



OPEN

Genome sequencing and functional analysis of a multipurpose medicinal herb *Tinospora cordifolia* (Giloy)

Shruti Mahajan, Abhisek Chakraborty, Manohar S. Bisht, Titas Sil & Vineet K. Sharma

Tinospora cordifolia (Willd.) Hook.f. & Thomson, also known as Giloy, is among the most important medicinal plants that have numerous therapeutic applications in human health due to the production of a diverse array of secondary metabolites. To gain genomic insights into the medicinal properties of *T. cordifolia*, the genome sequencing was carried out using 10× Genomics linked read and Nanopore long-read technologies. The draft genome assembly of *T. cordifolia* was comprised of 1.01 Gbp, which is the genome sequenced from the plant family Menispermaceae. We also performed the genome size estimation for *T. cordifolia*, which was found to be 1.13 Gbp. The deep sequencing of transcriptome from the leaf tissue was also performed. The genome and transcriptome assemblies were used to construct the gene set, resulting in 17,245 coding gene sequences. Further, the phylogenetic position of *T. cordifolia* was also positioned as basal eudicot by constructing a genome-wide phylogenetic tree using multiple species. Further, a comprehensive comparative evolutionary analysis of gene families contraction/expansion and multiple signatures of adaptive evolution was performed. The genes involved in benzyl iso-quinoline alkaloid, terpenoid, lignin and flavonoid biosynthesis pathways were found with signatures of adaptive evolution. These evolutionary adaptations in genes provide genomic insights into the presence of diverse medicinal properties of this plant. The genes involved in the common symbiosis signalling pathway associated with endosymbiosis (Arbuscular Mycorrhiza) were found to be adaptively evolved. The genes involved in adventitious root formation, peroxisome biogenesis, biosynthesis of phytohormones, and tolerance against abiotic and biotic stresses were also found to be adaptively evolved in *T. cordifolia*.

Abbreviations

SARS-COV-2	Severe Acute Respiratory Syndrome Coronavirus 2
COVID-19	Coronavirus disease 19
BLAST	Basic local alignment search tool
ITS	Internal transcribed spacer
MatK	Maturase K
LTR	Long terminal repeat
PEX	Peroxisome
CAM	Crassulacean acid metabolism
MYA	Million years ago
CI	Confidence interval

Tinospora cordifolia is a climbing shrub belonging to the Menispermaceae family that includes more than 400 plant species¹. The present scientific name of this plant is *Tinospora cordifolia* (Willd.) Hook.f. & Thomson according to the 'USDA-ARS GRIN Taxonomy', whereas its scientific name as per the 'International Plant Names Index' is *Tinospora cordifolia*. Thus, we have referred to it as *Tinospora cordifolia* in the subsequent text. This plant has gained tremendous therapeutic interest and significance during the COVID-19 pandemic due to its enormous therapeutic potential². It perhaps originated in Africa in the Oligocene epoch (28.57 million years ago) and was spread to Asia in the early Miocene epoch (21.54 million years ago)³. It is found in tropical and

MetaBioSys Group, Department of Biological Sciences, Indian Institute of Science Education and Research Bhopal, Bhopal, Madhya Pradesh 462066, India. email: vineetks@iiserb.ac.in

sub-tropical regions including India, China, Sri Lanka, Bangladesh, Myanmar, Thailand, Malaysia, etc., and is also known as ‘Giloy’, ‘Amrita’, ‘Guduchi’, and ‘heart leaved moonseed’¹. It is a perennial deciduous dioecious plant with morphological characteristics of twining branches, succulent stem with papery bark, alternatively arranged heart-shaped leaves, aerial roots, and greenish-yellow tiny flowers in the form of racemes inflorescence^{1,4}. Being a climber, it needs a supportive plant like *Jatropha curcas* (Jatropha), *Azadirachta indica* (Neem), *Moringa oleifera* (Moringa), etc., for its growth⁴. These supportive plants also play an important role in enhancing the production of various secondary metabolites of *T. cordifolia*^{4,5}.

T. cordifolia produces an array of secondary metabolites in response to stress conditions, and their concentration also varies based on seasons and its dioecy nature⁶. The chemical constituents of this plant are broadly categorized as alkaloids, terpenoids, phenolics, polysaccharides, steroids, essential oils, aliphatic compounds, etc., which can be obtained from various parts of the plant^{7,8}. A study reported that among the two species of *Tinospora* (i.e. *T. cordifolia* and *T. sinensis*), *T. cordifolia* produces three times higher concentration of an alkaloid, berberine, than *T. sinensis*, and thus the former is preferred in therapeutics⁹. Its bioactive compounds have known biological properties such as anti-oxidant, antipyretic, anti-diabetic, anti-inflammatory, anti-microbial, anti-viral, anti-arthritis, anti-osteoporotic, anti-HIV, anti-cancer, hepatoprotective, anti-malarial, immunomodulation, etc.^{10,11}. These properties make this species useful in the traditional treatment of several ailments, including fevers, cough, diabetes, general debility, ear pains, jaundice, asthma, heart diseases, burning sensation, bone fracture, urinary problems, chronic diarrhoea, dysentery, leucorrhoea, skin diseases, cancer, helminthiasis, leprosy, and rheumatoid arthritis^{12,13}. Further pre-clinical and clinical studies have been carried out to indicate its potential to treat leucopenia induced by breast cancer chemotherapy, hepatic disorders, post-menopausal syndrome, obstructive jaundice, etc.¹².

These diverse and important therapeutic applications of *T. cordifolia*, due to the presence of secondary metabolites such as alkaloids, terpenoids, lignans, flavonoids, etc., make it a species of broad interest⁸. Further the alkaloids, namely berberine, columbamine, magnoflorine, and other compounds from this plant have proven immunomodulatory activity¹⁴. The compounds like berberine, cordifoliside C, magnoflorine, tembetarine, cholesterol, melezitose, allopyranose, etc., are reported to possess antimicrobial activity against various bacterial and fungal species^{15–20}. These phytochemicals are reported to be effective against chikungunya, dengue, and also against SARS-CoV2 viruses, due to which this herb gained enormous attention and popularity during the COVID-19 pandemic^{21–27}. Ten compounds from *T. cordifolia* were predicted as drug candidates against COVID-19 and also found inactive towards the central nervous system, which is essential for a good drug candidate²⁸. Not only in humans, it also reduced the mortality rate in chickens infected with the infectious bursal disease virus²⁹.

The bioactive compounds in medicinal plants are not only of plant origin but are also produced by the endophytic microorganisms present in them. Plants are associated with a large number of symbiotic microorganisms like endophytic bacteria and fungi. These endophytic microorganisms play roles in the growth, development, and stress tolerance of associated plants^{30,31}. One of such microorganisms is arbuscular mycorrhiza fungi (AMF) that forms an association (endosymbiosis) of fungi with the roots of higher plants. AMF have branched structures called arbuscules that facilitate the exchange of minerals between plants and fungi and also improve the nutrient uptake of plants³². AMF are known to have impact on secondary metabolite production by increasing the plant biomass or stimulating the biosynthetic pathways for secondary metabolite production. This endosymbiosis occurs in plants through a pathway known as the common symbiosis signalling pathway (CSSP). They have been reported to enhance the production of secondary metabolites, such as alkaloids, terpenoids, phenolics, saponins, etc., in various medicinal plants³³. The presence of endophytic bacteria and fungi has previously been reported in the leaves, stems, petiole, and roots of this plant^{34–37}. These endophytic microorganisms thus potentially contribute to the medicinal properties of this multifaceted plant.

However, despite the widely known medicinal properties of this plant, its genome assembly was still unavailable. A preliminary study reported the transcriptome (482 Mb data) of this species from leaf and stem tissues using 454 GS-FLX pyrosequencing³⁸. A recent karyological study reported $2n = 26$ as the chromosome number in *T. cordifolia*, which was also supported by earlier studies^{39–41}. Thus, to uncover the genomic basis of its medicinal properties and to further explore its therapeutic potential, we carried out the genome sequencing and assembly of *T. cordifolia* using 10× Genomics linked reads and Oxford Nanopore long reads. This is the draft genome assembly of *T. cordifolia*, which is also the genome sequenced so far from the medicinally important genus *Tinospora* and its family⁴². We also carried out a comprehensive deep sequencing and assembly of the leaf transcriptome using Illumina technology. The genome-wide phylogenetic analysis was also carried out for *T. cordifolia* to determine its phylogenetic position. A comprehensive evolutionary comparative analysis including identification of gene families expansion/contraction and signatures of adaptive evolution was performed to gain genomic insights into its medicinal properties.

Results

Sampling and sequencing of *T. cordifolia* genome and transcriptome

The species was confirmed as *T. cordifolia* with the highest identity (98.32% for ITS and 99.41% for *matK*) with sequences of *T. cordifolia* available at NCBI (nt) database. The total genomics data generated was 102 Gb which comprised 79.4 Gb of 10× genomics linked read and 22.6 Gb of Nanopore read data (Fig. 1a and Supplementary Table S1). Additionally, 34.7 Gb of transcriptomic data from the leaf tissue was generated and used with genomic data for the genome assembly and analysis of *T. cordifolia* (Fig. 1a and Supplementary Table S2). Since the genome size was not available for this plant, the genome size for *T. cordifolia* was computationally estimated to be 1.13 Gbp. Considering this genome size, the sequenced genomic data provided 90.2× genome coverage. Further, the *T. cordifolia* genome was found to have high heterozygosity (1.55%) (Fig. 1a).

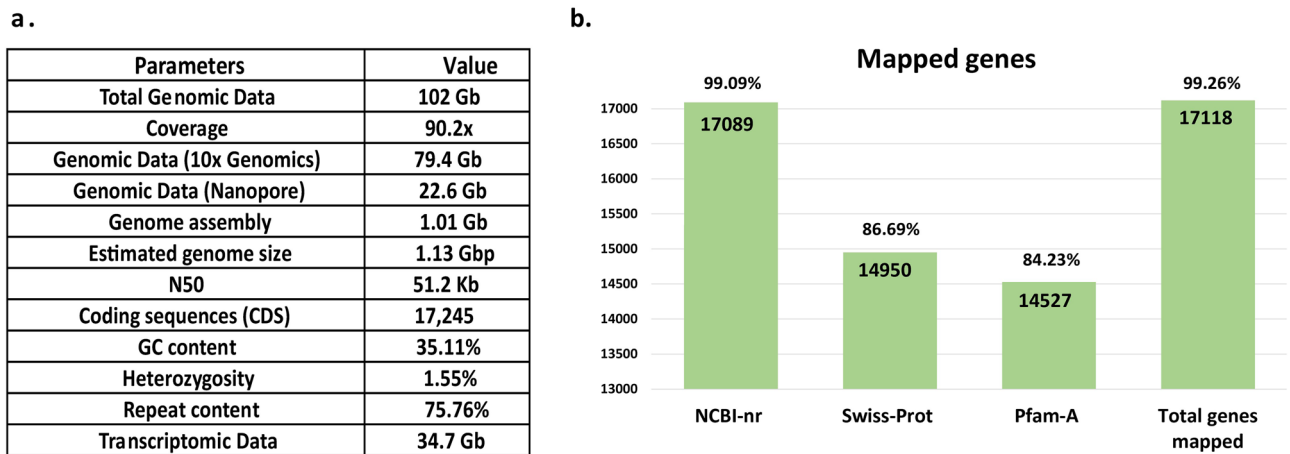


Figure 1. Summary of *T. cordifolia* genome assembly and genes. (a) Summary of generated data and genome assembly (b) Statistical summary of mapped genes of *T. cordifolia*.

After scaffolding, misassembly rectification, gap-filling, and polishing, the final draft genome assembly of *T. cordifolia* resulted in a 1.01 Gbp assembly size. The assembled genome consisted of 35.11% GC content, 55,140 scaffolds, and N50 of 51.2 Kb (Fig. 1a and Supplementary Table S3). The BUSCO completeness was 83% in the final polished *T. cordifolia* genome assembly, and the complete plus fragmented BUSCO score was 94.1%, and the missing BUSCOs was 5.9% (Supplementary Table S4). Similarly, the BUSCO completeness was 60.2% in the final polished *T. cordifolia* genome annotation, and the complete plus fragmented BUSCO score was 80.9%, and the missing BUSCOs was 19.1% (Supplementary Table S4). The de novo transcriptome assembly resulted in a 2,764,154 bp size, and 8208 transcripts were predicted in the assembled transcriptome.

Annotation of genome and gene set formation

The final polished genome assembly consisted of 1875 repeat families, which were further clustered into 1557 representative repeat family sequences. *T. cordifolia* genome was predicted with 75.75% of repetitive sequences. Among these repetitive sequences, 70.73% were characterized as interspersed repeats comprising 53.29% retroelements (48.39% of LTR repeats), 2.29% DNA transposons, and 15.14% unclassified repeats. The LTR repeats comprised 39.48% Gypsy/DIRS1 and 6.84% Ty1/Copia elements (Supplementary Table S5). Among the non-coding RNAs, 394 hairpin miRNAs, 1306 rRNAs, and 2164 tRNAs were also predicted.

A total of 19,264 coding sequences were obtained in the genome assembly of *T. cordifolia*. Among these coding genes, 17,245 genes were present in the high-confidence coding gene set after filtering. Overall, 17,118 genes (99.26%) could be mapped against publicly available databases (Fig. 1b). The functional annotations of these high-confidence coding genes are provided in Supplementary Table S6–S8. A large fraction of these genes was involved in KEGG pathways like ribosome, spliceosome, protein processing in endoplasmic reticulum, etc., and COG categories like signal transduction mechanisms, post translational modification, protein turnover, chaperones, carbohydrate transport and metabolism, transcription, etc.

Phylogenetic analysis

Among the identified 152,176 orthogroups, 224 fuzzy one-to-one orthogroups were predicted across 29 selected plant species. The filtered 224 fuzzy one-to-one orthogroups contained 201,201 alignment positions and were used to construct a phylogenetic tree. The position of *T. cordifolia* in the phylogenetic tree was found close to *Papaver somniferum* with estimated divergence time of ~122 MYA (95% CI), which also concurs with the species divergence time obtained from the TimeTree database⁴³. *T. cordifolia* and *P. somniferum* form a separate clade from all the core eudicots and monocot species because the clade of *T. cordifolia* is basal eudicots that diverged very early from the core eudicots (145 MYA with 95% CI)^{44,45} (Fig. 2 and Supplementary Figure S1).

Gene family contraction and expansion analysis

The CAFÉ analysis resulted in 5169 contracted and 782 expanded gene families in *T. cordifolia*. Among these gene families, 12 were highly contracted, and 25 were highly expanded in *T. cordifolia*. The highly contracted genes families were involved in phenylpropanoid biosynthesis, mitochondrial biogenesis, protein processing in endoplasmic reticulum, MAPK signalling pathway, and plant-pathogen interaction (Supplementary Table S9). The highly expanded gene families were majorly involved in phenylpropanoid biosynthesis, zeatin biosynthesis, carbohydrate metabolism, xenobiotic degradation, membrane trafficking, and MAPK signalling pathway (Supplementary Table S10).

Orthologous gene clustering

The phylogenetic tree constructed using the peptide sequences of six Ranunculales species showed that *T. cordifolia* is closer to Ranunculaceae members i.e., *T. thalictroides* and *C. chinensis* compared to the other considered

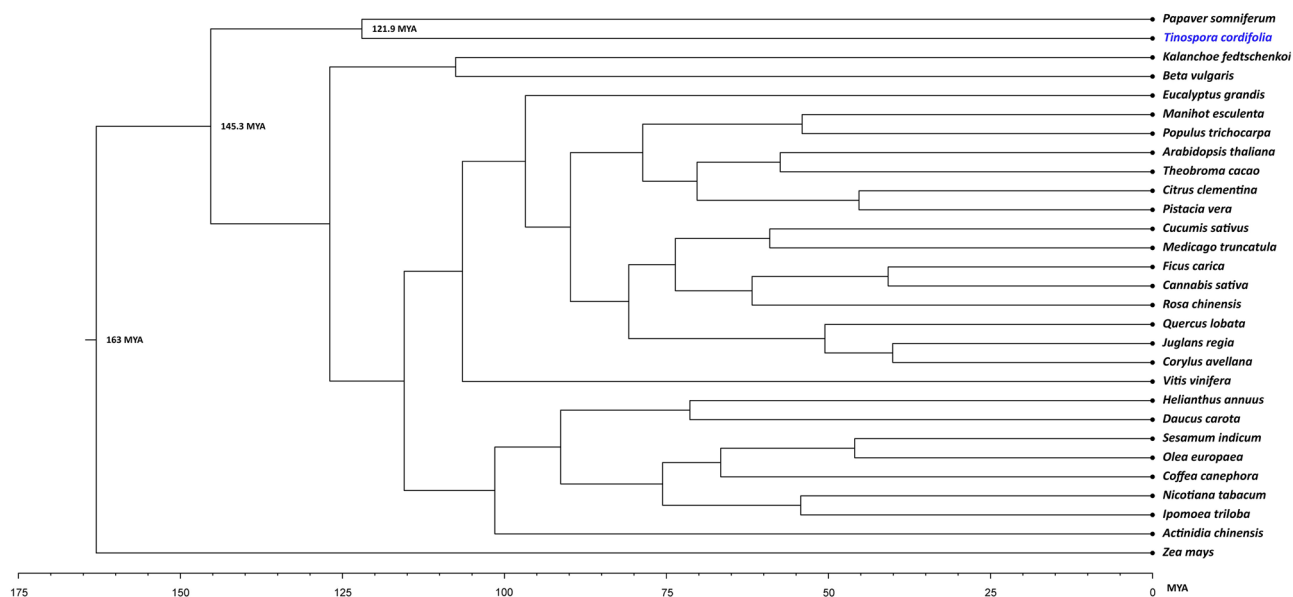


Figure 2. Phylogenetic tree showing the position of *T. cordifolia* with other species. The phylogenetic tree includes 26 core eudicot species, two basal eudicot species (*T. cordifolia* and *Papaver somniferum*) and a monocot species.

species (Fig. 3a). Among all the orthologous gene set, 66.9% (7321) gene set were shared across all six species, indicating their conservation. 2.7% (304) gene set were specific to *T. cordifolia* (Fig. 3b). The GO functions of these species-specific genes are mentioned in Fig. 3c,d, and e.

Genes with signatures of adaptive evolution

The comparative evolutionary analysis of the six selected species of order Ranunculales revealed 724 genes with high rate of evolution, 3432 genes with unique amino acid substitutions with functional impact and 3566 genes with positive selection in *T. cordifolia* (Supplementary Table S11 and Supplementary Table S12). Among these genes, 290 showed all three signatures of adaptive evolution in *T. cordifolia* (Supplementary Figure S2 and Supplementary Table S13).

Benzyl iso-quinoline alkaloid (BIA) biosynthesis pathway in *T. cordifolia*

Alkaloids are the main bioactive compounds of Menispermaceae and Ranunculales plants. The genes involved in the BIA biosynthesis pathway were observed in *T. cordifolia* with transcriptomic evidences (Fig. 5, Supplementary Figure S3, and Supplementary Table S14). BIAs such as berberine and magnoflorine were also reported to have pharmaceutical activities against the COVID-19 virus^{14,23}. Along with these, palmatine, tembetarine, and jatrorrhizine are the other pharmaceutically active biomolecules in *T. cordifolia*⁷. Among the genes involved in the biosynthesis of BIAs in *T. cordifolia*, two genes, Tyrosine decarboxylase (*TYDC*) and Canadine synthase (*CAS*), showed multiple signatures of adaptive evolution. Additionally, tyrosine aminotransferase (*TAT*) showed unique amino acid substitutions in *T. cordifolia*. The analysis of orthologous genes found seven genes (*4OMT*, *CNMT*, *CoOMT*, *RNMT*, *CST*, *NOMT*, and *SOMT*) with bacterial orthologs and three genes (*CNMT*, *RNMT*, and *TYDC*) with fungal orthologs (Supplementary Figures S4–18). *CAS* gene had orthologs from bryophytes and gymnosperms. *RNMT* had the highest number of bacterial and fungal orthologs among all the other BIA biosynthesis genes. Further, all genes except *CST* had orthologs from monocots.

Other secondary metabolites biosynthesis pathway

Apart from BIA biosynthesis pathway, genes involved in terpenoid backbone biosynthesis, lignin biosynthesis, and flavonoid biosynthesis were found in *T. cordifolia* genome with transcriptomic evidences (Supplementary Figure S3). Among the 19 genes involved in terpenoid backbone biosynthesis, eight genes showed at least one signature of adaptive evolution. Out of the eight genes, *DXS*, *AACT* and *FPPS* showed MSA, *PMK*, and *FOLK* showed positive selection, *ispE* and *GGPS1* showed unique amino acid substitutions, and *ispH* showed high nucleotide divergence (Fig. 5, Supplementary Table S15, and Supplementary Figure S3).

Genes *PAL*, *C4H*, and *4CL*, involved in both flavonoid and lignin biosynthesis pathways, showed signatures of adaptive evolution in *T. cordifolia*. Along with these three genes, five other genes of flavonoid biosynthesis showed at least one signature of adaptive evolution. Among these eight genes, *PAL*, *4CL*, *DFR*, and *ANR* showed multiple signatures of adaptive evolution. *UFGT* showed only positive selection, and *C4H*, *F3H* and *FLS* showed only unique amino acid substitutions. Among the 13 genes involved in lignin biosynthesis, eight genes showed at least one signature of adaptive evolution. *PAL*, *4CL*, *C3H*, *CCR*, and *CAD* showed multiple signatures of

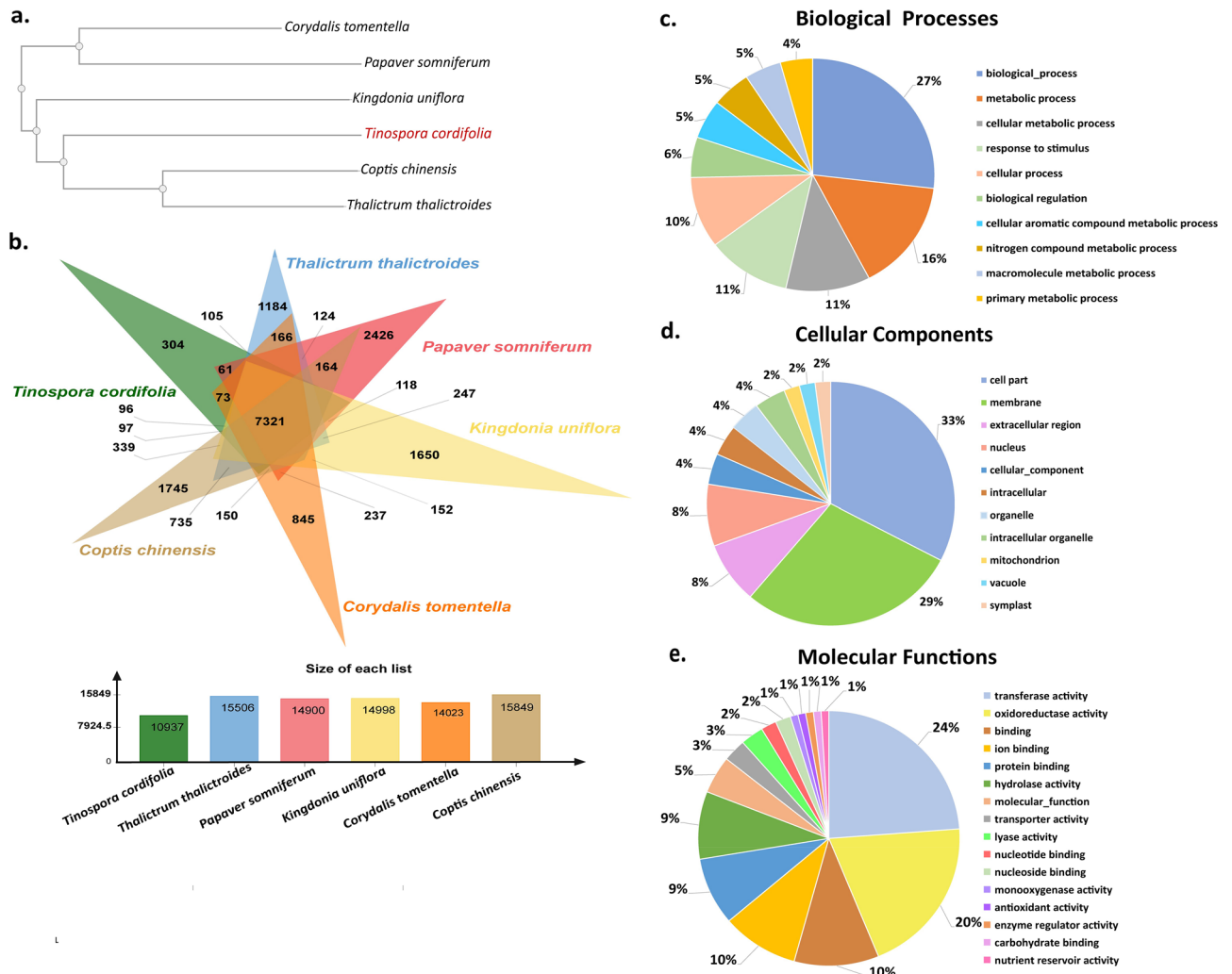


Figure 3. Comparative analysis of *T. cordifolia* genome. (a) Phylogeny of six Ranunculales species including *T. cordifolia* (b) Venn diagram of ortholog gene clustering of six species (c) Biological processes of *T. cordifolia* specific genes (d) Cellular component of *T. cordifolia* specific genes (e) Molecular functions of *T. cordifolia* specific genes Figs. 3a and 3b are obtained from Orthovenn3¹³⁸.

adaptive evolution. *HCT* and *CCoAOMT* showed only positive selection, and *C4H* showed only unique amino acid substitutions (Supplementary Table S15 and Supplementary Figure S3).

Common symbiosis signaling pathway (CSSP) in *T. cordifolia*

Several plant parts of *T. cordifolia* have been associated with endosymbiotic microorganisms^{35,37,46}. Therefore, we traced the genes involved in CSSP and found signatures of adaptive evolution in these genes. The angiosperm-specific conserved genes (*Castor*, *EPP1*, *kinF*, *LIN*, *SYN*, and *VAPYRIN*) responsible for plant endosymbiosis were present in *T. cordifolia*. However, all the land plant-specific conserved genes (*CCaMK*, *RAD*, *STR*, *STR2*, and *SymRK*) for plant endosymbiosis except *CYCLOPS* were also found in *T. cordifolia*⁴⁷. Among these conserved genes, *SYN* showed positive selection in *T. cordifolia*. Twenty-four genes showed signatures of adaptive evolution, of which *NUP133*, *NENA*, *MCA8*, *D3*, *DIS*, *MSBP1*, *Sk11*, *SUNN*, *RFCb*, and *CCD7* showed multiple signatures of adaptive evolution. Among these 24 genes, *LYK3*, *CNCG15*, *SYP132A*, *MIG1*, *KIN3* and *SUT2* showed only positive selection, and *NSP1*, *D14*, *FatM*, *LOM1*, *VAMP721d*, *HA1*, *EXO70I*, and *VTI12* showed only unique amino acid substitutions (Fig. 6 and Supplementary Table S16).

Adventitious root associated genes in *T. cordifolia*

Adventitious roots are aerial roots that emerge from the plant shoot system and function as storage, support, etc.⁴⁸. *T. cordifolia*, being a climbing plant, bears assimilatory roots (a type of adventitious roots) that help plants anchorage on other trees or structures, and perform photosynthesis^{49,50}. 25 genes associated with adventitious root development in plants showed signatures of adaptive evolution in *T. cordifolia* (Supplementary Table S17). Among these 25 genes, a few genes also showed evidence of their gene expression in *T. cordifolia* (Supplementary Figure S3). Out of the 25 genes, 17 genes showed MSA, and eight showed only one signature of adaptive evolution.

Among these genes, 10 were involved in auxin biosynthesis, transport, and response. Thirteen genes that showed signatures of adaptive evolution were involved in the formation of meristem and primordium. Genes *PRP1* and *CEP1*, involved in cell wall modification, showed two and three signatures of adaptive evolution, respectively. Additionally, an auxin-responsive gene *SAUR15* was highly expanded in *T. cordifolia* (Supplementary Table S17). Additionally, three miRNAs associated with adventitious root formation were present in *T. cordifolia*.

Peroxisome associated genes in *T. cordifolia*

Peroxisomes play key roles in several processes crucial to plant growth and development. It plays an essential role in fatty acid β -oxidation, phytohormones biosynthesis (auxin, salicylic acid and jasmonic acid), photorespiration, transportation of metabolites and cofactors, etc.⁵¹. Fourteen peroxisome-associated genes were found with at least one signature of adaptive evolution and a few genes among them also showed gene expression in *T. cordifolia* (Supplementary Table S18 and Supplementary Figure S3). All 14 genes showed unique amino acid substitutions except solute carrier family 25 member 17 (*PMP34*) and *PEX-19*, which showed positive selection. Among these 14 genes, *ALD* (member 3) and catalase showed only unique amino acid substitutions in *T. cordifolia*. Except these three genes, all other genes (11) showed MSA. Among these 11 MSA genes, five genes (*PEX5*, *PEX7*, *PEX14*, acyl-CoA oxidase, and L-pipecolate oxidase) showed all three signatures of adaptive evolution. Four genes, long-chain acyl-CoA synthetase, *PEX-3*, and two superoxide dismutases, showed only two signatures of adaptive evolution. Among these 14 genes, *PEX3*, *PEX5*, *PEX7*, *PEX14*, *PEX19*, *MPV17*, and *PMP34* were involved in peroxisome biogenesis (Fig. 7 and Supplementary Table S18).

Plant growth, phytohormones and stress tolerance associated genes in *T. cordifolia*

76 genes with all three signatures of adaptive evolution were associated with plant growth and development involved in leaf development, root development, fruit ripening, tissue development, pollen germination, flowering regulation, seed development, cell wall modification, etc. Among these, ten genes were associated with root development in *T. cordifolia*. 28 genes that were associated with phytohormones such as auxin, gibberellin, abscisic acid, ethylene, jasmonic acid, salicylic acid, cytokinin, and brassinosteroids showed all three signatures of adaptive evolution. Among these, nine genes were associated with auxin, and eight genes were associated with abscisic acid. Among plant stress tolerance-associated genes, 49 and 34 genes were involved in abiotic and biotic stress tolerance, respectively. These 49 abiotic stress tolerance genes were involved in light/heat tolerance, cold tolerance, salt tolerance, osmotic stress tolerance and oxidative stress regulation and tolerance (Supplementary Table S19).

Discussion

Tinospora cordifolia is known to produce several phytochemicals as secondary metabolites that are responsible for its various medicinal properties^{11,52}. Being a medicinally important herb with therapeutic applications in multiple health conditions, the genome and transcriptome sequencing of *T. cordifolia* was much needed to gain genomic insights into the secondary metabolites production responsible for its medicinal properties. This study reports the draft genome assembly of *T. cordifolia* which is also the genome from the *Tinospora* genus and its family. This study also deliberates the genome assembly, transcriptome assembly, gene set, and phylogeny of *T. cordifolia* that will help in understanding the medicinal properties of this multipurpose herbal plant. The genome assembly was constructed using a hybrid approach that has emerged as a promising methodology to significantly increase the contiguity in the genome assembly by increasing scaffold N50 and decreasing the number of scaffolds compared to short-read technology. This hybrid approach helped in the sequencing and assembly of this complex genome of *T. cordifolia* with high heterozygosity and high repetitive content⁵³. Further, this observed high heterozygosity of *T. cordifolia* could also be associated with its dioecious nature as reported in earlier studies^{54–56}.

T. cordifolia belongs to the order Ranunculales and our phylogenetic tree finds its position as a distinct branch separate from all the considered core eudicots and the monocot (Fig. 2). This could be due to the early divergence of order Ranunculales from all the other core eudicots. Order Ranunculales is regarded as an early diverging eudicot order and is among a few other eudicot orders (collectively known as basal eudicots) that are found to be sister lineage to the core eudicots, which was also observed in the case of *T. cordifolia* that showed early divergence from all other dicot species^{44,57,58}. According to the phylogenetic tree, *T. cordifolia* was found in a separate clade as basal eudicot with *P. somniferum* (122 MYA) diverged from core eudicots (145 MYA), which was also supported by other studies (Fig. 2)^{59–64}.

Alkaloids, terpenoids, lignans, and flavonoids are the main secondary metabolites in *T. cordifolia*. Among the alkaloids, benzyl iso-quinoline alkaloid compounds like berberine, palmatine, magnoflorine, and tembetarine are the key bioactive compounds in *T. cordifolia*. Among the BIAs, berberine and magnoflorine have been reported to have immunomodulatory, antibacterial, and anti-viral activities^{14,16,65}. Among a few tested compounds from *T. cordifolia*, berberine was reported to have the highest affinity against SARS-CoV-2 targets⁶⁶. It also inhibited DNA and protein synthesis in bacteria⁶⁵. Another BIA, tembetarine, possesses anti-diabetic and antibacterial activities, whereas columbamine has immunomodulatory activity^{16,67}. Additionally, its terpenoids and flavonoids are reported to possess antidiabetic activity, whereas lignans are reported to have anti-oxidant activity in *T. cordifolia*^{8,11}. These alkaloids, terpenoids, lignans, flavonoids along with various other phytochemicals, thus contribute to the diverse medicinal properties of *T. cordifolia*. Therefore, we traced the genes involved in the biosynthesis pathways of benzyl iso-quinoline alkaloids, terpenoids, lignans, and flavonoids in *T. cordifolia*, and found signatures of adaptive evolution along with transcriptomic evidences in a few of these genes (Figs. 4, 5 and Supplementary Figure S3). The genes with signatures of adaptive evolution involved in BIA (*TAT*, *TYDC* and *CAS*) and terpenoids (*DXS*, *AACT* and *FPPS*) are also reported in other plant species^{64,68–71}. The lignin and flavonoid biosynthesis genes with evolutionary signatures (*PAL*, *4CL*, *C3H*, *CCR*, and *CAD*) are also reported to

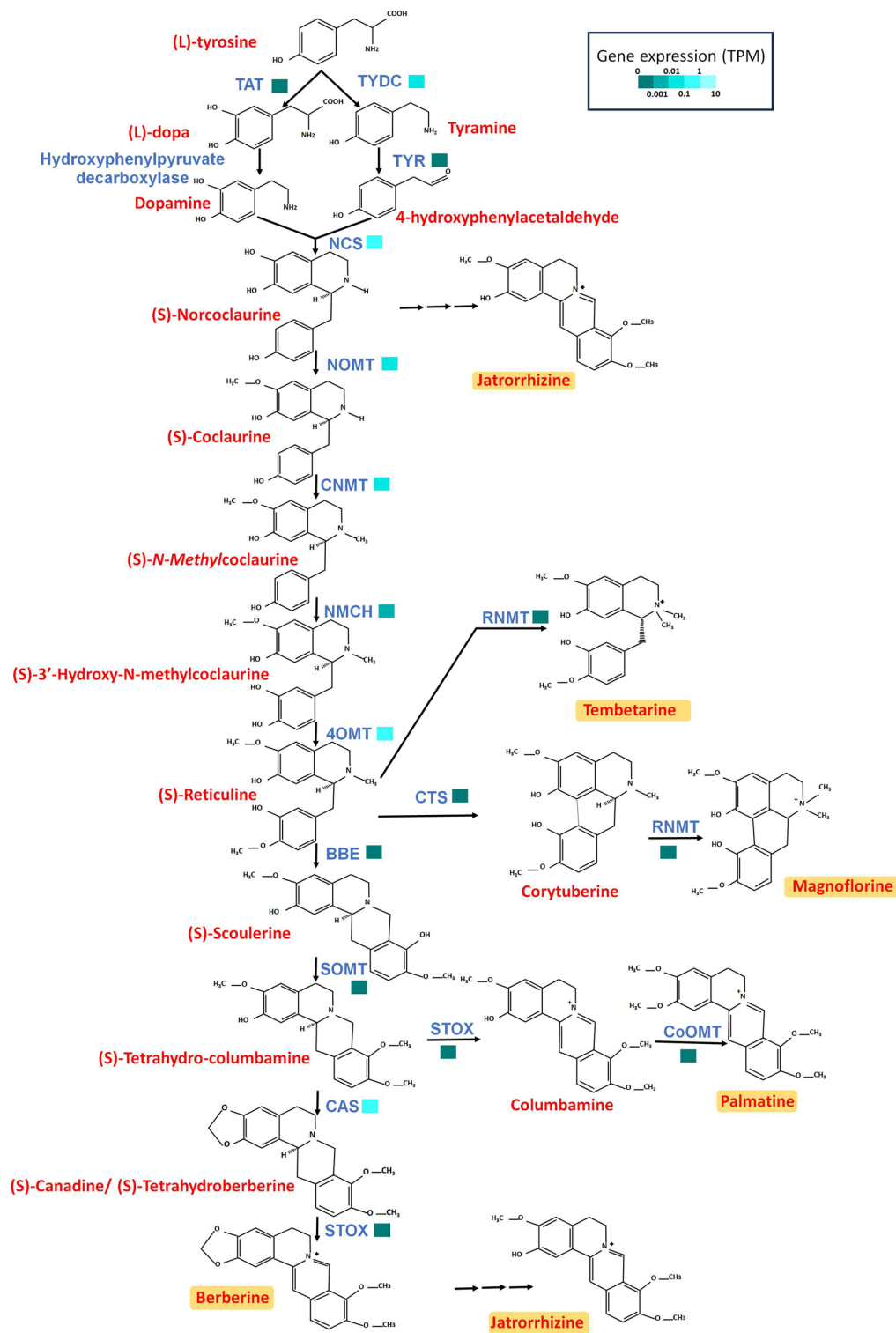


Figure 4. Benzyl iso-quinoline alkaloid biosynthesis pathway in *T. cordifolia*. This pathway indicates the genes involved in biosynthesis of alkaloids like berberine, magnoflorine, palmatine, and tembetarine. It includes tyrosine aminotransferase (*TAT*), Hydroxyphenylpyruvate decarboxylase, Tyrosine decarboxylase (*TYDC*), Tyrosinase/Phenolase (*TYR/3OHase*), (S)-Norcoclaurine synthase (*NCS*), Norcoclaurine 6-O-methyltransferase (*NoMT*), (S)-Coclaurine N-methyltransferase (*CNMT*), N-Methylcoclaurine 3'-hydroxylase (*NMCH*), 3'-Hydroxy-N-methylcoclaurine 4'-O-methyltransferase (*4OMT*), Corytuberine synthase (*CTS*), (S)-Corytuberine-N-methyltransferase (*SCNMT/RNMT*), Berberine bridge enzyme (*BBE*), S-Scoulerine 9-O-methyltransferase (*SOMT/SMT*), (S)-Tetrahydroprotoberberine oxidase (*STOX*), Columbamine O-methyltransferase (*CoOMT*), and (S)-Canadine synthase (*CAS*).

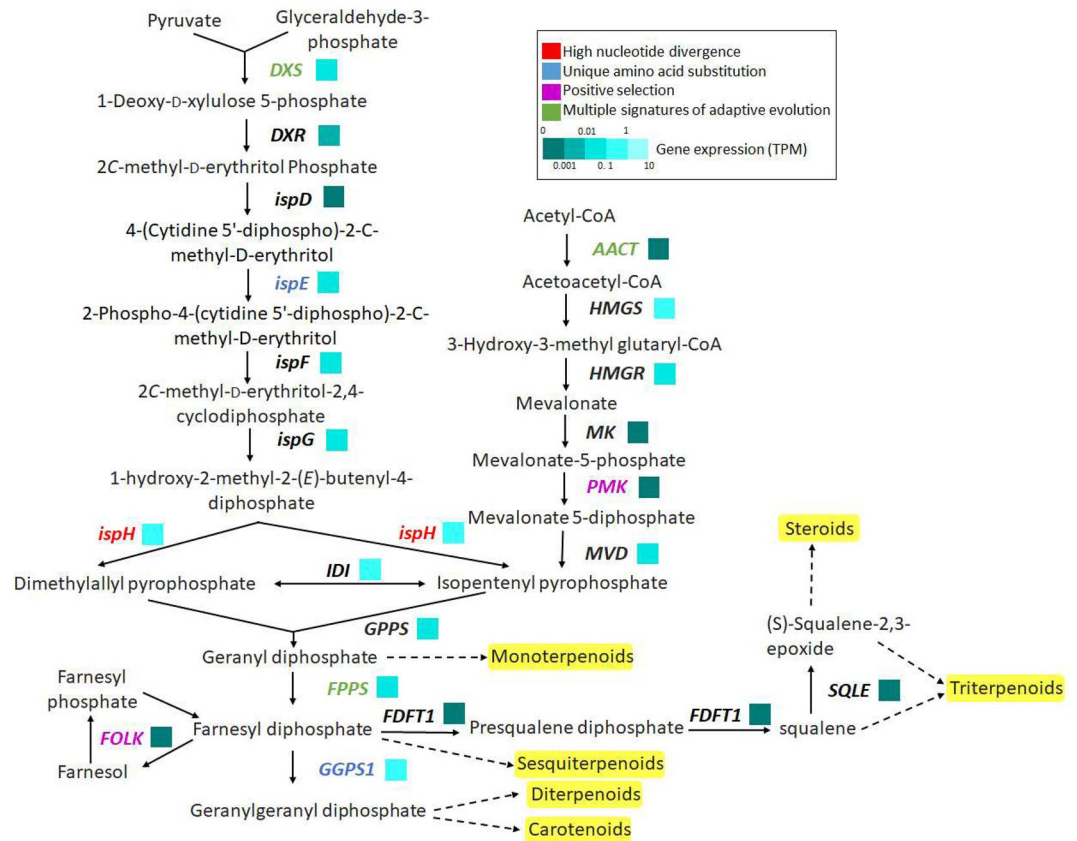


Figure 5. Terpenoid biosynthesis pathway in *T. cordifolia*. This pathway indicates the genes involved in biosynthesis of terpenoids in *T. cordifolia*. It includes 1-deoxy-D-xylulose-5-phosphate synthase (*DXS*), 1-deoxy-D-xylulose-5-phosphate reductoisomerase (*DXR*), 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase (*ispD*), 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (*ispE*), 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (*ispF*), (E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (*ispG*), 4-hydroxy 3-methylbut-2-en-1-yl diphosphate reductase (*ispH*), Acetoacetyl-CoA thiolase (*AACT*), HMG-CoA synthase (*HMGS*), HMG-CoA reductase (*HMGR*), Mevalonate kinase (*MK*), Phosphomevalonate kinase (*PMK*), Mevalonate-5-diphosphate decarboxylase (*MDD*), Isopentenyl diphosphate isomerase (*IDI*), Geranyl diphosphate synthase (*GPPS*), Farnesyl diphosphate synthase (*FPPS*), Geranylgeranyl diphosphate synthase (*GGPS1*), Farnesol kinase (*FOLK*), Farnesyl-diphosphate farnesyltransferase 1 (*FDFT1*), and Squalene monooxygenase (*SQLE*). Genes in red, blue, pink and green represents genes showing high nucleotide divergence, unique amino acid substitutions, positive selection, and multiple signatures of adaptive evolution, respectively.

have significance in other plant genomes^{72–75}. The observed evolutionary signatures in genes involved in these pathways might contribute to the production of diverse secondary metabolites in *T. cordifolia*.

The presence of endophytic fungi in plants helps in the availability of essential mineral nutrients, which improve plant growth via phytohormone production, tolerance against stressed conditions like drought, salinity, heavy metals, high or low temperatures, etc.^{33,76,77}. However, among the 12 conserved Arbuscular Mycorrhiza (AM) endosymbiosis-associated genes in plants, only one gene *CYCLOPS* is found to be absent in *T. cordifolia*, which was similar to other AM endosymbiosis-associated plant species such as *Fragaria x ananassa*, *Castanea mollissima* and *Quercus robur*⁴⁷. Most genes mentioned above are directly involved in CSSP, which is responsible for AM endosymbiosis in plants. The genes involved in this pathway showed signatures of adaptive evolution in *T. cordifolia*, which indicates towards the evolution of AM symbiosis (Fig. 6 and Supplementary Table S16). These genes with evolutionary signatures were also reported in other plant species and associated with endosymbiosis^{78,79}. The AM endosymbiosis is also influenced by phytohormones like auxin, cytokinin, jasmonic acid, brassinosteroids, etc.^{80,81}. Thus, the genes associated with these phytohormones also showed signatures of adaptive evolution which further suggests the evolution of AM symbiosis in *T. cordifolia* (Supplementary Table S13). Moreover, endophytic bacteria and fungi present in *T. cordifolia* are *Bacillus* spp., *Aneurinibacillus* spp., *Pseudomonas* spp., *Penicillium* spp., *Nigrospora oryzae*, *Chaetomium globosum*, *Colletotrichum* spp., *Curvularia* spp., *Alternaria alternata*, *Phome* spp., etc.^{36,37}. These endophytes are known to produce secondary metabolites, and phytohormones like indole acetic acid and gibberellins that helps in the growth and stress tolerance of their host plants^{82–91}. Thus, evolutionary signatures in genes associated with common symbiosis signalling pathway might contribute to the evolution of medicinal properties in *T. cordifolia*.

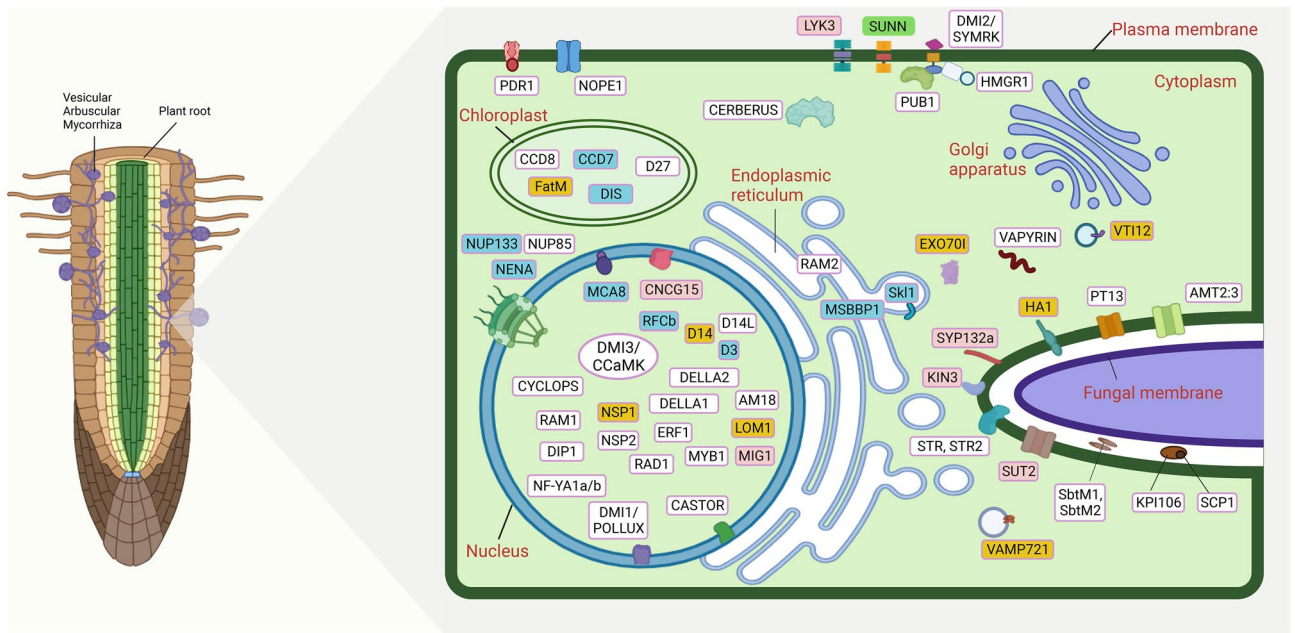


Figure 6. Common symbiosis signalling pathway in *T. cordifolia*. The figure (Created with BioRender.com) indicates the key genes of common symbiosis signalling pathway and the full name of genes is provided in the Supplementary Table S16. Genes in pink indicates genes only with positive selection, yellow indicates genes showing only amino acid substitutions, blue indicates genes with two signatures, and green indicates genes with all three signatures of adaptive evolution.

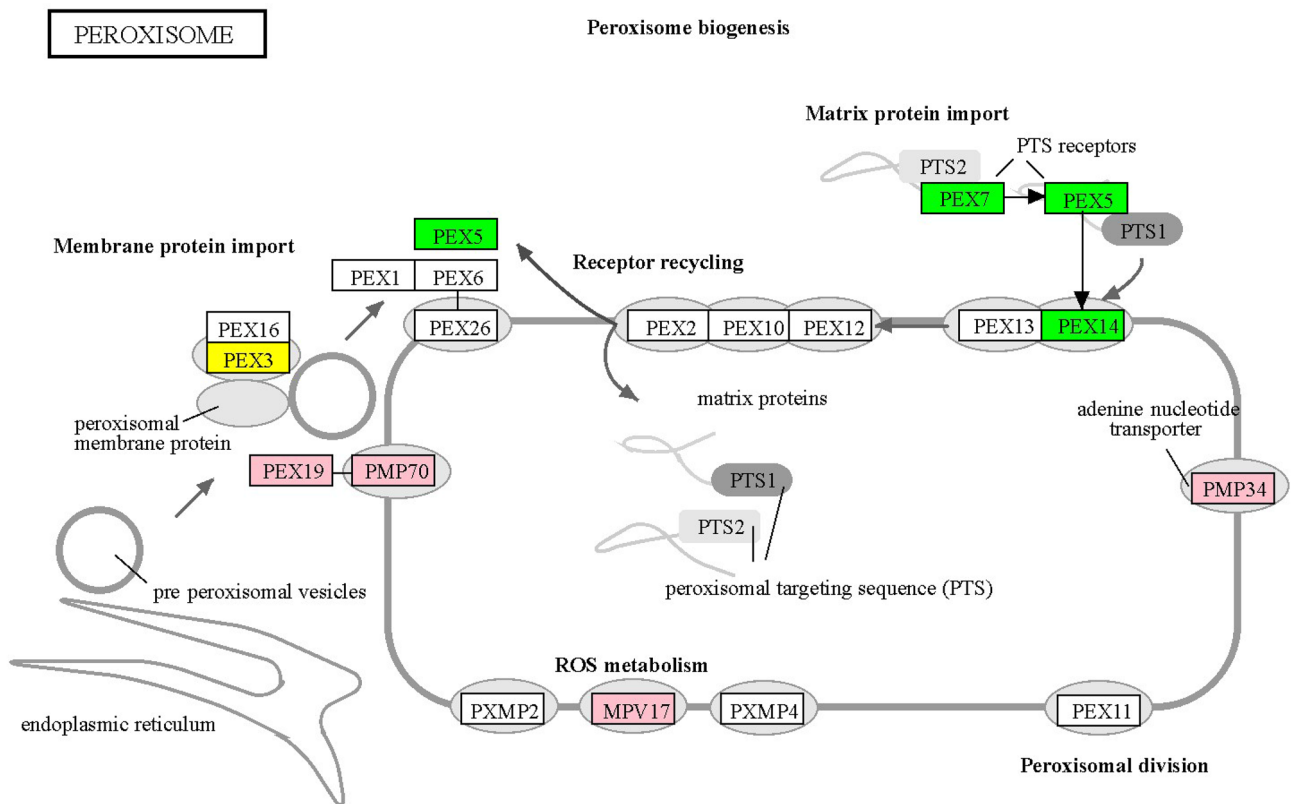


Figure 7. Peroxisome biogenesis pathway in *T. cordifolia*. The figure of peroxisome biogenesis is generated from KEGG database¹³⁷. It includes genes *PEX1*, *PEX2*, *PEX3*, *PEX5*, *PEX6*, *PEX7*, *PEX10*, *PEX11*, *PEX12*, *PEX13*, *PEX14*, *PEX16*, *PEX19*, *PEX26*, *PXMP2*, *MPV17*, *PXMP4*, and *PMP34*. Genes in pink, yellow and green colour indicate genes showing one, two and all three signatures of adaptive evolution, respectively.

T. cordifolia, being a climbing plant, remains in the shades of the supportive tree where lesser sunlight might reach the plant⁴⁹. Thus, these plants increase their photosynthesis capacity and anchorage by forming assimilatory roots. These aerial roots perform photosynthesis and provide anchorage that contributes to plant growth⁴⁸. The adventitious roots in *T. cordifolia* were found to be evolved due to the presence of signatures of adaptive evolution in these genes involved in adventitious root formation (Supplementary Table S17). These genes with evolutionary signatures were also reported in other plant species and play key roles in adventitious root formation^{92–94}. Additionally, phytohormones also play an important role in regulating adventitious root formation in plants⁹⁵. Along with the genes involved in adventitious root formation, phytohormone associated genes were also found to be evolved in *T. cordifolia*. Peroxisomes play an important role in photorespiration and plant growth that occurs in all C3 plants⁵¹. *T. cordifolia*, a C3 plant, performs photorespiration and lacks adaptations like the CAM pathway of CAM plants. Therefore, the signatures of adaptive evolution observed in genes associated with peroxisome biogenesis appear to be significant for C3 photosynthesis contributing to plant growth in *T. cordifolia* (Fig. 7 and Supplementary Table S18). These evolutionary signatures in genes associated with adventitious roots, phytohormones, and peroxisomes contribute towards the overall growth of *T. cordifolia*. Taken together, the comprehensive comparative evolutionary analysis helped to reveal the evolution in genes involved in common symbiosis signalling pathway, adventitious root formation, peroxisome associated, abiotic and biotic stress tolerance, and in genes associated with various phytohormones; thus, providing genomic clues on the diverse medicinal properties of this species.

The study reports the draft genome of *T. cordifolia* and provides important insights on the genomic basis of its medicinal properties due to the presence of diverse secondary metabolites. However, the presence of these metabolites also posed problems during the DNA and RNA extraction steps, and also severely reduced the data throughput from the ONT nanopore sequencing technologies. Thus, it is required to develop improved nucleic acid extraction protocols to deal with such medicinal plants. Further, the presence of high heterozygosity and repetitive content in *T. cordifolia* genome posed challenges during the assembly process and limited the assembly statistics such as the N50 score. The availability of more data from both second and third generation sequencing technologies is much needed to achieve a better coverage and a more contiguous assembly of this species. Further, due to the absence of genomes from the *Tinospora* genus and Menispermaceae family, species belonging to the same order were used for evolutionary analysis to eliminate the effect of large genetic distances. The availability of more genomes from closely related plant species would offer a deeper understanding of the species-specific or family-specific characteristics of *T. cordifolia*. Additionally, conducting experimental validation on the role of adaptively evolved genes in secondary metabolites biosynthesis will provide further support to the results of these analysis and will help in developing better approaches to benefit from the medicinal properties of this plant.

Conclusion

The insights obtained from this study and the availability of the genome sequence of *T. cordifolia* will help bridge the missing link between its genomic and medicinal properties and provide leads for exploring the genomic basis of these properties. It will also aid in various comparative genomic studies and act as a reference for future species sequenced from its genus and family. It will also help in the genome-wide phylogenetic assessments and evolutionary analyses of this species. The knowledge of mechanisms and pathways involved in producing its numerous medicinally important secondary metabolites will help better exploit these pathways and resultant metabolites for medicinal purposes and therapeutic applications.

Methods

Sample collection, species identification, nucleic acids extraction, and sequencing

The plant was brought from a nursery in Bhopal, Madhya Pradesh, India (23.2599° N, 77.4126° E). All methods were performed in accordance with the national guidelines concerning research involving plants. The leaves from the young plant were cleaned and used for DNA and RNA extractions. For DNA extraction, the leaves were homogenized in liquid nitrogen. For genomic DNA extraction, the powdered leaf was washed with 70% ethanol and distilled water to eliminate any such compounds that may hinder the extraction process, and employed CTAB-based lysis buffer for the isolation⁹⁶. Two genes: one nuclear gene and one chloroplast gene (internal transcribed spacer and Maturase K, respectively), were used for the species identification. These genes were amplified and sequenced at an in-house Sanger sequencing facility. The total RNA was extracted with TRIzol reagent from the leaf sample using multiple reactions and the RNA was then pooled before proceeding for the library preparation⁹⁷. The TruSeq Stranded Total RNA library preparation kit with Ribo-Zero Plant workflow (Illumina, Inc., USA) was deployed for preparing the transcriptomic library. The genomic library for linked reads was prepared using a Gel Bead kit and Chromium Genome library kit on a Chromium Controller instrument (10× Genomics, CA, USA). The quality of both the libraries (transcriptomic and genomic) was checked on TapeStation 4150 (Agilent, Santa Clara, CA) and sequenced on an Illumina platform, NovaSeq 6000 (Illumina, Inc., USA) for producing paired-end reads. For Nanopore sequencing, the genomic DNA libraries were prepared using SQK-LSK109 and 110, and sequenced on MinION sequencer. The comprehensive DNA and RNA extraction procedure is mentioned in the Supplementary Text.

Genome assembly

An array of Python scripts (<https://github.com/ucdavis-bioinformatics/proc10xG>) were used to remove the barcode sequences from the 10× Genomics linked reads. GenomeScope v2 was employed to calculate heterozygosity similar to other studies and SGA-preqc for genome size estimation of *T. cordifolia* using the k-mer count distribution method^{98–100}. To reduce the impact of sequencing errors on the estimation, only the k-mers with higher occurrences are considered in this k-mer count distribution method. Initially, the sga preprocess was run

with the option "-pe-mode 1" to consider all the paired-end barcode-filtered 10× Genomics reads. Following this step, the pre-processed reads were indexed by sga index with 'ropebwt' algorithm and "-no-reverse" options. Finally, the SGA-preqc was run with default options for the genome size estimation. The de novo assembly was generated by Supernova assembler v.2.1.1 (with maxreads=all options and other default settings) using 499.36 million raw reads¹⁰¹. The 'pseudohap2' style in Supernova mkoutput was implemented to assemble the haplotype-phased genome.

The barcodes of linked reads were processed using Longranger basic v2.2.2 (<https://support.10xgenomics.com/genome-exome/software/pipelines/latest/installation>) and these processed reads were used by Tigmint v1.2.1 to rectify the misassemblies present in Supernova assembled genome¹⁰². AGOUTI v0.3.3 with quality-filtered transcriptome reads was used to accomplish the initial scaffolding¹⁰³. To construct a more contiguous assembly, ARCS v1.1.1 with the barcode-processed linked reads was used to provide additional scaffolding and enhance the contiguity of the genome assembly¹⁰⁴. LINKS v1.8.6 was also used with the adapter-filtered Nanopore reads (using Porechop) for additional scaffolding¹⁰⁵. LR_Gapcloser used Nanopore reads to perform gap-closing of the assembly¹⁰⁶. Using a bloom filter-based method and k-mer value ranging from 30 to 120 with a ten bp interval, Sealer v2.1.5 used the linked reads (barcode processed) for gap-closing in the assembly¹⁰⁷. Performing scaffolding multiple times could give rise to local misassemblies, small indels, or distinct base errors that were overwhelmed using Pilon v1.23 which utilized the linked reads (barcode-processed) to increase the assembly quality¹⁰⁸. The obtained assembly was length based filtered such that scaffolds with length ≥ 2000 bp were retained. The completeness of genome assembly was evaluated with BUSCO v5.4.4 which used viridiplantae_odb10 database for the assessment¹⁰⁹.

Transcriptome assembly

The de novo transcriptome assembly was carried out using RNA-Seq data generated in this study. Trimmomatic v0.38 was used to process raw data reads, i.e., adapter removal and quality filtration¹¹⁰. The de novo transcriptome assembly was constructed using Trinity v2.9.1 with a strand-specific option and other default parameters using the processed paired-end reads¹¹¹. A Perl script offered in Trinity software package was utilized to evaluate the assembly statistics.

Genome annotation

The genome annotation was performed on the polished assembly. RepeatModeler v2.0.1 used this genome to construct a de novo repeat library¹¹². The obtained repeat sequences were clustered using CD-HIT-EST v4.8.1 with sequence identity as 90% and 8 bp seed size¹¹³. Repeat library obtained from RepeatModeler was further curated using TEclass2 software with a probability threshold of > 0.65 ¹¹⁴. Using this curated repeat library, RepeatMasker v4.1.0 (RepeatMasker Open-4.0, <http://www.repeatmasker.org>) soft-masked the genome that was used for the construction of the gene set. MAKER pipeline that employs ab initio-based gene prediction programs and evidence-based approaches to predict the final gene model was used for genome annotation¹¹⁵. As empirical evidence in the MAKER pipeline, the de novo transcriptome assembly of *T. cordifolia* and protein sequences of its phylogenetically closer species were used. The ab initio gene prediction, evidence-based alignments, and polishing of alignments were achieved using AUGUSTUS v3.2.3, BLAST and Exonerate v2.2.0, respectively, with the MAKER pipeline^{116,117}. The coding genes were filtered based on the criteria of AED (Annotation Edit Distance) value < 0.5 and gene length ≥ 150 bp¹¹⁸. This coding gene set was annotated against NCBI-nr and Swiss-Prot database using blastp (e-value 10^{-5}), and against Pfam-A database using HMMER v3.1 (e-value 10^{-5})^{116,119-121}. The completeness of genome annotation was evaluated with BUSCO v5.4.3 which used viridiplantae_odb10 database for the assessment¹⁰⁹. de novo tRNAs prediction, de novo rRNAs prediction, and miRNAs identification (homology-based) were performed using tRNAscan-SE v2.0.7, Barrnap v0.9 (<https://github.com/tseemann/barrnap>) and miRBase database, respectively¹²²⁻¹²⁴.

Functional annotation

T. cordifolia high-confidence gene set was annotated using blastp (e-value 10^{-5}) with NCBI non-redundant (nr) and SWISS-PROT databases, and HMMER v3.3 with Pfam-A v32.0 database^{116,119,120,125}. This coding gene set was functionally annotated and assign Gene Ontology (GO) categories, Cluster of Orthologous Groups (COG) categories, and KEGG pathways using WebGestalt web server, eggNOG-mapper v2.1.12, and KAAS v2.1, respectively¹²⁶⁻¹²⁸.

Phylogenetic analysis

Among all the eudicot species accessible on Ensembl Plants release 56, a total of 27 species were selected by choosing one species from each offered order. The 27 eudicot species and an outgroup species, *Zea mays*, were used to identify orthologs¹²⁹. The 28 selected species, along with the outgroup species, were—*Actinidia chinensis*, *Arabidopsis thaliana*, *Beta vulgaris*, *Cannabis sativa*, *Citrus clementina*, *Coffea canephora*, *Corylus avellana*, *Cucumis sativus*, *Daucus carota*, *Eucalyptus grandis*, *Ficus carica*, *Helianthus annuus*, *Ipomoea triloba*, *Juglans regia*, *Kalanchoe fedtschenkoi*, *Manihot esculenta*, *Medicago truncatula*, *Nicotiana attenuata*, *Olea europaea*, *Papaver somniferum*, *Pistacia vera*, *Populus trichocarpa*, *Quercus lobata*, *Rosa chinensis*, *Sesamum indicum*, *Theobroma cacao*, *Vitis vinifera*, and *Zea mays*.

The orthogroups were formed using the proteome files of 27 selected eudicot species along with *Zea mays* and MAKER retrieved protein sequences of *T. cordifolia*. Among all the protein sequences, the longest isoforms were retrieved for all the species and provided to OrthoFinder v2.5.4 for orthogroups construction¹³⁰. The orthogroups comprising genes from all 29 species were retrieved from all the identified orthogroups. KinFin v1.1 was used to increase the genes in one-to-one orthogroups that identified and extracted fuzzy one-to-one orthogroups among

these retrieved orthogroups¹³¹. In cases where multiple genes were present for a single species in any orthogroup, the longest gene among them was selected as representative.

MAFFT v7.310 was used to discretely align all the identified fuzzy one-to-one orthogroups to construct the phylogenetic tree¹³². The multiple sequence alignments were trimmed to eradicate empty sites, and the alignments were concatenated using BeforePhylo v0.9.0 (<https://github.com/qiyunzhu/BeforePhylo>). The concatenated alignments were used by RAXML v8.2.12, based on a rapid hill climbing algorithm, to create the maximum likelihood-based phylogenetic tree (100 bootstrap values and amino acid substitution model ‘PROTGAMMAAUTO’)¹³³. Similarly, another phylogenetic tree was constructed using a basal angiosperm species i.e., *Amborella trichopoda* as an outgroup instead of *Zea mays*. Additionally, MCMCTree was implemented to estimate the time of divergence between species with -nsample 1,000,000 and -burnin 10,000, and two calibration point was taken between basal eudicots—eudicots divergence (126–132 MYA) and between monocot—eudicot (142–163 MYA) from TimeTree database v5^{43,134}.

Gene family expansion and contraction analysis

The proteome files from 29 selected species and the above-generated phylogenetic tree were used to analyse the evolution of gene families using CAFÉ v5¹³⁵. The proteome files provided contained the longest isoform of each protein for all 29 species. Using TimeTree database v5.0, the divergence time between *T. cordifolia* and *Beta vulgaris* was obtained. This obtained time was used as a calibration point for adjusting the phylogenetic tree to an ultrametric species tree⁴³. These protein sequences were aligned using blastp in all-versus-all mode¹¹⁶. The clustering of blastp results was performed using MCL v12-137¹³⁶. Gene families with genes from < 2 species of the specified clades and ≥ 100 gene copies for ≥ 1 species were eliminated. Using the two-lambda (λ) model, the obtained gene families and the ultrametric species tree were used to evaluate the gene family contraction and expansion, where λ indicates a random birth–death parameter. Those with > 10 genes were considered as highly expanded or contracted gene families among the obtained gene families. The resultant highly contracted and expanded gene families were annotated using eggNOG-mapper v2.1.9¹²⁷. Further, the functional annotation of these highly expanded or contracted gene families were check manually on the Kyoto Encyclopaedia of Genes and Genomes (KEGG) database¹³⁷.

Orthologous gene clustering

The comparative genome-wide identification of orthologous genes across six genomes of Ranunculales order i.e., *Coptis chinensis*, *Corydalis tomentella*, *Kingdonia uniflora*, *Papaver somniferum*, *Thalictrum thalictroides*, and *Tinospora cordifolia* was conducted through OrthoVenn3 (<https://orthovenn3.bioinfotoolkits.net/home>)¹³⁸. The unique genes of *T. cordifolia* were obtained from this analysis.

Genes with signatures of adaptive evolution

The above mentioned six Ranunculales species were used for identification of genes with evolutionary signatures. OrthoFinder v2.5.4 was used to construct orthologous gene sets using the proteome files from six selected species¹³⁰. Orthogroups containing protein sequences from all six species were taken further for analysis. In case of more than one protein sequence occurrence for a species, the longest isoform was retained for analysis.

Genes with high nucleotide divergence

MAFFT v7.310 was used to individually align the orthogroups across six species¹³². These alignments were used by RAXML v8.2.12 to create orthogroups specific phylogenetic tree with the ‘PROTGAMMAAUTO’ amino acid substitution model and 100 bootstrap value¹³³. The root-to-tip branch length distance was calculated using the R package “adephylo” for every gene of all six species in the phylogenetic trees¹³⁹. The genes of *T. cordifolia* with relatively higher root-to-tip branch length distance values were extracted and considered as genes with higher evolution rates or nucleotide divergence.

Genes with unique amino acid substitutions with functional impact

Multiple sequence alignments were used to identify the protein sequences with identical amino acid positions in all species except *T. cordifolia*. The gaps and 10 positions around the gaps in the alignments were not considered for this analysis. Sorting Intolerant From Tolerant (SIFT) with UniProt database was used to evaluate the functional impact of extracted sequences with amino acid substitutions. The obtained genes were considered genes with unique amino acid substitutions with functional impact¹⁴⁰.

Positively selected genes

The nucleotide sequences of all orthogroups across selected six species were used by MAFFT v7.310 to obtain individual alignments¹³². The obtained alignments in PHYLIP format along with species phylogenetic tree were used by PAML v4.9a with “codeml” program¹³⁴. PAML v4.9a utilized the branch-site model to identify positively selected genes. Likelihood-ratio tests were performed on these PAML-identified genes and genes with < 0.05 FDR corrected p-values were referred to as positively selected genes of *T. cordifolia*. The identification of positively selected codon sites was performed using Bayes Empirical Bayes (BEB) analysis with > 95% probability criteria for the foreground lineage. These positive selected genes were functionally annotated and assigned to Gene Ontology (GO) categories and KEGG pathways using WebGestalt web server and KAAS v2.1, respectively^{126,128}.

Genes with all three signatures of adaptive evolution

The three evolutionary signatures of adaptation considered were higher evolution rate, positive selection, and unique amino acid substitutions with functional impact. *T. cordifolia* genes that indicated two of the three evolutionary signatures were suggested as genes with multiple signatures of adaptive evolution (MSA)^{118,124,141–143}. The *T. cordifolia* genes showing all three signatures were considered genes with three signatures of adaptive evolution. The eggNOG-mapper v2.1.9 was used to assign KEGG Orthology (KO) and KEGG pathway to the gene sets showing a higher rate of evolution, unique amino acid substitutions with functional impact, positive selection and three signatures of adaptive evolution¹²⁷. The KEGG database was used to retrieve the functional annotation of gene sets from the comparative evolutionary analysis¹³⁷.

Benzyl iso-quinoline alkaloid (BIA) biosynthesis pathway

The protein sequences of genes involved in BIA biosynthesis for *Papaver somniferum* or *Coptis chinensis* were downloaded from UniProt database¹⁴⁴. The BIA biosynthetic genes were checked for presence in *T. cordifolia* using blastp¹¹⁶. The exon–intron structures of these BIA biosynthetic genes in *T. cordifolia* and other five Ranunculales species were constructed using Exonerate v2.4.0 (<https://github.com/nathanweeks/exonerate>).

The distant orthologs of 15 genes involved in BIA biosynthesis pathway were identified to understand their origin. The distant orthologs were identified by HHblits web server using UniRef30 database and default parameters^{143,145}. The top 20 hits were extracted considering unique genus for each of 15 genes. These 20 hits along with *T. cordifolia* sequence were aligned for each of the 15 genes using MAFFT v.7.310¹³². RAxML v8.2.12 with 100 bootstrap value and ‘PROTGAMMAAUTO’ amino acid substitution model was used to construct the phylogenetic trees for each of the genes using the obtained alignments¹³³.

Other secondary metabolites biosynthesis pathway

The protein sequences of genes involved in terpenoid, lignin (backbone pathway for lignans biosynthesis), and flavonoid biosynthesis pathway for Ranunculales order members were obtained from UniProt database. The obtained genes were searched in *T. cordifolia* using blastp¹¹⁶. Among these *T. cordifolia* genes, these genes with signatures of adaptive evolution were identified.

Analysis of Common symbiosis signalling pathway genes

For CSSP, the genes reported in MacLean et al. (2017) were found in UniProt database and checked for their presence in *T. cordifolia* using blastp^{116,144,146}. The genes were manually checked for signatures of adaptive evolution in *T. cordifolia*.

Analysis of adventitious root formation genes

The genes reported by Li et al. (2021) were obtained from Uniprot database and checked for their presence in *T. cordifolia* using blastp^{116,144,147}. The genes were manually checked for signatures of adaptive evolution in *T. cordifolia*.

Gene expression analysis

Kallisto was used to quantify the transcripts of *T. cordifolia*¹⁴⁸. The TPM (Transcripts Per Million) values of genes involved in BIA biosynthesis, terpenoid biosynthesis, lignin biosynthesis, flavonoid biosynthesis, CSSP, formation of adventitious root formation, and genes showing all three signatures of adaptive evolution were analyzed.

Data availability

The raw data has been submitted to NCBI database with BioProject accession number PRJNA749156, Biosample accession SAMN20355817, and SRA accessions SRR15221491 and SRR15221490.

Received: 28 July 2023; Accepted: 29 January 2024

Published online: 02 February 2024

References

1. Devi, G. Medicinal Plant: Giloy (2020).
2. Chopra, D., Bhandari, B. & Dwivedi, S. Beneficial role of Indian medicinal plants in COVID-19. *MGM J. Med. Sci.* **8**, 166 (2021).
3. Lian, L., Ortiz, R. D. C., Jabbour, F., Chen, Z. D. & Wang, W. Re-delimitation of *Tinospora* (Menispermaceae): Implications for character evolution and historical biogeography. *Taxon* **68**, 905–917 (2019).
4. Mittal, J., Sharma, M. M. & Batra, A. *Tinospora cordifolia*: A multipurpose medicinal plant-A. *J. Med. Plants* **2** (2014).
5. Sonavale, R., Narkhede, A., Mulik, A., Kunchiraman, B. & Jagtap, S. D. In International Conference on Contemporary Research in Chemical and Life Science (2015).
6. Choudhry, N., Singh, S., Siddiqui, M. B. & Khatoon, S. Impact of seasons and dioecy on therapeutic phytoconstituents of *Tinospora cordifolia*, a Rasayana drug. *BioMed. Res. Int.* **2014** (2014).
7. Singh, D. & Chaudhuri, P. K. Chemistry and pharmacology of *Tinospora cordifolia*. *Nat. Prod. Commun.* **12**, 1934578X1701200240 (2017).
8. Sharma, P., Dwivedee, B. P., Bisht, D., Dash, A. K. & Kumar, D. The chemical constituents and diverse pharmacological importance of *Tinospora cordifolia*. *Heliyon* **5**, e02437 (2019).
9. Srinivasan, G., Unnikrishnan, K., Shree, A. R. & Balachandran, I. HPLC estimation of berberine in *Tinospora cordifolia* and *Tinospora sinensis*. *Indian J. Pharm. Sci.* **70**, 96 (2008).
10. Upadhyay, A. K., Kumar, K., Kumar, A. & Mishra, H. S. *Tinospora cordifolia* (Willd.) Hook. F. and Thoms. (Guduchi)–validation of the Ayurvedic pharmacology through experimental and clinical studies. *Int. J. Ayurveda Res.* **1**, 112 (2010).
11. Saha, S. & Ghosh, S. *Tinospora cordifolia*: One plant, many roles. *Anc. Sci. Life* **31**, 151 (2012).

12. Sinha, K., Mishra, N., Singh, J. & Khanuja, S. *Tinospora cordifolia* (Guduchi), a reservoir plant for therapeutic applications: A review (2004).
13. Devi, P. Role of *Tinospora cordifolia* in metabolic health disorders: An updated review. *Himal. J. Health Sci.* 15–23 (2021).
14. Borse, S. *et al.* Ayurveda botanicals in COVID-19 management: An in silico multi-target approach. *Plos One* **16**, e0248479 (2021).
15. Nath, H. *et al.* Nano-functionalization and evaluation of antimicrobial activity of *Tinospora cordifolia* against the TolB protein of *Pseudomonas aeruginosa*—An antibacterial and computational study. *Front. Microbiol.* **14**, 1138106 (2023).
16. Taechowisan, T. Antibacterial and cytotoxicity activities of major compounds from *Tinospora cordifolia* Willd. Growing on *Mangifera indica* L. *Int. J. Nutr.* **3**, 32–42 (2019).
17. Sriwastava, A. R. & Srivastava, V. Identification of bioactive agent in *Tinospora cordifolia* by in-silico approach.
18. Jeyachandran, R., Xavier, T. F. & Anand, S. Antibacterial activity of stem extracts of *Tinospora cordifolia* (Willd) Hook. f & Thomson. *Anc. Sci. Life* **23**, 40 (2003).
19. Singh, G., Saxena, R. K. & Singh, N. K. Screening of potential antimicrobial activity of Indian medicinal plant of different solvent extract: *Tinospora cordifolia* and *Hymenocallis littoralis*. *Int. Res. J. Eng. Technol.* **3**, 928–932 (2016).
20. Patil, R., Kulkarni, C. & Pandey, A. Antifungal and phytochemical properties of *Tinospora cordifolia*, *Azadirachta indica* and *Ocimum sanctum* leaves extract. *J. Med. Plants Stud.* **5**, 23–26 (2017).
21. Mishra, L. *et al.* In silico analysis of compounds isolated from selected Indian medicinal plants against Chikungunya Viral proteins. *Indian J. Pharm. Sci.* **82**, 677–685 (2020).
22. Bency, B. & Helen, P. In silico identification of dengue inhibitors in Giloy (*Tinospora cordifolia*) and Papaya. *J. Emerg. Technol. Innov. Res.* **5**, 506–511 (2018).
23. Sagar, V. & Kumar, A. H. Efficacy of natural compounds from *Tinospora cordifolia* against SARS-CoV-2 protease, surface glycoprotein and RNA polymerase. *Virology* 1–10 (2020).
24. Muralikumar, V., Ramakrishnamacharya, C. & Seshachalam, C. Inhibitory effect of phytochemicals from *Azadirachta indica* juss. and *Tinospora cordifolia* (thumb.) miers against SARS-COV-2 mpro and spike protease—An in silico analysis. *Int. J. Eng. Appl. Sci. Technol.* **5**, 303–319 (2020).
25. Shree, P. *et al.* Targeting COVID-19 (SARS-CoV-2) main protease through active phytochemicals of ayurvedic medicinal plants—*Withania somnifera* (Ashwagandha), *Tinospora cordifolia* (Giloy) and *Ocimum sanctum* (Tulsi)—a molecular docking study. *J. Biomol. Struct. Dyn.* 1–14 (2020).
26. Jena, S., Munusami, P., Balamurali, M. & Chanda, K. Computationally approached inhibition potential of *Tinospora cordifolia* towards COVID-19 targets. *VirusDisease* **32**, 65–77 (2021).
27. Balkrishna, A., Khandrika, L. & Varshney, A. Giloy Ghanvati (*Tinospora cordifolia* (Willd.) Hook. F. and Thomson) Reversed SARS-CoV-2 Viral Spike-Protein Induced Disease Phenotype in the Xenotransplant Model of Humanized Zebrafish. *Front. Pharmacol.* **12**, 534 (2021).
28. Thakkar, S. S., Shelat, F. & Thakor, P. Magical bullets from an indigenous Indian medicinal plant *Tinospora cordifolia*: An in silico approach for the antidote of SARS-CoV-2. *Egypt. J. Pet.* **30**, 53–66 (2021).
29. Sachan, S. *et al.* Immunomodulatory potential of *Tinospora cordifolia* and CpG ODN (TLR21 Agonist) against the very virulent, infectious Bursal disease virus in SPF Chicks. *Vaccines* **7**, 106 (2019).
30. de Vries, F. T., Griffiths, R. I., Knight, C. G., Nicolitch, O. & Williams, A. Harnessing rhizosphere microbiomes for drought-resilient crop production. *Science* **368**, 270–274 (2020).
31. Berendsen, R. L., Pieterse, C. M. & Bakker, P. A. The rhizosphere microbiome and plant health. *Trends Plant Sci.* **17**, 478–486 (2012).
32. Chen, M., Arato, M., Borghi, L., Nouri, E. & Reinhardt, D. Beneficial services of arbuscular mycorrhizal fungi—from ecology to application. *Front. Plant Sci.* **9**, 1270 (2018).
33. Zhao, Y., Cartabia, A., Lalaymia, I. & Declerck, S. Arbuscular mycorrhizal fungi and production of secondary metabolites in medicinal plants. *Mycorrhiza* **32**, 221–256 (2022).
34. Mishra, Y., Mittal, J., Singh, A., Batra, A. & Sharma, M. M. In vivo and in vitro histological localization of endophytic fungi in *Tinospora cordifolia* (Willd.) Miers ex Hook F. & Thomas. *J. Appl. Res. Med. Aromat. Plants* **2**, 30–33 (2015).
35. Kapoor, N. & Saxena, S. Endophytic fungi of *Tinospora cordifolia* with anti-gout properties. *3 Biotech* **8**, 1–6 (2018).
36. Mishra, A. *et al.* Season and tissue type affect fungal endophyte communities of the Indian medicinal plant *Tinospora cordifolia* more strongly than geographic location. *Microb. Ecol.* **64**, 388–398 (2012).
37. Duhana, P., Bansal, P. & Rani, S. Isolation, identification and characterization of endophytic bacteria from medicinal plant *Tinospora cordifolia*. *S. Afr. J. Bot.* **134**, 43–49 (2020).
38. Singh, R. *et al.* De novo transcriptome sequencing facilitates genomic resource generation in *Tinospora cordifolia*. *Funct. Integr. Genom.* **16**, 581–591 (2016).
39. Abraham, A. Chromosome number in *Tinospora*. *Curr. Sci.* **11**, 282–282 (1942).
40. Mathew, P. in *Proceedings of the Indian Academy of Sciences-Section B* 274–286 (Springer, 1958).
41. Santra, I., Halder, T. & Ghosh, B. Research article Somatic and gametic chromosomal characterization with fluorescence banding of Giloy (*Tinospora cordifolia*): A berberine synthesizing important medicinal plant of India. *Caryologia. Int. J. Cytol. Cytosyst. Cytogenet.* (2021).
42. Haque, M. A., Jantan, I. & Bukhari, S. N. A. *Tinospora* species: An overview of their modulating effects on the immune system. *J. Ethnopharmacol.* **207**, 67–85 (2017).
43. Kumar, S. *et al.* TimeTree 5: An expanded resource for species divergence times. *Mol. Biol. Evol.* **39**, msac174 (2022).
44. Worberg, A. *et al.* Phylogeny of basal eudicots: Insights from non-coding and rapidly evolving DNA. *Org. Divers. Evol.* **7**, 55–77 (2007).
45. Wang, W., Lu, A.-M., Ren, Y., Endress, M. E. & Chen, Z.-D. Phylogeny and classification of Ranunculales: Evidence from four molecular loci and morphological data. *Perspect. Plant Ecol. Evol. Syst.* **11**, 81–110 (2009).
46. Yadav, V., Singh, A., Mathur, N. & Yadav, R. Isolation and characterization of *Alternaria* GFAV15, an endophytic fungus from green fruit of *Tinospora cordifolia* (Willd.) Miers from semi-arid region. *S. Afr. J. Bot.* **134**, 343–348 (2020).
47. Radhakrishnan, G. V. *et al.* An ancestral signalling pathway is conserved in intracellular symbioses-forming plant lineages. *Nat. Plants* **6**, 280–289 (2020).
48. Steffens, B. & Rasmussen, A. The physiology of adventitious roots. *Plant Physiol.* **170**, 603–617 (2016).
49. Subbiah, V. R. & Dayanandan, P. Growth kinetics, vascular differentiation and functions of aerial roots of *Tinospora cordifolia* (1990).
50. Mishra, A. *et al.* Characterization of *Pseudofusicoccum adansoniae*, an endophytic fungus residing in photosynthetic root of *Tinospora cordifolia*, a medicinal plant. *Proc. Natl. Acad. Sci. India Sect. B Biol. Sci.* **89**, 1319–1326 (2019).
51. Hu, J. *et al.* Plant peroxisomes: Biogenesis and function. *Plant Cell* **24**, 2279–2303 (2012).
52. Begum, H. J. & Ramamurthy, V. Evaluation of secondary metabolites and antioxidant activity of ethanolic leaves extract of *Tinospora Cordifolia*.
53. Mostovoy, Y. *et al.* A hybrid approach for de novo human genome sequence assembly and phasing. *Nat. Methods* **13**, 587–590 (2016).
54. Kalpesh, I. & Mohan, J. Assessment of genetic diversity in medicinal climber of *Tinospora cordifolia* (Willd.) Miers (Menispermaceae) from Gujarat, India. *Asian J. Biotechnol.* **1**, 93–103 (2009).

55. Paliwal, R. *et al.* Development of genomic simple sequence repeats (g-SSR) markers in *Tinospora cordifolia* and their application in diversity analyses. *Plant Gene* **5**, 118–125 (2016).
56. Lade, S. *et al.* Diversity in a widely distributed dioecious medicinal plant, *Tinospora cordifolia* (Willd.) Miers. Ex. Hook F. and Thomas. *Curr. Sci.* **114**, 1520 (2018).
57. Soltis, D. Angiosperm phylogeny inferred from 18S rDNA, rbcL, and atpB sequences. *Bot. J. Linn. Soc.* **133**, 381–461 (2000).
58. Yang, L. *et al.* Phylogenomic insights into deep phylogeny of angiosperms based on broad nuclear gene sampling. *Plant Commun.* **1**, 100027 (2020).
59. Sun, Y. *et al.* Genome sequencing of the endangered *Kingdonia uniflora* (Circaceasteraceae, Ranunculales) reveals potential mechanisms of evolutionary specialization. *Iscience* **23** (2020).
60. Liu, Y. *et al.* Analysis of the *Coptis chinensis* genome reveals the diversification of protoberberine-type alkaloids. *Nat. Commun.* **12**, 3276 (2021).
61. Chen, D.-x. *et al.* The chromosome-level reference genome of *Coptis chinensis* provides insights into genomic evolution and berberine biosynthesis. *Hortic. Res.* **8** (2021).
62. Pei, L. *et al.* Genome and transcriptome of Papaver somniferum Chinese landrace CHM indicates that massive genome expansion contributes to high benzyloquinoline alkaloid biosynthesis. *Hortic. Res.* **8** (2021).
63. Liu, X. *et al.* The genome of medicinal plant *Macleaya cordata* provides new insights into benzyloquinoline alkaloids metabolism. *Mol. Plant* **10**, 975–989 (2017).
64. Xu, Z. *et al.* The genome of *Corydalis* reveals the evolution of benzyloquinoline alkaloid biosynthesis in Ranunculales. *Plant J.* **111**, 217–230 (2022).
65. Gnanam, S. In vitro and in silico approaches on the antibacterial activity of *Tinospora cordifolia* methanolic stem extract. *Asian J. Pharm. Clin. Res.* 126–131 (2020).
66. Chowdhury, P. In silico investigation of phytoconstituents from Indian medicinal herb '*Tinospora cordifolia* (giloy)' against SARS-CoV-2 (COVID-19) by molecular dynamics approach. *J. Biomol. Struct. Dyn.* 1–18 (2020).
67. Khanal, P., Patil, B. M., Mandar, B. K., Dey, Y. N. & Duyu, T. Network pharmacology-based assessment to elucidate the molecular mechanism of anti-diabetic action of *Tinospora cordifolia*. *Clin. Phytosci.* **5** (2019).
68. Beaudoin, G. A. & Facchini, P. J. Benzyloquinoline alkaloid biosynthesis in opium poppy. *Planta* **240**, 19–32 (2014).
69. Hagel, J. M. & Facchini, P. J. Benzyloquinoline alkaloid metabolism: A century of discovery and a brave new world. *Plant Cell Physiol.* **54**, 647–672 (2013).
70. Boutanaev, A. M. *et al.* Investigation of terpene diversification across multiple sequenced plant genomes. *Proc. Natl. Acad. Sci.* **112**, E81–E88 (2015).
71. Chakraborty, A., Mahajan, S., Bisht, M. S. & Sharma, V. K. Genome sequencing of *Syzygium cumini* (jamun) reveals adaptive evolution in secondary metabolism pathways associated with its medicinal properties. *Front. Plant Sci.* **14** (2023).
72. Chen, R. *et al.* Gene-to-metabolite network for biosynthesis of lignans in MeJA-elicited *Isatis indigotica* hairy root cultures. *Front. Plant Sci.* **6**, 952 (2015).
73. Bagniewska-Zadworna, A., Barakat, A., Łakomy, P., Smoliński, D. J. & Zadworny, M. Lignin and lignans in plant defence: Insight from expression profiling of cinnamyl alcohol dehydrogenase genes during development and following fungal infection in *Populus*. *Plant Sci.* **229**, 111–121 (2014).
74. Li, C. *et al.* Chromosome-scale genome assembly provides insights into the molecular mechanisms of tissue development of *Populus wilsonii*. *Commun. Biol.* **5**, 1125 (2022).
75. Mahajan, S., Bisht, M. S., Chakraborty, A. & Sharma, V. K. Genome of *Phyllanthus emblica*: The medicinal plant Amla with super antioxidant properties. *Front. Plant Sci.* **14** (2023).
76. Begum, N. *et al.* Role of arbuscular mycorrhizal fungi in plant growth regulation: Implications in abiotic stress tolerance. *Front. Plant Sci.* **10**, 1068 (2019).
77. Solanki, M.K. *et al.* in *Microbiomes and Plant Health* 205–223 (Elsevier, 2021).
78. Jayakodi, M. *et al.* The giant diploid faba genome unlocks variation in a global protein crop. *Nature* **615**, 652–659 (2023).
79. Thiergart, T. *et al.* *Lotus japonicus* symbiosis genes impact microbial interactions between symbionts and multikingdom commensal communities. *MBio* <https://doi.org/10.1128/mbio.01833-19> (2019).
80. Liao, D. *et al.* Phytohormones regulate the development of arbuscular mycorrhizal symbiosis. *Int. J. Mol. Sci.* **19**, 3146 (2018).
81. Gutjahr, C. Phytohormone signaling in arbuscular mycorrhiza development. *Curr. Opin. Plant Biol.* **20**, 26–34 (2014).
82. Khan, A. L. *et al.* Endophytes from medicinal plants and their potential for producing indole acetic acid, improving seed germination and mitigating oxidative stress. *J. Zhejiang Univ. Sci. B* **18**, 125 (2017).
83. Verma, A. *et al.* Fungal endophytes to combat biotic and abiotic stresses for climate-smart and sustainable agriculture. *Front. Plant Sci.* **13**, 953836 (2022).
84. Kedar, A., Rathod, D., Yadav, A., Agarkar, G. & Rai, M. *Endophytic Phoma* sp. isolated from medicinal plants promote the growth of *Zea mays*. *Nusant. Biosci.* **6** (2014).
85. Kamli, M. R., Alzahrani, N. A., Hajrah, N. H., Sabir, J. S. & Malik, A. Genome-driven discovery of enzymes with industrial implications from the genus *Aneurimibacillus*. *Microorganisms* **9**, 499 (2021).
86. Vig, R., Bhadra, F., Gupta, S. K., Sairam, K. & Vasundhara, M. Neuroprotective effects of quercetin produced by an endophytic fungus *Nigrospora oryzae* isolated from *Tinospora cordifolia*. *J. Appl. Microbiol.* **132**, 365–380 (2022).
87. Dwibedi, V. *et al.* Key insights into secondary metabolites from various *Chaetomium* species. *Appl. Microbiol. Biotechnol.* 1–17 (2023).
88. Kim, J. W. & Shim, S. H. The fungus *Colletotrichum* as a source for bioactive secondary metabolites. *Arch. Pharm. Res.* **42**, 735–753 (2019).
89. Mehta, T., Meena, M. & Nagda, A. Bioactive compounds of *Curvularia* species as a source of various biological activities and biotechnological applications. *Front. Microbiol.* **13**, 1069095 (2022).
90. Zhao, S. *et al.* Secondary metabolites of *Alternaria*: A comprehensive review of chemical diversity and pharmacological properties. *Front. Microbiol.* **13**, 1085666 (2023).
91. Xu, T. *et al.* Secondary metabolites of the genus *Nigrospora* from terrestrial and marine habitats: Chemical diversity and biological activity. *Fitoterapia* 105254 (2022).
92. Mhimdi, M. & Pérez-Pérez, J. M. Understanding of adventitious root formation: What can we learn from comparative genetics?. *Front. Plant Sci.* **11**, 582020 (2020).
93. Kidwai, M., Mishra, P. & Bellini, C. Species-specific transcriptional reprogramming during adventitious root initiation. *Trends Plant Sci.* **28**, 128–130 (2023).
94. Devi, J., Kaur, E., Swarnkar, M. K., Acharya, V. & Bhushan, S. D. De novo transcriptome analysis provides insights into formation of in vitro adventitious root from leaf explants of *Arnebia euchroma*. *BMC Plant Biol.* **21**, 1–16 (2021).
95. Bellini, C., Pacurar, D. I. & Perrone, I. Adventitious roots and lateral roots: Similarities and differences. *Annu. Rev. Plant Biol.* **65**, 639–666 (2014).
96. Gaytán-Oyarzún, C.S.-H. Two mini-preparation protocols to DNA extraction from plants with high polysaccharide and secondary metabolites. *Afr. J. Biotechnol.* **5**(20), 1864–1867 (2006).
97. Johnson, M. T. *et al.* Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. *Plos One* **7**, e50226 (2012).

98. Chakraborty, A. *et al.* Genome sequencing and de novo and reference-based genome assemblies of *Bos indicus* breeds. *Genes Genom.* **1–10** (2023).
99. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **11**, 1–10 (2020).
100. Simpson, J. T. & Durbin, R. Efficient de novo assembly of large genomes using compressed data structures. *Genome Res.* **22**, 549–556 (2012).
101. Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M. & Jaffe, D. B. Direct determination of diploid genome sequences. *Genome Res.* **27**, 757–767 (2017).
102. Jackman, S. D. *et al.* Tigmint: Correcting assembly errors using linked reads from large molecules. *BMC Bioinform.* **19**, 1–10 (2018).
103. Zhang, S. V., Zhuo, L. & Hahn, M. W. AGOUTI: Improving genome assembly and annotation using transcriptome data. *GigaScience* **5**, s13742-016-0136-3 (2016).
104. Yeo, S., Coombe, L., Warren, R. L., Chu, J. & Birol, I. ARCS: Scaffolding genome drafts with linked reads. *Bioinformatics* **34**, 725–731 (2018).
105. Warren, R. L. *et al.* LINKS: Scalable, alignment-free scaffolding of draft genomes with long reads. *GigaScience* **4**, s13742-015-0076-3 (2015).
106. Xu, G.-C. *et al.* LR_Gapcloser: A tiling path-based gap closer that uses long reads to complete genome assembly. *GigaScience* **8**, giy157 (2019).
107. Paulino, D. *et al.* Sealer: A scalable gap-closing application for finishing draft genomes. *BMC Bioinform.* **16**, 1–8 (2015).
108. Walker, B. J. *et al.* Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963 (2014).
109. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
110. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
111. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
112. Flynn, J. M. *et al.* RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci.* **117**, 9451–9457 (2020).
113. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
114. Bickmann, L., Rodriguez, M., Jiang, X. & Makalowski, W. TEclass2: Classification of transposable elements using Transformers. *BioRxiv*, 2023.10.13.562246 (2023).
115. Campbell, M. S., Holt, C., Moore, B. & Yandell, M. Genome annotation and curation using MAKER and MAKER-P. *Curr. Protoc. Bioinform.* **48**, 4.11.1–4.11.39 (2014).
116. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
117. Stanke, M. *et al.* AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439 (2006).
118. Chakraborty, A., Mahajan, S., Bisht, M. S. & Sharma, V. K. Genome sequencing and comparative analysis of *Ficus benghalensis* and *Ficus religiosa* species reveal evolutionary mechanisms of longevity. *Iscience* 105100 (2022).
119. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**, 45–48 (2000).
120. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37 (2011).
121. Sharma, A. K., Gupta, A., Kumar, S., Dhakan, D. B. & Sharma, V. K. Woods: A fast and accurate functional annotator and classifier of genomic and metagenomic sequences. *Genomics* **106**, 1–6 (2015).
122. Griffiths-Jones, S., Saini, H. K., Van Dongen, S. & Enright, A. J. miRBase: Tools for microRNA genomics. *Nucleic Acids Res.* **36**, D154–D158 (2007).
123. Chan, P. P. & Lowe, T. M. in *Gene Prediction 1–14* (Springer, 2019).
124. Jaiswal, S. K., Mahajan, S., Chakraborty, A., Kumar, S. & Sharma, V. K. The genome sequence of *Aloe vera* reveals adaptive evolution of drought tolerance mechanisms. *Iscience* **24**, 102079 (2021).
125. Bateman, A. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **32**, D138–D141 (2004).
126. Liao, Y., Wang, J., Jaehnic, E. J., Shi, Z. & Zhang, B. WebGestalt 2019: Gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* **47**, W199–W205 (2019).
127. Huerta-Cepas, J. *et al.* Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).
128. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C. & Kanehisa, M. KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **35**, W182–W185 (2007).
129. Bolser, D., Staines, D. M., Pritchard, E. & Kersey, P. in *Plant bioinformatics 115–140* (Springer, 2016).
130. Emms, D. M. & Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 1–14 (2019).
131. Laetsch, D. R. & Blaxter, M. L. KinFin: Software for Taxon-Aware analysis of clustered protein sequences. *G3 Genes Genomes Genet.* **7**, 3349–3357 (2017).
132. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
133. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
134. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
135. Mendes, F. K., Vanderpool, D., Fulton, B. & Hahn, M. W. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* **36**, 5516–5518 (2020).
136. Van Dongen, S. & Abreu-Goodger, C. Using MCL to extract clusters from networks. *Bact. Mol. Netw. Methods Protoc.* 281–295 (2012).
137. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
138. Sun, J. *et al.* OrthoVenn3: An integrated platform for exploring and visualizing orthologous data across genomes. *Nucleic Acids Res.* gkad313 (2023).
139. Jombart, T. & Dray, S. Adephylo: Exploratory analyses for the phylogenetic comparative method. *Bioinformatics* **26**, 1–21 (2010).
140. Ng, P. C. & Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).
141. Jaiswal, S. K. *et al.* Genome sequence of peacock reveals the peculiar case of a glittering bird. *Front. Genet.* **9**, 392 (2018).
142. Mittal, P., Jaiswal, S. K., Vijay, N., Saxena, R. & Sharma, V. K. Comparative analysis of corrected tiger genome provides clues to its neuronal evolution. *Sci. Rep.* **9**, 1–11 (2019).

143. Chakraborty, A., Mahajan, S., Jaiswal, S. K. & Sharma, V. K. Genome sequencing of turmeric provides evolutionary insights into its medicinal properties. *Commun. Biol.* **4**, 1193 (2021).
144. Consortium, U. UniProt: A hub for protein information. *Nucleic Acids Res.* **43**, D204–D212 (2015).
145. Remmert, M., Biegert, A., Hauser, A. & Söding, J. HHblits: Lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods* **9**, 173–175 (2012).
146. MacLean, A. M., Bravo, A. & Harrison, M. J. Plant signaling and metabolic pathways enabling arbuscular mycorrhizal symbiosis. *Plant Cell* **29**, 2319–2335 (2017).
147. Li, S.-W. Molecular bases for the regulation of adventitious root generation in plants. *Front. Plant Sci.* **12**, 614072 (2021).
148. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).

Acknowledgements

SM and AC acknowledge Council of Scientific and Industrial Research (CSIR) for supporting research fellowships. MSB thanks the Ministry of Education, Govt. of India for Prime Minister Research Fellowship (PMRF). TS thanks the Department of Science and Technology and IISc Bangalore for providing the KVPY scholarship. The authors also thank Central Instrumentation Facility, IISER Bhopal and intramural research funds provided by IISER Bhopal.

Author contributions

V.K.S. conceived and coordinated the project and collected the plant sample. S.M. extracted the nucleic acids (DNA and RNA) from the collected sample, prepared the DNA and RNA samples for sequencing, carried out the species identification assay and performed Nanopore sequencing. AC and V.K.S. conceived the computational outline of the study. A.C. and M.S.B. accomplished all the computational analyses presented in the study. S.M., A.C. and M.S.B. interpreted the results. S.M. constructed the figures. S.M., A.C., M.S.B., T.S. and V.K.S. wrote the manuscript. All the authors have read and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-53176-z>.

Correspondence and requests for materials should be addressed to V.K.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024