

RESEARCH ARTICLE

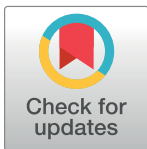
The chloroplast genome of *Cerasus humilis*: Genomic characterization and phylogenetic analysis

Xiaopeng Mu¹✉, Pengfei Wang¹✉, Junjie Du^{1,2*}, Yu Gary Gao³, Jiancheng Zhang¹

1 College of Horticulture, Shanxi Agricultural University, Jinzhong, Shanxi, P. R. China, **2** Shanxi Key Laboratory of Germplasm Improvement and Utilization in Pomology, Jinzhong, Shanxi, P. R. China, **3** OSU South Centers, College of Food, Agricultural, and Environmental Sciences, The Ohio State University, Shyville Road, Piketon, Ohio, United States of America

✉ These authors contributed equally to this work.

* djj738@163.com



OPEN ACCESS

Citation: Mu X, Wang P, Du J, Gao YG, Zhang J (2018) The chloroplast genome of *Cerasus humilis*: Genomic characterization and phylogenetic analysis. PLoS ONE 13(4): e0196473. <https://doi.org/10.1371/journal.pone.0196473>

Editor: Shashi Kumar, International Centre for Genetic Engineering and Biotechnology, INDIA

Received: November 29, 2017

Accepted: April 13, 2018

Published: April 25, 2018

Copyright: © 2018 Mu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The complete *C. humilis* chloroplast genome was submitted to GenBank, as Accession Number MF405921.

Funding: This study was performed with the support of the Major Subject of Shanxi Science and Technology Research (Grant No. 20121101010), the Doctoral Research Fund of Shanxi Agriculture University (Grant No. 2015ZZ19) and China Postdoctoral Science Foundation (Grant No. 2017M621106). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Cerasus humilis is endemic to China and is a new fruit tree species with economic and environmental benefits, with potential developmental and utilization applications. We report the first complete chloroplast genome sequence of *C. humilis*. Its genome is 158,084 bp in size, and the overall GC content is 36.8%. An inverted repeats (IR) of 52,672 bp in size is separated by a large single-copy (LSC) region of 86,374 bp and a small single-copy (SSC) region of 19,038 bp. The chloroplast genome of *C. humilis* contains 131 genes including 90 protein-coding genes, 33 transfer RNA genes, and 8 ribosomal RNA genes. The genome has a total 510 simple sequence repeats (SSRs). Of these, 306, 149, and 55 were found in the LSC, IR, and SSC regions, respectively. In addition, a comparison of the boundaries of the LSC, SSC, and IR regions of ten other *Prunus* species exhibited an overall high degree of sequence similarity, with slight variations in the IR boundary region which included gene deletions, insertions, expansions, and contractions. *C. humilis* lost the *ycf1* gene at the IRA/SSC border and it has the largest *ycf1* gene at the IRB/SSC border among these *Prunus* species, whereas the *rps19* gene was inserted at the IRB/LSC junction. Furthermore, phylogenetic reconstruction using 61 conserved coding-protein genes clustered *C. humilis* with *Prunus tomentosa*. Thus, the complete chloroplast genome sequence of *C. humilis* provides a rich source of genetic information for studies on *Prunus* taxonomy, phylogeny, and evolution, as well as lays the foundation for further development and utilization of *C. humilis*.

Introduction

Cerasus humilis (Bge.) Sok is a bush fruit tree that is endemic to China; its fruits are rich in calcium and are thus also known as “Calcium fruit” [1]. It is mainly distributed in the northeast, northwest, north, and other northern areas of China [2]. *C. humilis* has long existed in the wild, and studies on this species were only initiated in the 1990s [3]. After nearly 30 years of research studies, identification of varieties, and establishment of cultivation techniques, our understanding of this fruit tree species has improved and facilitated its large-scale artificial

Competing interests: The authors have declared that no competing interests exist.

cultivation. It has a strong root system and it shows strong adaptability to saline soil, harsh winter and drought [4]. Furthermore, it has strong soil erosion resistance, therefore it can be used for the improvement of damaged soil and environmental greening [5]. *C. humilis* is also used as an ornamental plant, generating striking white or red flowers. Its fruits look like cherries but have various colors including red, yellow, white, green, purple, etc. Moreover, it is a high-yield crop and its fruit has a unique flavor. The fruits can be eaten fresh, preserved, or processed into wine or other by-products. Its stems can be used in weaving and as high-protein cattle feed. The seed kernels have been traditionally used in Chinese medicine as “Yuliren” [5,6]. Based on its ecological and economic benefits, *C. humilis* has become an emerging multi-purpose fruit trees with a broad developmental and utilization potential.

Similar to the mitochondria and nuclei, chloroplasts are semi-autonomous organelles that function in the transfer and expression of genetic material DNA [7,8]. Chloroplasts are generally situated in the cytoplasmic matrix and play an important role in photosynthesis and fatty acid, starch, and amino acid synthesis [9]. In 1986, the complete chloroplast genome sequence of *Marchantia polymorpha* and tobacco (*Nicotiana tabacum*) were released and to date, the chloroplast whole genome sequence of 700 plant species has been completed [10]. The chloroplast genome is a closed cyclic structure that consists of large single copy (LSC), small single copy (SSC), and two reverse repeats (IR); its genome size ranges from 100 kb to 200 kb. Protein-coding genes are important in photosynthesis and other functions of the chloroplast and are further divided into nine categories: Photosystem I, cytochrome b6-f complex, Photosystem II, ATP synthase, ribulose bis-phosphate carboxylase, ribosome large subunit, ribosome small subunit, RNA polymerase, and a hypothetical protein [11]. The overall structure and genome of the chloroplast are relatively conserved, although in-depth studies in chloroplast genomics and biology have identified several mutations and small structural changes among various species including insertions, deletions, reversals, and translocations [12,13].

A famous Chinese fruit scientist earlier classified *C. humilis* as *Rosaceae cerasus* [14], whereas peach, apricot, plum, and cherry are classified as *Prunus*. Although *Prunus* consists of numerous species, only the chloroplast genome of ten species including *Prunus kansuensis*, *Prunus maximowiczii*, *Prunus pseudocerasus*, *Prunus yedoensis*, *Prunus persica*, *Prunus tomentosa*, *Prunus serotina*, *Prunus dulcis*, *Prunus cerasoides* and *Prunus padus* have been completely sequenced to date, with genome sizes ranging from 157,685 bp–158,955 bp. *C. humilis* may be further developed and potentially utilized in various industries. The complete chloroplast genome sequence of *C. humilis* could be applied in facilitating evolutionary analysis, accelerating breeding, identification of commercial cultivars and determination of varieties’ purity. In this study, the whole genome sequence of the chloroplast of *C. humilis* was obtained for the first time, and its structural characteristics and phylogenetic relationships were analyzed.

Materials and methods

Plant material and DNA extraction

Fresh and young leaves of the upper part of the branches from a cultivated variety of *C. humilis* (cv. Nongda 4, Fig 1) were collected from the germplasm resources of *C. humilis* of Shanxi Agricultural University (37°26'N, 112°32'E). Using 100 mg of a fresh plant sample, total genomic DNA was isolated using HP Plant DNA Kit D2485-01 (Omega Bio-Tek, Santa Clara, CA, USA), according to the manufacturer’s protocol.

Chloroplast genome sequencing, assembly, and annotation

After quantification and qualification of total genomic DNA of the sample, a paired-end library of 2 × 150 bp and insert size of ~350 bp was constructed according to the Illumina



Fig 1. Whole plant of *C. humilis* (cv. 'Nongda 4'). (A) Young fruit stage. (B) Mature fruit stage.

<https://doi.org/10.1371/journal.pone.0196473.g001>

standard library, which was then sequenced on an Illumina Hiseq 2500 platform. The resulting low-quality raw data were filtered, and the linker sequences were removed, thereby resulting in 20.86 Gb of clean data, with a quality value $\geq$ Q30, accounting for 91.37%. The genome of *P. pseudocerasus* was used as reference (GenBank Accession No. NC_030599.1), and NOVO-Plasty software was used to assemble the chloroplast genome, the mistake parameter was set by default [15]. Three software were used for genome degenerate base correction, namely, bwa, samtools, and bcftools [16,17]. The complete *C. humilis* chloroplast genome sequence was annotated using the online software CpGAVAS developed by Liu et al [18] (<http://www.herbalgenomics.org/0506/cpgavas/analyzer/annotate>). TRNAscan-SE was used in tRNA confirmation [19]; <http://lowelab.ucsc.edu/tRNAscan-SE/>). Genomic circle graphs were drawn using OGDRAW [20] (<http://ogdraw.mpimp-golm.mpg.de/>). The complete *C. humilis* chloroplast genome was submitted to GenBank, as Accession Number MF405921.

Comparative genome analysis

To investigate the differences between the *C. humilis* genome sequence and those of the other members of the same genus, we used the LAGAN alignment strategy as implemented in the mVISTA software to compare the whole genome sequence of ten species of *Prunus* with *C. humilis*, using *C. humilis* as reference [21] (http://genome.lbl.gov/cgi-bin/VistaInput?num_seqs=2). In addition, the differences in the types and gene sizes of the IR border genes of five species were analyzed.

Simple sequence repeat (SSR) analysis

Microsatellites are also known as SSRs, which occur as mono-, di-, tri-, tetra-, penta-, and hexa-nucleotide repeats. The SSRs in the chloroplast genome of *C. humilis* were identified using PHOBOS 3.3.12, and the alignment was screened in terms of matches, mismatches, gaps, and N positions using the scores of 1, -5, -5, and 0, respectively [22]. At the same time, the SSR of the IR, LSC, SSC, and coding regions, introns, and intergenic regions corresponding to different regions were analyzed.

Phylogenetic analysis

The chloroplast genome sequences of ten *Prunus* species were downloaded from NCBI. Approximately 61 coding-protein genes were identified in *C. humilis* and the other ten *Prunus* species, which were then aligned using mafft software's FFT-NS-2 algorithm [23]. After the results were integrated, phylogenetic reconstruction was performed using RAXML Version 8 tool as implemented in CIPRES [24].

Results

General features of the *C. humilis* chloroplast genome

The genome of the *C. humilis* circular chloroplast is 158,084 bp in size and exhibits the typical quadripartite structure found in most land plants (Fig 2). The IR region spans 52,672 bp, whereas the large single copy (LSC) and small single copy (SSC) region span 86,374 bp and 19,038 bp, respectively. GC contents of the complete genome is 36.8%, the LSC and SSC region are 34.5% and 30.5%, respectively, whereas that of the IR regions is 42.6% (Table 1). Similar to other terrestrial plants such as *Cynara humilis*, a species of family Asteraceae, the IRS showed higher GC content, which was mainly due to the rRNA genes *rrn16*, *rrn23*, *rrn4.5*, and *rrn5* [25].

There are a total of 131 genes, including 90 protein-coding genes, 33 transfer RNA (tRNA), and 8 ribosomal RNA (rRNA) genes. Among these, 17 genes have two copies while *rps15* and *rps12* have more than copies (Tables 1 and 2). In the functional classification of genes, 44 genes (unique) are associated with photosynthetic functions such as *atpH*, *ndhD*, and *rbcL*. The function of the conserved open reading frames of *ycf* (*ycf1*, *ycf2*, *ycf3*, *ycf4* and *ycf15*) is unknown. The nine genes, namely, *atpF*, *ndhA*, *ndhB*, *petB*, *rpl16*, *rpl2*, *rpoC1*, *rps16*, and *ycf1*, contain a single intron. The two copies of the *rpl2* and *ndhB* genes all have one intron. In addition, there are two introns in the *ycf3* and *clpP* genes (Table 2).

Comparative genome analysis

To further analyze the characteristics of the *C. humilis* chloroplast genome, we compared the assembled genomes with the chloroplast genomes of ten other *Prunus* species in NCBI. First, we compared the basic characteristics of their genomes. The size of the *C. humilis* chloroplast genome is smaller than those of *P. padus*, *P. pseudocerasus* and *P. tomentosa*. However, it has the highest GC content and the second largest number of encoded proteins (N = 90). In addition, it has 131 genes in total and the number of tRNA genes is the lowest among these 11 members of the *Prunus* genus (Table 3). The sequence identity was analysed using program mVISTA by aligning the ten *Prunus* species' chloroplast genomes with *C. humilis* (Fig 3). As expected, the sequence identity of the eleven chloroplast genome sequences was high, indicating that these are highly conserved. However, several variations were found in noncoding and single copy regions of these eleven chloroplast genome (Fig 3).

In addition, we compared the SSC region and the IR-LSC and IR-SSC boundaries of *C. humilis* with other four species of *Prunus*: *P. maximowiczii*, *P. kansuensis*, *P. persica* and *P. pseudocerasus* (Fig 4). In the SSC region, all genes are conserved except *tmS-AGA*. The *rps19* gene is present at the LSC/IRA junction of the *C. humilis*, with a size of 142 bp which is shorter than other species' *rps19* gene. The IR boundary regions also showed slight variations, which included gene deletions, insertions, expansions, and contractions (Fig 4). *C. humilis* and *P. maximowiczii* lost the *ycf1* gene in IRA/SSC boundary region. Moreover, *rps19* gene was deleted in *P. maximowiczii*, but inserted in *C. humilis* at the IRB/LSC boundary. The size of these bordering genes also varied. The small *ycf1* gene ranged in size from 1,003 bp to 1,059

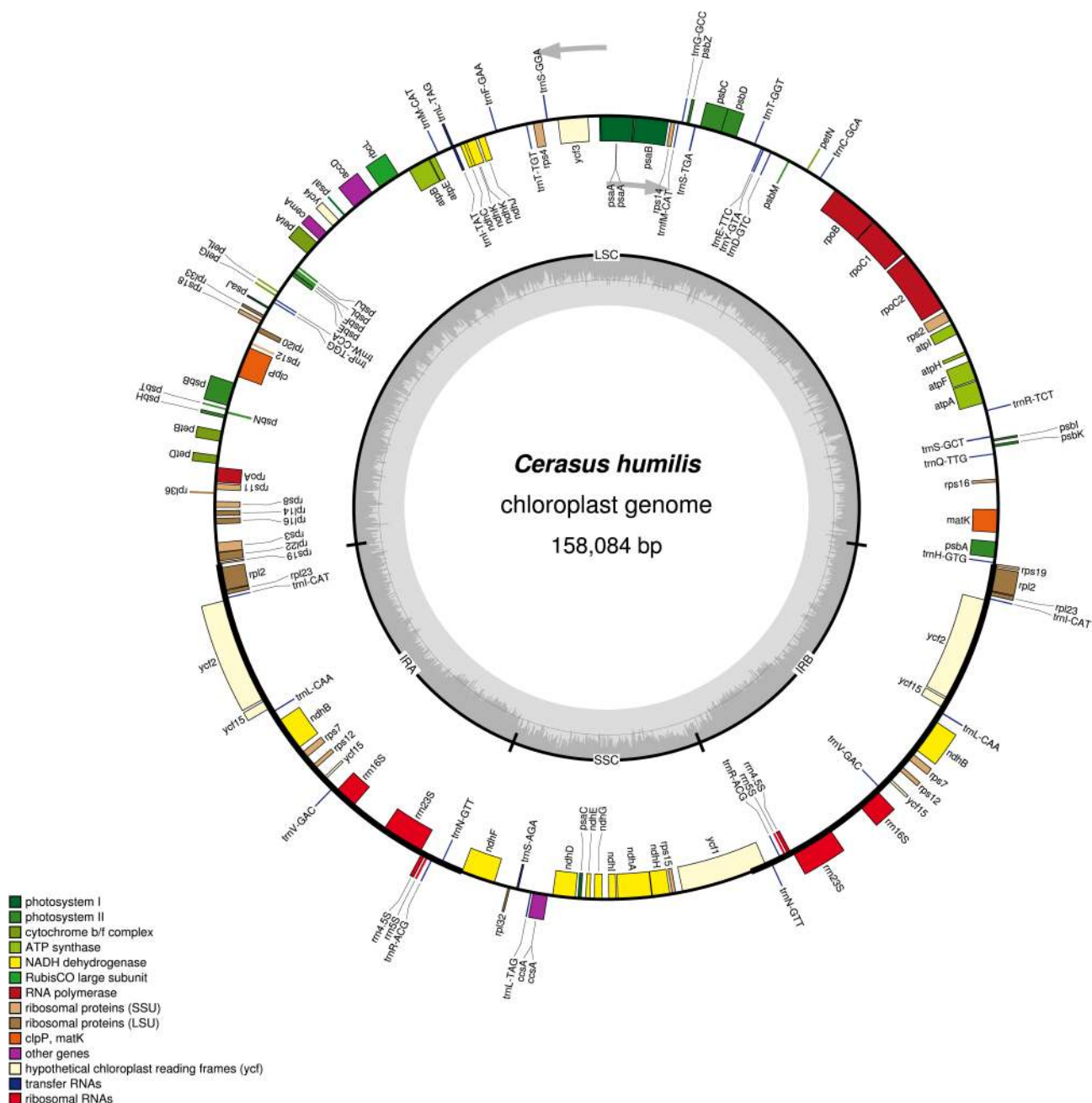


Fig 2. Gene map of the *C. humilis* chloroplast genome. Genes drawn inside the circle are transcribed clockwise, whereas genes outside are transcribed counterclockwise. Genes are color coded by their function in the legend. The area in darker gray and lighter gray in the inner circle indicates GC content and AT content, respectively.

<https://doi.org/10.1371/journal.pone.0196473.g002>

bp, and that of the larger *ycf1* gene ranged from 5,607 bp to 5,664 bp, with the largest observed in *C. humilis*. At the IRB/LSC junction, the *rps19* gene was within the range of 183 bp–210 bp, whereas *P. maximowiczii* has no *rps19* gene at this area.

No major errors in genome assembly were observed in this study, the *ycf1* and *rps19* genes showed deletions, insertions, expansions and contractions among the five species, indicating that these genes may be used as markers in studying the evolution of the IR/SSC and IR/LSC junction in the *Prunus*.

Table 1. Summary statistics for assembly of *C. humilis* chloroplast genomes.

Genome features	<i>Cerasus humilis</i>
Size (bp)	158,084
LSC length (bp)	86,374
SSC length (bp)	19,038
IR length (bp)	52,672
Number of gene	131
Protein-coding genes	90
tRNA genes	33
rRNA genes	8
GC content (%)	36.8
GC content in LSC (%)	34.5
GC content in SSC (%)	30.5
GC content in IR (%)	42.6

<https://doi.org/10.1371/journal.pone.0196473.t001>

SSR analysis

A total of 510 SSRs were identified in the chloroplast genome of *C. humilis*. Then, we studied the distribution, presence, and types of SSRs (Fig 5). Of the 510 SSRs, 306 (60%), 55 (11%), and 149 (29%) were discovered within the LSC, SSC, and IR regions, respectively (Fig 5A). In addition, 263, 510, and 19 SSRs were situated within the protein-coding regions, intergenic

Table 2. Gene content in *C. humilis* chloroplast genome.

Function	Family	Genes
Genes for photosynthesis	Subunits of ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> ^a , <i>atpH</i> , <i>atpI</i>
	Subunits of NADH-dehydrogenase	<i>ndhA</i> ^a , <i>ndhB</i> ^a (2), <i>ndhC</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhF</i> , <i>ndhG</i> , <i>ndhH</i> , <i>ndhI</i> , <i>ndhJ</i> , <i>ndhK</i>
	Subunits of cytochrome b/f complex	<i>petA</i> , <i>petB</i> ^a , <i>petD</i> , <i>petG</i> , <i>petL</i> , <i>petN</i>
	Subunits of photosystem I	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i>
	Subunits of photosystem II	<i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</i> , <i>psbH</i> , <i>psbI</i> , <i>psbJ</i> , <i>psbK</i> , <i>psbL</i> , <i>psbM</i> , <i>psbN</i> , <i>psbT</i> , <i>psbZ</i>
	Subunit of RuBisCO	<i>rbcL</i>
Other genes	Subunit of Acetyl-CoA-carboxylase	<i>accD</i>
	c-type cytochrome synthesis gene	<i>ccsA</i>
	Envelop membrane protein	<i>cemA</i>
	Protease	<i>clpP</i> ^b
	Maturase	<i>matK</i>
Self-replication	Large subunit of ribosome	<i>rpl14</i> , <i>rpl16</i> ^a , <i>rpl20</i> , <i>rpl22</i> , <i>rpl23</i> (2), <i>rpl2</i> ^a (2), <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>
	DNA-dependent RNA polymerase	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> ^a , <i>rpoC2</i>
	Small subunit of ribosome	<i>rps11</i> , <i>rps12</i> (3), <i>rps14</i> , <i>rps15</i> , <i>rps16</i> ^a , <i>rps18</i> , <i>rps19</i> (2), <i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> (2), <i>rps8</i>
	rRNA Genes	<i>rrn4.5S</i> (2), <i>rrn5S</i> (2), <i>rrn23S</i> (2), <i>rrn16S</i> (2)
	tRNA genes	<i>trnR</i> -ACG (2), <i>trnS</i> -TGA, <i>trnT</i> -TGT, <i>trnS</i> -GCT, <i>trnF</i> -GAA, <i>trnH</i> -GTG, <i>trnD</i> -GTC, <i>trnL</i> -CAA (2), <i>trnE</i> -TTC, <i>trnG</i> -GCC, <i>trnI</i> -CAT (2), <i>trnI</i> -TAT (2), <i>trnR</i> -TCT, <i>trnL</i> -TAG, <i>trnM</i> -CAT, <i>trnY</i> -GTA, <i>trnQ</i> -TTG, <i>trnT</i> -GGT, <i>trnS</i> -AGA (2), <i>trnN</i> -GTT (2), <i>trnW</i> -CCA, <i>trnM</i> -CAT, <i>trnS</i> -GGA, <i>trnP</i> -TGG, <i>trnV</i> -GAC (2), <i>trnC</i> -GCA
Unknown function	Conserved open reading frames	<i>ycf1</i> ^a , <i>ycf2</i> (2), <i>ycf3</i> ^b , <i>ycf4</i> , <i>ycf15</i> (4)

^a: Gene containing a single intron

^b: Gene containing two introns.

<https://doi.org/10.1371/journal.pone.0196473.t002>

Table 3. Comparison of chloroplast genome characteristics in various *Prunus* species.

Species	GenBank Acc. No.	Size (kb)	GC%	Protein	rRNA	tRNA	Gene	Pseudogene
<i>Cerasus humilis</i>	MF405921	158.084	36.80	90	8	33	131	-
<i>Prunus kansuensis</i>	NC_023956.1	157.736	36.76	86	8	39	133	-
<i>Prunus maximowiczii</i>	NC_026981.1	157.852	36.73	84	8	37	129	-
<i>Prunus persica</i>	NC_014697.1	157.79	36.76	85	8	37	131	1
<i>Prunus pseudocerasus</i>	NC_030599.1	157.834	36.73	86	8	37	131	-
<i>Prunus yedoensis</i>	NC_026980.1	157.859	36.73	84	8	37	129	-
<i>Prunus serotina</i>	NC_036133.1	158.778	36.60	82	8	36	126	-
<i>Prunus dulcis</i>	NC_034696.1	157.723	36.80	86	0	38	124	-
<i>Prunus cerasoides</i>	NC_035891.1	157.685	36.70	84	8	37	129	-
<i>Prunus tomentosa</i>	NC_036394.1	158.356	36.80	91	0	36	127	-
<i>Prunus padus</i>	NC_026982.1	158.955	36.59	84	8	37	129	-

<https://doi.org/10.1371/journal.pone.0196473.t003>

spacer regions, and introns, respectively. The presence of SSRs in the SSC and IR regions was the same as that above except for the LSC, i.e., the SSR number in the protein-coding regions was the highest, followed by intergenic spacers, and then introns. Furthermore, the predominant SSRs included dinucleotides, trinucleotides, and tetranucleotides, with the repeat-number equal to three, and dinucleotide span the repeat-number range from 3 to 9.

Phylogenetic analysis

The 61 conservative coding-protein genes were used to analyze the phylogenetic relationships among the members of *Prunus* (Fig 6). The phylogenetic tree demonstrated that *P. persica* and *P. kansuensis* formed a monophyletic clade while *C. humilis* and *P. tomentosa* formed another monophyletic clade. These two clades were then clustered with *P. dulcis* to form a larger branch (Fig 6). *C. humilis* and *P. tomentosa* were previously classified to the subgenus *Microcerasus* based on their morphological characteristics [14]. The phylogenetic results of this study provide new evidence for this classification at the genome level.

Discussion

In this study, the complete chloroplast genome of *C. humilis* was sequenced and annotated, and the SSRs and comparative genomics of the genome were analyzed. Moreover, phylogenetic reconstruction of *C. humilis* and ten other *Prunus* species was performed. Although the chloroplast genomes are highly conserved in terms of genomic structure and size, the IR/SC border genes varied in size and in type, which is a typical feature of chloroplast genomes [10,26,27]. The genomic structure of ten *Prunus* species and *C. humilis* are also relatively conserved, whereas differences in the presence of genes in the IR/SC border region were observed. *P. kansuensis* and *C. humilis* lost the *ycf1* gene in the IRA/SSC border region. *Ycf1* is one of the giant ORFs in most higher plants of chloroplast genomes and it usually spans the boundary of the IR and SSC regions of the plastid genome [28]. However, in the orchid genus *Phalaenopsis*, the entire *ycf1* gene is situated within the SSC region [29,30]. In addition, the function of the *ycf1* gene remains elusive [31]. In tobacco, it has been suggested that the *ycf1* gene does not encode a functional protein while some other studies have indicated that it encodes a protein that is essential for cell survival or related to the ABC-transporters [31,32]. The higher variation in the *ycf1* gene could provide superior resolution and support at lower taxonomic levels in *Orchidaceae* [28]. In our study, the small size of the *ycf1* gene in IRA/SSC boundary ranges from 1,003 bp to 1,059 bp which is significantly smaller than the larger *ycf1* gene (size range:

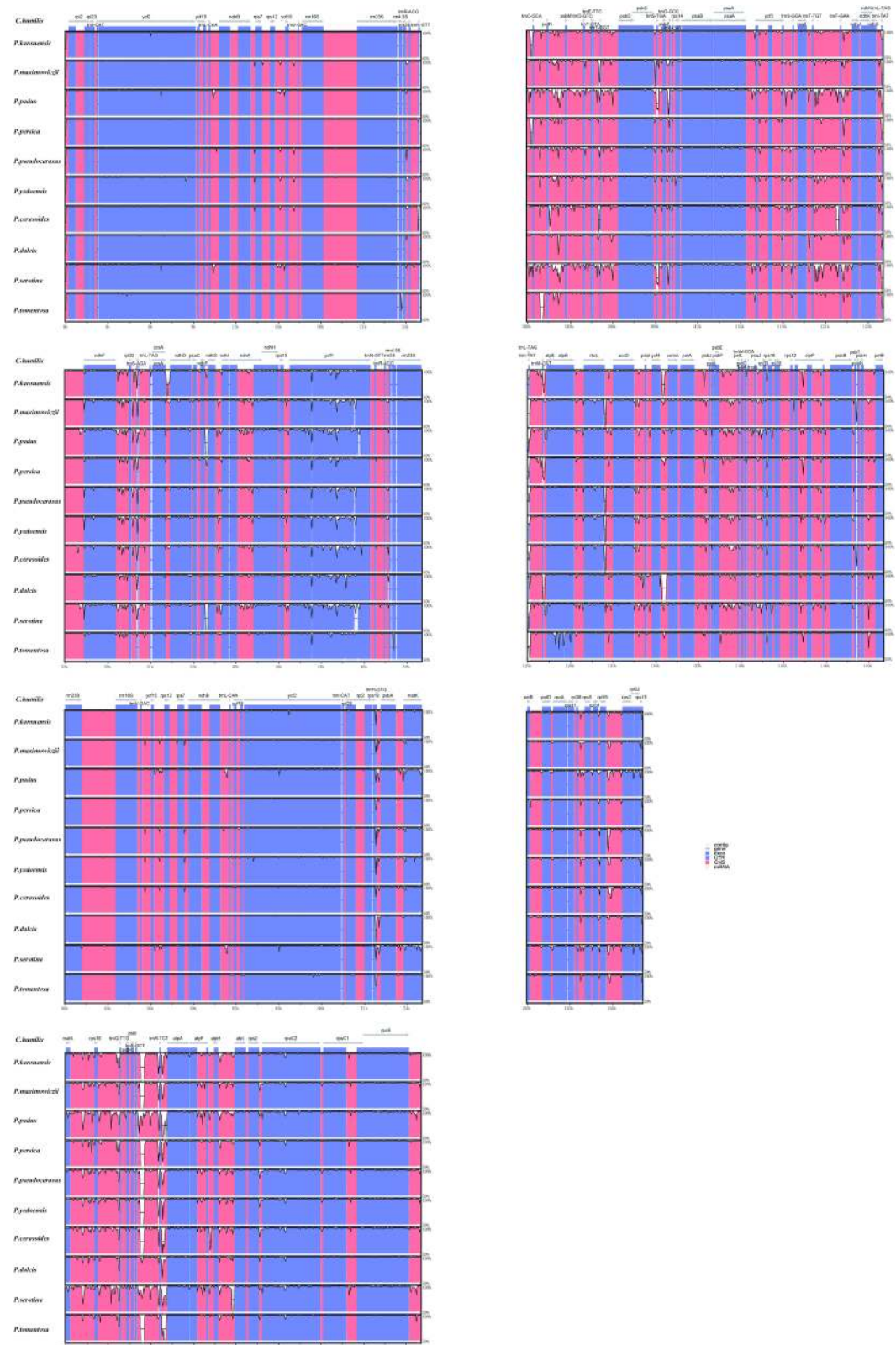


Fig 3. Visualization of genome alignment of the chloroplast genomes of ten *Prunus* species using *C. humilis* as reference using mVISTA. Vertical scale indicates the percentage of identity, which ranges from 50% to 100%. Coding regions are marked in blue, and non-coding regions are indicated in red. Gray arrows indicate the position and direction of each gene.

<https://doi.org/10.1371/journal.pone.0196473.g003>

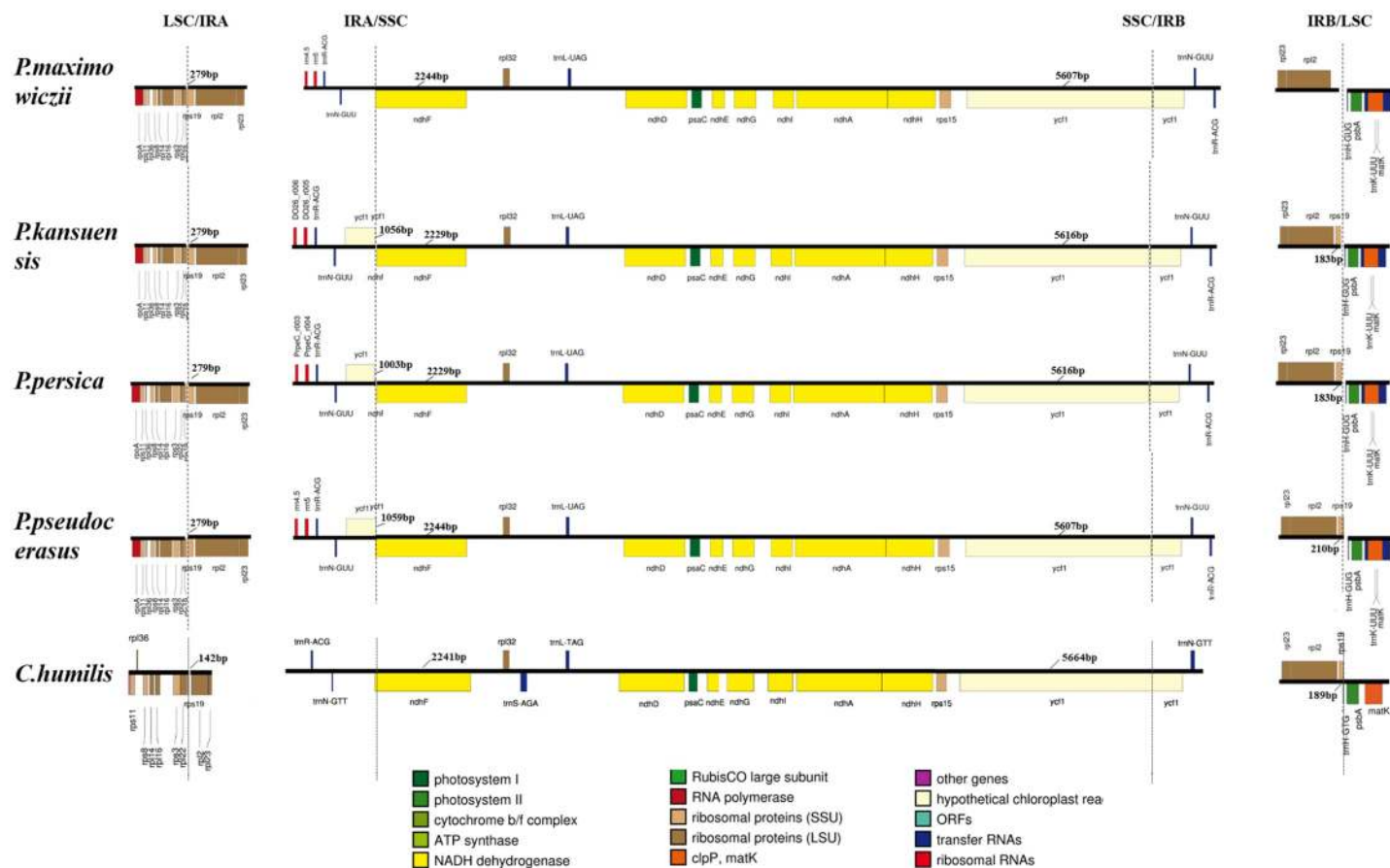


Fig 4. Comparison of the boundaries of the LSC, SSC, and IR regions in the *C. humilis* chloroplast genome with those of four other *Prunus* species.

<https://doi.org/10.1371/journal.pone.0196473.g004>

5,607 bp–5,662 bp). The role of this gene deletion in the evolution of chloroplast genome of *Prunus* requires further investigations.

In this study, two copies of *C. humilis* *rps19* gene were found in the boundaries of IRA/LSC and IRB/SSC respectively, whereas the *rps19* gene was deleted from the IRB/SSC boundary in *P. kansuensis*. Similar results were found in two leguminous plants *Cajanus* and *Milletia* [33–35]. In addition, it was reported that one of the two *rps19* genes in the IR/SC boundaries is usually a pseudogene. For example, *Dianthus* encodes one copy of the *rps19* gene at the IRB/SSC junction and a pseudogene *rps19* at IRA/LSC junction while the size of the pseudogene *rps19* is shorter than that of the regular *rps19* gene [34]. In three *Cardiocrinum* (Liliaceae) species, the *rps19* gene located in the boundary between LSC and IRA apparently lost its protein-coding ability due to partial gene duplication [36]. The *rps19* gene of *C. humilis* located in the boundary between LSC/IRA is much shorter than those of other four *Prunus* species which means it could also be a pseudogene. Roles of the *rps19* gene's insertion and deletion observed in the chloroplast genomes of *Prunus* species remains to be elucidated.

C. humilis belongs to the genus *Prunus* which consists of approximately 250 species. Many economically important fruit crops such as cherry, plum, apricot, almond, peach are in this genus. However, there are different opinions about how *C. humilis* should be grouped and named. A famous Chinese plant taxonomist earlier classified *C. humilis* as *Rosaceae cerasus* because he believed *C. humilis* is a close relative to *P. pseudocerasus* (cherry) [14]. However, the phylogenetic trees we built in this study showed that *C. humilis* is closer to *P. tomentosa*

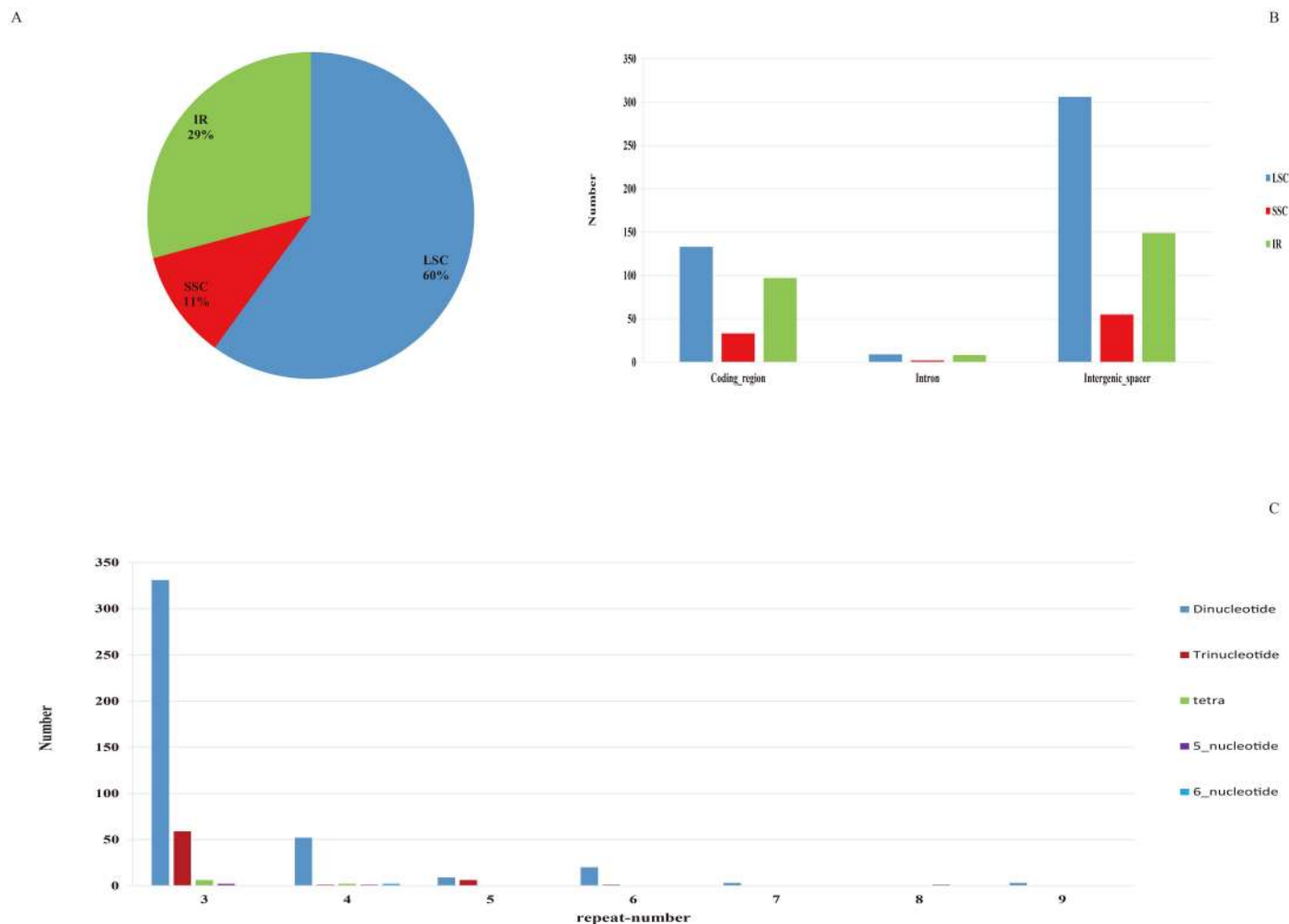


Fig 5. The distribution and type of SSRs in the chloroplast genome of *C. humilis*. (A) Percentage of SSRs in the LSC, SSC, and IR regions. (B) Distribution of SSRs in the protein-coding regions, introns, and intergenic spacers of the LSC, SSC, and IR regions. (C) Presence of polymers in the chloroplast genome of *C. humilis*.

<https://doi.org/10.1371/journal.pone.0196473.g005>

than to *P. pseudocerasus*. Our results agree with another classification which put *C. humilis* and *P. tomentosa* to the same subgenus *Microcerasus* based on their morphological characteristics [37]. Therefore, the results of this study might draw attentions of other scientists who have been working on clarifying the phylogenetic relationships among species of genus *Prunus*. More chloroplast genome sequences will be needed to improve our understanding of the evolution of *Prunus* species. Our findings not only provide the foundation for further studies on the evolution of chloroplast genome of *C. humilis* and *Prunus* species, but also serve as the basis for the molecular identification of *Prunus* species.

C. humilis was a wild species 30 years ago and it had sour and astringent taste which was quite different for people to eat. Our team started a program including domestication and breeding of this species in 1987 aiming at improving its fruit characters for commercial cultivation. One part of this program is to introduce specific advantageous traits into *C. humilis* from cultivated crops by crossing. However, little has been accomplished despite a lot of effort. Success in breeding is determined by genetic compatibility and chloroplast genomes serve as a valuable tool for identifying plants that are likely to be closely related and, therefore, genetically compatible [38]. We believe that chloroplast genome of *C. humilis* can provide useful

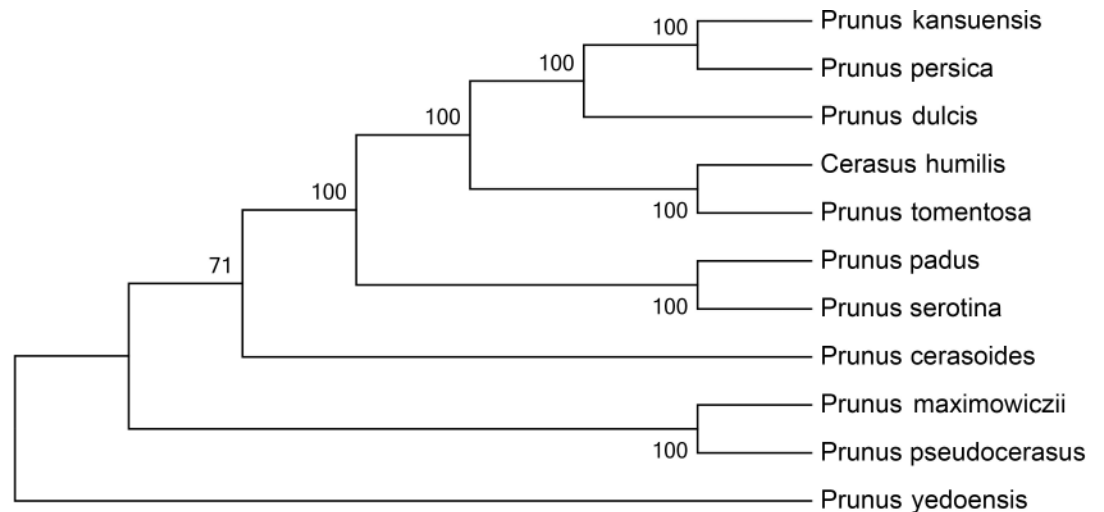


Fig 6. Phylogenetic reconstruction of the *Prunus* species based on 61 conserved coding-protein genes using the maximum likelihood approach.

<https://doi.org/10.1371/journal.pone.0196473.g006>

information for us to select proper parental combinations to increase breeding efficiency. Another key application of the chloroplast genome of *C. humilis* is the identification of commercial cultivars and the determination of their purity. DNA barcodes could be developed from its chloroplast genome and be used to identify varieties and their offsprings.

Acknowledgments

The authors would like to acknowledge Ms. Luting Jia and Ms. Xuanxuan Han who provided invaluable assistance in the samples collection. Dr. Juan Wang provided critical comments on an earlier version of the manuscript.

Author Contributions

Data curation: Pengfei Wang, Junjie Du, Jiancheng Zhang.

Formal analysis: Xiaopeng Mu, Pengfei Wang.

Funding acquisition: Junjie Du.

Investigation: Xiaopeng Mu.

Methodology: Junjie Du, Jiancheng Zhang.

Project administration: Junjie Du.

Resources: Jiancheng Zhang.

Validation: Pengfei Wang, Yu Gary Gao.

Writing – original draft: Xiaopeng Mu.

Writing – review & editing: Junjie Du, Yu Gary Gao.

References

1. Mu XP, Aryal N, Du JM, Du JJ. Oil content and fatty acid composition of the kernels of 31 genotypes of Chinese dwarf cherry [*Cerasus humilis* (Bge.) Sok.]. *Journal of Horticultural Science & Biotechnology*. 2015; 90:525–529. <https://doi.org/10.1080/14620316.2015.11668709>

2. Song XS, Shang ZW, Yin ZP, Ren J, Sun MC, Ma XL. Mechanism of xanthophyll-cycle-mediated photo-protection in *Cerasus humilis* seedlings under water stress and subsequent recovery. *Photosynthetica*. 2011; 49:523–530. <https://doi.org/10.1007/s11099-011-0065-4>
3. Du J, Yang H, Chi J. Distribution and groups of Chinese dwarf cherry (*Cerasus humilis*) in Shanxi province. *Crop Variety Resources*. 1993; 2:6–7.
4. Yin ZP, Li S, Ren J, Song XS. Role of spermidine and spermine in alleviation of drought-induced oxidative stress and photosynthetic inhibition in Chinese dwarf cherry (*Cerasus humilis*) seedlings. *Plant Growth Regulation*. 2014; 74:209–218. <https://doi.org/10.1007/s10725-014-9912-1>
5. Ha XL, Han ZJ, Liang JY. Comparison on photosynthetic and water use characteristics of different varieties (Lines) in *Prunus humilis*. *Northern Horticulture*. 2017; 7–11. <https://doi.org/10.11937/bfy.201702002>
6. Li W, Li O, Zhang A, Li L, Hao J, Jin J, et al. Genotypic diversity of phenolic compounds and antioxidant capacity of Chinese dwarf cherry (*Cerasus humilis* (Bge.) Sok.) in China. *Scientia Horticulturae*. 2014; 175:208–213. <https://doi.org/10.1016/j.scienta.2014.06.015>
7. Leliaert F, Smith DR, Moreau H, Herron MD, Verbruggen H, Delwiche CF, et al. Phylogeny and molecular evolution of the green Algae. *Critical Reviews in Plant Sciences*. 2012; 31:1–46. <https://doi.org/10.1080/07352689.2011.615705>
8. Wolfe KH, Li WH, Sharp PM. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast and nuclear DNA. *Proceedings of the National Academy of Sciences*. 1988; 84:9054–9058.
9. Neuhaus HE, Emes MJ. Nonphotosynthetic metabolism in plastids. *Annual Review of Plant Physiology & Plant Molecular Biology*. 2000; 51:111. <https://doi.org/10.1146/annurev.arplant.51.1.111> PMID: 15012188
10. Zuo LH, Shang AQ, Zhang S, Yu XY, Ren YC, Yang MS, et al. The first complete chloroplast genome sequences of *Ulmus* species by *de novo* sequencing: Genome comparative and taxonomic position analysis. *PLoS ONE*. 2017; 12:e0171264. <https://doi.org/10.1371/journal.pone.0171264> PMID: 28158318
11. Howe CJ. Chloroplast Genome. 2005; eLS. <https://doi.org/10.1038/npg.els.0002016>
12. Park S, Jansen RK, Park S. Complete plastome sequence of *Thalictrum coreanum* (Ranunculaceae) and transfer of the *trp32* gene to the nucleus in the ancestor of the subfamily Thalictrioideae. *BMC Plant Biology*. 2015; 15:40. <https://doi.org/10.1186/s12870-015-0432-6> PMID: 25652741
13. Reith M, and Munholland J. Complete nucleotide sequence of the *Porphyra pupurea* chloroplast. *Plant Molecular Biology Reporter*. 1995; 13:333–335. <https://doi.org/10.1007/BF02669187>
14. Yu DJ. Chinese fruit taxonomy. China agriculture Publishing; 1979.
15. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. *Nucleic Acids Research*. 2017; 45(4):e18. <https://doi.org/10.1093/nar/gkw955> PMID: 28204566
16. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011; 27:2987–2993. <https://doi.org/10.1093/bioinformatics/btr509> PMID: 21903627
17. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324> PMID: 19451168
18. Liu C, Shi LC, Zhu YJ, Chen HM, Zhang JH, Lin XH, et al. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genomics*. 2012; 13: 715. <https://doi.org/10.1186/1471-2164-13-715> PMID: 23256920
19. Lowe T. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. 1997; 25:955–964. <https://doi.org/10.1093/nar/25.5.955> PMID: 9023104
20. Lohse M, Drechsel O, Kahlau S, Bock R. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Research*. 2013; 41:W575–581. <https://doi.org/10.1093/nar/gkt289> PMID: 23609545
21. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: Computational tools for comparative genomics. *Nucleic Acids Research*. 2004; 32:W273–279. <https://doi.org/10.1093/nar/gkh458> PMID: 15215394
22. Leese F, Mayer C, Held C. Isolation of microsatellites from unknown genomes using known genomes as enrichment templates. *Limnology & Oceanography Methods*. 2008; 6(5):412–426.
23. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic Acids Research*. 2002; 30(14):3059–3066. <https://doi.org/10.1093/nar/gkf436> PMID: 12136088

24. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014; 30:1312–1313. <https://doi.org/10.1093/bioinformatics/btu033> PMID: 24451623
25. Curci PL, De Paola D, Danzi D, Vendramin GG, Sonnante G. Complete chloroplast genome of the multi-functional crop globe artichoke and comparison with other Asteraceae. *PLoS ONE*. 2015; 10: e0120589. <https://doi.org/10.1371/journal.pone.0120589> PMID: 25774672
26. Kim KJ, Lee HL. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Research*. 2004; 11:247–261. <https://doi.org/10.1093/dnares/11.4.247> PMID: 15500250
27. Silva SR, Diaz YC, Penha HA, Pinheiro DG, Fernandes CC, Miranda VF, et al. The chloroplast genome of *Utricularia reniformis* sheds light on the evolution of the *ndh* gene complex of terrestrial carnivorous plants from the Lentibulariaceae Family. *PLoS ONE*. 2016; 11:e0165176. <https://doi.org/10.1371/journal.pone.0165176> PMID: 27764252
28. Neubig KM, Whitten WM, Carlsward BS, Blanco MA, Endara L, Williams NH, et al. Phylogenetic utility of *ycf1* in orchids: a plastid gene more variable than *matK*. *Plant Systematics and Evolution*. 2006; 277:77–84. <https://doi.org/10.1007/s00606-008-0105-0>
29. Chang CC, Lin HC, Lin I, Chow TY, Chen HH, Chen WH, et al. The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): Comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. *Molecular Biology & Evolution*. 2006; 23:279. <https://doi.org/10.1093/molbev/msj029> PMID: 16207935
30. Raubeson LA, Jansen RK. Plant diversity and evolution: genotypic and phenotypic variation in higher plants. In: Henry RJ, ed. *Biologist*. Cambridge: CABI Publishing. 2005; p45–78.
31. Doerks T, Copley RR, Schultz J, Ponting CP, Bork P. Systematic identification of novel protein domain families associated with nuclear functions. *Genome Research*. 2002; 12:47–56. <https://doi.org/10.1101/gr.203201> PMID: 11779830
32. Drescher A, Ruf S, Calsa T, Carrer H, Bock R. The two largest chloroplast genome-encoded open reading frames of higher plants are essential genes. *Plant Journal*. 2002; 22:97–104. <https://doi.org/10.1046/j.1365-3113x.2000.00722.x>
33. Kaila T, Chaduvla PK, Saxena S, Bahadur K, Gahukar SJ, Chaudhury A, et al. Chloroplast genome sequence of Pigeonpea (*Cajanus cajan* (L.) Millspaugh) and *Cajanus scarabaeoides* (L.) Thouars: Genome organization and comparison with other Legumes. *Frontier in Plant Science*. 2016; 7:1847. <https://doi.org/10.3389/fpls.2016.01847> PMID: 28018385
34. Raman G, Park S. Analysis of the complete chloroplast genome of a medicinal plant, *Dianthus superbus* var. *longicalyncinus*, from a comparative genomics perspective. *PLoS ONE*. 2015; 10:e0141329. <https://doi.org/10.1371/journal.pone.0141329> PMID: 26513163
35. Williams AV, Boykin LM, Howell KA, Nevill PG, Small I. Correction: The complete sequence of the *Aca-cia ligulata* chloroplast genome reveals a highly divergent *clpP1* gene. *PLoS ONE*. 2015; 10:e0125768. <https://doi.org/10.1371/journal.pone.0125768> PMID: 25955637
36. Lu RS, Li P, Qiu YX. The complete chloroplast genomes of three *Cardiocrinum* (Liliaceae) species: Comparative genomic and phylogenetic analyses. *Front Plant Science*. 2016; 7:2054. <https://doi.org/10.3389/fpls.2016.02054> PMID: 28119727
37. Zhang L, Chen L, Guo M, Song J, Li S, Liu J. Studies on peroxidase isozymes in *cerasus humilis* (Bge.) Sok. *Acta Horticulturae Sinica*. 1994; 1:94–96
38. Daniell H, Lin C, Yu M, Chang W. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biology*. 2016; 17:134. <https://doi.org/10.1186/s13059-016-1004-2> PMID: 27339192