

## Adaptation of a parasitic lifestyle by *Cuscuta gronovii* Willd. ex Roem. & Schult.: large scale gene deletion, conserved gene orders, and low intraspecific divergence

Yang Ni<sup>a\*</sup> , Mei Jiang<sup>b\*</sup>, Haimei Chen<sup>b</sup>, Linfang Huang<sup>b</sup>, Pinghua Chen<sup>a</sup> and Chang Liu<sup>b</sup> 

<sup>a</sup>Key Laboratory of Ministry of Education for Genetics, Breeding and Multiple Utilization of Crops, National Engineering Research Center of Sugarcane, College of Agriculture, Fujian Agriculture and Forestry University, Fuzhou, P. R. China; <sup>b</sup>Key Laboratory of Bioactive Substances and Resource Utilization of Chinese Herbal Medicine from Ministry of Education, Engineering Research Center of Chinese Medicine Resources from Ministry of Education, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, P. R. China

### ABSTRACT

*Cuscuta gronovii* Willd. ex Roem. & Schult. is a parasitic plant and widely distributed in Europe and North America. Here, we compared the plastome sequences of a *C. gronovii* plant collected from China and other 16 *Cuscuta* plastomes, including one *C. gronovii* plastome from North America. The plastome was 86,727 bp in size with a pair of inverted repeats (IRs) of 14,354 bp, a large single-copy (LSC) region of 50,956 bp, and a small single-copy (SSC) region of 7,063 bp. We predicted 97 genes in the plastome, including 61 protein-coding genes, eight ribosomal RNA genes, and 28 transfer RNA genes. We detected a total of 21 microsatellite, 16 tandem, and ten interspersed repeats in the genome. Gene contents analysis of 17 *Cuscuta* plastomes showed the loss of the entire *ndh* gene family. The phylogenomic analysis using 22 shared protein sequences shows that the two *C. gronovii* formed a cluster. Thirteen of *Cuscuta* plastomes showed no rearrangement. The other three showed smaller inversions and smaller numbers of gene deletion. Next, we identified the two INDELS, one SNP site between the two *C. gronovii* plastomes. And we identified two putative RNA editing sites. Lastly, to distinguish between *C. gronovii* and two medicinal *Cuscuta* species, *C. chinensis* and *C. australis*, we identified ten molecular markers based on the genome sequences. The overall large-scale gene loss, conserved gene order, low intraspecific divergence are consistent with the adaptation of *C. gronovii* to a parasitic lifestyle.

### ARTICLE HISTORY

Received 30 September 2020  
Accepted 29 March 2021

### KEYWORDS





*Cuscuta gronovii*; Plastome;  
SNP; RNA editing site;  
Phylogenetic analysis;  
Molecular markers

## Introduction


The *Cuscuta* spp. (dodders) belongs to the genus *Cuscuta* (*Convolvulaceae*, *Solanales*), which contains about 200 plant species, most of which are parasitic (Dawson et al. 1994). *Cuscuta* spp. exhibit massive body plan changes, being leaf- and rootless throughout their life cycles (Sun et al. 2018). The adaptation of the parasitic lifestyle appears to be a continuous process (Cooke and Black 1987). The gradual loss of the chloroplast contents and activities coincide with the degree of parasitism. Based on the chloroplast morphology and photosynthetic capacity, the *Cuscuta* spp. can be divided into three groups. The first group has functional chloroplasts and photosynthetic capacity. A second group has degenerated chloroplasts with restricted photosynthetic activity. A third group are devoid of chloroplasts and are achlorophyllous.

Photosynthesis is a key biological process in plants that are carried out by the organelle chloroplasts. As a result, the loss of chloroplast and photosynthetic activity is critical in adapting to the parasitic lifestyle. The chloroplast genome (plastome) is a circular DNA molecule and encodes proteins, which plays an essential role in photosynthesis (Keeling 2004). One manifestation of the loss of the photosynthetic activity is the deletion of genes from the plastomes. As a result, the plastomes' content and structure were often used to study the parasitic plants' evolution and adaptation.

*C. gronovii* Willd. ex Roem. & Schult. (the plantlist.org, last accessed: 17 August 2020) is initially distributed in Europe and North America. A previous study showed that the plastid genome of *C. gronovii* had suffered extensive deletions by comparing the relative sizes and coding capacities of the

**CONTACT** Pinghua Chen  [phcemail@126.com](mailto:phcemail@126.com)  Key Laboratory of Ministry of Education for Genetics, Breeding and Multiple Utilization of Crops, National Engineering Research Center of Sugarcane, College of Agriculture, Fujian Agriculture and Forestry University, Fuzhou, Fujian Province 350002, P. R. China; Chang Liu  [cliu6688@yahoo.com](mailto:cliu6688@yahoo.com)  Key Laboratory of Bioactive Substances and Resource Utilization of Chinese Herbal Medicine from Ministry of Education, Engineering Research Center of Chinese Medicine Resources from Ministry of Education, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing 100193, P. R. China

\*These authors contributed equally to this work.

 Supplemental data for this article can be accessed [here](#).

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

plastid DNAs of three *Cuscuta* species based on hybridization patterns with tobacco plastid DNA (Berg et al. 2003). Funk et al. compared the plastome differences between *C. gronovii* and *Cuscuta reflexa*, found (a) the parallel losses of genes for the subunits of the plastid-encoded RNA polymerase and its promoters; (b) the loss of *matK* gene, a putative splicing factor; and (c) a significant reduction of RNA editing (Funk et al. 2007).

Our original goal is to study *C. australis*' parasitism by analyzing its plastome's content and structure. *C. australis* is a species listed in the Chinese Pharmacopeia. However, sequence comparison with those from the database suggests that what we have sequenced is most likely to be *C. gronovii*. As a result, we carried out a systematic comparative genomic study of *C. gronovii* plastome and those from another sixteen *Cuscuta* species. We also characterized the intraspecific variations, including Insertion/deletions (INDELs), SNPs, and post-transcriptional modification events such as RNA-editing. Overall, the levels of genome deletion, rearrangement, intraspecific variation, and post-transcriptional modifications are consistent with the adaptation of the *Cuscuta* species to a parasitic lifestyle. The results provide valuable information regarding the gradual transformation of parasitism for *Cuscuta* species.

## Materials and methods

### Plant material, DNA extraction, and sequencing

The fresh stems of *C. gronovii* were collected from the Institute of Medicinal Plant Development (IMPLAD), Beijing, China. Professor Zhao Zhang of IMPLAD identified the samples. The genomic DNA was extracted using a Plant Genomic DNA kit (Tiangen Biotech, Beijing, China) and stored at the Herbarium of IMPLAD with accession number: Implad20170662. The DNA library was constructed with 1ug DNA using the library preparation kit (New England BioLabs, America) and was sequenced on a Hiseq 2500 platform (Illumina, San Diego, CA, USA).

### Genome assembly, validation, and annotation

The raw data pre-processing was conducted with two procedures by the Trimmomatic software (Bolger et al. 2014). First, we cut the adapters for the base reads. The sequence quality standard is then set at quality values of  $Q < 19$  and with more than 5% bases being 'N' according to the reads. The plastome was de novo assembled using NOVOPlasty (v.4.0) (Dierckxsens et al. 2016). We employed 14 *Cuscuta* plastomes as reference sequences (Supplementary Table S1). We used Gepard software to draw Dot plots between the assembled genome and the reference genome to identify the plastome structure (Krumsiek et al. 2007). To validate the assembly's correctness, we mapped sequence reads to the assembled genome and then examined and edited the mapping results using the consed software (Gordon et al. 1998). We used CpGAVAS2 to annotate the plastome. We manually edited the annotations with problems using the Apollo (Misra and Harris 2005) software. The genome sequence and annotation

results have been submitted to the GenBank with accession number MT872375. In the following text, *C. gronovii* plastome refers to the MT872375 sequence without otherwise specified.

### Characteristics and repeat analysis

We used modules embedded in the CpGAVAS2 webserver to analyze codon usage and repeat contents. We identified the microsatellite sequences using Misa software (Sebastian et al. 2017). We identified the tandem repeats using TRF software (Benson 1999) with the size of the repeat unit  $\geq 7$ . We calculated the GC content using Editseq from the DNASTAR Lasergene package (v9) (Burland 2000). We identified the interspersed repeats using VMATCH software (Kurtz et al. 2001).

### Phylogenetic analysis

The plastome sequences of 16 *Cuscuta* species and two out-group species (*Ipomoea batatas* and *Ipomoea purpurea*) were downloaded from GenBank. In total, we found 22 shared DNA sequences (*accD*, *atpA*, *atpB*, *atpH*, *clpP*, *rpl2*, *rpl14*, *rpl16*, *rpl20*, *rpl22*, *rpl33*, *rps2*, *rps3*, *rps4*, *rps7*, *rps8*, *rps11*, *rps12*, *rps15*, *rps18*, *rps19*, *yfc2*) in the 19 plastomes and aligned them with MAFFT software (Katoh et al. 2005). 'TVM + F + G4' was found to be the best evolutionary model, according to BIC (Bayesian Information Criterion) scores. The phylogenetic tree of *C. gronovii* and its closest relatives were built using the IQ-TREE with this model (Nguyen et al. 2015). We used UBoot to perform the bootstrap analysis with 1000 replicates.

### Gene deletion identification of *Cuscuta* genus

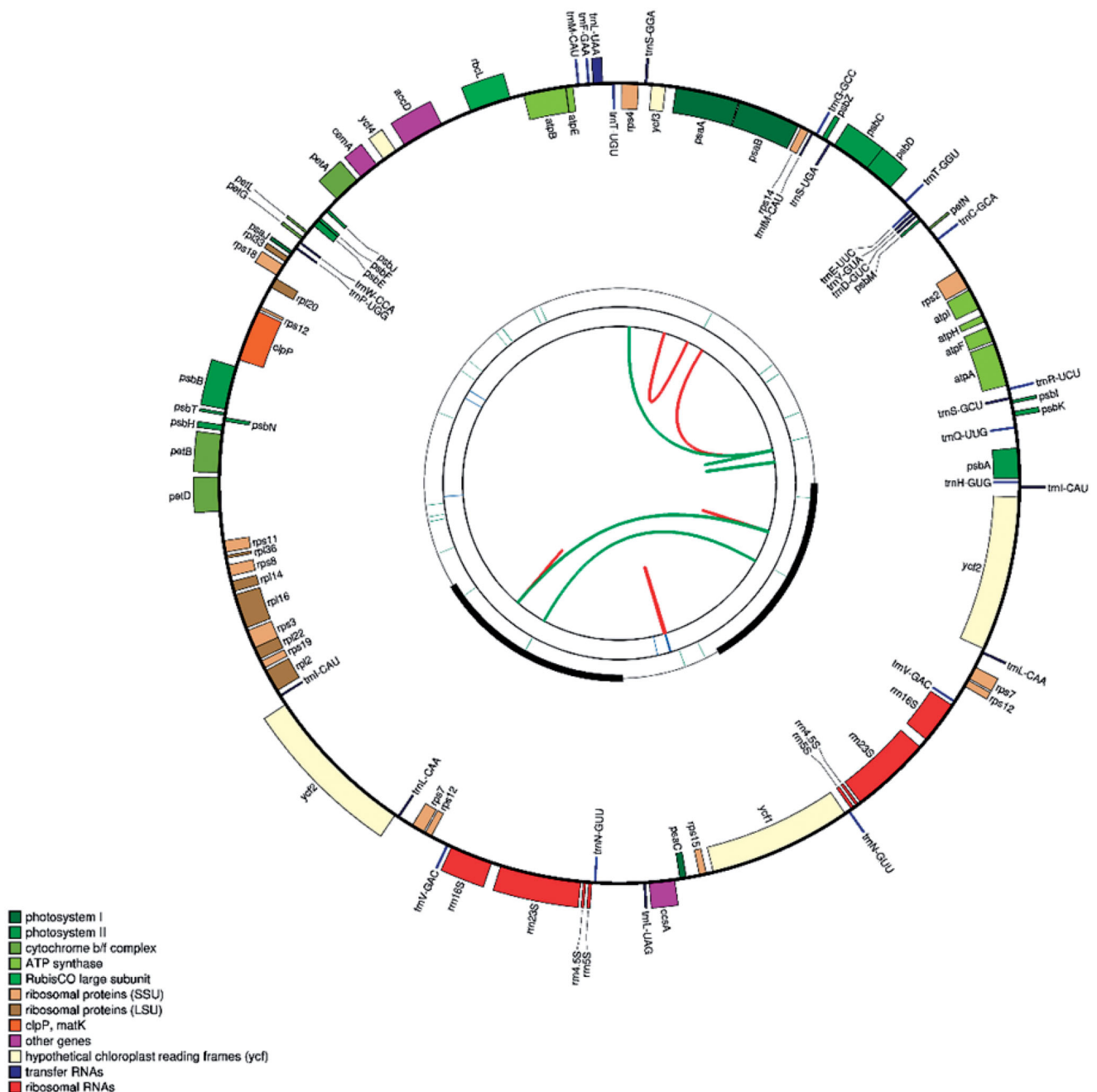
We constructed a blastable database of all plastome sequences of the *Cuscuta* genus using the tool formatdb. The gene belongs to the *ndh* gene family from *Arabidopsis thaliana* were selected as the query to search the database using BLAST with the e-value cutoff of  $1e-10$  and alignment length of more than 50 bp (Madden 2013).

### Discovery of intraspecific divergence between the *C. gronovii* plastomes

Previously, another *C. gronovii* plastome has been published (NC\_009765) (Funk et al. 2007). The two plastome sequences were aligned using the module seqman from the DNASTAR Lasergene package (v9) (Burland 2000) with default parameters. The INDELs and SNPs were then identified manually.

### Identification of RNA editing sites in *C. gronovii* plastome

We downloaded the raw data (SRR SRP200788) of RNA-seq experiments for *C. gronovii* from the GenBank SRA database (<http://www.ncbi.nlm.nih.gov/sra>). The dataset contains reads from the stems of three biological replicates. RNA editing



**Figure 1.** The schematic representation of the plastome of *C. gronovii* created using CPGAVAS2. The map contains four rings. From the center going outward, the first circle shows the forward and reverse repeats connected with red and green arcs, respectively. The next circle shows the tandem repeats marked with short bars. The third circle shows the microsatellite sequences identified using MISA. The fourth circle shows the gene structure on the plastome. The genes were colored based on their functional categories, which are shown in the left corner.

sites were analyzed using REDIttools (Picardi et al. 2015) with the following parameters: coverage  $\geq 5$ , frequency  $\geq 0.1$ , and  $p$ -value  $\leq .05$ .

### Molecular markers development

To identify molecular markers to distinguish three species, we used ecoprimer to identify primers based on the reference genome sequences with the following parameters: -l 300 -L 600 -e 0 -3 2 -t species -T 1 -U -f -O 25. Ecoprimer tend to predict many primer pairs, many of which overlap with each other. These overlapping primer sequences were

merged to longer sequences for primer design using custom scripts.

## Results

### Validation of the assembly and general features of the plastome

We assembled the raw data using the chloroplast genome sequence in the Ref-seq database as reference (Supplementary Table S1). The assembly results of ten of them are in complete agreement. However, the assembly using the four chloroplast genome sequences of *Cuscuta*

*reflexa* (NC\_009766.1), *Cuscuta australis* (NC\_045885.1), *Cuscuta obtusiflora* (NC\_009949.1), *Cuscuta exaltata* (NC\_009963.1) showed different result. The alignment of the assembly results obtained using the different reference sequences were shown in [supplementary file 1](#). We eventually chose the majority consensus results for the analysis.

The entire length of the *C. gronovii* plastome is 86,727 bp, significantly shorter than most plastomes in GenBank. To confirm the correctness of the assembly, we first compared the *C. gronovii* plastome with those of *A. thaliana* ([Supplementary Figure S1](#)) and itself ([Supplementary Figure S2](#)). All the Dot plots have a diagonal line and two perpendicular lines, representing the inverted repeats (IRs). The plastome of *C. gronovii* is largely colinear with that of *A. thaliana*, although it is significantly shorter. Then we mapped the reads to the assembly. There are no low coverage sites observed through the plastome ([Supplementary Figure S3](#)). The junction site result shows that the IR region's coverage depth is two times that of the LSC region ([Supplementary Figure S4, S5](#)). The above results support the high-quality of our plastome assembly.

The plastome of *C. gronovii* has a typical quadripartite structure, consisting of a pair of IR regions of 14,354 bp, a large single-copy (LSC) region of 50,956 bp, and a small single-copy (SSC) region of 7,063 bp ([Figure 1](#)). The genome encodes 97 genes, including 61 protein-coding genes, eight ribosomal RNA (rRNA) genes, and 28 transfer RNA (tRNA) genes ([Supplementary Table S2](#)). Among these genes, three genes (*petB*, *petD*, *rp16*) contain one intron, one gene *clpP* contains two introns, and one tRNA gene (*trnL-UAA*) contains one intron ([Supplementary Table S3](#)). The length of the protein-coding sequence (CDS) in the plastome is 49,104 bp, representing 56.62% of the plastome sequence's total length. In contrast, the rRNA and tRNA genes' size is 8,878 bp and 2,121 bp, representing 10.24% and 2.45% of the plastome sequence's entire length. The GC content analysis showed that the overall GC content is 37.72%, whereas those for the tRNA genes, rRNA genes and for the protein-coding regions are 51.01%, 54.79%, and 37.66%, respectively.

Moreover, a total of 22,437 codons were identified in the plastome of *C. gronovii*. Among these codons, 3,509 codons encode leucine, and 336 codes encode cysteine, respectively, representing the most and least abundant coded amino acids in the *C. gronovii* plastome ([Supplementary Table S4](#)).

### Gene pseudonization and deletion

To explore the *Cuscuta* species' gene deletion scale, we performed a two-way clustering analysis of 16 *Cuscuta* species and use *A. thaliana* as a reference ([Figure 2](#)). The tRNA (*trnK-UUU*) and genes from the *ndh* family (*ndhA*, *ndhB*, *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhI*, *ndhJ*, *ndhK*) were lost in all *Cuscuta* plastomes compared with that of *A. thaliana*. Based on the scale of gene deletion, the 16 *Cuscuta* species can be divided into three groups. *C. japonica*, *C. reflexa*, and *C. exaltata* form one group. They consisted of three tRNA genes (*trnI-GAC*, *trnA-UGC*, *trnR-ACG*) and six protein-coding genes

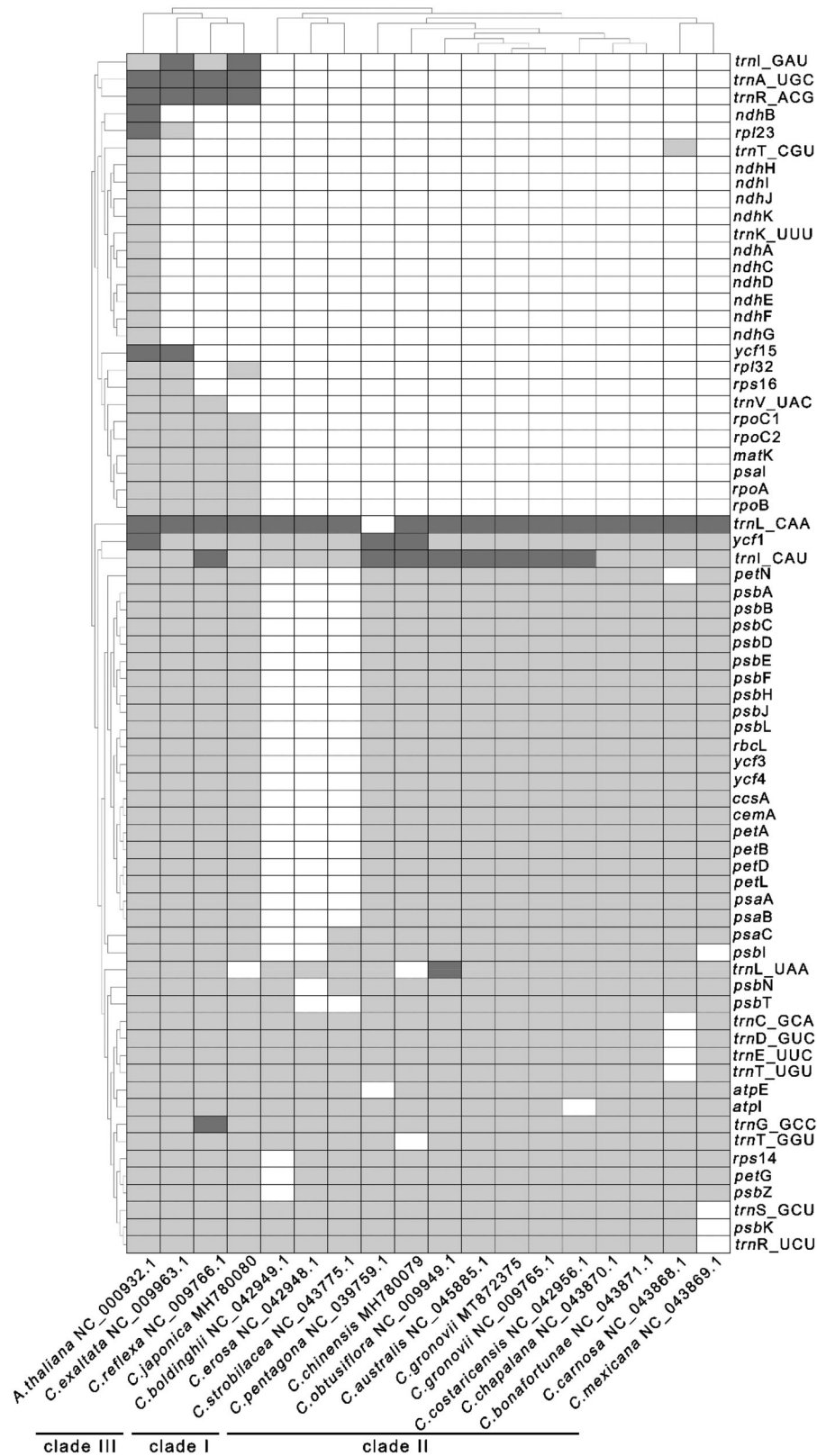
(*rpoA*, *rpoB*, *rpoC1*, *rpoC2*, *psaI*, *matK*) that are missing from the other *Cuscuta* species. *C. boldinghii*, *C. erosa*, *C. strobilacea* form one group. There are 21 common protein-coding genes (*petA*, *petB*, *petL*, *petN*, *psaA*, *psaB*, *psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbG*, *psbH*, *psbJ*, *psbK*, *rbcl*, *ycf3*, *ycf4*, *ccsA*, *cemA*) are missing in those species. And the others form another group. There is no difference in the gene contents between *C. gronovii* (MT872375) and *C. gronovii* (NC\_009765). The results show that there is lower intraspecific diversity in terms of gene contents. To further examine the gene loss in the *ndh* gene family of *Cuscuta* spp, we searched the *Cuscuta* spp sequences with those from *A. thaliana* using BLAST. We identified large fragments of the *ndhB* genes in the plastome sequences of *C. reflexa* and *C. exaltata* ([Supplementary Figure S6 and S7](#)). Compare to the *ndhB* gene in *A. thaliana*, we found a base G changed to T at position 22 resulted in the appearance of a stop codon in the *ndhB* coding sequence (CDS) in *C. reflexa* ([Supplementary Figure S8](#)). Similarly, a stop codon appeared in the CDS of *ndhB* at position 87–90 in *C. exaltata* due to a large indel of 25 bp ([Supplementary Figure S9](#)).

### Repeat analysis

Microsatellites, also known as simple sequence repeats (SSR), are abundant across genomes and show high polymorphism (Li et al. 2002). Microsatellite repeats, tandem repeats, and interspersed repeats were detected in the plastome. A total of 21 microsatellite repeats (A/T) was detected ([Supplementary Table S5](#)). All the microsatellite repeats are less than 100 bases. And eight microsatellite repeats were located in the CDS regions (*atpF*, *rps7*, *cemA*, *ycf1*, *ycf2*). Using the cutoff 90% similarity between repeat units, we detected 16 repeats in the plastome of *C. gronovii* finally ([Supplementary Table S6](#)). The lengths of repeat units range from 11 bp to 135 bp. For interspersed repeats, four palindromic repeats and six direct repeats were identified ([Supplementary Table S7](#)). The most extended interspersed repeat unit is 36 bp long and located in the intergenic region of *psbI* and *trnS-GCU*. In the future, when more plastome sequences become available, these repeats can be compared, and hopefully, some of them will serve as the polymorphic markers for intraspecific discrimination.

### Phylogenetic analysis

We selected 16 *Cuscuta* species to construct the phylogenetic tree. *I. batatas* and *I. purpurea* were selected as the outgroup taxa. The complete plastome sequences ([Supplementary Figure S10A](#)), all CDS sequences ([Supplementary Figure S10B](#)), all shared gene protein sequences ([Supplementary Figure S10C](#)) and all shared gene nucleic acid sequences ([Figure 3](#)) were used to construct four trees. All the trees have the identical topology. However, the tree constructed with all shared gene nucleic acid sequences had the highest bootstrap values. According to the phylogenetic analysis, these *Cuscuta* species formed two clades. *C. exaltata*, *C. japonica*, and *C. reflexa* belong to one clade, and the others

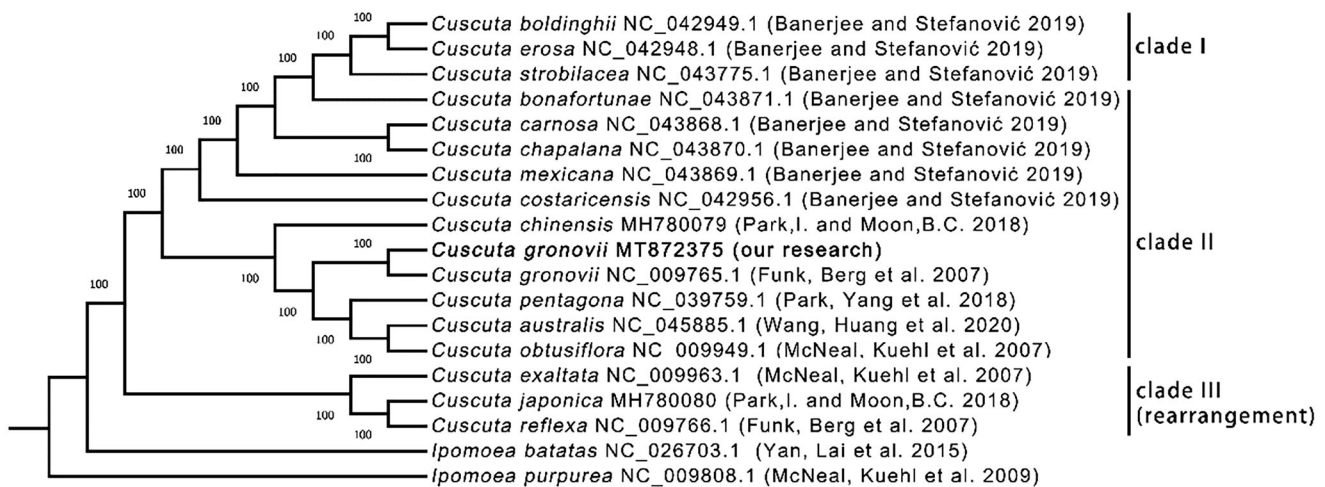


**Figure 2.** Gene deletion in *Cuscuta* genus. Each column represents a *Cuscuta* species. Each row represents a gene. The cells are colored as dark gray, light gray, and white, which represents two, one, and zero-copy of genes.

belong to another clade, consistent with previous results from the gene deletion analysis. As expected, two *C. gronovii* were clustered together. Furthermore, the bootstrap scores at all nodes in the phylogenetic tree were 100, indicating the tree's high reliability.

### **Snp identification from the plastome**

To discover SNPs from the plastome, we compared the two plastomes of *C. gronovii* and identified 2 INDELS and 1 SNP (Supplementary Table S8). Among the INDELS, the first one is



**Figure 3.** The phylogenomic tree of species from genus constructed using the maximum likelihood (ML) method. The number next to each node represents the bootstrap values. The GenBank accession numbers are shown after the Latin names. The sequence obtained from this study was highlighted in Bold and noted with 'our research.' The clade I, II and III are the same as those shown in Figure 2. The clade III containing three species whose plastom sequences contain several rearranged regions.

an 18 bp fragment inserted in the IGS region between the genes *trnS-UGA* and *psbZ*. The second one is located in the IGS region between the genes *rbcL* and *accD*. The SNP located in the CDS of gene *accD* is from C to A, causing the encoded amino acids to change from Asp to Glu. The molecular markers based on the intergenic region between the *trnS-UGA* and the *psbZ* gene might help distinguish *C. gronovii* under the species level.

### Identification of RNA editing sites in *C. gronovii* plastome

To determine the possible presence of RNA editing events in the *C. gronovii* plastid, we downloaded the RNA-seq data from the Public database and compared them with the plastome sequence. The RNA-seq data were obtained from the stems of *C. gronovii*. We identified two putative RNA editing sites using the REDtools software (Supplementary Table S9). The corresponding editing events cause one G to A transition and one U to C transition. The two RNA editing sites are located in the coding regions of gene *rrn4.5S*. Since *rrn4.5S* does not encode any protein, it is unclear if these RNA-editing sites will have any functional consequence. It appears that RNA-editing is not used to correct the miss-coding of proteins in the chloroplast of *C. gronovii*.

### Structural variation of *Cuscuta* plastomes

All the *Cuscuta* plastome sequences in GenBank (Last accessed: 31 December 2020) were compared with that of *A. thaliana* for structural variation analysis. All the Dotplot results showed a high degree of colinearity between the *Cuscuta* plastomes and *A. thaliana* (Supplementary Figure S1 and Figure S11A-P). It should be noted that the starting positions of these *Cuscuta* plastomes were set differently. The SSC region's first position was selected as the starting position for *C. bonafortunae*, *C. carnososa*, *C. boldinghii*, *C. mexicana*, *C. erosa*, *C. strobilacea*, *C. costaricensis* and *C. chapalana*.

In contrast, the LSC regions' first positions were the first position for all other *Cuscuta* species (Supplementary Figure S11A-H).

*C. bonafortunae* has an inverted fragment in its SSC region (Supplementary Figure S11D). The three plastomes of *C. exaltata*, *C. japonica*, and *C. reflexa* are much longer than those of the plastomes because they contain an additional nine genes (Figure 2). *C. australis*, *C. obtusiflora*, *C. pentagona*, *C. gronovii* and *C. chinensis* show the similar structure compared with *A. thaliana* (Supplementary Figure S1 and Figure S11 I-M). In terms of the structure, *C. exaltata* has two inverted fragments with different length (Supplementary Figure S11N). In contrast, *C. japonica* and *C. reflexa* have three inverted fragments (Supplementary Figure S11O-P). The results of the structural variations are consistent with that of the phylogenetic analysis. It showed that *C. exaltata*, *C. japonica*, and *C. reflexa* belonged to one clade. And the rest species belonged to another clade (Figure 3).

### Molecular marker development

*C. chinensis* and *C. australis* are two species listed in the Chinese Pharmacopeia. They are frequently mixed up with other *Cuscuta* species due to similar morphological characters. To distinguish *C. gronovii* and two medicinal *Cuscuta* species: (*C. chinensis* and *C. australis*), we developed molecular markers based on their plastome sequences based on the aligned plastome sequences. In total, ten molecular markers were found (Supplementary Table S10). These molecular markers will be invaluable for the molecular discrimination of these species.

### Discussion

Here, we systematically studied the intraspecific and interspecific variations of *C. gronovii* plastomes. We analyzed the gene content, genome rearrangement, and intraspecific variations at DNA and post-transcription levels. The results

painted a vivid picture of the adaptation of *Cuscuta* species to a parasitic lifestyle.

Foremost, the plastomes of all sixteen species have undergone extensive gene losses. Interestingly, these sixteen species can be divided into two groups. The first group comprises *C. australis*, *C. obtusiflora*, *C. pentagona*, *C. gronovii*, *C. chinensis*, *C. bonafortunae*, *C. camosa*, *C. boldinghii*, *C. mexicana*, *C. erosa*, *C. strobilacea*, *C. costaricensis*, and *C. chapalana*. The second group includes *C. exaltata*, *C. japonica*, and *C. reflexa*. Plastomes in the first group had ten more common genes comparing with those in the second group. Furthermore, species in the first group are closer to the out-group taxa, suggesting that these three species in the second group had a much higher gene deletion rate. However, some genes are only missing in one species. They are likely to result from different parameters used in the annotation step. And further work is needed to confirm whether these genes are actually missing (Figure 2).

Secondly, when the eleven plastomes were compared with that of *A. thaliana*, the first group's plastomes showed 2–3 inverted regions. Examining the nine genes in group one but not in group two found that these genes are spread out in the genome (date is not shown), suggesting that the genes were deleted independently in multiple runs in the evolutionary history.

Very few data were available to explore the intraspecific variations for *Cuscuta* species. Here, we compared the two *C. gronovii* plastome sequences. In total, we found 2 INDELS and 1 SNP. The largest variation comes from an 18 insertion between the *trnS*-UGA and *psbZ* gene, accounted for 90% of the total variation identified in terms of the number of nucleotides involved. The SNP located at the *accD* is from C to A, causing the amino acid to change from Asp to Glu. The functional consequence of this amino acid change remains to be elucidated.

RNA-editing is a post-transcriptional mechanism to modify the mRNA sequence. It will either increase the translated proteins' diversity in mitochondria or serve as an 'error-correction' mechanism to maintain the conserved sequences of the translated proteins in plastids. Previously, Funk et al. found 12 RNA-editing sites in *C. gronovii* (NC\_009765) from the Sanger sequencing result of PCR products amplified from cDNA (Funk et al. 2007). Our scrutiny of the data suggests that only two of them (*rps2*, *rps14*) had strong signals. In this study, mapping RNA-seq data obtained from a public database to the plastome revealed two potential RNA-editing sites. Both sites locate in the rRNA coding region. This observation raises several new questions. Are these RNA-editing sites functional relevant? While future work needs to clarify the discrepancy of the RNA editing sites, it is consistent that the RNA editing level is relatively low.

The discrimination of closely related species has been vital for their correct classification and the plant materials' downstream application. Our original plan is to sequence the medicinal species *C. australis*. However, sequence comparison with those in the database suggests that our sample is most likely from *C. gronovii*. *C. gronovii* is a parasitic plant and widely spread in Europe and North America. Our results indicate that it might have been introduced into China by route

unidentified. Regardless of what has happened, it suggests that there might be a mixed usage of different species in the production of health products involving *Cuscuta* materials. To ensure the efficacy and safety of these products, accurate identification of the source material is critical. To serve this purpose, we identified ten regions that can be used to design primers. The corresponding products can be used to discriminate *C. gronovii* and two medicinal species, *C. chinensis* and *C. australis*, which are listed in Chinese Pharmacopeia.

In conclusion, the identification and characterization of the plastome sequences of *C. gronovii* provide a valuable resource to analyze the intraspecific diversity among different lines of *C. gronovii* and to study the genomic evolution of *Cuscuta* species. Overall, low levels of intraspecific diversity were observed. This is possible in the context of the parasitic lifestyle of *C. gronovii*. As a parasitic plant, *C. gronovii* depends on the host cells' interaction to obtain the nutrients. As a result, having a lower mutation rate will help maintain stable interactions between the plant and its host.

## Author contributions

CL conceived the study; HMC collected samples of *C. gronovii*, MJ extracted DNA for next-generation sequencing, NY assembled and validated the genome; NY and JM performed data analysis and drafted the manuscript; PHC and LFH reviewed the manuscript critically. All authors have read and agreed on the contents of the manuscript.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

The National Mega-Project supported this work for Innovative Drugs of China [2019ZX09735-002], and National Science & Technology Fundamental Resources Investigation Program of China [2018FY100705], National Natural Science Foundation of China [81872966]. The funders were not involved in the study design, data collection and analysis, decision to publish, or manuscript preparation.

## ORCID

Yang Ni  <http://orcid.org/0000-0002-3472-386X>  
Chang Liu  <http://orcid.org/0000-0003-3879-7302>

## Data availability statement

The genome sequence data that support the findings of this study are openly available in GenBank of NCBI at (<https://www.ncbi.nlm.nih.gov/>) under the accession number MT872375. The associated BioProject, SRA, and Bio-Sample numbers are PRJNA661415, SAMN16057401, and SRR12597239, respectively. The sample has been deposited in the Herbarium of the Institute of Medicinal Plant Development in Beijing, China, with the accession number: implad20170662.

## References

- Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27(2):573–580.
- Berg S, Krupinska K, Krause K. 2003. Plastids of three *Cuscuta* species differing in plastid coding capacity have a common parasite-specific RNA composition. *Planta.* 218(1):135–142.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 30(15):2114–2120.
- Burland TG. 2000. DNASTAR's Lasergene sequence analysis software. In: *Bioinformatics methods and protocols.* Clifton: Springer; p. 71–91.
- Cooke D, Black I. 1987. Biology and control of *Cuscuta campestris* and other *Cuscuta* spp.: a bibliographic review. Technical Paper, Department of Agriculture, South Australia.
- Dawson JH, Musselman LJ, Wolswinkel P, Dörr I. 1994. Biology and control of *Cuscuta*. *Rev Weed Sci.* 6:265–317.
- Dierckxsens N, Mardulyn P, Smits G. 2016. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* 45(4):e18.
- Funk HT, Berg S, Krupinska K, Maier UG, Krause K. 2007. Complete DNA sequences of the plastid genomes of two parasitic flowering plant species, *Cuscuta reflexa* and *Cuscuta gronovii*. *BMC Plant Biol.* 7(1):45.
- Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. *Genome Res.* 8(3):195–202.
- Katoh K, Kuma K-i, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 33(2): 511–518.
- Keeling PJ. 2004. Diversity and evolutionary history of plastids and their hosts. *Am J Bot.* 91(10):1481–1493.
- Krumsiek J, Arnold R, Rattei T. 2007. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics.* 23(8): 1026–1028.
- Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. 2001. REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res.* 29(22):4633–4642.
- Li YC, Korol AB, Fahima T, Beiles A, Nevo E. 2002. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol Ecol.* 11(12):2453–2465.
- Madden T. 2013. The BLAST sequence analysis tool. In: *The NCBI handbook [Internet].* 2nd ed. Bethesda: National Center for Biotechnology Information (US).
- Misra S, Harris N. 2005. Using Apollo to browse and edit genome annotations. *Curr Protoc Bioinformatics.* 12(1):9.5. 1–9.5. 28.
- Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. 2015. evolution: IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32(1):268–274.
- Picardi E, D'Erchia AM, Montalvo A, Pesole G. 2015. Using REDtools to detect RNA editing events in NGS Datasets. *Curr Protoc Bioinformatics.* 49:12.12.1–12.12.15.
- Sebastian B, Thomas T, Thomas M, Uwe S, Martin M. 2017. MISA-web: a web server for microsatellite prediction. *Bioinformatics.* 33(16):2583.
- Sun G, Xu Y, Liu H, Sun T, Zhang J, Hettenhausen C, Shen G, Qi J, Qin Y, Li J, et al. 2018. Large-scale gene losses underlie the genome evolution of parasitic plant *Cuscuta australis*. *Nat Commun.* 9(1):2683.