Physiol Mol Biol Plants (May 2023) 29(5):709–724
https://doi.org/10.1007/s12298-023-01312-w

RESEARCH ARTICLE

Decoding the complete chloroplast genome of *Cissus quadrangularis*: insights into molecular structure, comparative genome analysis and mining of mutational hotspot regions

Alok Senapati[1] · Bimal K. Chetri[2] · Sudip Mitra[2] · Rahul G. Shelke[1] · Latha Rangan[1]

Received: 14 December 2022 / Revised: 25 March 2023 / Accepted: 24 April 2023 / Published online: 14 May 2023

Abstract *Cissus quadrangularis* L., a member of the Vitaceae family, is an important medicinal plant with widespread application in Indian traditional medicines. *C. quadrangularis* L. whole chloroplast genome of 160,404 bp was assembled using a genome skimming approach from the whole genome library. The assembled chloroplast genome contained a large single-copy region (88,987 bp), a small single-copy region (18,621 bp), and pairs of inverted repeat regions (26,398 bp). It also comprised 133 genes, including 37 tRNAs, eight rRNAs, and 88 protein-coding genes. Aside from that, we annotated three genes *atp*H, *pet*B, and *psb*L, as well as one duplicated copy of the *ycf*1 gene in *C. quadrangularis* L. that had previously been missing from the annotation of compared *Cissus* chloroplast genomes. Five divergent hotspot regions such as *pet*A_*psb*J (0.1237), *rps*16_*trn*Q-UUG (0.0913), *psb*C_*trn*S-UGA (0.0847), *rps*15_*ycf*1 (0.0788), and *rps*2_*rpo*C2 (0.0788) were identified in the investigation that could aid in future species discrimination. Surprisingly, we found the overlapping genes *ycf*1 and *ndh*F on the IRb/SSC junction, rarely seen in angiosperms. The results of the phylogenetic study showed that the genomes of the *Cissus* species under study formed a single distinct clade. The detailed annotations given in this study could be useful in the future for genome annotations of *Cissus* species. The current findings of the study have the potential to serve as a useful resource for future research in the field of population genetics and the evolutionary relationships in the *Cissus* genus.

Keywords *Cissus quadrangularis* L. · Inverted repeats · Phylogeny · Simple sequence repeats · Tandem repeats

Introduction

Cissus quadrangularis L. is a perennial evergreen fleshy climber that belongs to the family Vitaceae. In common parlance, it is referred to as veldt grape, adamant creeper, and hadjod etc. The genus *Cissus*, formerly placed in the family Ampelidaceae, is thought to contain around 350 species that can be found in various habitats around the world (Onyeweaku et al. 2020). The family Vitaceae is comprised of 14 different genera, with *Cissus* being the most abundant genus (Onyeweaku et al. 2020). Members of this genus are found in South-East Asia, the Arabian Peninsula, tropical and subtropical Africa, and the Indian subcontinent. The plant grows in plain coastal regions, woodlands, and wastelands up to 500 m in elevation. It generates adventitious roots at nodes, which makes the process of vegetative propagation quite simple.

Since antiquity, people have used *C. quadrangularis* L. for therapeutic purposes in a number of traditional Ayurvedic medications. Its name, asthisamharaka, translates to "that which avoids the deterioration of bones" since it is supposed to help cure shattered bones and is used as a tonic and painkiller in Siddha medicine (Sivarajan and Balachandran 1994). Some compounds in *Cissus* act as glucocorticoid

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12298-023-01312-w.

✉ Rahul G. Shelke
rahul.sg98@gmail.com

✉ Latha Rangan
latha_rangan@yahoo.com; lrangan@iitg.ac.in

[1] Applied Biodiversity Laboratory, O Block, Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Guwahati, Assam 781039, India

[2] School of Agro and Rural Technology, Indian Institute of Technology Guwahati, Guwahati, Assam 781039, India

receptor antagonists and promote bone development and fracture healing due to their anabolic and/or androgenic properties (Tiwari et al. 2018). In addition, different parts of the plant have been utilized to treat conditions such as piles, blindness, tumours, muscular discomfort, asthma, chronic ulcers, epileptic fits, appetite loss, constipation, and sexually transmitted diseases (Sundaran et al. 2020). The stem of *C. quadrangularis* L. contains various important antioxidants, including vitamin C, carotenoids, calcium, and steroids (Dhanasekaran 2020). Previously, researchers isolated bioactive compounds from *C. quadrangularis* L. and its associated endophytes, including quercetin, daidzein quadrangularin-A, myristic acid, 2-phenylethanol, 2-naphthol, methyl palmitate, betulinaldehyde, and -amyrin acetate (Bafna et al. 2021; Purohit et al. 2022). Interestingly, there are no harmful consequences associated with its consumption (Sawangjit et al. 2017; Sundaran et al. 2020).

The amount of genomic data available online has grown dramatically in recent years as a result of the low cost and advancement of second and third-generation high-throughput DNA sequencing technologies. These large whole genome-sequencing projects are regularly used to infer phylogenetic relations, evolutionary trends, genetic diversity studies, etc. (Gichuki et al. 2019; Muraguri et al. 2020; Shelke et al. 2020). Due to the chloroplast genome's modest size (107–218 kb), it is considerably easier to sequence and assemble than the nuclear genome, resulting in the submission of 3823 chloroplast genomes to the Comprehensive Database of Chloroplast Genomes (http://www.gndu.ac.in/CpGDB) (Daniell et al. 2016). Chloroplast genomes differ from mitochondrial and nuclear genomes in a number of significant ways, including low nucleotide substitution rates, uniparental heredity, little homologous recombination, and a significantly preserved genomic makeup (Daniell et al. 2016). In contrast to nuclear genes, homologous genetic elements in chloroplast genomes evolved on average at a rate that was about one-third that of nuclear and three times that of mitochondrial genes, making it the perfect for molecular taxonomic investigations in resolving phylogenetic connections (Palmer et al. 1988). Molecular markers such as simple sequence repeats (SSRs) and single-nucleotide polymorphisms (SNPs), as well as intergenic spacers and introns, have a higher mutation rate and are thus useful for population genetics studies (Muraguri et al. 2020; Shelke et al. 2020; Li et al. 2021). Since the majority of medicinal plants are rare species with only a limited amount of information available to confirm their identity, chloroplast DNA is widely used in the field of herbal medicine for species identification and evolutionary research (Li et al. 2021). Based on chloroplast genome sequencing, it is possible to design species- or genus-specific barcodes that are extremely reliable, durable, and affordable (Li et al. 2021; Shelke and Rangan 2022). Additionally, using the complete chloroplast genome as a super barcode in situations when universal or single barcodes are unable to offer sufficient information for species discrimination opens up new possibilities for resolving the intricate connections between taxa (Krawczyk et al. 2018; Zhang et al. 2021).

The current study involved the assembly of the *C. quadrangularis* L. chloroplast genome and its comparison to the chloroplast genomes of five other *Cissus* sp. Furthermore, we performed pairwise comparisons of the protein-coding sequences from these species to identify candidate genes that had been subjected to either positive or negative selection. Through our data analysis, we found a plethora of species-specific repeat sequences and barcodes that could be used as markers. The evolutionary connections among the members of the Vitoideae subfamily were also examined using the chloroplast genome data. In general, the findings of our study provide in-depth genome annotations of *C. quadrangularis* L. and contribute to resources that can be used in future evolutionary research for *Cissus* and other related species in the Vitoideae subfamily.

Materials and methods

C. quardrangularis L. chloroplast genome assembly and annotation

In this investigation, we retrieved SRA read of *C. quadrangularis* L. (SRR8573652) from the SRA database submitted by Gichuki et al. (2019). The reads were generated on Novogene's Illumina HiSeq X Ten platform as paired-end 2X150 bp reads with 400–450 bp insert size (Gichuki et al. 2019). The downloaded raw Illumina reads were quality inspected using the FastQC (0.11.8) tool before being de novo assembled using the NOVOplasty assembler (version 3.7) with a *rbc*L sequence as a seed (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Using the NCBI BLASTN software, the resulting chloroplast genome assembly sequence was checked against the non-redundant (nr) database. After that, the GeSeq tool was used to annotate the whole chloroplast genome (Tillich et al. 2017). Eventually, the OGDRAWv1.2 tool was used to create the circular genome map for *C. quadrangularis* L. (Dierckxsens et al. 2017). The whole chloroplast genome sequence of *C. quadrangularis* L. was submitted to GenBank along with its gene annotation (OP414588).

Structural comparison of *Cissus* chloroplast genomes

We have included six genomes of *Cissus* species in the current investigation for structural comparison of chloroplast genomes: *C. quadrangularis* L. (OP414588), *C. antarctica* Vent. (NC_061724.1), *C. discolor* Blume (NC_061723.1),

C. microcarpa Vahl (NC_061722.1), *C. trifoliata* (L.) L. (NC_061720.1), and *C. tuberosa* Moc. & Sessé ex DC. (NC_061719.1). Synteny studies of *Cissus* chloroplast genomes were carried out through the MAUVE tool (https://darlinglab.org/mauve/mauve.html). While the borders between inverted repeats (IR) and short single copy (SSC) regions in these species were also compared and studied to assess the evolution of genes found in IR boundary areas.

Tandem repeats and simple sequence repeats mining

The six *Cissus* chloroplast genomes were examined for the presence of forward, reverse, palindromic, and complementary repeats using the REPuter software. The parameters used to identify repeats were more than 30 bp and a hamming distance of three (https://bibiserv.cebitec.uni-bielefeld.de/reputer/). MISA software v1.0.6 was used to identify SSRs in the six compared genomes using the following parameters: 10 for mono-, 5 for di-, 4 for tri-, and 3 for each of the following nucleotides: penta-, hexa-, septa-, octa-, nona-, and deca- (https://pgrc.ipk-gatersleben.de/misa/misa.html).

Identification of divergent hotspot regions

C. quadrangularis L. was compared using the mVISTA program with five *Cissus* genomes such as by keeping the *C. quadrangularis* L. annotation as a reference (http://genome.lbl.gov/vista/mvista/submit.shtml). The DnaSP v5.10 tool was utilized to determine the level of nucleotide diversity of genic and intergenic regions separately present in the compared chloroplast genomes (http://www.ub.edu/dnasp/).

Characterization of substitution rates and synonymous codon usage analysis

In six comparable species of the genus *Cissus*, the evolutionary forces operating on common protein-coding genes were examined. The rate of synonymous to nonsynonymous nucleotide substitution (Ka/Ks) was calculated with the help of KaKs_calculator 3.0 (Zhang 2022). Common protein-coding genes from the six *Cissus* species were selected for codon usage bias analysis. Average guanine-cytosine (GC) content and codon usage bias were determined for all protein-coding genes by nucleotide composition and relative synonymous codon usage (RSCU) analysis using MEGA X programme (https://www.megasoftware.net/).

Phylogenetic analyses

The newly assembled *C. quadrangularis* L. chloroplast genomes, as well as other Vitoideae genomes, were used in the study to construct a phylogenetic tree. We chose one chloroplast genome of *Leea guineensis* G. Don as an out-group species belonging to the Leeoideae subfamily of Vitaceae. The compared sequences were first aligned using MAUVE to find the locally collinear blocks (LCBs) in the HomBlocks pipeline (Bi et al. 2018). Then the model of substitution was figured out with the assistance of the MEGA-X tool (version 10.2.4), and it was done so in accordance with what the Akaike information criterion (AIC) indicated. Finally, the phylogenetic tree was built using the maximum-likelihood method (ML) in the MEGA X programme with 500 bootstraps (https://www.megasoftware.net/). The resulting phylogenetic tree was then depicted visually via the iTOL server (https://itol.embl.de/).

Results

C. quadrangularis L. chloroplast genome assembly and annotation features

Using the NOVOplasty assembler with 18,978,233 raw paired-end reads, the whole chloroplast genome of *C. quadrangularis* L. was successfully assembled. The size of the *C. quadrangularis* L. chloroplast genome was 160,404 bp in length (Fig. 1). And its average mean coverage depth was around 2793 X. The newly assembled chloroplast genome was highly comparable in sequence to the *Cissus* chloroplast genomes included in the Genbank database. The whole chloroplast genome sequence of *C. quadrangularis* L. was submitted to GenBank with the accession number OP414588. The chloroplast genome of *C. quadrangularis* L. is arranged in the usual quadripartite pattern, with two IR regions (26,398 bp) that are separated by LSC (88,987 bp) and SSC (18,621 bp) regions. The average percentage of GC in the whole chloroplast genome was 37.2%, which was 42.8% in IR, 35.07% in LSC, and 31.3% in SSC.

In the *C. quadrangularis* L. chloroplast genome, there are a total of 133 genes; 88 of these genes code for proteins, while the remaining 37 and 8 code for tRNA and rRNA, respectively (Table 1). The genome was revealed to have duplications in four rRNA, seven tRNA and nine protein-coding genes. In total, 18 genes were observed in IRb and IRa regions, respectively. The genes that were found duplicated in IR regions were seven protein-coding (*ndh*B, *rpl*2, *rpl*23, *rps*7, *rps*12, *ycf*2, and *ycf*15), four rRNA (rRNA 4.5, rRNA 5, rRNA 16, and rRNA 23), and seven tRNA (*trn*A-ACG, *trn*A-UGC, *trn*M-CAU, *trn*I-GAU, *trn*L-CAA, *trn*N-GUU, and *trn*V-GAC) genes. Two genes (*clp*P1 and *ycf*3) have double introns, and the rest twenty genes (*atp*F, *rpo*C1, *pet*B, *pet*D, *rpl*2, *rpl*2_copy, *rpl*16, *ndh*A, *ndh*B, *ndh*B_copy, *rps*12, *rps*12_copy, *rps*16, *rps*16_copy, *trnA*-UGC, *trnA*-UGC_copy, *trnI*-GAU, *trnI*-GAU_copy, *trnK*-UUU and *trnV*-UAC) include single intron. The LSC and IRa

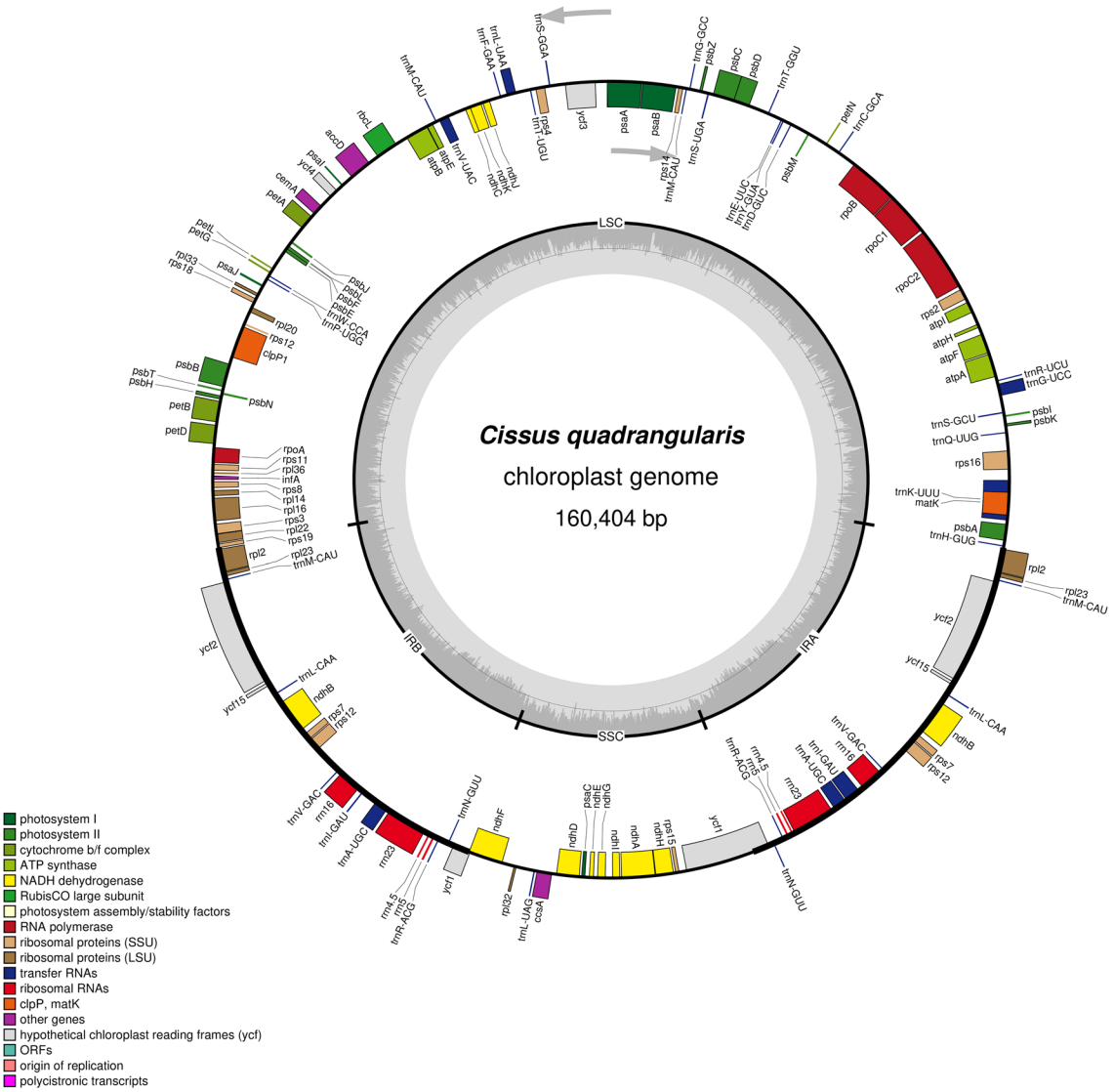


Fig. 1 Circular map of *C. quadrangularis* whole chloroplast genome along with gene annotations. Genes shown outside the outer circle are transcribed anti-clockwise, and those inside the circle are transcribed clockwise. Genes belonging to different functional groups are colour coded. The grey colour inside the circle indicates the GC content and the lighter grey depicts the AT content

area contains the trans-spliced *rps*12 gene. The *mat*K gene is located in the biggest intron of the gene *trn*K-UUU. No gene loss was observed in the *C. quadrangularis* L. genome.

Structural comparison of *Cissus* chloroplast genomes

The Mauve alignment of *Cissus* chloroplast genomes found no major inversions; with the exception of one inversion that was around 250 bp long in *C. trifoliata* (L.) L. and *C. tuberosa* Moc. & Sessé ex DC. (Fig. 2). Remaining compared genomes showed no inversion. The observed inversion that encompassed the intergenic region can be found in the LSC region of *C. trifoliata* (L.) L. and *C. tuberosa* Moc. & Sessé ex DC. (Online Resource 1).

In all six studied genomes, the IR/SSC boundary included two genes, *ycf*1 and *ndh*F (Fig. 3). The contraction and expansion of the *ycf*1 gene at the LSC/IRa boundary in six *Cissus* species showed distinct patterns occupying the varying number of bp in SSC and IRa regions, with a larger part in SSC region in all cases. The *ycf*1 gene was found in the *Cissus* species at positions 4447 to 4568 bp in the SSC region and 965 to 1082 bp in the IRa region. The *ndh*F gene was discovered at the junction of IRb/SSC in *C. quadrangularis* L., *C. tuberosa* Moc. & Sessé ex DC., *C. trifoliata* (L.) L., and *C. microcarpa* Vahl. It extended 2–38 bp in the IRb area and the bulk remained in the SSC region. On the other

Table 1 Annotated genes and their functions in *C. quandrangularis* chloroplast genome

Gene category	Group of genes	Name of genes				
Self-replication	rRNA genes	*rrn*16 [D]	*rrn*23 [D]	*rrn*4.5 [D]	*rrn*5 [D]	
	tRNA genes	*trn*A-*UGC* [D*]	*trn*C-*GCA*	*trn*D-*GUC*	*trn*E-*UUC*	*trn*F-*GAA*
		*trn*G-*GCC*	*trn*G-*UCC*	*trn*H-*GUG*		
		*trn*I-*GAU* [D*]	*trn*K-*UUU* [*]	*trn*L-*CAA* [D]	*trn*L-*UAA*	*trn*L-*UAG*
		*trn*M-*CAU* [#]	*trn*N-*GUU* [D]	*trn*P-*UGG*	*trn*Q-*UUG*	*trn*R-*ACG* [D]
		*trn*R-*UCU*	*trn*S-*GCU*	*trn*S-*GGA*	*trn*S-*UGA*	*trn*T-*GGU*
		*trn*T-*UGU*	*trn*V-*GAC* [D]	*trn*V-*UAC* [*]	*trn*W-*CCA*	*trn*Y-*GUA*
	Large subunit of ribosome (LSU)	*rpl*2 [D*]	*rpl*14	*rpl*16 [*]	*rpl*20	*rpl*22
		*rpl*23 [D]	*rpl*32	*rpl*33	*rpl*36	
	Small subunit of ribosome (SSU)	*rps*2	*rps*3	*rps*4	*rps*7 [D]	*rps*8
		*rps*11	*rps*12 [D*]	*rps*14	*rps*15	*rps*16 [*]
		*rps*18	*rps*19 [D]			
	RNA polymerase	*rpo*A	*rpo*B	*rpo*C1 [*]	*rpo*C2	
	Translational initiation factor	*inf*A				
Photosynthetic genes	Photosystem I	*psa*A	*psa*B	*psa*C	*psa*I	*psa*J
	Photosystem II	*psb*A	*psb*B	*psb*C	*psb*D	*psb*E
		*psb*F	*psb*H	*psb*I	*psb*J	*psb*K
		*psb*L	*psb*M	*psb*N/*pbf*1	*psb*T	*psb*Z
	Cytochrome b/f complex	*pet*A	*pet*B [*]	*pet*D [*]	*pet*G	*pet*L
		*pet*N				
	ATP synthase	*atp*A	*atp*B	*atp*E	*atp*F [*]	*atp*H
		*atp*I				
	ATP-dependent protease (p subunit)	*clpP*1 [**]				
	RubisCO large subunit	*rbcL*				
	NADH dehydrogenase	*ndh*A [*]	*ndh*B [D*]	*ndh*C	*ndh*D	*ndh*E
		*ndh*F	*ndh*G	*ndh*H	*ndh*I	*ndh*J
		*ndh*K				
Other genes	Maturase	*mat*K				
	Protein in envelop membrane	*cem*A				
	Acetyl-CoA-carboxylase subunit	*acc*D				
	Cytochrome synthesis gene (C-type)	*ccs*A				
Genes with unknown function	Hypothetical reading frames for chloroplast	*ycf*1 [D]	*ycf*2 [D]	*ycf*3 [**]	*ycf*4	*Ycf*15 [D]

[*]Genes having one intron

[**]Gene having two introns

[D]Gene having duplicated copies

[#]Gene having quadruplicated copies

hand, the *ndh*F of *C. discolor* Blume and *C. antarctica* Vent. was seen in the SSC region.

In all species, the *rps*19 gene extended 23–45 bp in the IRb area across the LSC/IRb junction except *C. trifoliata* (L.) L., which was present entirely at 17 bp inside the LSC region from the border. The IRa and IRb regions included the full *rpl*2 gene. The *trn*H gene was entirely localized in the LSC area just outside the IRa/LSC border in all the species that were investigated. The findings demonstrated that these six species of *Cissus* shared conserved chloroplast genomes, with only minor changes at the junctions.

Tandem repeats and simple sequence repeats mining

There were 32 tandem repeats in the chloroplast genome of *C. quadrangularis* L., 16 of which were forward repeats, 15 were palindromic, single complement, and there were no reverse repeats (Fig. 4a). In general, the total number

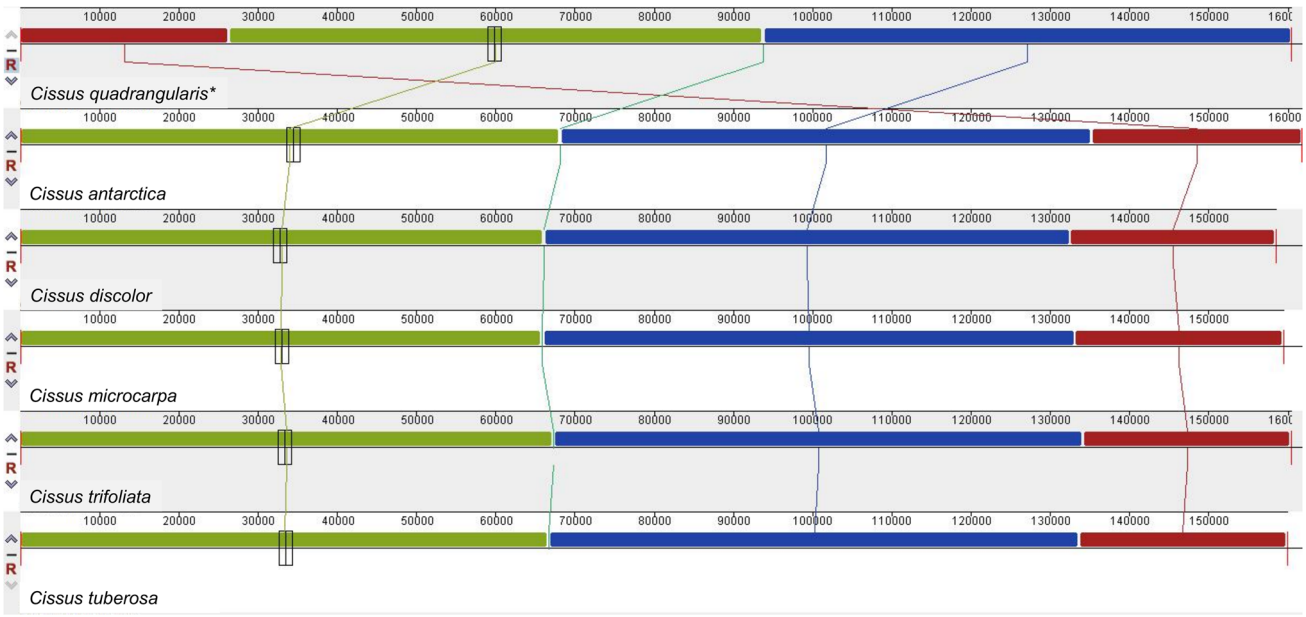


Fig. 2 MAUVE alignment of *C. quadrangularis* and five chloroplast genomes of *Cissus* species. The gene arrangement in chloroplast genomes is represented by different coloured local collinear blocks. (* denotes target species)

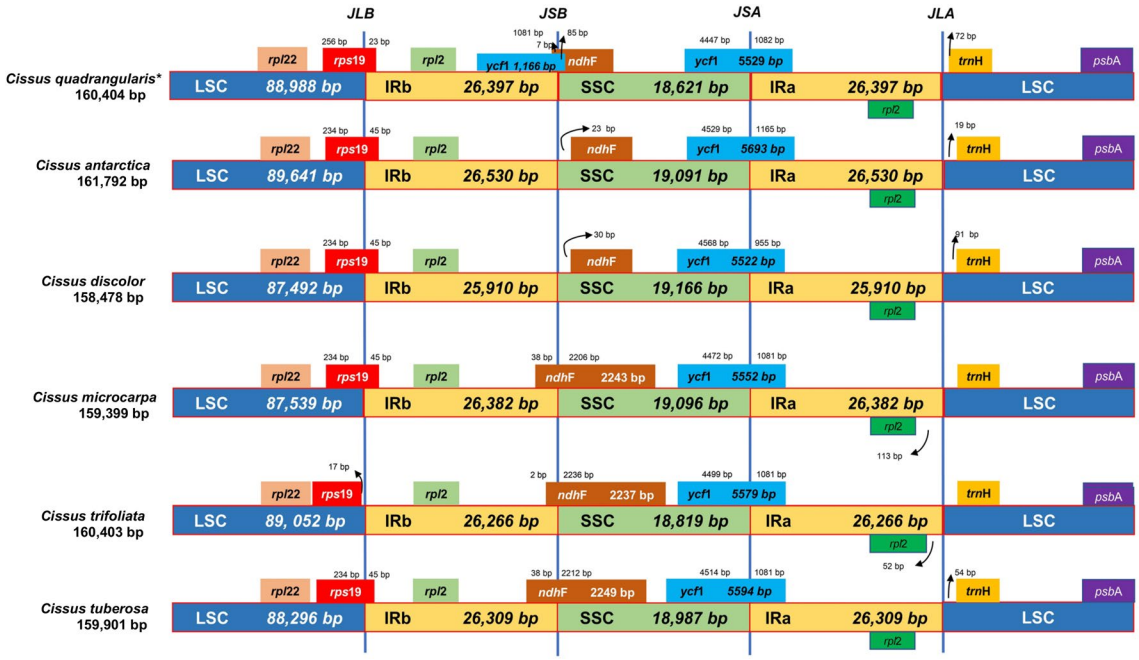


Fig. 3 Comparisons of LSC, SSC, and IR region borders among six chloroplast genomes of *Cissus* species. Colour boxes present the genes. (* denotes target species)

of tandem repeats ranged from 32 in *C. quadrangularis* L. to 49 in *C. trifoliata* (L.) L. (Fig. 4b). The majority of tandem repeats were found to be of the forward type across all six analysed genomes, with maximum forward repeats in *C. antarctica* Vent. (27) and the minimum in *C. discolor* Blume and *C. quadrangularis* L. (16). On the other hand, the highest numbers of palindromic repeats were recorded in *C. antarctica* Vent. and *C. trifoliata* (L.) L. (20 repeats) and the lowest in *C. microcarpa* Vahl (13 repeats). *C. tuberosa* Moc. & Sessé ex DC. had the most reverse repeats (9) and no reverse repeat was observed in *C. quadrangularis* L. The complimentary repetitions were absent in both *C. antarctica*

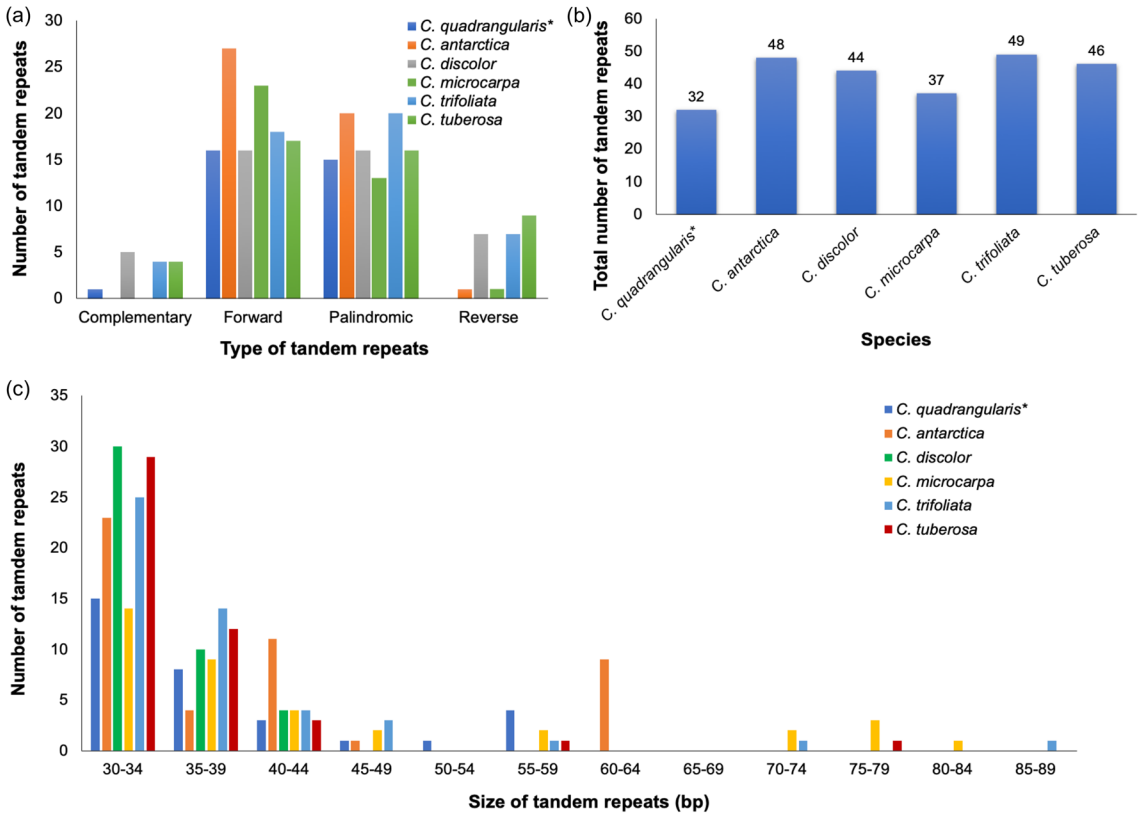


Fig. 4 Comparison of tandem repeats of six chloroplast genomes of *Cissus* species, **a** Number of tandem repeats by repeat types; **b** Number of tandem repeats by sequence length. (* denotes target species)

Vent. and *C. microcarpa* Vahl, while *C. discolor* Blume has a maximum of five repeats. Tandem repeats of length 30–34 bp emerged as the most frequent repetitions, followed by 35–39 bp and 40–44 bp fragments (Fig. 4c). Tandem repetitions within 45–49 bp and 55–59 bp ranges were uncommon in all species except *C. quadrangularis* L., and only *C. discolor* Blume had nine repetitions within 60–64 bp range. Additionally, a few repetitions can be observed in the range of 70–74 bp, 75–79 bp, 80–84 bp, and 85–89 bp, respectively.

In the chloroplast genome of *C. quadrangularis* L., we found 70 simple and 9 compound SSRs (Fig. 5a). Out of 70 simple SSRs, 40 (50.6%) were mononucleotides, 18 (22.8%) were dinucleotides, 7 (8.9%) were tri-nucleotides, and 14 (14.1%) were tetra-nucleotides (Fig. 5b). Comparative analysis of SSRs found that *C. antarctica* Vent., *C. discolor* Blume, *C. microcarpa* Vahl, *C. trifoliata* (L.) L., and *C. tuberosa* Moc. & Sessé ex DC. contained 67, 81, 72, 95, and 72 SSRs, respectively. *C. microcarpa* Vahl was discovered to have the highest percentage of mononucleotide SSRs (73.6%), followed by *C. discolor* Blume (67.9%) and *C. trifoliata* (L.) L. (67.37%). The least mononucleotide percentage was recorded in *C. quadrangularis* L. (50.6%), followed by *C. tuberosa* Moc. & Sessé ex DC. (59.72%). In all cases, A/T motifs were abundant, whereas C/G repeat motifs were rare. *C. trifoliata* (L.) L. had the most A/T rich repeats (64), followed by *C. discolor* Blume (54), and *C. antarctica* Vent. along with *C. quadrangularis* L. had the fewest (38). Similarly, the AT/AT repeats were higher in *C. quadrangularis* L. (18) than *C. discolor* Blume (15) and least in *C. microcarpa* Vahl (7). AAT/ATT repeats were the most common trinucleotide repeats, whereas others were rarely present (Fig. 5c). *C. trifoliata* (L.) L. had a higher SSR density (0.6 SSR/kb) than the other species. However, *C. antarctica* Vent. had the lowest SSR density (0.41 SSR/kb) (Fig. 5d).

Comparative genome analysis and DNA diversity

Comparing the nucleotide diversity in coding and non-coding regions identified hotspot regions in the *Cissus* chloroplast genomes. Protein-coding genes' nucleotide diversity (*Pi*) ranged from 0.0013 (*ndh*B) to 0.0579 (*psa*I) (Fig. 6a). The average diversity in protein-coding regions was calculated to be 0.0191. The five most diverged coding regions in *Cissus* genomes were *psa*I (0.0579), *psa*J (0.0486), *rpl*32 (0.0439), *ycf*1 (0.0437), and *mat*K (0.03721). The least

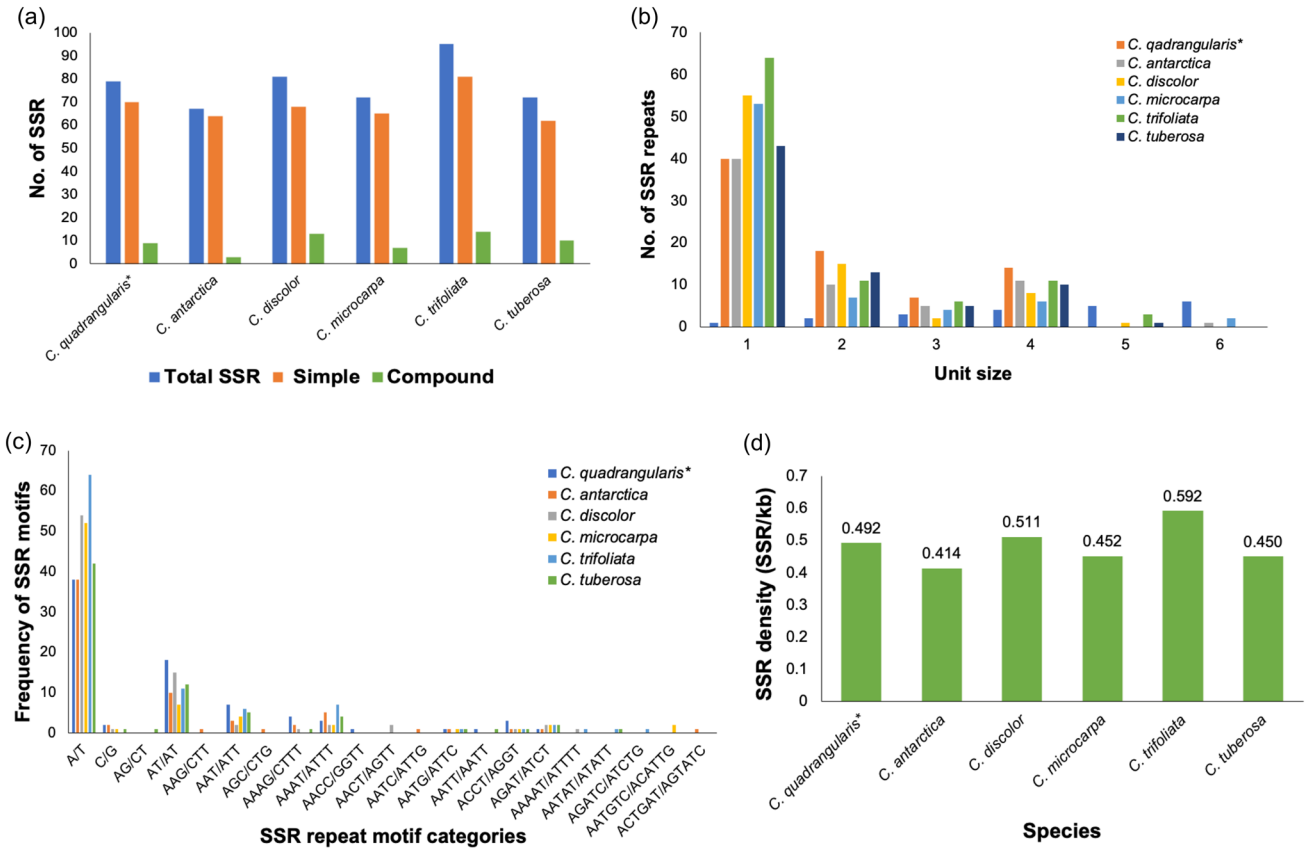


Fig. 5 Comparison of SSR repeats between *C. quadrangularis* and five chloroplast genomes of *Cissus* species. **a** Number of SSR repeats in regular and compound formation; **b** The total number of SSR repeat types; **c** The number of SSR motifs, and **d** SSR density. (* denotes target species)

diverse coding regions were *ndh*B (0.0013), *rps*7 (0.0027), *pet*N (0.0040), *ycf*15 (0.0048), and *psb*N (0.0050).

Non-coding region nucleotide diversity (*Pi*) was found to be higher in *Cissus* genomes than in coding regions, ranging from 0.0034 (*rrn*16_copy2_*trn*V_GAC_copy2) to 0.1237 (*pet*A_*psb*J) (Fig. 6b). It was discovered that non-coding regions had an average diversity of 0.0391. The five regions that had the maximum nucleotide diversity were as follows: *pet*A_*psb*J (0.1237), *rps*16_*trn*Q-UUG (0.0913), *psb*C_*trn*S-UGA (0.0847), *rps*15_*ycf*1 (0.0788), and *rps*2_*rpo*C2 (0.0788). Non-coding regions that had the lowest *Pi* value were as follows: *rrn*16_copy2_*trn*V-GAC_copy2 (0.0034), *rps*7_*trn*V-GAC (0.0036), *trn*V-GAC_*rrn*16 (0.0039), *trn*V-GAC_copy2_*rps*7_copy2 (0.0040), and *rrn*16_*trn*I-GAU (0.0044). The investigation of the chloroplast genome of *Cissus* performed with mVISTA produced remarkably similar results with DNA diversity, with a significant drop in the percentage of similarity at non-coding regions and high conservation in coding regions (Fig. 7).

Ka/Ks ratios in *Cissus* genomes

The Ka/Ks ratio was determined by examining 82 common genes from compared *Cissus* genomes. Out of 82 genes, 12 had a 0 or no Ka/Ks ratio, 70 genes had a Ka/Ks ratio, wherein three genes had more than one ratio (*atp*F, *rpl*2, and *rpl*2_copy) and sixty-seven had less than one ratio (Fig. 8). The Ka/Ks ratio varied between the six *Cissus* species, ranging from 0.009 (*rps*11) to 1.077. (*rpl*2). Purifying selection is therefore acting on the vast majority of genes with a Ka/Ks ratio lower than one. When the Ka/Ks ratio is greater than 1, the selection is working positively on the gene under study; when it is equal to 1, the selection is neutral. There were a few genes with Ka/Ks ratios that were either zero or non-existent; these values were obtained when Ks values were exceedingly low or when there was no substitution in matched sequences.

Synonymous codon usage

The six *Cissus* chloroplast protein-coding genes were discovered to preferentially use A/T bases or A/U-ending

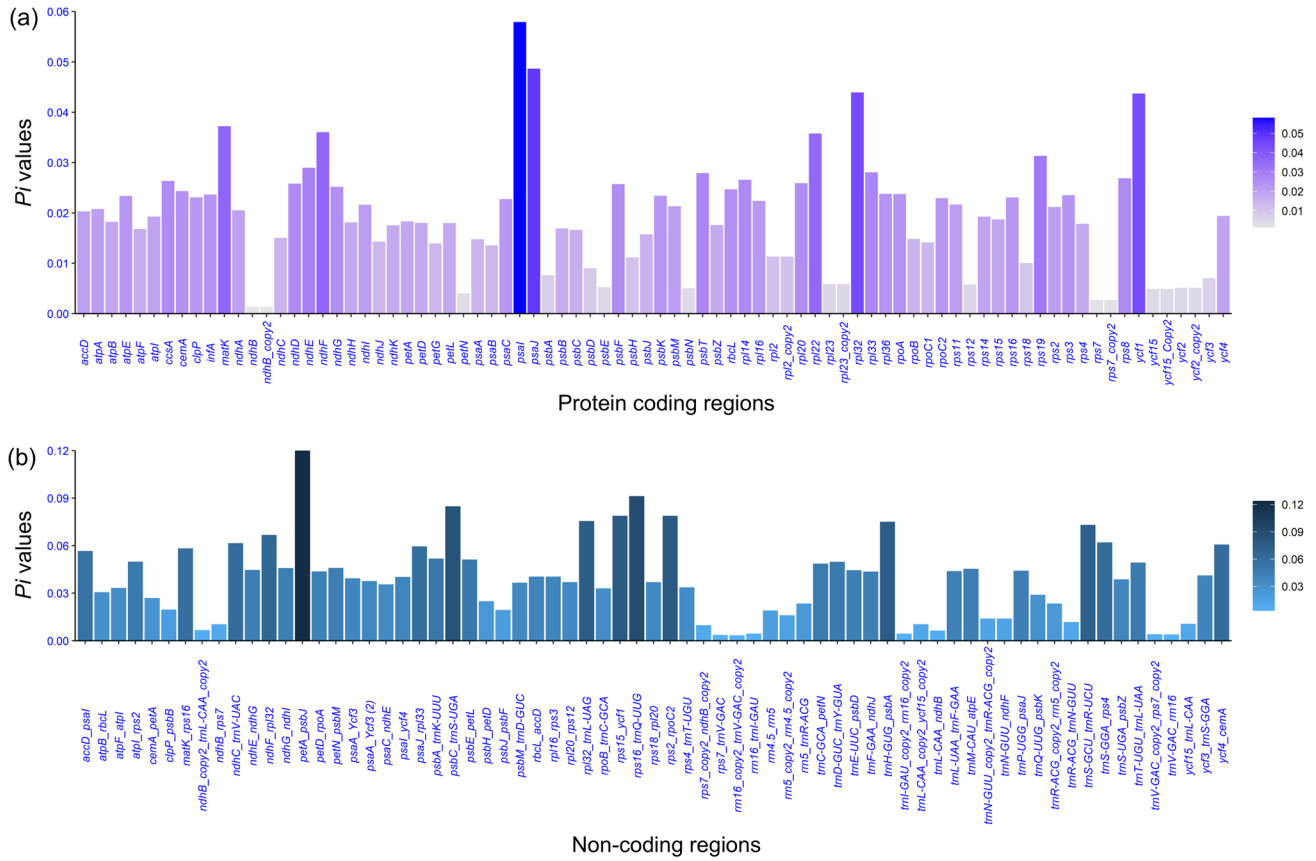
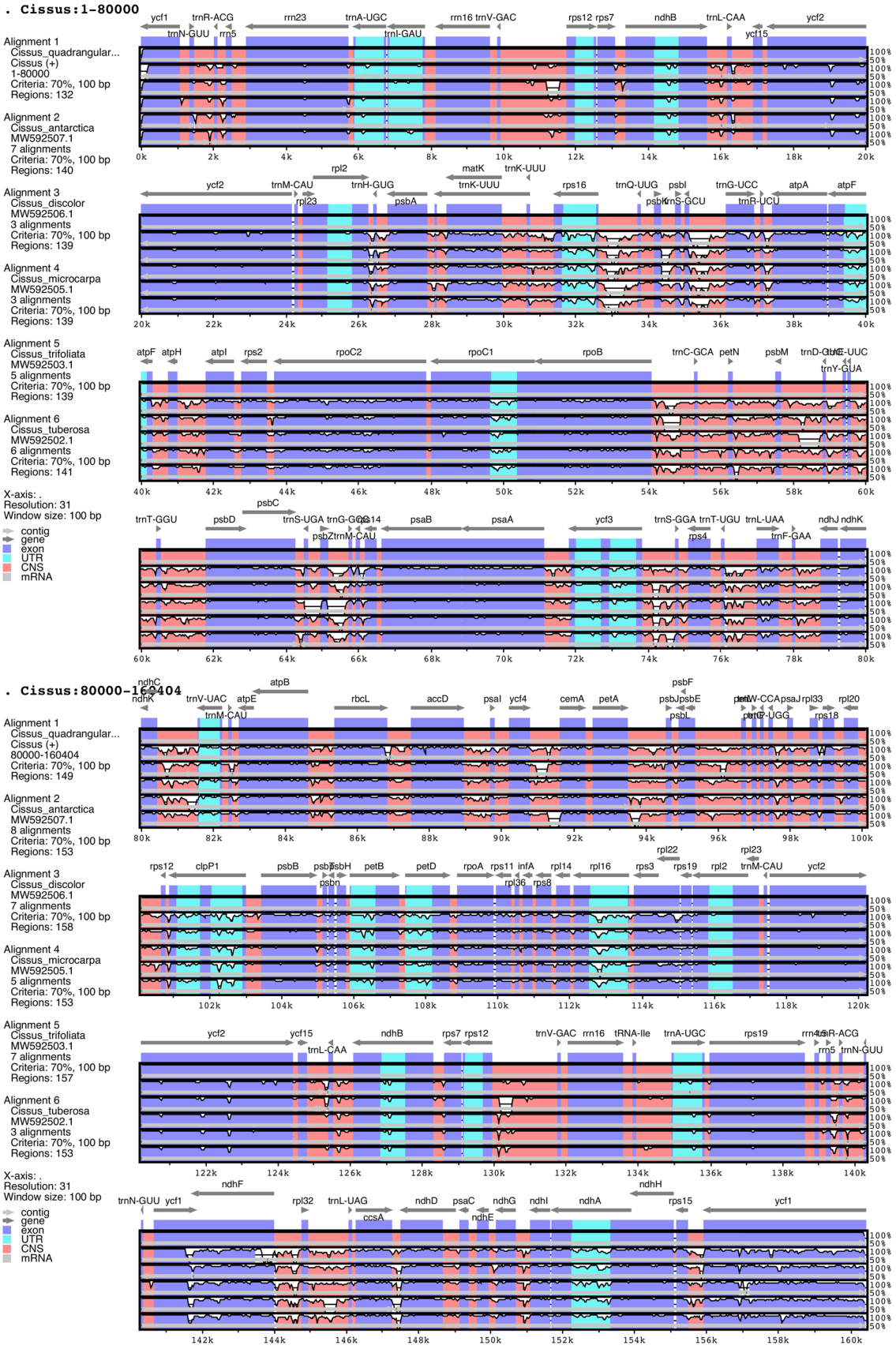


Fig. 6 Nucleotide diversity (*Pi*) values of *Cissus* species. The nucleotide diversity of, **a** protein-coding regions; **b** non-coding regions

codons over GC bases at all three codon positions; as a result, the overall average GC content of codons was less than 40% (Online Resource 2). In the compared genomes, the GC content in the first position (GC1) ranged from 36.2% to 39.5% in *C. antarctica* Vent and *C. microcarpa* Vahl, respectively. While the GC content in the second position (GC2) was lowest in *C. discolor* Blume (36.4%) and highest in *C. tuberosa* Moc. & Sessé ex DC. (39.7%). The highest concentration of GC for the third position (GC3) was found in *C. antarctica* Vent (38.8%), whereas the lowest concentration was found in *C. tuberosa* Moc. & Sessé ex DC. (36.4%). Mutations are reported to have an essential role in codon usage bias (Li et al. 2017). There were 64 codons including three stop codons, present across the six *Cissus* species encoding 20 amino acids. *C. quadrangularis* L. and *C. antarctica* Vent each had 25,256 codons, *C. microcarpa* Vahl and *C. tuberosa* Moc. & Sessé ex DC. each had 25,174 codons, *C. discolor* Blume had 25,010 codons, and *C. trifoliata* (L.) L. had 25,420 codons. In the studied genomes, A and U bases were prevalent at the third position of codons (RCSU > 1). Analyses revealed that leucine was the most prevalent amino acid (9.01%—10.85%), followed by tryptophan (1.89%—2.11%) and serine (1.73%—2.34%). Tryptophan and methionine had only one codon each, so they had no preference for codon usage (Fig. 9). In contrast, the remaining amino acids contained multiple synonymous codons. For leucine and arginine, there were six synonymous codons, with biasness towards UUA and AGA respectively. While valine, serine, proline, threonine, alanine, and glycine had four synonymous codons each (Online Resource 3). Out of the six *Cissus* species, *C. discolor* Blume used 28 codons, *C. antarctica* Vent and *C. tuberosa* Moc. & Sessé ex DC. used 30 codons frequently (RSCU > 1), and the other three species used 29 codons more often than expected at equilibrium (RSCU > 1). The start codons AUG and UGG, which code for the amino acids methionine and tryptophan, were used equally often by all species of *Cissus* (RSCU = 1).

Phylogenetic analysis

The ML phylogenetic tree was created using the chloroplast genomes of 24 different species as a super barcode. Unaligned sequences were removed from the final alignment. The GTR + G + I model was determined to be the best-fit model for further phylogenetic tree construction based on the results of an alignment of 1,37,163 bp of nucleotides


. Cissus:1-80000
Alignment 1
Cissus_quadrangular...
Cissus (+)
1-80000
Criteria: 70%, 100 bp
Regions: 132
Alignment 2
Cissus_antarctica
MW592507.1
7 alignments
Criteria: 70%, 100 bp
Regions: 140
Alignment 3
Cissus_discolor
MW592506.1
3 alignments
Criteria: 70%, 100 bp
Regions: 139
Alignment 4
Cissus_microcarpa
MW592505.1
3 alignments
Criteria: 70%, 100 bp
Regions: 139
Alignment 5
Cissus_trifoliata
MW592503.1
5 alignments
Criteria: 70%, 100 bp
Regions: 139
Alignment 6
Cissus_tuberosa
MW592502.1
6 alignments
Criteria: 70%, 100 bp
Regions: 141
X-axis: .
Resolution: 31
Window size: 100 bp
contig
gene
exon
UTR
CNS
mRNA
. Cissus:80000-160404
Alignment 1
Cissus_quadrangular...
Cissus (+)
80000-160404
Criteria: 70%, 100 bp
Regions: 149
Alignment 2
Cissus_antarctica
MW592507.1
8 alignments
Criteria: 70%, 100 bp
Regions: 153
Alignment 3
Cissus_discolor
MW592506.1
7 alignments
Criteria: 70%, 100 bp
Regions: 158
Alignment 4
Cissus_microcarpa
MW592505.1
5 alignments
Criteria: 70%, 100 bp
Regions: 153
Alignment 5
Cissus_trifoliata
MW592503.1
7 alignments
Criteria: 70%, 100 bp
Regions: 157
Alignment 6
Cissus_tuberosa
MW592502.1
3 alignments
Criteria: 70%, 100 bp
Regions: 153
X-axis: .
Resolution: 31
Window size: 100 bp
contig
gene
exon
UTR
CNS
mRNA

◂**Fig. 7** Chloroplast genome alignment between *C. quadrangularis* and five chloroplast genomes using the mVISTA tool with *C. quadrangularis* as a reference. The grey arrows represent the direction of gene transcription. The y-axis depicts the percentage identity ranging between 50 and 100%. (* denotes target species)

from the 24 whole genome sequences. The ML analysis tree was resolved, and the majority of nodes received maximum support (Fig. 10). The Leeoideae, a subfamily of the Vitaceae family served as an out-group. The phylogenetic tree recovered the clustering of *Cissus* species in one group clade that included *C. quadrangularis* L., *C. discolor* Blume, *C. tuberosa* Moc. & Sessé ex DC. and *C. trifoliata* (L.) L., sharing a common node with the *Tetrastigma* and *Cyphostemma* genera. Within the *Cissus* group, the species *C. quadrangularis* L. was found to have a close association with *C. discolor* Blume. On the other hand, the *Vitis* and *Amplelopsis* genus formed their own separate group but shared a node.

Discussion

The chloroplast genome of *C. quadrangularis* L. is circular and exhibits a quadripartite genome structure consistent with the angiosperm chloroplast genomes (Shelke and Rangan 2022). The chloroplast genome of *C. quadrangularis* L. reported in this study was 160,404 bp in length, which is within predicated range of angiosperm chloroplast genomes (107–218 kb) (Daniell et al. 2016). Among the *Cissus* genomes, the highest genome size of 161,792 bp was recorded in *C. antarctica* Vent. and lowest 158,478 bp in *C. discolor* Blume.

The chloroplast genome of *C. quadrangularis* L. encoded 37 tRNA genes and 8 rRNA genes, similar to the majority of angiosperms, including *Cissus* species (Daniell et al. 2016; Shelke and Rangan 2022). The fact that previous *Cissus* genomes were missing four genes was intriguing. If they were lost, were their genes still in the nuclear or mitochondrial genomes? While looking for an answer, we were able to identify four extra protein-coding genes, including *atp*H, *pet*B, and *psb*L, as well as one duplicated copy of the *ycf*1 gene, which had previously gone unnoticed in the annotation of compared *Cissus* chloroplast genomes. The genes in question are, in fact, present in the *Cissus* chloroplast genomes; however, it's possible that previous reports failed to include them in their annotation database. Through BLASTX analysis, we confirmed the presence of these genes in the compared genomes in the current study. We conclude that the *Cissus* genomes have not lost any protein-coding genes. When introns are located in certain regions, they have a significant impact on the expression of genes. The chloroplast genomes of typical angiosperms contain around 23 introns in tRNA and protein-coding genes (Wang et al. 2022). From the 133 genes found in the *C. quadrangularis* L. genome, 22 were found to have introns which included 20 genes with a single intron and two genes (*clp*P1 and *ycf*3) with two introns. Intriguingly, we noticed two copies of the *ycf*15 gene in *Cissus* genomes, a gene that is typically absent from many angiosperm genomes (Jin et al. 2020; Wang et al. 2021; Shelke and Rangan 2022). However, the presence of the *ycf*15 gene was observed in many members of the Vitaceae family, with its size ranging from 234 to 488 bp. It is also present in full copies in the plastome genomes of *Nicotiana*, *Epifagus*, *Cuscuta* etc. (Schmitz-Linneweber et al. 2001).

Many chloroplast genome inversions have been fully characterized, revealing that this process is critical in genome rearrangement (Kim et al. 2005; Jin et al. 2020; Jiang et al. 2022). We were able to observe an inversion of approximately 250 bp in length in both *C. trifoliata* (L.) L. and *C. tuberosa* Moc. & Sessé ex DC. A similar, albeit smaller, inversion in an intergenic region was documented in *Mesua ferrea* L. in the past (Shelke and Rangan 2022). On the other hand, larger inversions of up to 50 kb and even more have been reported previously in some angiosperm species (Jin et al. 2020; Tian et al. 2021; Wu et al. 2022). The inversion appears to be conserved because it is located in the same place between the *pet*A gene and the *psb*J gene in both *C. trifoliata* (L.) L. and *C. tuberosa* Moc. & Sessé ex DC. genomes. Such inversion plays a crucial role in plant evolution as a mechanism for the establishment and preservation of interspecific differentiation and is likely to be related to the formation of plant groupings (Kim et al. 2005). Overall the *Cissus* chloroplast genomes had identical gene orders, and most of the chloroplast region's sequence identities were also similar across species. These findings imply that the chloroplast genomes of *Cissus* are substantially conserved.

The IR junctions in chloroplast genomes are typically the site of structural changes. The expansion and contraction of IR junctions in higher plants usually contribute to these changes. It was discovered that the diversity of the genome's size and structure was influenced by the length of the IR regions (Jin et al. 2020). On the IR junctions, there was no evidence of gene loss in any of the six compared species of *Cissus*, with the exception of the *ycf*1 gene, which was present on the IRb/LSC border of *C. quadrangularis* L. but absent in other species. There is a high degree of homogeneity between LSC/IRs and SSC/IRs borders in *Cissus* and angiosperm chloroplast genomes, with most of these boundaries located in *rps*19 or *ycf*1 (Downie and Jansen 2015; Alzahrani et al. 2020; Shelke and Rangan 2022). In *C. quadrangularis* L., it was found that the *ndh*F gene and the *ycf*1 gene both crossed the IRb/SSC junction by 7 and 85 bp, respectively. In this event, the two genes overlapped each other and shared a 92 bp coding region, as documented in some species of *Andrographis*, *Barleria* and *Justicia* genera

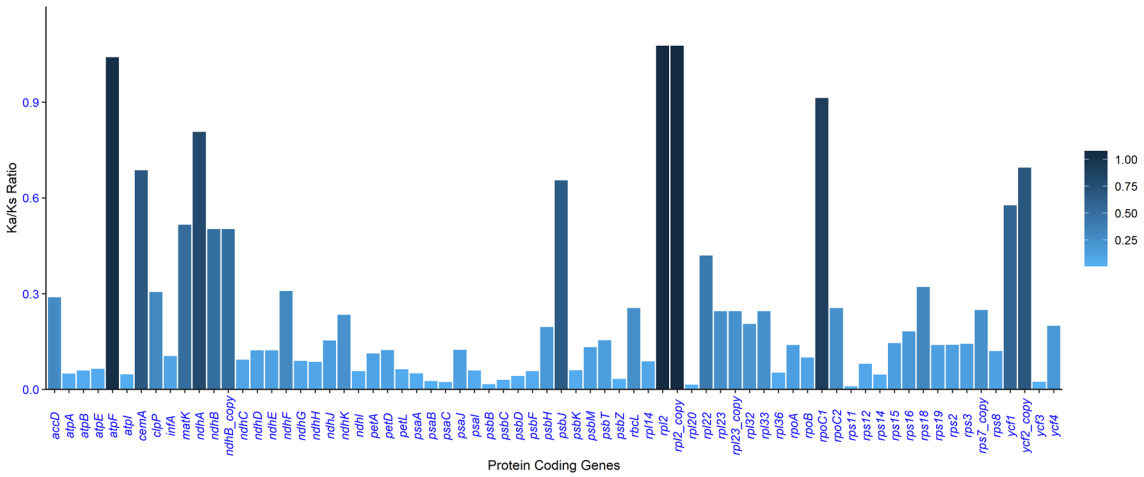


Fig. 8 Comparison of K_a/K_s ratio generated in protein-coding genes aligned from *Cissus* species genomes

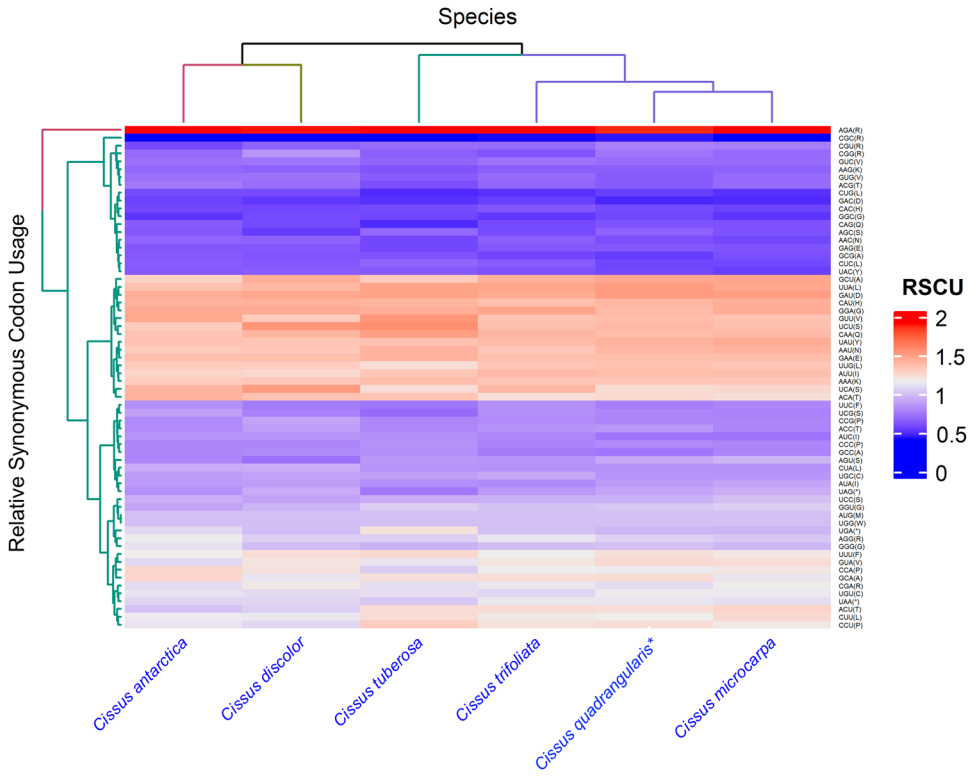


Fig. 9 Heat map of relative synonymous codon usage (RSCU) values among six *Cissus* species blue to red colour indicates low to high RSCU values of codons. (* denotes target species)

(Alzahrani et al. 2020). We also observed such overlapping sequences for *atp*E_*atp*B, *ndh*K_*ndh*C and *psb*D_*psb*C genes in *Cissus* chloroplast and other angiosperm genomes (Rossini et al. 2021). In contrast to *rps*19 and *ndh*F genes, which were occasionally found crossing the IR regions in some compared genomes, *rpl*2 and *trn*H were found to be located entirely within their respective regions. The *ycf*1 gene, which lies at the IRa/SSC intersection, entered the IRa area in the compared genomes with varying lengths ranging from 965 to 1082 bp, which is a regular phenomenon in angiosperms (Alzahrani et al. 2020; Li et al. 2022; Xu et al. 2022). These instances of *ycf*1 crossing the IRb/SSC junction were found to occur frequently in the genomes of numerous angiosperm species (Jin et al. 2020; Shelke and Rangan 2022). Here, there was no noticeable difference in IR length between *Cissus* genomes. It was found that the length of the *ycf*1 gene ranged from 5522 bp (*C. discolor* Blume) to 5693 bp (*C. antarctica* Vent.) across the analyzed genomes. The *ycf*1 gene was found in most of the Vitaceae species and varied from 3617 to 5811 bp. Despite its presence in

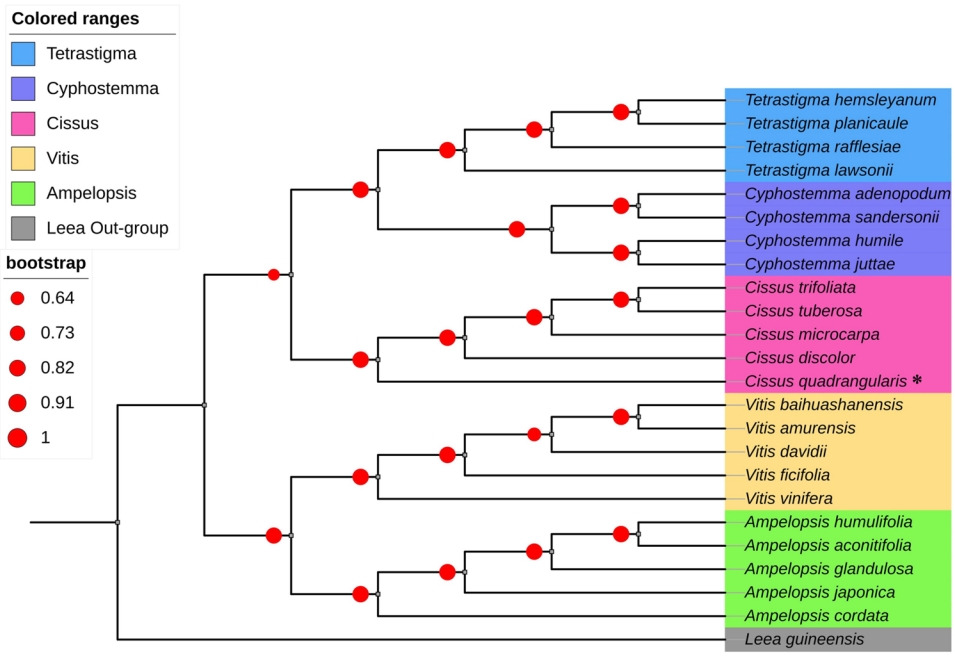


Fig. 10 Phylogenetic tree was drawn based on the whole chloroplast genomes from 24 Vitaceae species constructed through the maximum likelihood method. (* denotes target species)

many angiosperm species, it has been confirmed to be absent in some members of the Poaceae family (Goremykin et al. 2005; Guisinger et al. 2011).

The chloroplast genome contains a variety of repeated sequences, including SSRs, homo-polymeric repeats, and long repeats. These sequences serve as a source of variations for genome evolution and rearrangement (Das et al. 2018; Shelke and Rangan 2020; Shelke et al. 2020). Therefore, repeat elements may be used to make molecular markers for analyzing population genetic structure and evolutionary relationships between distinct species (Shelke and Das 2015; Shelke and Rangan 2019; Shelke et al. 2020; Wu et al. 2021). Long tandem repeats are frequently found in the chloroplast genome in association with intermolecular recombination to produce sequence diversity (Guo et al. 2021). The number of tandem repetitions in all of the *Cissus* genomes ranged from 32 in *C. quadrangularis* L. to 49 in *C. trifoliata* (L.) L. Despite having the largest genome size, *C. antarctica* Vent. appears to have the fewest number of SSRs of all the genomes studied, which is consistent with previous research in *M. ferrea* L. (Khan et al. 2019; Shelke and Rangan 2022). Therefore, the sequence variability caused by the tandem repeats could aid in the design of species-specific markers.

SSRs are also tandem repeats that are found in the genome in significantly greater abundances than long tandem repeats (Khan et al. 2019; Wang et al. 2021; Shelke and Rangan 2022). SSRs ranged from 67 in *C. antarctica* Vent. to 95 in *C. trifoliata* (L.) L. in the genomes of the *Cissus* species. It is interesting to note that the genome of *C. trifoliata* (L.) L. was found to contain the highest number of tandem and SSR repeats among those that were compared. Similarly, although the genomes of *C. trifoliata* (L.) L. and *C. quadrangularis* L. are roughly the same size, but there was a substantial difference in the number of repeats. This fact is supported further by the observation that *C. trifoliata* (L.) L. has the highest SSR density (0.6 SSR/kb). Generally, it is difficult to compare the results of repeat populations obtained from different publications because different authors use a wide range of mining parameters for their repeat analysis. Such comparisons do not provide sufficient information regarding comparative studies in such circumstances. The emergence of a significant number of SSRs in the genome is most likely the result of adaptations that were developed over the course of evolution in response to specific factors of the surrounding environment (Shelke et al. 2020). These new SSR mutations can lead to variations that are species-specific and can assist in the designing of barcodes as a marker for species delimitation.

A significant level of sequence similarity was observed among the *Cissus* chloroplast genomes. The non-coding areas were more diverged than the coding sections, and the IR regions were more preserved than the LSC and SSC regions, which is consistent with other angiosperms (Guan et al. 2022; Shelke and Rangan 2022). Then, comparable outcomes were also seen in DNA diversity analysis, where it was shown that coding regions were more conserved than non-coding areas (Rossini et al. 2021; Guan et al. 2022). The *Cissus* chloroplast genome's most varied coding and non-coding regions have been identified in the current study. However, some parts of the chloroplast genome change at a significantly faster rate and hence meet the criteria for being

used as a DNA barcode (Senapati et al. 2021; Guan et al. 2022). In light of the fact that many universal DNA barcodes are often insufficient to resolve species-level distinctions in many angiosperms due to low levels of nucleotide diversity. Therefore such genus-specific DNA barcodes have the potential to be utilised in the future for discriminating *Cissus* species (Li et al. 2021).

Nucleotide substitution rate changes have been linked to selective pressures on protein-coding regions of genomes over evolutionary time (Shidhi et al. 2021). The *Cissus* genus was found to be under purifying selection on the vast majority of its protein-coding genes, a trend that has been outlined frequently in many plant species, indicating that the overall conserved nature of plastid genes in angiosperms (Shidhi et al. 2021; Shelke and Rangan 2022). As a result, plastid function is under less selective pressure leading to purifying selection on genes producing proteins for DNA maintenance and expression that may eventually be lowered.

Codon usage is extremely important in expressing genetic information correctly. The codons used by all six *Cissus* species were the nearly same, with 61 amino acid codons, three termination codons (UAA, UAG, and UGA). In addition to preferred codon usage, the number and type of codons that code for the 20 amino acids varied among *Cissus* species and similar observations were also reported in *Physalis* species (Feng et al. 2020). Most codons that were preferred for encoding amino acids had either an A or a T(U) as their third nucleotide. Out of all the biased codons (RSCU > 1) in *Cissus* species 45% had A and 48% had U at third position. Such observations have previously been made in a number of angiosperms, including the Poaceae, Asteraceae, Euphorbiaceae, and *Oryza* species families (Zhang et al. 2012; Nie et al. 2014; Ma et al. 2020; Chakraborty et al. 2020). Codon usage frequencies differ in the chloroplast genomes of *Cissus* species, which may be related to hydrophilicity, base substitution, natural selection, tRNA abundance, gene length, mRNA secondary structure, and random genetic drift (Feng et al. 2020; Li et al. 2022). The genetic makeup of a population, as well as selection in favor of enhanced translation, are said to be determinants of the trend in codon usage (Lee et al. 2019). There was also evidence from early studies that monocots favored GC-rich codons and dicots favored AT-rich ones (Wang and Roossinck 2006).

Chloroplast genomes have been found to be a very helpful tool for understanding the evolutionary connections of many angiosperm plants (Jin et al. 2020; Li et al. 2021; Shelke and Rangan 2022). In an attempt to provide light on the evolutionary position of the *Cissus* species, an ML tree was built utilizing the whole chloroplast genome sequences of 5 *Cissus* and 18 additional species from the Vitoideae subfamily. The phylogenetic grouping of *Cissus* genomes demonstrated that the chloroplast genome is sufficient to establish the relationships between *Cissus* species and the remainder of the Vitoideae subfamily with significant bootstrap support. Within the *Cissus* clade, *C. quadrangularis* L. was found closely related to *C. discolor* Blume. The phylogenetic analysis utilizing different DNA barcodes, as well as the study based on the entire chloroplast genomes, are in agreement with our findings (Zhang et al. 2016a; Lu et al. 2018; Wen et al. 2018). In the current study, the *Cissus* genus was discovered to be the sister to the Cayratieae tribe, which comprises the *Tetrastigma* species. This relationship was also brought up in the earlier exploration that was conducted to understand the phylogeny of the Vitoideae subfamily (Zhang et al. 2016b; Wen et al. 2018). However, we have removed the *C. antarctica* Vent. species from the current phylogenetic tree since it was difficult to resolve within the *Cissus* clade. A similar type of resolution difficulties involving *C. antarctica* Vent. were also documented in earlier studies, and for that reason, the *Cissus* is considered a polyphyletic clade as well (Rossetto et al. 2002, 2007; Liu et al. 2013). In the future, we believe that resolving this evolutionary uncertainty will be made easier by increasing the number of *Cissus* species, as well as with the inclusion of organellar and nuclear markers.

Conclusion

In summary, we assembled a whole chloroplast genome of *C. quadrangularis* L. spanning 160,404 bp from low-coverage whole genome data using a genome skimming approach. Our investigation has identified the previously missing annotations of three genes *atp*H, *pet*B, and *psb*L, as well as an additional copy of the *ycf*1 gene in *C. quadrangularis* L., which was previously reported to be absent from *Cissus* genomes. *C. quadrangularis* L. has the second largest genome size after *C. antarctica* Vent. among the genomes studied. Comparative genomics identifies the genetic differences between the chloroplast genomes of the different *Cissus* species in terms of the number of repeat sequences and the level of DNA diversity found in the non-coding sequences. The present knowledge of the evolutionary lineage of *C. quadrangularis* L. in the Vitoideae subfamily is supported by the phylogenetic study focused on 24 complete chloroplast genomes. The comprehensive annotation of the genome discussed in this study could be useful as a reference for the annotation of *Cissus* chloroplast genomes in future research.

Acknowledgements AS expresses gratitude to the Ministry of Education, Government of India, for supporting the student fellowship. BKC is grateful to Sherubtse College, Royal University of Bhutan, for providing the fellowship. Authors would like to thank the IIT Guwahati, Assam, India, for providing an institutional computational facility. No funding was received to assist with the preparation of this manuscript.

Author contributions AS, and RGS assembled and annotated the chloroplast genomes. AS, RGS, SM and BKC analyzed and interpreted the data. LR and RGS conceived and designed the study. LR conceptualized, supervised and revised the manuscript. All authors have read and approved the final manuscript.

Declarations

Conflict of interests Authors have no competing interests to declare that are relevant to the content of this article.

References

Alzahrani DA, Yaradua SS, Yaradua SS et al (2020) Complete chloroplast genome sequence of *Barleria prionitis*, comparative chloroplast genomics and phylogenetic relationships among Acanthoideae. BMC Genom 21:1–20. https://doi.org/10.1186/s12864-020-06798-2

Bafna PS, Patil PH, Maru SK, Mutha RE (2021) *Cissus quadrangularis* L.: A comprehensive multidisciplinary review. J Ethnopharmacol 279:114355. https://doi.org/10.1016/j.jep.2021.114355

Bi G, Mao Y, Xing Q, Cao M (2018) HomBlocks: A multiple-alignment construction pipeline for organelle phylogenomics based on locally collinear block searching. Genomics 110:18–22. https://doi.org/10.1016/j.ygeno.2017.08.001

Chakraborty S, Yengkhom S, Uddin A (2020) Analysis ofcodon usage bias of chloroplast genes in *Oryza* species. Planta 252:67. https://doi.org/10.1007/s00425-020-03470-7

Daniell H, Lin CS, Yu M, Chang WJ (2016) Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol 17:1–29. https://doi.org/10.1186/s13059-016-1004-2

Das R, Shelke RG, Rangan L, Mitra S (2018) Estimation of nuclear genome size and characterization of Ty1-copia like LTR retrotransposon in *Mesua ferrea* L. J Plant Biochem Biotechnol 27:478–487. https://doi.org/10.1007/s13562-018-0457-7

Dhanasekaran S (2020) Phytochemical characteristics of aerial part of *Cissus quadrangularis* (L.) and its in-vitro inhibitory activity against leukemic cells and antioxidant properties. Saudi J Biol Sci 27:1302–1309. https://doi.org/10.1016/j.sjbs.2020.01.005

Dierckxsens N, Mardulyn P, Smits G (2017) NOVOPlasty: *De novo* assembly of organelle genomes from whole genome data. Nucleic Acids Res. https://doi.org/10.1093/nar/gkw955

Downie SR, Jansen RK (2015) A comparative analysis of whole plastid genomes from the Apiales: expansion and contraction of the inverted repeat, mitochondrial to plastid transfer of DNA, and identification of highly divergent noncoding regions. Syst Bot 40:336–351. https://doi.org/10.1600/036364415X686620

Feng S, Zheng K, Jiao K et al (2020) Complete chloroplast genomes of four *Physalis* species (Solanaceae): lights into genome structure, comparative analysis, and phylogenetic relationships. BMC Plant Biol 20:242. https://doi.org/10.1186/s12870-020-02429-w

Gichuki DK, Ma L, Zhu Z et al (2019) Genome size, chromosome number determination, and analysis of the repetitive elements in *Cissus quadrangularis*. PeerJ 7:e8201. https://doi.org/10.7717/peerj.8201

Goremykin VV, Holland B, Hirsch-Ernst KI, Hellwig FH (2005) Analysis of *Acorus calamus* chloroplast genome and its phylogenetic implications. Mol Biol Evol 22:1813–1822. https://doi.org/10.1093/molbev/msi173

Guan Y, Liu W, Duan B et al (2022) The first complete chloroplast genome of *Vicatia thibetica* de Boiss.: genome features, comparative analysis, and phylogenetic relationships. Physiol Mol Biol Plants 28:439–454. https://doi.org/10.1007/s12298-022-01154-y

Guisinger MM, Kuehl JV, Boore JL, Jansen RK (2011) Extreme reconfiguration of plastid genomes in the angiosperm family Geraniaceae: rearrangements, repeats, and codon usage. Mol Biol Evol 28:583–600. https://doi.org/10.1093/molbev/msq229

Jiang H, Tian J, Yang J et al (2022) Comparative and phylogenetic analyses of six Kenya *Polystachya* (Orchidaceae) species based on the complete chloroplast genome sequences. BMC Plant Biol 22:177. https://doi.org/10.1186/s12870-022-03529-5

Jin D-M, Jin J, Yi T (2020) Plastome structural conservation and evolution in the clusioid clade of Malpighiales. Sci Rep 10:9091. https://doi.org/10.1038/s41598-020-66024-7

Khan A, Asaf S, Khan AL et al (2019) Complete chloroplast genomes of medicinally important *Teucrium* species and comparative analyses with related species from Lamiaceae. PeerJ 7:7260. https://doi.org/10.7717/peerj.7260

Kim K-J, Choi K-S, Jansen RK (2005) Two chloroplast DNA inversions originated simultaneously during the early evolution of the sunflower family (Asteraceae). Mol Biol Evol 22:1783–1792. https://doi.org/10.1093/molbev/msi174

Krawczyk K, Nobis M, Myszczyński K et al (2018) Plastid superbarcodes as a tool for species discrimination in feather grasses (Poaceae: *Stipa*). Sci Rep 8:1–10. https://doi.org/10.1038/s41598-018-20399-w

Lee SR, Kim K, Lee BY et al (2019) Complete chloroplast genomes of all six *Hosta* species occurring in Korea: molecular structures, comparative, and phylogenetic analyses. BMC Genomics 20:833. https://doi.org/10.1186/s12864-019-6215-y

Li B, Lin F, Huang P et al (2017) Complete chloroplast genome sequence of *Decaisnea insignis*: Genome organization, genomic resources and comparative analysis. Sci Rep 7:10073. https://doi.org/10.1038/s41598-017-10409-8

Li H, Xiao W, Tong T et al (2021) The specific DNA barcodes based on chloroplast genes for species identification of Orchidaceae plants. Sci Rep 11:1424. https://doi.org/10.1038/s41598-021-81087-w

Li Y, Zhang LN, Wang TX et al (2022) The complete chloroplast genome sequences of three lilies: Genome structure, comparative genomic and phylogenetic analyses. J Plant Res 135:723–737. https://doi.org/10.1007/s10265-022-01417-5

Liu XQ, Ickert-Bond SM, Chen LQ, Wen J (2013) Molecular phylogeny of *Cissus* L. of Vitaceae (the grape family) and evolution of its pantropical intercontinental disjunctions. Mol Phylogenet Evol 66:43–53. https://doi.org/10.1016/j.ympev.2012.09.003

Lu L, Cox CJ, Mathews S et al (2018) Optimal data partitioning, multispecies coalescent and Bayesian concordance analyses resolve early divergences of the grape family (Vitaceae). Cladistics 34:57–77. https://doi.org/10.1111/cla.12191

Ma W, Lv C, Jiang D, Kang C, Zhao D (2020) The complete chloroplast genome sequence of *Euphorbia lathyris* L. Mito DNA B Resour5(3): 3678–3680 https://doi.org/10.1080/23802359.2020.1832601

Muraguri S, Xu W, Chapman M et al (2020) Intraspecific variation within Castor bean (*Ricinus communis* L.) based on chloroplast genomes. Ind Crops Prod 155:112779. https://doi.org/10.1016/j.indcrop.2020.112779

Nie X, Deng P, Feng K et al (2014) Comparative analysis of codon usage patterns in chloroplast genomes of the Asteraceae family. Plant Mol Biol Rep 32:828–840. https://doi.org/10.1007/s11105-013-0691-z

Onyeweaku GC, Nyananyo BL, Ozimede CO (2020) Taxonomic studies on the genus *Cissus* L. (Vitaceae) present in Obio/Akpor local government area of Rivers state Nigeria. J Appl Sci Environ Manag 24:139. https://doi.org/10.4314/jasem.v24i1.20

Palmer JD, Jansen RK, Michaels HJ et al (1988) Chloroplast DNA variation and plant phylogeny. Annals Missouri Bot. Gard. 75:1180–1206

Purohit S, Bohra MK, Jain R (2022) Identification of bioactive pentacyclic triterpenoids and fatty acid derivatives from *Cissus quadrangularis* and C. *rotundifolia* through untargeted metabolite profiling. Appl Biochem Biotechnol. https://doi.org/10.1007/s12010-022-03940-6

Rossetto M, Jackes BR, Scott KD, Henry RJ (2002) Is the genus *Cissus* (Vitaceae) monophyletic? evidence from plastid and nuclear ribosomal DNA. Syst Bot 27:522–533. https://doi.org/10.1043/0363-6445-27.3.522

Rossetto M, Crayn DM, Jackes BR, Porter C (2007) An updated estimate of intergeneric phylogenetic relationships in the Australian Vitaceae. symposium on Vitis at the XVII international botanical congress-2005, Vienna. Austria Can J Bot 85:722–730. https://doi.org/10.1139/B07-022

Rossini BC, de Moraes MLT, Marino CL (2021) Complete chloroplast genome of *Myracrodruon urundeuva* and its phylogenetics relationships in Anacardiaceae family. Physiol Mol Biol Plants 27:801–814. https://doi.org/10.1007/s12298-021-00989-1

Sawangjit R, Puttarak P, Saokaew S, Chaiyakunapruk N (2017) Efficacy and safety of *Cissus quadrangularis* L. in clinical use: A systematic review and meta-analysis of randomized controlled trials. Phyther Res 31:555–567. https://doi.org/10.1002/ptr.5783

Schmitz-Linneweber C, Maier RM, Alcaraz JP et al (2001) The plastid chromosome of spinach (*Spinacia oleracea*): complete nucleotide sequence and gene organization. Plant Mol Biol 45:307–315. https://doi.org/10.1023/A:1006478403810

Senapati A, Basak S, Rangan L (2021) A review on application of DNA barcoding technology for rapid molecular diagnostics of adulterants in herbal medicine. Drug Saf. https://doi.org/10.1007/s40264-021-01133-4

Shelke RG, Das AB (2015) Analysis of genetic diversity in 21 genotypes of Indian banana using RAPDs and IRAPs markers. Proc Natl Acad Sci India Sect B Biol Sci 85:1027–1038. https://doi.org/10.1007/s40011-015-0505-1

Shelke RG, Rangan L (2019) Isolation and characterisation of Ty1-copia retrotransposons from *Pongamia pinnata*. Trees 33:1559–1570. https://doi.org/10.1007/s00468-019-01878-7

Shelke RG, Rangan L (2020) The role of transposable elements in *Pongamia* unigenes and protein diversity. Mol Biotechnol 62:31–42. https://doi.org/10.1007/s12033-019-00223-0

Shelke RG, Rangan L (2022) The whole chloroplast genome of *Mesua ferrea*: Insight into the dynamic pattern of evolution and its comparison with species from recently diverged families. Gene 846:146866. https://doi.org/10.1016/j.gene.2022.146866

Shelke RG, Basak S, Rangan L (2020) Development of EST-SSR markers for *Pongamia pinnata* by transcriptome database mining: Cross-species amplification and genetic diversity. Physiol Mol Biol Plants 26:2225–2241. https://doi.org/10.1007/s12298-020-00889-w

Shidhi PR, Nadiya F, Biju VC et al (2021) Complete chloroplast genome of the medicinal plant *Evolvulus alsinoides*: Comparative analysis, identification of mutational hotspots and evolutionary dynamics with species of Solanales. Physiol Mol Biol Plants 27:1867–1884. https://doi.org/10.1007/s12298-021-01051-w

Sivarajan VV, Balachandran I (1994) Ayurvedic drugs and their plant sources. Oxford and IBH publishing

Sundaran J, Begum R, Vasanthi M et al (2020) A short review on pharmacological activity of *Cissus quadrangularis*. Bioinformation 16(8):579–585. https://doi.org/10.6026/97320630016579

Tian S, Lu P, Zhang Z et al (2021) Chloroplast genome sequence of Chongming lima bean (*Phaseolus lunatus* L.) and comparative analyses with other legume chloroplast genomes. BMC Genom 22:194. https://doi.org/10.1186/s12864-021-07467-8

Tillich M, Lehwark P, Pellizzer T et al (2017) GeSeq - versatile and accurate annotation of organelle genomes. Nucleic Acids Res 45:W6–W11. https://doi.org/10.1093/nar/gkx391

Tiwari M, Gupta PS, Sharma N (2018) Ethnopharmacological, phytochemical and pharmacological review of plant *Cissus quadrangularis* L. Res J Pharmacogn Phytochem 10:81. https://doi.org/10.5958/0975-4385.2018.00014.6

Wang L, Roossinck MJ (2006) Comparative analysis of expressed sequences reveals a conserved pattern of optimal codon usage in plants. Plant Mol Biol 61:699–710. https://doi.org/10.1007/s11103-006-0041-8

Wang L, Liang J, Sa W, Wang L (2021) Sequencing and comparative analysis of the chloroplast genome of *Ribes odoratum* provide insights for marker development and phylogenetics in Ribes. Physiol Mol Biol Plants 27:81–92. https://doi.org/10.1007/s12298-021-00932-4

Wang X, Wang D, Gao N et al (2022) Identification of the complete chloroplast genome of *Malus zhaojiaoensis* Jiang and its comparison and evolutionary analysis with other *Malus* species. Genes. https://doi.org/10.3390/genes13040560

Wen J, Lu LM, Nie ZL et al (2018) A new phylogenetic tribal classification of the grape family (Vitaceae). J Syst Evol 56:262–272. https://doi.org/10.1111/jse.12427

Wu F, Zhang S, Gao Q et al (2021) Genetic diversity and population structure analysis in a large collection of *Vicia amoena* in China with newly developed SSR markers. BMC Plant Biol 21:544. https://doi.org/10.1186/s12870-021-03330-w

Wu H-Y, Wong K-H, Kong BL et al (2022) Comparative analysis of chloroplast genomes of *Dalbergia* species for identification and phylogenetic analysis. Plants 11:1109

Xu S, Sun M, Mei Y et al (2022) The complete chloroplast genome sequence of the medicinal plant *Abrus pulchellus* subsp. cantoniensis: genome structure, comparative and phylogenetic relationship analysis. J Plant Res 135:443–452. https://doi.org/10.1007/s10265-022-01385-w

Zhang Z (2022) KaKs_calculator 30: Calculating selective pressure on coding and non-coding sequences. Genom Proteom Bioinform. https://doi.org/10.1016/j.gpb.2021.12.002

Zhang T, Fang Y, Wang X et al (2012) The complete chloroplast and mitochondrial genome sequences of *Boea hygrometrica*: Insights into the evolution of plant organellar genomes. PLoS ONE 7(1):e30531. https://doi.org/10.1371/journal.pone.0030531

Zhang N, Wen J, Zimmer EA (2016a) Another look at the phylogenetic position of the grape order Vitales: chloroplast phylogenomics with an expanded sampling of key lineages. Mol Phylogenet Evol 101:216–223. https://doi.org/10.1016/j.ympev.2016.04.034

Zhang N, Wen J, Zimmer EA (2016b) Correction: congruent deep relationships in the grape family (Vitaceae) based on sequences of chloroplast genomes and mitochondrial genes via genome skimming. PLoS ONE 11:1–12. https://doi.org/10.1371/journal.pone.0152059

Zhang W, Sun Y, Liu J et al (2021) DNA barcoding of Oryza: conventional, specific, and super barcodes. Plant Mol Biol 105:215–228. https://doi.org/10.1007/s11103-020-01054-3