ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/348478615

Comparative analysis and implications of the chloroplast genomes of three thistles (Carduus L., Asteraceae)

Article *in* PeerJ · January 2021

DOI: 10.7717/peerj.10687

CITATIONS
0

READS
27

5 authors, including:

Hoang Dang Khoa Do
Gachon University
25 PUBLICATIONS **83** CITATIONS

SEE PROFILE

Changkyun Kim
Gachon University
67 PUBLICATIONS **517** CITATIONS

SEE PROFILE

Joo-Hwan Kim
Gachon University
199 PUBLICATIONS **1,848** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Tracheophyte Phylogeny Poster (TPP) View project

Project Genome research project of endemic and endangered plants of Korea, Korea National Arboretum (2015-) View project

All content following this page was uploaded by Hoang Dang Khoa Do on 16 January 2021.

The user has requested enhancement of the downloaded file.

Comparative analysis and implications of the chloroplast genomes of three thistles (*Carduus* L., Asteraceae)

Joonhyung Jung[1,*], Hoang Dang Khoa Do[1,2,*], JongYoung Hyun[1], Changkyun Kim[1] and Joo-Hwan Kim[1]

[1] Department of Life Science, Gachon University, Seongnam, Gyeonggi, Korea
[2] Nguyen Tat Thanh Hi-Tech Institute, Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam
[*] These authors contributed equally to this work.

ABSTRACT

Background. *Carduus*, commonly known as plumeless thistles, is a genus in the Asteraceae family that exhibits both medicinal value and invasive tendencies. However, the genomic data of *Carduus* (i.e., complete chloroplast genomes) have not been sequenced.

Methods. We sequenced and assembled the chloroplast genome (cpDNA) sequences of three *Carduus* species using the Illumina Miseq sequencing system and Geneious Prime. Phylogenetic relationships between *Carduus* and related taxa were reconstructed using Maximum Likelihood and Bayesian Inference analyses. In addition, we used a single nucleotide polymorphism (SNP) in the protein coding region of the *matK* gene to develop molecular markers to distinguish *C. crispus* from *C. acanthoides* and *C. tenuiflorus*.

Results. The cpDNA sequences of *C. crispus, C. acanthoides*, and *C. tenuiflorus* ranged from 152,342 bp to 152,617 bp in length. Comparative genomic analysis revealed high conservation in terms of gene content (including 80 protein-coding, 30 tRNA, and four rRNA genes) and gene order within the three focal species and members of subfamily Carduoideae. Despite their high similarity, the three species differed with respect to the number and content of repeats in the chloroplast genome. Additionally, eight hotspot regions, including *psbI-trnS_GCU*, *trnE_UUC-rpoB*, *trnR_UCU-trnG_UCC*, *psbC-trnS_UGA*, *trnT_UGU-trnL_UAA*, *psbT-psbN*, *petD-rpoA*, and *rpl16-rps3*, were identified in the study species. Phylogenetic analyses inferred from 78 protein-coding and non-coding regions indicated that *Carduus* is polyphyletic, suggesting the need for additional studies to reconstruct relationships between thistles and related taxa. Based on a SNP in *matK*, we successfully developed a molecular marker and protocol for distinguishing *C. crispus* from the other two focal species. Our study provides preliminary chloroplast genome data for further studies on plastid genome evolution, phylogeny, and development of species-level markers in *Carduus*.

Subjects Genomics, Plant Science

Keywords *Carduus crispus*, Chloroplast genome, Invasive species, Molecular markers, Plumeless thistles

Submitted 28 February 2020
Accepted 11 December 2020
Published 14 January 2021

Corresponding author
Joo-Hwan Kim,
kimjh2009@gachon.ac.kr

Academic editor
Jonathan Thomas

Additional Information and Declarations can be found on page 14

DOI **10.7717/peerj.10687**

Copyright
2021 Jung et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

How to cite this article Jung J, Do HDK, Hyun JY, Kim C, Kim J-H. 2021. Comparative analysis and implications of the chloroplast genomes of three thistles (*Carduus* L., Asteraceae). *PeerJ* **9**:e10687 http://doi.org/10.7717/peerj.10687

INTRODUCTION

Carduus L. (subfamily Carduoideae; Asteraceae), commonly known as plumeless thistles, comprises 90 species native to Eurasia and Africa (*Angiosperm Phylogeny Group, 2016*). Several *Carduus* species are invasive, noxious weeds on other continents (*Doing, Biddiscombe & Knedlhans, 1969*). Four species, including *C. acanthoides* L. (spiny plumeless thistle), *C. tenuiflorus* Curtis (sheep thistle), *C. pycnocephalus* L. (Italian thistle), and *C. crispus* Guirão ex Nyman (welted thistle), all of which originate in Eurasia and Africa, are considered invasive in North America (*Dunn, 1976*; *Verloove, 2014*). *Carduus crispus*, also called curly plumeless thistle, is also considered an invasive species in Korea (*Jung et al., 2017*). This species differs from other *Carduus* in having soft, sparsely arachnoid-hairy leaves with short marginal bristles, and apically recurved involucral bracts (*Todorov et al., 2018*). Among *Carduus* species, *C. crispus* contains chemicals with the potential to treat various diseases (*Xie, Li & Jia, 2005*; *Davaakhuu, Sukhdolgor & Gereltu, 2010*; *Lee et al., 2011*; *Tunsag, Davaakhuu & Batsuren, 2011*). Specifically, certain compounds extracted from *C. crispus* have potential value in the treatment of obesity and cancer (*Davaakhuu, Sukhdolgor & Gereltu, 2010*; *Lee et al., 2011*). While *Carduus* has been studied from various perspectives (i.e., invasion, phylogeny, and medicinal effects), its chloroplast genome has not been sequenced. It is therefore worthwhile to study the genome of *Carduus*, and particularly that of *C. crispus*, which has potential medicinal benefits.

In most angiosperms, the chloroplast genome (cpDNA) contains genes essential to photosynthesis (*Sugiura, 1992*). Genomic events (i.e., gene deletion, inversion, or duplication) in cpDNA may provide information about species' evolutionary history (*Cosner, Raubeson & Jansen, 2004*; *Do & Kim, 2017*; *Haberle et al., 2008*). For example, the Fabaceae includes clades that are characterised by large inversions and the loss of inverted repeat regions (*Choi & Choi, 2017*). Inversions have also been recorded in the cpDNA of Asteraceae (*Kim, Choi & Jansen, 2005*). Specifically, a large inversion comprising a 22.8 kb sequence occurred simultaneously with a small inversion of a 3.3 kb fragment; this event coincided with the split between major clades (excluding Barnadesioideae) in the evolution of Asteraceae. cpDNA data can also be used to develop molecular markers based on nucleotide polymorphisms (i.e., single nucleotide polymorphism (SNP) markers and microsatellite markers). Molecular authentication has been reported for various plant species, with a focus on invasive plants, endangered species, and taxa with potential medicinal value (*Kim et al., 2012*; *Ishikawa, Sakaguchi & Ito, 2016*; *Luo et al., 2016*; *Park et al., 2016*; *Marochio et al., 2017*; *Han et al., 2018*; *Do et al., 2019*). Specific regions of cpDNA have been identified for developing molecular markers in plants, including the commonly-used *matK* region (*Poovitha et al., 2016*; *Vu et al., 2017*). Among the Asteraceae, studies on molecular markers have been conducted for rubber dandelion (*Taraxacum kok-saghyz* LE Rodin), horseweed (*Conyza* sp.), Indian Chrysanthemum (*Chrysanthemum indicum* L.), the endemic herb *Aster savatieri* Makino, and the invasive plant *Tithonia diversifolia* (Hemsl.) A Gray (*Ishikawa, Sakaguchi & Ito, 2016*; *Luo et al., 2016*; *Zhang et al., 2017*; *Marochio et al., 2017*; *Han et al., 2018*). In addition to the development of these molecular markers, complete cpDNA sequences have been reported for various Asteraceae species

(*Kim, Choi & Jansen, 2005*; *Choi & Park, 2015*; *Wang et al., 2015a*; *Wang et al., 2015b*; *Yun, Gil & Kim, 2017*; *Liu et al., 2018*; *Ma, Sun & Zhao, 2018*; *Su et al., 2018*). cpDNA sequences may be used to elucidate the phylogeny of angiosperms from the clade that is basal to monocots and eudicots (*Angiosperm Phylogeny Group, 2016*). Previous investigations into phylogenetic relationships among members of the Asteraceae have been conducted using a range of molecular data types, including *rbcL*, *ndhF*, *matK*, chloroplast DNA restriction sites, ITS sequence data, and nuclear loci (*Jansen, Michaels & Palmer, 1991*; *Häffner & Hellwig, 1999*; *Fu et al., 2016*; *Mandel et al., 2019*). However, the paucity of available sequence data may have resulted in ambiguous relationships between *Carduus* and related taxa (*Häffner & Hellwig, 1999*; *Fu et al., 2016*). In particular, ITS sequence data suggest that *C. leptacanthus* is sister to *Cirsium* and *Notobasis*, whereas another *Carduus* species is sister to *Cirsium* and *Tyrimmus* (*Häffner & Hellwig, 1999*). As such, clarification of relationships between *Carduus* and related species will require studies that include a larger number of *Carduus* species and different data types (i.e., chloroplast and mitochondrial genomes).

We used next-generation sequencing (NGS) to sequence and characterise the chloroplast genomes of *Carduus crispus*, *C. acanthoides*, and *C. tenuiflorus*, which exhibit both invasive tendencies and potential medical utility (particularly *C. crispus*). We then conducted comparative genomic analyses to explore genomic diversity among the three species with respect to highly variable regions, and the types and numbers of repeats. In addition, we reconstructed the formerly ambiguous relationship between *Carduus* and related taxa based on 78 protein-coding regions and non-coding sequences. Finally, we developed a specific molecular marker for *C. crispus* based on a SNP in the *matK* gene. This molecular marker provides useful information for managing *C. crispus* invasions, particularly with respect to the identification of immature (vegetative) individuals, which tend to be morphologically similar to other *Carduus* species (i.e., having winged stems with apical spines and spiny leaves). This molecular marker may also support positive identification of *C. crispus* for medical usage.

MATERIALS & METHODS

Taxon sampling, total DNA extraction, chloroplast genome assembly, and comparative analysis

Leaves of *C. crispus*, *C. acanthoides*, and *C. tenuiflorus* were collected and dried in silica gel powder for NGS analysis. In addition, to test the efficiency of molecular markers (Table 1), leaves of 22 individuals of the three species were sampled from herbaria at the Korea National Arboretum (KNA), National Institute of Biological Resources (NIBR), New York Botanical Garden (NYBG), Carnegie Museum Herbarium (CM), and the United States National Herbarium (US). A modified cetyl trimethylammonium bromide (CTAB) protocol was used to extract total DNA from collected samples (*Doyle & Doyle, 1987*). High-quality DNA was used to conduct NGS using the Miseq platform and a Miseq Reagent Kit v3 (Illumina, Seoul, South Korea). Raw NGS data (2×300 bp paired-end reads) were cleaned by cutting adapter sequences, removing duplicate and chimeric reads, and trimming ends with > 0.05 probability of error per base. Cleaning was conducted using

Table 1 List of *Carduus* species for NGS and developing molecular marker.

No.	Species	Voucher	Location
1	*Carduus crispus* Guirão ex Nyman	Korea National Arboretum (LK0908)	Mt. Seokbyeong, Imgye-myeon, Jeongseon-gun, Gangwon-do, Republic of Korea
2	*Carduus crispus* Guirão ex Nyman*	Korea National Arboretum (LK0943)	Mt. cheongog, Hajang-myeon, Samcheok-si, Gangwon-do, Republic of Korea
3	*Carduus crispus* Guirão ex Nyman	Korea National Arboretum (LK1497)	Mt. Nochu, Yeoryang-myeon, Jeongseon-gun, Gangwon-do, Republic of Korea
4	*Carduus crispus* Guirão ex Nyman	Korea National Arboretum (CNUFR0470)	295, Sinseong-ri, Bukha-myeon, Jangseong-gun, Jeollanam-do, Republic of Korea
5	*Carduus crispus* Guirão ex Nyman	Korea National Arboretum (LK0430)	Mt. cheongog, Imgye-myeon, Jeongseon-gun, Gangwon-do, Republic of Korea
6	*Carduus crispus* Guirão ex Nyman	National Institute of Biological Resources (NIBRVP0000619716)	Mt. Jaam, Namhu-myeon, Andong-si, Gyeongsangbuk-do, Republic of Korea
7	*Carduus crispus* Guirão ex Nyman	National Institute of Biological Resources (NIBRVP0000601328)	Sanghwa-ri, Danchon-myeon, Uiseong-gun, Gyeongsangbuk-do, Republic of Korea
8	*Carduus crispus* Guirão ex Nyman	National Institute of Biological Resources (NIBRVP0000524207)	Mt. Beophwa, Yugu-eup, Gongju-si, Chungcheongnam-do, Republic of Korea
9	*Carduus crispus* Guirão ex Nyman	National Institute of Biological Resources (NIBRVP0000613325)	Mt. Mani, Hwado-myeon, Ganghwa-gun, In-cheon, Republic of Korea
10	*Carduus crispus* Guirão ex Nyman	National Institute of Biological Resources (NIBRVP0000580633)	Mt. Jeonggwang, Mohyeon-eup, Cheoin-gu, Yongin-si, Gyeonggi-do, Republic of Korea
11	*Carduus crispus* Guirão ex Nyman	National Institute of Biological Resources (NIBRVP0000578538)	Mt. Jangam, Pyeongchang-eup, Pyeongchang-gun, Gangwon-do, Republic of Korea
12	*Carduus crispus* Guirão ex Nyman	National Institute of Biological Resources (NIBRVP0000580501)	Mt. Gwangdeok, Dongnam-gu, Cheonan-si, Chungcheongnam-do, Republic of Korea
13	*Carduus acanthoides* L.*	Kim 2018-001, Nevada city, Ca, USA	USA, California, Nevada city
14	*Carduus acanthoides* L.	NewYork Botanical Garden (00532263)	USA, Colorado, Pitkin Co., South side of State Highway 82, 5 miles W of Aspen, Airport., 2256 - 2256m
15	*Carduus acanthoides* L.	NewYork Botanical Garden (00532244)	USA, Wisconsin, Richland Co., 3 miles SE of Richland Center., 43.300443 -90.321019
16	*Carduus acanthoides* L.	NewYork Botanical Garden (00532262)	USA, Wyoming, Platte Co., 1402 - 1402m
17	*Carduus acanthoides* L.	Carnegie Museum Herbarium (267884)	Romania, Oltenia, distr. Dolj, inter vicos Lascar Catargiu et Popoveni ad Canalul colector, 75m
18	*Carduus acanthoides* L.	Carnegie Museum Herbarium (528144)	USA, Pennsylvania, Mifflin, Maple & Walnut Sts, Belleville
19	*Carduus acanthoides* L.	United States National Herbarium (1944312)	Czech Republic, Bohemia centralis: Paraha-Troja. In ruderatis
20	*Carduus acanthoides* L.	United States National Herbarium (3419733)	Ukraine, Prov. Czerkassy, prope opp. Umanj, in ruderatis
21	*Carduus tenuiflorus* Curtis*	Kim 2018-002, Nevada city, Ca, USA	USA, California, Nevada city
22	*Carduus tenuiflorus* Curtis	NewYork Botanical Garden (00366662)	Mexico, El Cercado. Santiago, N. L., 495–495 m
23	*Carduus tenuiflorus* Curtis	Carnegie Museum Herbarium (519240)	USA, California, Humboldt, Angels Ranch, to-ward Hungry Hollow, Bald Mountain
24	*Carduus tenuiflorus* Curtis	Carnegie Museum Herbarium (282243)	USA, California, Alameda, Berkeley

Notes.

Asterisks indicate samples for next generation sequencing (NGS) analysis.

Geneious Prime (*Kearse et al., 2012*). Raw data were submitted to NCBI (accession number PRJNA645567). Using Geneious Prime, filtered NGS data for each species were mapped to the reference chloroplast genome sequences of *Arctium lappa* (NCBI accession number MH375874), *Saussurea polylepis* (MF695711), *Carthamus tinctorius* (KP404628), *Centaurea diffusa* (KJ690264), *Silybum marianum* (KT267161), *Cirsium arvense* (KY562583), *Cynara humilis* (KP299292), and *Atractylodes chinensis* (MG874805) to isolate cpDNA reads of which the similarity to reference was over 95%. Isolated cpDNA reads were then assembled using the de novo function in Geneious to create various contigs of chloroplast genome sequences. The newly created contigs were de novo re-assembled to construct complete cpDNA sequences for each focal species. We confirmed the results from Geneious Prime by reconstructing the complete cpDNA of *Carduus* using NOVOplasty and following Dierckxsens et al. (2017). Using Geneious Prime, gene content and order of sequenced cpDNA were annotated based on existing complete cpDNA sequences of other Asteraceae taxa. Annotations that had over 95% similarity in comparison with references were retained, and the start and stop codons in the protein coding regions were verified (Data S1). tRNA sequences were assessed using tRNA Scan-SE (*Chan & Lowe, 2019*). Complete cpDNA sequences of the three species were submitted to GenBank; accession numbers were MK652229 for *C. crispus*, MK652228 for *C. acanthoides*, and MK652230 for *C. tenuiflorus.* The cpDNA map was visualised using OGDraw (*Greiner, Lehwark & Bock, 2019*). Complete chloroplast genomes of *Carduus* species were aligned with other Asteraceae and related species (*Nicotiana tabacum* (NC_001879) was used as an outgroup; Table S1), and gene loss and rearrangement were identified using MAUVE (*Darling, Mau & Perna, 2010*). In addition, Geneious Prime was used to calculate the pairwise identities of cpDNA sequences in the focal species. Small single repeats (SSRs) were analysed using Phobos embedded in Geneious Prime (*Christoph, 2006–2010*). Minimum repeat numbers were 10, 5, 4, and 3 for mono-, di-, tri-, and tetra-nucleotides, respectively. REPuter (*Kurtz et al., 2001*) was used to analyse the large repeat sequences (minimum length = 20 bp) in each species. To explore nucleotide diversity, 131 coding and non-coding regions were extracted from the complete cpDNA (Table S2). Following alignment using MUSCLE embedded in Geneious Prime (*Edgar, 2004*), aligned sequences were imported into DnaSP 6 (*Rozas et al., 2017*) for calculation of Pi values.

Phylogenetic analysis of *Carduus* and related taxa

A total of 78 protein-coding regions were extracted from the complete cpDNA of the focal species and other related taxa (Table S1). Sequences were aligned using MUSCLE embedded in Geneious Prime (*Edgar, 2004*). We used jModeltest (*Posada, 2008*) to find the best model for the aligned DNA sequences; GTR + I + R was selected as the most suitable model and was used in Maximum Likelihood (ML) and Bayesian Inference (BI) analyses. The ML analysis was conducted with the IQ-tree web server (http://iqtree.cibiv.univie.ac.at), using 1,000 bootstrap replications to calculate branch support values (*Trifinopoulos et al., 2016*). We used MrBayes v3.2 (*Ronquist et al., 2012*) for BI analyses. The Markov chain Monte Carlo (MCMC) analysis was run for 1,000,000 generations, and a tree was assembled every 1000 generations. A 25% burn-in setting was used for summarising trees. Figtree v4.0

(http://tree.bio.ed.ac.uk/software/figtree/) was used to visualise phylogenetic trees. Other datasets, including whole chloroplast genomes (excluding one IR region), non-coding regions of cpDNA, and hotspot regions derived from the cpDNA of *Carduus* species, were used in phylogenetic analysis in addition to protein-coding regions (Table S1). Analytical procedures for these additional datasets were identical to those used for the protein coding regions; however, we used the TVM+I+G model for the whole chloroplast genome (excluding one inverted repeat [IR] region) and all non-coding regions, and the TVM+G model for the hotspot regions dataset.

SNP identification, primer design, and multiplex PCR

The complete *matK* gene, extracted from the cpDNA of the three focal species, was aligned using MUSCLE to identify SNPs (*Edgar, 2004*; Fig. S1). The selected SNP for *C. crispus* was then confirmed by aligning the available *matK* sequences of other *Carduus* species on NCBI to those of the focal species (Fig. S2). Based on SNP data, primer pairs were designed using Primer3 to distinguish *C. crispus* from other *Carduus* species (*Untergasser et al., 2012*). Primer sequences included matK_463F (5′-CATCTGGAAATCTTGGTTCAG-3′), matK_1162R (5′-GATGCCCCAATGCGTTACAA-3′), CD_SNP_F1 (5′-AATTCTTGCTTCAAAAGG GTCC- 3′), CD_SNP_R1 (5′-TTCCATTTATTCATCAA AAGATAC-3′), CD_SNP_F2 (5′-AATTCTTGCTTCAAAAGGGTCG-3′), and CD_SNP_R2 (5′-TTCCATTTATTCATCAAAAGATAG- 3′). The multiplex PCR of matK_463F, matK_1162R, CD_SNP_F1, and CD_SNP_R1 was designed to yield the 323 bp band for *C. crispus*, the 421 bp band for other *Carduus*, and the 700 bp band for all examined samples (Fig. S3). By contrast, the combination of matK_463F, matK_1162R, CD_SNP_F2, and CD_SNP_R2 yielded a 421 bp PCR product for *C. crispus* and a 323 bp band for other *Carduus* species (Fig. S3). Reactions were conducted in 25 μl solution consisting of 50 ng of template DNA, 2.5 μl of 10× reaction buffer, 0.5 U of E-taq DNA polymerase, 50 mM $MgCl_2$, and 5 mM dNTPs. Concentrations of outer primer pairs (matK_463F and matK_1162R) and inner primer pairs (CD_SNP_F1 and CD_SNP_R1, CD_SNP_F2 and CD_SNP_R2) were 0.75 pM and 0.5 pM, respectively. The PCR procedure consisted of 1 min at 94 °C, followed by denaturing for 1 min at 94 °C, annealing for 40 s at 55 °C, an extension stage of 50 s at 72 °C, and an additional extension of 7 min at 72 °C.

RESULTS

Comparative chloroplast genome analysis of the focal species

Differing numbers of reads were obtained from NGS data, resulting in varying cpDNA coverage rates among the three focal species (Table 2). Total cpDNA length differed among species, ranging from 152,342 bp to 152,617 bp, and included a large single copy (LSC), a small single copy (SSC), and two IR regions (Fig. 1). By contrast, all three species had identical numbers of protein coding (80), tRNA (30), and rRNA (4) genes (Table 2, Table S3). The IR-LSC and IR-SSC junctions were located in the *rps19* and *ycf1* coding regions, respectively, but were longer in *C. acanthoides* (*rps19* = 61 bp and *ycf1* = 568 bp) than in the other two species (*rps19* = 60 bp and *ycf1* = 565 bp). In addition,

Table 2 Features of chloroplast genomes, assembly information, and pairwise identity among three *Carduus* species and related taxa.

Species	*Carduus crispus* (MK652229)	*C. tenuiflorus* (MK652230)	*C. acanthoides* (MK652228)	*Cynara humilis* (KP299292)	*C. baetica* (KP842706)	*C. cornigera* (KP842707)	*C. cardunculus var. scolymus* (KP842708)	*C. cardunculus var sylvestris* (KP842721)	*Cirsium arvense* (KY562583)	*Helianthus annus* (NC007977)
Total reads	21,118,624	4,621,758	4,300,959	–	–	–	–	–	–	–
Assemble read	805,076 (3.8%)	189,215 (4.1%)	180,245 (4.2%)	–	–	–	–	–	–	–
Coverage	1,585	372	354	–	–	–	–	–	–	–
Number of contigs	126	119	14	–	–	–	–	–	–	–
N50 value (bp)	95,199	50,873	128,044	–	–	–	–	–	–	–
Total length	152,342	152,426	152,617	152,585	152,548	152,550	152,529	152,528	152,855	151,104
LSC	83,254	83,360	83,532	83,622	83,599	83,580	83,578	83,577	83,859	83,530
SSC	18,706	18,674	18,693	18,651	18,639	18,660	18,641	18,641	18,633	18,308
IR	25,191	25,196	25,196	25,156	25,155	25,155	25,155	25,155	25,182	24,633
Protein-coding genes	80	80	80	80	80	80	80	80	80	80
tRNA	30	30	30	30	30	30	30	30	30	30
rRNA	4	4	4	4	4	4	4	4	4	4
LSC-IR junction	*rps19* (60 bp)	*rps19* (60 bp)	*rps19* (61 bp)	*rps19* (60 bp)	*rps19* (60 bp)	*rps19* (60 bp)	*rps19* (60 bp)	*rps19* (60 bp)	*rps19* (60 bp)	*rps19* (101 bp)
SSC-IR junction	*ycf1* (565 bp)	*ycf1* (565 bp)	*ycf1* (568 bp)	*ycf1* (567 bp) (*ycf1-ndhF* overlap 17 bp)	*ycf1* (567 bp) (*ycf1-ndhF* overlap 17 bp)	*ycf1* (567 bp) (*ycf1-ndhF* overlap 17 bp)	*ycf1* (567 bp) (*ycf1-ndhF* overlap 17 bp)	*ycf1* (567 bp) (*ycf1-ndhF* overlap 17 bp)	*ycf1* (565 bp)	*ycf1* (576 bp)
Pairwise identity (%)										
C. crispus	100			98.5	98.5	98.5	98.5	98.5	98.8	92.9
C. tenuiflorus	99.6	100		98.5	98.5	98.5	98.5	98.5	98.9	92.3
C. acanthoides	99.2	99.3	100	98.7	98.6	98.7	98.7	98.7	98.7	92.3

Notes.

The dashes (–) mean no data in this study.

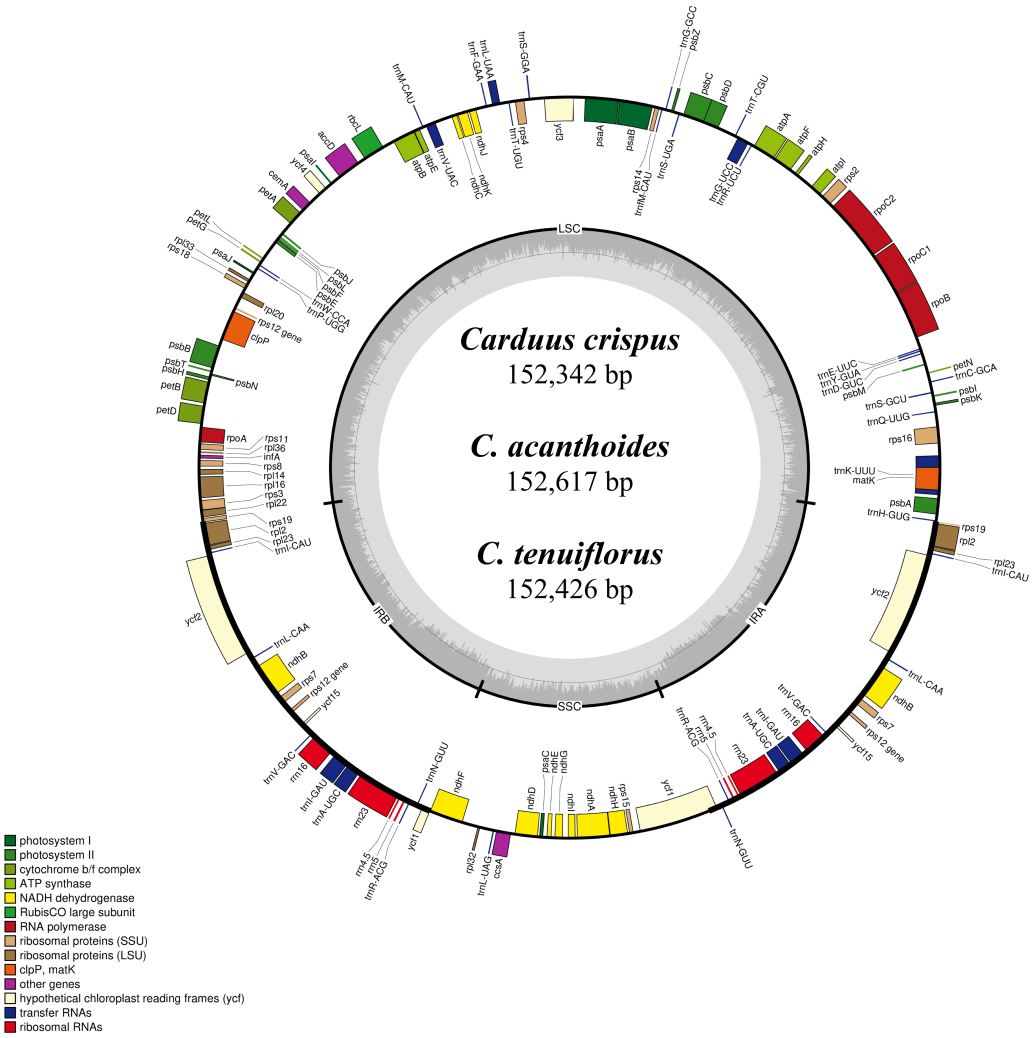

Figure 1 **Map of the chloroplast genomes of *Carduus*.** Genes inside the circle are transcribed clockwise whereas genes outside the circle are transcribed counterclockwise. LSC, large single copy; SSC, small single copy; IRA–IRB, inverted repeat regions.

Full-size DOI: 10.7717/peerj.10687/fig-1

pairwise identity indicated that *C. crispus* is more similar to *C. tenuiflorus* (99.6%) than *C. acanthoides* (99.2%). Observations of nucleotide diversity indicated that 119 of 131 surveyed regions differ among the three focal species (Table S2, Fig. 2). Compared to coding regions, non-coding sequences had higher Pi values (Fig. 2). The highest Pi values were found in the *psbC-trnS* (0.0171) and *psbH-petB* (0.0161) regions. The highest value in coding regions was 0.00696 for *ycf1* (Fig. 2). High nucleotide diversity regions (Pi values >0.008) included *psbI-trnS_GCU*, *trnE_UUC-rpoB*, *trnR_UCU-trnG_UCC*, *psbC-trnS_UGA*, *trnT_UGU-trnL_UAA*, *psbT-psbN*, *petD-rpoA*, and *rpl16-rps3*.

Features of cpDNA repeats

Analysis of SSRs yielded 43 SSRs in *C. crispus*, 40 in *C. tenuiflorus*, and 31 in *C. acanthoides* (Table S4). SSRs occupied the same position in all three species and were mostly located

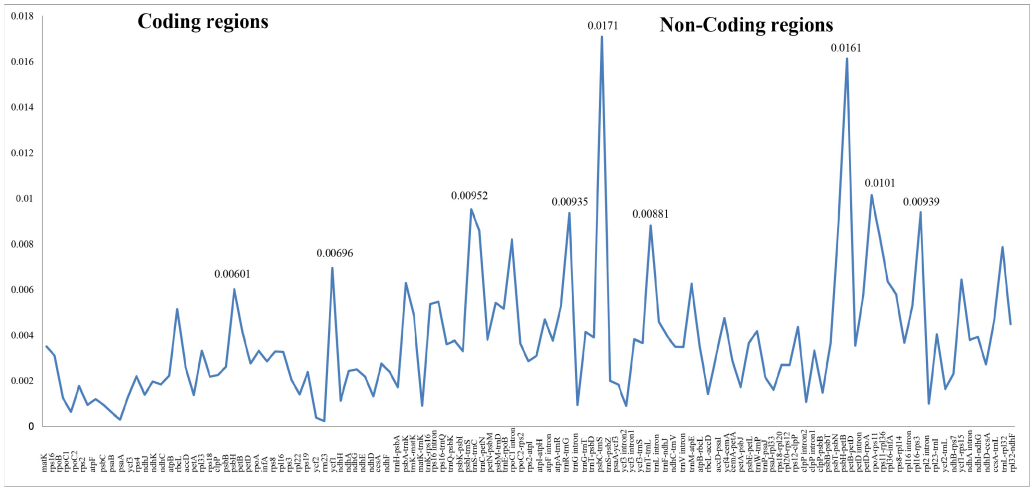

Figure 2 **Nucleotide diversity (Pi values) among the three *Carduus* species.**
Full-size DOI: 10.7717/peerj.10687/fig-2

in non-coding regions. Although four types of SSR (i.e., mono-, di-, tri-, and tetra-nucleotides) were identified, most SSRs were mononucleotides composed of A and T nucleotides (Table S4). All 31 SSRs found in *C. acanthoides* were also present in the other two species (Table S4). By contrast, *C. crispus* had three unique SSRs and shared nine SSRs with *C. tenuiflorus*. There were no unique SSRs in *C. acanthoides* or specific shared SSRs between *C. acanthoides* and either *C. crispus* or *C. tenuiflorus*.

Among the focal species, 16 repeats were identified for both *C. crispus* and *C. tenuiflorus*, compared to 15 for *C. acanthoides* (Fig. 3A, Table S5). There were more repeats in coding regions than in non-coding areas, with the exception of *C. acanthoides*. Three types of repeats (i.e., forward, reverse, and palindrome) were identified in *C. tenuiflorus* and *C. acanthoides*; by contrast, only forward and palindrome repeats were found in *C. crispus*. Forward repeats were more abundant than reverse and palindrome repeats (Fig. 3B). Among recorded repeats, nine were common among all three species (Fig. 3C). *Carduus acanthoides* had four unique repeats, whereas *C. tenuiflorus* and *C. crispus* each had a single unique repeat. *Carduus acanthoides* shared one specific repeat with both *C. crispus* and *C. tenuiflorus*, whereas *C. crispus* and *C. tenuiflorus* shared five specific repeat regions (Fig. 3C).

Phylogenetic relationships between *Carduus* and related taxa

The ML and BI analyses, based on 78 protein-coding genes from *Carduus* and related taxa, yielded identical topologies (Fig. 4). In particular, both analyses confirmed the monophyly of Asteraceae subfamilies (i.e., Carduoideae, Chichorioideae, and Asteroideae). In contrast to the high support for Carduoideae and Cichorioideae (Bootstrap = 100/Posterior Probability = 1), low support was found for Asteroideae clades (Fig. 4). Notably, monophyly of the three *Carduus* species was not supported by either analysis. For example, *C. acanthoides* was sister to *Silybum marianum*, whereas *C. crispus* and *C. tenuiflorus* were

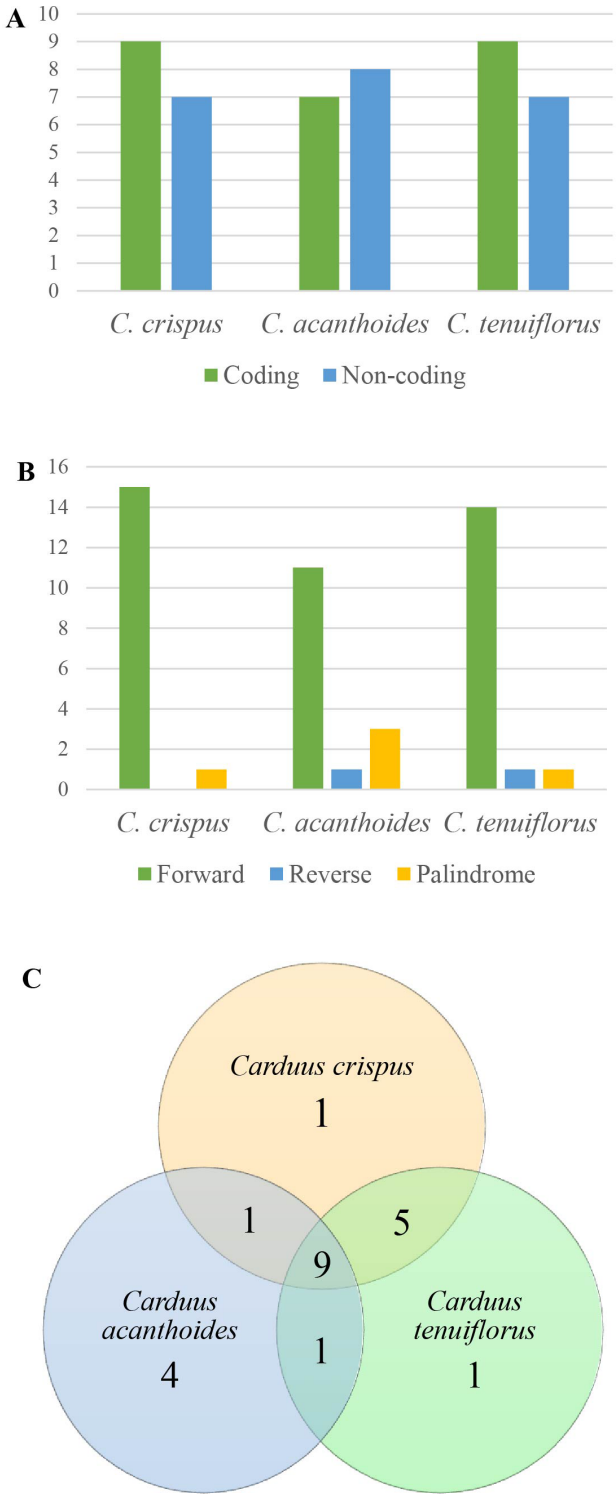

Figure 3 **Repeats in the three *Carduus* species.** (A) Distribution of repeats in the coding and non-coding regions, (B) composition of three types of repeat, and (C) unique and shared repeats in the three *Carduus* species.

Full-size DOI: 10.7717/peerj.10687/fig-3

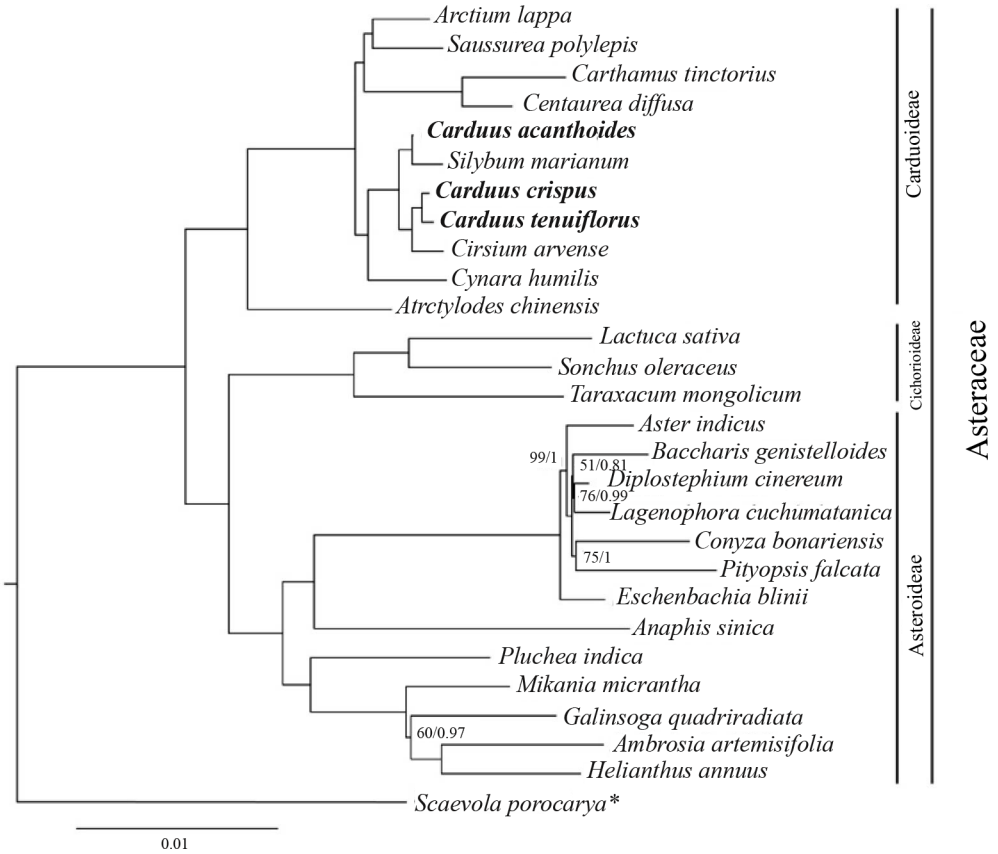

Figure 4 **BI tree of *Carduus* and related taxa inferred from 78 protein coding cpDNA regions.** Numbers indicate supporting values (BP, Bootstrap; PP, Posterior Probability). The asterisk indicates the *Scaevola porocarya* branch, which is compared with other species in the tree; this branch was reduced five times. Only supporting values under 100 (BP) or 1 (PP) are shown.

Full-size DOI: 10.7717/peerj.10687/fig-4

sister to *Cirsium arvense*. Additional ML and BI analyses of full cpDNA sequences and non-coding regions suggested similar relationships (Fig. S4).

Multiplex PCR and specific markers for *C. crispus*

The results of multiplex PCR for the two groups of primer pairs yielded similar products, both of which were designed to identify *C. crispus*. In the first group, a 323 bp band was found in *C. crispus,* whereas a 421 bp band was identified in *C. acanthoides* and *C. tenuiflorus* (Fig. 5A). By contrast, the combination of matK_463F, matK_1162R, CD_SNP_F2, and CD_SNP_R2 yielded a longer PCR product for *C. crispus* in comparison with the other species (Fig. 5B). The designed primer pairs were specific to all *C. crispus* samples examined in this study (Figs. S5 and S6).

DISCUSSION

Conservatism of *Carduus* cpDNA

Chloroplast genomes are highly conserved in angiosperms with respect to gene content and order (*Sugiura, 1992*). This conservative tendency was observed in the three newly

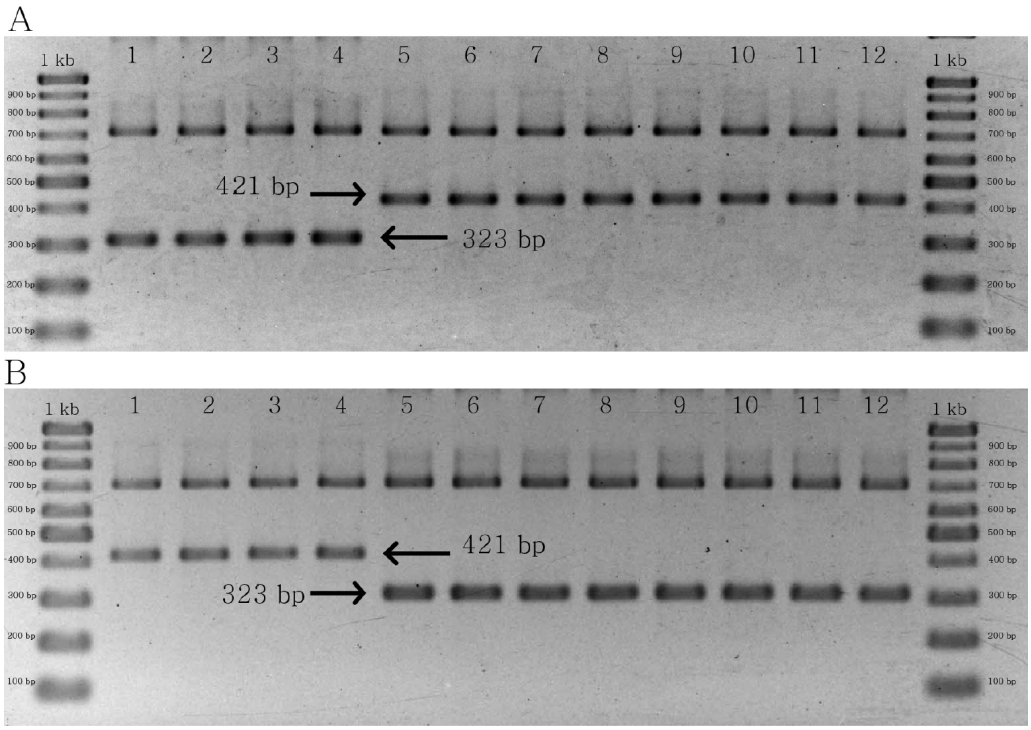

Figure 5 **PCR results of specific primer pairs for *C. crispus*.** (A) Combination of matK_463F, matK_1162R, CD_SNP_F1, and CD_SNP_R1, and (B) combination of matK_463F, matK_1162R, CD_SNP_F2, and CD_SNP_R2. Numbers 1–4 = *C. crispus*; 5–8 = *C. acanthoides*, and 9–12 = *C. tenuiflorus*.

Full-size DOI: 10.7717/peerj.10687/fig-5

sequenced *Carduus* cpDNA genomes, compared to other Asteraceae members (Table 2). Other cpDNA sequences have revealed unique genomic events in Asteraceae. For example, the *atpB* gene, which encodes the CF1 ATPase beta subunit, is annotated as a pseudogene in *Aster spathulifolius* due to a deletion within the coding region (Choi and Park, 2015). Similarly, *trnT_GGU* was completely deleted or pseudogenised in the tribe Gnaphalieae (*Lee et al., 2017*). Duplication of *trnF_GAA* has been identified in *Taraxacum* (*Salih et al., 2017*). No comparable genomic events are present in *Carduus* or other members of subfamily Carduoideae (Table 2, Fig. S7). However, nucleotide diversity data pointed to potential regions for further study of phylogeny and population genetics, and the development of *Carduus*-specific molecular markers (Table S2, Fig. 2). The number of species we examined for this study was low relative to the approximately 32,000 known Asteraceae species. Therefore, additional studies that include the majority of Asteraceae species should be conducted to explore the overall evolutionary trends in the chloroplast genomes of this globally-distributed family.

Chloroplast genomes provide useful molecular data for reconstructing phylogeny, exploring biogeography, and estimating divergence time in angiosperm lineages (*Do, Kim & Kim, 2014*; *Nguyen, Kim & Kim, 2015*; *Kim, Kim & Kim, 2016*; *Kim & Kim, 2018*). Repetitive sequences in the chloroplast genome provide useful information for studying

genomic rearrangement and phylogeny (*Cavalier-Smith, 2002*; *Nie et al., 2012*; *Yi et al., 2013*; *Kim & Kim, 2018*). In addition, existing repeats might result in the accumulation of new repeats in cpDNA (*Asano et al., 2004*). One of the crucial molecular data types in cpDNA is SSR sequences. Other studies have used SSRs to develop specific markers for different species and to study the genetic diversity of angiosperms (*Ishikawa, Sakaguchi & Ito, 2016*; *Luo et al., 2016*; *Marochio et al., 2017*; *Han et al., 2018*). In this study, although cpDNA sequences were highly conserved, the three *Carduus* species were found to have different numbers of SSRs (Table S4). While we did not develop SSR markers or conduct population studies of *Carduus,* the SSR information we obtained may be useful in future studies on *Carduus* species. In addition to the repeats shared among the three species, *C. crispus* had three unique repetitive sequences (Table S4), which may be useful in population studies, phylogenetic analyses, and the development of additional molecular markers.

Uncertain relationships among *Carduus*

Phylogenetic analyses of the Asteraceae have identified ambiguous relationships between *Carduus*, *Cirsium*, and *Silybum* (*Fu et al., 2016*; *Panero, 2016*; *Arnelas et al., 2018*); for example, ITS data suggests that *Carduus* is polyphyletic (*Häffner & Hellwig, 1999*). Although three coding regions (*matK*, *rbcL*, and *ndhF*) were used to reconstruct phylogenetic relationships, the position of *Carduus* remained unresolved (*Fu et al., 2016*). We used 78 protein-coding regions to clarify these relationships; however, the phylogeny of *Carduus* and related taxa remains unclear (Fig. 4). Specifically, *C. acanthoides* was found to be close to *Silybum marianum* whereas *C. crispus* and *C. tenuiflorus* form a clade with *Cirsium arvense*. While non-coding regions can be useful in reconstructing the phylogeny of lower taxa, we were unable to recover the monophyly of *Carduus* using data from non-coding regions, including the eight hotspot areas as well as the combined data from coding and non-coding regions (Fig. S4). These issues suggest the need for additional studies on the phylogeny of *Carduus* and other members of the subfamily Carduoideae using supplementary molecular data and morphology.

Implications of SNP data for developing molecular markers for *Carduus*

SNPs are useful in population studies due to its extremely abundant presence in the angiosperms genomes (*Cui et al., 2017*; *Fischer et al., 2017*; *Pantoja et al., 2017*), and are effective in phylogenetic analysis (*Leaché & Oaks, 2017*). In addition, various molecular markers have been developed for different angiosperm species based on SNP data from chloroplast genomes (*Khlestkina & Salina, 2006*; *Wang et al., 2015a*; *Wang et al., 2015b*; *Hyun et al., 2019*; *Xia et al., 2019*). We successfully developed a molecular marker, inferred from SNP data, to distinguish *C. crispus* from *C. acanthoides* and *C. tenuiflorus* (Fig. S3). Our marker demonstrates that nucleotide sequence variations can provide rapid molecular identification of *C. crispus*. We focused on *C. crispus* because it exhibits the characteristics of an invasive species (*Dunn, 1976*; *Verloove, 2014*; *Jung et al., 2017*), and may also have value for the treatment of obesity and cancer (*Xie, Li & Jia, 2005*; *Davaakhuu, Sukhdolgor & Gereltu, 2010*; *Lee et al., 2011*; *Tunsag, Davaakhuu & Batsuren, 2011*). Various DNA-based

markers (i.e., inter-simple sequence repeats [ISSRs], sequence characterisation of amplified regions [SCARs], and SSRs) have been developed to authenticate medicinal plants to ensure safety and efficacy (*Hao et al., 2010*; *Sarwat et al., 2012*; *Ganie et al., 2015*; *Ward, Gaskin & Wilson, 2008*). Additionally, molecular data are useful for understanding invasion processes of alien plants (Ward et al., 2008). We developed a SNP-based molecular marker for *C. crispus* (Fig. 5, Figs. S5 and S6) that may be used to detect *C. crispus* invasions in their early stages, and to develop suitable management strategies. Our alignment results also identified specific SNPs for *C. acanthoides* and *C. tenuiflorus*, which may be used to create molecular markers for these species (Fig. S1). Although we only used the SNP in *matK*, use of the complete cpDNA sequences of *Carduus* will enable the mining of SNPs from other regions for developing molecular markers for *C. crispus* and related species.

CONCLUSIONS

In this study, we provided the first complete cpDNA sequences for *Carduus* species. Despite the absence of significant differences (i.e., inversions, deletions, and duplications) between the chloroplast genomes of *Carduus* and those of related taxa, the newly acquired cpDNA sequences have value as a resource in future studies of the evolution of the chloroplast genome in Carduoideae and Asteraceae. Additionally, the 78 protein-coding regions of the chloroplast genome revealed uncertainty regarding the position of *Carduus* within the subfamily Carduoideae, and suggested the need for additional studies to reconstruct relationships not only among thistles, but among other members of the Asteraceae as well. The methods and protocols used in developing molecular markers for *C. crispus* are easy to apply and may be useful as a standard method in other studies of Asteraceae species.

ACKNOWLEDGEMENTS

The authors thank the Korea National Arboretum (KNA), National Institute of Biological Resources (NIBR), New York Botanical Garden (NYBG), Carnegie Museum Herbarium (CM), and the United States National Herbarium (US) for providing materials in support of this study. We also appreciate the helpful comments of anonymous reviewers.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This research was funded by grants from scientific research (2017-2018) of the Animal and Plant Quarantine Agency of the Republic of Korea and the Gachon University research fund of 2019 (GCU-2019-0821). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Animal and Plant Quarantine Agency of the Republic of Korea and Gachon University.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Joonhyung Jung and Hoang Dang Khoa Do performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- JongYoung Hyun performed the experiments, prepared figures and/or tables, and approved the final draft.
- Changkyun Kim and Joo-Hwan Kim conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

The complete cpDNA sequences of *C. crispus, C. acanthoides*, and *C. tenuiflorus* are available at GenBank: MK652229, MK652228, and MK652230, respectively.

Data Availability

The following information was supplied regarding data availability:

The three *Carduus* species' sequence data are available at NCBI: PRJNA645567. Specimen voucher numbers and deposition information are available in Table 1.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.10687#supplemental-information.

REFERENCES

Angiosperm Phylogeny Group. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* **181(1)**:1–20 DOI 10.1111/boj.12385.

Arnelas I, Pérez-Collazos E, Devesa JA, López E, Catalan P. 2018. Phylogeny of highly hybridogenous Iberian *Centaurea* L. (Asteraceae) taxa and its taxonomic implications. *Plant Biosystems* **152(5)**:1182–1190 DOI 10.1080/11263504.2018.1435569.

Asano T, Tsudzuki T, Takahashi S, Shimada H, Kadowaki K. 2004. Complete nucleotide sequence of the sugarcane (*Saccharum officinarum*) chloroplast genome: a comparative analysis of four monocot chloroplast genomes. *DNA Research* **11**:93–99 DOI 10.1093/dnares/11.2.93.

Cavalier-Smith T. 2002. Chloroplast evolution: secondary symbiogenesis and multiple losses. *Current Biology* **12**:R62–R64 DOI 10.1016/S0960-9822(01)00675-3.

Chan PP, Lowe TM. 2019. tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods in Molecular Biology* **1962**:1–14 DOI 10.1007/978-1-4939-9173-0_1.

Choi IS, Choi BH. 2017. The distinct plastid genome structure of *Maackia fauriei* (Fabaceae: Papilionoideae) and its systematic implications for genistoids and tribe Sophoreae. *PLOS ONE* **12(4)**:e0173766 DOI 10.1371/journal.pone.0173766.

Choi KS, Park SJ. 2015. The complete chloroplast genome sequence of *Aster spathulifolius* (Asteraceae); genomic features and relationship with Asteracea. *Gene* **572(2)**:214–221 DOI 10.1016/j.gene.2015.07.020.

Christoph M. 2006–2010. Phobos 3.3.11. *Available at http://www.rub.de/ecoevo/cm/cm_phobos.htm*.

Cosner ME, Raubeson LA, Jansen RK. 2004. Chloroplast DNA rearrangements in Campanulaceae: phylogenetic utility of highly rearranged genomes. *BMC Evolutionary Biology* **4**:27 DOI 10.1186/1471-2148-4-27.

Cui C, Mei H, Liu Y, Zhang H, Zheng Y. 2017. Genetic diversity, population structure, and linkage disequilibrium of an association-mapping panel revealed by genome-wide SNP markers in sesame. *Frontiers in Plant Science* **8**:1189 DOI 10.3389/fpls.2017.01189.

Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLOS ONE* **5(6)**:e11147 DOI 10.1371/journal.pone.0011147.

Davaakhuu G, Sukhdolgor J, Gereltu B. 2010. Lipid lowering effect of ethanolic extract of *Carduus crispus* L. on hypercholesterolemic rats. *Mongolian Journal of Biological Sciences* **8(2)**:49–51.

Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research* **45(4)**:e18 DOI 10.1093/nar/gkw955.

Do HDK, Jung J, Hyun J, Yoon SJ, Lim C, Park K, Kim J-H. 2019. The newly developed single nucleotide polymorphism (SNP) markers for a potentially medicinal plant, *Crepidiastrum denticulatum* (Asteraceae), inferred from complete chloroplast genome data. *Molecular Biology Reports* **46**:3287–3297 DOI 10.1007/s11033-019-04789-5.

Do HDK, Kim J-H. 2017. A dynamic tandem repeat in monocotyledons inferred from a comparative analysis of chloroplast genomes in Melanthiaceae. *Frontiers in Plant Science* **8**:693 DOI 10.3389/fpls.2017.00693.

Do HDK, Kim JS, Kim JH. 2014. A *trnI_CAU* triplication event in the complete chloroplast genome of *Paris verticillata* M.Bieb. (Melanthiaceae, Liliales). *Genome Biology and Evolution* **6(7)**:1699–1706 DOI 10.1093/gbe/evu138.

Doing H, Biddiscombe EF, Knedlhans S. 1969. Ecology and distribution of the *Carduus nutans* group (nodding thistles) in Australia. *Vegetatio* **17**:313–351 DOI 10.1007/BF01965915.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**:11–15.

Dunn P. 1976. Distribution of *Carduus nutans, C. acanthoides, C. pycnocephalus*, and *C. crispus*, in the United States. *Weed Science* **24(5)**:518–524 DOI 10.1017/S0043174500066595.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32(5)**:1792–1797 DOI 10.1093/nar/gkh340.

Fischer MC, Rellstab C, Leuzinger M, Roumet M, Gugerli F, Shimizu KK, Widmer A. 2017. Estimating genomic diversity and population differentiation—an empirical comparison of microsatellite and SNP variation in *Arabidopsis halleri*. *BMC Genomics* **18**:69 DOI 10.1186/s12864-016-3459-7.

Fu Z, Jiao B, Nie B, Zhang G, Gao T-G. China Phylogeny Consortium. 2016. A comprehensive generic-level phylogeny of the sunflower family: implications for the systematics of Chinese Asteraceae. *Jnl of Sytematics Evolution* **54**:416–437 DOI 10.1111/jse.12216.

Ganie SH, Upadhyay P, Das S, Sharma MP. 2015. Authentication of medicinal plants by DNA markers. *Plant Gene* **4**:83–99 DOI 10.1016/j.plgene.2015.10.002.

Greiner S, Lehwark P, Bock R. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research* **47**:W59–W64 DOI 10.1093/nar/gkz238.

Haberle RC, Fourcade HM, Boore JL, Jansen RK. 2008. Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. *Journal of Molecular Evolution* **66**:350–361 DOI 10.1007/s00239-008-9086-4.

Häffner E, Hellwig FH. 1999. Phylogeny of the tribe Cardueae (Compositae) with emphasis on the subtribe Carduinae: an analysis based on ITS sequence data. *Willdenowia* **29(1/2)**:27–39 DOI 10.3372/wi.29.2902.

Han Z, Ma X, Wei M, Zhao T, Zhan R, Chen W. 2018. SSR marker development and intraspecific genetic divergence exploration of *Chrysanthemum indicum* based on transcriptome analysis. *BMC Genomics* **19**:291 DOI 10.1186/s12864-018-4702-1.

Hao DC, Chen SL, Xiao PG, Peng Y. 2010. Authentication of medicinal plants by DNA-based markers and genomics. *Chinese Herbal Medicines* **2(4)**:250–261 DOI 10.3969/j.issn.1674-6384.2010.04.003.

Hyun J, Do HDK, Jung J, Kim JH. 2019. Development of molecular markers for invasive alien plants in Korea: a case study of a toxic weed, Cenchrus longispinus L., based on next generation sequencing data. *PeerJ* **7**:e7965 DOI 10.7717/peerj.7965.

Ishikawa N, Sakaguchi S, Ito M. 2016. Development and characterization of SSR markers for *Aster savatieri* (Asteraceae). *Applications in Plant Sciences* **4(6)**:apps.1500143 DOI 10.3732/apps.1500143.

Jansen R, Michaels H, Palmer J. 1991. Phylogeny and character evolution in the asteraceae based on chloroplast dna restriction site mapping. *Systematic Botany* **16(1)**:98–115 DOI 10.2307/2418976.

Jung SY, Lee JW, Shin HT, Kim SJ, An JB, Heo TI, Chung JM, Cho YC. 2017. *Invasive Alien Plants in South Korea*. Pocheon: Korea National Arboretum.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Drummond A. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28(12)**:1647–1649 DOI 10.1093/bioinformatics/bts199.

Khlestkina EK, Salina EA. 2006. SNP markers: methods of analysis, ways of development, and comparison on an example of common wheat. *Russian Journal of Genetics* **42**:585–594 DOI 10.1134/S1022795406060019.

Kim K-J, Choi K, Jansen R. 2005. Two chloroplast DNA inversions originated simultaneously during the early evolution of the sunflower family (Asteraceae). *Molecular Biology and Evolution* **22**:1783–1792 DOI 10.1093/molbev/msi174.

Kim JS, Jang HW, Kim JS, Kim J-H. 2012. Molecular identification of *Schisandra chinensis* and its allied species using multiplex PCR based on SNPs. *Genes Genom* **34**:283–290 DOI 10.1007/s13258-011-0201-3.

Kim JS, Kim J-H. 2018. Updated molecular phylogenetic analysis, dating and biogeographical history of the lily family (Liliaceae: Liliales). *Botanical Journal of the Linnean Society* **187(4)**:579–593 DOI 10.1093/botlinnean/boy031.

Kim SC, Kim JS, Kim JH. 2016. Insight into infrageneric circumscription through complete chloroplast genome sequences of two *Trillium* species. *AoB PLANTS* **8**:plw015 DOI 10.1093/aobpla/plw015.

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 Research* **29(22)**:4633–4642 DOI 10.1093/nar/29.22.4633.

Leaché AD, Oaks JR. 2017. The utility of Single Nucleotide Polymorphism (SNP) data in phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* **48(1)**:69–84 DOI 10.1146/annurev-ecolsys-110316-022645.

Lee DH, Cho WB, Choi BH, Lee JH. 2017. Characterization of two complete chloroplast genomes in the tribe gnaphalieae (Asteraceae): gene loss or pseudogenization of *trn* T-GGU and implications for phylogenetic relationships. *Horticultural Science and Technoloy* **35(6)**:769–783 DOI 10.12972/kjhst.20170081. .

Lee EJ, Joo EJ, Hong YN, Kim YS. 2011. Inhibition of adipocyte differentiation by MeOH extract from *Carduus crispus* through ERK and p38 MAPK pathways. *Natural Product Sciences* **17(4)**:273–278.

Liu X, Zhou B, Yang H, Li Y, Yang Q, Lu Y, Gao Y. 2018. Sequencing and analysis of *Chrysanthemum carinatum* Schousb and *Kalimeris indica.* The complete chloroplast genomes reveal two inversions and *rbcL* as barcoding of the vegetable. *Molecules* **23**:1358 DOI 10.3390/molecules23061358.

Luo L, Zhang P, Ou X, Geng Y. 2016. Development of EST-SSR markers for the invasive plant *Tithonia diversifolia* (Asteraceae). *Applications in Plant Sciences* **4(7)**:apps.1600011 DOI 10.3732/apps.1600011.

Ma YP, Sun SY, Zhao L. 2018. The complete chloroplast genome of the marsh plant-*Leucanthemella linearis* (Asteraceae). *Mitochondrial DNA Part B* **3(1)**:237–238 DOI 10.1080/23802359.2018.1437801.

Mandel JR, Dikow RB, Siniscalchi CM, Thapa R, Watson LE, Funk VA. 2019. A fully resolved backbone phylogeny reveals numerous dispersals and explosive diversifications throughout the history of Asteraceae. *Proceedings of the National Academy of Sciences of the United States of America* **116(28)**:14083–14088 DOI 10.1073/pnas.1903871116.

Marochio CA, Bevilaqua MRR, Takano HK, Mangolim CA, Oliveira Junior RS, Machado M, Silva FP. 2017. Genetic admixture in species of *Conyza* (Asteraceae) as revealed by microsatellite markers. *Acta Scientiarum. Agronomy* **39(4)**:437–445 DOI 10.4025/actasciagron.v39i4.32947.

Nguyen PAT, Kim JS, Kim J-H. 2015. The complete chloroplast genome of colchicine plants (*Colchicum autumnale* L. and *Gloriosa superba* L.) and its application for identifying the genus. *Planta* **242**:223 DOI 10.1007/s00425-015-2303-7.

Nie X, Lv S, Zhang Y, Du X, Wang L, Biradar SS, Tan X, Wan F, Weining S. 2012. Complete chloroplast genome sequence of a major invasive species, crofton weed (*Ageratina adenophora*). *PLOS ONE* **7**:e36869 DOI 10.1371/journal.pone.0036869.

Panero JL. 2016. Phylogenetic uncertainty and fossil calibration of Asteraceae chronograms. *Proceedings of the National Academy of Sciences of the United States of America* **113(4)**:E411 DOI 10.1073/pnas.1517649113.

Pantoja PO, Simón-Porcar VI, Puzey JR, Vallejo-Marín M. 2017. Genetic variation and clonal diversity in introduced populations of *Mimulus guttatus* assessed by genotyping at 62 single nucleotide polymorphism loci. *Plant Ecology & Diversity* **10(1)**:5–15 DOI 10.1080/17550874.2017.1287785.

Park H, Kim C, Lee Y, Kim J-H. 2016. Development of chloroplast microsatellite markers for the endangered *Maianthemum bicolor* (Asparagaceae s.l.). *Applications in Plant Sciences* **4(8)**:1600032 DOI 10.3732/apps.1600032.

Posada D. 2008. jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* **25(7)**:1253–1256 DOI 10.1093/molbev/msn083.

Poovitha S, Stalin N, Balaji R, Parani M. 2016. Multi-locus DNA barcoding identifies matK as a suitable marker for species identification in Hibiscus L. *Genome* **59(12)**:1150–1156 DOI 10.1139/gen-2015-0205.

Ronquist F, Teslenko M, Van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61(3)**:539–542 DOI 10.1093/sysbio/sys029.

Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A. 2017. DnaSP 6: DNA sequence polymorphism analysis of large datasets. *Molecular Biology and Evolution* **34**:3299–3302 DOI 10.1093/molbev/msx248.

Salih RH, L'Majeský, Schwarzacher T, Gornall R, Heslop-Harrison P. 2017. Complete chloroplast genomes from apomictic *Taraxacum* (Asteraceae): identity and variation between three microspecies. *PLOS ONE* **12(2)**:e0168008 DOI 10.1371/journal.pone.0168008.

Sarwat M, Nabi G, Das S, Srivastava PS. 2012. Molecular markers in medicinal plant biotechnology: past and present. *Critical Reviews in Biotechnology* **32(1)**:74–92 DOI 10.3109/07388551.2011.551872.

Su Y, Huang L, Wang Z, Wang T. 2018. Comparative chloroplast genomics between the invasive weed *Mikania micrantha* and its indigenous congener *Mikania cordata*: Structure variation, identification of highly divergent regions, divergence time

estimation, and phylogenetic analysis. *Molecular Phylogenetics and Evolution* **126**:181–195 DOI 10.1016/j.ympev.2018.04.015.

Sugiura M. 1992. The chloroplast genome. *Plant Molecular Biology* **19**:149–168 DOI 10.1007/BF00015612.

Todorov KT, Cheshmedzhiev IV, Stoyanov PS, Mladenova TR, Dimitrova-Dyulgerova IZ. 2018. Taxonomic Structure of Genus Carduus L. in Bulgaria. *Ecologia Balkanica* **10(2)**:229–233.

Trifinopoulos J, Nguyen LT, Von Haeseler A, Minh BQ. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research* **44(W1)**:W232–W235 DOI 10.1093/nar/gkw256.

Tunsag J, Davaakhuu G, Batsuren D. 2011. New isoquinoline alkaloid from *Carduus crispus* L. *Mongolian Journal of Chemistry* **12(38)**:85–87.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3-new capabilities and interfaces. *Nucleic Acids Research* **40(15)**:e115 DOI 10.1093/nar/gks596.

Verloove F. 2014. Carduus acanthoides (Asteraceae), a locally invasive alien species in Belgium. *Dumortiera* **105**:23–28.

Vu THT, Le TL, Nguyen TK, Tran DD, Tran HD. 2017. Review on molecular markers for identification of Orchids. *Vietnam Journal of Science, Technology and Engineering* **59(2)**:62–75 DOI 10.31276/VJSTE.59(2).62.

Wang M, Cui L, Feng K, Deng P, Du X, Wan F, Song W, Nie X. 2015b. Comparative analysis of asteraceae chloroplast genomes: structural organization, RNA editing and evolution. *Plant Molecular Biology Reporter* **33**:1526–1538 DOI 10.1007/s11105-015-0853-2.

Wang B, Tan H, Fang W, Meinhardt LW, Mischke S, Matsumoto T, Zhang D. 2015a. Developing single nucleotide polymorphism (SNP) markers from transcriptome sequences for identification of longan (*Dimocarpus longan*) germplasm. *Horticulture Research* **2**:14065 DOI 10.1038/hortres.2014.65.

Ward S, Gaskin J, Wilson L. 2008. Ecological Genetics of Plant Invasion: What Do We Know? *Invasive Plant Science and Management* **1(1)**:98–109 DOI 10.1614/IPSM-07-022.1.

Xia W, Luo T, Zhang W, Mason AS, Huang D, Huang X, Tang W, Dou Y, Zhang C, Xiao Y. 2019. Development of high-density SNP markers and their application in evaluating genetic diversity and population structure in *Elaeis guineensis*. *Frontiers in Plant Science* **10**:130 DOI 10.3389/fpls.2019.00130.

Xie WD, Li PL, Jia ZJ. 2005. A new flovone glycoside and other constituents from *Carduus crispus*. *Pharmazie* **60**:233–236.

Yi X, Gao L, Wang B, Su Y-J, Wang T. 2013. The complete chloroplast genome sequence of *Cephalotaxus oliveri* (Cephalotaxaceae): evolutionary comparison of *Cephalotaxus* chloroplast DNAs and insights into the loss of inverted repeat copies in gymnosperms. *Genome Biology and Evolution* **5**:688–698 DOI 10.1093/gbe/evt042.

Yun SA, Gil H, Kim SC. 2017. The complete chloroplast genome sequence of *Saussurea polylepis* (Asteraceae), a vulnerable endemic species of Korea. *Mitochondrial DNA Part B* **2(2)**:650–651 DOI 10.1080/23802359.2017.1375881.

Zhang Y, Iaffaldano BJ, Zhuang X, Cardina J, Cornish K. 2017. Chloroplast genome resources and molecular markers differentiate rubber dandelion species from weedy relatives. *BMC Plant Biology* **17**:34 DOI 10.1186/s12870-016-0967-1.

View publication stats