Phyton-International Journal of Experimental Botany
DOI:10.32604/phyton.2020.08831
Article

The Chloroplast Genome of Cerasus Campanulata Diverges from Other Prunoideae Genomes

Yuhao Weng[1], Liming Zhu[2], Yan Ma[3], Hao Su[1], Lu Lu[1], Pengkai Wang[1], Jinhui Chen[1], Asif Ali[1], Renhua Zheng[4,*] and Jisen Shi[1,*]

[1]Key Laboratory of Forest Genetics & Biotechnology of Ministry of Education, Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing, 210037, China
[2]College of Biology and the Environment, Nanjing Forestry University, China Forestry University, Nanjing, 210037, China
[3]Jinling Institute of Technology, Nanjing, 210037, China
[4]Fujian Academy of Forestry, Fuzhou, 350012, China
*Corresponding Authors: Renhua Zheng. Email: zrh08@126.com; Jisen Shi. Email: jshi@njfu.edu.cn
Received: 15 October 2019; Accepted: 07 January 2020

Abstract: *Cerasus Campanulata* is one of several species belonging to the Prunoideae focke, a subfamily of the flowering plant Rosaceae. We investigated the details of its chloroplast genome which may reveal its genus independent of morphological determination. Here, we determined the complete chloroplast (cp) genome sequence of *C. campanulata* and performed sequence analysis to reveal the presence of 18 forward repeats, 20 palindrome repeats, 2 complement repeats, 4 reverse repeats and 93 simple sequence repeats (SSRs). We additionally performed a comparative study of *C. campanulata* and seven other Prunoideae focke species. Then, maximum parsimony (MP) and maximum likelihood (ML) phylogenetic analyses were carried out in the little part of Rosaceae, respectively. The results strongly support a position of *C. campanulata* as a member of the Cerasus in the Rosaceae family. Moreover, the complete cp genome can be used for plant phylogenetic and evolutionary studies that will provide insight into the degree of gene conservation.

Keywords: *Cerasus campanulata*; chloroplast genome sequence; repeat and SSR; phylogenetic analysis

1 Introduction

Cerasus campanulata (Maxim.) A. N. Vassiljeva is a cultivated tree species that is a member of the Cerasus, Prunoideae focke, Rosaceae family. *C. campanulata* is widely grown as an ornamental tree, while in some areas it is also cultivated for its fruit. *C. campanulata* thrives best in a luminous and warm environment, has resistance to high temperature, yet grows preferentially at temperatures from 15°C to 28°C. It originated from secondary broad-leaved forests from the north to north-east part of China, and is widely distributed in Fujian, Guangdong, Guangxi, Jiangxi, Taiwan and in some provinces within Vietnam and Japan. Because of its decorative flowers, some cities or schools have used it as an avenue tree [1]. Individual plants of *C. campanulata* grow either as shrubs or trees, depending on environmental conditions, standing 3 to 8 m tall. Its bark is blackish brown, while its branches are grayish brown to

purplish brown. Its winters buds are ovoid and glabrous, the stipules caducous. The flower petiole is 0.8–1.3 centimeter long and has an apex containing 2 nectaries. The leaf blade may be ovate, ovate-elliptic, or obovate-elliptic in shape; on the abaxial side it is either pale green and glabrous or possesses tufts of hairs in vein axils, while on the adaxial side it is green and glabrous. The leaf has a rounded base, a usually somewhat irregular, serrate margin and an acuminate apex. Its inflorescences are umbellate, precede leaf buds in opening and contain 2–4 flowers with pink petals that darken towards the apex. The flowers contain 39–41 stamens and have a glabrous style that is usually longer than the stamens. *C. campanulate* grows well in loose red and yellow soil, yet poorly in arid barren hills. It is primarily distributed over low altitude areas, possesses a strong adaptability and antipollution effect. China has a large number of wild Cerasus species, such as *Cerasus serrulate* var.sontanea, *Cerasus maximowiczii*, *Cerasus subhirtella* var.ascensens, *Cerasus cerasoides*, *Cerasus craetaegifolia*, *Cerasus setulosa*, *Cerasus yedoensis*, *Cerasus subhirtella* var.pendula etc. However, as opposed to the body of research on Cerasus existing in Japan, the phylogenetic and evolutionary relationship among the undomesticated species widely distributed in China is unclear, making it difficult to perform taxonomic determination and breeding activities. Therefore, to proceed with breeding, it is crucial to study their taxonomy, which can be done based on chloroplast sequence information.

The chloroplast is a plant-specific circular molecule, with a size of 120–160 kb in most higher plants, that has its own mechanism for genetic replication. It contains the entire machinery necessary for the process of photosynthesis and also participates in the biosynthesis of amino acids, nucleotides, lipids and starch [2]. In addition to sustaining life by providing oxygen, reacting and responding to signaling as a metabolic center [3,4], chloroplasts can also be used as a tool to confer valuable agronomic traits and serve as bioreactors for the production of industrial enzymes, biopharmaceuticals, bioproducts, or vaccines [5]. The angiosperm chloroplast genome contains 2 copies of inverted repeats (IRs), separating the large single copy region (LSC) and the small single copy region (SSC) [6]. The gene content of the angiosperm chloroplast genome is highly conserved between species. Because of its slow evolutionary rate, it is convenient enough to perform comparative studies of the chloroplast genome across different species and molecular phylogeny studies. It provides a new way to resolving evolutionary relationships between plants which have variation in their sequence or structure. As a result, it serves to improve our understanding of plant biology and diversity [7].

The chloroplast genome sequence of *C. campanulata* has been published without an accompanying detailed analysis [8]. Its codon preference, repeat structure and differences to other Prunoideae species remain undetermined. We found that a disagreement exists on how *C. campanulata* should be classified, as a Prunus or Cerasus species. Based on a morphological approach, it was described as a Cerasus species in the book "Flora reipublicae popularis subucae." We initially followed the conclusion taken in this book and classified *C. campanulata* as a Cerasus. We then reanalysed the sequence of the *C. campanulata* chloroplast genome in attempt to determine its actual phylogenetic position of *C. campanulata*. The results strongly support a position of *C. campanulata* as a member of the Cerasus in the Rosaceae family.

2 Materials and Methods

2.1 Genome Annotation and Codon Usage

The data of the chloroplast sequence of *C. campanuata* could be downloaded from the NCBI database (BankIt2081270 Cerasus MG827394) and it was annotated using DOGMA [9], available online (http://dogma.ccbb.utexas.edu/). ApE, a freely software (available online: https://jorgensen.biology.utah.edu/wayned/ape/), verified questionable regions, coupled with manually corrected start and stop codons, against other Prunoideae chloroplast genomes (GenBank: KP732472.1, HQ336405.1, NC_030599.1, KP760071.1, NC_023956.1, KP760075.1, KP760073.1, KP760072.1 etc.). The transfer RNA genes were

identified by using DOGMA and tRNAscan-SE2.0 [10]. In addition, GC content and codon usage were analyzed with MEGA7 [11] and CondonW [12].

2.2 Repeat Structure, Genome Comparison and Phylogenetic Analysis

Repeat structure within the *C. campanulata* chloroplast genome was analyzed following the method of Nie et al. [13]. We used the Tandem repeat Finder program [14] with default parameter settings. Disperse repeats, including forward, palindrome, reverse and complement sequences, were identified by using REPuter [15] with hamming distance set to 3 and repeat size to more than 30 bp. To identify simple sequence repeats (SSRs), we used MISA [16] (http://pgrc.ipkgatersleben.de/misa/) with the minimum number of repeats set to 10, 5, 4, 3, 3 and 3 for mono-, di-, tri-, tetra-, penta- and hexanucleotides, respectively [17]. Moreover, we ran the same analysis to detect disperse repeats and SSRs, on the chloroplast genomes of another seven Prunoideae species. Finally, tandem repeats less than 15 bp in length and redundant results between REPuter and SSRs were removed manually. We used mVISTA [18] to compare the full chloroplast genome of *C. campanulata* with other complete Prunoideae chloroplast genomes using the annotation of *C. campanulata*; including *Prunus pseudocerasus*, NC_030599.1; *Prunus maximowiczii*, KP760071.1; *Prunus kansuensis*, NC_023956.1; *Prunus subhirtella*, KP760075.1; *Prunus serrulata*, KP760073.1; *Prunus persica*, HQ336405.1 and *Prunus padus*, KP760072.1. The code behind the name of species is the accession number of the species in the NCBI database (https://www.ncbi.nlm.nih.gov/). We used the MEGA7 program to perform maximum likelihood (ML) and maximum parsimony (MP) phylogenetic analyses, comparing to the other Prunoideae species. For ML analysis the GTR model was used [19], setting gamma distributed of rates among sites and bootstrap analysis to 1,000 replicates. For MP analysis the TBR model was used, setting bootstrap replications to 1,000 [13].

3 Results

3.1 Genome Organization and Features

We started by identifying four parts of whole chloroplast genome. The total nucleotide length of different parts of the chloroplast genome and their GC contents are shown in Tab. 1. The IR regions have a higher GC content. One of the possible reasons is that 8 rRNAs with a high GC content exist in both IR regions, while most repeats in chloroplast genomes are AT rich regions, and a low number of repeats occurs within the IR regions. The complete chloroplast genome of *C. campanulata* contains 115 unique genes, including 79 protein-coding genes, 32 tRNAs and 4 rRNAs. Additionally, 8 protein-coding genes,

Table 1: Nucleotide composition of the *C. campanulata* chloroplast genome

	T(U)%	C%	A%	G%	Length(bp)
LSC	33.6	17.8	31.8	16.8	85929
SSC	34.9	15.8	34.9	14.4	19113
IRA	28.7	20.5	28.8	22	26432
IRB	28.8	22	28.7	20.5	26432
Total	32.1	18.7	31.2	18	157906
CDS	31.5	17.5	30.9	20.1	79899
First position	24	18.5	30.9	26.8	26633
Second position	33	20.1	29.6	17.7	26633
Third position	38	13.9	32.2	15.9	26633

Note: CDS: Protein-coding sequence.

7 tRNAs and 4 rRNAs are duplicated within the IR regions. 17 genes contain introns, of which 14 genes only contain one intron and 3 genes contain two introns. Interestingly, the *rps12* gene is trans-spliced, with the first exon located in the LSC region, while the other two exons are duplicated in both IR regions.

We combined all sequences of protein-coding genes to calculate their length and the type of codons that are represented. T has the highest utilization rate at the second and third codon positions, while A has the highest utilization rate at the first codon position. Among 26,633 codons, the most encoded amino acids are leucine (2,780, 10.5%) and isoleucine (2,289, 8.66%), while the least encoded are cysteine (311, 1.17%) and tryptophan (452, 1.71%). Like most other angiosperm chloroplast genomes [20], the codon usage is biased toward a high representation of T (Fig. 1T).

3.2 Repeat Structure and Sequence Analysis

There are 26 tandem repeats longer than 15 bp in the complete genome sequence. Most of them are within the intergenic space (IGS), excluding 5 repeats in the *ycf1* and *ycf2* genes, respectively. Twenty-one tandem repeats have 2 copies, which accounts for nearly 80.8% of all tandem repeats, 4 tandem repeats have 3 copies and just one tandem repeat has 4 copies. It is remarkable that all of the tandem repeats in *ycf2* have more than 2 copies because 4 out of 5 tandem repeats which have 3 or 4 copies are in the *ycf2*. 43 dispersed repeats were identified, of which 17 were forward repeats, 20 were palindrome repeats, 4 were reverse repeats and 2 were complement repeats. Forward and palindrome repeats are in the majority, accounting for approximately 86% of the total repeat number. Among these dispersed repeats, 30 (69.8%) are 30–34 bp, 9 (20.9%) are 35–39 bp and 4 (9.3%) are 40–44 bp in length. Not a single repeat longer than 42 bp in length has been observed. Most repeats are distributed within the IGS and about 39.5% can be found within genes and introns. SSRs are an important type of repeat, also known as a microsatellite, and can be used as a molecular marker. We identified 93 SSRs in the chloroplast genome of *C. campanulata*, including 66 mononucleotide SSRs (71.0%), 16 dinucleotide SSRs (17.2%) and 11 tetranucleotide SSRs (11.8%), excluding trinucleotide, pentanucleotide and hexanucleotide SSRs. 64 SSRs (68.8%) are located within the IGS, 13 SSRs (14.0%) within introns and 16 SSRs (17.2%) within a CDS, including *matK*, *rpoC2*, *rpoB*, *atpB*, *cemA*, *psbF*, *rps18*, *clpP*, *ndhF*, *ndhE*, *ndhI* and *ycf1*. Moreover, 77 SSRs (82.8%) are located within the LSC, 10 SSRs (10.7%) within the SSC and 6 SSRs (6.5%) within the IRs (Fig. 1B). In general, the vast majority of SSRs belong to the AT type which contribute to AT richness [21].

3.3 Contraction and Expansion of IRs

Contraction or expansion of IRs over time often results in length variation of the chloroplast genome, leading to significant differences between plant species [22]. We compared the IR-LSC and IR-SSC borders, as well as adjacent genes of the chloroplast genomes of *C. campanulata* and another seven Prunoideae species (*P. kansuensis*, *P. maximowiczii*, *P. padus*, *P. persica*, *P. pseudocerasus*, *P. serrulata* and *P. subjirtella*), all of which are published in the NCBI database (Fig. 2). All species contain the long *ycf1* pseudogene that is located at the border between the IR and SSC regions, all with a similar length of approximately 1,050 bp. The *ndhF* and *ycf1* pseudogenes overlapped in all species except in *P. padus* and *P. subhirtella*.

In addition, in a single species, *P. padus*, the IR region has expanded towards the *rps19* gene, truncating the sequence and thereby creating an *rps19* pseudogene at the border between the IR and LSC regions. Furthermore, the distance from the border between the LSC and IRa regions until the start of the *trnH*-GUG gene differs between all eight species we compared. All species vary considering the degree of contraction or expansion of the IR regions, yet the content and length of the IRs are not consistent with the complete size of chloroplast genome.

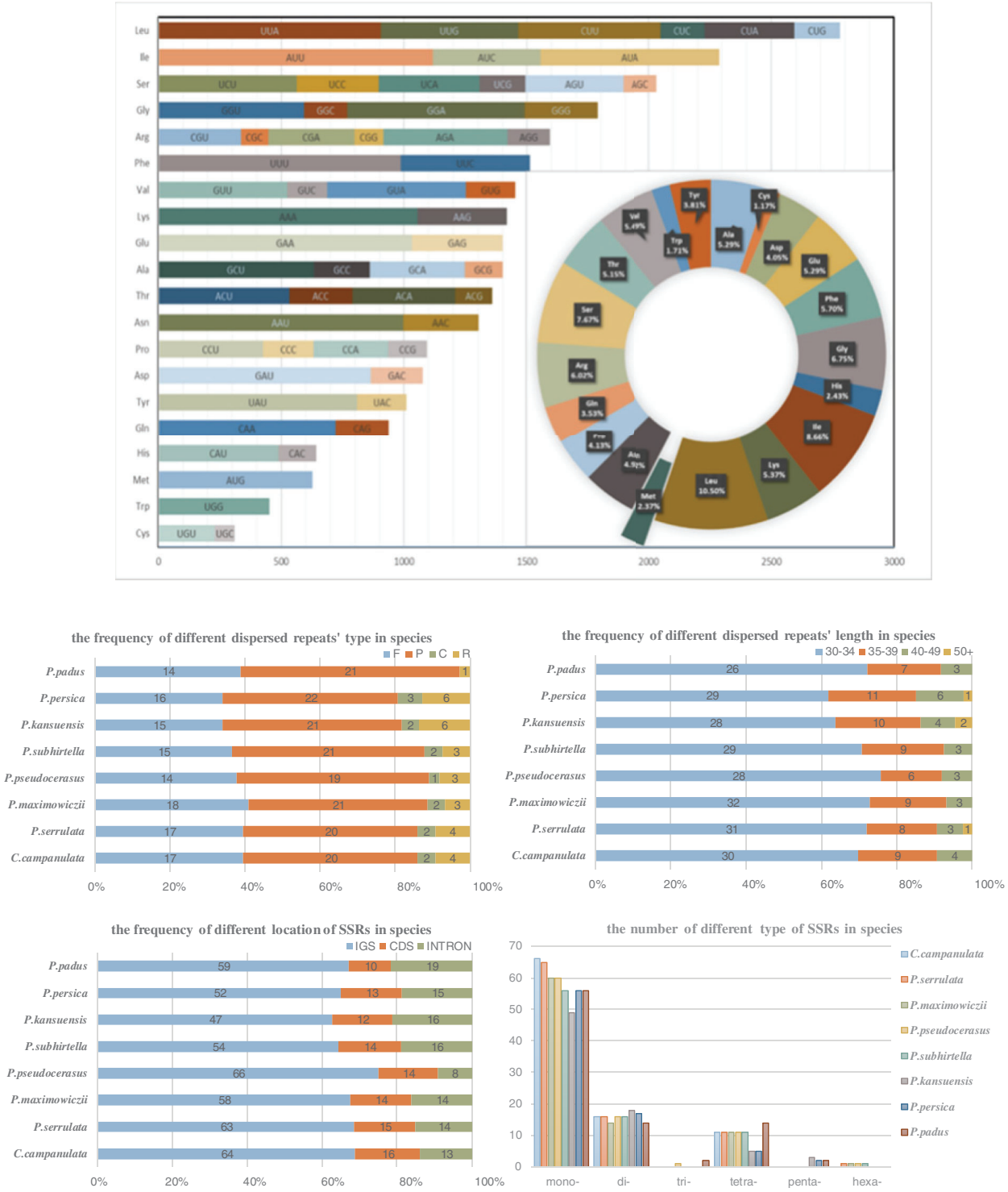

Figure 1: Top: (T) The codon usage and frequency of amino acids in the *C. campanulata* cp protein-coding genes. Bottom: (B) Frequency of identified SSR motifs in different repeat type classes. F in bule represents forward repeats, P in orange represents palindrome repeats, C in green represents complement repeats and R in yellow represents reverse repeats

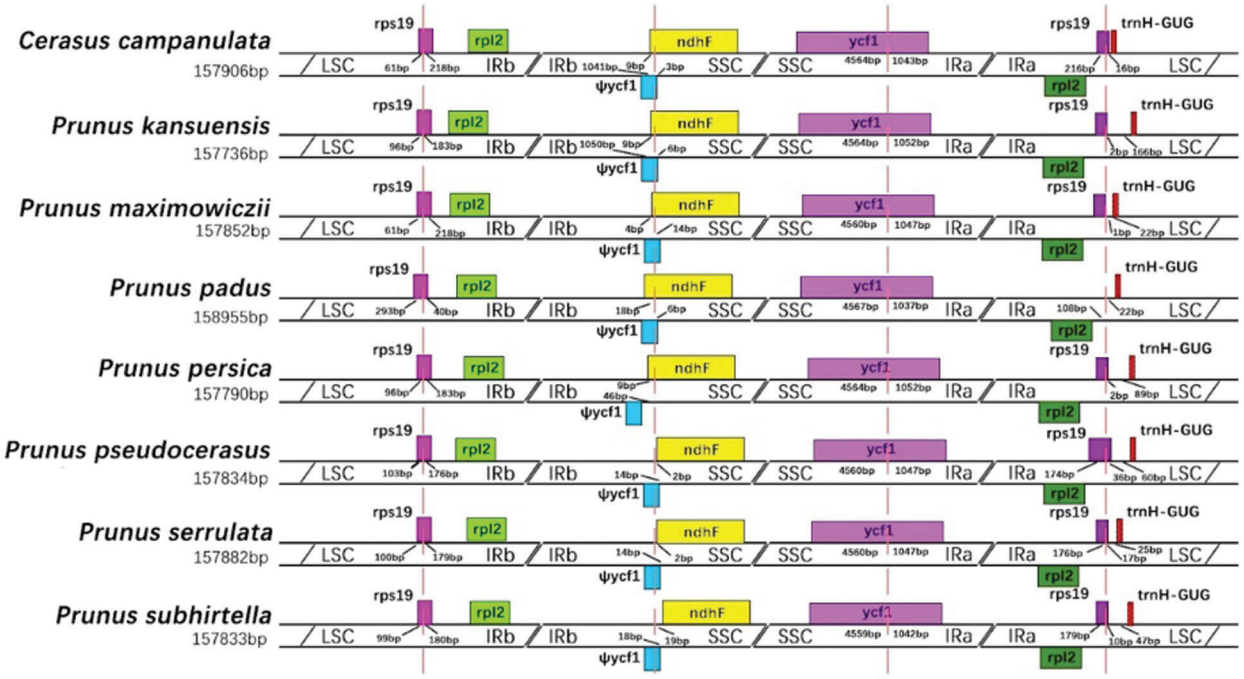

Figure 2: Comparison of the borders of the LSC, SSC and IR regions among eight Prunoideae chloroplast genomes

3.4 Repeat Sequence Comparison between Prunoideae Chloroplast Genomes

We continued to compare the dispersed repeats and SSRs of the other seven Prunoideae species' cp genomes to *C. campanulata*. Our comparison shows that *P. padus*, *P. kansuensis* and *P. persica* are the most divergent of all eight species. *C. campanulata* is the most similar to *P. serrulata*, with which it rarely shows any differences in the number and type of repeats. It is interesting to note that the species can be divided into 3 separate groups, with one of the groups consisting of *P. padus*, another of *P. persica* and *P. kansuensis*, and the last one consisting of the remaining species, corresponding with our phylogenetic analysis. The *C. campanulata* cp genome contains a total of 93 SSRs, more than any of the other species. Considering cp genome size, *C. campanulata* has the second biggest cp genome compared to the other species, just below *P. padus*. However, the relative sizes of its LSC, SSC and IRs regions are fourth, fifth and second, respectively, in comparison (Tab. 2). The sequence identity of all eight Prunoideae chloroplast genome was plotted using mVISTA, with the annotation of *C. campanulata* as a reference (Fig. 3). Results indicate a high degree of sequence conservation between Prunoideae species, yet a few divergent regions can be identified. In general, coding regions are more conserved than non-coding regions, while the LSC and SSC regions are more divergent than the IR regions. Regions containing rRNA genes are most conserved across the complete genome. We found that several stretches of CNS separating gene pairs, such as between *tRNK-UUU* and *atpA*, *rpoB* and *psbD*, *ndhC* and *atpB*, *ndhF* and *ccsA*, as well as *rpl2* and *trnH-GUG*, show a lot of sequence divergence. This might be related to the expression of nearby genes. Our sequence identity analysis suggests that *P. padus* as a single species diverged the most, whereas the remaining seven Prunoideae species can be divided into 2 distinct groups, corresponding with our phylogenetic analysis. Furthermore, *P. serrulata* is highly similar to *C. campanulata* at the sequence level, perhaps due to their similar geographical distribution, resulting in similar diversification. One possible conclusion from these data is that *C. campanulata* and *P. serrulata* both diverged from the same species, yet more experiments are required to confirm this hypothesis.

Table 2: Length of different parts of eight species

Species	Length(bp)			
	Total genome	LSC	SSC	IR
Cerasus campanulata	157906	85929	19113	26432
Prunus kansuensis	157736	85753	19122	26387
Prunus maximowiczii	157852	85846	19133	26436
Prunus padus	158955	87665	18871	26209
Prunus persica	157790	85881	19060	26381
Prunus pseudocerasus	157834	85928	19084	26393
Prunus serrulata	157882	85967	19120	26397
Prunus subhirtella	157833	85950	19120	26381

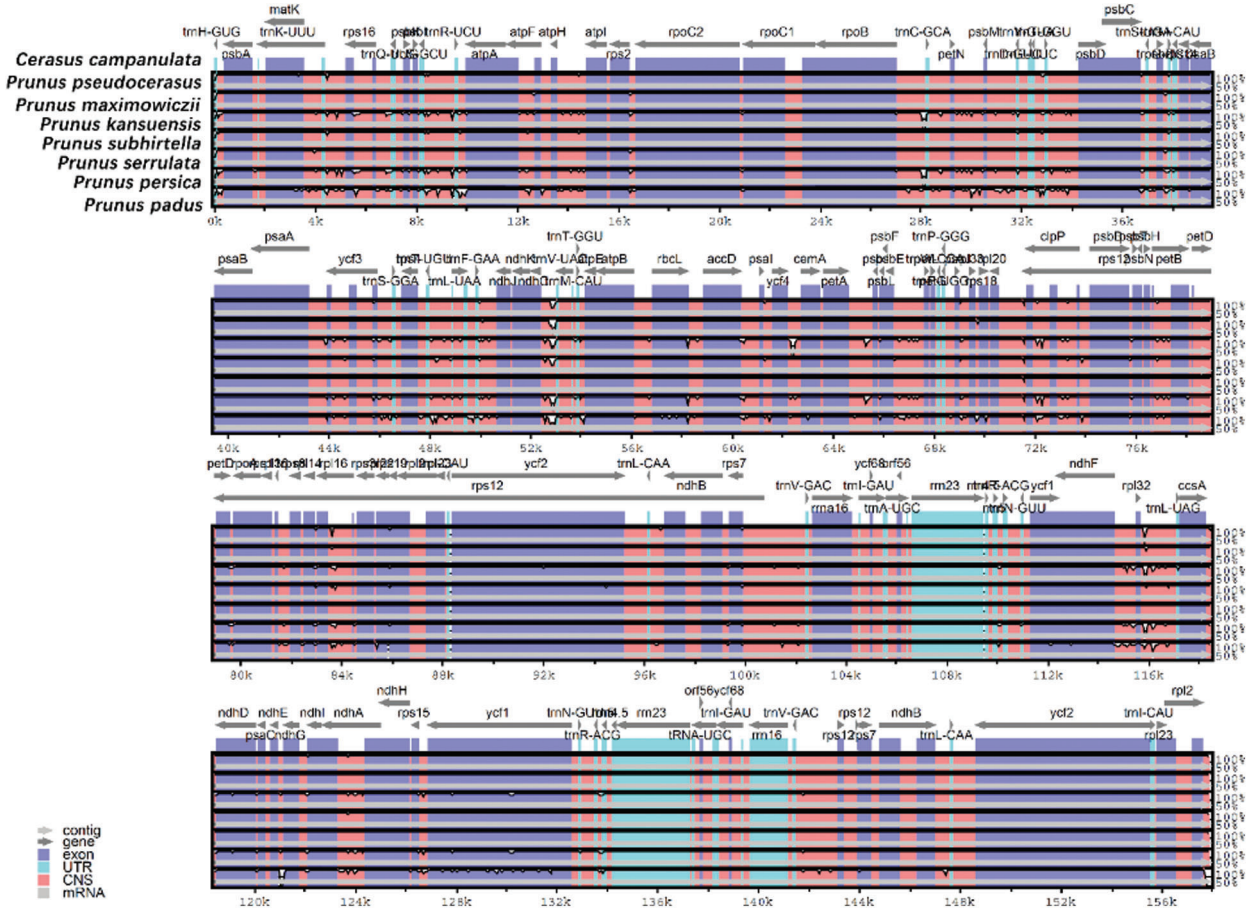

Figure 3: Percent identity plot for comparison of eight Prunoideae chloroplast genomes using mVISTA (see materials and methods). The top line shows genes in order (transcriptional direction indicated with arrow). The *x*-axis represents the coordinate in the chloroplast genome and the *y*-axis shows the average percentage of sequence identity between *C. campanulata* and the other seven species. Genome regions are color coded as exon, intron and conserved non-coding sequences (CNS) (see legend on the bottom left)

3.5 Phylogenetic Analysis

To verify the phylogenetic position of *C. campanulata* amongst the Prunoideae, we constructed Maximum Likelihood (ML) and Maximum Parsimony (MP) trees. In both ML and MP trees (Fig. 4), bootstrap values are high with values of ≥ 95% for most nodes, excluding the *C. campanulata* and *P. serrulata* pait with a value of approximately 72%. Considering the similarity between *C. campanulata* and *P. serrulata*, we approbate the phylogenetic position of *C. campanulata* and the result of both the ML and MP trees. It is divided into three groups that were compared with *C. campanulata* before. We consider that *P. utils* belongs to Prinsepia, *P. padus* belongs to Padus, *P. kansuensis* and *P. persica* belong to Amygdalus and *C. campanulata*, *P. serrulata*, *P. maximowiczii*, *P. pseudocersus*, *P. subhirtella* and *P. yedoensis* belong to Cerasus. Moreover, *C. campanulata* has a very close relationship with *P. serrulata* and the specific relationship and the history of the evolution of them need to be studied further with more samples and related sequences.

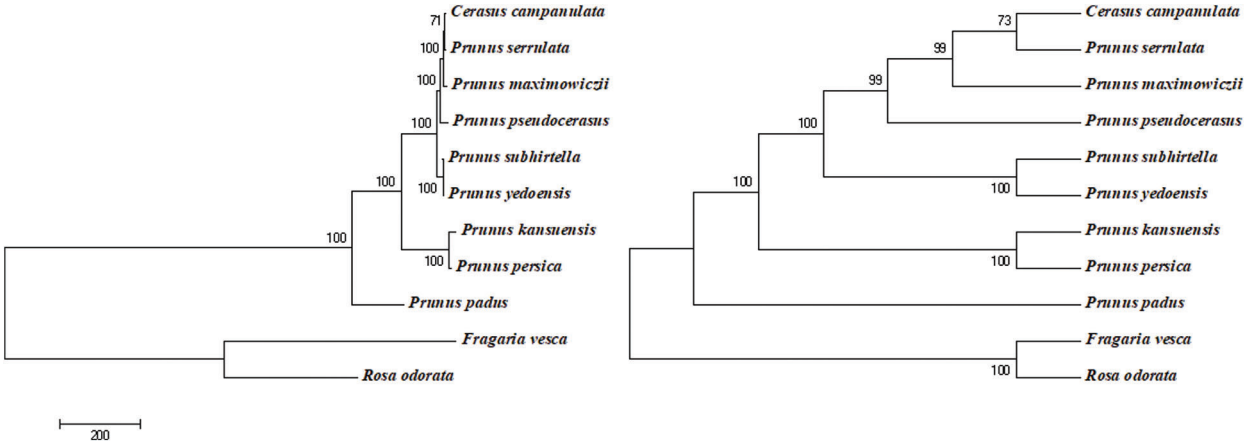

Figure 4: The Maximum Parsimony (MP) phylogenetic tree (left) and Maximum Likelihood (ML) phylogenetic tree (right). Numbers above node are bootstrap support values

4 Discussion

We analyzed the complete chloroplast genome sequence of *C. campanuata*, a member of the Prunoideae, a subfamily within the Rosaceae. The data of the chloroplast sequence of *C. campanuata* could be downloaded from the NCBI database (BankIt2081270 Cerasus MG827394). In this study we analyzed its repeat structure, SSRs, as well as IR contraction and expansion. The genomic organization and gene order of the *C. campanulata* chloroplast genome are generally similar to the previously reported Prunoideae. We made a detailed comparison with other Prunoideae chloroplast genomes, and we found that *C. campanuata* cp genome has a very high similarity to that of *P. serrulate* especially. This result is consistent with previously reported phylogenetic analyses and strongly supports the taxonomic position of *C. campanulate* [8]. Since the chloroplast sequence we used for our analysis originated from NCBI, we leave the species name unchanged. According to the morphological description of Cerasus and Prunus by "Flora reipublicae popularis subucae," as well as our results concerning comparison of chloroplast genome structure and phylogenetic analysis based on chloroplast genome information, we believe that *C. campanuata* should be classified as a Cerasus rather than the previously recognized Prunus. For the same reason, *P. serrulata*, *P. maximowiczii*, *P. pseudocersus*, *P. subhirtella* and *P. yedoensis* should likely be reclassified to the Cerasus family. This could entitle them to a scientific name change into *C. serrulata*, *C. maximowiczii*, *C. pseudocersus*, *C. subhirtella* and *C. yedoensis*. The future release of additional

genome-wide chloroplast data could strengthen our results and increase confidence for determining the exact taxonomic status of *C. campanuata* and the phylogenetic interrelationship of Prunoideae and even of Rosaceae species. However, our study does show that taxonomic determinations based on morphological description alone are not necessarily enough for accurate phylogenetic classification of a species. Molecular phylogeny is a powerful tool to aid such taxonomic efforts.

Funding Statement: This work was supported by key research and development plan of Foundation of Jiangsu Forestry Bureau (LYKJ [2017] 42), Jiangsu Province (BE2017376), the Nature Science Foundation of China (31770715), the Qinglan project of Jiangsu Province, and Priority Academic Program Development of Jiangsu Higher Education Institutions.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. dos Santos, J. M., Sanson, D., da Mata, J. B., Bobrowski, R. (2019). Vigor vegetativo de árvores adultas de *Cerasus campanulata (Maxim.) AN Vassiljeva* sob influência de alterações paisagísticas. *Revista Mundi Meio Ambiente e Agrárias, 4(1),* 82–96.
2. Sugiura, M. (1992). *The chloroplast genome. 10 Years Plant Molecular Biology.* Springer, Germany.
3. Bobik, K., Burch-Smith, T. M. (2015). Chloroplast signaling within, between and beyond cells. *Frontiers in Plant Science, 6*(*1872*)*,* 781. DOI 10.3389/fpls.2015.00781.
4. Daniell, H., Chan, H. T., Pasoreck, E. K. (2016). Pasoreck, Vaccination via chloroplast genetics: affordable protein drugs for the prevention and treatment of inherited or infectious human diseases. *Annual Review of Genetics, 50(1),* 595–618. DOI 10.1146/annurev-genet-120215-035349.
5. Jin, S., Daniell, H. (2015). The engineered chloroplast genome just got smarter. *Trends in Plant Science, 20(10),* 622–640. DOI 10.1016/j.tplants.2015.07.004.
6. Palmer, J. D. (1991). Plastid chromosomes: structure and evolution. *Molecular Biology of Plastids, 7,* 5–53.
7. Daniell, H., Lin, C. S., Yu, M., Chang, W. J. (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biology, 17(1),* 134. DOI 10.1186/s13059-016-1004-2.
8. Cheng, T., Weng, Y., Yang, L., Lu, L., Hao, Z. et al. (2018). The chloroplast genome of *Cerasus campanulata* (Maxim.) AN Vassiljeva. *Mitochondrial DNA Part B, 3(1),* 222–224. DOI 10.1080/23802359.2018.1437799.
9. Wyman, S. K., Jansen, R. K., Boore, J. L. (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics, 20(17),* 3252–3255. DOI 10.1093/bioinformatics/bth352.
10. Lowe, T. M., Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic acids research, 25(5),* 955–964. DOI 10.1093/nar/25.5.955.
11. Kumar, S., Stecher, G., Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution, 33(7),* 1870–1874. DOI 10.1093/molbev/msw054.
12. Peden, J. F. (1999). *Analysis of codon usage*. University of Nottingham: Nottingham, UK.
13. Nie, X., Lv, S., Zhang, Y., Du, X., Wang, L. et al. (2012). Complete chloroplast genome sequence of a major invasive species, Crofton weed (*Ageratina adenophora*). *PLoS One, 7(5),* e36869. DOI 10.1371/journal.pone.0036869.
14. Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research, 27(2),* 573–580. DOI 10.1093/nar/27.2.573.
15. Kurtz, S., Choudhuri, J. V., Ohlebusch, E., Schleiermacher, C., Stoye, J. et al. (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.

16. Thiel, T., Michalek, W., Varshney, R., Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics, 106(3),* 411–422. DOI 10.1007/s00122-002-1031-0.
17. Ng, P. K., Lin, S. M., Lim, P. E., Liu, L. C., Chen, C. M. et al. (2017). Complete chloroplast genome of *Gracilaria firma* (Gracilariaceae, Rhodophyta), with discussion on the use of chloroplast phylogenomics in the subclass Rhodymeniophycidae. *BMC Genomics, 18(1),* 40. DOI 10.1186/s12864-016-3453-0.
18. Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M., Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Research, 32(suppl_2),* W273–W279. DOI 10.1093/nar/gkh458.
19. Tavaré, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. *Lectures on Mathematics in the Life Sciences, 17(2),* 57–86.
20. Tangphatsornruang, S., Sangsrakru, D., Chanprasert, J., Uthaipaisanwong, P., Yoocha, T. et al. (2009). The chloroplast genome sequence of mungbean (*Vigna radiata*) determined by high-throughput pyrosequencing: structural organization and phylogenetic relationships. *DNA Research, 17(1),* 11–22. DOI 10.1093/dnares/dsp025.
21. Kuang, D. Y., Wu, H., Wang, Y. L., Gao, L. M., Zhang, S. Z. et al. (2011). Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): implication for DNA barcoding and population genetics. *Genome, 54(8),* 663–673. DOI 10.1139/g11-026.
22. Chung, H. J., Jung, J. D., Park, H. W., Kim, J. H., Cha, H. W. et al. (2006). The complete chloroplast genome sequences of *Solanum tuberosum* and comparative analysis with Solanaceae species identified the presence of a 241-bp deletion in cultivated potato chloroplast DNA sequence. *Plant Cell Reports, 25(12),* 1369–1379. DOI 10.1007/s00299-006-0196-4.