Complete chloroplast genome sequencing and comparative analysis of threatened dragon trees Dracaena serrulata and Dracaena cinnabari

Waqar Ahmad1,4, Sajjad Asaf1, Arif Khan2, Ahmed Al-Harrasi1,2, Abdulraqeb Al-Okaishi3 & Abdul Latif Khan4

Dracaena (Asparagaceae family) tree is famous for producing "dragon blood"—a bioactive red-colored resin. Despite its long history of use in traditional medicine, little knowledge exists on the genomic architecture, phylogenetic position, or evolution. Hence, in this study, we sequenced the whole chloroplast (cp) genomes of D. serrulata and D. cinnabari and performed comparative genomics of nine genomes of the genus Dracaena. The results showed that the genome sizes range from 155,055 (D. elliptica) to 155,449 (D. cochinchinensis). The cp genomes of D. serrulata and D. cinnabari encode 131 genes, each including 85 and 84 protein-coding genes, respectively. However, the D. hokouensis had the highest number of genes (133), with 85 protein coding genes. Similarly, about 80 and 82 repeats were identified in the cp genomes of D. serrulata and D. cinnabari, respectively, while the highest repeats (103) were detected in the cp genome of D. terniflora. The number of simple sequence repeats (SSRs) was 176 and 159 in D. serrulata and D. cinnabari cp genomes, respectively. Furthermore, the comparative analysis of complete cp genomes revealed high sequence similarity. However, some sequence divergences were observed in accD, matK, rpl16, rpoC2, and ycf1 genes and some intergenic spacers. The phylogenomic analysis revealed that D. serrulata and D. cinnabari form a monophyletic clade, sister to the remaining Dracaena species sampled in this study, with high bootstrap values. In conclusion, this study provides valuable genetic information for studying the evolutionary relationships and population genetics of Dracaena, which is threatened in its conservation status.

Dracaena is an important genus from the family Asparagaceae that includes wild and indoor exquisite plants1. The genus comprises 190 species2 and is also known as Dragon trees. These are distributed across the drylands in Africa, Arabia, and the Americas3. In response to incisions, these plants produce a red resin called "Dragon Blood" that is medicinally important and has an ancient history in traditional herbal medicine4. The resin has been known to act as an anti-cancer, hemostatic, anti-ulcer, anti-viral, anti-microbial, anti-inflammatory, and anti-oxidant5. Dracaena resin is also used for giving colors to certain materials like toothpaste, varnishes, and plasters6. The highest levels of species diversity occur in tropical Africa and Southeast Asia. These species grow in various habitats, including tropical monsoon, semi-evergreen, and evergreen rain forests. Some species grow in specialized habitats such as escarpments, littoral forest edges, and riverbeds with strongly fluctuating water levels, where they become facultative rheophytes7.

Among Dracaena species, D. serrulata and D. cinnabari (Fig. 1) are regional, endemic species found in southern Oman, Saudi Arabia, and Yemen (Socotra Island). These endangered species are currently threatened by mining operations, agriculture, drought, and possibly climate change. The known populations are threatened by grazing (camels, goats, and sheep) during the dry season8–10. Dracaena, along with other globally important genera Sansevieria Thunb and Pleomele Salisb (family Asparagaceae and Nolinoideae subfamily) are collectively

1Natural and Medical Sciences Research Centre, University of Nizwa, 616 Nizwa, Oman. 2Genomics Group, Faculty of Biosciences and Aquaculture, Nord University, 8049 Bode, Norway. 3Environmental Protection Agency, Socotra, Yemen. 4Department of Engineering Technology, University of Houston, Sugar Land, TX 77479, USA. 5*email: aharrasi@unizwa.edu.om; alkhan@uh.edu
shifted from one family to another like Agavaceae, Liliaceae, Dracaenaceae, Ruscaceae, and lately in genomes such as... and discrimination of sub-species.

bari

Authority Socotra, Yemen. The voucher specimen numbered UoN-DS1 (D. serrulata) and UoN-DC1 (D. cinnabari) were donated by the Environmental Protection Affairs, Sultanate of Oman. The fresh specimens of were donated by the Environmental Protection

compliance with local, institutional, national, or international regulations at the time of collection. A permission letter was retrieved from the Director-General of Nature Conservation, Ministry of Environment and Climate

(25–46 °C). All plant specimens used for this study were collected from the wild to the best of our knowledge in the Dhofar region (wild) of Oman. The habitat climate is arid with a low precipitation rate and temperate

looking at these challenges, in the current study, we aim to resolve identification at lower taxonomic levels. Complete chloroplast genome sequencing coupled with comparative analysis allows advanced phylogenetic reconstruction and can be used as super-barcodes to resolve identification at lower taxonomic levels. Looking at these challenges, in the current study, we aim to sequence the D. serrulata and D. cinnabari and perform comparative chloroplast genome analysis to explain the basis of genome architecture and divergences across Dracaena species. Hence, we report the complete chloroplast genomes sequences of D. serrulata and D. cinnabari. Both the species and other species in the genus possess the least genic information. Hence, current datasets will help understand the genome architecture, comparative genomics with related species, and in-depth phylogeny of Dracaena species.

Methodology

Sample collection. Fresh young leaves were collected from the D. serrulata and D. cinnabari plants growing in the Dhofar region (wild) of Oman. The habitat climate is arid with a low precipitation rate and temperate (25–46 °C). All plant specimens used for this study were collected from the wild to the best of our knowledge in compliance with local, institutional, national, or international regulations at the time of collection. A permission letter was retrieved from the Director-General of Nature Conservation, Ministry of Environment and Climate Affairs, Sultanate of Oman. The fresh specimens of D. cinnabari were donated by the Environmental Protection Authority Socotra, Yemen. The voucher specimen numbered UoN-DS1 (D. serrulata) and UoN-DC1 (D. cinnabari) were deposited in the University of Nizwa herbarium center. The identification of plants was carried out by...
DNA extraction and sequencing. With brief modifications, the cp DNA was isolated from collected samples as described in Shi et al.31. The construction of genomic libraries was carried out as per provided instructions (Life Technologies USA, Eugene, OR, USA). To arrange the cp DNA into 400 bp fragments (enzymatically) for libraries, the Ion Shear™ Plus Reagents kit and Ion Xpress™ Plus gDNA Fragment Library kit were used. Qubit 3.0 fluorometer and bioanalyzer (Agilent 2100 Bioanalyzer system, Life Technologies USA) were used to quantify the prepared libraries. The amplification of the template was performed using Ion OneTouch™ 2. The Ion OneTouchTM ES enrichment system enriched the amplified templates using Ion 530 and 520 OT2 reagents. Ion S5 protocol of sequencing was followed for loading the sample on S5 530 chip.

Genome assembly and annotation. The number of raw reads obtained for *D. serrulata* and *D. cinnabari* were 14,654,144 and 16,888,126, respectively. The reads were first screened for a Phred score < 30 to remove low-quality sequences. To ensure the accuracy of cp genome assembly, we employed two methods to assemble the cp genome. In the first method, obtained reads of cp genomes *D. serrulata* and *D. cinnabari* were mapped to the reference genome of *D. cochinchinensis* (MF943127) and *D. comobiana* (MN20094), respectively, by Geneious Pro (v.10.2.3) software using Bowtie2 (v.2.2.3)23,33. Assembly means coverage of *D. serrulata* was 876X, and *D. cinnabari* was 768X. In the second method, the cp genome of *D. serrulata* and *D. cinnabari* were de novo assembled using the GetOrganelle pipeline34, with SPAdes 3.10.1 assembler35. The cp genomes *D. serrulata* and *D. cinnabari* were annotated using CpGAVAS and DOGMA (http://dogma.cbb.utexas.edu/, China)36. The tRNAs can-SE detected the tRNA genes (v.1.21)37. Intron boundaries, manual alteration and start and stop codon adjustments of genomes were carried out using Geneious Pro (v.10.2.3)33 and tRNAs can-SE37 by comparing the cp genomes to reference genomes. OGDRAW38 was utilized to illustrate the structural features in cp genomes.

Repeat identification. The determination of palindromic, forward and reverse repeats was performed using the online tool REPuter39 with 8 bp minimum repeat size and 50 maximum computed repeats. Furthermore, MISA software40 with conditions of ≥ 10 repeat units for 1 bp repeats; ≥ 8 repeat units for 2 bp repeats; ≥ 4 repeat units for 3 and 4 bp repeats and ≥ 3 repeat units for 5 and 6 bp repeats was used to calculate Simple sequence repeats (SSRs) and tandem repeats were calculated by Tandem Repeats Finder v.4.0931. The sequence divergence in shared genes and complete cp genomes of *D. serrulata* and *D. cinnabari*, and other closely related species were determined. Multiple sequence alignment was performed via comparative analysis, and the gene order was compared to clarify the missing and ambiguous gene annotation. The cp genomes were aligned with default parameters using MAFFT version 7.22242 with default window analysis using DnaSP software version 6.13.0344. The mVISTA45 in shuffle-LAGAN mode was used to determine the genomic divergence while using cp genome of *D. serrulata* as a reference.

Phylogenetic analysis. To resolve the phylogenetic position of *D. serrulata* and *D. cinnabari* within the subfamily of Nolinoideae a total of 44 cp genomes were retrieved from NCBI database. Four *Asparagus* species, *A. scholberoides*, *A. officinalis*, *A. racemoses* and *A. setaceus* were selected as outgroups. The first tier alignment of complete cp genomes was performed according to the cp genome structure and conserved gene order46. The phylogenetic trees were constructed using four methods by employing the setting described previously by Asaf et al.48. Neighbour-joining (NJ) and maximum likelihood (ML) were implemented in MEGA 649; Bayesian phylogenetic trees were constructed using four methods by employing the setting described previously by Asaf et al.53. The branch-swapping tree search criterion. In the second tier, 66 shared protein-coding genes from 46 cp genomes from subfamily Nolinoideae were aligned using MAFFT version 7.22242 under default parameters and making various manual adjustments to preserve and improve reading frames in the second tiers of phylogenetic analysis. The above four aforementioned phylogenetic inference models (ML, NJ, BI and MP) were employed to construct trees using 66 concatenated genes as mentioned above and suggested by Asaf et al.53.

Results and discussion. The results showed that the cp genomes of *D. serrulata* (MT408026) and *D. cinnabari* (OK235335) have the typical quadripartite structures like other related plants44–46 with a genome size of 155,398 bp and 155,351 bp respectively. Both the cp genomes comprised of 4 distinctive parts in which the LSC (83,871 bp, 83,818 bp) and SSC (19,247 bp, 18,579 bp) are separated by two IRs (26,140 bp, 26,477 bp) (Fig. 2, Table 1). The cp genomes of *D. serrulata* and *D. cinnabari* were analyzed and compared with *D. angustifolia*, *D. cambodiana*, *D. cochinchinensis*,
Figure 2. Genome Map of the *D. serrulata* and *D. cinnabari* cp genomes. Thick lines represent inverted repeat regions (IRs). IRs split the cp genome into large single copies (LSC) and single small copies (SSC) regions. The counter-clockwise transcribing genes are drawn outside while the clockwise are drawn inside the circle. Genes related to different functional groups are color coded. The GC and AC content is represented by the circle’s dark and light green shades.

Table 1. Chloroplast genomes features summary of *D. serrulata*, *D. cinnabari* and related species of *Dracaena* genus.
D. cochinchinensis2, D. draco, D. elliptica, D. fragrans, D. hokouensis and D. terniflora (Table 1), which are closely related and belongs to the same genus. The sizes of these cp genomes range from 155,055 bp (D. elliptica) to 155,449 bp (D. cochinchinensis), as shown in Table 1. The cp genomes of D. serrulata and D. cinnabari encodes a total of 131 genes like all compared cp genomes except D. cambodiana, D. cochinchinensis and D. elliptica, which encode 130 genes while D. hokouensis encodes 133 genes. Similarly, among the total genes encoded by cp genomes of D. serrulata and D. cinnabari, 85 and 84 are protein-coding genes, respectively (Table 1). Furthermore, D. serrulata and D. cinnabari's cp genomes encode eight rRNA and 38 tRNA genes, respectively (Fig. 2).

Repeats and simple sequence repeats SSR analysis in Cp genomes. A total of 80 and 82 repeats were identified in D. serrulata and D. cinnabari, respectively. In contrast, the cp genome of D. terniflora had the highest number of total repeats (103) and D. elliptica had the minimum (79). In D. serrulata and D. cinnabari, the palindromic repeats were 29 and 26, respectively (Fig. 3A). Similarly, both sequenced cp genomes had the forward repeats of 20 each (Fig. 3B) whereas the reverse repeats identified were zero in D. serrulata and 3 in D. cinnabari (Fig. 3C). Furthermore, the tandem repeats were also identified for both sequenced genomes, 31 and 33, respectively (Fig. 3D). Although, the highest and lowest number of forward repeats were detected in cp genome of D. terniflora (36) and D. hokouensis (19), while the reverse repeats were highest in D. cochinchinensis, D. draco and D. elliptica (4) and zero in D. serrulata. Most palindromic repeats were detected in D. serrulata and D. hokouensis i.e. 29. Similarly, the tandem repeats were most in the cp genome of D. cochinchinensis (38) and least in D. elliptica (30). The total number of repeats was highest (87) in D. cochinchinensis (Fig. 3E).

Simple sequence repeats (SSR) are used as genetic markers in evolutionary and population genetics studies45. These repeats also known as microsatellites are usually comprised of 1–6 bp repeat units46. Furthermore, SSRs are important because their relative lack of recombination, maternal inheritance, and haploid nature make them potential candidates for phylogenetic studies. SSRs play a role in estimating genetic variation, gene flow analysis, and studying the population history in plants and animals45,46. In this study, we analyzed SSRs in the cp genomes of D. serrulata and D. cinnabari and studying the population history in plants and animals65,66. In this study, we analyzed SSRs in the cp genomes of D. serrulata and D. cinnabari, and found SSRs in the cp genome of D. serrulata (5), followed D. serrulata (4) (Fig. 4D), while the tri-nucleotide SSRs were 3 in cp genomes of D. serrulata and D. cinnabari along with other compared cp genomes except in D. fragrance which were two and D. cochinchinensis2 with no

Table 2. Gene composition in Dracaena species cp genomes.

| Category of genes | Group of genes |
|-------------------|----------------|
| Genes for photosynthesis | Subunits of ATP synthase |
| Genes for photosynthesis | Subunits of photosystem II |
| Genes for photosynthesis | Subunits of NADH-dehydrogenase |
| Genes for photosynthesis | Subunits of cytochrome b/f complex |
| Genes for photosynthesis | Subunits of photosystem I |
| Genes for photosynthesis | Subunits of rubisco |
| Self-replication | Large subunit of ribosome |
| Self-replication | DNA dependent RNA polymerase |
| Self-replication | Small subunit of ribosome |
| Other genes | Subunit of Acetyl-CoA-carboxylase |
| Other genes | c-type cytochrom synthesis gene |
| Other genes | Envelop membrane protein |
| Other genes | Maturase |
| Unknown | Conserved open reading frames |

Self-replication DNA dependent RNA polymerase (rpoA, rpoB, rpoC1, rpoC2)
Self-replication Small subunit of ribosome (rpl14, rpl16, rpl2, rpl20, rpl22, rpl23, rpl32, rpl33, rpl36)
Other genes Subunit of Acetyl-CoA-carboxylase (accD)
Other genes c-type cytochrom synthesis gene (ccsA)
Other genes Envelop membrane protein (cemA)
Other genes Maturase (matK)
Unknown Conserved open reading frames (ycf1, ycf2, ycf3, ycf4)
tri-nucleotides (Fig. 4E). A total of 2 tetra-nucleotide SSRs were detected only in the cp genome of Dracaena cochinchinensis. In contrast, in this study, the remaining cp genomes had no tetra-nucleotide SSRs, including the sequenced cp genomes of Dracaena serrulata and Dracaena cinnabari (Fig. 4F). The penta-nucleotide SSRs detected in Dracaena serrulata were 5, while the Dracaena cinnabari had zero (Fig. 4G). Contrastingly the hexanucleotide SSRs was found in only the Dracaena cinnabari cp genome, as shown in Fig. 4H. Likewise, patterns in Dracaena and other angiosperms cp genomes

Table 3. Introns and exons lengths for the splitting genes in cp genomes of D. serrulata and D. cinnabari.

| Gene       | Start DS | End DS | ExonI (bp) DS | IntronI (bp) DS | ExonII (bp) DS | IntronII (bp) DS | ExonIII (bp) DS | Start DC | End DC | ExonI (bp) DC | IntronI (bp) DC | ExonII (bp) DC | IntronII (bp) DC | ExonIII (bp) DC |
|------------|----------|--------|---------------|-----------------|----------------|-----------------|-----------------|----------|--------|---------------|----------------|----------------|----------------|-----------------|
| trnK-UUU   | 1513     | 1513   | 4157          | 37              | 2568           | 2568            | 37              |          |        |               |                |                |                |                 |
| rps16      | 4789     | 4789   | 5910          | 46              | 867            | 867             | 46              |          |        |               |                |                |                |                 |
| trnG-GCC   | 9131     | 9131   | 9906          | 23              | 716            | 716             | 23              |          |        |               |                |                |                |                 |
| atpF       | 11,854   | 11,854 | 13,230        | 145             | 828            | 828             | 145             |          |        |               |                |                |                |                 |
| rpsCl      | 20,640   | 20,640 | 23,415        | 432             | 718            | 718             | 432             |          |        |               |                |                |                |                 |
| ycfD       | 42,150   | 42,150 | 44,126        | 126             | 731            | 731             | 126             |          |        |               |                |                |                |                 |
| trnL-UAG   | 46,962   | 46,962 | 47,593        | 35              | 547            | 547             | 35              |          |        |               |                |                |                |                 |
| trnV-UAC   | 52,093   | 52,093 | 52,754        | 39              | 586            | 586             | 39              |          |        |               |                |                |                |                 |
| clpP       | 70,044   | 70,497 | 72,016        | 69              | 825            | 819             | 69              |          |        |               |                |                |                |                 |
| petB       | 74,979   | 74,979 | 76,381        | 7               | 752            | 752             | 7               |          |        |               |                |                |                |                 |
| petD       | 76,586   | 76,586 | 77,830        | 8               | 732            | 732             | 8               |          |        |               |                |                |                |                 |
| rpl2       | 84,455   | 84,455 | 85,928        | 391             | 652            | 652             | 391             |          |        |               |                |                |                |                 |
| ndhB       | 94,954   | 94,954 | 97,185        | 77             | 699            | 699             | 77              |          |        |               |                |                |                |                 |
| trnA-UGC   | 103,803  | 103,803| 104,690       | 38              | 815            | 815             | 38              |          |        |               |                |                |                |                 |
| ndhA       | 120,271  | 120,271| 122,444       | 559             | 1076           | 1076            | 559             |          |        |               |                |                |                |                 |
| trnA-UGC   | 134,580  | 134,580| 135,467       | 38              | 815            | 815             | 38              |          |        |               |                |                |                |                 |
| ndhB       | 142,085  | 142,085| 144,316       | 775             | 699            | 699             | 775             |          |        |               |                |                |                |                 |
| rpl12      | 1513     | 1513   | 9811          | 23              | 716            | 706             | 23              |          |        |               |                |                |                |                 |
| trnG-GCC   | 9035     | 9035   | 9906          | 42              | 937            | 947             | 42              |          |        |               |                |                |                |                 |
| trnL-GAU   | 102,671  | 102,671| 136,545       | 42              | 937            | 947             | 42              |          |        |               |                |                |                |                 |

Figure 3. Repetitive sequences in D. serrulata, D. cinnabari and related Dracaena species cp genomes. (A) A total number of repetitive sequences in cp genomes, (B) Lengthwise frequency of palindromic repeats (C) Lengthwise frequency of forward repeats (D) Lengthwise frequency of reverse repeats (E) Lengthwise frequency of tandem repeats.
Figure 4. Simple sequence repeats (SSRs) in *D. serrulata*, *D. cinnabari*, and related *Dracaena* species cp genomes. (A) Total number of SSRs in cp genomes, (B) SSR motif frequency in cp genomes, (C) Mono-nucleotides SSRs (D) Di-nucleotides SSRs, (E) Tri-nucleotides SSRs, (F) Tetra-nucleotides SSRs, (G) Penta-nucleotides SSRs and (H) Hexa-nucleotides SSRs.

were also reported previously. Our results agree with the recent studies reporting that identified SSRs in cp genomes are made of polyadenine or polythymine repeats. The contrary is with guanine (G) and cytosine (C). Therefore, the cp genomes of *D. serrulata* and *D. cinnabari* are rich in 'AT' content, as reported previously.
per earlier reports, the SSRs are randomly distributed across the cp genomes, revealing important information for selecting molecular markers for polymorphism (inter and intra-specific) \(^72,73\). The current results are in synergy with previous reports of angiosperms indicating the dominating abundance of ‘A’ or ‘T’ mono-nucleotides SSRs in cp genomes and resulting in ‘AT’ rich cp genomes\(^74,75\).

**Comparative analysis and sequence divergence analyses.** Comparative analysis of the cp genome plays a pivotal role in understanding plant species’ genetic diversity and evolutionary relationships\(^22,76\). The cp genomes \(D. serrulata\) and \(D. cinnabari\) were compared to the closely related species for sequence divergence. The cp genome of \(D. serrulata\) was selected as a reference genome. The cp genomes of \(D. serrulata\) and \(D. cinnabari\) along with all the compared cp genomes, were highly conserved. All aligned sequences exhibit high similarities with only a few regions sequence variations in non-coding regions (Fig. 5). Interestingly, a higher degree of divergence was observed in non-coding regions in all cp genomes compared to the coding areas reported previously\(^77,78\). The current results revealed the high similarity of cp genome sequences of all species included in the study, suggesting that the cp genomes of \(Dracaena\) genus are highly conserved as reported for \(Dracaena\)\(^28\) and \(Camellia\) genus\(^79\). The \(petD\), and \(clpP\) genes in the LSC region, and the \(ycf1\) gene in the SSC region showed sequence divergence in the coding areas across all compared species, and these results agree with\(^21,28,71,80\).

Moreover, in IR regions, the \(rps19\) gene showed sequence divergence in the cp genome of \(D. cinnabari\) and \(D. cochinchinensis\). In contrast, the \(ycf2\) gene showed variation in the cp genome of \(D. cambodiana\). In the LSC region, \(accD\), \(atpF\), \(ycf3\), and \(rps15\) genes showed sequence divergence in some cp genomes compared to the \(D. serrulata\) cp genome (Fig. 5). Furthermore, in the non-coding areas such as \(rps16-trnT\), \(rps4-trnL\), and \(petE-trnG\) in LSC while \(rps7-trnV\) in SSC showed sequence divergence across all the compared cp genomes, likewise pattern of divergence was also reported previously\(^78,79,81\).

---

**Figure 5.** Visual alignment of \(D. serrulata\), \(D. cinnabari\), and related \(Dracaena\) species cp genomes. VISTA-based identity plot showing sequence identities among eleven \(Dracaena\) species, using \(D. serrulata\) as a reference. Genome regions are color-coded as protein-coding, rRNA coding, tRNA coding, or conserved non-coding sequences (CNS). The x-axis represents the coordinate in the chloroplast genome. Annotated genes are displayed along the top. The sequences similarity of the aligned regions is shown as horizontal bars indicating the average percent identity between 50 and 100%.
Moreover, the average pairwise sequence divergence among the complete cp genomes (Table S2) and shared genes (Table S3) was calculated. *D. cinnabari* cp genome showed an average pairwise sequence divergence of 0.003. The cp genome of *D. cinnabari* showed the highest average pairwise sequence divergence with *D. cochinchinensis* and *D. fragrance* (0.0077). Other cp genomes included in the study and previous reports also showed similar results48,71. The most divergent genes were accD, matK, rpl16, rpoC2, and ycf1. The highest pairwise sequence divergence was identified for rpl16 (0.03) (Table S3). Similar results are also reported by Zhang, et al.28.

Similarly, the relative synonymous codon usage (RSCU) value analysis was performed using coding regions of 10 *Dracaena* cp genomes. The most abundantly used codons were A/U-ending codons. These results exhibited a higher codon usage toward A/U- endings than G/C-ended codons in all cp genomes of *Dracaena* species28,82,83. Codons like CAA, GCU, GCA, and GU UAC (yellow colored) have less than one RSCU value in one or more cp genomes (Fig. 6). Whereas the highest RSCU value was recorded for AGA (2) across all cp genomes, similar results were reported for *Punica granatum*84 and *D. draco*28. The codon characteristic pattern and frequency of usage are given in Table S1. The most frequently used codon was AAA (n = 2,036, 51.5%) in these genomes, which encodes lysine amino acid. In contrast, the least used codon was GCG (n = 257, 5.2%), coding the arginine amino acid (Table S1); these results agree with earlier reports28,85.

Similarly, the nucleotide diversity (Pi) values were calculated in these cp genomes (Fig. 7). The Pi values ranged from 0 to 0.024 (LSC), 0 to 0.027 (LSC), and 0 to 0.049 (IRs) with a mean of 0.0030, which indicates that the variation is slight among these cp genomes and are highly conserved, similar variation patterns were previously reported in angiosperm cp genomes86. Furthermore, the IR region showed higher Pi values than LSC and SSC reported87. However, some genes like accD, psbl (LSC), and ycf1 (SSC) showed higher Pi values of 0.02, 0.02, and 0.026 than other protein-coding genes. Similarly, the trnV-rps7 (IR region) showed the highest Pi value of 0.05. these results also agree with mVISTA divergence analysis and previous reports21,88,89.

**Contraction and expansion of IRs and single copy regions.** Inverted repeat regions are considered the most conserved regions. The size variations in cp genomes occur due to expansion/contraction of IRs and single copy regions90,91. The four junctions (JLA, JLB, JS, and JSB) between the single copy regions (LSC, SSC) and IRs (IRA, IRB) in cp genomes of *D. serrulata, D. cinnabari*, and others were comprehensively assessed.
The IRs regions are remarkably conserved across all the cp genomes in the current study. The IRs regions’ lengths correlate across all the compared cp genomes with only slight expansion and contraction (Fig. 8). The cp genomes of D. serrulata and D. cinnabari have the shortest IRs regions of 26,140 bp and 26,477 bp, respectively. In comparison, D. angustifolia and D. terniflora possess the most extended IRs regions of 26,530 bp. The positions of rpl22 and rps19 genes at JLB are similar in all the cp genomes with only one base (D. elliptica, D. cochinchenensis) and three bases (D. draco) differences. Interestingly, the ndhF gene was found 40 and 22 base pairs away from the JSB in SSC in cp genome of D. serrulata and D. cinnabari (Fig. 8). In contrast, in other compared cp genomes it is found extended into IRb regions and overlaps with ycf1 (28–80 bp) as found previously. Similarly, the ycf1 and rpl22 genes at JSA and JSB are slightly variable across some cp genomes. Previous reports support the results.

Phylogenetic analysis. Since the eighteenth century, the phylogenetic relationships among the Dracaena species have not been completely clarified and are still unclear. In Dracaena, significant morphological variation has been shown, with species generally. Until recently, limited number of genetic markers such as chloroplast genes (matK, rbcL) and intergenic spacer regions such as rpl32-trnL, trnQ-rps16, psbA-trnH, trnL-trnF etc.) were used to infer the phylogenetic relationships between the various Dracaena species such as D. serrulata, D. cinnabari and D. draco, etc. Therefore, additional genetic markers are required to determine the phylogenetic position of D. cinnabari and D. serrulata. Cp genomes as a super-barcode and concatenated protein coding genes with sufficient informative sites have been proven effective in resolving complicated phylogenetic relationships in various complex plant species. Therefore, this study determined the phylogenetic dispositions of D. serrulata and D. cinnabari within the subfamily Nolinoideae by analyzing 46 complete cp genomes from subfamily Nolinoideae and four complete cp genomes from subfamily Asparagoidea as outgroups (Fig. 9) and 66 shared protein-coding genes (Fig. S1). Phylogenetic analysis using ML, BI, NJ and MP methods was performed. The phylogenetic analysis of complete genomes and shared protein coding genes revealed almost the same phylogenetic signals. In these phylogenetic trees (Figs. 9, S1), D. serrulata and D. cinnabari formed a single clade with high bootstrap value and BI support.

Moreover, the tree topology enabled inference of the relationship based on the phylogenetic studies conducted by Durán et al. The position of both D. serrulata and D. cinnabari confirms the previously published phylogeny described by Durán et al. that D. serrulata is more closely related to D. cinnabari than D. draco (Fig. 9). Furthermore, both trees revealed that D. draco is more closely related to D. cochinchenensis and D. cambodiana. Similar results were reported previously by Durán et al. on the basis of chloroplast barcode genes such as rbcL and matK genes and intergenic spacer regions(trnQ-rps16 and rpl32-trnL). However, another study by Lu and Morden based on combined chloroplast intergenic spacer regions using Bayesian analysis showed contradictory results to our study. In this study, where D. serrulata was closely related to D. draco. Furthermore, The earlier finding of Wang et al., who placed Liriope and Ophiopogon in the tribe Ophiopogoneae, is also supported by our phylogenetic trees Lun-Kai et al., Song-Yun and Lun-Kai, and proposed that Ophiopogon and Liriope were closely related based on characteristics of the leaf epidermis and pollen as well as chromosomal counts. Even though our findings, based on the available cp genomes, clarified the phylogenetic relationships of some D. serrulata and D. cinnabari, more complete cp genome sequences are needed to resolve the comprehensive phylogenies of this genus because limited taxon sampling may produce discrepancies in tree topologies as reported earlier.
Conclusion

In the current study, the complete chloroplast genomes of *D. serrulata* and *D. cinnabari* were sequenced and elucidated for the first time. The overall gene order and cp genome organization were similar to nine *Dracaena* species. Repetitive sequences and SSRs were identified from the sequenced data and nine related cp genomes. In contrast, the highest number of repeats and SSRs were identified in *D. terniflora* and *D. serrulata*. Moreover, divergence is detected in intergenic spaces greater than in protein-coding regions of these cp genomes. Current results showed that the *D. serrulata* and *D. cinnabari* form a single clade. The whole cp genome sequencing of *D. serrulata* and *D. cinnabari* gives exciting insights and valuable data that may facilitate the identification of related species and answer taxonomic questions.

Figure 8. Distances between adjacent genes and junctions of the small single-copy (SSC), large single-copy (LSC), and two inverted repeats (IR) regions among *D. serrulata*, *D. cinnabari*, and related *Dracaena* species cp genomes. Boxes above and below the primary line indicate the adjacent border genes. The figure is not scaled regarding sequence length and only shows relative changes at or near the IR/SC borders.
Data availability
All data generated or analyzed during this study are included in this published article. The *D. serrulata* and *D. cinnabari* cp genomes were submitted to NCBI with accession numbers MT408026 and OK235335 respectively.

Received: 13 March 2022; Accepted: 12 September 2022
Published online: 06 October 2022

References
1. Bogawski, P. et al. Current and future potential distributions of three *Dracaena* Vand. ex L. species under two contrasting climate change scenarios in Africa. *Ecol. Evol.* 9, 6833–6848 (2019).
2. Madera, P. et al. What we know and what we do not know about dragon trees?. *Forests* 11, 236 (2020).
3. Marrero, A., Almeida, R. S. & González-Martín, M. A new species of the wild dragon tree, *Dracaena* (Dracaenaceae) from Gran Canaria and its taxonomic and biogeographic implications. *Bot. J. Linn. Soc.* 128, 291–314 (1998).
47. Kumar, S., Neti, M., Dudley, J. & Tamura, K. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief. Bioinform. 9, 299–306 (2008).
48. Asaf, S., Khan, A. L., Khan, A. & Al-Harrasi, A. Unraveling the chloroplast genomes of two prosopis species to identify its genomic information, comparative analyses and phylogenetic relationship. Int. J. Mol. Sci. 21, 3280 (2020).
49. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729 (2013).
50. Kuan, L., Pratas, E., Sousa, L. & Tomás, P. MrBayes sMC3: Accelerating Bayesian inference of phylogenetic trees. Int. J. High Perform. Comput. Appl. 32, 246–265 (2018).
51. Swofford, D. L. Phylogenetic analysis using parsimony. (1998).
52. Gascuel, O. BIOM: an improved version of the NJ algorithm based on a simple model of sequence data. Mol. Biol. Evol. 14, 685–695 (1997).
53. Asaf, S. et al. Expanded inverted repeat region with large scale inversion in the first complete plastid genome sequence of Plantago ovata. Sci. Rep. 10, 1–16 (2020).
54. Jin, J. et al. Complete chloroplast genome of a medicinal species Polygonatum kingianum in China (Asparagaceae, Asparagales). Mitochondrial DNA Part B 5, 959–960. https://doi.org/10.1080/23802359.2020.1721373 (2020).
55. Gu, L., Su, T., Luo, G.-L. & Hu, G.-X. The complete chloroplast genome sequence of Heteropolygonum ginchufashanicum (Asparagaceae) and phylogenetic analysis. Mitochondrial DNA Part B 6, 1799–1802. https://doi.org/10.1080/23802359.2021.1933636 (2021).
56. Jang, J.-H. et al. Characterization of the complete chloroplast genome of Liriope platyphylla (Asparagaceae: Nolinoideae) isolated in Korea. Mitochondrial DNA Part B 5, 2874–2875. https://doi.org/10.1080/23802359.2020.1787989 (2020).
57. Cay, S. B. et al. Genome skimming approach reveals the gene arrangements in the chloroplast genomes of the highly endangered Crocus L. species: Crocus istanbulensis (B. Mathew) Ruthsâns. PLoS ONE 17, e0269747 (2022).
58. Hishamuddin, M. S. et al. Comparison of eight complete chloroplast genomes of the endangered Aquilaria tree species (Thymelaeaceae) and their phylogenetic relationships. Sci. Rep. 10, 1–13 (2020).
59. Qian, S.-J., Zhang, Y.-H. & Li, G.-D. The complete chloroplast genome sequence of a medicinal plant, Wikstroemia chamaedaphne (Thymelaeaceae) and its phylogenetic relationship. Mitochondrial DNA Part B 5, 648–649 (2020).
60. Tao, X., Ma, L., Zhang, Z., Liu, W. & Liu, Z. Characterization of the complete chloroplast genome of alfalfa (Medicago sativa) (Leguminoseae). Gene Rep. 6, 67–73 (2017).
61. Xiang, B. et al. The complete chloroplast genome sequence of the medicinal plant Swertia mussotii using the PacBio RS II platform. Molecules 21, 1029 (2016).
62. Duffy, A. M., Kelchner, S. A. & Wöll, P. G. Conservation of selection on matK following an ancient loss of its flanking intron. Gene 438, 17–25 (2009).
63. Huang, Y.-Y., Matzke, A. I. & Matzke, M. Complete sequence and comparative analysis of the chloroplast genome of coconut palm (Cocos nucifera). PLoS ONE 8, e74736 (2013).
64. Mason, A. S. in Plant Genotyping 77–89 (Springer, 2015).
65. Vieira, L. D. N. et al. The complete chloroplast genome sequence of Podocarpus lambertii: genome structure, evolutionary aspects, gene content and SSR detection. PLoS ONE 9, e90618 (2014).
66. Ebert, D. & Peakall, R. Chloroplast simple sequence repeats (cpSSRs): technical resources and recommendations for expanding cpSSR discovery and applications to a wide array of plant species. Mol. Ecol. Resour. 9, 673–690 (2009).
67. Khan, A. L., Asaf, S., Lee, I.-J., Al-Harrasi, A. & Al-Rawahi, A. First reported chloroplast genome sequence of Punica granatum (cultivar Helow) from Jabal Al-Akhdar, Oman: phylogenetic comparative assessment with Lagerstroemia. Genetica 146, 681–694 (2018).
68. Du, X. et al. The complete chloroplast genome sequence of yellow mustard (Sinapis alba L.) and its phylogenetic relationship to other Brassicaceae species. Gene 731, 144340 (2020).
69. Liu, X., Li, Y., Yang, H. & Zhou, B. Chloroplast genome of the folk medicine and vegetable plant Talinum paniculatum (Jacq.) Gaertn.: gene organization, comparative and phylogenetic analysis. Molecules 23, 857 (2018).
70. Zhou, J. et al. Molecular structure and phylogenetic analyses of complete chloroplast genomes of two Aristolochia medicinal species. Int. J. Mol. Sci. 18, 1839 (2017).
71. Qian, S.-J. et al. The complete chloroplast genome sequence of the medicinal plant Salvia miltiorrhiza. PLoS ONE 8, e57607 (2013).
72. Powell, W., Morgante, M., McDevitt, R., Vendramin, G. & Ralfski, J. Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. Proc. Natl. Acad. Sci. 92, 7759–7763 (1995).
73. Prowan, J., Corbett, G., Powell, W. & Nicot, J. Chloroplast DNA variability in wild and cultivated rice (Oryza spp.) revealed by polymorphic chloroplast simple sequence repeats. Genome 40, 104–119 (1997).
74. Khan, A. L., Asaf, S., Al-Rawahi, A. & Al-Rassasi, A. Decoding first complete chloroplast genome of toothbrush tree (Salvadora persica L.): insights into genome evolution, sequence divergence and phylogenetic relationship within Brassicales. BMC Genomics 22, 312. https://doi.org/10.1186/s12864-021-07626-x (2021).
75. Ma, Q. et al. Complete chloroplast genome sequence of a major economic species, Ziziphus jujuba (Rhamnaceae). Curr. Genet. 63, 117–129 (2017).
76. Asaf, S. et al. Comparative analysis of complete plastid genomes from wild soybean (Glycine soja) and nine other Glycine species. PLoS ONE 12, e0182281 (2017).
77. Wu, C.-S. & Chaw, S.-M. Evolutionary stasis in cycad plastomes and the first case of plastome GC-biased gene conversion. Genome Biol. Evol. 7, 2000–2009 (2015).
78. Khan, A. L. et al. Comparative chloroplast genomics of endangered Euphorbia species: Insights into hotspot divergence, repetitive sequence variation, and phylogeny. Plants 9, 199 (2020).
79. Huang, H., Shi, C., Liu, Y., Mao, S.-Y. & Gao, L.-Z. Thirteen Camellia chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. BMC Evol. Biol. 14, 1–17 (2014).
80. Asaf, S., Yan, J., Khan, A. L. & Lee, I.-J. Complete chloroplast gene characterization of oxtails crucifera and its comparison with related species from family oxalidaceae. Plants 9, 928 (2020).
81. Yang, J.-B., Yang, S.-X., Li, H.-T., Yang, J. & Li, D.-Z. Comparative chloroplast genomes of Camellia species. Molecules 20, 437 (2015).
82. Deng, N. et al. Complete chloroplast genome sequences and codon usage pattern among three wetland plants. Agron. J. 113, 840–851 (2021).
83. Zhou, J. et al. Complete chloroplast genomes of papaver rheas and papaver orientale: Molecular structures, comparative analysis, and phylogenetic analysis. Molecules 23, 437 (2018).
84. Yan, M. et al. The complete chloroplast genomes of punica granatum and a comparison with other species in lythraceae. Int. J. Mol. Sci. 20, 2886 (2019).
85. Somaratne, Y., Guan, D.-L., Wang, W.-Q., Zhao, L. & Xu, S.-Q. The complete chloroplast genomes of two lispediae species: insights into codon usage bias, RNA editing sites, and phylogenetic relationships in desmodieae (fabeceae: papilionoidaeae). Plants 9, 51 (2020).
86. Asaf, S. et al. Complete chloroplast genome of nicotiana otophora and its comparison with related species. Front. Plant Sci. https://doi.org/10.3389/fpls.2016.00843 (2016).
87. Hong, S.-Y. et al. Complete chloroplast genome sequences and comparative analysis of chenopodium quinoa and C. album. *Front. Plant Sci.* https://doi.org/10.3389/fpls.2017.01696 (2017).

88. Asaf, S. et al. Complete chloroplast genome of Nicotiana otophora and its comparison with related species. *Front. Plant Sci.* 7, 843 (2016).

89. Zhao, X.-L. & Zhu, Z.-M. Comparative genomics and phylogenetic analyses of christia vespertilionis and urariopsis brevissima in the tribe desmodieae (fabaceae: papilionoideae) based on complete chloroplast genomes. *Plants* 9, 1116 (2020).

90. Guo, Y.-Y., Yang, J.-X., Bai, M.-Z., Zhang, G.-Q. & Liu, Z.-J. The chloroplast genome evolution of Venus slipper (Paphiopedilum): IR expansion, SSC contraction, and highly rearranged SSC regions. *BMC Plant Biol.* 21, 1–14 (2021).

91. Raubeson, L. A. et al. Comparative chloroplast genomics: analyses including new sequences from the angiosperms Nuphar advena and Ranunculus macranthus. *BMC Genomics* 8, 1–27 (2007).

92. Park, I. et al. The complete chloroplast genome sequences of fritillaria ussuriensis maxim. and Fritillaria cirrhosa D. Don, and comparative analysis with other fritillaria species. *Molecules* 22, 982 (2017).

93. Shaw, J., Lickey, E. B., Schilling, E. E. & Small, R. L. Comparison of whole chloroplast genome sequences to choose non-coding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *Am. J. Bot.* 94, 275–288 (2007).

94. Khakhlova, O. & Bock, R. Elimination of deleterious mutations in plastid genomes by gene conversion. *Plant J.* 46, 85–94 (2006).

95. Wang, G.-Y., Meng, Y., Huang, J.-L. & Yang, Y.-P. Molecular phylogeny of ophiopogon (Asparagaceae) inferred from nuclear and plastid DNA sequences. *Syst. Bot.* 39, 776–784 (2014).

96. Lun-Kai, D., Song-Yun, L., Lun-Kai, D. & Song-Yun, L. Epidermal features of leaves and their taxonomic significance in subfamily Ophiopogonoideae (Liliaceae). *J. Syst. Evol.* 29, 335 (1991).

97. Song-Yun, L. & Lun-Kai, D. Pollen morphology and generic phylogenetic relationships in Ophiopogonoideae (Liliaceae). *J. Syst. Evol.* 30, 427 (1992).

98. Zhang, D. Chromosomal study and an insight into systematics of the tribe Ophiopogoneae (Endl.) Kunth. *Ph. D. Dissertation, Inst Bot, Chinese Acad Sci* (1991).

99. Leebens-Mack, J. et al. Identifying the basal angiosperm node in chloroplast genome phylogenies: sampling one’s way out of the Felsenstein zone. *Mol. Biol. Evol.* 22, 1948–1963 (2005).

**Author contributions**

A.L.K., A.K. and S.A. performed experiments; A.L.K., S.A. and W.A. wrote the original draft and Bioinformatics analysis: A.R. collected samples, A.L.K. and A.H. supervision and arranging resources. All authors have read and approved the manuscript.

**Competing interests**

The authors declare no competing interests.

**Additional information**

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-20304-6.

**Correspondence** and requests for materials should be addressed to A.A.-H. or A.L.K.

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

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

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

© The Author(s) 2022