types of repeats and statistics is shown in the supplementary material (Additional file 16: Table S5.2 and Additional file 17: Table S5.3).

### Transcription factors (TFs)

Transcription factors act in regulating gene expression driven by several external and internal signals by activating or suppressing the downstream genes. The MAKER annotated protein sequences of Peepal tree genome assembly were used for BLAST analysis with the Plant Transcription Factor Database v5.0 [20] using the *A. thaliana* protein sequence as a reference. A total of 1,264 protein sequences from 35,093 protein-coding genes with genome annotation shows evidence for 56 families of Transcription factors (Additional file 18: Table S6). The TFs families include the ERF, M-type MADS, ARF, DBB, MIKC MADS, WOX, C3H, G2-like, MYB, TALE, B3, HB-other, and MYB-related family proteins. The transcription factors play an important role in regulating growth, developmental processes, and environmental responses in the plant's [21].

**Table 2** Statistics of Uni-genes in Peepal Transcript

| Features | Sample (Day – 2 PM) | Sample (Night – 2 AM) | Combined Sample |
|---|---|---|---|
| Number of genes | 22,597 | 23,360 | 26,479 |
| Number of transcripts | 16,912 | 16,780 | 18,173 |
| GC content (%) | 46.37 | 46.24 | 46.25 |
| Contig N50 (bp) | 1,404 | 1,407 | 1,374 |
| Median contig length (bp) | 813 | 801 | 753 |
| Average Contig (bp) | 1060.06 | 1055.95 | 1017.44 |
| Total assembled bases (bp) | 23,954,145 | 24,667,005 | 27,055,237 |

### Transcriptome sequencing, assembly, and annotation

*De novo* transcriptome assembly was performed for the mature leaf samples of the Peepal tree collected during the day and night periods. The assembly was performed for each sample and also a combined assembly was performed for the reads of both samples. The combined transcriptome assembly resulted in 152.8 Mb assembled bases with an N50 length of 2,076 bp and an average transcript length of 1316.83 bp and 42.17% GC content. The transcriptome assembly and annotation workflow are given in the supplementary material (Additional file 19: Figure S5). The statistics of assembly contigs and sequence assembled contigs are also provided in the supplementary material (Additional file 20: Table S7).

The *de novo* assembled transcript sequences (116,038) were processed for annotation. *De novo* assembled transcripts were clustered to exclude the redundant transcripts and identified 26,479 unique transcripts sequences. The statistics of Unigenes are given in Table 2. *De novo* assembled transcripts and unigenes were annotated to find the structural and functional genes. The protein families were identified for the uniquely characterized transcripts of RNA data. Out of 26,479 transcripts, Pfam IDs for 19,175 were distributed across 3,977 types of protein family (Pfam) domains and their gene ontology (GO) terms were also identified (Additional file 21: Table S8). Pfam IDs and GO terms were assigned to predict the function of unique gene sequences and encoded translated proteins.

The differentially expressed genes (35,182) from the day and night periods of leaf tissue samples of the Peepal tree (Additional file 22: Table S9.1) were used for the pathway analysis. The top 272 highly up-regulated differentially expressed transcripts were identified for diurnal and nocturnal periods (Additional file 23: Table S9.2).

The TFs were identified from the differentially expressed transcripts. From the day sample, 2 transcripts coded for specific TFs like C3H family protein and nuclear factor Y, subunit A7 (NF-YA7), and in the night period sample, the 6 transcripts coded for specific TFs like ERF family protein, CONSTANS-like 2, MYB-related family protein (Additional file 24: Table S9.3). In plants, the nuclear factor-YA has a role in drought stress

responses. In rice, NF-YA7 is involved in the drought tolerance pathway which is independent of the Abscisic acid manner [22]. Expression of C3H and NF during the day could influence plant growth and development in the Peepal tree. The Ethylene response factor ERF105 showed as the cold-regulated transcription factor gene of Arabidopsis [23]. In the Peepal tree, ERF, MYB -related family proteins like REVEILLE 1 (RVE1) [24] and late elongated hypocotyl gene (LHY) are expressed during the night period. RVE1 functions in the circadian clock and auxin pathways and LHY maintains the circadian rhythm in Arabidopsis [25]. Both the RVE1 and LHY are found expressed in night-specific Peepal tree transcripts indicating the active circadian rhythms and pathways during the dark time.

#### Non-coding RNA genes in the Peepal tree genome

Based on a coding potential calculator (CPC), *de novo*-based assembled transcripts (26,479) were further categorized into protein-coding (19,911) and non-coding (6,568). Based on BLASTN analysis, out of 6,568 non-coding transcripts, 4,219 transcripts targeted genome-annotated genes and 2,349 remained non-coding. A total of 30,973 Cufflinks assembled transcripts (reference-based alignment with genome assembly) were further categorized into protein-coding (7,163) and non-coding (23,810). Out of 23,810 non-coding transcripts, 14,605 were having alignment to genome-annotated genes using BLASTN. Further, categorization of specific day and night sample transcripts resulted in 6,628 (day) and 7,339 (night) protein-coding and 20,528 and 25,494 non-coding transcripts respectively. From these non-coding transcripts, 18,893 (day) and 22,232 (night) transcripts were aligned to MAKER-P predicted genes using BLASTN. The remaining transcripts were considered to be non-coding transcripts, as we did not find any match to predicted gene evidence to support them. Hence, the majority of RNA sequences are found to have protein-coding sequences, while the non-coding genes have been shown biologically relevant in recent years, and deepen studies are needed to understand their functions.

**miRNAs** microRNAs are a major class of non-coding RNAs. Based on the homology search, we identified the microRNA precursors using the miRbase database (http://www.mirbase.org). These microRNAs belong to MIR396, MIR2916, MIR156, MIR164, MIR6236, MIR166, MIR168, and MIR395 families. Among the identified miRNAs, MIR408 was found to be specific to the night period transcripts of the Peepal tree. MIR 408 was identified on the genes like TPK5 Two-pore potassium channel 5, prfA peptide chain release factor 1, and also on proteins of unknown function in the Peepal genome. MIR408 is a highly conserved microRNA in plants and is involved in enhancing photosynthesis by mitigating the efficiency of irradiation utilization and the capacity for carbon dioxide fixation [26].

The unigene transcripts were used to identify the microRNAs. MIR168 and MIR166 homologs were identified on the two transcripts. We identified the miRNAs on genomic scaffolds based on mapping the transcriptome data to the genome. This provides information on miRNAs specific to the day and night leaf tissue transcriptome. We identified 23 and 25 pre-miRNA expressions in the day and night period respectively (Additional file 25: Table S10.1). The statistics of transfer RNAs (tRNA) were identified in the genome and their details are given in the supplementary material (Additional file 26: Table S10.2).

#### Elucidation of carbon fixation pathway in Peepal tree

The study was conducted to analyze the gene expression patterns in the leaf tissues of the Peepal tree under the diurnal (2 PM) and the nocturnal period (2 AM). Through the pathway analysis, the candidate genes for carbon fixation pathways like the CAM pathway, Calvin-Benson cycle (C3) pathway, and C4 - Dicarboxylic pathway were identified and estimated based on their transcript abundance. The transcriptome data contained 20 putative genes involved in the carbon fixation module of CAM, C3, and C4 including the key genes fructose-bisphosphate aldolase class I, fructose-1,6-bisphosphate, phosphoenolpyruvate carboxylase (PEPC/PPC), phosphoenolpyruvate carboxylase kinase (PPCK), NAD+and NADP+, malate dehydrogenase (MDH) and pyruvate orthophosphate dikinase (PPDK) genes (Additional file 27: Table S11). Gene mapping was completed for CAM and C4 cycle pathways and could not find a mapping for three genes in the C3 cycle pathway. Those three genes, fructose-6-phosphate phosphoketolase (EC 4.1.2.22) and phosphoketolase (EC 4.1.2.9) were purified in *Acetobacter xylinum* [27] and sedoheptulokinase (EC 2.7.1.14) was shown in *Bacillus* species [28]. These three genes were not found in the Peepal tree for the C3 cycle.

The differentially expressed genes from transcriptomic data were mapped to the reference carbon fixation in photosynthetic organisms pathway on the KEGG database (Additional file 28: Figure S6) [29], [30], and [31]. The diagrammatic representation of the genes involved in the carbon fixation pathway is shown in the supplementary material (Additional file 29: Figure S7).

The important genes expressed in the C3 cycle are rubisco and glyceraldehyde-3-phosphate dehydrogenase. Ribulose-bisphosphate carboxylase (RuBP carboxylase or *rubisco*) small chain enzyme that is enriched in leaf tissue collected during the day (2 PM). Rubisco is the most abundant protein in chloroplasts. The glyceraldehyde-3-phosphate dehydrogenase (NADP+) is enriched in day sample leaf tissue, the enzyme responsible for the

reversible conversion of glyceraldehyde 3-phosphate to ribulose bisphosphate using ATP, the acceptor for CO2. The trancriptomic genes have been mapped to the C3 cycle except for the three genes mentioned above [32].

The signature genes responsible for the CAM cycle were expressed in the Peepal tree during the night. The phosphoenolpyruvate carboxylase kinase (PPCK), NAD(P)-ME (maeB), and Malate dehydrogenase (MDH) transcripts were highly enriched in the photosynthetic leaf tissue collected during the night period than the day. It indicates that the Peepal tree adapts to the CAM pathway and can fix nocturnal carbon dioxide using the PEP carboxylase (PEPC) enzyme and accumulate malate by the enzyme malate dehydrogenase. The transcriptomic genes of the Peepal tree have been completely mapped to the KEGG pathway of the CAM cycle.

In the C4 Dicarboxylic cycle, the high expression of glutamate-glyoxylate aminotransferase enzyme (GGAT) in the leaf tissue collected during the night period (2 AM) indicates the photorespiration in the Peepal tree. The carbon fixation begins in the mesophyll cells, where CO2 is converted into bicarbonate. It adds the 3-carbon acid phosphoenolpyruvate (PEP) by an enzyme called phosphoenolpyruvate carboxylase. The product of this reaction is the four-carbon acid oxaloacetate, which is reduced to malate another four-carbon acid [33]. The second highest expression is NADP-malate dehydrogenase (MDH), which converts the oxaloacetate generated by PEPC to malate. The differentially expressed genes from the Peepal tree had a complete mapping to the C4 cycle. The gene expression pattern of the carbon-fixation pathway in the Peepal tree suggests that the plant switches between the C3, C4, and CAM cycles during the diurnal and nocturnal periods. The FPKM and Trimmed Mean of M-values (TMM) values of the differentially expressed genes for the carbon fixation pathway are shown in Fig. 2A, B, C, and 3 A, B, and C.

## Discussion

This study generated and annotated the genomics and transcriptomics data for the keystone species Peepal tree (*F. religiosa*). We used two next-generation sequencing technologies to sequence the Peepal genome and characterized the hybrid whole genome. The assembled

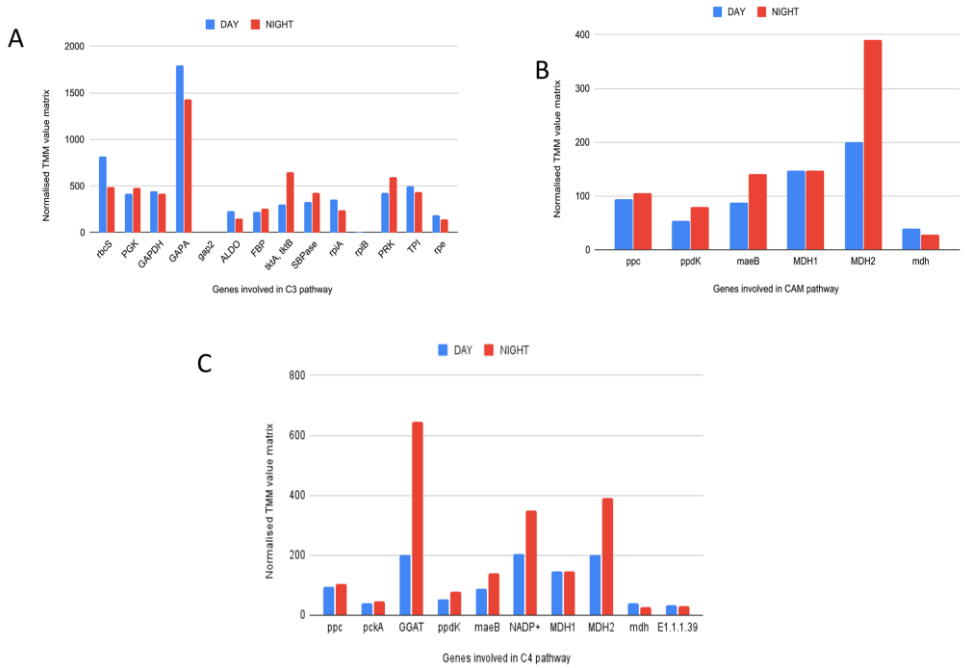



**Fig. 2** The candidate genes involved in the C3, CAM, and C4 cycles. (A) Calvin- Benson (C3) cycle B) Crassulacean acid metabolism (CAM) cycle C) C4 cycle. The X-axis represents the genes involved in pathways, Y-axis is the matrix of normalized expression trimmed mean of M (TMM) values; the Blue graph - leaf tissue collected during the day (2 PM), Red graph - leaf tissue collected during the night period (2 AM).

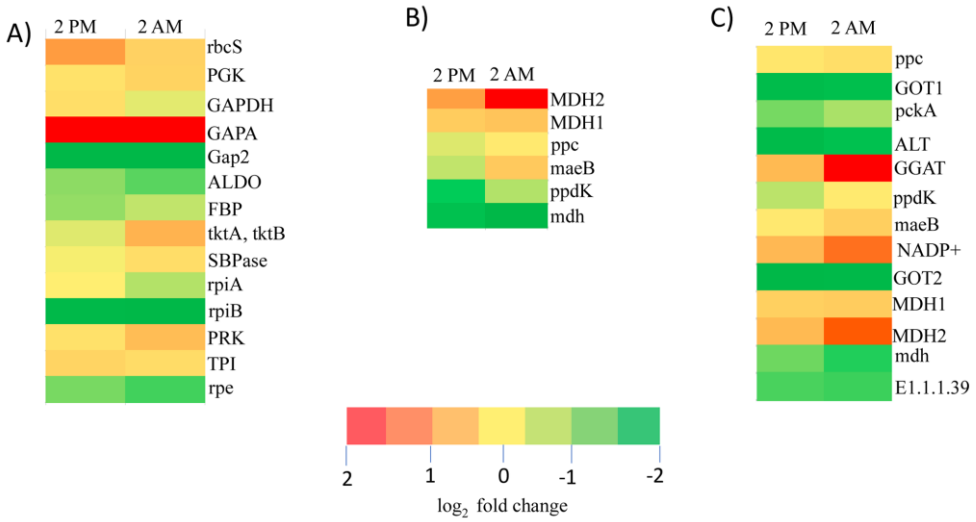



**Fig. 3** Gene expression pattern of *F. religiosa* carbon fixation genes across the diurnal (2 PM) and nocturnal (2 AM) expression data. (A) C3 cycle, (B) CAM cycle (C) C4 cycle log2-transformed Fragments Per Kilobase of transcript per Million mapped reads (FPKM) value-based expression profiles are shown

genome resulted in a size of 406 Mb with 35,093 protein-coding genes, based on ab initio, homology, and mRNA evidence used for annotation. Photosynthetic tissues at distinct conditions (diurnal and nocturnal) were used for RNA sequencing to understand the genes, proteins, and molecular pathways. The combined transcriptome analysis yielded 26,479 unique transcripts. The completeness of the Peepal tree genome was confirmed based on BUSCO analysis and comparative analysis of transcriptome data.

We performed the downstream analysis of genomic and transcriptomic data to understand the microRNAs, TFs, and molecular pathways of the Peepal tree. The miRNA MIR408 was identified to be specially expressed in the in Peepal leaf tissue during the night period. The miRNA MIR408 responds to copper deficiency and light in Arabidopsis [34]. In *O. sativa*, MiR408 plants were efficient at saving and converting light energy into sugars, suggesting that miR408 can promote photosynthesis by down-regulating the uclacyanin (UCL8) gene [35]. Thus, MIR408 found specific expression in transcripts of night period leaf tissue of the Peepal tree indicating the similar conversion of light energy and accumulation of sugars at night. It may also aid in photosynthesis by enhancing carbon fixation.

In the current study, Catalase gene expression was found to be high in the night period transcripts of Peepal tree leaf tissue. A previous study on the Peepal tree showed that leaf tissue collected at night time exhibited the scotoactive opening of stomata during the night, which indicates that through the stomatal opening molecular oxygen ($O_2$) is released by the action of catalase enzyme on hydrogen peroxide ($H_2O_2$) [36]. The physiological interaction between catalase and its substrate $H_2O_2$ in the plant was determined by quantifying $H_2O_2$ and assaying the catalase, in which catalase showed a 4-fold increase in activity, especially during the night. Peepal tree has a higher amount of $H_2O_2$ deposition during the night than day [36], which is an indication of pathway switching between carbon fixation pathways.

The RNA sequencing from diurnal (2 PM) and nocturnal (2 AM) leaf samples showed the gene expression patterns of the carbon fixation pathway. The day mRNA expression data suggested Peepal tree can carry out the diurnal carbon fixation by the C3 cycle. GGAT is involved in the photorespiratory process. High expression of GGAT in the C4 cycle indicates that there could be photorespiration in the Peepal tree during the night. Plants adapt to the CAM cycle to grow during water constraints and increase the level of carbon dioxide uptake than their C3 and C4 cycles [37]. The Peepal tree study provides information on plants using the CAM pathway to fix nocturnal carbon dioxide using the PEP carboxylase (PEPC) enzyme and the accumulation of malate by the enzyme malate dehydrogenase.

The Peepal tree gene expression analysis for the C3, C4, and CAM cycles suggested that plants could switch between these three cycles depending on the

carbohydrate, amino acids biosynthesis, metabolism, and environmental conditions. In the *Kalanchoë fedtschenkoi* genome study, the convergence in protein sequence and re-scheduling of diel transcript expression of genes was reported to be involved in nocturnal $CO_2$ fixation, stomatal movement, heat tolerance, circadian clock, and carbohydrate metabolism with the other CAM species in comparison with non-CAM species [14]. Some of the previous studies in the pineapple genome revealed the gene lineage transition from C3 photosynthesis to CAM, and CAM-related genes exhibit a diel expression pattern in photosynthetic tissues [13]. The evolution of CAM in *Agave* from C3 photosynthesis shows that the core metabolic components required for CAM have ancient genomic origins which could be traceable to non-vascular plants while regulatory proteins required for diel re-programming of metabolism have shared among the recent origin of C3, C4, and CAM species [38].

The plant model Arabidopsis encodes several orthologues of human proteins that function in mechanisms similar to those in other eukaryotes [39] [40]. Previous findings showed that 70% of oncogenes involved in cancer have orthologs in Arabidopsis, 67% in *D. melanogaster*, 72% in *C. elegans*, and 41% in *S. cerevisiae*. This 'disease gene' similarity is comparable to that observed in other model organisms [41]. The research in Arabidopsis and many other model systems has led to the discovery or analysis of genes and processes important to human health. Previous studies showed that the Peepal tree has been tested for the treatment of neurodegenerative disorders such as Parkinson's disease and Huntington's disease [6], [7], as well as anti-ulcer activity in animal models. [8]. Another study showed that methanol extract of *F. religiosa* has proven anti-inflammatory properties in LPS-induced activation of BV2 microglial cells, and it might have therapeutic potential for various neurodegenerative diseases [42]. In the present work, we found that Peepal genes show similarities with human disease pathways, which can be utilized to further understand the traditional medicinal practices and Buddha's meditation practices. Thus, plant research opens up new frontiers in terms of drug development and treatment of diseases of great importance to human health. Plants seem to be a part of this diverse portfolio of tools necessary to understand fundamental cellular processes.

In summary, the genome data and transcript abundance evidence indicate the molecular switch in the carbon fixation pathway of the Peepal tree (*F. religiosa*) during the day and night periods depending on its physiological and environmental conditions. Our study is a foundation for further experiments to determine the underlying mechanisms in C3, C4, and CAM metabolism.

## Conclusions

In this study, we generated the genomic and transcriptomic data for Peepal/Bodhi tree. Genomic data pathway analyses identified the genes associated with several physiological, biochemical, metabolic, and disease pathways. Differential expression data from diurnal and nocturnal leaf tissue samples of Peepal revealed gene expression patterns in the carbon fixation pathway during light and dark. The transcript abundance indicates that plants could switch between the three C3, C4, and CAM pathways. The well-annotated genome for the Peepal tree will have broader implications for studies regarding the physiology, evolution, conservation of species, and human neurological diseases.

## Methods

### Collection of leaf samples and extraction of nucleic acids

The mature leaves were collected from a cultivated Peepal tree (15 years old) at a Private property, Anuganalu village, Hassan District, India (13.0647° N, 76.0363° E). We have followed a non-invasive method for collecting leaf samples. Genomic DNA was extracted from the leaves using the Qiagen DNeasy Plant Mini kit (Catalog #69,106), and the quality and quantity of DNA were confirmed using the Nanodrop. From the same Peepal tree, the leaf samples were collected and immediately placed on dry ice during the day (2 PM) and night (2 AM) periods. Total RNA was isolated from the leaf samples using the QiagenRNeasy Plant Mini kit (Catalog #74,904) method and was treated with RNase-free DNase I (Catalog #M0303S) from New England BioLabs for 30 min at 37 °C to remove residual DNA. RNA integrity and quantity were confirmed on Qubit and Tape station using a dsDNA HS (Catalog #32,854) kit from Invitrogen and RNA screen tape from Agilent respectively.

### DNA and RNA library preparation and sequencing

Whole-genome shotgun DNA library preparation was performed using the Illumina TrueSeq DNA sample preparation kit (FC-121-2001). The paired-end (PE) (2×100 nts) sequencing was carried out using Illumina HiSeq-1000. Also, to increase the size of genome data, we sequenced the genome with paired-end (PE) (2×100 nts) using the MGISEQ-2000 platform.

The RNA libraries were prepared using "TruSeq RNA Library Prep Kit v2 from Illumina®" with Illumina standardized protocol. The RNA libraries were quantified on Qubit (dsDNA HS kit) and validated on the TapeStation instrument (D1000 screen tape). These RNA libraries were used for sequencing with the Illumina HiSeq-2500 platform.

#### Genome size estimation and assembly

Each of the Illumina and MGISEQ-2000 raw reads were processed for a quality check using the FastQC v0.11.6 tool [8]. Then filtering and trimming of raw reads were done to remove the low complexity bases using the TrimGalore-0.4.5 (https://www.bioinformatics.babraham.ac.uk/projects/trimgalore/) and reads having quality value Q>20 and length above 20 bases were taken for constructing the assembly. To estimate the genome size, filtered reads were taken for the k-mer distribution (different k-mers from 21 to 77) and abundance analysis using Jellyfish v1.1.12 [43] and GenomeScope v2.0 [44]. The separate Illumina and MGI Seq generated raw reads were used to construct the assembly using the tools SPAdes-3.13.0 [45] and MaSuRCA-3.2.9 [46] respectively. The parameters were the default k-mer sizes of 21, 33, and 55 for Illumina assembly. The constructed assemblies were used to build the super scaffolds using the tool SSPACE standard v3.0 [47].

The combined Illumina HiSeq and MGISEQ raw reads were used to construct the hybrid assembly using the assembler SPAdes-v3.13.0 [45]. The parameters were the default k-mer sizes 21, 33, and 55, with a 77 mer also set. The gaps in the assembly were closed by GMcloser-1.6.2 [48]. The assembly statistics were obtained using the tool Quast v4.6.1 [49]. The completeness and evaluation of the assembly were done by BUSCOv3 tool [50] with the Embryophyte and Eukaryota database and by aligning the RNA-seq reads to the genome.

#### Structural gene prediction and functional annotation

Peepal tree assembled scaffolds were processed for structural and functional gene annotation using the MAKER-P v.2.31.10 software [51]. The RNA-sequenced data of *Morus notabilis* [9] consists of expression sequence tags (ESTs) and the GFF (Gene finding format) file which contains the gene features and structures of genes, protein data of *A. thaliana* and RNA-sequence data of Peepal tree were imported as evidence for annotation support. The structural and functional annotation of predicted genes and proteins was performed using BLASTP in the Uniprot database. The protein family, structures, and gene ontology (GO) terms were identified for protein-coding genes using InterProScan-V5.27-66.0 [18].

#### Gene family construction, identification of homologous and orthologous genes

Protein sequences of *A. thaliana*, *M. notabilis*, *P. persica*, *C. sativa*, *Z. jujuba*, and the protein sequences of the current study *F. religiosa* were taken for the homologous and orthologous gene identification. The homologous genes were identified in *F. religiosa* proteome sequences using the BLASTP analysis against the other 5 proteomes of *A. thaliana*, *M. notabilis*, *P. persica*, *C. sativa*, and *Z. jujuba*. OrthoVenn2 [52] was used to cluster orthologous genes and identify the single-copy orthologous genes in all six proteomes. Further, these single-copy orthologous genes were used for constructing the phylogenetic tree using the tool MAFFT-v7 [53].

#### Comparative genome analysis

We downloaded the genomes of *F. carica*, *F. microcarpa*, and *M. notabilis*. We aligned these genomes against the Peepal tree genome assembly to understand their relationships using the BWA-V0.7.17 (Burrows-Wheeler Aligner) [54] and Samtools v1.7 [55].

#### Prediction of repetitive elements: TEs and SSR

The RepeatModeller-open-1.0.11and RepeatMasker-4.0 tools were used for repeat library building and repeat identification in the assembly respectively. The MicroSAtellite identification tool (MISA) [56] was used for the identification of SSRs from assembled genome sequences of *F. religiosa*. The parameters were set to identify perfect di-, tri-, tetra-, penta-, and hexa nucleotide motifs with a minimum threshold of 6, 5, 5, 5, and 5 repeats, respectively.

#### Prediction of transcription factor families

The families of transcription factors (TFs) were predicted in genome annotations and differentially expressed transcripts of the Peepal tree using Plant Transcription Factor Database v5.0 [20].

#### Non-coding RNA genes

The transfer RNAs in the Peepal tree genome were found using tRNAscan-SE (v2.0.3) [57] with the 'eukaryotes' option. tRNAscan-SE software deployed with the covariance models identifies the primary sequence and secondary structure information of tRNA and gives the complete tRNA genes for the query genome and transcriptome sequences. tRNAscan-SE software is integrated with Infernal v1.1 to enhance the tRNA search with better covariance and other updated models. Using the isotype-specific covariance model provides the functional classification of tRNAs and in the first pass search cutoff score, 10 is set. The miRbase database (http://www.mirbase.org) was used for the identification of putative miRNAs in the genome and unique identified transcripts sequence data based on the homology search. The long non-coding RNAs (lncRNAs) were identified with the Coding Potential Calculator tools [58].

#### Transcriptome sequencing, assembly, and annotation

High-quality stranded RNA sequencing (ssRNA-seq) reads were assembled into putative transcripts using Trinity v2.9.0 [59]. Assembled transcripts were passed through Transdecoder v5.02 [60] to predict the coding

sequences. The transcripts were clustered to find the unigenes by removing the redundant transcripts using the tool CD-HIT-est v.0.0.1 [61] with a 95% sequence identity threshold. Transcripts assembled from Trinity and CD-HIT-v0.0.1 were used in downstream analyses for gene prediction. Unigenes were used to predict the putative genes using the NCBI non-redundant (nr) database using the BLASTX program and proteins were predicted from the Uniprot database using the BLASTP program. The Trinity assembled transcripts were annotated using Trinotate- V3.11. The raw reads were mapped to scaffold assembled genome using Cufflinks-v2.2.1 [62] and considered as reference assembly.

### Transcript quantification and differential gene expression analysis

The estimation of transcripts abundance was determined using RNA-Seq by Expectation-Maximization (RSEM) tool [63], which quantifies transcript level abundance from RNA-seq data. RSEM first generates and preprocesses a set of reference transcript sequences and then aligns reads to reference transcripts followed by an estimation of transcript abundances. Normalized transcripts obtained from the transcript quantification methods were used in the next step for the differential gene expression analysis. FPKM and Trimmed Mean of M-values (TMM) are calculated to understand the expression levels of genes in day and night samples of the Peepal tree. For further analysis, the gene expression was estimated using FPKM and TMM value minimum $\geq 1$. The TMM value was used to cluster the genes according to their expression pattern using the edgeR package in the R tool. The parameters used in the differential expression analysis were a probability value P-value of 0.001 and a fold change value of log2. The expression value was also determined for assembled transcripts to verify the expression of genes predicted from gene models. The differentially expressed genes were annotated using BLAST2GO Annotation software [64].

### Pathway analysis

The annotated genes from the assembled genome and the differentially expressed genes from the Peepal tree leaf tissues collected during the day (2 PM) and night (2 AM) were used for pathway analysis in the KAAS (KEGG Automatic Annotation Server) (KEGG) server [17] using the BBH (bi-directional best hit) method and the search against a default set of 40 eukaryotic organisms. It provided the list of pathways where the candidate genes were mapped based on the orthologous homology alignment.

**Abbreviations**

| | |
|---|---|
| BUSCO | Benchmarking Universal Single-Copy Orthologous |
| ESTs | Expressed Sequence Tags |
| PPR repeat | Pentatricopeptide repeat |
| FPKM | Fragments Per Kilobase of transcript per Million mapped reads |
| MSA | Multiple Sequence Alignment |

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-023-09270-z.

**Additional file 1: Table S1.1.** Details on raw sequence data of F. religiosa genome and transcriptome

**Additional file 2: Figure S1A and S1B.** The kmer histogram distribution for the Illumina reads using the GenomeScope tool and The kmer histogram distribution for the MGI reads using the GenomeScope tool

**Additional file 3: Figures S2A.** Flow chart of De Novo Whole Genome Analysis (WGA) of Ficus religiosa

**Additional file 4: Table S1.2.** Contig and scaffold assembly statistics of F. religiosa genome

**Additional file 5: Text File S1.1.** The statistics of IlluminaMGI reads mapped to Peepal hybrid whole genome

**Additional file 6: Figures S3A and S3B.** BUSCO Assessment results using the plant universal single-copy orthologs (Embryophyta database) and BUSCO Assessment results using the eukaryote universal single-copy orthologs (Eukaryota database).

**Additional file 7: Text File S1.2.** The statistics of transcriptome reads mapped to Peepal hybrid whole genome

**Additional file 8: Text File S2.1.** Annotated gene sequence of F. religiosa

**Additional file 9: Text File S2.2.** Annotated protein sequences of F. religiosa

**Additional file 10: Table S2.** Top 10 pathways with the highest gene counts in the F. religiosa genome

**Additional file 11: Table S3.** Protein family IDs identified for the annotated genes of F. religiosa

**Additional file 12: Figure S4.** Phylogenetic analysis of Ficus religiosa with other plant species like A. thaliana, C. sativa, M. notabilis, Z. jujuba, P. persica

**Additional file 13: Text file S3.** The statistics of genome sequence alignment of F. religiosa against F. carica, M. notabilis, F. macrocarpa genomes

**Additional file 14: Table S4.** Repeat content in the assembled F. religiosa genome

**Additional file 15: Table S5.1.** Simple sequence repeats (SSR) prediction in the genome of F. religiosa

**Additional file 16: Table S5.2.** Distribution to different SSR repeat-type classes

**Additional file 17: Table S5.3.** Statistics of SSR repeat type classes

**Additional file 18: Table S6.** Families of Transcription factors (TFs) identified in the annotated genes

**Additional file 19: Figure S5.** Flow chart of De novo Transcriptome Analysis of Ficus religiosa.

**Additional file 20: Table S7.** Assembly Statistics of F. religiosa Transcriptome

**Additional file 21: Table S8.** Protein family IDs identified for the unique transcripts of F. religiosa

**Additional file 22: Table S9.1.** The total differentially expressed genes from F. religiosa leaf tissue sample of day and night periods

**Additional file 23: Table S9.2.** The top highly up-regulated differentially expressed genes from the day and night periods with their FPKM values

**Additional file 24: Table S9.3.** Families of Transcription factors (TFs)

identified in transcripts from day and night leaf samples

Additional file 25: Table S10.1. Details on long non-coding RNAs, miRNAs in De novo transcripts, reference-based transcripts, and genome

Additional file 26: Table S10.2. Statistics of Transfer RNAs predicted in the genome

Additional file 27: Table S11. Peepal tree genes mapped to the KEGG database pathway

Additional file 28: Figure S6. Graphical representation of differentially expressed genes of Ficus religiosa mapped to reference Carbon fixation in photosynthetic organisms pathway (map00710) from KEGG.

Additional file 29: Figure S7. Diagrammatic representation of candidate genes of Carbon fixation pathway

Acknowledgements
We would like to thank Tata Education and Development Trust fellowship (01/2018–2019) provided for Ashalatha KL for her Ph.D. program. We acknowledge the Next Generation Genomics Facility at the Centre for Cellular and Molecular Platforms (C-CAMP) and Bangaluru Genomics Center Pvt. Ltd for support in sequencing, We thank Mr. Ravindra Raut, TDU lab member for his help in collecting the leaf samples.

Author contribution
AKL performed the DNA and RNA isolation from leaf tissues, genome assembly and functional annotation, gene prediction, repeat prediction, orthologous gene clustering, Genome and Transcriptome pathway analysis, submitted WGS and RNA-seq data to NCBI, prepared the genomic and transcriptomic study tables and figures; other bioinformatic analysis and wrote the manuscript; MG, AKL, and AKP designed the genomic and transcriptomic experiments of *Ficus religiosa*; MG conceived and conceptualized the project; reviewed and edited the manuscript. AKP reviewed and edited the manuscript. All authors have read and reviewed the final draft of the manuscript.

Funding
Not Applicable (NA).

Data Availability
The raw sequence reads have been deposited under NCBI Sequence Read Archive (SRA) accession numbers SRR7244210 for Illumina sequenced *F. religiosa* WGS, SRR13827064 for MGISEQ sequenced *F. religiosa* WGS–(https://trace.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA474013), SRR7343291 (https://trace.ncbi.nlm.nih.gov/Traces/study/?acc=SRR7343291) for *F. religiosa* transcriptome). The Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JAFMPE00000000 (https://www.ncbi.nlm.nih.gov/nuccore/JAFMPE00000000.1/) and Transcriptome Shotgun Assembly project has been deposited at DDBJ/ENA/GenBank under the accession GJAV00000000 (https://www.ncbi.nlm.nih.gov/nuccore/GJAV00000000.1/). Other published data used in the study: The protein sequences of *Arabidopsis thaliana* (https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/001/735/GCF_000017354_TAIR10.1/GCF_000017354_TAIR10.1_protein.faa.gz), *Morus notabilis* (https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/414/095/GCF_000414095.1_ASM41409v2/GCF_000414095.1_ASM41409v2_protein.faa.gz), *Prunus persica* (https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/346/465/GCF_000346465.2_Prunus_persica_NCBIv2/GCF_000346465.2_Prunus_persica_NCBIv2_protein.faa.gz), *Cannabis sativa* (https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/900/626/175/GCF_900626175.2_cs10/GCF_900626175.2_cs10_protein.faa.gz), *Ziziphus jujuba* (https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/020/796/205/GCF_020796205.1_ASM2079620v1/GCF_020796205.1_ASM2079620v1_protein.faa.gz).

## Declarations

Statement
Experimental Plant details: *Ficus religiosa* a cultivated plant was used in this study. The mature leaves were collected from this plant at Private property, Anuganalu village, Hassan District, India (13.0647° N, 76.0363° E) were kindly provided by Prof. Malali Gowda, DNA Life Foundation, Anuganalu village, Hassan District, India (13.0647° N, 76.0363° E). The formal identification of the plant was undertaken by Prof. Malali Gowda, who owns the place of plant location and is also a corresponding author of this study. Hence, we had his permission to collect the plant material. The specimen of this material has not been deposited in any publicly available herbarium.

Ethics approval and consent to participate
The study is compiled with relevant institutional, national, and international guidelines and legislation.

Consent for publication
Not Applicable (NA).

Competing interests
The authors declare that they have no competing interests.

Received: 11 August 2022 / Accepted: 23 March 2023
Published online: 12 April 2023

References

1. Chri D, Khoshoo TN. Nuclear DNA contents in the genus Ficus (Moraceae). Plant Syst Evol. 1987;156(1–2):1–4.
2. Van Noort S, Gardiner AJ, Tolley KA. New records of Ficus (Moraceae) species emphasize the conservation significance of inselbergs in Mozambique. South Afr J Bot. 2007;73(4):642–9.
3. Ashalatha KL, Gowda M. Heritage of Neem–Peepal Tree Resides a Profound Scientific Facts: The Neem Genome. Springer, Cham; 2019 pp. 13–9.
4. Chantarasuwan B, Baas P, Van Heuven B-J, Baider C, Van Welzen PC. Leaf anatomy of Ficus subsection Urostigma (Moraceae). Bot J Linn Soc. 2014;175(2):259–81.
5. Singh D, Singh B, Goel RK. Traditional uses, phytochemistry and pharmacology of Ficus religiosa: a review. J Ethnopharmacol. 2011;134(3):565–83.
6. Bhangale JO, Acharya SR. Anti-Parkinson Activity of Petroleum Ether Extract of Ficus religiosa (L.) Leaves. Adv Pharmacol Sci. 2016;2016:1–9.
7. Bhangale JO, Acharya NS, Acharya SR. Protective effect of Ficus religiosa (L.) against 3-nitropropionic acid induced Huntingt on disease. Orient Pharm Exp Med. 2016;16(3):165–74.
8. Gregory M, Divya B, Mary RA, Viji MMH, Kalaichelvan VK, Palanivel V. Anti–ulcer activity of Ficus religiosa leaf ethanolic extract. Asian Pac J Trop Biomed. 2013;3(7):554–6.
9. He N, Zhang C, Qi X, Zhao S, Tao Y, Yang G, et al. Draft genome sequence of the mulberry tree Morus notabilis. Nat Commun. 2013;4(1):2445.
10. Usai G, Mascagni F, Giordani T, Vangelisti A, Bosi E, Zuccolo A, et al. Epigenetic patterns within the haplot ype phased fig (Ficus carica L.) genome. Plant J. 2020;102(3):600–14.
11. Zhang X, Wang G, Zhang S, Chen S, Wang Y, Wen P, et al. Genomes of the Banyan Tree and Pollinator Wasp provide insight s into fig-wasp coevolution. Cell. 2020;183(4):875–889e17.
12. Chakraborty A, Mahajan S, Bisht MS, Sharma VK. Genome sequencing and comparative analysis of Ficus benghalensis and Ficus religiosa species reveal evolutionary mechanisms of longe vity. iScience. 2022;25(10):105100.
13. Ming R, VanBuren R, Wai CM, Tang H, Schatz MC, Bowers JE, et al. The pineapple genome and the evolution of CAM photosynthesis. Nat Genet. 2015;47(12):1435–42.
14. Yang X, Hu R, Yin H, Jenkins J, Shu S, Tang H, et al. The Kalanchoë genome provides insights int o convergent evolution and building blocks of crassu - lacean acid metabolism. Nat Commun. 2017;8(1):1899.
15. Guan Q, Tan B, Kelley TM, Tian J, Chen S. Physiological changes in Mesembryanthemum crystallinum during the C3 to CAM t ransition Induced by Salt stress. Front Plant Sci. 2020;11:283.
16. Schlüter U, Denton AK, Bräutigam A. Understanding metabolit e transport and metabolism in C4 plants through RNA-seq. Curr Opin Plant Biol. 2016;31:83–90.
17. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 2007;35 Web Server:W182–5.

18. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30(9):1236–40.
19. Italiani VCS, da Silva Neto JF, Braz VS, Marques MV. Regulation of catalase-peroxidase KatG is OxyR dependent and Fur Independent in Caulobacter crescentus. J Bacteriol. 2011;193(7):1734–44.
20. Jin J, Tian F, Yang D-C, Meng Y-Q, Kong L, Luo J, et al. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res. 2017;45(D1):D1040–5.
21. Liu C, Xu X, Kan J, Cheng Z ming, Chang Y, Lin J et al. Genome-wide analysis of the C3H zinc finger family reveals its functions in salt stress responses of Pyrus betulaefolia. PeerJ. 2020;8:e9328.
22. Lee D-K, Kim H, Il, Jang G, Chung PJ, Jeong JS, Kim YS, et al. The NF-YA transcription factor OsNF-YA7 confers drought stress tolerance of rice in an abscisic acid independent manner. Plant Sci. 2015;241:199–210.
23. Bolt S, Zuther E, Zintl S, Hincha DK, Schmülling T. ERF105 is a transcription factor gene of Arabidopsis thaliana required for freezing tolerance and cold acclimation. Plant Cell Environ. 2017;40(1):108–20.
24. Rawat R, Schwartz J, Jones MA, Sairanen I, Cheng Y, Andersson CR, et al. REVEILLE1, a myb-like transcription factor, integrates the circadian clock and auxin pathways. Proc Natl Acad Sci. 2009;106(39):16883–8.
25. Mizoguchi T, Wheatley K, Hanzawa Y, Wright L, Mizoguchi M, Song H-R, et al. LHY and CCA1 are partially redundant genes required to maintain circadian rhythms in Arabidopsis. Dev Cell. 2002;2(5):629–41.
26. Pan J, Huang D, Guo Z, Kuang Z, Zhang H, Xie X, et al. Overexpression of microRNA408 enhances photosynthesis, growth, and seed yield in diverse plants. J Integr Plant Biol. 2018;60(4):323–40.
27. Schramm M, Klybas V, Racker E. Phosphorolytic cleavage of fructose-6-phosphate by fructose-6-phosphate phosphoketolase from Acetobacter xylinum. J Biol Chem. 1958;233(6):1283–8.
28. Ebata M, Sato R, Bak T. The enzymatic phosphorylation of sedoheptulose. J Biochem. 1955;42(6):715–25.
29. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30.
30. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. Protein Sci. 2019;28(11):1947–51.
31. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. 2023;51(D1):D587–92.
32. McFarlane CR, Shah NR, Kabasakal BV, Echeverria B, Cotton CAR, Bubeck D, et al. Structural basis of light-induced redox regulation in the Calvin–Benson cycle in cyanobacteria. Proc Natl Acad Sci. 2019;116(42):20984–90.
33. Cousins AB, Baroli I, Badger MR, Ivakov A, Lea PJ, Leegood RC, et al. The role of Phospho enol pyruvate carboxylase during C4 photosynthetic Isotope Exchange and Stomatal Conductance. Plant Physiol. 2007;145(3):1006–17.
34. Abdel-Ghany SE, Pilon M. MicroRNA-mediated systemic down-regulation of copper protein expression in response to low copper availability in Arabidopsis. J Biol Chem. 2008;283(23):15932–45.
35. Zhang J-P, Yu Y, Feng Y-Z, Zhou Y-F, Zhang F, Yang Y-W, et al. MiR408 regulates Grain Yield and Photosynthesis via a phytocyanin protein. Plant Physiol. 2017;175(3):1175–85.
36. Smitha RB, Bennans T, Mohankumar C, Benjamin S. Oxidative stress enzymes in Ficus religiosa L.: biochemical, histochemical and anatomical evidences. J Photochem Photobiol B Biol. 2009;95(1):17–25.
37. Nobel PS. Achievable productivities of certain CAM plants: basis for high values compared with C 3 and C 4 plants. New Phytol. 1991;119(2):183–205.
38. Yin H, Guo H-B, Weston DJ, Borland AM, Ranjan P, Abraham PE, et al. Diel rewiring and positive selection of ancient plant proteins enabled evolution of CAM photosynthesis in Agave. BMC Genomics. 2018;19(1):588.
39. Hays JB. Arabidopsis thaliana, a versatile model system for study of eukaryotic genome-maintenance functions. DNA Repair. 2002;1(8):579–600.
40. Spampinato CP, Gomez-Casati DF. Research on plants for the understanding of Diseases of Nuclear and mitochondrial origin. J Biomed Biotechnol. 2012;2012:1–12.
41. Jones AM, Chory J, Dangl JL, Estelle M, Jacobsen SE, Meyerowitz EM, et al. The impact of Arabidopsis on Human Health: diversifying our portfolio. Cell. 2008;133(6):939–43.
42. Jung HW, Son HY, Minh C, Van, Kim YH, Park Y-K. Methanol extract of Ficus leaf inhibits the production of nitric oxide and proinflammatory cytokines in LPS-stimulated microglia via the MAPK pathway. Phyther Res. 2008;22(8):1064–9.
43. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011;27(6):764–70.
44. Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, et al. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics. 2017;33(14):2202–4.
45. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a New Genome Assembly Algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455–77.
46. Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. The MaSuRCA genome assembler. Bioinformatics. 2013;29(21):2669–77.
47. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics. 2011;27(4):578–9.
48. Kosugi S, Hirakawa H, Tabata S. GMcloser: closing gaps in assemblies accurately with a likelihood-based selection of contig or long-read alignments. Bioinformatics. 2015;31(23):3733–41.
49. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29(8):1072–5.
50. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210–2.
51. Campbell MS, Law M, Holt C, Stein JC, Moghe GD, Hufnagel DE, et al. MAKER-P: a Tool Kit for the Rapid Creation, Management, and Quality Control of Plant Genome annotations. Plant Physiol. 2014;164(2):513–24.
52. Xu L, Dong Z, Fang L, Luo Y, Wei Z, Guo H, et al. OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res. 2019;47(W1):W52–8.
53. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. 2019;20(4):1160–6.
54. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.
55. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10(2):giab008.
56. Thiel T, Michalek W, Varshney R, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L). Theor Appl Genet. 2003;106(3):411–22.
57. Lowe TM, Eddy SR. tRNAscan-SE: a program for Improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25(5):955–64.
58. Kong L, Zhang Y, Ye Z-Q, Liu X-Q, Zhao S-Q, Wei L, et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. 2007;35(suppl2):W345–9.
59. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29(7):644–52.
60. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013;8(8):1494–512.
61. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22(13):1658–9.
62. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28(5):511–5.
63. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12(1):323.
64. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21(18):3674–6.


## Publisher's Note

**APPENDIX II**

**BOOK CHAPTER:**

## Heritage of Neem - Peepal tree resides a profound scientific facts

# Heritage of Neem–Peepal Tree Resides a Profound Scientific Facts

2

K. L. Ashalatha and Malali Gowda

**Abstract**

India's mythology and traditional medicines have great significance to Neem (*Azadirachta indica*) and Peepal (*Ficus religiosa*). Some of the historic values, rituals, and festival practices taken for these trees are mentioned here. After the meditation of great Indian sage Buddha, the Peepal tree is named as Bhodi tree. Neem–Peepal trees are widely used for treating many diseases in traditional medicines like Ayurveda, Homeopathy, Unani, and Siddha. In Ayurveda, few of the shlokas (quotes) describing the medicinal properties of Neem–Peepal trees are given in this chapter.

## 2.1 Introduction

The divine trees of India, Neem (*Azadirachta indica*) belongs to Meliaceae family and Peepal or Bodhi tree (*Ficus religiosa*) belongs to Moraceae family. There are several vernacular names of Neem: Nimba, Bevina mara, etc. and Peepal are Ashwatta, Arali, Bodhi, etc. These tree species are considered as keystone and holy in Indian subcontinent. These are semi-evergreen, terrestrial, and adopt to xerophyte nature. Since the ancient days to contemporary world, these trees are believed to be sacred and worshiped by various communities like Hindus, Buddhists, Jains, etc.

## 2.2 Ashwattakatte or Aralikatte Platform

We could encounter the Neem and Peepal trees that are grown together near Indian temples and villages. The **Ashwattakatte** or **Aralikatte** platform is often constructed using rocks around these trees (Fig. 2.1). In the early dawn of morning, devotees do pradakshina (circumambulation) and prayer for these trees (Fig. 2.2).

In southern parts of India, people perform marriage between Neem and Peepal trees which are planted and grown together. Hence they coin Neem as female and Peepal as male. The above image depicts the marriage of Neem and Peepal tree (Fig. 2.1). The white-colored cloth (Panche—in Kannada language) been tied to the left side Peepal tree and considered it as male (groom) and the Indian traditional saree is been tied to the middle Neem tree, as female (bride) (Fig. 2.1) and also we can witness the Nagas idols (Snake sculpture) are placed under these trees. Later, the traditional marriage rituals were made for these trees.

K. L. Ashalatha
Centre for Functional Genomics and Bioinformatics, The University of TransDisciplinary Health Sciences and Technology, Bengaluru, India

M. Gowda (✉)
The University of TransDisciplinary Health Sciences and Technology, Bengaluru, India
e-mail: malalig@tdu.edu.in


M. Gowda et al. (eds.), *The Neem Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-030-16122-4_2

**Fig. 2.1** Ashwatta or Aralikatte platform near India's temple. *Credit* Ashalatha K. L.

## 2.3 Neem is Symbolic to Mariamman Temple

Neem is associated with Indian culture. For example, Mariamman festival celebrated across the Tamil Nadu state during the Adi or April month. Mari is called as Goddess Parvathi, Durge, Kali, Shitaladevi, etc. (Fig. 2.3).

Thousands of people gather during Mariamman festival and carry neem leaves from one house to another for worshiping goddess Mari. People walk for miles, carrying water mixed with neem leaves and turmeric powder to ward-off illnesses or diseases like measles, cholera, smallpox, and chickenpox. The outbreak of these diseases is so high during summer period. There is a logical and scientific thorough process discovered by ancient people in India and converted these ideas into traditional and cultural practices or festival.

## 2.4 Bevu-Bella During Ugadi Festival

In Karnataka and Andhra Pradesh, people celebrate the Ugadi festival during March–April (Chaitra Masa). This is considered as the New year for local people. On Ugadi day, people consume Neem and Jaggery (Bevu-Bella) together which signifies that life is formed with happiness and sorrows or like the taste of bitter and sweet. They also add Neem leaves into the hot water and take bath on that day which is good for skin health. This fest is celebrated among the Hindu families (Fig. 2.4).

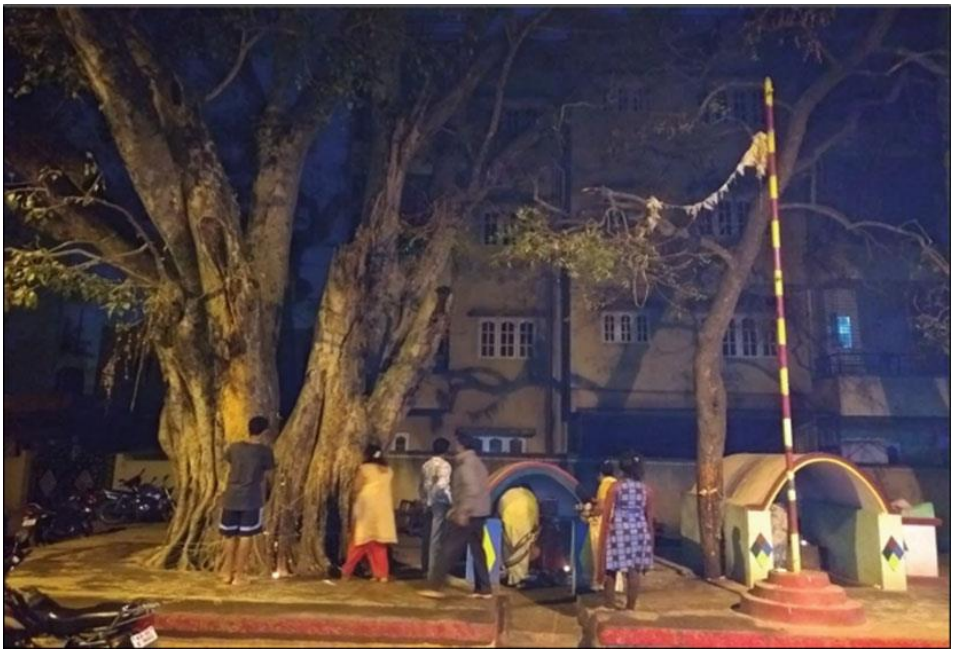


**Fig. 2.2** Devotees circumambulation to the Aralikatte. *Credit* Dr. Malali Gowda

**Fig. 2.3** Use of Neem leaves during Mariamman festival. *Credit* Kamala

**Fig. 2.4** Neem and jaggery (Bevu-Bella) consumed on Ugadi festival. *Credit* Shreyas

Thus, this concept needs to be translated among the local communities to create an awareness of scientific importance of this festival. Neem has several properties like antimicrobial, antiviral, antifungal, and antidiabetic properties and also widely used as a pesticide in agriculture (Brahmachari et al. 2004). The different tissues of Neem like twig, leaves, roots, bark, flowers, fruits, seeds, and bark have been used for preparation of traditional medicines and bio-pesticides (Brahmachari et al. 2004). Further, we have sequenced the Neem genome to understand the metabolic pathways (Kuravadi et al. 2015).

Peepal tree is believed to harbor God Trimurtis (Lord Shiva, Lord Vishnu, and Lord Brahma) in Hinduism (mentioned in ancient traditional texts). The Buddhist believes that Buddha attained enlightenment underneath the Peepal tree. After his meditation, this tree became largely popularized as "**Bodhi tree**". The Mahabodhi Temple is located in Bodh Gaya in Bihar, India is known as Buddhist pilgrimage in the world. Bodhi tree at Bodh Gaya temple was the offshoot of the original propagule planted in 288 BCE (CABI 2018) (Fig. 2.5).

Other than religious beliefs and usage, these trees are privileged to have scientific facts. Most of the traditional medicines such as Ayurveda, Siddha, and Homeopathy describe the usage of various tissues (leaf, bark, roots, fruits, twigs, etc.) of Neem and Peepal for making medicines and formulations.

Peepal tree has various medicinal properties including antioxidants, antimicrobial, anti-acetylcholinesterase, wound healing, antidiabetic, anti-inflammatory, etc. (Gautam et al. 2014). It is used for many medicinal purposes such as diabetes, ulcers, gastrointestinal problems, neurological disorders, skin diseases, urinary disorders, respiratory problems, etc. Some of the Ayurvedic formulations, which use Peepal tissues, are like Nalpamardi thailam, Nyagrodhadi churna, Sarivadyasava, Panchavalkadi tailam. Recently, we have sequenced genome of Peepal tree to decode the biomedical properties (*unpublished* data) (Fig. 2.6).

Behind every traditional practice, there is a hidden scientific knowledge, yet to explore, observe, understand, and narrate it carefully. Concept of Aralikatte structure is an interesting biological knowledge where roots are trained towards the ground. If this structure is not constructed, then roots of Peepal tree pierce aerially and damage houses. Thus, Aralikatte structure is covered with rocky material, which guide the Peepal roots towards the underground soil. It is the best example where traditional knowledge that has scientific thought process.

Some of the useful properties of Neem and Peepal tree are been described in the Ayurveda texts. And few of those shlokas (quotes) were taken from Ayurveda listed below.

**Neem Tree**:

**Neem Tree:**

**||भावप्रकाशनघिण्टु||**

नम्बः शीतो लघुर्ग्राही कटुपाकोऽग्नवितनुत् |
अहृद्यः श्रमतड्कासज्वरारुचकृिमप्रिणुत् ||८२||
व्रणपत्तिकफच्छर्दकुिष्ठहृल्लासमेहनुत् |

नम्बः शीतो - Neem tree is cold in condition
लघुर्ग्राही - Quick and easy to digest and absorbs the moisture from intestine. It is used treat the ulcer and heals the wounds.
कटुपाकोऽग्नवितनुत् - Having a pungent taste, bitter taste, helps in digestion, increase Vata prakriti (ayurveda classified type)
अहृद्यः - It is not good for heart
श्रमतड्कासज्वरारुचकृिमप्रिणुत् - It helps in relieving tiredness and thirst. It decreases the Pitta (ayurveda classified type) hence associated with fever and thirst. It helps in respiratory problems, relieves cough and also has antimicrobial property. Useful for relieving anorexia disorder, removes worms in intestine and heals wounds quickly. It balances the Pitta and Kapha prakriti. Helps to relieve nausea and vomiting. Useful for skin diseases, diabetes problem and urinary problems.

*Source* e-Nighantu (Collection of Āyurvedic Lexicons)

**Peepal Tree**:

**Peepal tree:**
|| **कैयदेवनघिण्टु** ||
अश्वत्थः शीतलो रूक्षः कषायो दुर्जरो गुरुः ||४३२||
व्रणपत्तिकफास्रघ्नो वर्ण्यो योनविशोधनः |४३३|

अश्वत्थः शीतलो रूक्षः - Ashvattha tree has coolant property
कषायो दुर्जरो गुरुः - Having an astringent taste and difficult to digest
व्रणपत्तिकफास्रघ्नो - It is used to treat the ulcers and wounds. And also balances the pitta, kapha doshas [types of prakriti (nature) classified in ayurveda]
वर्ण्यो - It gives the good complexion to skin and improves the skin color.
योनविशोधनः - It also cures the female reproductive diseases

**||राजनघिण्टु||**
पप्पिलः सुमधुरस्तु कषायः शीतलश्च कफपत्तिवनिाशी |
रक्तदाहश्रमनः स हि सद्यो योनदिोषहरणः कलि पक्वः ||११४||
अश्वत्थवृक्षस्य फलानि पक्वान्यतीवहृद्यानि च शीतलानि |
कुरुवन्ति पत्तिास्रवषिारतदिाहवच्छिर्दशिोषारुचदिोषनाशम् ||११५||
सुमधुरस्तु - sweet in taste
रक्तदाहश्रमनः - Useful for bleeding disorder
सद्यो योनदिोषहरणः - Quick relieve from vaginal and urinary tract problems
कलि पक्वः - Bark is used
फलानि पक्वान्यतीवहृद्यानि - Fruits are tasty and cold
पत्तिास्रवषिारतदिाहवच्छिर्दशिोषारुचदिोषनाशम् - Removes toxic and poisonous condition. And it relieves burning, nausea, weakness anorexia and cures skin diseases

*Source* e-Nighantu (Collection of Āyurvedic Lexicons)

**Fig. 2.5** Buddha meditating under the Peepal tree and circumambulation by devotees. *Credit* Ashalatha K. L.

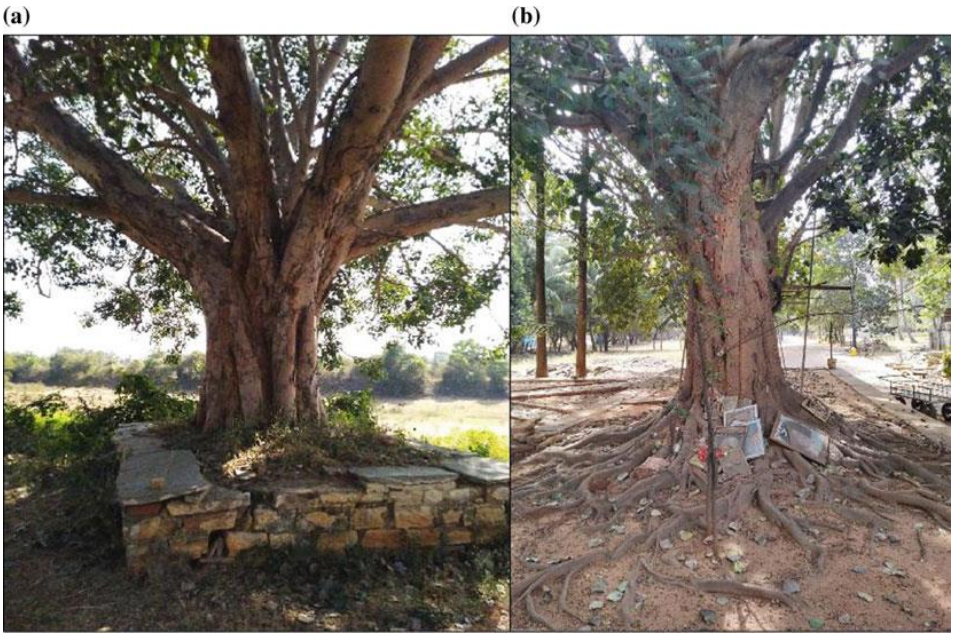



**Fig. 2.6** **a** Aralikatte built around the Peepal using rock and **b** showing roots of Peepal tree. *Credit* Sachin and Rakshith

**Acknowledgements** I would like to thank Dr. Malali Gowda for giving an opportunity to contribute in this book and his motivation for writing this article. I would like to thank Mr. Shreyas, Ms. Kamala, Mr. Sachin, and Mr. Rakshith for contributing the photographs, which are used in this chapter.

## References

Brahmachari G et al (2004) Neem—an omnipotent plant: a retrospection. ChemBioChem 5:408–421

CABI (2018) *Ficus religiosa* (sacred fig tree). Retrieved 23 July 2018. Invasive Species Compendium, CABI

e-Nighantu (Collection of Āyurvedic Lexicons), Central Council for Research in Ayurvedic Sciences (CCRAS), New Delhi. http://niimh.nic.in/ebooks/e-Nighantu/

Gautam S, Meshram A, Bhagyawant SS, Srivastava N (2014) *Ficus religiosa*—potential role in pharmaceuticals. Int J Pharm Sci Res 5(5):1616

Kuravadi NA, Yenagi V, Rangiah K, Mahesh HB, Rajamani A, Shirke MD, Russiachand H, Loganathan RM, Shankara Lingu C, Siddappa S, Ramamurthy A, Sathyanarayana BN, Gowda M (2015) Comprehensive analyses of genomes, transcriptomes and metabolites of neem tree. PeerJ 3: e1066