Title: Genome sequencing and comparative analysis of *Ficus benghalensis* and *Ficus religiosa* trees reveal evolutionary mechanisms of longevity

Authors: Abhisek Chakraborty, Shruti Mahajan, Manohar Singh Bisht, Vineet K. Sharma *

Affiliation: MetaBioSys Group, Department of Biological Sciences, Indian Institute of Science Education and Research Bhopal

*Corresponding Author email:

Vineet K. Sharma - vineetks@iiserb.ac.in

E-mail addresses of authors:

Abhisek Chakraborty - abhisek18@iiserb.ac.in, Shruti Mahajan - shruti17@iiserb.ac.in, Manohar Singh Bisht - manohar21@iiserb.ac.in, Vineet K. Sharma - vineetks@iiserb.ac.in

ABSTRACT

Ficus benghalensis (Indian banyan tree) and *Ficus religiosa* (Peepal) trees are well-known for their long lifespan, traditional significance, and medicinal properties. Therefore, to understand the genomic and evolutionary aspect of these characteristics the whole genomes of these two *Ficus* species were sequenced using 10x Genomics and Oxford Nanopore sequencing platforms. The draft genome assemblies of *F. benghalensis* (392.89 Mbp genome containing 25,016 high-confidence coding genes), and *F. religiosa* (332.97 Mbp genome containing 23,929 high-confidence coding genes) were constructed. We also established the genome-wide phylogenetic position of the two *Ficus* trees with respect to 46 other Angiosperm plant species and studied the comparative population demographic history of these two species to show a population bottleneck event ~0.8 Mya for both the species. We also identified 7,468 orthogroups across 16 phylogenetically closer Eudicot plant species including *F. benghalensis* and *F. religiosa*. Comparative evolutionary analyses using these orthogroups, gene family expansion/contraction analysis, and gene duplication analysis showed adaptive evolution in genes involved in cellular pathways and mechanisms that are central to plant growth and development and provide genomic insights into longevity and ecological significance of these large woody trees.

INTRODUCTION

Ficus is one of the most diverse genus that originated around 80-90 million years ago, and consists of ~800 species^{1,2}. *Ficus* plants belong to the plant family Moraceae, also known as fig family consisting of over 1,100 species³. Plants from this genus are also known as fig plants for the presence of their characteristic inflorescence known as syconium or fig that helps in pollination of fig plants with the help of pollinators such as fig wasps¹. *Ficus* plants are abundant in a wide range of tropical and subtropical ecosystem. Fig plants possess several characteristics such as smaller seed size, higher growth rate, higher fecundity etc. that make them evolutionarily flexible to grow in diverse environmental habitats². Species from the *Ficus* genus occupy diverse ecological niches, and different life forms such as free-living, epiphytes, semi-epiphytes, lithophytes, or rheophytes⁴. Fig plants are considered as keystone species in a tropical ecosystem as they sustain a large number of animal species with fig fruits that are produced throughout the year⁵. Large *Ficus* plants (including *F. benghalensis* and *F. religiosa*) are crucial components of tropical ecosystem as they can be considered as forest restoration agents due to their larger canopy areas that favor plant species dispersed by frugivores as well as other species⁶. Fig plants, and their pollinator species such as fig wasps coevolved in course of time to show obligate mutualism⁷.

For pollination purpose, fig plants interact with their wasp pollinators by attracting them with volatile organic compounds (VOCs)⁷.

F. benghalensis, also known as Indian banyan tree, is the National tree of India, and is native to tropical, subtropical regions in South Asia⁸. It is also known for its traditional religious, mythological significance in Hinduism and Buddhism, and their wide-spread branches signify eternal life⁸. *F. benghalensis* (a strangler fig) is a large woody plant with around 30 meters height, and typically have a long life spanning several centuries⁹. The largest banyan tree (Thimmamma Marrimanu) is located in Southern India, and is 550 years old⁹. It starts growing as an epiphyte in early days, and in course of time it disperses thick plentiful aerial roots from the spreading branches to cover several hundred meters and grows independently¹⁰. The extracts from bark, leaf and fruit of this plant possess various bioactive compounds e.g., flavonoid, saponin, sterol, anthocyanidin derivatives etc. that confer a large number of pharmacological activities e.g., anti-inflammatory, anti-helminthic, anti-oxidant, anti-microbial, anti-tumor, antipyretic, immunomodulatory, and wound healing properties⁸.

F. religiosa, also known as 'sacred fig', 'Bodhi tree', 'Peepal', or 'Ashwattha', is a large hemi epiphyte, deciduous tree with long average lifespan of 900-1,500 years^{11,12}. It is also a well-known tree for its mythological and religious importance in Buddhism and Hinduism, and is known as the oldest tree in Hindu mythology^{12,13}. Similar to *F. benghalensis*, the *F. religiosa* is also a large perennial tree with up to 20 meters height, is epiphytic in young stages and the adult tree have widespread branches but without the aerial roots^{11,14}. Different parts of this plant such as root bark, stem bark, fruits, and seeds contain essential compounds e.g., steroids, alkaloids, flavonoids, phenols, tannins, essential amino acids that have a wide range of pharmacological applications¹¹. Phytochemical analysis of the barks of *F. religiosa* showed the presence of bioactive compounds such as flavonoids, glycosides, terpenoids, tannins, saponins, wax, and others¹⁵ that can also be used as potential therapeutic agents¹⁶. These components are responsible for conferring anti-oxidant, anti-bacterial, hypoglycemic, hypolipidemic, wound-healing, antihelminthic, immunomodulatory, anti-diabetic, anti-convulsant, and anti-ulcer activities to *F. religiosa* plant^{11,15}.

F. benghalensis and *F. religiosa* genomes are diploid (2n=26) and the estimated genome size is 686 Mbp for both the species according to Plant DNA C-values Database¹⁷. A few *Ficus* species such as *F. carica*, *F. hispida*, *F. microcarpa* and *F. erecta* have been sequenced^{7,18,19}; however these two significant *Ficus* genomes are yet not sequenced. In this study, we carried out the first genome sequencing and analysis of *F. benghalensis* and *F. religiosa* using 10x Genomics linked reads and Oxford Nanopore long reads to

get evolutionary insights into the mechanisms related to medicinal properties and long lifespan. We also performed transcriptome sequencing of these species that helped in comprehensive genome annotation. The phylogenetic position of the two fig species was resolved using other available Eudicot species, and comparative evolutionary analysis with selected Angiosperm plant species revealed that genes required for plant growth, development, and stress tolerance mechanisms are highly evolved in these species, which are perhaps responsible for the longevity of these two plant species.

RESULTS

Genome and Transcriptome Sequencing

Genome sizes of *F. benghalensis* and *F. religiosa* were estimated as 431.53 Mbp and 413.14 Mbp, respectively, using SGA-preqc²⁰. Supernova²¹ v2.1.1 also predicted the genome sizes as 420.13 Mbp and 388.12 Mbp for *F. benghalensis* and *F. religiosa*, respectively. A total of 49 Gb and 9 Gb of genomic data was generated from *F. benghalensis* leaf tissue using 10x Genomics (Chromium) and Oxford Nanopore technology, respectively. Genomic data corresponded to 113.6X coverage of 10x Genomics linked read data, and 20.9X coverage of Nanopore long read data based on predicted genome size using SGA-preqc. Further, a total of 7.4 Gb of RNA-Seq data was also generated for *F. benghalensis* plant species (**Supplementary Tables 1-2**). For *F. religiosa* plant species, 10x Genomics and Oxford Nanopore sequencing technologies were used to generate 48.7 Gb (117.8X) and 7.2 Gb (17.4X) of genomic data, respectively. Further, a total of 10.8 Gb of RNA-Seq data was also generated (**Supplementary Tables 1-2**).

Genome assembly

The heterozygosity content for *F. benghalensis* was estimated to be 1.78% (**Supplementary Figure 1a**). The final genome assembly of *F. benghalensis* (assembly v3) had a total size of 392.89 Mbp with 4,822 contigs, N50 value of 486.9 Kbp, longest contig size of 5.4 Mbp, and a GC-content of 34.54% (**Supplementary Table 3**). Final genome assembly size was close to the estimated genome size of 381.9 Mbp predicted using GenomeScope²² (**Supplementary Figure 1a**). *F. benghalensis* genome assembly had 96.4% complete BUSCOs and 1.7% of fragmented BUSCOs (**Supplementary Figure 1c, Supplementary Table 4**). 98.39% of the barcode-filtered linked-reads and 93.11% adapter-processed Nanopore reads could be mapped on the final genome assembly.

The heterozygosity content for *F. religiosa* genome was estimated to be 1.62% (**Supplementary Figure 1b**). *F. religiosa* final genome assembly had a total size of 332.97 Mbp consisted of 6,087 contigs (largest contig of 4.51 Mbp), an N50 value of 553.4 Kbp, and a GC-content of 34.31% (**Supplementary Table 3**). This final *F. religiosa* genome assembly had 95.6% complete BUSCOs, and 1.1% of fragmented BUSCOs (**Supplementary Figure 1c, Supplementary Table 4**). Further, 98.14% of the barcode-filtered 10x Genomics linked reads and 93.99% of the adapter-processed Nanopore long reads could be mapped to the final genome assembly of *F. religiosa*. Final assembled genome size of *F. religiosa* was close to the predicted genome size of 337.7 Mbp using GenomeScope²² (**Supplementary Figure 1b**).

Sequence variation analysis in these final genome assemblies using barcode-filtered 10x Genomics reads identified 1,772,802 nucleotide positions (0.45%) as variant sites in *F. benghalensis*, and 1,466,212 variant sites (0.44%) in *F. religiosa*. These variant sites were located in 3,254 and 3,859 contigs (after scaffolding) in *F. benghalensis* and *F. religiosa*, respectively.

Genome annotation

A total of 2,159 repeat family sequences were identified in *F. benghalensis* genome assembly, which were clustered to obtain 1,879 repeat families that were used to repeat-mask the *F. benghalensis* genome. Using this *de novo* repeat library, 51.36% of the genome assembly was masked as repetitive regions using RepeatMasker. Among these, 47.08% was identified as interspersed repeats, which was further divided into 32.88% unclassified, 11.81% retroelements (including 2.70% Ty1/Copia, and 8.28% Gypsy/DIRS1 elements), and 2.39% DNA transposons (**Supplementary Figure 1d, Supplementary Table 5**). 7.01% of the genome was also detected as tandem repeats using TRF²³ v4.09. Thus, ~54% of the *F. benghalensis* genome was predicted to consist of simple and interspersed repeat regions. Additionally, 1,028 rRNAs, 679 tRNAs (decoding standard amino acids), and 195 miRNAs were identified in *F. benghalensis* genome.

Similarly, 2,002 *de novo* repeat family sequences were identified in *F. religiosa* final genome assembly, which were clustered to obtain 1,735 repeat sequences that were used as *de novo* repeat library. Repeat-masking of *F. religiosa* genome assembly was performed using this library. RepeatMasker analysis masked 45.88% of the *F. religiosa* genome assembly as repetitive regions (41.68% interspersed repeats). Additionally, Tandem Repeat Finder (TRF) also predicted 8.84% of the genome sequence as simple repeats. Among the interspersed repeats, 6.70% was retroelements (1.55% Ty1/Copia, 4.51% Gypsy/DIRS1 elements), 1.64% was DNA transposons, and 33.35% was unclassified (**Supplementary**

Figure 1d, Supplementary Table 6). Therefore, ~51% of the *F. religiosa* genome assembly consisted of repetitive regions (including both simple and interspersed repeats). Also, 668 tRNAs (decoding standard amino acids), 1,135 rRNAs, and 185 miRNAs were detected in the final *F. religiosa* genome assembly.

Prior to gene set construction, quality-filtered RNA-Seq data for *F. benghalensis* plant were *de novo* assembled, which resulted in a total of 10,156 transcripts (used as a set of empirical evidence in MAKER pipeline²⁴). A total of 29,524 gene models were identified from MAKER genome annotation pipeline. Among these, 27,674 genes (93.73%) having an AED value of <0.5 were extracted. Among these 27,674 genes, 25,016 were filtered out based on the length-based criteria of ≥ 300 bp and were retained in the high-confidence gene set of *F. benghalensis*. In case of *F. religiosa* species, quality-filtered RNA-Seq data was used for *de novo* transcriptome assembly resulting in 74,514 assembled transcripts, which were used for evidence-based alignments in MAKER genome annotation pipeline. The pipeline predicted a total of 27,544 coding gene sequences, among which, 26,126 genes (94.85%) had AED values <0.5. After extracting these genes, and further length-based filtering (≥ 300 bp), a total of 23,929 coding genes were retained to construct the high-confidence gene set for *F. religiosa* genome.

In the resultant coding gene set of *F. benghalensis*, 86.3% genes were identified as complete BUSCOs and 8.1% genes were identified as fragmented BUSCOs. Out of 25,016 high-confidence genes, 24,909 genes, 20,947 genes, and 20,541 genes could be mapped against NCBI-nr, Swiss-Prot, and Pfam-A databases, respectively, and 90 genes (0.36%) could not find a match in any of these databases. In case of *F. religiosa*, the resultant coding gene set had 90.1% complete BUSCOs, and 4.2% of fragmented BUSCOs. Among the 23,929 high-confidence coding genes, 23,638 genes, 19,850 genes, and 19,437 genes found matches against NCBI-nr, Swiss-Prot, and Pfam-A databases, respectively. Overall, a total of 262 genes (1.09%) could not be annotated using any of these databases.

Synteny analysis

Intra-species collinear blocks were identified using MCScanX²⁵, which showed 23.72% collinearity (5,933 out of 25,016 genes) among *F. benghalensis* coding genes and 18.65% collinearity (4,463 out of 23,929 genes) among *F. religiosa* coding genes. Further, 48.44% dispersed, 4.37% proximal, 12.93% tandem, and 23.72% segmental duplicated genes, and 10.54% singleton genes were detected in *F. benghalensis* coding gene set. *F. religiosa* contained 53.21% dispersed, 3.63% proximal, 12.99% tandem, 18.65% segmental duplicated genes, and 11.52% singleton coding genes.

503 inter-species syntenic blocks were identified, which involved 58.33% of the total coding genes predicted for *F. benghalensis* and *F. religiosa*, including 14,688 *F. benghalensis* syntelogs (58.71% of the predicted genes) and 13,861 *F. religiosa* syntelogs (57.93% of the predicted genes).

Demographic history and Phylogenetic position of *Ficus* species

Effective population size (N_e) estimation for the two *Ficus* species using PSMC²⁶ analysis showed that both the species suffered one major population bottleneck in a similar time period - around 0.8 Million years ago (**Figure 1a**) in the Pleistocene epoch of the Geological time scale, which is similar to the population demographic history of other tropical forest plants²⁷.

Phylogenetic position of *F. benghalensis* was resolved in comparison with 46 other Eudicot species spanning a total of 15 plant orders, along with *Zea mays* as an out-group species. A total of 507 fuzzy one-to-one orthogroups were constructed across all 48 selected species including *F. benghalensis* and *F. religiosa*, and maximum likelihood-based phylogenetic tree was constructed (**Figure 1b**) using the concatenated, filtered (for undetermined values) protein sequence alignments of these orthogroups, containing a total of 559,964 alignment positions.

F. religiosa and *F. benghalensis* that belong to the same genus and same plant family (Moraceae) were found to be the closest species of each other and were placed in the same clade. Four other species – *Rosa chinensis*, *Malus domestica*, *Prunus avium*, and *Cannabis sativa* from the same phylogenetic order Rosales were placed in the adjacent clades to the two *Ficus* species. Among these four species, *Cannabis sativa* belonging to the plant family Cannabaceae was the closest of *Ficus* species and shared the most recent common ancestor. *Ficus* plants belonging to Rosales plant order were closer to species from Fabales and Cucurbitales orders in comparison to other selected species, which is also supported by studies based on 18s rDNA, atpB and rbcL gene sequences²⁸, and other previous reports²⁹.

Relative phylogenetic positions of the other selected plant orders in this study were also supported by previous studies^{28,30}. The phylogenetic positions of the plant orders – Rosales, Fabales and Cucurbitales were compared and found to be comparatively closer to the plant orders Brassicales, Malpighiales, Sapindales, and Malvales. The plant orders Apiales, Asterales, Solanales, Gentianales, Lamiales, Ericales, Vitales, and Caryophyllales diverged earlier, and among them species from the plant order Caryophyllales (*Beta vulgaris*) was found to diverge the earliest among all the selected Eudicot species, and hence it was the most distant Eudicot species to the two *Ficus* species.

Expanded/contracted gene families

A total of 60,204 gene families were obtained using the coding gene sequences of the two *Ficus* species and 46 other plant species that were used for phylogenetic tree construction. 26,471 families were retained after the clade and size-based filtering of gene families. 1,710 families were expanded and 1,311 families were contracted in gene number in *F. benghalensis*. In *F. religiosa* species, 1,320 and 1,526 families were found to be expanded and contracted, respectively.

Genes with signatures of adaptive evolution

A total of 7,468 orthogroups were identified across the 16 species including *F. benghalensis* and *F. religiosa* (see Methods) that were used for the evolutionary analyses to identify the *Ficus* species-specific genes showing signatures of adaptive evolution. Among the *F. benghalensis* species-specific genes, 1,449 genes had unique amino acid substitutions compared to the other plant species, and 1,001 of these genes were identified to have a functional impact. Also, 108 genes showed positive selection (with p-values < 0.05), and 26 genes showed higher rate of evolution i.e., higher branch length distance values compared to other species. Additionally, 268 genes containing positively selected codon sites were found with >95% probability, which is a site-specific evolutionary signature. A total of 45 genes showed at least two of the three evolutionary signatures – positive selection, higher rate of evolution, and unique substitution having functional impact, and thus were identified as the genes with Multiple Signs of Adaptive evolution (MSA). All these 45 genes were found to be associated with functions that are essential for long-time survival of plants e.g., root development, flowering and reproduction, cellular machinery, metabolism etc. (**Figure 2**).

Similarly, in *F. religiosa* species a total of 111 genes were positively selected (with p-values < 0.05), and 30 genes showed higher rate of nucleotide divergence. Among the 1,360 genes that showed unique amino acid substitution, 978 genes were predicted to have functional impacts, and were considered for further analysis. Also, 255 genes possessed positively selected codon sites with >95% probability. A total of 55 genes were identified as MSA genes. All these 55 genes were found to be involved in functions that could be essential for providing longevity in plants such as root development, reproduction, seedling growth, metabolism etc. (**Figure 2**).

Enzymes involved in Shikimate and Mevalonate pathway

Shikimate and Mevalonate pathways are the two important pathways for fig pollination by production of VOCs, secondary metabolism, and plant development⁷. All seven enzymes involved in shikimate pathway and mevalonate pathway were present in the high-confidence coding gene set of *F. religiosa* and *F. benghalensis*. In *F. religiosa*, shikimate kinase, *DHQS* synthase (involved in shikimate pathway), and mevalonate kinase (involved in mevalonate pathway) showed unique amino acid substitution with functional impact compared to their orthologs in other 15 species. Furthermore, mevalonate pathway is also the precursor to terpenoid backbone biosynthesis. Two enzymes – *FPPS* (unique substitution with functional impact) and *GGPPS* (unique substitution with functional impact and positively selected codon sites) that are involved in the downstream terpenoid backbone biosynthesis pathway also showed evolutionary signatures. In *F. benghalensis*, *EPSPS* (involved in shikimate pathway) showed unique amino acid substitution with functional impact, and *GGPPS* (involved in terpenoid biosynthesis pathway) showed unique substitution with functional impact and positively selected codon sites.

Enzymes involved in mevalonate and shikimate pathway were also analyzed for gene family expansion/contraction. Gene families containing the genes of mevalonate pathway were not expanded/contracted in *F. benghalensis*, whereas one family was contracted (*MK*) in *F. religiosa*. For the gene families containing genes involved in shikimate pathway, three expanded (*DAHP* synthase, 3-dehydroquinate dehydratase, shikimate dehydrogenase) gene families were found in *F. benghalensis*, whereas one contracted family (shikimate kinase) was found in *F. religiosa*. Gene family expansion of the genes involved in shikimate pathway of *F. benghalensis* was accounted by gene duplication events. 3-dehydroquinate dehydratase and shikimate dehydrogenase genes were observed to be originated by tandem duplication, whereas *DAHP* synthase gene underwent segmental duplication.

DISCUSSION

In this study, we have performed the first whole genome sequencing and analysis of *F. benghalensis* and *F. religiosa*, two dicot fig plants belonging to Moraceae family. *Ficus* plants are broadly known for their well-developed morphological characteristics as well as their ecological, evolutionary, medicinal, and religious importance. Thus, the whole genome sequences of these two *Ficus* species will serve as a valuable resource to study the significant characteristics of *Ficus* genus as well as Moraceae plant family. The hybrid sequencing and assembly approach using 10x Genomics linked reads and Oxford Nanopore long reads helped in successful construction of *F. benghalensis* and *F. religiosa* genomes with N50 values

of 486.9 Kbp and 553.4 Kbp, respectively. A comparatively higher N50 value of *F. religiosa* genome assembly might be a result of lower heterozygosity (1.62%) and lower repeat content (51%) compared to *F. benghalensis* genome (1.78% heterozygosity and 54% repeat content). Higher BUSCO completeness and higher percentage of reads mapped to the genome assembly also attests to the genome assembly quality.

The genome size of these two *Ficus* species is similar to *F. microcarpa* genome and slightly larger than other *Ficus* - *F. carica*, *F. hispida* and *F. erecta* genomes^{7,18,19}. However, the estimated and assembled genome sizes of the *Ficus* plants in this study is smaller than their genome sizes reported at Plant C-values Database, which was obtained using experimental methodologies that depend on various factors such as the usage of reference standard plant species while estimating the genome size³¹. Thus, in this study, three different approaches - GenomeScope (**Supplementary Figures 1a-1b**), Supernova assembler, and SGA-preqc were used to estimate the genome sizes of both the *Ficus* species that provide additional confirmations on smaller genome sizes compared to the sizes reported at Plant C-values Database.

The final genome assemblies were used for coding gene prediction along with transcriptome and protein sequences as empirical evidence, and AED value criteria of <0.5 was used as quality control metric to construct the high-confidence gene set. Presence of 94.4% (complete and fragmented) and 94.3% (complete and fragmented) BUSCO genes and all the genes present in shikimate and mevalonate pathway in this high-confidence gene set attests to the comprehensiveness of the coding gene set of *F. benghalensis* and *F. religiosa*, respectively. The number of high-confidence coding genes is also similar to *F. hispida* and *F. microcarpa* genomes⁷. Further, the observed high percentage of duplicated genes in these genomes may have occurred due to expansion of gene families that provide long lifespan to plants, as in case of Oak genome³². Also, the high percentage of inter-species collinearity between the two *Ficus* genomes might be a result of *Ficus* hybridization events caused by frequent host switching of the fig pollinators³³. Among the LTR retrotransposon class of repeats, Gypsy/DIRS1 elements were higher than Ty1/Copia elements in both the species (**Supplementary Tables 5-6**), similar to other *Ficus* genomes^{7,18,34}.

The comprehensive genome-wide phylogeny including 48 species (with two fig plants) covering 15 phylogenetic plant orders constructed using a large number of one-to-one orthogroups and a decent bootstrap value (no polytomy was observed) will serve as a valuable reference for future phylogenetic studies. The predicted population bottleneck events in case of both these *Ficus* plants might be the

result of major glaciation events in Pleistocene³⁵, and the environmental changes (warm and cold conditions prevailing for longer time, lack of rainfall etc.) caused in the Middle Pleistocene transition (MPT)^{36,37}. PSMC analysis also showed that these *Ficus* species could not recover from the long-term decrease in effective population size caused by the bottleneck events.

The inclusion of closely related plant species (according to the phylogenetic tree) for comparative analysis helped in precise identification of genes with evolutionary signatures in *F. benghalensis* and *F. religiosa* species. Further the inclusion of coding genes from both the fig plants in this analysis helped to identify plausible species-specific *Ficus* genes responsible for the morphological characteristics and evolutionary adaptation in these two plant species.

Evolutionary signatures as well as gene family expansion/contraction for the enzymes involved in shikimate pathway in *F. benghalensis*, and for the enzymes involved in both shikimate and mevalonate pathways in case of *F. religiosa* suggest adaptive evolution of these pathways. Evolution of shikimate pathway, which is also responsible for the production of pollinator attracting VOCs⁷, suggest presence of fig-wasp obligate mutualism in these plant species. Similarly, the evolution of mevalonate pathway in *F. religiosa* that regulates plant developmental processes by aiding metabolism, phytohormone biosynthesis, isoprenoid biosynthesis etc.^{38,39}, indicates the adaptive evolution of plant growth and development in this species. In addition, the adaptive evolution of shikimate and mevalonate pathways, which are also the precursors for production of phenolic and terpenoid classes of secondary metabolites^{40,41}, explain the presence of medicinal properties in these plants. To support this, evolutionary signatures were also observed in other enzymes involved in terpenoid backbone biosynthesis pathway in both the species.

Genes that are associated with providing longevity were identified with signatures of adaptive evolution in both the species. Among the MSA genes related to growth and development, the genes involved in plant root development, flowering, vegetative growth, and metabolism were notable. Flowering related MSA genes in both the species also might influence fig pollination as well as fig-wasp mutualism and co-evolution. The study also reports the presence of CAM pathway of photosynthesis in *F. benghalensis* and *F. religiosa*, which is a characteristic of epiphyte or hemi epiphyte plants to conserve water storage, when they survive as epiphytes⁴². Genes involved in CAM pathway showed signatures of adaptive evolution in both these species. Studies have shown that plants that use both C₃ and CAM pathway of photosynthesis exhibit long-lived leaves, thus the evolutionary signatures of CAM pathway in *F. benghalensis* and *F. religiosa* perhaps also contribute to leaf longevity⁴³.

84% and 91% of the MSA genes were also associated with tolerance against biotic and abiotic stress responses in *F. benghalensis* and *F. religiosa*, respectively, which further assist in plant survival against environmental challenges. Presence of diverse defense or stress tolerance-related genes is also an important factor in providing long-time survival of plants⁴⁴. Stress response mechanisms such as JA and SA-signaling pathways also regulate plant secondary metabolism⁴⁵, which is further related to the medicinal properties of plants⁴⁶. Taken together, it can be speculated that genes that help to survive in diverse and challenging environmental conditions e.g., genes involved in plant growth and development (root cell growth and proliferation, auxin signaling, CAM pathway, plant senescence regulating pathways, plant metabolism, cell wall and plasma membrane component biosynthesis, intracellular machinery) have evolved to provide longer life-span in plant species like *F. benghalensis* and *F. religiosa*, which are keystone plant species in tropical and sub-tropical ecosystems^{2,5}.

METHODS

DNA-RNA extraction and Sequencing

The leaves were collected from a *F. benghalensis* tree and a *F. religiosa* tree located in Bhopal, India (23.2280252°N 77.2088987°E). The leaves were cleaned and processed for DNA and RNA extraction. The DNA extraction was carried out using Carlson lysis buffer by following Oxford Nanopore suggested protocol using Genomic tip 20 of Blood and cell culture kit (Qiagen, CA, USA) with few modifications. The quality check and quantification of DNA was performed on Nanodrop 8000 spectrophotometry (Thermo Scientific) and Qubit 2.0 fluorometer (Invitrogen, Thermofisher Scientific, USA). The species was identified as *F. benghalensis* and *F. religiosa* by amplifying, sequencing and performing Nucleotide BLAST (Basic local Alignment Search Tool) of two marker genes i.e., Internal Transcribed Spacer (ITS) and Maturase K (MatK). The genomic library for linked reads was prepared on Chromium Controller instrument (10X Genomics, CA) and sequenced on Illumina Novaseq 6000 platform to produce 150 bp paired-end reads. The long read genomic library was prepared using SQK-LSK-109 kit followed by sequencing on MinION instrument (Oxford Nanopore Technologies, UK). The total RNA was extracted using TriZol reagent (Invitrogen, Thermofisher Scientific, USA). The transcriptomic library was prepared using Illumina TruSeq Stranded Total RNA library preparation kit with Ribo-Zero workflow and 150 bp paired-end reads were generated on Illumina Novaseq 6000 instrument (Illumina Inc., CA, USA). The

detailed information related to species identification and DNA-RNA extraction is mentioned in **Supplementary Notes 1.**

Genome assembly

Genome sizes of the two *Ficus* plant species were predicted using 10x Genomics linked reads. Barcode sequences were extracted from the raw reads using available proc10xG set of python scripts (<https://github.com/ucdavis-bioinformatics/proc10xG>), and these pre-processed linked reads were used to estimate the genome size using a k-mer distribution-based approach executed by SGA-preqc²⁰ (in paired-end mode). Genomic heterozygosity was also estimated for the two *Ficus* species in order to assess the complexity of the genomes. Jellyfish⁴⁷ v2.2.10 was used to generate k-mer count histograms using the barcode-filtered paired-end reads, and GenomeScope²² v2.0 was used to calculate the heterozygosity profiles.

The *de novo* genome assembly of both the *Ficus* species was performed following the same approach. Oxford Nanopore raw data was basecalled using Guppy v3.2.1 (Oxford Nanopore Technologies) and adapter removal of the raw reads were performed using Porechop v0.2.3 (Oxford Nanopore Technologies). Genome assembly using these pre-processed reads was performed using wtdbg v2.0.0⁴⁸ with default settings, SMARTdenovo (<https://github.com/ruanjue/smartdenovo>) with minimum read length of 0, Canu v2.0⁴⁹ with default settings (after reads correction using Canu v2.0) and Flye v2.4.2⁵⁰ with default settings. Quality statistics of all these assemblies were calculated using Quast v5.0.2⁵¹, and based on the assembled genome size and assembly statistics, the assembly derived from Flye v2.4.2 was selected for further analysis. 10x Genomics linked reads, which were processed using Longranger basic v2.2.0 (<https://support.10xgenomics.com/genome-exome/software/pipelines/latest/installation>), were used to polish this assembly three times using Pilon v1.23⁵². RNA-Seq data was quality-filtered using Trimmomatic⁵³ v0.38, and the filtered paired-end data was used for scaffolding using AGOUTI v0.3.3⁵⁴, to improve the assembly contiguity. The barcode-processed linked reads and pre-processed Nanopore long reads (Canu v2.0-corrected reads) were also used for further scaffolding of this assembly using ARCS v1.1.1⁵⁵ and LINKS v1.8.6⁵⁶ with default settings. Contigs with length of $\geq 1,000$ bp (after scaffolding) were extracted (assembly v0.1).

A total of 327.41 million (116.67x raw reads coverage) and 161.11 million (~56x raw reads coverage) barcoded 10x Genomics linked reads were used for *de novo* assembly of *F. benghalensis* genome using Supernova v2.1.1²¹. For *F. religiosa* genome assembly, a total of 324.53 million reads (125.43X raw

coverage) and 154.24 million reads (59X raw coverage) were used in Supernova v2.1.1. The assemblies generated using 327.41 million barcoded linked reads (116.67x raw coverage) for *F. benghalensis*, and 324.53 million 10x Genomics linked reads (125.43X raw coverage) for *F. religiosa* were considered for downstream analysis based on better assembly quality parameters. The mis-assembly regions were corrected using Tigrint v1.2.1⁵⁷ with the barcode-processed linked reads resulted from Longranger basic v2.2.0. Quality-filtered RNA-Seq paired-end reads were used to improve the contiguity of this assembly using AGOUTI v0.3.3⁵⁴. Barcode-processed linked reads, and error-corrected Nanopore long reads were used for further scaffolding using ARCS v1.1.1⁵⁵ (default settings) and LINKS v1.8.6⁵⁶, respectively. Contigs with length of $\geq 1,000$ bp (after scaffolding) were retained (assembly v0.2).

For each of the two species, assembly v0.1 and assembly v0.2 were merged using Quickmerge v0.3⁵⁸, which joins the gaps between the sequences in one assembly by using sequences from the other assembly, and thus improve the assembly contiguity (assembly v1). Further, contigs from assembly v0.2 were searched against the contigs from assembly v1 using BLASTN, and the unique sequences that did not match even with stringent parameters (**Supplementary Notes 2**) were directly added to assembly v1^{59,60}. Gap-closing of these assemblies (assembly v2) was performed in three iterations using LR_Gapcloser⁶¹ with error-corrected Oxford Nanopore long reads. Finally, barcode-processed 10x Genomics linked reads were used to fix any further local mis-assemblies, small indels, and base errors using Pilon v1.23⁵², to generate the final *F. benghalensis* and *F. religiosa* genome assemblies (assembly v3). Completeness of these genome assemblies was checked using embryophyta_odb10 single-copy orthologs dataset from BUSCO v5.2.1⁶². Also, to check the comprehensiveness of the final genome assembly, barcode-filtered 10x Genomics linked reads and adapter-processed Nanopore long reads were mapped to the final genome assemblies using BWA-MEM⁶³ v0.7.17, and the mapping percentage was calculated using SAMtools⁶⁴ v1.9 “flagstat”. The detailed information related to genome assembly and processing steps are mentioned in **Supplementary Figure 2 and Supplementary Notes 2**.

Sequence variation in the final *Ficus* genome assemblies was analyzed by mapping the barcode-filtered 10x Genomics linked reads to their respective whole genome assemblies (assembly v3) using BWA-MEM⁶³ v0.7.17, and SAMtools⁶⁴ v1.9. Variant calling was performed using BCFtools⁶⁵ v1.9 “mpileup” with the quality-filtration criteria - sequencing depth ≥ 30 , quality of variant sites ≥ 30 , and mapping quality ≥ 50 .

Genome annotation and Gene set construction

Genome annotation for *F. benghalensis* and *F. religiosa* species was performed using their corresponding final genome assemblies (assembly v3) generated in the previous step. For repeat annotation, a *de novo* repeat library was constructed for these genomes using RepeatModeler v2.0.1⁶⁶, which was further clustered to remove redundant sequences using CD-HIT-EST v4.8.1⁶⁷ with sequence identity of 90%, and seed size of 8 bp. This resultant repeat library was used to soft-mask the two *Ficus* genomes using RepeatMasker v4.1.0 (<http://www.repeatmasker.org>). Additionally, simple repeat sequences were also predicted in the final genome assemblies (assembly v3) using Tandem Repeats Finder (TRF) v4.09²³ (**Supplementary Notes 2**).

The repeat-masked genomes of *F. benghalensis* and *F. religiosa* were used for gene set construction. MAKER v2.31.10²⁴, the genome annotation pipeline was used to construct the coding gene sets using *ab initio* and evidence alignment-based approaches. Quality-filtered RNA-Seq reads (using Trimmomatic v0.38) of *F. benghalensis* from this study, and quality-filtered transcriptome reads of *F. religiosa* from this study and previous report⁶⁸ were used to construct *de novo* transcriptome assemblies using Trinity v2.9.1⁶⁹ with default parameters (**Supplementary Notes 2**). The resultant transcriptome assemblies, and the protein sequences of phylogenetically closer species - *Cannabis sativa*, *Malus domestica*, *Prunus avium*, and *Rosa chinensis* that belong to the same plant order Rosales (obtained from Ensembl Plants release 48) were used as EST and protein evidence for gene set construction of the two respective *Ficus* plant species. For *ab initio* gene finding, AUGUSTUS v3.2.3⁷⁰ and were used, and for empirical evidence alignments, BLAST⁷¹ and EXONERATE v2.2.0 (<https://github.com/nathanweeks/exonerate>) were used. The gene models were filtered based on their Annotation Edit Distance (AED) values and length-based criteria. Only the genes with AED values <0.5, and length (≥300 bp) were selected for the final gene set construction of *F. benghalensis* and *F. religiosa*, and termed as high-confidence gene sets⁷². Completeness of these gene sets was assessed using BUSCO v5.2.1⁶² with embryophyta_odb10 dataset of single-copy orthologous genes.

Further, *de novo* prediction of tRNA and rRNA genes was performed using tRNAscan-SE v2.0.7⁷³, and Barrnap v0.9 (<https://github.com/tseemann/barrnap>). Homology-based identification of miRNA gene sequences was performed using miRBase database⁷⁴ (**Supplementary Notes 2**).

PSMC analysis

Demographic history was constructed for *F. benghalensis* and *F. religiosa* plant species for comparative study using Pairwise Sequentially Markovian Coalescent (PSMC)²⁶ with the whole genome sequences of

the corresponding *Ficus* species. The barcode-filtered 10x Genomics linked reads of the two species, which were mapped to their respective whole genome assemblies (assembly v3) using BWA-MEM⁶³, were used to construct the consensus diploid sequences using SAMtools v1.9⁶⁴ “mpileup”, and BCFtools v1.9⁶⁵ in a manner that the minimum read depth is a third of the average read depth, and the maximum read depth is set to the twice of the average read depth. The consensus sequences were filtered for sites with quality score <20. PSMC analysis was performed using the parameters “N30 -t5 -r5 -p4+25*2+4+6” with 100 bootstrap values. Generation time of 7 years, and a mutation rate per site per generation of 1.75e-08 was considered for this analysis, which was used for another *Ficus* species *F. carica* in a previous study²⁷.

Phylogenetic analysis

Protein sequences of all 45 Eudicot species, available on Ensembl⁷⁵ plants release 48, were used to resolve the phylogenetic position of *F. benghalensis* and *F. religiosa*. The selected Eudicot species were – *Actinidia chinensis*, *Arabidopsis halleri*, *Arabidopsis lyrata*, *Arabidopsis thaliana*, *Arabis alpina*, *Beta vulgaris*, *Brassica napus*, *Brassica oleracea*, *Brassica rapa*, *Camelina sativa*, *Cannabis sativa* female, *Capsicum annuum*, *Citrullus lanatus*, *Citrus clementina*, *Coffea canephora*, *Corchorus capsularis*, *Cucumis melo*, *Cucumis sativus*, *Cynara cardunculus*, *Daucus carota*, *Glycine max*, *Gossypium raimondii*, *Helianthus annuus*, *Ipomoea triloba*, *Lupinus angustifolius*, *Malus domestica* Golden, *Manihot esculenta*, *Medicago truncatula*, *Nicotiana attenuata*, *Olea europaea* var. *sylvestris*, *Phaseolus vulgaris*, *Pistacia vera*, *Populus trichocarpa*, *Prunus avium*, *Prunus persica*, *Prunus dulcis*, *Rosa chinensis*, *Solanum lycopersicum*, *Solanum tuberosum*, *Theobroma cacao* Belizian Criollo B97-61/B2, *Theobroma cacao* Matina 1-6, *Trifolium pratense*, *Vigna angularis*, *Vigna radiata*, and *Vitis vinifera*. Proteome files of these 45 Eudicot species along with MAKER-derived protein sequences of *F. benghalensis* and *F. religiosa* were used for phylogenetic analysis. *Zea mays*, which is a monocot plant, was used as an outgroup species, and protein sequences of *Zea mays* were also obtained from Ensembl plants release 48.

Proteome files of all species were used to extract the longest isoform for each protein. The resultant protein sequences were used to construct the orthologous gene sets using OrthoFinder v2.4.1⁷⁶. From these orthogroups, the fuzzy one-to-one orthogroups that contained proteins from all 48 species were extracted using KinFin v1.0⁷⁷. For further analysis, these fuzzy one-to-one orthologous gene sets were processed to extract and consider only the longest protein sequence for each species. Each of these orthogroups was individually aligned using MAFFT v7.467⁷⁸. These multiple sequence alignments were

filtered to remove the empty sites and were concatenated using BeforePhylo v0.9.0 (<https://github.com/qiyunzhu/BeforePhylo>). RAXML v8.2.12⁷⁹, which uses rapid hill climbing algorithm, was utilized to construct the maximum likelihood-based species phylogenetic tree with the concatenated protein sequence alignment. 'PROTGAMMAAUTO' amino acid substitution model and a bootstrap value of 100 were used for phylogenetic tree construction.

Evolution of gene families

CAFÉ v4.2.1⁸⁰ was used to estimate the evolution of gene families across all 48 selected species (*Zea mays* as an outgroup species) using the proteome files containing the longest isoform for each protein, and the species phylogenetic tree generated in the previous step. Prior to the analysis, the species phylogenetic tree (with *Zea mays* as an outgroup species) was converted to ultrametric tree based on the calibration point of 121 Million years between *F. benghalensis*, and *Beta vulgaris*, obtained from TimeTree database⁸¹. All-versus-All BLASTP search for the protein sequences of all 48 species was performed, which was followed by clustering using MCL v14.137⁸², and removal of gene families that contained genes in <2 species of the specified clades and ≥ 100 gene copies for at least one species. These filtered gene families, and the ultrametric species tree was utilized to analyze the expansion or contraction of gene families using two-lambda (λ) model, where λ is a random birth-death parameter. In this analysis, *F. benghalensis*, *F. religiosa*, and their phylogenetically closer species – *Cannabis sativa*, *Rosa chinensis*, *Malus domestica*, *Prunus avium*, *Prunus persica*, and *Prunus dulcis* were given the same λ -value, while the other species were given another λ -value. The considerably expanded (with ≥ 10 expanded genes) and contracted (with ≥ 5 contracted genes) gene families for *F. benghalensis* and *F. religiosa* were extracted and compared.

Further, gene family expansion/contraction for the enzymes involved in mevalonate and shikimate pathway that play important role in *Ficus* plants⁷ were analyzed using CAFÉ results. EC numbers of these enzymes were used to extract their sequence (specific to *Arabidopsis thaliana*, a model Eudicot plant) from Swiss-Prot database⁸³, and the sequences were queried against the proteome of *F. benghalensis* and *F. religiosa* using BLASTP (e-value of 10^{-5}) to identify the enzymes in these two *Ficus* species. Gene families of these enzymes were identified from CAFÉ results, and the expansion/contraction of these families was identified in two *Ficus* species compared to their immediate ancestor phylogenetic node.

Synten analysis

Collinear blocks between *F. benghalensis* and *F. religiosa* species (inter-species), and intra-species syntenic blocks for these two species were identified using the All-versus-All BLASTP (e-value of 10^{-5}) homology alignments of the respective protein sequences and the GFF annotations. Inter-species and intra-species synteny analysis was performed using MCScanX²⁵, 10 genes were required to detect a collinear block⁸⁴. Also, the duplicated genes in each of the *Ficus* species were detected and classified into various gene duplication categories (proximal, tandem, segmental, or dispersed) using the same parameter.

Identification of Evolutionary signatures

For identifying the genes with evolutionary signatures, only the Eudicot species from the phylogenetically closer plant orders of the two *Ficus* species (one species from each order) were considered to avoid erroneous results and to analyze adaptive evolution of these two *Ficus* species across dicot plant orders. The selected 14 Eudicot species were - *Arabidopsis thaliana* (order Brassicales), *Actinidia chinensis* (order Ericales), *Coffea canephora* (order Gentianales), *Cucumis melo* (order Cucurbitales), *Daucus carota* (order Apiales), *Glycine max* (order Fabales), *Gossypium raimondii* (order Malvales), *Helianthus annuus* (order Asterales), *Olea europaea* var. *sylvestris* (order Lamiales), *Pistacia vera* (order Sapindales), *Populus trichocarpa* (order Malpighiales), *Rosa chinensis* (order Rosales), *Solanum tuberosum* (order Solanales), and *Vitis vinifera* (order Vitales). Proteome files of these 14 species along with that of *F. benghalensis* and *F. religiosa* were used to construct the orthologous gene sets across 16 species using OrthoFinder v2.4.1⁷⁶. Orthogroups containing protein sequences from all 16 species were identified, and only the longest protein sequence for each species was extracted from these orthogroups.

Genes with unique amino acids substitution

Each of the orthogroups across 16 species constructed in the previous step was aligned using MAFFT v7.467⁷⁸. In the multiple sequence alignments, *F. benghalensis* genes showing amino acid positions that were identical in other species but different in this species were considered as genes with unique amino acid substitutions in *F. benghalensis*. Similarly, *F. religiosa* genes that showed different amino acids in the sequence alignments compared to the other species were also considered as the *F. religiosa* genes with unique substitutions. Any gap in these alignments and ten sites around the gap were not considered in this analysis. The genes showing unique amino acid substitutions were extracted, and

impact of these unique amino acid substitutions on the proteins function was predicted using Sorting Intolerant From Tolerant (SIFT)⁸⁵ with UniProt reference database⁸⁶.

Genes with higher nucleotide divergence

The protein sequence alignments of the orthogroups across 16 species generated in the previous step were used to construct individual maximum likelihood-based phylogenetic tree for each orthogroup using RAxML v8.2.12⁷⁹ with 'PROTGAMMAAUTO' amino acid substitution model and 100 bootstrap values. Root-to-tip branch length distances were estimated for genes from all selected species in these phylogenetic trees using "adephylo"⁸⁷ package in R. *F. benghalensis* and *F. religiosa* specific genes with relatively higher branch length distance values compared to the other species were extracted separately as the genes with higher rate of evolution.

Positively selected genes

The nucleotide sequences of all the orthologous gene sets constructed across 16 species were individually aligned using MAFFT v7.467⁷⁸. Further, RAxML⁷⁹ v8.2.12 was used to construct the species phylogenetic tree for these selected 16 species in a similar way as mentioned in the earlier steps (see Phylogenetic analysis step). This species phylogenetic tree and the nucleotide alignments for the orthogroups in phylip format were used to separately detect the *F. benghalensis* and *F. religiosa* genes with positive selection using a branch-site model utilized by "codeml" program available in PAML v4.9a⁸⁸. Likelihood-ratio tests were performed from the obtained log-likelihood values and genes qualifying against the null model having FDR-corrected p-values of <0.05 (obtained from chi-square analysis) were termed as positively selected genes. Additionally, Bayes Empirical Bayes (BEB) analysis was carried out to identify the codon sites of *F. benghalensis* and *F. religiosa* genes that showed positive selection with >95% probability for the foreground lineage.

The genes that showed more than one of these above mentioned evolutionary signatures, namely - positive selection, higher nucleotide divergence rate, and unique substitution with functional impact, were extracted and considered as the genes with multiple signatures of adaptive evolution (MSA genes)^{89,90}.

Functional annotation

The complete high-confidence gene sets of *F. benghalensis* and *F. religiosa* was annotated against NCBI Non-Redundant (nr) database using BLASTP (e-value of 10^{-5}), Swiss-Prot⁸³ database using BLASTP (e-

value of 10^{-5}), and Pfam-A v32.0⁹¹ database using HMMER v3.3⁹² (e-value of 10^{-5}). The species-specific gene sets with evolutionary signatures were assigned to KEGG Orthology (KO) categories, and KEGG pathways using KAAS⁹³ v2.1 genome annotation server. These genes were also annotated to different COG categories using eggNOG-mapper v2⁹⁴. Gene Ontology (GO) categories were also assigned to these genes using WebGeStalt⁹⁵ web server, and in the over-representation analysis, only the GO categories with p-values <0.05 were retained. Further, functional annotations of these gene sets were screened, and curated manually. The MSA genes were also checked for protein-protein interaction data in STRING v11.0 database⁹⁶ using information from previously reported studies, and the interacting genes were analyzed. Enzymes involved in shikimate pathway and Eukaryotic mevalonate pathway were searched for their presence in the gene sets showing evolutionary signatures, and in the expanded/contracted gene families.

ACKNOWLEDGEMENTS

AC, SM, and MSB thank Council of Scientific and Industrial Research (CSIR) for fellowship. The authors also thank the NGS facility at IISER Bhopal and the intramural research funds provided by IISER Bhopal.

AUTHORS' CONTRIBUTIONS

VKS conceived and coordinated the project. SM performed DNA-RNA extraction, prepared the samples for sequencing, performed the Nanopore sequencing, and the species identification assays. AC and VKS designed the computational framework of the study. AC performed all the computational analysis presented in the study, performed the functional annotation of gene sets, and constructed the figures. AC, MSB and VKS analyzed the data and interpreted the results. AC, SM and VKS wrote the manuscript. All the authors have read and approved the final version of the manuscript.

COMPETING INTERESTS

The authors declare no competing financial and non-financial interest.

REFERENCES

1. Singh, D., Singh, B. & Goel, R. K. Traditional uses, phytochemistry and pharmacology of *Ficus religiosa*: A review. *Journal of Ethnopharmacology* (2011) doi:10.1016/j.jep.2011.01.046.
2. Harrison, R. D. Figs and the diversity of tropical rainforests. *BioScience* (2005) doi:10.1641/0006-3568(2005)055[1053:FATDOT]2.0.CO;2.
3. Christenhusz, M. J. M. & Byng, J. W. The number of known plants species in the world and its annual increase. *Phytotaxa* (2016) doi:10.11646/phytotaxa.261.3.1.
4. Shi, Y. *et al.* The genus *Ficus* (Moraceae) used in diet: Its plant diversity, distribution, traditional uses and ethnopharmacological importance. *Journal of Ethnopharmacology* (2018) doi:10.1016/j.jep.2018.07.027.
5. Shanahan, M., Samson, S. O., Compton, S. G. & Corlett, R. Fig-eating by vertebrate frugivores: A global review. *Biological Reviews of the Cambridge Philosophical Society* (2001) doi:10.1017/S1464793101005760.
6. Cottee-Jones, H. E. W., Bajpai, O., Chaudhary, L. B. & Whittaker, R. J. The Importance of *Ficus* (Moraceae) Trees for Tropical Forest Restoration. *Biotropica* **48**, 413–419 (2016).
7. Zhang, X. *et al.* Genomes of the Banyan Tree and Pollinator Wasp Provide Insights into Fig-Wasp Coevolution. *Cell* (2020) doi:10.1016/j.cell.2020.09.043.
8. Gopukumar, S. T. & Praseetha, P. K. *Ficus benghalensis* Linn – the sacred Indian medicinal tree with potent pharmacological remedies. *International Journal of Pharmaceutical Sciences Review and Research* (2015).
9. Thimmamma Marrimanu – Edurudona, India - Atlas Obscura. <https://www.atlasobscura.com/places/thimmamma-marrimanu>.
10. A. Moustafa, A. A Threatened Introduced Species (*Ficus benghalensis* L.) in Ismailia, Egypt . in *Modern Fruit Industry* (2020). doi:10.5772/intechopen.88164.
11. Devanesan, E. B., Anand, A. V., Kumar, P. S., Vinayagamoorthy, P. & Basavaraju, P. Phytochemistry and pharmacology of *ficus religiosa*. *Systematic Reviews in Pharmacy* (2018) doi:10.5530/srp.2018.1.9.
12. Vidyapeeth, J. A Review on *Ficus religiosa* (Sacred Fig). **6**, 901–906 (2019).
13. Sitaramam, V., Jog, S. R. & Tetali, P. Ecology of *Ficus religiosa* accounts for its association with religion. *Curr. Sci.* (2009).
14. GALIL, J. *Ficus religiosa* L.–the tree-splitter. *Bot. J. Linn. Soc.* (1984) doi:10.1111/j.1095-8339.1984.tb01570.x.
15. Kumar Makhija, I., Sharma, I. P. & Khamar, D. Phytochemistry and Pharmacological properties of *Ficus religiosa*: an overview. *Sch. Res. Libr. Ann. Biol. Res.* (2010).
16. Chaudhary, A. S. *et al.* Docking-based screening of *Ficus religiosa* phytochemicals as inhibitors of

623 human histamine H2 receptor. *Pharmacogn. Mag.* (2017) doi:10.4103/pm.pm_49_17.

624 17. Pellicer, J. & Leitch, I. J. The Plant DNA C-values database (release 7.1): an updated online
625 repository of plant genome size data for comparative studies. *New Phytologist* (2020)
626 doi:10.1111/nph.16261.

627 18. Shirasawa, K. *et al.* The Ficus erecta genome aids Ceratocystis canker resistance breeding in
628 common fig (F. carica). *Plant J.* (2020) doi:10.1111/tpj.14703.

629 19. Mori, K. *et al.* Identification of RAN1 orthologue associated with sex determination through
630 whole genome sequencing analysis in fig (Ficus carica L.). *Sci. Reports* 2017 71 7, 1–15 (2017).

631 20. Simpson, J. T. & Durbin, R. Efficient de novo assembly of large genomes using compressed data
632 structures. *Genome Res.* (2012) doi:10.1101/gr.126953.111.

633 21. Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M. & Jaffe, D. B. Direct determination of diploid
634 genome sequences. *Genome Res.* (2017) doi:10.1101/gr.214874.116.

635 22. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for
636 reference-free profiling of polyploid genomes. *Nat. Commun.* (2020) doi:10.1038/s41467-020-
637 14998-3.

638 23. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.*
639 (1999) doi:10.1093/nar/27.2.573.

640 24. Campbell, M. S., Holt, C., Moore, B. & Yandell, M. Genome Annotation and Curation Using
641 MAKER and MAKER-P. *Curr. Protoc. Bioinforma.* (2014) doi:10.1002/0471250953.bi0411s48.

642 25. Wang, Y. *et al.* MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and
643 collinearity. *Nucleic Acids Res.* (2012) doi:10.1093/nar/gkr1293.

644 26. Li, H. & Durbin, R. Inference of human population history from individual whole-genome
645 sequences. *Nature* (2011) doi:10.1038/nature10231.

646 27. Bharatraj, A. *et al.* The genome sequence of Mesua ferrea and comparative demographic
647 histories of forest trees. *Gene* 145214 (2020) doi:10.1016/j.gene.2020.145214.

648 28. Soltis, D. E. *et al.* Angiosperm phylogeny inferred from 18S rDNA, rbcL, and atpB sequences. *Bot.*
649 *J. Linn. Soc.* (2000) doi:10.1006/bojl.2000.0380.

650 29. Lee, J. Y. *et al.* Recruitment of CRABS CLAW to promote nectary development within the eudicot
651 clade. *Development* (2005) doi:10.1242/dev.02067.

652 30. Mahajan, S., Chakraborty, A., Sil, T. & Sharma, V. K. Genome sequencing and assembly of
653 Tinospora cordifolia (Giloy) plant. *bioRxiv* 2021.08.02.454741 (2021)
654 doi:10.1101/2021.08.02.454741.

655 31. Doležel, J. & Bartoš, J. Plant DNA flow cytometry and estimation of nuclear genome size. in
656 *Annals of Botany* (2005). doi:10.1093/aob/mci005.

657 32. Plomion, C. *et al.* Oak genome reveals facets of long lifespan. *Nat. Plants* 2018 47 4, 440–452
658 (2018).

- 659 33. Wang, G. *et al.* Genomic evidence of prevalent hybridization throughout the evolutionary history
660 of the fig-wasp pollination mutualism. *Nat. Commun.* 2021 121 **12**, 1–14 (2021).
- 661 34. Usai, G. *et al.* Epigenetic patterns within the haplotype phased fig (*Ficus carica* L.) genome. *Plant*
662 *J.* **102**, 600–614 (2020).
- 663 35. Verbitsky, M. Y., Crucifix, M. & Volobuev, D. M. A theory of Pleistocene glacial rhythmicity. *Earth*
664 *Syst. Dyn.* (2018) doi:10.5194/esd-9-1025-2018.
- 665 36. Clark, P. U. *et al.* The middle Pleistocene transition: characteristics, mechanisms, and implications
666 for long-term changes in atmospheric pCO₂. *Quat. Sci. Rev.* (2006)
667 doi:10.1016/j.quascirev.2006.07.008.
- 668 37. Dupont, L. M., Donner, B., Schneider, R. & Wefer, G. Mid-Pleistocene environmental change in
669 tropical Africa began as early as 1.05 Ma. *Geology* (2001) doi:10.1130/0091-
670 7613(2001)029<0195:MPECIT>2.0.CO;2.
- 671 38. Kasahara, H. *et al.* Contribution of the mevalonate and methylerythritol phosphate pathways to
672 the biosynthesis of Gibberellins in *Arabidopsis*. *J. Biol. Chem.* (2002)
673 doi:10.1074/jbc.M208659200.
- 674 39. Venkateshwaran, M. *et al.* A role for the mevalonate pathway in early plant symbiotic signaling.
675 *Proc. Natl. Acad. Sci. U. S. A.* (2015) doi:10.1073/pnas.1413762112.
- 676 40. Francenia Santos-Sánchez, N., Salas-Coronado, R., Hernández-Carlos, B. & Villanueva-Cañongo, C.
677 Shikimic Acid Pathway in Biosynthesis of Phenolic Compounds. in *Plant Physiological Aspects of*
678 *Phenolic Compounds* (2019). doi:10.5772/intechopen.83815.
- 679 41. Wang, Q., Quan, S. & Xiao, H. Towards efficient terpenoid biosynthesis: manipulating IPP and
680 DMAPP supply. *Bioresources and Bioprocessing* (2019) doi:10.1186/s40643-019-0242-z.
- 681 42. Martin, S. L. *et al.* The occurrence of crassulacean acid metabolism in epiphytic ferns, with an
682 emphasis on the vittariaceae. *Int. J. Plant Sci.* **166**, 623–630 (2005).
- 683 43. Lüttge, U. Ability of crassulacean acid metabolism plants to overcome interacting stresses in
684 tropical environments. *AoB Plants* **2010**, 1–15 (2010).
- 685 44. Tobias, P. A. & Guest, D. I. Tree immunity: growing old without antibodies. *Trends Plant Sci.* **19**,
686 367–370 (2014).
- 687 45. Isah, T. Stress and defense responses in plant secondary metabolites production. *Biological*
688 *research* (2019) doi:10.1186/s40659-019-0246-3.
- 689 46. Chakraborty, A., Mahajan, S., Jaiswal, S. K. & Sharma, V. K. Genome sequencing of turmeric
690 provides evolutionary insights into its medicinal properties. *bioRxiv* 2020.09.07.286245 (2020)
691 doi:10.1101/2020.09.07.286245.
- 692 47. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of
693 occurrences of k-mers. *Bioinformatics* (2011) doi:10.1093/bioinformatics/btr011.
- 694 48. Ruan, J. & Li, H. Fast and accurate long-read assembly with wtdbg2. *Nat. Methods* (2020)
695 doi:10.1038/s41592-019-0669-3.

- 696 49. Koren, S. *et al.* Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and
697 repeat separation. *Genome Res.* (2017) doi:10.1101/gr.215087.116.
- 698 50. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using
699 repeat graphs. *Nat. Biotechnol.* (2019) doi:10.1038/s41587-019-0072-8.
- 700 51. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUASt: Quality assessment tool for genome
701 assemblies. *Bioinformatics* (2013) doi:10.1093/bioinformatics/btt086.
- 702 52. Walker, B. J. *et al.* Pilon: An integrated tool for comprehensive microbial variant detection and
703 genome assembly improvement. *PLoS One* (2014) doi:10.1371/journal.pone.0112963.
- 704 53. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence
705 data. *Bioinformatics* (2014) doi:10.1093/bioinformatics/btu170.
- 706 54. Zhang, S. V., Zhuo, L. & Hahn, M. W. AGOUTI: Improving genome assembly and annotation using
707 transcriptome data. *Gigascience* (2016) doi:10.1186/s13742-016-0136-3.
- 708 55. Yeo, S., Coombe, L., Warren, R. L., Chu, J. & Birol, I. ARCS: Scaffolding genome drafts with linked
709 reads. *Bioinformatics* (2018) doi:10.1093/bioinformatics/btx675.
- 710 56. Warren, R. L. *et al.* LINKS: Scalable, alignment-free scaffolding of draft genomes with long reads.
711 *Gigascience* (2015) doi:10.1186/s13742-015-0076-3.
- 712 57. Jackman, S. D. *et al.* Tigrint: Correcting assembly errors using linked reads from large molecules.
713 *BMC Bioinformatics* (2018) doi:10.1186/s12859-018-2425-6.
- 714 58. Chakraborty, M., Baldwin-Brown, J. G., Long, A. D. & Emerson, J. J. Contiguous and accurate de
715 novo assembly of metazoan genomes with modest long read coverage. *Nucleic Acids Res.* (2016)
716 doi:10.1093/nar/gkw654.
- 717 59. Schmidt, H. *et al.* A high-quality genome assembly from short and long reads for the non-biting
718 midge chironomus riparius (diptera). *G3 Genes, Genomes, Genet.* (2020)
719 doi:10.1534/G3.119.400710.
- 720 60. Jaiswal, S. K., Mahajan, S., Chakraborty, A., Kumar, S. & Sharma, V. K. The genome sequence of
721 Aloe vera reveals adaptive evolution of drought tolerance mechanisms. *iScience* (2021)
722 doi:10.1016/j.isci.2021.102079.
- 723 61. Xu, G. C. *et al.* LR-Gapcloser: A tiling path-based gap closer that uses long reads to complete
724 genome assembly. *Gigascience* (2018) doi:10.1093/gigascience/giy157.
- 725 62. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO:
726 Assessing genome assembly and annotation completeness with single-copy orthologs.
727 *Bioinformatics* (2015) doi:10.1093/bioinformatics/btv351.
- 728 63. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. **00**, 1–3
729 (2013).
- 730 64. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* (2009)
731 doi:10.1093/bioinformatics/btp352.

- 732 65. Narasimhan, V. *et al.* BCFtools/RoH: A hidden Markov model approach for detecting autozygosity
733 from next-generation sequencing data. *Bioinformatics* (2016)
734 doi:10.1093/bioinformatics/btw044.
- 735 66. Flynn, J. M. *et al.* RepeatModeler2 for automated genomic discovery of transposable element
736 families. *Proc. Natl. Acad. Sci. U. S. A.* (2020) doi:10.1073/pnas.1921046117.
- 737 67. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: Accelerated for clustering the next-generation
738 sequencing data. *Bioinformatics* (2012) doi:10.1093/bioinformatics/bts565.
- 739 68. Matasci, N. *et al.* Data access for the 1,000 Plants (1KP) project. *GigaScience* (2014)
740 doi:10.1186/2047-217X-3-17.
- 741 69. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity
742 platform for reference generation and analysis. *Nat. Protoc.* (2013) doi:10.1038/nprot.2013.084.
- 743 70. Stanke, M. *et al.* AUGUSTUS: A b initio prediction of alternative transcripts. *Nucleic Acids Res.*
744 (2006) doi:10.1093/nar/gkl200.
- 745 71. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool.
746 *J. Mol. Biol.* (1990) doi:10.1016/S0022-2836(05)80360-2.
- 747 72. Zhang, L. *et al.* A high-quality apple genome assembly reveals the association of a
748 retrotransposon and red fruit colour. *Nat. Commun.* (2019) doi:10.1038/s41467-019-09518-x.
- 749 73. Chan, P. P. & Lowe, T. M. tRNAscan-SE: Searching for tRNA genes in genomic sequences. in
750 *Methods in Molecular Biology* (2019). doi:10.1007/978-1-4939-9173-0_1.
- 751 74. Griffiths-Jones, S., Saini, H. K., Van Dongen, S. & Enright, A. J. miRBase: Tools for microRNA
752 genomics. *Nucleic Acids Res.* (2008) doi:10.1093/nar/gkm952.
- 753 75. Bolser, D., Staines, D. M., Pritchard, E. & Kersey, P. Ensembl plants: Integrating tools for
754 visualizing, mining, and analyzing plant genomics data. in *Methods in Molecular Biology* (2016).
755 doi:10.1007/978-1-4939-3167-5_6.
- 756 76. Emms, D. M. & Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative
757 genomics. *Genome Biol.* (2019) doi:10.1186/s13059-019-1832-y.
- 758 77. Laetsch, D. R. & Blaxter, M. L. KinFin: Software for taxon-aware analysis of clustered protein
759 sequences. *G3 Genes, Genomes, Genet.* (2017) doi:10.1534/g3.117.300233.
- 760 78. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7:
761 Improvements in performance and usability. *Mol. Biol. Evol.* (2013) doi:10.1093/molbev/mst010.
- 762 79. Stamatakis, A. RAXML version 8: A tool for phylogenetic analysis and post-analysis of large
763 phylogenies. *Bioinformatics* (2014) doi:10.1093/bioinformatics/btu033.
- 764 80. De Bie, T., Cristianini, N., Demuth, J. P. & Hahn, M. W. CAFE: A computational tool for the study
765 of gene family evolution. *Bioinformatics* (2006) doi:10.1093/bioinformatics/btl097.
- 766 81. Hedges, S. B., Dudley, J. & Kumar, S. TimeTree: A public knowledge-base of divergence times
767 among organisms. *Bioinformatics* (2006) doi:10.1093/bioinformatics/btl505.

- 768 82. Van Dongen, S. & Abreu-Goodger, C. Using MCL to extract clusters from networks. *Methods Mol.*
769 *Biol.* (2012) doi:10.1007/978-1-61779-361-5_15.
- 770 83. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence database and its supplement
771 TrEMBL in 2000. *Nucleic Acids Research* (2000) doi:10.1093/nar/28.1.45.
- 772 84. Bi, Q. *et al.* Pseudomolecule-level assembly of the Chinese oil tree yellowhorn (*Xanthoceras*
773 *sorbifolium*) genome. *Gigascience* (2019) doi:10.1093/gigascience/giz070.
- 774 85. Ng, P. C. & Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic*
775 *Acids Res.* (2003) doi:10.1093/nar/gkg509.
- 776 86. Bateman, A. *et al.* UniProt: A hub for protein information. *Nucleic Acids Res.* (2015)
777 doi:10.1093/nar/gku989.
- 778 87. Jombart, T. & Dray, S. Adephylo: Exploratory Analyses for the Phylogenetic Comparative Method.
779 *Bioinformatics* (2010) doi:10.1093/bioinformatics/btq292.
- 780 88. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* (2007)
781 doi:10.1093/molbev/msm088.
- 782 89. Jaiswal, S. K. *et al.* Genome Sequence of Peacock Reveals the Peculiar Case of a Glittering Bird.
783 *Front. Genet.* (2018) doi:10.3389/fgene.2018.00392.
- 784 90. Mittal, P., Jaiswal, S. K., Vijay, N., Saxena, R. & Sharma, V. K. Comparative analysis of corrected
785 tiger genome provides clues to its neuronal evolution. *Sci. Rep.* (2019) doi:10.1038/s41598-019-
786 54838-z.
- 787 91. Bateman, A. The Pfam protein families database. *Nucleic Acids Res.* (2004)
788 doi:10.1093/nar/gkh121.
- 789 92. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: Interactive sequence similarity
790 searching. *Nucleic Acids Res.* (2011) doi:10.1093/nar/gkr367.
- 791 93. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C. & Kanehisa, M. KAAS: An automatic genome
792 annotation and pathway reconstruction server. *Nucleic Acids Res.* (2007)
793 doi:10.1093/nar/gkm321.
- 794 94. Huerta-Cepas, J. *et al.* Fast genome-wide functional annotation through orthology assignment by
795 eggNOG-mapper. *Mol. Biol. Evol.* (2017) doi:10.1093/molbev/msx148.
- 796 95. Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z. & Zhang, B. WebGestalt 2019: gene set analysis toolkit with
797 revamped UIs and APIs. *Nucleic Acids Res.* (2019) doi:10.1093/nar/gkz401.
- 798 96. Szklarczyk, D. *et al.* The STRING database in 2017: Quality-controlled protein-protein association
799 networks, made broadly accessible. *Nucleic Acids Res.* (2017) doi:10.1093/nar/gkw937.

800