The first complete chloroplast genome sequences in Resedaceae: Genome structure and comparative analysis

Dhafer Alzahrani¹, Enas Albokhari¹,2, Abidina Abba¹,3 and Samaila Yaradua¹,4
¹Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia
²Department of Biological Sciences, Umm Al-Qura University, Makkah, Saudi Arabia
³Department of Biological Sciences, Federal University Lokoja, Lokoja, Kogi State, Nigeria
⁴Centre for Biodiversity and Conservation, Department of Biology, Umaru Musa Yaradua University, Katsina, Nigeria

Abstract
Caylusea hexagyna and Ochradenus baccatus are two species in the Resedaceae family. In this study, we analysed the complete plastid genomes of these two species using high-throughput sequencing technology and compared their genomic data. The length of the plastid genome of C. hexagyna was 154,390 bp while that of O. baccatus was 153,380 bp. The lengths of the inverted repeats (IR) regions were 26,526 bp and 26,558 bp, those of the large single copy (LSC) regions were 83,870 bp and 83,023 bp; and those of the small single copy (SSC) regions were 17,468 bp and 17,241 bp in C. hexagyna and O. baccatus, respectively. Both genomes consisted of 113 genes: 79 protein-coding genes, 30 tRNA genes and 4 rRNA genes. Repeat analysis showed that the plastid genome included all types of repeats, with more frequent occurrences of palindromic sequences. Comparative studies of SSR markers showed that there were 256 markers in C. hexagyna and 255 in O. baccatus; the majority of the SSRs in these plastid genomes were mono-nucleotide repeats (A/T). All the clusters in the phylogenetic tree had high support. This study reported the first complete plastid genomes of the genera Caylusea and Ochradenus and the first for the Resedaceae family.

Corresponding authors:
Dhafer Alzahrani, Department of Biological Sciences, Faculty of Sciences, King Abdulaziz University, P.O. Box 80203, Jeddah 21589, Saudi Arabia.
Email: dalzahrani@kau.edu.sa, dhaferalzahrani@hotmail.com
Enas Albokhari, Department of Biological Sciences, Faculty of Sciences, King Abdulaziz University, P.O. Box 80203, Jeddah 21589, Saudi Arabia.
Email: ejbokhary@uqu.edu.sa

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (https://us.sagepub.com/en-us/nam/open-access-at-sage).
Keywords
Caylusea hexagyna, Ochradenus baccatus, Resedaceae, genetics, chloroplast genome (cp), phylogenetic relationships

Introduction

Resedaceae S.F. Gray is the smallest family in the Brassicales comprising 107 species in 12 genera,¹ and is distributed in arid to subtropical regions of Europe, Asia, the Middle East, North America, Mesoamerica, the Caribbean, India and Africa.²,³ Resedaceae was in the order Rhoedales,⁴ Later, Rhoedales was renamed to Resedales.⁵ Resedaceae was then moved to the order Capparidales.⁶ This order (Capparidales) was renamed taxonomically into Capparales.⁷–¹⁰ Finally, the order Capparales was renamed as Brassicales.¹¹–¹⁴ Serval taxonomic studies have moved some genera from other families into the Resedaceae family such as Borthwickia (formerly Borthwickiaceae), Neothorelia, Stixis and Tirania (formerly Stixidaceae) and Forchhammeria (formerly in Capparaceae) which has caused the Resedaceae family to expand.¹³,¹⁴

Caylusea hexagyna (Forsk.) M. L. is a green annual or perennial herb that is distributed in North and East Africa, the Middle East and Iran.¹⁵ Ochradenus baccatus Delile is a perennial shrub that is distributed in Ethiopia, Egypt, Libya, the Middle East, Iran and parts of Pakistan.¹⁶–¹⁸ Ochradenus baccatus has antioxidant and anti-inflammatory activities, with uses in the treatment of diarrhea and colonic ulceration.¹⁹

Chloroplast genomes are similar across most angiosperms; in terms of their coding regions, sizes, tRNA genes, and other related genes, even though the nature of chloroplasts may vary, as reported by different authors.²⁰–²² Complete chloroplast genomes have been used as a vital tool for re-examining the phylogenetic relationships of complex taxa; these genomes have recently been used to resolve some unanswered questions in plant taxonomy.²³

This study provides the first insights into the systematic position of two species (C. hexagyena and O. baccatus) and provides valuable resources for further phylogenetic and evolutionary relationship studies of the species in the Resedaceae family to resolve the taxonomic problems of species such as: Borthwickia, Neothorelia, Stixis, Tirania and Forchhammeria.

Materials and methods

Plant material and DNA extraction

Fresh young leaf materials for C. hexagyena were collected from the Al-Shafa region (21.1561°N 40.3672°E) and those for O. baccatus were collected from the Al-Taif region (21.2999°N 40.3768°E), Saudi Arabia. Total genomic DNA was extracted from the samples using a Qiagen genomic DNA extraction kit (Qiagen, Inc., Germany) according to the manufacturer’s protocols.

Library construction, sequencing and assembly

A total amount of 1.0 µg DNA was used as input material for the DNA sample preparations. The NEBNext DNA Library Prep Kit was used to generate sequence libraries
according to the manufacturer’s recommendation; indices were also added to each sample. Genomic DNA was randomly fragmented by shearing to a size of 350 bases in length. The ends of randomly fragmented DNA were repaired and A-tailed, adapters were ligated with NEBNext for Illumina sequencing, then the PCR improved by P5 and indexed P7 oligo sequences. The AMPure XP system was used to purify the PCR products; subsequent findings were analyzed by the Agilent 2100 Bioanalyzer for size distribution and later quantified using real-time PCR. The total genomic DNA was sequenced using an Illumina Hiseq 2500. The raw reads data were subjected to PRINSEQlite to obtain clean data (5Gb). The generated reads were assembled using NOVOPlasty with kmer (K-mer = 33) to assemble the complete chloroplast genome from the whole genome sequence. The \textit{ndhF} intergenic spacer from \textit{Cleome chrysantha} (KF923088) was used as the seed, and the plastome sequence of \textit{Stapelia gigantea} (MG963259) was used as the reference for the assembly of the \textit{C. hexagena} cp genome. For \textit{O. baccatus}, the \textit{accD} gene from \textit{Brassica juncea} (JF801163) was used as the seed, and the plastome sequence of \textit{Brassica juncea} (NC_028272) was used as the reference. Finally, for each species, one contig containing the complete chloroplast genome sequence was generated.

**Gene annotation**

In \textit{C. hexagyna} and \textit{O. baccatus} genes were annotated using DOGMA (Dual Organellar GenoMe Annotator, University of Texas at Austin, Austin, TX, USA). The positions of start and stop codon were adjusted manually. tRNA genes were identified by the \textit{trnAscan-SE} server (http://lowelab.ucsc.edu/tRNAscan-SE/). The circular chloroplast genome maps were drawn using OGDRAW (Organellar Genome DRAW). The sequences of the chloroplast genome were deposited in the GenBank database: \textit{C. hexagyna} (MT948187) and \textit{O. baccatus} (MT948189).

**Sequence analysis**

The relative synonymous codon usage (RSCU) values, base composition and codon usage were analysed using MEGA 6.0 software. Potential RNA editing sites present in the protein-coding genes were predicted by the PREP suite with the cut-off value set to 0.8.

**Repeat analysis of the chloroplast genome**

The online software MLicroSAtellite (MISA) was used to identify simple sequence repeats (SSRs) in the chloroplast genome with the following parameters: eight, five, four and three repeat units for mononucleotide, dinucleotide, trinucleotide and tetra-, penta-, and hexa-nucleotides SSR motifs, respectively. REPutter software (https://bibiserv.cbi.technion.ac.il/repeter) was used with default settings to detect the size and location of the long repeats in the two Cleomaceae species.
**Genome comparison**

Comparisons were made against the border region between inverted repeat (IR), large single copy (LSC) and small single copy (SSC) among the two species of Resedaceae chloroplast genomes.

**Characterization of substitution rate**

The nonsynonymous (dN) and synonymous (dS) substitution rates were calculated using DNAsp v5.10.01. The chloroplast genome of *C. hexagyna* was compared with the chloroplast genome of *O. baccatus* to identify the genes that are under selective pressure. Geneious software v. 8.1.3 (Biomatters, Ltd, Auckland, New Zealand) was used to align the individual protein-coding genes separately, and the aligned sequences were then translated into protein sequences.

**Phylogenetic analysis**

The analysis was conducted based on the complete chloroplast genome sequences of two Resedaceae species (*C. hexagyna* and *O. baccatus*), three Capparaceae species, four Cleomaceae species, eight Brassicaceae species and two Malvaceae species, which were used as outgroups. All of the sequences were aligned using MAFFT with default settings. The phylogenetic trees were reconstructed based on the maximum parsimony (MP) method using PAUP version 4.0b10 and a heuristic search strategy of 1000 random sequence addition replicates with tree bisection–reconnection (TBR) branch swapping. A maximum of 100 trees per replicate were saved, with MulTrees turned on and gaps treated as missing data. Statistical support was assessed for clades with nonparametric bootstrap analysis using 1000 bootstrap replicates. A 50% majority-rule consensus tree was calculated from all the most parsimonious trees. Bayesian inference (BI) analysis were performed in MrBayes version 3.2.6 and the GTR models were selected using jModelTest version 3.7.

**Results**

**Characteristics of *C. hexagyna* and *O. baccatus* chloroplast genomes**

The complete chloroplast genome sequence has a circular and quadripartite structure. The total length of the *C. hexagyna* chloroplast genome is 154,390 bp, and that of *O. baccatus* is 153,380 bp. The chloroplast genome has four distinct regions: a large single copy (LSC), which a length of 83,870 bp in *C. hexagyna* and 83,023 bp in *O. baccatus*, a small single copy (SSC) region with a length of 17,468 bp in *C. hexagyna* and 17,241 bp in *O. baccatus*; and a pair of inverted repeats (IRa and IRb) that separate the SSC and LSC with lengths of 26,526 bp in *C. hexagyna* and 26,558 bp in *O. baccatus* (Figure 1). The region coding for genes *C. hexagyna* is 76,620 bp in length, which constitutes 59.62% of the chloroplast genome, in *O. baccatus*, this region is 77,268 bp in length, which constitutes 50.37% of the chloroplast genome. The remainder of the
chloroplast genome is the noncoding region, which includes introns and intergenic spacers with lengths of 68,293 bp (44.23%) in *C. hexagyna* and is 67,177 bp (43.79%) in *O. baccatus*. The GC percentage in the inverted repeat regions is higher than that in the large single copy and small single copy regions. The IRa and IRb regions have a GC content of 42.2% in both species, while the LSC and SSC regions have GC contents of 34.25% and 29.8%, respectively, in *C. hexagyna* and 34.16% and 29.49%, respectively, in *O. baccatus* (Table 1).

The complete chloroplast genomes of *C. hexagyna* and *O. baccatus* contained a total of 134 genes, 113 were unique, and 19 genes were duplicated in the pair of IR regions. There were 79 protein-coding genes, 30 tRNAs and four rRNAs in the chloroplast genomes (Figure 1 and Table 2). Eight protein-coding genes were located in the IR region, along with seven tRNAs and four rRNAs. Sixty protein-coding genes and 22...
tRNA genes were located in the LSC region, while the SSC region contained 12 protein-coding genes and one tRNA. Similar to most flowering plants (angiosperms), all the protein-coding genes in the chloroplast genome start with methionine (ATG codon), whereas some genes start with other codons, such as ATC, GTG and ACG.35–37

Similar to those in all angiosperms, the chloroplast genomes in both species contained introns in some of the protein-coding and tRNA genes.35,36 Among the 113 unique genes, 13 contained introns, eight of those were protein-coding genes and five were tRNAs (Table 3). Eight genes were found in the LSC region, with four genes in the IR region and only one gene in the SSC region. Eleven genes had only one intron, and two genes had two introns, *ycf3* and *clpP*. The *trnK-UUU* gene had the longest intron (2559 bp in *C. hexagyna* and 2554 bp in *O. baccatus*) (Table 3).

### Codon usage

The nucleotide sequences of protein-coding genes and tRNA genes were used to compute the frequency of the codon usage in chloroplast genomes; these sequence lengths were 79,613 bp in *C. hexagyna* and 80,236 bp in *O. baccatus*. The relative synonymous codon usage of the genes in these genomes is presented in S1 and S2 Tables. The results showed that the genes in the chloroplast genome were encoded by 26,439 codons in *C. hexagyna* and by 26,646 codons in *O. baccatus*. Codons coding for leucine were the highest in both genomes, with 2719 (10.28%) in *C. hexagyna* and 3697 (13.87%) in *O. baccatus*, whereas those coding for cysteine were the least abundant, with 325 (1.22%) in *C. hexagyna* and 477 (1.79%) in *O. baccatus* (Figure 2). The results (S1, S2 Table) showed that all amino acids have codon usage bias, except for tryptophan and methionine, which have RSCU values of 1.

---

**Table 1. Base composition in the *C. hexagyna* and *O. baccatus* chloroplast genomes.**

| Species    | *C. hexagyna* | *O. baccatus* |
|------------|---------------|---------------|
| Genome size (bp) | 154,390 | 153,380 |
| IR (bp)   | 26,526 | 26,558 |
| LSC (bp)  | 83,870 | 83,023 |
| SSC (bp)  | 17,468 | 17,241 |
| Total number of genes | 134 | 134 |
| rRNA      | 4 | 4 |
| tRNA      | 30 | 30 |
| Protein-coding genes | 79 | 79 |
| T (U) %   | 32.12 | 32.13 |
| C %       | 18.42 | 18.51 |
| A %       | 31.36 | 31.44 |
| G %       | 18.07 | 17.91 |
| Overall GC content % | 36.49 | 36.42 |
| GC content in LSC % | 34.25 | 34.16 |
| GC content in SSC % | 29.8 | 29.49 |
| GC content in IR % | 42.2 | 42.2 |
RNA editing sites

The program PREP suite was used to predict the RNA editing sites in both chloroplast genomes. The first nucleotide of the first codon was used in all analyses. The results showed that most of the conversions in the codon positions were from the amino acid serine to that of leucine (S3, S4 Table). In total, the program revealed 41 editing sites distributed within 15 protein-coding genes in the genome of *C. hexagyna* and 39
editing sites distributed within 14 protein-coding genes in the genome of *O. baccatus*. In both genomes: *ndhD* had the highest number of editing sites (nine sites), *ndhB* had seven editing sites, and *ndhF* had five editing sites, the remaining genes ranged from four to one editing site: *accD, atpF, clpP, matK, ndhA, ndhG, psaI, rpoB, rpoC2, rps2* and *rps14* in both genomes and *rps16* in *C. hexagyna* (S3, S4 Table). Conversions of proline

**Table 3.** Length of introns and exons in the *C. hexagyna* and *O. baccatus* chloroplast genomes.

| Gene Location | Genes     | Exon I (bp) | Intron I (bp) | Exon II (bp) | Intron II (bp) | Exon III (bp) |
|---------------|-----------|-------------|---------------|--------------|----------------|---------------|
|               |           | C. h        | O. b          | C. h         | O. b           | C. h          | O. b         |
| **rps16**     | LSC       | 230         | 224           | 913          | 935            | 35            | 40           |
| **atpF**      | LSC       | 410         | 410           | 713          | 715            | 143           | 143          |
| **rpoC1**     | LSC       | 1612        | 1612          | 785          | 801            | 429           | 429          |
| **ycf3**      | LSC       | 154         | 154           | 804          | 808            | 225           | 225          |
| **clpP**      | LSC       | 227         | 227           | 576          | 581            | 291           | 291          |
| **rpl2**      | IR        | 433         | 433           | 684          | 684            | 390           | 390          |
| **ndhB**      | IR        | 752         | 752           | 686          | 686            | 776           | 776          |
| **ndhA**      | SSC       | 551         | 530           | 1070         | 1064           | 539           | 551          |
| **trnK-UUU**  | LSC       | 34          | 34            | 2559         | 2554           | 36            | 36           |
| **trnL-UAC**  | LSC       | 36          | 36            | 433          | 435            | 49            | 49           |
| **trnI-GAU**  | IR        | 41          | 31            | 946          | 947            | 34            | 34           |
| **trnA-UGC**  | IR        | 37          | 37            | 797          | 797            | 34            | 34           |

**Figure 2.** Amino acid frequencies in the protein-coding sequences of *C. hexagyna* and *O. baccatus* chloroplast genomes protein-coding sequences.
(nonpolar) to serine (polar) were observed in the RNA editing site. The remaining genes (atpA, atpB, atpI, ccsA, petB, petD, petG, petL, psaB, psbB, psbE, psbF, psbL, rpl2, rpl20, rpl23, rpoA, rpoC1, rps8 and ycf3, in either genome and rps16 in O. baccatus) do not have predicted RNA editing sites in the first nucleotide of the first codon.

**Repeat analysis**

**Long repeats.** The program REPuter was used to identify long repeat sequences present in the *C. hexagyna* and *O. baccatus* chloroplast genomes using default settings; from the results, it was discovered that all four types of repeats (palindromic, forward, reverse and complement) were present in both chloroplast genomes (S5, S6 Table). The analysis of *C. hexagyna* and *O. baccatus* showed 20 and 22 palindromic repeats, 11 and 10 forward repeats, 14 and 12 reverse repeats and 4 and 5 complement repeats, respectively (S5, S6 Table and Figure 3). In total, there were 49 repeats in both chloroplast genomes. The majority of the repeat sizes in *C. hexagyna* were between 10 and 19 bp (51.02%), followed by those between 20 and 29 bp (38.77%) and 30 and 39 bp (10.2%); in *O. baccatus*, the majority of the repeat sizes were between 20 and 29 bp (93.87%), followed by those between 30 and 39 bp (6.12%). The intergenic spacer region in *C. hexagyna* and *O. baccatus* harboured 65.3% and 83.67% of the repeats, respectively. In *C. hexagyna*, the tRNA region contained nine repeats (18.36%), while 12 repeats (24.48%) were located in the protein-coding genes; in *O. baccatus*, the tRNA region contained five repeats (10.2%), and the protein-coding genes contained six repeats (12.24%) (S5, S6 Table).

**Simple sequence repeats (SSRs).** Short sequence repeats, or microsatellites (SSRs) are dispersed throughout the genome. The chloroplast genome of *C. hexagyna* contained 256 microsatellites, and there were 255 microsatellites in *O. baccatus* (S7, S8 Table).

![Figure 3. Number of different repeats in the chloroplast genomes of *C. hexagyna* and *O. baccatus*. P = palindromic, F = forward, R = reverse and C = complement.](image)
In the chloroplast genome of both species, the majority of SSRs were mononucleotides (89.4%), of which the majority were poly A or poly T (Figure 4 and Table 4). Poly A (polyadenine) constituted 38.67% of the mononucleotides in C. hexagyna and 38.43% in O. baccatus, whereas poly T (polythymine) constituted 48.43% of the mononucleotides in C. hexagyna and 49% in O. baccatus. Only three poly C (polycytosine) mononucleotides were present in C. hexagyna (1.17%), with four in O. baccatus (1.56%); there were two poly G (polyguanine) mononucleotides in both species (0.78%). In the genomes of both species, only one dinucleotide (AT/AT) was found in the genome, with one trinucleotide (AAT/ATT), two tetranucleotides (AAAG/CTTT and AAAT/ATT). Only one pentanucleotide (AAAAT/ATTT) was found in the genome of O. baccatus.

![Figure 4. Frequency of different SSR motifs in different repeat types in the C. hexagyna and O. baccatus chloroplast genomes.](image)

![Table 4. The SSRs in two chloroplast genomes of Resedaceae.](table)

| SSR type | Repeat unit     | Species     |
|----------|----------------|-------------|
| Mono     | A/T            | 223         |
|          | C/G            | 6           |
| Di       | AT/AT          | 16          |
| Tri      | AAT/ATT        | 3           |
| Tetra    | AAAG/CTTT      | 1           |
|          | AAAT/ATTT      | 5           |
| Penta    | AAAAG/CTTTTT   | 1           |
|          | AAAAT/ATTTTT   | 1           |
| Hexa     | AATAGG/ATTCCT  | 0           |

Table 4. The SSRs in two chloroplast genomes of Resedaceae.
ATTT), one pentanucleotide (AAAAG/CTTTT in *C. hexagyna* and AAAAT/ATTTT in *O. baccatus*), and one hexanucleotide (AATAGG/ATTCCT in *O. baccatus* but none in *C. hexagyna*) (Figure 4 and Table 4). Microsatellites were more prevalent in the intergenic spacer region in *C. hexagyna* and *O. baccatus* (83.59% and 82.35% respectively) than in the coding region (17.96% and 18%, respectively) (Figure 5).

**Figure 5.** Number of SSR types in the complete genomes, protein-coding regions and noncoding genes in *C. hexagyna* and *O. baccatus.*

Comparative analysis. In the current study, the IR-LCS and IR-SSC boundaries of the genomes of the two Resedaceae species were compared. Although the results showed that there were similarities between the cp genomes of both species (Figure 6), *O. baccatus* had the smallest chloroplast genome (153,380 bp), whereas *C. hexagyna*...
had the largest chloroplast genome (154,390 bp). The IR region was 26,526 bp in C. hexagyna and 26,251 bp in O. baccatus. Furthermore, the lengths of the LSC regions were 83,870 bp in C. hexagyna and 83,023 bp in O. baccatus, and those of the SSC regions were 17,468 bp in C. hexagyna and 17,241 bp in O. baccatus. Additionally, comparative analysis of the cp genome of these two Resedaceae species revealed that the location of the rpsl9 gene is between the LSC and IRb regions. The ycf1 gene was located at the boundary of the IRb/SSC regions in C. hexagyna (1051 bp/4225 bp) and O. baccatus (1036 bp/145 bp) and the boundary of the SSC/IRa regions in C. hexagyna (2 bp/45 bp) and O. baccatus (4247 bp/1036 bp). The ndhF was found at the boundary of the IRb/SSC regions, with 43 bp in the IRb region and 2195 bp in the SSC region, in O. baccatus; it was found at the boundary of the SSC/IRa regions, with 2192 bp in the SSC region and 45 bp in the IRa region, in C. hexagyna.

**Divergence of protein-coding gene sequence.** The rates of synonymous (dS) and nonsynonymous (dN) substitutions and the dN/dS ratio were calculated to detect the selective pressure on 79 protein-coding genes in the plastid genome of two Resedaceae species. The results showed that the dN/dS ratio was less than 1 in all of the paired genes except rpl22 and ycf2 in C. hexagyna vs O. baccatus, with values of 1.14 and 5.57, respectively. The synonymous (dS) substitution values in all the genes ranged from 0 to 0.12 (Figure 7).

**Phylogenetic analysis.** Phylogenetic relationships based on the Bayesian and maximum parsimony analysis placed all samples into four main clades, and the results matched across the two analyses with strong support for nodes PP and MP, both 1.00 (Figure 8). The two Resedaceae species formed one clade, while the other three families formed in one main clade includes three subclades. The first subclade contained three species of the Capparaceae family. The second subclade contained four Cleomaceae species which was found to be sister to Brassicaceae, while the third clade included eight species from the Brassicaceae family. The phylogenetic tree showed the taxonomic position of the family Resedaceae in the order Brassicales.13,14
Discussion

The complete chloroplast genome has provided much genetic information and many molecular markers as valuable tools to investigate obscure phylogenetic relationships among land plants.\(^{38}\) The plastid genomes of \textit{C. hexagyna} and \textit{O. baccatus} have similar structures to the chloroplast genomes of other angiosperms.\(^ {35,36,39}\) This study is the first to report the complete plastid genomes for the genera \textit{Caylusea} and \textit{Ochradenus} and for Resedaceae in general.

The plastid genome size of \textit{C. hexagyna} is 154,390 bp, and that of \textit{O. baccatus} is 153,380 bp (Figure 1). The chloroplast genome sequences of \textit{C. hexagyna} and \textit{O. baccatus} had GC contents of 36.49\% and 36.42\% and AT contents of 63.48\% and 63.57\%, respectively (Table 1). The GC content in the IR region was 42.2\% in both species, which was higher than that of the LSC and SSC regions (Table 1).

There were 113 genes in both genomes, which included 79 protein-coding genes, 30 tRNA genes and four rRNA genes. The IR region contained eight protein-coding genes, seven tRNA genes and four rRNA genes; the LSC region contained 60 protein-coding genes and 22 tRNA genes; and the SSC region contained 12 protein-coding genes and one tRNA gene. Introns are present in several of the protein-coding and tRNA genes in both chloroplast genomes, similar to other chloroplast genomes of flowering plants.\(^ {35,36,39}\) Out of the total genes in the cp genomes of both species, 13 genes contained introns, eight were protein-coding genes and five were tRNA genes (Table 3).

The results showed that the genes in the chloroplast genome were encoded by 26,439 codons in \textit{C. hexagyna} and 26,646 codons in \textit{O. baccatus}. The most common codons

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure7.png}
\caption{The synonymous (dS) subsitution and dN/dS ratio values of 79 protein-coding genes from two Resedaceae plastid genomes.}
\end{figure}
coded for the amino acid leucine, which has been previously stated for several cp genomes of flowering plants. The RNA editing site results revealed that most of the amino acid conversions in the codon positions were serine to leucine, presenting editing sites distributed among 15 protein-coding genes in *C. hexagyna* and 39 editing sites distributed among 14 protein-coding genes in *O. baccatus*. The repeat sequences were identified in the chloroplast genomes of *C. hexagyna* and *O. baccatus* using default settings, and the long repeat analysis showed 20 and 22 palindromic repeats, 11 and 10 forward repeats, 14 and 12 reverse repeats and 4 and 5 complement repeats, respectively (Figure 3). The length of repeated sequences in the two chloroplast genomes ranged from 10 to 69 bp, which is analogous to the lengths in other angiosperm

Figure 8. Phylogenetic tree reconstruction based on the complete chloroplast genome of sixteen taxa inferred from Bayesian inference (BI) methods showing relationships within resedaceae and other families in brassicales. Numbers in the clade represent posterior probability (PP) values.
plants. The majority of SSRs in the cp genomes were mononucleotides, of which most were poly T and A (Figure 4 and Table 4) as in most flowering plant cp genome.

This study compared the IR-LCS and IR-SSC boundaries of the two plastid genomes of Resedaceae, and the results showed that O. baccatus had the smallest chloroplast genome. The smallest IR region was in C. hexagyna, and the smallest LSC and SSC regions were in O. baccatus. The results of the rates of the selective pressure among 79 protein-coding genes in the chloroplast genomes of the two Resedaceae species revealed that the dN/dS ratio was less than 1 in all of the paired genes, with a few exceptions (Figure 7). The identical (dS) values in all of the protein-coding genes range from 0 to 0.12 (Figure 7). Phylogenetic relationships based on the Bayesian and maximum parsimony analysis placed all samples into four main clades, where every family was in a separate clade (Figure 8). The phylogenetic tree showed that the family Resedaceae were sister to all other families in the Brassicales.

**Conclusion**

This current study used the Illumina HiSeq 2500 platform to sequence the complete chloroplast genome of Caylusea hexagyna and Ochradenus baccatus, which provided valuable plastid genomic resources for these plants. We annotated the cp genome for both species; moreover, we identified the base composition, codon usage and RNA editing sites, SSRs and long repeats in these genomes. This study has also provided the first insight into the systematic position of two species (C. hexagyna and O. baccatus) whereby these valuable resources could be used for further phylogenetic and evolutionary relationship studies of the species in the Resedaceae family in order to resolve the taxonomic problems of species such as: Borthwickia, Neothorelia, Stixis, Tirania and Forchhammeria.

**Authors’ Note**

Conceptualization, D.A., E.A. and S.Y.; collected data, D.A and E.A; formal analysis, E.A. and S.Y.; writing—original draft preparation, D.A. and E.A.; writing—review and editing, S.Y. and A.A.; All authors have read and agreed to the published version of the manuscript. The data that support the findings of this study are openly available in the NCBI GenBank at https://www.ncbi.nlm.nih.gov, reference number (C. hexagyna MT948187; O. baccatus MT948189).

**Declaration of conflicting interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The author(s) received no financial support for the research, authorship and/or publication of this article.

**ORCID iD**

Dhafer Alzahrani https://orcid.org/0000-0003-3549-5907
Supplemental material

Supplemental material for this article is available online.

References

1. Christenhusz MJM and Byng JW. The number of known plants species in the world and its annual increase. *Phytotaxa* 2016; 261: 201–217.
2. Willis JC. *A dictionary of the flowering plants & ferns*, 7th ed. Cambridge, UK: Cambridge University Press, 1973.
3. Mabberley DI. *The plant book*. Cambridge, UK: Cambridge University Press, 1987.
4. Engler A and Prantle K. *Die natürlichen pflanzen familien*, vol. 3 (4 & 5). Leipzig: Engelmann, 1897.
5. Hutchinson J. *Evolution and phylogeny of flowering plants. Dicotyledons: facts and theory with over 550 illustrations and maps by the author*. London, New York: Academic Press, 1969.
6. Takhtajan AL. *Flora armenii 1*. Yerevan: Akademia Nauk, 1954.
7. Cronquist A. *An integrated system of classification of flowering plants*. New York: Columbia University Press, 1981.
8. Dahlgren RMT. General aspects of angiosperm evolution and macrosystematics. *Nord J Bot* 1983; 3: 119–149.
9. Thorne RF. Proposed new alignments in the angiosperms. *Nord J Bot* 1983; 3: 85–117.
10. Takhtajan AL. *Flowering plant*. Russia: Komarov Botanical Institute, 1987.
11. APG I (Angiosperm Phylogeny Group). An ordinal classification for the families of flowering plants. *Ann Mo Bot Gard* 1998; 85: 531–553.
12. APG II (Angiosperm Phylogeny Group). An update of the angiosperm phylogeny group classification for the orders and families of flowering plants. *Bot J Linn Soc* 2003; 141: 399–436.
13. APG III (Angiosperm Phylogeny Group). An update of the angiosperm phylogeny group classification for the orders and families of flowering plants. *Bot J Linn Soc* 2009; 161: 105–121.
14. APG IV (Angiosperm Phylogeny Group). An update of the angiosperm phylogeny group classification for the orders and families of flowering plants. *Bot J Linn Soc* 2016; 181: 1–20.
15. Chaudhary SA. *Flora of the kingdom of Saudi Arabia* vol. 3. Riyadh, Saudi Arabia: Ministry of Agriculture and Water, 2001.
16. Al-Fredan MA. Sexual polymorphism in the dioecious plant species of eastern Saudi Arabia. *I J Plant R Biol* 2010; 2: 85–91.
17. El-Wahab RH A, Zaghoul MS, Kamel WM, et al. Diversity and distribution of medicinal plants in North Sinai, Egypt. *African J Environ Sci Tech* 2008; 2: 157–171.
18. Khan S, Al-Qurainy F and Nadeem M. Biotechnological approaches for conservation and improvement of rare and endangered plants of Saudi Arabia. *Saudi J Biol Sci* 2012; 19: 1–11.
19. Alqasoumi SI, Soliman GA, Awaad AS, et al. Anti-inflammatory activity, safety and protective effects of *Leptadenia pyrotechnica*, *Haloxylon salicornicum* and *Ochradenus baccatus* in ulcerative colitis. *Phuopharmacology* 2012; 2: 58–71.
20. Raubeson LA, Peery R, Chumley TW, et al. Comparative chloroplast genomics: analyses including new sequences from the angiosperms Nuphar advena and Ranunculus macranthus. *BMC Genomics* 2007; 8: 174.
21. Daniell H, Lin CS, Yu M, et al. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol* 2016; 17, article. 2016; 134: 1–29.
22. Yaradua S, Alzahrani D, Albokhari E, et al. Complete chloroplast genome sequence of justicia flava: genome comparative analysis and phylogenetic relationships among Acanthaceae. *BioMed Res Int* 2019; 21: 1–17.
23. Ruhsam M, Rai HS, Mathews S, et al. Does complete plastid genome sequencing improves species discrimination and phylogenetic resolution in araucaria? *Mol Ecol Resour* 2015; 15: 1067–1078.

24. Dierckxsens N, Mardulpyn P and Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res* 2016; 45: 1–9.

25. Wyman S, Jansen R and Boore J. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 2014; 20: 3252–3255.

26. Schattner P, Brooks AN and Lowe TM. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res* 2005; 33: 686–689.

27. Lohse M, Drechsel O and Bock R. Organellar genome DRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr Genet* 2007; 52: 267–274.

28. Kurtz S, Choudhuri JV, Ohlebusch E, et al. Reputer: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res* 2001; 29: 4633–4642.

29. Thiel T, Michalek W, Varshney R, et al. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet* 2003; 106: 411–422.

30. Librado P and Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 2009; 25: 1451–1452.

31. Katoh K and Standley DM. MAFFT Multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013; 30: 772–780.

32. Felsenstein J. Cases in which parsimony or compatibility methods will be positively misleading. *Syst Zool* 1978; 27: 401–410.

33. Fredrik R, Maxim T, Paul VM, et al. Mrbayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic* 2012; 61: 539–542.

34. Posada D. Jmodeltest: phylogenetic model averaging. *Mol Biol Evol* 2008; 25: 1253–1259.

35. Raman G and Park S. The complete chloroplast genome sequence of ampelopsis: gene organization, comparative analysis, and phylogenetic relationships to other angiosperms. *Front Plant Sci* 2016; 341: 7.

36. Park I, Kim WJ, Yeo S-M, et al. The complete chloroplast genome sequences of Fritillaria ussuriensis maxim. In addition, Fritillaria cirrhosa D. Don, and comparative analysis with other Fritillaria species. *Molecules* 2017; 282: 22.

37. Li B, Lin F, Huang P, et al. Complete chloroplast genome sequence of *Decaisnea insignis*: genome organization, genomic resources and comparative analysis. *Sci Rep* 2017; 7: 1–10.

38. Luo J, Hou BW, Niu ZT, et al. Comparative chloroplast genomes of photosynthetic orchids: insights into evolution of the Orchidaceae and development of molecular markers for phylogenetic applications. *PLoS One* 2014; 9: e99016.

39. Chen H, Shao J, Zhang H, et al. Sequencing and analysis of *Strobilanthes cusia* (nees) Kuntze chloroplast genome revealed the rare simultaneous contraction and expansion of the inverted repeat region in angiosperm. *Front Plant Sci* 2018; 9: 324.

40. Liu HJ, Ding CH, He J, et al. Complete chloroplast genomes of *Archiclematis*, *Naravelia* and *Clematis* (Ranunculaceae), and their phylogenetic implications. *Phytotaxa* 2018; 343: 214–226.

41. Greiner S, Wang X, Rauwolf U, et al. The complete nucleotide sequences of the five genetically distinct plastid genomes of *Oenothera*, subsection Oenothera: I. Sequence evaluation and plastome evolution. *Nucleic Acids Res* 2008; 36: 2366–2378.

42. Song Y, Wang S, Ding Y, et al. Chloroplast genomic resource of *Paris* for species discrimination. *Sci Rep* 2017; 7: 3427.
Author biographies

Dhafer Ahmed M. Alzahrani was born in Saudi Arabia. He obtained his PhD in plant systematics, from the Centre for Plant Diversity and Systematics, School of Biological Sciences, Plant Sciences, University of Reading, United Kingdom, in 2011. He is now an associate professor in the Department of Biological Sciences, Faculty of Sciences, King Abdulaziz University, Jeddah, Saudi Arabia. His specialization includes botany, systematics of flowering plants, and molecular phylogeny. His field of interest includes plant diversity, conservation of plants, flora of Saudi Arabia, and plant ecology. He has published more than 34 research papers in reputable journals and presented about 11 papers in various conferences.

Enas Jameel Albokhari was born in Saudi Arabia. He obtained his PhD in Biology with specialization in plant systematics from King Abdulaziz University, Jeddah, Saudi Arabia, in 2021. He is an assistant professor in the Department of Biological Sciences, Faculty of Applied Sciences, Umm Al-Qura University, Makkah, Saudi Arabia. His specialization includes botany, systematics of flowering plants, molecular phylogeny, and next generation. His field of interest includes plant taxonomy, plant diversity, conservation of plants, flora of Saudi Arabia, plant ecology and plant genetics. He has published 18 research papers in reputable journals and presented about four papers in various conferences.

Abidina Abba was born in 1987 at Batsari town in Katsina state Nigeria. He attended Umaru Musa Yaradua University Katsina for BSc Biology, Bayero University Kano for MSc Applied Biology, and then attended King Abdulaziz University, Jeddah, Saudi Arabia, for PhD in Biology (Botany) in 2021. His research interests include molecular systematics and chloroplast evolution in angiosperm.

Samaila Samaila Yaradua was born in Katsina, Katsina state, Nigeria. He obtained his PhD in biology with specialization in plant systematics from King Abdulaziz University, Jeddah. He schooled at Umaru Musa Yaradua University, Katsina, where he obtained his Bachelor’s and Master’s degrees. His area of research interest includes taxonomy, biogeography, genetic diversity, etc. He has published more than 20 research papers in reputable journals and presented about 10 papers in various conferences.