SHORT TITLE

Genome assembly and annotation of Peepal tree

ARTICLE TITLE

Genome Sequencing Unravelled C3, C4 and CAM Photosynthetic Pathways in *Ficus religiosa*

KL Ashalatha [1] , KP Arunkumar[2], Malali Gowda[1, 3]

CORRESPONDING AUTHOR

Malali Gowda

The University of Trans-Disciplinary Health Sciences and Technology

Bengaluru - 560064

Email: malali.gene@gmail.com

[1] The University of Trans-Disciplinary Health Sciences and Technology, Bengaluru - 560064, India

[2] Central Muga Eri Research and Training Institute (CMER&TI), Central Silk Board, Ministry of Textiles Lahdoigarh, Jorhat, Assam, India - 785700

[3] DNA Life Organization, Yelahanka, Bengaluru - 560064, India

ONE SENTENCE SUMMARY

Genome and transcriptome analyses revealed switches the between the Calvin-Benson (C3) cycle, C4-Dicarboxylic cycle and Crassulacean acid metabolism (CAM) cycle for carbon fixation in the Peepal tree (*Ficus religiosa*) during the day and night periods.

AUTHOR CONTRIBUTIONS

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, 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.

Abstract

Peepal / Bodhi tree (*Ficus religiosa* L.) is an important long-lived keystone ecological species. This plant has been widely used by communities in Indian subcontinent in traditional medicine, Ayurveda and spirituality. Our study aimed to generate molecular resources from whole-genome and transcriptome sequencing approaches. The whole genome of Peepal tree was sequenced using Illumina and MGISEQ-2000 sequencers. We assembled the draft genome (380-Mb) of Peepal tree and annotated 35,093 protein-coding genes; 53% of its genome consists of repetitive sequences. To understand the photosynthetic pathways in leaf tissues, we analysed photosynthetically distinct conditions: bright sunny days and dark nights. The RNA-seq analysis supported the expression of 26,691 genes. The gene expression analysis of the day and night period revealed the molecular switches between the Calvin-Benson (C3) cycle, C4-Dicarboxylic cycle and Crassulacean acid metabolism (CAM) cycle for carbon fixation in leaves of Peepal tree.

Keywords: *Ficus religiosa,* genome sequencing, transcriptome sequencing, Carbon fixation pathway genes

Introduction

India's plant biodiversity is vast, rich, diverse and associated with profound science, culture and tradition. Peepal tree (*Ficus religiosa* L.,), a sacred fig that belongs to Moraceace family, is a long-lived deciduous species related to 755 fig species widespread worldwide (Van Noort et al., 2007). Peepal tree is cosmopolitan species that is a culturally and spiritually sacred plant in Buddhism, Hinduism and Jainism. This plant is popularly known as the Bodhi tree, where Buddha is believed to have undergone spiritual enlightenment underneath this tree. Hence, the culture is spread across Asia and it has been worshipped as Bodhi and Ashwatha. Generally, Peepal tree has a special significance in communities across India that are believed to produce oxygen day and night. Although the special type of stoma is called *sunken, giant or hydathode* at the lower leaf epidermis of *F. religiosa*. These are larger than the normal stomata and occur over the veins or are mixed with normal stomata, it indicates that such stomata hold gaseous and water molecules for a longer time (Chantarasuwan et al., 2014). According to our knowledge, there are no scientific evidence to claim oxygen production during a dark period.

In Ayurveda, Peepal tree has been classified as a Rasayana (a type of drug), whereby rejuvenators and antioxidants aid in reliving the body's stress (Singh et al., 2011). Peepal tree alleviates Pitta and Kapha (Ayurvedic classifications), hence prescribed as an Ayurvedic treatment for the disorders like respiratory and inflammatory disorders, ulcers, stomatitis, hiccup, arthritis, gout, skin diseases, allergies, bone fracture, diabetes etc., (Singh et al., 2011). Peepal tree has been tested for the treatment of neurodegenerative disorders like Parkinson's disease and Huntington's disease in animal models like rats (Bhangale and Acharya, 2016) and (Bhangale et al., 2016).

Next-generation sequencing (NGS) technologies have accelerated the generation of draft genomes of Moraceace plant species including mulberry (*Morus notabilis*) (He et al., 2013)

and fig (*Ficus carica*) (Usai et al., 2020). The whole-genome and transcriptome sequencing of non-model plant species revealed the Crassulacean acid metabolism pathway in pineapple (Ananas comosus) (Ming et al., 2015a) and *Kalanchoë* (Yang et al., 2017). The whole-genome sequencing of common or crystalline ice plant (*Mesembryanthemum crystallinum*) has shown to switch from Calvin- Benson Cycle (C_3) to CAM photosynthesis under salt stress (Guan et al., 2020). Using RNA-seq, the study described by comparing both species with and without the C_4 trait and different tissues within a C_4 plant suggests the ways of integration into the underlying C_3 metabolism (Schlüter et al., 2016).

In spite of its medicinal, cultural and historic importance, the molecular biology and genomics studies on the Peepal tree are scanty. Hence, the present study was hypothesized to elucidate the genome of the Peepal tree for future scientific studies. The objective of the present study was to generate a genome sequence for the Peepal tree and annotate genes. To understand the photosynthetic activity of the Peepal tree by identifying genes involved in various physiological, biochemical and other pathways.

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, Peepal tree. Paired-end reads from Illumina HiSeq1000 and MGISEQ-2000 were 132 and 322 million respectively; a total data of 88.44 billion high-quality bases (Quality>20) used for data assembly. A hybrid assembly was performed using 48.92X of Illumina reads and 118.34X MGI reads. The raw data details are given (Supplemental Table S1). The hybrid assembly analysis resulted in a 380-Mb genome. The contig N50 length is 6,385 bp and the largest contig length is 174Kb. The GC content of the Peepal tree genome is 33.70%. The gap-

closing step was performed for the hybrid assembly. There were 35,811 (5.5%) misassembled contigs and 6,04,807 (94.4%) truly assembled contig sequences in the final assembled genome. The workflow of the genome assembly is represented in (Supplemental Fig S1A). The statistics of assembly contigs and sequence assembled contigs are shown (Supplemental Table S2.1) and the scaffold summary of the genome is given in Table 1.

The completeness of the Peepal tree genome assembly was assessed with the BUSCO tool. This resulted that 76.5% (232 out of 303 genes) and 84.1% (1210 out of 1440 genes) of genes being conserved 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 1210 complete genes in the Embryophyte database 1173 are single-copy orthologs, 37 are duplicates, 105 are fragmented and 125 are missing (Supplemental Fig 2A and 2B).

Genome and pathways annotation

We identified 35,093 protein-coding genes with the complete structures in the Peepal tree genome (Supplemental File S7.1 and S7.2). RNA-seq data from two leaf tissue 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. Of the 35,093 genes predicted, 81.86 % (28,729 genes) were having the RNA-seq evidence. About 76.3 per cent of RNA-seq reads from the leaf tissue from night and day were mapped to the annotated genes of the Peepal tree. The complete set of annotated genes and their amino acid sequences were used in the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis (Moriya et al., 2007), resulting in metabolism, biosynthesis of secondary metabolites, genetic and environmental information processing, signal transduction pathway genes were common

and several others were found. 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). As importantly 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 listed in (Supplemental Table S2.4).

Protein family and Gene Ontology analysis

The protein family (Pfam) ID and Gene Ontology (GO terms) were assigned to genes using an InterProScan module (Jones et al., 2014). Out of 35,093 genes, 24,163 consisted of Pfam IDs that were distributed across 3759 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 were listed in (Supplemental File S8). Similarly, the protein families were identified for the unique characterised transcripts of RNA data, out of 26,691 transcripts, 19175 consists Pfam IDs that were distributed across 3977 types of Pfam domains and their gene ontology terms were also identified (Supplemental File S9).

The Catalase is an antioxidant enzyme known to catalyze H_2O_2 into water and oxygen. The Catalase gene (FRLM_016351-RA) and its isozyme CAT1 Catalase isozyme 1 (FRLM_016350-RA), (FRLM_012250-RA) structural gene were identified in the genome of the Peepal tree. Two catalase genes were identified in differential expression data, KatE enciphers as a monofunctional catalase and KatG enciphers as a catalase-peroxidase. KatE is known as CatB expression in the day (FPKM – 937.49) and night (FPKM – 1786.02) transcript abundance. KatG expression in the day (FPKM – 162.03) and night (FPKM – 81.53) transcript

abundance. KatE gene has been involved in physiological pathways like Glyoxylate and dicarboxylate metabolism, Tryptophan metabolism, MAPK signalling pathway – plant, FoxO signalling pathway, serine-pyruvate transaminase and KatG Tryptophan metabolism, Tyrosine metabolism, Biosynthesis of secondary metabolism, Drug metabolism.

Identification of orthologous and singleton genes

To understand the orthologous gene evolution and relationships across the species, we performed the orthologous gene detection for the Peepal tree with the other 5 species. Orthologous gene prediction with multiple species could be called part of the comparative genomic study. The predicted proteins from the genome study were considered to be proteomes of the species that were used for orthologous gene prediction and clustering. Clustering of genes provides information about taxonomic and phylogenetic classifications of genes, it helps in understanding the origin, evolution and other molecular characteristics of the genus and species. Orthologous clustering of the proteomes of six species inclusive of model organism *A. thaliana* and Ficus closest species *M. notabilis* and other species of closely related Moraceae family were chosen for the analysis. *F. religiosa, A. thaliana*, *M. notabilis, Cannabis sativa, Prunus persica*, and *Ziziphus jujuba* formed 24310 orthologous gene clusters and these genes are conserved within the species. The number of specific orthologous gene clusters in *F. religiosa, A. thaliana, M. notabilis, C. sativa, P. persica*, and *Z. jujuba* were 15016, 16170, 16517, 15921, 16655 and 17235 respectively. However, 1184 single-copy gene clusters were found across the six species and specific singletons for *F. religiosa, A. thaliana, M. notabilis, C. sativa, P. persica,* and *Z. jujuba* were 10154, 4469, 2284, 1912, 1802, 4209, respectively (Fig 3A). The identified single copy clusters were able to suggest the taxonomic and phylogenetic relationships among a group of species. Based on single-copy orthologous clustering, we deduced the phylogenetic tree for the *F. religiosa* and the other five species. A

phylogenetic tree was constructed based on a multiple sequence alignment (MSA) and Neighbour-Joining (NJ) method. It was found that *F. religiosa* is closely related to *M. notabilis* followed by *C. sativa, Z. jujuba, P. persica,* and *A. thaliana* (Supplemental Fig 3B).

Repeats in the genome of 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% (269.62Mb) repetitive sequences in the Peepal tree genome. The RNA elements, long terminal repeats (LTR) high constitute about 5% repeats and 43.71% of the repeats did not belong to any of the annotated repeats families. The percentage of repetitive sequences in the Peepal tree genome is similar to its Moraceae family, the closest species mulberry (*M. notabilis*) genome (47%) and *Ficus macrocarpa* which is composed of 46.5%. 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%) (Supplemental Table S2.3).

Simple sequence repeats (SSRs)

We predicted SSRs from the assembled Peepal tree genome. In total, 7,99,992 SSRs were identified on 2,67,593 sequences, which is composed of mono- (6,06,169), di- (1,43,113), tri- (34,327), tetra- (11,791), penta- (2,911), and hexa- (1,681) type repeats (Supplemental Table S3). Among mono repeats, the 'A/T' (73.91%) type was the highest followed by 'C/G' (1.87%). Similarly, 'AT /TA', 'AG/CT', 'AC/GT' and 'CG/CG' type 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 most predominant in tetra repeats. The detailed distribution of all types of repeats is shown in (Supplemental File S11.1 and S11.2).

Transcription Factors (TFs)

Transcription factors act in regulating the gene expression due to several external and internal signals by activating or suppressing 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 (Jin et al., 2017) using *A. thaliana* protein sequence as a reference.

A total of 1,264 protein sequences of genome annotated 35,093 protein-coding genes have a shred of evidence for 56 families of Transcription factors (TFs) (Supplemental File S10.1). The TFs families include the abundant ERF, M-type MADS, ARF, DBB, MIKC MADS, WOX, C3H, G2-like, MYB, TALE, B3, HB-other and MYB-related family proteins. The differential gene expressed data from the day and night period collected leaf tissue transcriptome were analysed to identify the TFs. From the day period sample, the 2 transcripts had specific TFs like C3H family protein and nuclear factor Y, subunit A7 (NF-YA7) and in the night period sample, the 6 transcripts had specific TFs like ERF family protein, CONSTANS-like 2, MYB-related family protein (Supplemental File S10.2).

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 combined assembly for the reads of both samples. The combined transcriptome assembly was performed and contributed to 152.8-Mb assembled bases with the N50 length of 2076 bp and an average transcript length of 13.16 Kb and 42.17% GC content. The transcriptome assembly and annotation workflow are given in (Supplemental Fig 1B). The statistics of assembly contigs and sequence assembled contigs are shown in Supplementary Table S4.1.

The *de novo* assembled transcript (1,16,038) sequences were processed for annotation. *De novo* assembled transcripts were clustered to exclude the redundant transcripts and identified the unique transcripts (26,691) 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 differential expressed genes from the day and night periods with their FPKM values are given in Supplementary Table S6.

Non-coding RNA genes in 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). For the non-coding (6,780) transcripts, 4219 transcripts were aligned on genome annotated genes using BLASTN, the remaining 2561 transcripts were considered 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, 14605 were having alignment to genome annotated genes using BLASTN. Further, categorization of specific day and night sample transcripts resulted in protein-coding 6,628 and 7,339 and non-coding of 20,528 and 25,494 transcripts respectively. From those non-coding transcripts of day and night, 18,893 and 22,232 transcripts got aligned 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.

miRNAs: miRNAs are a major class of non-coding RNAs. Based on the homology search, we have identified the microRNA precursors using the miRbase database (http://www.mirbase.org). These microRNAs belong to MIR396, MIR2916, MIR156, MIR164, MIR6236, MIR166, MIR168, MIR395 families. Among the identified miRNAs, MIR408 was found specific to the

night period transcripts of the Peepal tree. On the genome, MIR 408 was found on the genes like TPK5 Two-pore potassium channel 5, prfA peptide chain release factor 1 and also on proteins of unknown function. MIR408 is a highly conserved microRNA in plants. miR408 has enriched photosynthesis by mitigating the efficiency of irradiation utilization and the capacity for carbon dioxide fixation ((Pan et al., 2018).

In the *de novo*-based transcriptome data, the non-redundant unigenes were used to identify the microRNAs MIR168 and MIR166 and their homologs were found on two of the transcripts. We identified the miRNAs on genomic scaffolds based on mapping the transcriptome data to the genome. This provides the information of 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 (Supplemental Table S5.1). The statistics of transfer RNAs (tRNA) were identified and their details are given in the Supplemental Table S5.2.

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 two conditions the day period (2 PM) and the night period (2 AM). Through the pathway analysis, the candidate genes for carbon fixation pathways like Crassulacean Acid Metabolism (CAM) pathway, Calvin-Benson cycle (C3) pathway and C4 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 for the key fructose-bisphosphate aldolase class I, fructose-1,6-bisphosphate, phosphoenolpyruvate carboxylase (PEPC/PPC), phosphoenolpyruvate carboxylase kinase (PPCK), NAD+ and NADP+, malate dehydrogenase (MDH), pyruvate orthophosphate dikinase (PPDK) (Supplemental Table S7). The differential genes mapped to the reference

carbon fixation pathway on the KEGG database are provided in (Supplemental Fig 4) and the diagrammatic representation of the genes involved in the pathway is shown in (Supplemental Fig 5A).

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) genes were highly enriched in the photosynthetic leaf tissue collected during the night period than day period. It indicates that the Peepal tree adapts to the CAM pathway and is able to fix nocturnal carbon dioxide using the PEP carboxylase (PEPC) enzyme and accumulate malate by the enzyme malate dehydrogenase in the large vacuoles of their cells. The transcriptomic genes from the Peepal tree were mapped to the complete pathway of the CAM cycle in the KEGG database.

Ribulose-bisphosphate carboxylase (RuBP carboxylase or *rubisco*) small chain enzyme is enriched in leaf tissue collected during day period (2 PM). Rubisco is the most abundant protein in chloroplasts. The glyceraldehyde-3-phosphate dehydrogenase (NADP+) is enriched in day collected leaf tissue, the enzyme responsible for the reversible conversion of glyceraldehyde 3-phosphate to ribulose bisphosphate using ATP, the acceptor for CO2 and completes the Calvin-Benson Cycle (C3 cycle). For C3 cycle pathway except 3 genes, all other genes got mapped in the database. The 3 genes were EC 4.1.2.22 fructose-6-phosphate phosphoketolase and EC 4.1.2.9 phosphoketolase evolved from fungal, and EC 2.7.1.14 sedoheptulokinase evolved from animal kingdom and these were not found in Peepal tree for C3 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. The second highest expression is NADP-malate dehydrogenase (MDH), which converts the oxaloacetate generated by PEPC to malate. The genes from Peepal tree differential expression data had a complete map 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 period. The Fragments Per Kilobase of transcript per Million mapped reads (FPKM) and Trimmed Mean of M-values (TMM) values for the genes expressed for carbon fixation pathway are shown in Fig 6A, B, C and 7A, B and C.

Discussion

In this study the genomics and transcriptomics data was generated and also annotated for 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 genome of Peepal tree and described the whole genome. The photosynthetic tissues of distinct condition were used for transcriptome sequencing and to understand the genes, proteins and molecular pathways. The assembled genome resulted in size of 381-Mb with 35,093 protein-coding genes based on mRNA evidence used for annotation. The N50 value 6385bp was observed. A total of about 53% of the genome consisted of the repetitive sequences. The GC content of the genome is 33.70%, indicating that Peepal tree genome is AT rich. The transcriptome analysis resulted in the 26,691 unique

317 transcripts. Based on BUSCO analysis, 84.1% of genes had completeness of Peepal tree
318 genome assembly for the conserved genes in plant universal single-copy orthologs data set
319 (embryophyta database) and 81.5% had single-copy orthologs in the genome; and 76.5%
320 which indicates the completeness of Peepal tree genome assembly for the conserved genes in
321 eukaryotic universal single-copy orthologs data set (eukaryota database) and 70.6% had single-
322 copy orthologs in the genome, it has provided the confidence for the downstream analyses. The
323 protein-coding genes identified in the Peepal tree are related to other plant genomes sequenced
324 over the decades. Similarly, in our laboratory *Santalum album* (Sandalwood) genome was
325 sequenced with the prediction of 38,119 protein-coding genes based on mRNA and peptide
326 evidence (Mahesh et al., 2018).

327 To understand the molecular functions in the photosynthesis of the Peepal tree in the diurnal
328 and nocturnal periods. We have performed a downstream analysis of the genome and
329 transcriptome of the Peepal tree. Analysis of microRNAs, TFs, and carbon fixation pathways
330 has led to the understanding of the photosynthesis in the Peepal tree. The long non-coding
331 RNAs have important roles in metabolic, physiological and biological processes of the cell
332 (Jones-Rhoades et al., 2006). MIR408 has been reported to target various blue copper protein
333 members and phytocyanin family (Jones-Rhoades et al., 2006), (Sunkar et al., 2005). In
334 Arabidopsis, miR408 responds to copper deficiency and light (Abdel-Ghany and Pilon, 2008).
335 In *Oryza sativa* MiR408 plants were efficient at saving and converting light energy into sugars,
336 suggesting that miR408 might promote grain yield through regulating the phytocyanin (PC)
337 content and photosynthesis by down-regulating UCL8 uclacyanin (UCL) gene of the
338 phytocyanin family. So, MIR408 found specific expression in transcripts of night period leaf
339 tissue of Peepal tree indicated the absence of light (no expression in day period) and
340 accumulation of sugars in the plant during dark. It also aids in photosynthesis by enhancing the
341 carbon fixation.

342

343 Transcription factors like C3H family proteins were identified by their motif of three cysteines
344 and one histidine residue, and they play an important role in the regulation of growth,
345 developmental processes and environmental responses in plants (Liu et al., 2020). In plants,
346 the nuclear factor-YA has a role in drought stress responses. In rice, NF-YA7 is involved in
347 the drought tolerance pathway which is independent of Abscisic acid manner (Lee et al., 2015).
348 Expression of C3H and NF during the day period could be referred for plant growth and
349 development in the Peepal tree. The Ethylene response factor ERF105 showed the cold-
350 regulated transcription factor gene of Arabidopsis and also functions in the freezing tolerance
351 and cold acclimation (Bolt et al., 2017). In the Peepal tree, ERF is expressed during the night
352 period indicating the accumulation of humidity in the dark. MYB -related family proteins like
353 REVEILLE 1 (RVE1) and late elongated hypocotyl gene (LHY) are expressed during the night
354 period. RVE1 functions in the circadian clock and auxin pathways (Rawat et al., 2009) and
355 LHY maintains the circadian rhythm in Arabidopsis (Mizoguchi et al., 2002). Both the RVE1
356 and LHY are found expressed in night-specific Peepal tree transcripts indicating the active
357 circadian rhythms and pathways during the dark time.

358 In the current study, Catalase gene expression was found high in night period transcript
359 abundance of Peepal tree leaf tissue. A previous study on the Peepal tree showed that leaf tissue
360 collected in night time exhibited the scotoactive opening of stomata during the night, which
361 indicates that through the stomatal opening molecular oxygen (O_2) is released by the action of
362 catalase enzyme on hydrogen peroxide (H_2O_2) (Smitha et al., 2009). Peepal tree plants grown
363 in adverse habitats showed 55% higher H_2O_2 production with about a 30% increase in
364 peroxidase activity (Smitha et al., 2009). The physiological interaction between catalase and
365 its substrate H_2O_2 in the plant was determined by quantifying H_2O_2 and assaying the catalase,

366 in which catalase showed a 4-fold increase in activity, especially during the night. Peepal tree
367 has a higher amount of H_2O_2 deposition during the night than day time, which is an indication
368 of pathway switching between carbon fixation pathways. The transcriptomic sequencing from
369 time-structure diurnal (2 PM) and nocturnal (2 AM) leaf samples showed the gene expression
370 patterns of the carbon fixation pathway. The gene expression of mRNA in the C_3, C_4 and CAM
371 cycles indicated that depending on the carbohydrate, amino acids biosynthesis and metabolism
372 and environmental conditions the plant switches between these three cycles in a time-structured
373 manner.

374 The enzymatic steps involved in the CAM cycle are similar to the C4 plant's (Christin et al.,
375 2014). Plants adapt to the CAM cycle to grow during water constraints and increase the level
376 of carbon dioxide uptake than their C_3 and C_4 cycles (NOBEL, 1991). The photorespiration in
377 C3 plants dissipates more than 25% of the carbon fixed by means of the photosynthesis
378 (Sharkey, 1988). The day period mRNA data showed that the Peepal tree may carry out the
379 diurnal carbon fixation by the C3 cycle. GGAT plays an important role in the biosynthesis and
380 metabolism of major amino acids. GGAT is involved in the photorespiratory process. It
381 catalyzes the reaction of glutamate and glyoxylate in the 2-oxoglutarate and glycine (Igarashi
382 et al., 2003). High expression of GGAT in the C_4 cycle indicates that there could be
383 photorespiration in the Peepal tree during the night.
384
385 This study of Peepal tree provides the information on plants using the CAM pathway to fix
386 nocturnal carbon dioxide using the PEP carboxylase (PEPC) enzyme and the accumulation of
387 malate by the enzyme malate dehydrogenase in the large vacuoles of their cells. The key CAM
388 genes in C3 species expressed in time periods suggested that ancestral expression patterns
389 required for CAM may have pre-dated its origin in *Yucca* (Heyduk et al., 2019). In *Kalanchoë*
390 *fedtschenkoi* genome study, the convergence in protein sequence and re-scheduling of diel

391 transcript expression of genes was reported to be involved in nocturnal CO_2 fixation, stomatal
392 movement, heat tolerance, circadian clock, and carbohydrate metabolism with the other CAM
393 species in comparison with non-CAM species (Yang et al., 2017).
394
395 Some of the previous studies in the pineapple genome revealed the gene lineage transition from
396 C_3 photosynthesis to CAM, and CAM-related genes exhibit a diel expression pattern in
397 photosynthetic tissues (Ming et al., 2015b). The evolution of CAM in *Agave* from
398 C3 photosynthesis shows that the core metabolic components required for CAM have ancient
399 genomic origins which could be traceable to non-vascular plants while regulatory proteins
400 required for diel re-programming of metabolism have shared among the recent origin of C3,
401 C4 and CAM species (Yin et al., 2018).
402
403 In summary, the genome data and transcript abundance evidence indicate the molecular switch
404 in the carbon fixation pathway of the Peepal tree (*F. religiosa*) during the day and night periods
405 depending on its physiological and environmental conditions. Our study is a foundation for
406 further experiments to determine the underlying mechanisms in C3, C4 and CAM metabolism.

407 **Conclusions**

408 In this study, we generated the genomic and transcriptomic data for Peepal/Bodhi tree.
409 Genomic data pathway analyses identified the genes associated with several physiological,
410 biochemical, metabolic and disease pathways. Differential expression data from diurnal and
411 nocturnal leaf tissue samples of Peepal revealed gene expression patterns in the Carbon fixation
412 pathway during light and dark. The transcript abundance indicates that plants could switch
413 between all the three C3, C4 and CAM pathways. Further in-vitro studies have to be
414 experimented with in future to understand clear. 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 sample and extraction of nucleic acids

The mature leaves were collected from a Peepal tree (15 years old) at DNA foundation, Anuganalu village, Hassan District, India (13.0647° N, 76.0363° E). Genomic DNA was extracted from the leaves using the Qiagen Plant Genomic DNA Miniprep kit, and the quality and quantity of DNA were confirmed using the Nanodrop. From the same Peepal tree, during the day (2 PM) and night (2 AM) periods, the leaf samples were collected and immediately placed on the dry ice. Total RNA was isolated from the leaf samples using the Qiagen Plant RNA isolation kit method and was treated with RNase-free DNase I for 30 min at 37 °C (New England BioLabs) to remove residual DNA. RNA integrity and quantity were confirmed on Qubit and Tape station using dsRNA HS 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 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 at the Next Generation Genomics Facility at the Centre for Cellular and Molecular Platforms (C-CAMP), Bengaluru. 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 at Bengaluru Genomics Center (BGC) Pvt. Ltd, Bengaluru.

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 platform at the Bengaluru Genomics Centre (BGC) Pvt. Ltd, Bengaluru.

Genome 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. The separate Illumina and MGI Seq generated raw reads were used to construct the assembly using the tools SPAdes-3.13.0 (Bankevich et al., 2012) and MaSuRCA-3.2.9 (Zimin et al., 2013) respectively. The parameters were the default kmer 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 (Boetzer et al., 2011).

The combined Illumina HiSeq and MGISEQ raw reads were used to construct the hybrid assembly using the assembly tool SPAdes-3.13.0 (Bankevich et al., 2012). The parameters were the default k-mer size 21, 33, 55 and also set 77 mer for it. The gaps in the assembly were closed by GMcloser-1.6.2. tool (Kosugi et al., 2015). The statistics of the assembly were obtained using the tool Quast version4.6.1 (Gurevich et al., 2013). The completeness and evaluation of the assembly were assessed using the tool BUSCO v3 (Simão et al., 2015) with the Embryophyte and Eukaryota database.

460 **Structural gene prediction and functional annotation:**

461 Peepal tree assembled scaffolds were processed for structural and functional gene annotation
462 using the MAKER-P v.2.31.10 software (Campbell et al., 2014). The RNA-sequenced data of
463 *Morus notabilis* (He et al., 2013) consists of expression sequence tags (ESTs) and the GFF
464 (Gene finding format) file which contains the gene features and structures of genes, protein
465 data of *A. thaliana* and RNA-sequence data of Peepal tree were imported as evidence for
466 annotation support. The structural and functional annotation of predicted genes and proteins
467 were performed using BLASTP in the Uniprot database. The protein family, structures and
468 gene ontology (GO) terms were identified for protein-coding genes using InterProScan-V5.27-
469 66.0 (Jones et al., 2014).

470 **Prediction of repetitive elements: TEs and SSR**

471 The RepeatModeller-open-1.0.11[17] and RepeatMasker-4.0 tools [18] were used for repeat
472 library building and repeat identification in the assembly respectively. The MicroSAtellite
473 Identification tool (MISA) (Thiel et al., 2003) was used for the identification of SSRs from
474 genome assembled sequences of *F. religiosa*. The parameters were set to identify perfect di-,
475 tri-, tetra-, penta- and hexa nucleotide motifs with a minimum threshold of 6, 5, 5, 5 and 5
476 repeats, respectively.

477 **Prediction of transcription factor families**

478 The prediction of transcription factor (TF) families in Peepal tree transcriptome was done using
479 Plant Transcription Factor Database v5.0 (Jin et al., 2017).

480

481 **Non-coding RNA genes**

The transfer RNAs in the Peepal tree genome were found using tRNAscan-SE (v2.0.3) (Lowe and Eddy, 1997) 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 it 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 (Kong et al., 2007).

Transcriptome sequencing, assembly and annotation

High-quality stranded RNA sequencing (ssRNA-seq) reads were assembled into putative transcripts using Trinity v2.9.0 (Grabherr et al., 2011). Assembled transcripts were passed through Transdecoder v5.02 to predict the coding sequences and also the transcripts were clustered to find the unigenes by removing the redundant transcripts using the tool CDHIT-est v.0.0.1 (Li and Godzik, 2006). Transcripts assembled from Trinity and CDHIT-v0.0.1 were used in downstream analyses for gene prediction. Unigenic transcripts 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 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 (Li and Dewey, 2011), 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. Normalised transcripts obtained from the transcript quantification methods were used in the next step for analyzing the differential gene expression of genes. FPKM and Trimmed Mean of M-values (TMM) are calculated to understand the expression levels of genes in day and night collected samples of Peepal tree. The gene expression was estimated using FPKM and TMM value minimum ≥1 for further analysis. 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 analysis were a probability value P-value of 0.001 and a fold change value of 2. 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 (Conesa et al., 2005).

Pathway Analysis

The annotated genes from the assembled genome and the differentially expressed genes from the Peepal tree leaf tissues collected during day (2.00 PM) and night (2.00 AM) were used for pathway analysis in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Moriya et al., 2007) using the BBH (bi-directional best hit) method. It provided the list of pathways where the candidate genes were mapped based on the orthologous homology alignment.

Supplementary Information

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

Supplemental Fig S1A: Flow chart of *De Novo* Whole Genome Analysis (WGA) of *Ficus religiosa*.

Supplemental Table S2.1 and S2.2: Contig assembly statistics of *F. religiosa* genome and Scaffold assembly and annotation statistics of *F. religiosa* genome

Supplemental Text File S7.1 and S7.2: Annotated gene sequence and protein sequences of *F. religiosa*

Supplemental Fig 2A: BUSCO Assessment results using the plant universal single-copy orthologs (embryophyta database).

Supplemental Fig 2B: BUSCO Assessment results using the eukaryote universal single-copy orthologs (eukaryota database).

Supplemental Table S2.4: Top 10 pathways with highest gene counts in *F. religiosa* genome

Supplemental Text File S8: Protein family ID's for the annotated genes of *F. religiosa*

Supplemental Text File S9: Protein family ID's for the unique transcripts of *F. religiosa*

Supplemental Fig 3B: Phylogenetic analysis of *Ficus religiosa* with other plant species like *A. thaliana, C. sativa*, *M. notabilis, Z. jujuba, P. persica*

Supplemental Table S2.3: Repeat content in the assembled *F. religiosa* genome

Supplemental Table S3: Simple sequence repeats (SSR) prediction in the genome of *F. religiosa*

Supplemental File S11.1 and S11.2: Distribution to different SSR repeat type classes and statistics

Supplemental File S10.1: Transcription factors identified in the annotated genes

Supplemental File S10.2: Transcription factors identified in Day and Night period transcripts

Supplemental Fig 1B: Flow chart of *De novo* Transcriptome Analysis of *Ficus religiosa*.

Supplementary Table S4.1: Assembly Statistics of *F. religiosa* Transcriptome

554 **Supplementary Table S4.2**: Uni-genes Analysis from *F. religiosa* Transcripts

555 **Supplementary Table S6:** Differential expressed genes from day and night period with their
556 FPKM values

557 **Supplemental Table S5.1:** Details on long non-coding RNAs, miRNAs in *De novo* transcripts,
558 reference-based transcripts and genome

559 **Supplemental Table S5.2:** Statistics of Transfer RNAs predicted in the genome

560 **Supplemental Fig 4:** Graphical representation of candidate genes of *Ficus religiosa* involved
561 in Carbon fixation pathway.

562 **Supplemental Fig 5A:** Diagrammatic representation of candidate genes of *of* Carbon fixation
563 pathway (C_3, C_4, CAM cycle).

564 **Data Availability**

565 The raw sequence reads have been deposited under NCBI Sequence Read Archive (SRA)
566 accession numbers SRR7244210 (Illumina sequenced *F. religiosa* WGS), SRR13827064
567 (MGISEQ sequenced *F. religiosa* WGS), SRR7343291 (*F. religiosa* transcriptome). The
568 whole-genome shotgun projects have been deposited at DDBJ/EMBL/Genbank under the
569 accession JAFMPE000000000 and transcriptome data has been deposited at
570 DDBJ/EMBL/Genbank under the accession GJAV00000000.

571 **Declaration**

572 The authors declare no conflict of financial interests.

573 **Acknowledgement**

574 We would like to thank Tata Education and Development Trust fellowship (01/2018-2019)
575 provided for Ashalatha K L for her PhD programme. We acknowledge the Next Generation

Genomics Facility at the Centre for Cellular and Molecular Platforms (C-CAMP) and Bengaluru Genomics Center Pvt. Ltd for supporting sequencing and analysis. We are grateful for the MGI team, Hong Kong for sequencing the whole genome.

Tables

Table 1: Summary of genome assembly and annotation of Peepal genome

Features	**Illumina assembly**	**MGI assembly**	**Hybrid assembly**	**Gap closed assembly**
No. of scaffolds	87060	118027	121895	121696
Largest scaffold (bp)	56337	54685	174746	174006
Total assembled bases	264662650	345020555	386980121	381047120
Minimum scaffold length (bp)	56	300	78	78
Maximum scaffold length (bp)	56337	54685	174748	174748
GC %	32.99	33.44	33.69	33.70
N50 (bp)	5340	4771	6482	6385
L50	13703	20196	15619	15539
Number of annotated genes	-	-	-	35093

Table 2: Statistics of Uni-genes in Peepal Transcripts

	Sample (Day - 2 PM)	**Sample (Night - 2 AM)**	**Combined Sample**
Number of genes	22597	23360	26691
Number of transcripts	16912	16780	18173
Percent GC	46.37	42.42	46.25
Contig N50 (nts)	1404	2053	1374
Median contig length	813	1011	753
Average Contig (nts)	1060.06	1332.73	1017.44
Total assembled bases	23954145	120608982	27156384

Figure legends

584 **Figure 3A:** Orthologous clustering of 6 species using proteome data deduced 24310
585 orthologous gene clusters and 1184 single-copy gene clusters across the above 6 species.

586 **Figure 6:** The candidate genes involved in the C3, CAM and C4 cycle.

587 A) Calvin- Benson (C3) cycle B) Crassulacean acid metabolism (CAM) cycle C) C4 cycle. X
588 – axis represents the genes involved pathways, Y – axis is the matrix of normalised expression
589 trimmmed mean of M (TMM) values; Blue graph – leaf tissue collected during day period (2
590 PM), Red graph – leaf tissue collected during night period (2 AM).

591 **Figure 7:** Gene expression pattern of *F. religiosa* carbon fixation genes across the diurnal (2
592 PM) and nocturnal (2 AM) expression data. A) C3 cycle, B) CAM cycle C) C4 cycle log2-
593 transformed Fragments Per Kilobase of transcript per Million mapped reads (FPKM) value
594 based expression profiles are shown.

595 **References**

596 **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM,**
597 **Nikolenko SI, Pham S, Prjibelski AD, et al** (2012) SPAdes: A New Genome Assembly
598 Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol **19**: 455–477
599 **Bhangale JO, Acharya NS, Acharya SR** (2016) Protective effect of Ficus religiosa (L.)
600 against 3-nitropropionic acid induced Huntington disease. Orient Pharm Exp Med **16**:
601 165–174
602 **Bhangale JO, Acharya SR** (2016) Anti-Parkinson Activity of Petroleum Ether Extract of
603 Ficus religiosa (L.) Leaves. Adv Pharmacol Sci **2016**: 1–9
604 **Boetzer M, Henkel C V., Jansen HJ, Butler D, Pirovano W** (2011) Scaffolding pre-
605 assembled contigs using SSPACE. Bioinformatics **27**: 578–579
606 **Bolt S, Zuther E, Zintl S, Hincha DK, Schmülling T** (2017) ERF105 is a transcription factor
607 gene of Arabidopsis thaliana required for freezing tolerance and cold acclimation. Plant
608 Cell Environ **40**: 108–120
609 **Campbell MS, Law M, Holt C, Stein JC, Moghe GD, Hufnagel DE, Lei J,**
610 **Achawanantakun R, Jiao D, Lawrence CJ, et al** (2014) MAKER-P: A Tool Kit for the
611 Rapid Creation, Management, and Quality Control of Plant Genome Annotations. Plant
612 Physiol **164**: 513–524
613 **Chantarasuwan B, Baas P, van Heuven B-J, Baider C, van Welzen PC** (2014) Leaf
614 anatomy of Ficus subsection U rostigma (Moraceae). Bot J Linn Soc **175**: 259–281
615 **Christin P-A, Arakaki M, Osborne CP, Bräutigam A, Sage RF, Hibberd JM, Kelly S,**

Covshoff S, Wong GK-S, Hancock L, et al (2014) Shared origins of a key enzyme during the evolution of C4 and CAM metabolism. J Exp Bot **65**: 3609–3621

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics **21**: 3674–3676

Guan Q, Tan B, Kelley TM, Tian J, Chen S (2020) Physiological Changes in Mesembryanthemum crystallinum During the C3 to CAM Transition Induced by Salt Stress. Front Plant Sci. doi: 10.3389/fpls.2020.00283

Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics **29**: 1072–1075

He N, Zhang C, Qi X, Zhao S, Tao Y, Yang G, Lee T-H, Wang X, Cai Q, Li D, et al (2013) Draft genome sequence of the mulberry tree Morus notabilis. Nat Commun **4**: 2445

Heyduk K, Ray JN, Ayyampalayam S, Moledina N, Borland A, Harding SA, Tsai C-J, Leebens-Mack J (2019) Shared expression of crassulacean acid metabolism (CAM) genes pre-dates the origin of CAM in the genus Yucca. J Exp Bot **70**: 6597–6609

Igarashi D, Miwa T, Seki M, Kobayashi M, Kato T, Tabata S, Shinozaki K, Ohsumi C (2003) Identification of photorespiratory glutamate:glyoxylate aminotransferase (GGAT) gene in Arabidopsis. Plant J **33**: 975–987

Jin J, Tian F, Yang D-C, Meng Y-Q, Kong L, Luo J, Gao G (2017) PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res **45**: D1040–D1045

Jones-Rhoades MW, Bartel DP, Bartel B (2006) MicroRNAs AND THEIR REGULATORY ROLES IN PLANTS. Annu Rev Plant Biol **57**: 19–53

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics **30**: 1236–1240

Kong L, Zhang Y, Ye Z-Q, Liu X-Q, Zhao S-Q, Wei L, Gao G (2007) CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res **35**: W345–W349

Kosugi S, Hirakawa H, Tabata S (2015) GMcloser: closing gaps in assemblies accurately with a likelihood-based selection of contig or long-read alignments. Bioinformatics btv465

Lee D-K, Kim H Il, Jang G, Chung PJ, Jeong JS, Kim YS, Bang SW, Jung H, Choi Y Do, Kim J-K (2015) The NF-YA transcription factor OsNF-YA7 confers drought stress tolerance of rice in an abscisic acid independent manner. Plant Sci **241**: 199–210

Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics **12**: 323

Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics **22**: 1658–1659

Liu C, Xu X, Kan J, Cheng Z ming, Chang Y, Lin J, Li H (2020) Genome-wide analysis of the C3H zinc finger family reveals its functions in salt stress responses of Pyrus betulaefolia. PeerJ **8**: e9328

Lowe TM, Eddy SR (1997) tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. Nucleic Acids Res **25**: 955–964

Mahesh HB, Subba P, Advani J, Shirke MD, Loganathan RM, Chandana SL, Shilpa S, Chatterjee O, Pinto SM, Prasad TSK, et al (2018) Multi-Omics Driven Assembly and Annotation of the Sandalwood (Santalum album) Genome. Plant Physiol **176**: 2772–2788

Ming R, VanBuren R, Wai CM, Tang H, Schatz MC, Bowers JE, Lyons E, Wang M-L, Chen J, Biggers E, et al (2015a) The pineapple genome and the evolution of CAM

666 photosynthesis. Nat Genet **47**: 1435–1442
667 **Ming R, VanBuren R, Wai CM, Tang H, Schatz MC, Bowers JE, Lyons E, Wang ML,**
668 **Chen J, Biggers E, et al** (2015b) The pineapple genome and the evolution of CAM
669 photosynthesis. Nat Genet **47**: 1435–1442
670 **Mizoguchi T, Wheatley K, Hanzawa Y, Wright L, Mizoguchi M, Song H-R, Carré IA,**
671 **Coupland G** (2002) LHY and CCA1 Are Partially Redundant Genes Required to
672 Maintain Circadian Rhythms in Arabidopsis. Dev Cell **2**: 629–641
673 **Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M** (2007) KAAS: an automatic
674 genome annotation and pathway reconstruction server. Nucleic Acids Res **35**: W182–
675 W185
676 **NOBEL PS** (1991) Achievable productivities of certain CAM plants: basis for high values
677 compared with C 3 and C 4 plants. New Phytol **119**: 183–205
678 **Van Noort S, Gardiner AJ, Tolley KA** (2007) New records of Ficus (Moraceae) species
679 emphasize the conservation significance of inselbergs in Mozambique. South African J
680 Bot **73**: 642–649
681 **Pan J, Huang D, Guo Z, Kuang Z, Zhang H, Xie X, Ma Z, Gao S, Lerdau MT, Chu C, et**
682 **al** (2018) Overexpression of microRNA408 enhances photosynthesis, growth, and seed
683 yield in diverse plants. J Integr Plant Biol **60**: 323–340
684 **Rawat R, Schwartz J, Jones MA, Sairanen I, Cheng Y, Andersson CR, Zhao Y, Ljung K,**
685 **Harmer SL** (2009) REVEILLE1, a Myb-like transcription factor, integrates the circadian
686 clock and auxin pathways. Proc Natl Acad Sci **106**: 16883–16888
687 **Schlüter U, Denton AK, Bräutigam A** (2016) Understanding metabolite transport and
688 metabolism in C4 plants through RNA-seq. Curr Opin Plant Biol **31**: 83–90
689 **Sharkey TD** (1988) Estimating the rate of photorespiration in leaves. Physiol Plant **73**: 147–
690 152
691 **Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM** (2015) BUSCO:
692 assessing genome assembly and annotation completeness with single-copy orthologs.
693 Bioinformatics **31**: 3210–3212
694 **Singh D, Singh B, Goel RK** (2011) Traditional uses, phytochemistry and pharmacology of
695 Ficus religiosa: A review. J Ethnopharmacol **134**: 565–583
696 **Smitha RB, Bennans T, Mohankumar C, Benjamin S** (2009) Oxidative stress enzymes in
697 Ficus religiosa L.: Biochemical, histochemical and anatomical evidences. J Photochem
698 Photobiol B Biol **95**: 17–25
699 **Sunkar R, Girke T, Jain PK, Zhu J-K** (2005) Cloning and Characterization of MicroRNAs
700 from Rice. Plant Cell **17**: 1397–1411
701 **Thiel T, Michalek W, Varshney R, Graner A** (2003) Exploiting EST databases for the
702 development and characterization of gene-derived SSR-markers in barley (Hordeum
703 vulgare L.). Theor Appl Genet **106**: 411–422
704 **Usai G, Mascagni F, Giordani T, Vangelisti A, Bosi E, Zuccolo A, Ceccarelli M, King R,**
705 **Hassani-Pak K, Zambrano LS, et al** (2020) Epigenetic patterns within the haplotype
706 phased fig (Ficus carica L.) genome. Plant J **102**: 600–614
707 **Yang X, Hu R, Yin H, Jenkins J, Shu S, Tang H, Liu D, Weighill DA, Cheol Yim W, Ha**
708 **J, et al** (2017) The Kalanchoë genome provides insights into convergent evolution and
709 building blocks of crassulacean acid metabolism. Nat Commun **8**: 1899
710 **Yin H, Guo H-B, Weston DJ, Borland AM, Ranjan P, Abraham PE, Jawdy SS, Wachira**
711 **J, Tuskan GA, Tschaplinski TJ, et al** (2018) Diel rewiring and positive selection of
712 ancient plant proteins enabled evolution of CAM photosynthesis in Agave. BMC
713 Genomics **19**: 588
714 **Zimin A V., Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA** (2013) The MaSuRCA
715 genome assembler. Bioinformatics **29**: 2669–2677

716

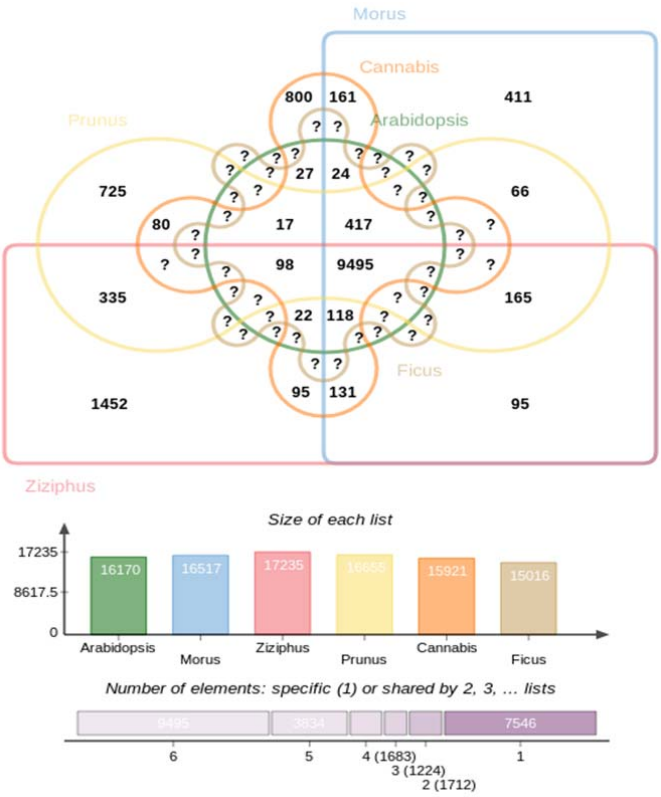

Figure 3A: Orthologous clustering of 6 species using proteome data deduced 24310 orthologous gene clusters and 1184 single-copy gene clusters across the above 6 species

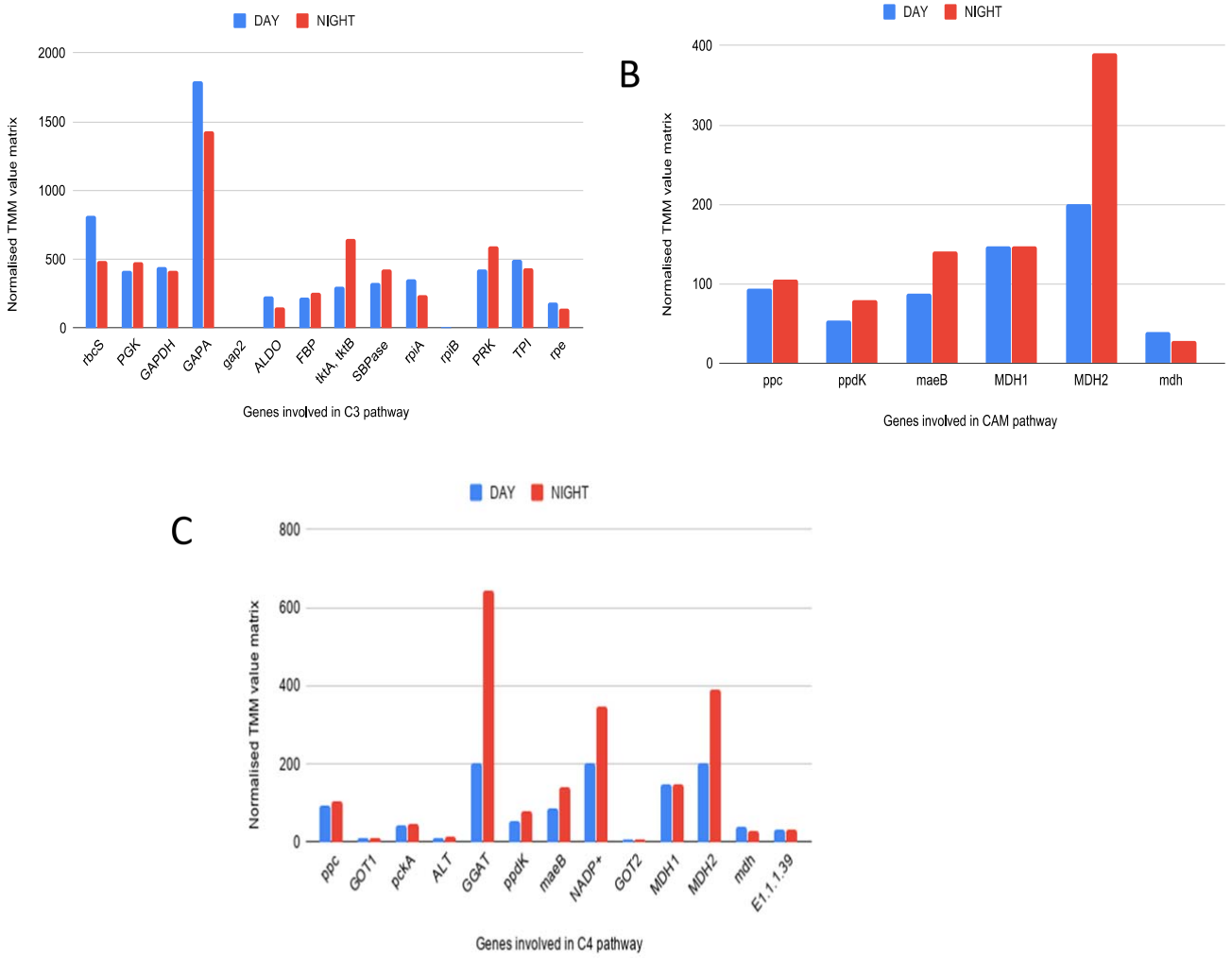

Figure 6: The candidate genes involved in the C3, CAM and C4 cycle.

A) Calvin- Benson (C3) cycle B) Crassulacean acid metabolism (CAM) cycle C) C4 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)

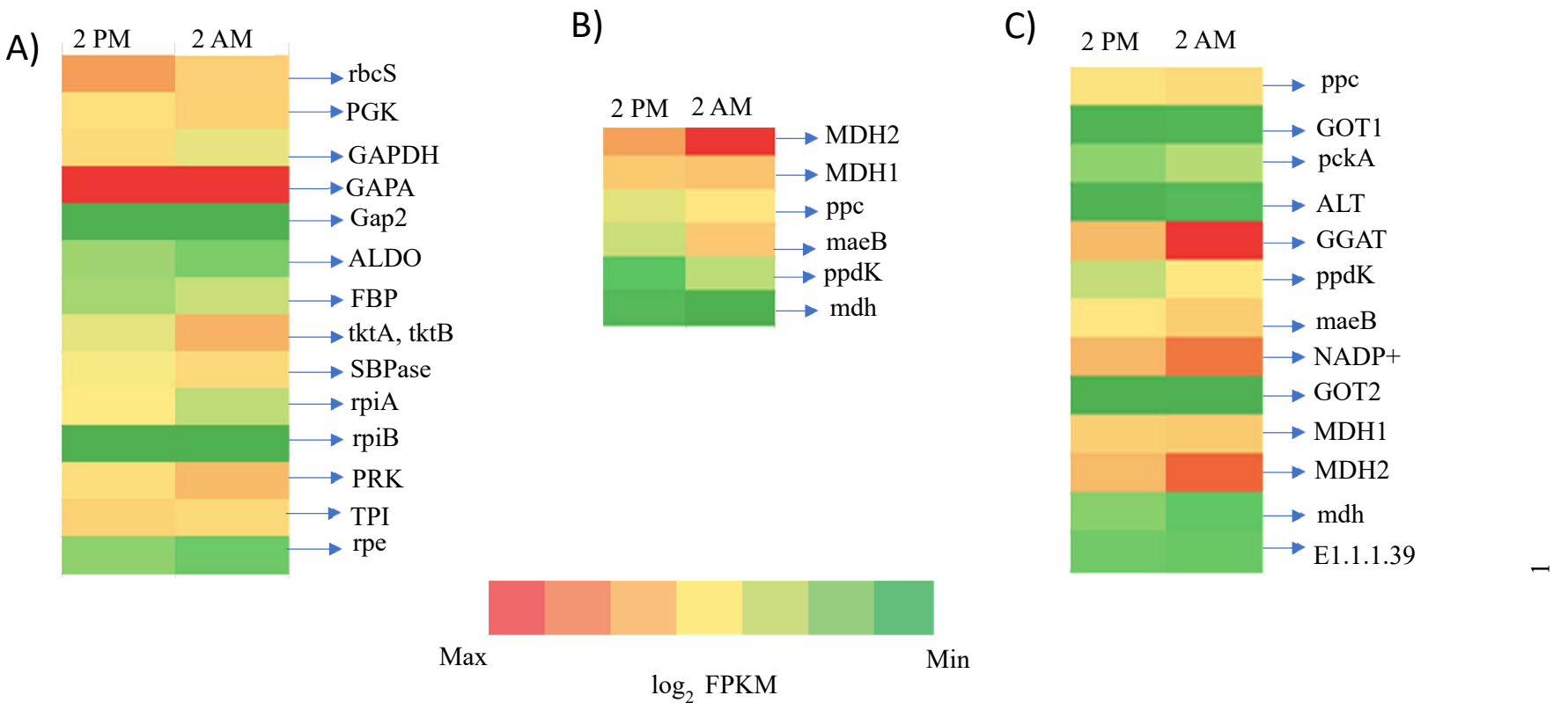

Figure 7: Gene expression pattern of *F. religiosa* carbon fixation genes across the diurnal (2 PM) and nocturnal (2 AM) expression data. A) C3 cycle, B) CAM cycle C) C4 cycle log2-transformed Fragments Per Kilobase of transcript per Million mapped reads (FPKM) value based expression profiles are shown

Parsed Citations

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al (2012) SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J Comput Biol 19: 455–477
Google Scholar: Author Only Title Only Author and Title

Bhangale JO, Acharya NS, Acharya SR (2016) Protective effect of Ficus religiosa (L.) against 3-nitropropionic acid induced Huntington disease. Orient Pharm Exp Med 16: 165–174
Google Scholar: Author Only Title Only Author and Title

Bhangale JO, Acharya SR (2016) Anti-Parkinson Activity of Petroleum Ether Extract of Ficus religiosa (L.) Leaves. Adv Pharmacol Sci 2016: 1–9
Google Scholar: Author Only Title Only Author and Title

Boetzer M, Henkel C V., Jansen HJ, Butler D, Pirovano W (2011) Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27: 578–579
Google Scholar: Author Only Title Only Author and Title

Bolt S, Zuther E, Zintl S, Hincha DK, Schmülling T (2017) ERF105 is a transcription factor gene of Arabidopsis thaliana required for freezing tolerance and cold acclimation. Plant Cell Environ 40: 108–120
Google Scholar: Author Only Title Only Author and Title

Campbell MS, Law M, Holt C, Stein JC, Moghe GD, Hufnagel DE, Lei J, Achawanantakun R, Jiao D, Lawrence CJ, et al (2014) MAKER-P: A Tool Kit for the Rapid Creation, Management, and Quality Control of Plant Genome Annotations. Plant Physiol 164: 513–524
Google Scholar: Author Only Title Only Author and Title

Chantarasuwan B, Baas P, van Heuven B-J, Baider C, van Welzen PC (2014) Leaf anatomy of Ficus subsection U rostigma (Moraceae). Bot J Linn Soc 175: 259–281
Google Scholar: Author Only Title Only Author and Title

Christin P-A, Arakaki M, Osborne CP, Bräutigam A, Sage RF, Hibberd JM, Kelly S, Covshoff S, Wong GK-S, Hancock L, et al (2014) Shared origins of a key enzyme during the evolution of C4 and CAM metabolism. J Exp Bot 65: 3609–3621
Google Scholar: Author Only Title Only Author and Title

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21: 3674–3676
Google Scholar: Author Only Title Only Author and Title

Guan Q, Tan B, Kelley TM, Tian J, Chen S (2020) Physiological Changes in Mesembryanthemum crystallinum During the C3 to CAM Transition Induced by Salt Stress. Front Plant Sci. doi: 10.3389/fpls.2020.00283
Google Scholar: Author Only Title Only Author and Title

Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics 29: 1072–1075
Google Scholar: Author Only Title Only Author and Title

He N, Zhang C, Qi X, Zhao S, Tao Y, Yang G, Lee T-H, Wang X, Cai Q, Li D, et al (2013) Draft genome sequence of the mulberry tree Morus notabilis. Nat Commun 4: 2445
Google Scholar: Author Only Title Only Author and Title

Heyduk K, Ray JN, Ayyampalayam S, Moledina N, Borland A, Harding SA, Tsai C-J, Leebens-Mack J (2019) Shared expression of crassulacean acid metabolism (CAM) genes pre-dates the origin of CAM in the genus Yucca. J Exp Bot 70: 6597–6609
Google Scholar: Author Only Title Only Author and Title

Igarashi D, Miwa T, Seki M, Kobayashi M, Kato T, Tabata S, Shinozaki K, Ohsumi C (2003) Identification of photorespiratory glutamate:glyoxylate aminotransferase (GGAT) gene in Arabidopsis. Plant J 33: 975–987
Google Scholar: Author Only Title Only Author and Title

Jin J, Tian F, Yang D-C, Meng Y-Q, Kong L, Luo J, Gao G (2017) PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res 45: D1040–D1045
Google Scholar: Author Only Title Only Author and Title

Jones-Rhoades MW, Bartel DP, Bartel B (2006) MicroRNAs AND THEIR REGULATORY ROLES IN PLANTS. Annu Rev Plant Biol 57: 19–53
Google Scholar: Author Only Title Only Author and Title

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics 30: 1236–1240
Google Scholar: Author Only Title Only Author and Title

Kong L, Zhang Y, Ye Z-Q, Liu X-Q, Zhao S-Q, Wei L, Gao G (2007) CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res 35: W345–W349
Google Scholar: Author Only Title Only Author and Title

Kosugi S, Hirakawa H, Tabata S (2015) GMcloser: closing gaps in assemblies accurately with a likelihood-based selection of contig or long-read alignments. Bioinformatics btv465
Google Scholar: Author Only Title Only Author and Title

Lee D-K, Kim H Il, Jang G, Chung PJ, Jeong JS, Kim YS, Bang SW, Jung H, Choi Y Do, Kim J-K (2015) The NF-YA transcription factor OsNF-YA7 confers drought stress tolerance of rice in an abscisic acid independent manner. Plant Sci 241: 199–210
Google Scholar: Author Only Title Only Author and Title

Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12: 323
Google Scholar: Author Only Title Only Author and Title

Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22: 1658–1659
Google Scholar: Author Only Title Only Author and Title

Liu C, Xu X, Kan J, Cheng Z ming, Chang Y, Lin J, Li H (2020) Genome-wide analysis of the C3H zinc finger family reveals its functions in salt stress responses of Pyrus betulaefolia. PeerJ 8: e9328
Google Scholar: Author Only Title Only Author and Title

Lowe TM, Eddy SR (1997) tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. Nucleic Acids Res 25: 955–964
Google Scholar: Author Only Title Only Author and Title

Mahesh HB, Subba P, Advani J, Shirke MD, Loganathan RM, Chandana SL, Shilpa S, Chatterjee O, Pinto SM, Prasad TSK, et al (2018) Multi-Omics Driven Assembly and Annotation of the Sandalwood (Santalum album) Genome. Plant Physiol 176: 2772–2788
Google Scholar: Author Only Title Only Author and Title

Ming R, VanBuren R, Wai CM, Tang H, Schatz MC, Bowers JE, Lyons E, Wang M-L, Chen J, Biggers E, et al (2015a) The pineapple genome and the evolution of CAM photosynthesis. Nat Genet 47: 1435–1442
Google Scholar: Author Only Title Only Author and Title

Ming R, VanBuren R, Wai CM, Tang H, Schatz MC, Bowers JE, Lyons E, Wang ML, Chen J, Biggers E, et al (2015b) The pineapple genome and the evolution of CAM photosynthesis. Nat Genet 47: 1435–1442
Google Scholar: Author Only Title Only Author and Title

Mizoguchi T, Wheatley K, Hanzawa Y, Wright L, Mizoguchi M, Song H-R, Carré IA, Coupland G (2002) LHY and CCA1 Are Partially Redundant Genes Required to Maintain Circadian Rhythms in Arabidopsis. Dev Cell 2: 629–641
Google Scholar: Author Only Title Only Author and Title

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res 35: W182–W185
Google Scholar: Author Only Title Only Author and Title

NOBEL PS (1991) Achievable productivities of certain CAM plants: basis for high values compared with C 3 and C 4 plants. New Phytol 119: 183–205
Google Scholar: Author Only Title Only Author and Title

Van Noort S, Gardiner AJ, Tolley KA (2007) New records of Ficus (Moraceae) species emphasize the conservation significance of inselbergs in Mozambique. South African J Bot 73: 642–649
Google Scholar: Author Only Title Only Author and Title

Pan J, Huang D, Guo Z, Kuang Z, Zhang H, Xie X, Ma Z, Gao S, Lerdau MT, Chu C, et al (2018) Overexpression of microRNA408 enhances photosynthesis, growth, and seed yield in diverse plants. J Integr Plant Biol 60: 323–340
Google Scholar: Author Only Title Only Author and Title

Rawat R, Schwartz J, Jones MA, Sairanen I, Cheng Y, Andersson CR, Zhao Y, Ljung K, Harmer SL (2009) REVEILLE1, a Myb-like transcription factor, integrates the circadian clock and auxin pathways. Proc Natl Acad Sci 106: 16883–16888
Google Scholar: Author Only Title Only Author and Title

Schlüter U, Denton AK, Bräutigam A (2016) Understanding metabolite transport and metabolism in C4 plants through RNA-seq. Curr Opin Plant Biol 31: 83–90
Google Scholar: Author Only Title Only Author and Title

Sharkey TD (1988) Estimating the rate of photorespiration in leaves. Physiol Plant 73: 147–152
Google Scholar: Author Only Title Only Author and Title

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31: 3210–3212
Google Scholar: Author Only Title Only Author and Title

Singh D, Singh B, Goel RK (2011) Traditional uses, phytochemistry and pharmacology of Ficus religiosa: A review. J Ethnopharmacol 134: 565–583
Google Scholar: Author Only Title Only Author and Title

Smitha RB, Bennans T, Mohankumar C, Benjamin S (2009) Oxidative stress enzymes in Ficus religiosa L.: Biochemical, histochemical and anatomical evidences. J Photochem Photobiol B Biol 95: 17–25
Google Scholar: Author Only Title Only Author and Title

Sunkar R, Girke T, Jain PK, Zhu J-K (2005) Cloning and Characterization of MicroRNAs from Rice. Plant Cell 17: 1397–1411
Google Scholar: Author Only Title Only Author and Title

Thiel T, Michalek W, Varshney R, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). Theor Appl Genet 106: 411–422
Google Scholar: Author Only Title Only Author and Title

Usai G, Mascagni F, Giordani T, Vangelisti A, Bosi E, Zuccolo A, Ceccarelli M, King R, Hassani-Pak K, Zambrano LS, et al (2020) Epigenetic patterns within the haplotype phased fig (Ficus carica L.) genome. Plant J 102: 600–614
Google Scholar: Author Only Title Only Author and Title

Yang X, Hu R, Yin H, Jenkins J, Shu S, Tang H, Liu D, Weighill DA, Cheol Yim W, Ha J, et al (2017) The Kalanchoë genome provides insights into convergent evolution and building blocks of crassulacean acid metabolism. Nat Commun 8: 1899
Google Scholar: Author Only Title Only Author and Title

Yin H, Guo H-B, Weston DJ, Borland AM, Ranjan P, Abraham PE, Jawdy SS, Wachira J, Tuskan GA, Tschaplinski TJ, et al (2018) Diel rewiring and positive selection of ancient plant proteins enabled evolution of CAM photosynthesis in Agave. BMC Genomics 19: 588
Google Scholar: Author Only Title Only Author and Title

Zimin A V., Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA (2013) The MaSuRCA genome assembler. Bioinformatics 29: 2669–2677
Google Scholar: Author Only Title Only Author and Title