Title: Genome sequencing and assembly of *Tinospora cordifolia* (Giloy) plant

Authors: Shruti Mahajan, Abhisek Chakraborty, Titas Sil, Vineet K Sharma*

Affiliation: MetaBioSys Group, Dept. of Biological Sciences, Indian Institute of Science Education and Research Bhopal, Bhopal, Madhya Pradesh, 462066, India

*Corresponding Author email:
Vineet K Sharma - vineetks@iiserb.ac.in

E-mail addresses of authors: Shruti Mahajan - shruti17@iiserb.ac.in, Abhisek Chakraborty - abhisek18@iiserb.ac.in, Titas Sil - titas16@iiserb.ac.in, Vineet K Sharma - vineetks@iiserb.ac.in

ABSTRACT

During the ongoing COVID-19 pandemic *Tinospora cordifolia* also known as Giloy gained immense popularity and use due to its immunity-boosting function and anti-viral properties. *T. cordifolia* is among the most important medicinal plants that has numerous therapeutic applications in health due to the production of a diverse array of secondary metabolites. Therefore, to gain genomic insights into the medicinal properties of *T. cordifolia*, the first genome sequencing was carried out using 10x Genomics linked read technology and the draft genome assembly comprised of 1.01 Gbp. This is also the first genome sequenced from the plant family Menispermaceae. We also performed the first genome size estimation for *T. cordifolia* and was found to be 1.13 Gbp. The deep sequencing of transcriptome from the leaf tissue was also performed followed by transcriptomic analysis to gain insights into the gene expression and functions. The genome and transcriptome assemblies were used to construct the gene set in *T. cordifolia* that resulted in 19,474 coding gene sequences. Further, the phylogenetic position of *T. cordifolia* was also determined through the construction of a genome-wide phylogenetic tree using 35 other dicot species and one monocot species as an outgroup species.

INTRODUCTION

Tinospora cordifolia is a climbing shrub belonging to the Menispermaceae family that includes more than 400 plant species of high therapeutic properties [1, 2]. 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]. *T. cordifolia* is found in tropical and sub-tropical regions including India, China, Sri Lanka, Bangladesh, Myanmar, Thailand, Malaysia, etc. and also known as 'Giloy', 'Amrita', 'Guduchi', and 'heart leaved moonseed' [2]. 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 [2, 4]. Being a climber, *T. cordifolia* needs a supportive plant like *Jatropha curcas* (Jatropha), *Azadirachta indica* (Neem), *Moringa oleifera* (Moringa), etc. for its growth [4]. These co-occurring plants also play an important role in enhancing the production of various secondary metabolites of *T. cordifolia* [4, 5]. Previous reports also indicated the presence of endophytic fungi in the leaves and the stem of this plant but their ecological significance has yet to be studied [6, 7]. This plant produces the secondary metabolites in response to the stress conditions and their concentration also varies based on seasons and its dioecy status [8]. High genetic diversity has been reported in *T. cordifolia* due to the dioecious nature [9-11].

The chemical constituents of this plant have been broadly categorized as alkaloids (tinosporine, magnoflorine, berberine, etc.), terpenoids (tinosporide, furanolactone diterpene, cordifolioside, etc.), phenolics (lignans, flavonoids, phenylpropanoids etc.), polysaccharides (glucose, xylose, rhamnose, etc.), steroids (giloinsterol, ß-sitosterol, etc.), essential oils and aliphatic compounds along with a few other compounds such as giloin, tinosporidine, sinapic acid, tinosporone, tinosporic acid, etc. that are obtained from various parts of the plant [12, 13]. A terpene tinosporaside and an alkaloid berberine were found to be the most dominant compounds in *T. cordifolia* and suggested to use them as its chemical biomarkers [14, 15]. The bitter taste of *T. cordifolia* is due to the presence of tinosporic acid, tinosporol, giloin, giloinin, tinosporide, cordifolide, tinosporin and a few other compounds [12]. 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 berberine than *T. sinensis,* and thus the former is preferred in therapeutics [16].

The bioactive compounds found in *T. cordifolia* have known biological properties such as anti-pyretic, anti-diabetic, anti-inflammatory, anti-microbial, anti-allergic, anti-oxidant, anti-diabetic, anti-toxic, anti-arthritis, anti-osteoporotic, anti-HIV, anti-cancer, hepatoprotective, anti-malarial, and also in immunomodulation etc. [17, 18]. 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. 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. [19, 20]. These diverse and important therapeutic applications make it a species of interest for a broad scientific community. Interestingly, this Giloy plant has gained tremendous therapeutic interest and significance during the recent and ongoing COVID-19 pandemic [21].

However, despite the widely known and important medicinal properties of this plant its genome assembly is yet unavailable. A preliminary study reported the transcriptome (482 Mbp data) of this species from leaf and stem tissues using 454 GS-FLX pyrosequencing [22]. A recent karyological study reported 2n=26 as the chromosome number in *T. cordifolia*, which was also supported by the earlier studies [23-25]. Thus, to uncover the genomic basis of its medicinal properties and for further exploration of its therapeutic potential, we carried out the first genome sequencing and assembly of *T. cordifolia* using 10x Genomics linked reads. This is the first draft genome assembly of *T. cordifolia* which is also the first genome sequenced so far from the medicinally important genus *Tinospora* and its family [26]. We also carried out a comprehensive deep sequencing and assembly of the leaf

transcriptome using Illumina reads. The genome-wide phylogenetic analysis was also carried out for *T. cordifolia* with other dicot species and a monocot species as an outgroup to determine its phylogenetic position.

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). DNA and RNA extraction were carried out using a cleaned leaf, which was homogenized in liquid nitrogen. The TRIzol reagent was used for RNA extraction [27]. For genomic DNA extraction, the powdered leaf was washed with 70% ethanol and distilled water in order to eliminate any such compounds that may hinder the extraction process and employed CTAB based Carlson lysis buffer for the isolation [28]. 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 in-house sanger sequencing facility. 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 Gel Bead kit and Chromium Genome library kit on a Chromium Controller instrument (10x 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. The comprehensive DNA and RNA extraction procedure is mentioned in **Supplementary Text S1.**

Genome assembly

An array of Python scripts (https://github.com/ucdavis-bioinformatics/proc10xG) were used to remove the barcode sequences from the raw reads. SGA-preqc was employed for genome size estimation of *T. cordifolia* using k-mer count distribution method **(Supplementary text S2)** [29]. 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 reads [30]. The 'pseudohap2' style in Supernova mkoutput was implemented to assemble 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 mis-assemblies present in Supernova assembled genome [31]. AGOUTI v0.3.3 with quality-filtered transcriptome reads was used to accomplish the initial scaffolding [32]**.** In order to construct a more contiguous assembly ARCS v1.1.1 with its default parameters was used to provide additional scaffolding and enhance the contiguity of the genome assembly [33]. Using a bloom filter-based method and k-mer value ranging from 30 to 120 with 10 bp interval, Sealer v2.1.5 used the linked reads (barcode processed) for gap-closing in the assembly [34]. Performing scaffolding multiple times could give rise to local mis-assemblies, small indels or distinct base errors which were overwhelmed using Pilon v1.23 that utilized the linked reads (barcode-processed) to increase the assembly quality [35]. The completeness of genome assembly was evaluated with BUSCO v4.1.4 which used embryophyte_odb10 database for the assessment [36]. The additional information about the post-processing of genome assembly is provided in **Supplementary Text S2.**

Transcriptome assembly

The de novo transcriptome assembly was carried out using RNA-Seq data generated in this study. Trimmomatic v.0.38 was used for processing of raw data reads i.e., adapter removal and quality-filtration **[Supplementary Text S2]** [37]. The de novo transcriptome assembly was constructed using Trinity v2.9.1 with strand-specific option and other default parameters using the processed paired-end reads [38]. A Perl script offered in Trinity software package was utilized to evaluate the assembly statistics.

Genome annotation

The genome annotation was achieved on the polished assembly (length ≥1,000 bp). This genome was used by RepeatModeler v2.0.1 to construct a *de novo* repeat library [39]. The clustering of obtained repeat sequences was performed using CD-HIT-EST v4.8.1 with sequence identity as 90% and 8 bp seed size [40]. Using the repeat library, RepeatMasker v4.1.0 (RepeatMasker Open-4.0, http://www.repeatmasker.org) was used to soft-mask the genome that was used for the construction of gene set. MAKER pipeline that employs *ab initio*-based gene prediction programs as well as evidence-based approaches for prediction of final gene model was used for genome annotation [41]. As an empirical evidence in MAKER pipeline the de novo transcriptome assembly of *T. cordifolia* and protein sequences of its phylogenetically closer species *Beta vulgaris* (belonging to plant order Caryophyllales) 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 [42, 43]. The completeness of the coding gene set was also assessed using BUSCO v4.1.4 embryophyte_odb10 database. The tandem repeats detection, de novo tRNAs prediction, de novo rRNAs prediction and miRNAs identification (homology based) were performed using Tandem Repeat Finder (TRF) v4.09, tRNAscan-SE v2.0.7, Barrnap v0.9 (https://github.com/tseemann/barrnap) and miRBase database, respectively [44-46]. The detailed information about genome annotation is provided in **Supplementary Text S3.**

Orthogroups identification

Among all the eudicot species accessible on Ensembl Plants release 48, a total of 35 species were selected by choosing one species from each offered genus. The 35 eudicot species along with an outgroup species, *Zea mays* were used for the identification of orthologs [47]. The 36 selected species along with the outgroup species were - *Actinidia chinensis*, *Arabidopsis thaliana*, *Arabis alpina*, *Beta vulgaris*, *Brassica napus*, *Camelina sativa*, *Cannabis sativa* female, *Capsicum annuum*, *Citrullus lanatus*, *Citrus clementina*, *Coffea canephora*, *Corchorus capsularis*, *Cucumis melo*, *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*, *Rosa chinensis*, *Solanum tuberosum*, *Theobroma cacao* Matina 1-6, *Trifolium pratense*, *Vigna angularis*, *Vitis vinifera,* and *Zea mays*. The orthogroups were formed using the proteome files of 35 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 v2.4.1 for orthogroups construction [48].

Orthologous gene set construction

The orthogroups comprising genes from all 37 species were retrieved from all the identified orthogroups. KinFin v1.0 was used to increase the genes in one-to-one orthogroups that identified and extracted fuzzy one-to-one orthogroups among these retrieved orthogroups [49]. In cases where multiple genes were present for a single species in any orthogroup, the longest gene among them was selected as representative.

Phylogenetic tree construction

MAFFT v7.480 was used to discretely align all the identified fuzzy one-to-one orthogroups for construction of the phylogenetic tree [50]. 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 rapid hill climbing algorithm, to create the maximum likelihood-based phylogenetic tree (100 bootstrap values and amino acid substitution model 'PROTGAMMAAUTO') [51].

RESULTS

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

The *T. cordifolia* plant was brought from a plant nursery in Bhopal, Madhya Pradesh, India and the DNA and RNA extracted from leaf was used for sequencing. Two marker genes: ITS and MatK were amplified and sequenced at in-house sanger sequencing facility for the species identification (**Supplementary Text S1).** The sequenced reads of these marker genes were aligned using BLASTN that showed the highest identity (98.32% for ITS and 99.41% for MatK) with sequences of *T. cordifolia* available at NCBI database and confirmed the plant species as *T. cordifolia*. We generated 79.4 Gbp of genomic data and 34.7 Gbp of transcriptomic data for the genomic assembly and analysis of *T. cordifolia* **(Table 1 and Supplementary Tables S1-S2)**. The barcode sequences were removed from the raw reads and high-quality reads were selected for further analysis. The de novo assembly generated by Supernova assembler v.2.1.1 using 499.36 million raw reads [30], and the 'pseudohap2' style in Supernova mkoutput was implemented to assemble haplotype-phased genome. Since the genome size was not known for this plant, we performed the first genome size estimation for *T. cordifolia* using SGA-preqc processed linked reads, and the genome size was estimated to be 1.13 Gbp. Considering this genome size, the sequenced genomic data amounts to 70.2x genome coverage. After scaffolding, mis-assemblies rectification, gap-filling and polishing *T. cordifolia* genome assembly resulted in 1.01 Gbp assembly size (≥2,000 bp) as the final draft. The %GC of the assembled genome was 35.12%, and a total of 56,342 scaffolds were obtained having the N50 of 50.2 Kbp **(Table 1 and Supplementary Table S3)**. The BUSCO completeness was 78.9% in the final polished *T. cordifolia* genome assembly **(Supplementary Table S4)**. The *de novo* transcriptome assembly was constructed using Trinity v2.9.1 with strand-specific option and other default parameters using the processed paired end reads. The Trinity software assembled 2,764,154 bp of *de novo* transcriptome assembly, and a total of 8,208 transcripts were predicted in the transcriptome assembly.

Annotation of genome and gene set formation

The final polished genome assembly was analyzed by RepeatModeler v2.0.1 to construct a *de novo* repeat library consisting of 1,918 repeat families that were further clustered into 1,584 representative repeat family sequences. The repetitive sequences in *T. cordifolia* genome were predicted using RepeatMasker v4.1.0 and obtained 75.15% of repetitive sequences in this genome. Among these repetitive sequences, 73.75% were characterized as interspersed repeats comprising 32.12% retroelements (29.15% of LTR repeats), 2.48% DNA transposons and 39.14% unclassified repeats. The LTR repeats consisted of 25.38% Gypsy/DIRS1 and 3.22% Ty1/Copia elements. About 3.87% of the genome comprised of simple repeats found using TRF v4.09. The interspersed and simple repetitive sequences collectively comprised ~78% of the total genome. 392 hairpin miRNAs, 1,344 rRNAs and 2,186 tRNAs (decoding standard amino acids) were also predicted among the non-coding RNAs.

The soft-masked genome (generated using RepeatMasker) was used for the construction of gene set using the MAKER pipeline that employs *ab initio*-based gene prediction programs and evidence-based approaches for prediction of final gene model [41]. The de novo transcriptome assembly of *T. cordifolia* and protein sequences of its phylogenetically closer species *Beta vulgaris* (belonging to plant order Caryophyllales) were used as an empirical evidence in MAKER pipeline. The *ab initio* gene prediction, evidence-based alignments, and alignments polishing were achieved using AUGUSTUS v3.2.3, BLAST and Exonerate v2.2.0, respectively in the MAKER pipeline that predicted 19,730 coding sequences in the genome assembly [42, 43]. These coding genes were filtered based on length ≥150 bp and resulted in 19,474 coding gene sequences. The completeness of the coding gene set was assessed using BUSCO v4.1.4 embryophyte_odb10 database on the final gene set that showed 70% of the complete and fragmented BUSCOs to be present in the coding gene set.

Phylogenetic analysis of *T. cordifolia*

A total of 35 species were selected from the eudicot species accessible on Ensembl Plants release 48 by choosing one species from each offered genus, and included *Actinidia chinensis*, *Arabidopsis thaliana*, *Arabis alpina*, *Beta vulgaris*, *Brassica napus*, *Camelina sativa*, *Cannabis sativa* female, *Capsicum annuum*, *Citrullus lanatus*, *Citrus clementina*, *Coffea canephora*, *Corchorus capsularis*, *Cucumis melo*, *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*, *Rosa chinensis*, *Solanum tuberosum*, *Theobroma cacao* Matina 1-6, *Trifolium pratense*, *Vigna angularis*, and *Vitis vinifera*. Thus, the 36 selected species along

with one outgroup species (*Zea mays)*, and including *T. cordifolia* were used to detect 162,809 orthogroups using the protein sequences retrieved from 37 species.

The phylogenetic tree (maximum likelihood-based) was constructed using 454 fuzzy one-to-one orthogroups predicted by KinFin v1.0 across the 37 species. The missing or unknown values were predicted by aligning, concatenating and filtering all the fuzzy one-to-one orthogroups. The alignment data was filtered resulting in 422,034 alignment positions. The phylogenetic tree was constructed using this alignment data (filtered) of *T. cordifolia* along with the 35 dicot species available on Ensembl Plants release 48 and an outgroup species, *Zea mays* **(Figure 2)**. The resultant phylogenetic position of *T. cordifolia* was observed as a separate clade from all the other eudicots and monocot outgroup plausibly because the clade to which *T. cordifolia* belongs is considered as basal eudicots that diverged very early from the other eudicots [52, 53].

DISCUSSION

T. cordifolia is known to produce several phytochemicals as secondary metabolites such as alkaloids, terpenoids, tannins, phenolic compounds, steroids, flavonoids, phytosterols, volatile oils, etc. that are responsible for its diverse medicinal properties like anti-oxidant, immunomodulatory, anti-microbial, anti-viral, anticancer, anti-pyretic, anti-inflammatory, hepatoprotective, neuroprotective, anti-osteoporotic, anti-toxic, anti-diabetic, anti-arthritis, anti-ulcer, etc. [18, 54] **(Figure 1)**. Tinosporaside, tembetarine and phenolic compounds were suggested to responsible for its anti-inflammatory, anti-diabetic and anti-oxidant activities, respectively [55-57]. Likewise, a few other compounds from *Tinospora cordifolia* were found accountable for its various activities [58-62]. Being a medicinally important herb with therapeutic applications in multiple health conditions, the genome and transcriptome sequencing of *Tinospora cordifolia* was much needed to gain insights into its genetic information, transcriptome, and functional analysis of pathways associated with the secondary metabolites production responsible for its medicinal properties. This study reported the first draft genome assembly of *T. cordifolia* which is also the first genome from the genus *Tinospora* 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 plant. The linked read sequencing has recently emerged as a promising technology to significantly increase the contiguity in the genome assembly by increasing scaffold N50 and decreasing the number of scaffolds compared to short read technology, and was successfully used for the sequencing of *T. cordifolia* [63].

T. cordifolia belongs to order Ranunculales and our phylogenetic tree finds its position as a distinct branch separate from all the other eudicots. This could be due to the early-divergence of order Ranunculales from all the other core eudicots. Order Ranunculales is considered 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 [52, 64, 65], which was also observed in the case of *T. cordifolia* that showed early divergence from all other dicot species (**Figure 2**). Thus, the early divergence could be the reason for its distinct position relative to the other eudicots and monocot species. The correctness of this phylogeny is supported by the fact that the species belonging to the same clade shared common nodes, e.g., *Arabidopsis thaliana*, *Camelina sativa*, *Brassica napus,* and *Arabis alpina* shared the same clade because they belong to the same order Brassicales, and similarly *Gossypium raimondii*, *Theobroma cacao* Matina 1-6 and *Corchorus capsularis* belong to order Malvales and shared the same node in the phylogeny. Likewise, the members of order Fabales and order Solanales also shared their respective nodes.

CONCLUSION

Thus, as apparent from the above literature-based evidences on the medicinal and pharmaceutical properties of this plant, the availability of *T. cordifolia* genome will help in bridging 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 will act as a reference for the future species sequenced from its genus and family. It will also help in the genome-wide phylogenetic assessments as well as evolutionary analyses on this species. The knowledge of mechanisms and pathways involved in production of its numerous medicinally important secondary metabolites will help in better exploitation of these pathways and resultant metabolites for medicinal purposes and therapeutic applications.

LIST OF ABBREVIATIONS

COVID	Corona Virus Disease
Gbp	Giga base pair
RNA	Ribonucleic Acid
Mbp	Mega base pair
DNA	Deoxyribonucleic acid
CTAB	Cetyl trimethyl ammonium bromide

tRNA	Transfer RNA
rRNA	Ribosomal RNA
miRNA	micro-RNA
TRF	Tandem repeat finder
ITS	Internal transcribed spacer
MatK	Maturase K
BLAST	Basic Local Alignment Search Tool
BLASTN	Nucleotide BLAST
NCBI	National Center for Biotechnology Information
bp	Base pair
kbp	Kilo base pair
LTR	Long terminal repeats
DIRS1	Dictyostelium intermediate repeat sequence 1

COMPETING INTERESTS

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

AUTHORS' CONTRIBUTIONS

VKS perceived and coordinated the project and collected the plant sample. SM extracted the nucleic acids (DNA and RNA) from the collected sample, prepared the DNA and RNA samples for sequencing, and carried out the species identification assay. AC and VKS conceived the computational outline of the study. AC accomplished all the computational analysis presented in the study. SM, AC and VKS interpreted the results. SM and AC constructed the figures. SM, AC, TS and VKS wrote the manuscript. All the authors have read and approved the final version of the manuscript.

ACKNOWLEDGEMENT

SM and AC acknowledge Council of Scientific and Industrial Research (CSIR) for supporting research fellowships. TS thanks the Department of Science and Technology and IISc Bangalore for providing the KVPY scholarship. The authors also thank IISER Bhopal for providing the intramural research funds.

DATA AVAILABILITY

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

FIGURES

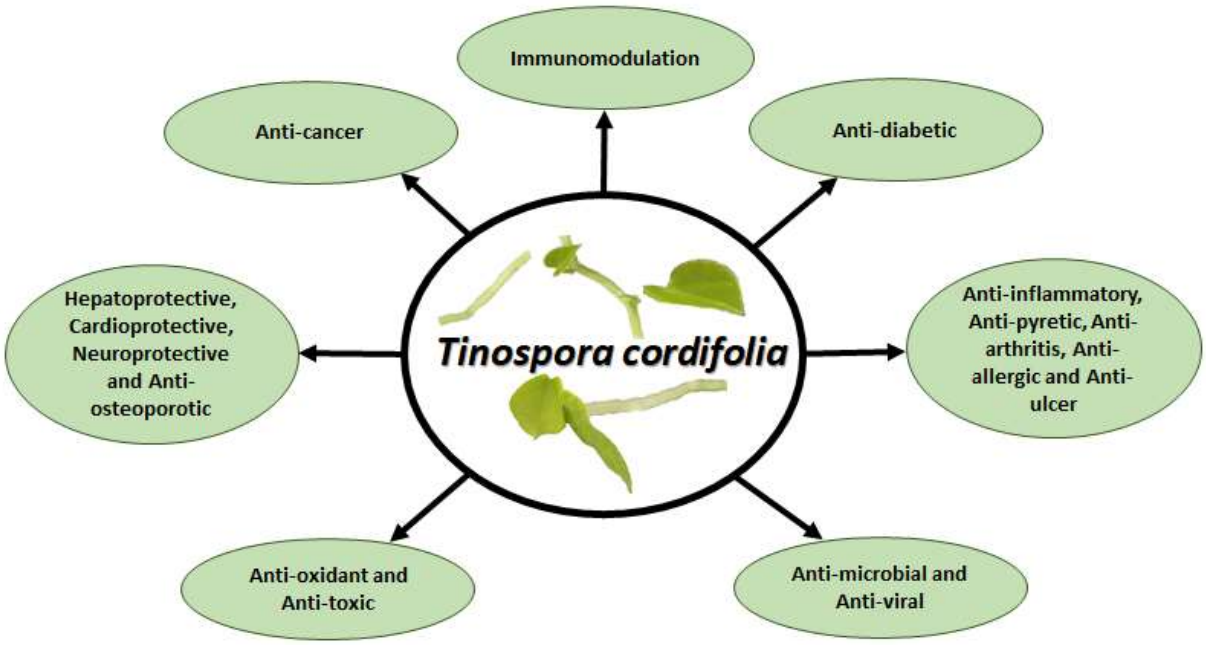

Figure 1. Medicinal properties of *Tinospora cordifolia*

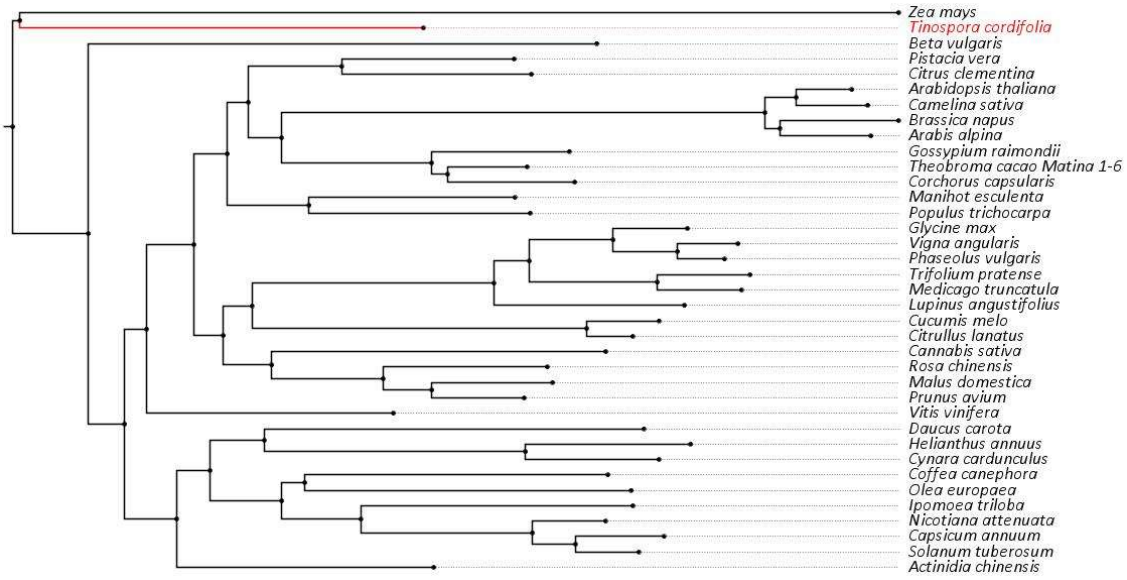

Figure 2. Phylogenetic tree describing the phylogenetic position of *Tinospora cordifolia* with other species

TABLE

Table 1. Summary of genome and transcriptome assemblies

Parameter	Value
Genomic data generated (Gbp)	79.4
Estimated genome size (Gbp)	1.13
Estimated genome coverage (x)	70.2
Size of draft genome assembly (Gbp)	1.01
N50 of genome assembly (Kbp)	50.2
GC content (%)	35.12
Percentage of repetitive sequences (%)	75.15
Total coding sequences	19,474
Transcriptome data generated (Gbp)	34.7
Size of transcriptome assembly (bp)	2,764,154
Total transcripts identified	8,208

REFERENCES

1. Thulasi, M.M. and R.K. Vasu, *The Menispermaceae family of plants and its action against infectious diseases: A review.* Mapana Journal of Sciences, 2020. **19**(2): p. 33-72.
2. Devi, G., *MEDICINAL PLANT: GILOY.* 2020.
3. Lian, L., et al., *Re-delimitation of Tinospora (Menispermaceae): Implications for character evolution and historical biogeography.* Taxon, 2019. **68**(5): p. 905-917.

4. Mittal, J., M.M. Sharma, and A. Batra, *Tinospora cordifolia: a multipurpose medicinal plant-A.* Journal of Medicinal Plants, 2014. **2**(2).
5. Sonavale, R., et al. *EFFECT OF EDAPHIC FACTORS ON MAJOR SECONDARY METABOLITES OF TINOSPORA CORDIFOLIA AND NEEM GUDUCHI WITH RESPECT TO THEIR IMMUNOMODULATORY EFFECT*. in *International Conference on Contemporary Research in Chemical and Life Sceince*. 2015.
6. Mishra, Y., et al., *In vivo and in vitro histological localization of endophytic fungi in Tinospora cordifolia (Willd.) Miers ex Hook F. & Thomas.* Journal of Applied Research on Medicinal and Aromatic Plants, 2015. **2**(1): p. 30-33.
7. Kapoor, N. and S. Saxena, *Endophytic fungi of Tinospora cordifolia with anti-gout properties.* 3 Biotech, 2018. **8**(6): p. 1-6.
8. Choudhry, N., et al., *Impact of seasons and dioecy on therapeutic phytoconstituents of Tinospora cordifolia, a Rasayana drug.* BioMed research international, 2014. **2014**.
9. Kalpesh, I. and J. Mohan, *Assessment of genetic diversity in medicinal climber of Tinospora cordifolia (Willd.) Miers (Menispermaceae) from Gujarat, India.* Asian Journal of Biotechnology, 2009. **1**(3): p. 93-103.
10. Paliwal, R., et al., *Development of genomic simple sequence repeats (g-SSR) markers in Tinospora cordifolia and their application in diversity analyses.* Plant Gene, 2016. **5**: p. 118-125.
11. Lade, S., et al., *Diversity in a widely distributed dioecious medicinal plant, Tinospora cordifolia (Willd.) Miers. ex. Hook F. and Thomas.* Current Science, 2018. **114**(7): p. 1520.
12. Singh, D. and P.K. Chaudhuri, *Chemistry and pharmacology of Tinospora cordifolia.* Natural product communications, 2017. **12**(2): p. 1934578X1701200240.
13. Sharma, P., et al., *The chemical constituents and diverse pharmacological importance of Tinospora cordifolia.* Heliyon, 2019. **5**(9): p. e02437.
14. Gahlaut, A., A. Gothwal, and R. Dabur, *TLC based analysis of allelopathic effects on tinosporoside contents in Tinospora cordifolia.* J. Chem. Pharm. Res, 2012. **4**: p. 3082-3088.
15. Choudhary, N., M. Siddiqui, and S. Khatoon, *Pharmacognostic evaluation of Tinospora cordifolia (Willd.) Miers and identification of biomarkers.* 2014.
16. Srinivasan, G., et al., *HPLC estimation of berberine in Tinospora cordifolia and Tinospora sinensis.* Indian journal of pharmaceutical sciences, 2008. **70**(1): p. 96.
17. Upadhyay, A.K., et al., *Tinospora cordifolia (Willd.) Hook. f. and Thoms.(Guduchi)–validation of the Ayurvedic pharmacology through experimental and clinical studies.* International journal of Ayurveda research, 2010. **1**(2): p. 112.
18. Saha, S. and S. Ghosh, *Tinospora cordifolia: One plant, many roles.* Ancient science of life, 2012. **31**(4): p. 151.
19. Sinha, K., et al., *Tinospora cordifolia (Guduchi), a reservoir plant for therapeutic applications: A Review.* 2004.
20. Devi, P., *Role of Tinospora cordifolia in metabolic health disorders: An updated review.* Himalayan Journal of Health Sciences, 2021: p. 15-23.
21. Chopra, D., B. Bhandari, and S. Dwivedi, *Beneficial role of Indian medicinal plants in COVID-19.* MGM Journal of Medical Sciences, 2021. **8**(2): p. 166.
22. Singh, R., et al., *De novo transcriptome sequencing facilitates genomic resource generation in Tinospora cordifolia.* Functional & integrative genomics, 2016. **16**(5): p. 581-591.
23. Abraham, A., *Chromosome number in Tinospora.* Current Science, 1942. **11**(7): p. 282-282.
24. Mathew, P. *Studies on the Menispermaceae*. in *Proceedings of the Indian Academy of Sciences-Section B*. 1958. Springer.
25. Santra, I., T. Halder, and B. Ghosh, *Research article Somatic and gametic chromosomal characterization with fluorescence banding of Giloy (Tinospora cordifolia): A berberine synthesizing important medicinal plant of India.* Caryologia. International Journal of Cytology, Cytosystematics and Cytogenetics, 2021.
26. Haque, M.A., I. Jantan, and S.N.A. Bukhari, *Tinospora species: An overview of their modulating effects on the immune system.* Journal of ethnopharmacology, 2017. **207**: p. 67-85.

27. Johnson, M.T., et al., *Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes.* Plos one, 2012. **7**(11): p. e50226.
28. Gaytán-Oyarzún, C.S.-H.a.J.C., *Two mini-preparation protocols to DNA extraction from plants with high polysaccharide and secondary metabolites.* African Journal of Biotechnology, 2006. **5 (20)**: p. 1864-1867.
29. Simpson, J.T. and R. Durbin, *Efficient de novo assembly of large genomes using compressed data structures.* Genome research, 2012. **22**(3): p. 549-556.
30. Weisenfeld, N.I., et al., *Direct determination of diploid genome sequences.* Genome research, 2017. **27**(5): p. 757-767.
31. Jackman, S.D., et al., *Tigmint: correcting assembly errors using linked reads from large molecules.* BMC bioinformatics, 2018. **19**(1): p. 1-10.
32. Zhang, S.V., L. Zhuo, and M.W. Hahn, *AGOUTI: improving genome assembly and annotation using transcriptome data.* GigaScience, 2016. **5**(1): p. s13742-016-0136-3.
33. Yeo, S., et al., *ARCS: scaffolding genome drafts with linked reads.* Bioinformatics, 2018. **34**(5): p. 725-731.
34. Paulino, D., et al., *Sealer: a scalable gap-closing application for finishing draft genomes.* BMC bioinformatics, 2015. **16**(1): p. 1-8.
35. Walker, B.J., et al., *Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement.* PloS one, 2014. **9**(11): p. e112963.
36. Simão, F.A., et al., *BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.* Bioinformatics, 2015. **31**(19): p. 3210-3212.
37. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.* Bioinformatics, 2014. **30**(15): p. 2114-2120.
38. Haas, B.J., et al., *De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis.* Nature protocols, 2013. **8**(8): p. 1494-1512.
39. Flynn, J.M., et al., *RepeatModeler2 for automated genomic discovery of transposable element families.* Proceedings of the National Academy of Sciences, 2020. **117**(17): p. 9451-9457.
40. Fu, L., et al., *CD-HIT: accelerated for clustering the next-generation sequencing data.* Bioinformatics, 2012. **28**(23): p. 3150-3152.
41. Campbell, M.S., et al., *Genome annotation and curation using MAKER and MAKER-P.* Current protocols in bioinformatics, 2014. **48**(1): p. 4.11. 1-4.11. 39.
42. Altschul, S.F., et al., *Basic local alignment search tool.* Journal of molecular biology, 1990. **215**(3): p. 403-410.
43. Stanke, M., et al., *AUGUSTUS: ab initio prediction of alternative transcripts.* Nucleic acids research, 2006. **34**(suppl_2): p. W435-W439.
44. Benson, G., *Tandem repeats finder: a program to analyze DNA sequences.* Nucleic acids research, 1999. **27**(2): p. 573-580.
45. Griffiths-Jones, S., et al., *miRBase: tools for microRNA genomics.* Nucleic acids research, 2007. **36**(suppl_1): p. D154-D158.
46. Chan, P.P. and T.M. Lowe, *tRNAscan-SE: searching for tRNA genes in genomic sequences*, in *Gene Prediction*2019, Springer. p. 1-14.
47. Bolser, D., et al., *Ensembl plants: integrating tools for visualizing, mining, and analyzing plant genomics data*, in *Plant bioinformatics*2016, Springer. p. 115-140.
48. Emms, D.M. and S. Kelly, *OrthoFinder: phylogenetic orthology inference for comparative genomics.* Genome biology, 2019. **20**(1): p. 1-14.
49. Laetsch, D.R. and M.L. Blaxter, *KinFin: software for Taxon-Aware analysis of clustered protein sequences.* G3: Genes, Genomes, Genetics, 2017. **7**(10): p. 3349-3357.
50. Katoh, K. and D.M. Standley, *MAFFT multiple sequence alignment software version 7: improvements in performance and usability.* Molecular biology and evolution, 2013. **30**(4): p. 772-780.
51. Stamatakis, A., *RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies.* Bioinformatics, 2014. **30**(9): p. 1312-1313.

52. Worberg, A., et al., *Phylogeny of basal eudicots: Insights from non-coding and rapidly evolving DNA.* Organisms Diversity & Evolution, 2007. **7**(1): p. 55-77.
53. Wang, W., et al., *Phylogeny and classification of Ranunculales: Evidence from four molecular loci and morphological data.* Perspectives in Plant Ecology, Evolution and Systematics, 2009. **11**(2): p. 81-110.
54. Begum, H.J. and V. Ramamurthy, *Evaluation of secondary metabolites and antioxidant activity of ethanolic leaves extract of Tinospora Cordifolia.*
55. Deepika, P.L., et al., *Analgesic, Antiinflammatory Activity of Tinospora Cordifolia (Guduchi) and Valeriana Wallichi (Tagara) In Albino Rats.* IOSR Journal of Pharmacy and Biological Sciences, 2016. **Volume 11**(Issue 2 Ver. IV): p. 18-22.
56. Khanal, P., et al., *Network pharmacology-based assessment to elucidate the molecular mechanism of anti-diabetic action of Tinospora cordifolia.* Clinical Phytoscience, 2019. **5**(1).
57. Balkrishna, A., M.H. Kumar, and A.K. Gupta, *COMPARATIVE ANALYSIS OF HPTLC, SECONDARY METABOLITES AND ANTIOXIDANT ACTIVITIES OF TINOSPORA CORDIFOLIA STEM POWDERS.* International Journal of Pharmaceutical Sciences and Research, 2016. **7**(10): p. 4263-4271.
58. Mishra, A., S. Kumar, and A.K. Pandey, *Scientific validation of the medicinal efficacy of Tinospora cordifolia.* The Scientific World Journal, 2013. **2013**.
59. Borse, S., et al., *Ayurveda botanicals in COVID-19 management: An in silico multi-target approach.* Plos one, 2021. **16**(6): p. e0248479.
60. Sharma, U., et al., *Immunomodulatory active compounds from Tinospora cordifolia.* Journal of ethnopharmacology, 2012. **141**(3): p. 918-926.
61. Palmieri, A., et al., *Berberine and Tinospora cordifolia exert a potential anticancer effect on colon cancer cells by acting on specific pathways.* International journal of immunopathology and pharmacology, 2019. **33**: p. 2058738419855567.
62. Bhutada, P., et al., *Anticonvulsant activity of berberine, an isoquinoline alkaloid in mice.* Epilepsy & Behavior, 2010. **18**(3): p. 207-210.
63. Mostovoy, Y., et al., *A hybrid approach for de novo human genome sequence assembly and phasing.* Nature methods, 2016. **13**(7): p. 587-590.
64. Soltis, D., *Angiosperm phylogeny inferred from 18S rDNA, rbcL, and atpB sequences.* Botanical Journal of the Linnean Society, 2000. **133**(4): p. 381-461.
65. Yang, L., et al., *Phylogenomic Insights into Deep Phylogeny of Angiosperms Based on Broad Nuclear Gene Sampling.* Plant Commun, 2020. **1**(2): p. 100027.