Characterization and comparative analysis of the complete chloroplast genome sequence from *Prunus avium* 'Summit'

Xueqing Zhao[1,2], Ming Yan[1,2], Yu Ding[1,2], Yan Huo[1,3] and Zhaohe Yuan[1,2]

[1] Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing, Jiangsu, China

[2] College of Forestry, Nanjing Forestry University, Nanjing, Jiangsu, China

[3] College of Landscape Architecture, Nanjing Forestry University, Nanjing, Jiangsu, China

ABSTRACT

Background. Sweet cherry (*Prunus avium*) is one of the most popular of the temperate fruits. Previous studies have demonstrated that there were several haplotypes in the chloroplast genome of sweet cherry cultivars. However, none of chloroplast genome of a sweet cherry cultivar were yet released, and the phylogenetic relationships among *Prunus* based on chloroplast genome data were unclear.

Methods. In this study, we assembled and annotated the complete chloroplast genome of a sweet cherry cultivar *P. avium* 'Summit' from high-throughput sequencing data. Gene Ontology (GO) terms were assigned to classify the function of the annotated genes. Maximum likelihood (ML) trees were constructed to reveal the phylogenetic relationships within *Prunus* species, using LSC (large single-copy) regions, SSC (small single-copy) regions, IR (inverted repeats) regions, CDS (coding sequences), intergenic regions, and whole cp genome datasets, respectively.

Results. The complete plastid genome was 157, 886 bp in length with a typical quadripartite structure of LSC (85,990 bp) and SSC (19,080 bp) regions, separated by a pair of IR regions (26,408 bp). It contained 131 genes, including 86 protein-coding genes, 37 transfer RNA genes and 8 ribosomal RNA genes. A total of 77 genes were assigned to three major GO categories, including molecular function, cellular component and biological process categories. Comparison with other *Prunus* species showed that *P. avium* 'Summit' was quite conserved in gene content and structure. The non-coding regions, *ndhc-trnV*, *rps12-trnV* and *rpl32-trnL* were the most variable sequences between wild Mazzard cherry and 'Summit' cherry. A total of 73 simple sequence repeats (SSRs) were identified in 'Summit' cherry and most of them were mononucleotide repeats. ML phylogenetic tree within *Prunus* species revealed four clades: *Amygdalus*, *Cerasus*, *Padus*, and *Prunus*. The SSC and IR trees were incongruent with results using other cp data partitions. These data provide valuable genetic resources for future research on sweet cherry and *Prunus* species.

Subjects Biotechnology, Evolutionary Studies, Genomics, Molecular Biology, Plant Science

Keywords *Prunus avium*, Chloroplast genome, Genome comparison, Phylogenetic analysis, SSR

Submitted 27 March 2019
Accepted 13 November 2019
Published 20 December 2019

Corresponding author
Zhaohe Yuan,
zhyuan88@hotmail.com

Academic editor
Jessica Kissinger

Additional Information and Declarations can be found on page 14

DOI **10.7717/peerj.8210**

How to cite this article Zhao X, Yan M, Ding Y, Huo Y, Yuan Z. 2019. Characterization and comparative analysis of the complete chloroplast genome sequence from *Prunus avium* 'Summit'. *PeerJ* 7:e8210 http://doi.org/10.7717/peerj.8210

INTRODUCTION

Chloroplast (cp) is generally situated in the cytoplasmic matrix and plays an important role in photosynthesis and fatty acid, starch, and amino acid synthesis (*Wicke et al., 2011*). The cp genome size ranges from 100 kb to 200 kb (*Daniell et al., 2016*). It is a typical quadripartite structure that consists of large single copy (LSC) region, small single copy (SSC) region, and two inverted repeats (IR) regions; It is well known that the cp genome is usually highly conserved in gene structure and content. Several mutations and small structural changes, such as insertions, deletions, reversals, and translocations have been identified in cp genomes. Therefore, the mutational changes in cp genome sequences provide valuable information for phylogenetic, genetic diversity analysis and molecular marker development.

Prunus L., a large and diverse genus, comprises more than 400 species, including most of economically important fruit crops as well as many ornamental species. Due to the parallel evolution of morphological traits, and interspecific hybridization, the botanical classification of the *Prunus* L. has long been controversial and complicated. As early as in 1,700, six subgenera within *Prunus* were recognized based on fruit morphology: *Amygdalus* L., *Armeniaca* Mill., *Cerasus* Mill., *Laurocerasus* Duhamel, *Persica* Mill., and *Prunus sensu stricto* (*Bouhadida et al., 2007*). Afterwards, different opinions, such as a single genus *Prunus* subdivided into seven sections by Bentham and Hooker in 1865, four subgenera within *Prunus* by Koehne in 1911, were also put forward. Currently, the most widely accepted classification of *Prunus* was defined by Rehder in 1940, in which five subgenera *Amygdalus*, *Cerasus*, *Laurocerasus*, *Padus*, and *Prunus* (=*Prundophora*) were divided (*Potter, 2011*).

As the plastid genome provides more accurate proofs to estimate genetic affinities and phylogenetic relationships, several plastomes of *Prunus* plants have been sequenced and reported, such as *P. persica* (*Jansen et al., 2011*), *P. yedoensis* (*Cho et al., 2016*), *P. mume* (*Wang, Gao & Gao, 2016*), *Amygdalus mira* (*Amar et al., 2018*), *P. tomentosa* (*Chen et al., 2018b*), *P. takesimensis* (*Cho, Yang & Kim, 2018*), *P. mongolica* (*Duan et al., 2018*), *P. pedunculata* (*Duan et al., 2018*; *Wang et al., 2018a*), *P. pseudocerasus* (*Feng et al., 2018*), *P. serotina* (*Luan et al., 2018*), *Cerasus humilis* (*Mu et al., 2018*), *P. cerasoides* (*Xu et al., 2018*), *P. davidiana* (*Zhang et al., 2018*), and *P. speciosa* (*Sun, Katsuki & Liu, 2019*).

Sweet cherry (*Prunus avium* L.) is an important *Prunus* fruit in temperate and sub-tropical regions. Traditional intraspecific and interspecific hybridizations have been carried out in this species for genetic improvement by introducing additional desirable characters (*Sansavini & Lugli, 2008*; *Potter, 2011*; *Carrasco et al., 2013*), and all kinds of cultivars were released to meet the market needs of all over the world. Due to natural multiplication and artificial cultivation for so long time, genetic diversity among cultivars and/or populations were verified by researchers (*Frascaria, Santi & Gouyon, 1993*; *Beaver, Iezzoni & Ramm, 1995*; *Lacis et al., 2009*). Because of the conserved properties of chloroplast DNA (cpDNA), many researchers believed that the chances of detecting intraspecific cpDNA variations were low. However, several haplotypes in sweet cherry populations or cultivars were reported previously (*Mohanty, Martín & Aguinagalde, 2001a*; *Mohanty, Martín & Aguinagalde, 2001b*; *Panda et al., 2003*), which provided a great opportunity to study plastome sequence

variation below species level. Recently, the complete chloroplast genome of wild Mazzard cherry (*P. avium*) has been deposited in GenBank (*Chen et al., 2018a*). However, none of chloroplast genome sequences of a sweet cherry cultivar have yet to be released, which was not conducive to cherry haplotypes variation studies.

In this study, we assembled and analyzed the chloroplast genome of a sweet cherry cultivar 'Summit' based on the next-generation sequencing method. Furthermore, we carried out comparative analysis with other *Prunus* species to obtain basic features of cp genomes in *Prunus*. Particularly, general cp genome features and sequence comparison between wild Mazzard cherry and 'Summit' were conducted. Phylogenetic trees were also constructed based on the LSC, SSC, IR, CDS (coding sequences), intergenic regions, and the whole chloroplast sequences to study the relationships in genus *Prunus*. The results might benefit the genetics and breeding of cultivated sweet cherries and related *Prunus* species.

MATERIALS & METHODS

Sampling and DNA extraction

The sample 'Summit' tree was grown in Baima Teaching and Research Base of Nanjing Forestry University, Jiangsu Province, China. The voucher specimen was deposited in Nanjing Forestry University Herbarium (NF0000016). Total genomic DNA was extracted from fresh leaves by a CTAB method (*Li et al., 2013*) with slight modifications. The concentration of DNA was checked by using a Nanodrop ND-2000 spectrometer (Nanodrop Technologies, Wilmington, DE, USA).

Sequencing, assembly, annotation, and Gene Ontology (GO) analysis

A shortgun DNA library was constructed and the subsequent high-throughput sequencing was carried out on the Illumina HiSeq 2500 Sequencing System (Illumina, CA, USA). Raw paired reads were retrieved, trimmed using Fastp 0.20.0 (*Chen et al., 2018b*) to obtain clean data. The de novo assembly of the complete cp genome was performed by NOVOPlasty v3.1 program (*Dierckxsens, Mardulyn & Smits, 2017*). The complete cp genome of *Prunus persica* (HQ336405) was selected as the reference, with *rbcL* as seeds sequence in the analysis. On-line program Geseq (*Tillich et al., 2017*) was used to annotate the cp genome, and the annotation results were inspected by Geneious 8.0.4 software (*Kearse et al., 2012*) while modified manually as needed. We deposited the sequence data into GenBank with the accession number MK622380. A physical map of the genome was obtained by using the online tool OGDRAW (*Lohse, Drechsel & Bock, 2007*). Gene Ontology (GO) annotation was performed by TBtools 1.6 (*Chen et al., 2018a*) to assign GO terms in our genome data.

Genome comparison

The cp genome sequence of 'Summit' cherry and other 12 additional reported *Prunus* species were compared to analyze the basic features of cp genomes in *Prunus* species. In order to show interspecific variation, five cp genome sequences, *P. avium* (MH756631), *P. persica* (HQ336405), *P. tomentosa* (MF624726), *P. padus* (KP760072), and *Malus prunifolia*

(KU851961), were aligned with *P. avium* 'Summit' respectively by mVISTA program (*Mayor et al., 2000*) using Shuffle-LAGAN mode (*Dubchak, 2007*). The IR expansion and contraction of cp genome among six species was visualized by on-line programme IRscope (*Amiryousefi, Hyvönen & Poczai, 2018*).

Simple sequence repeats (SSRs) analysis

Simple sequence repeats (SSRs) in the cp genome of *P. avium* 'Summit' were identified using the MsatCommander 0.8.2 program (*Faircloth, 2008*). The criteria for SSRs identification were 10, 5, 4, 3, 3, 3 repeats units for mono-, di-, tri-, tetra-, penta- and hexa-nucleotides, respectively.

Phylogenetic analysis

Besides *P. avium* 'Summit', an additional 19 *Prunus* species were chosen for phylogenetic analysis, using *M. prunifolia* (KU851961) as an outgroup. The complete cp genome sequences were downloaded from GenBank. Phylogenetic analysis was conducted using the whole genome data, as well as LSC, SSC, IR, CDS, and intergenic regions. The sequences of individual partition regions were aligned using MAFFT v7.308 (*Kazutaka & Standley, 2013*). A maximum likelihood (ML) tree was implemented in IQ-tree v1.6.8 (*Nguyen et al., 2015*) under the best-fitting model TVM + F + R2. We completed a bootstrap analysis with 1000 replicates. Phylogenetic trees were visualized using the FigTree v1.4.3 software.

RESULTS

Characteristics of the chloroplast genome of *P. avium* 'Summit'

In order to facilitate sequence annotation and subsequent analysis accurately, the sequencing low-quality reads were filtered, yielding 8.53 Gb data for *P. avium* 'Summit'. The clean data in the GenBank SRA archive were deposited with the accession number PRJNA579503. After sequence assembly, a circularized molecule of 157,886 bp cpDNA was obtained. The whole cp genome exhibited a typical quadripartite structure resembling to most of land plants, with a pair of IRs of 26,408 bp separated by a LSC region of 85,990 bp and a SSC region of 19,080 bp (Fig. 1, Table 1). There were 131 functional genes annotated in the cp genome, including 86 protein-coding genes, 37 tRNA genes and 8 rRNA genes. The majority of genes occurred as a single copy, while 17 of them duplicated, including six protein-coding species (*rps7*, *rps12*, *rpl2*, *rpl23*, *ndhB*, and *ycf2*), seven tRNA (*trnS-AGA*, *trnL-UAG*,*trnN-GUU*, *trnR-ACG*, *trnA-UGC*, *trnI-GAU*, and *trnV-GAC*), and all four rRNA species (*rrn4.5*, *rrn5*, *rrn16*, and *rrn* 23) (Table 2). Additionally, 13 genes, i.e., *trnA-UGC*, *trnG-UCC*, *trnI-GAU*, *trnK-UUU*, *trnL-UAA*, *trnV-UAC*, *rpoC1*, *rps12*, *rps16*, *rpl2*, *atpF*, *ndhA*, and *ndhB*, contained a single intron, while *ycf3* and *clpP* had two introns (Table 2).

When Gene Ontology (GO) was conducted, only 3 genes (*rps19*, *pbf1*, and *lhbA*) were unable to be annotated. According to the GO result, the most functional groups (15) were identified in *psaA* and *psbA* genes (Table S1). Predicted genes of cp genome were functionally classified according to the three main GO categories including 58 functional groups (Fig. 2, Table S2). Molecular functional categories were strongly represented by

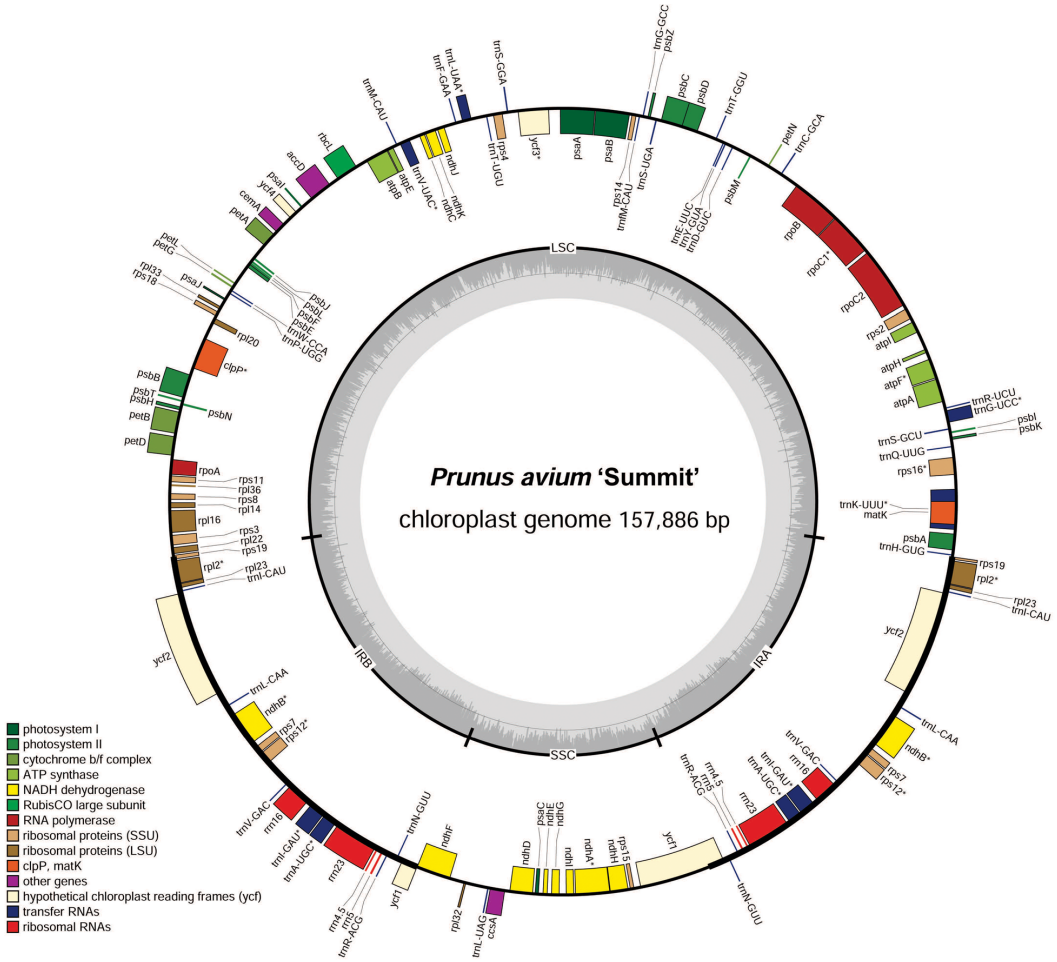

Figure 1 Chloroplast genome map of *P. avium* 'Summit'. Genes inside the circle are transcribed clockwise, and those outside are transcribed counterclockwise. Genes of different functions are color-coded. The darker gray in the inner circle shows the GC content, while the lighter gray shows the AT content.
Full-size DOI: 10.7717/peerj.8210/fig-1

terms related to organic cyclic compound binding (GO:0097159), heterocyclic compound binding (GO:1901363), followed by oxidoreductase activity (GO:0016491). The most common assignments in the cellular component category were membrane-bounded organelle (GO:0043227), intracellular (GO:0005622) and intracellular part (GO:0044424). Genes in the biological process category were primarily sorted into the metabolic process and biosynthetic process.

Comparative analysis of the cp genomes of genus *Prunus*

The complete cp genome sequence of *P. avium* 'Summit' was compared to that of reported *Prunus* species. The results (Table 1) showed that sequenced plastid genomes were similar in terms of organization, gene content, gene order, and GC content. From the aspect of genome size, *P. cerasoides* had the smallest cp genome with the smallest LSC region (85,792 bp), while *P. serotina* had the largest cp genome size with the largest LSC, at 87,289 bp

Table 1 Basic features of cp genomes of reported *Prunus* species.

Species	GenBank Accession No.	Size (kb)	LCS length (kb)	SSC length (kb)	IR length (kb)	Protein	tRNA	Gene	GC%	Reference
P. avium 'Summit'	MK622380	157.886	85.990	19.080	26.408	86	37	131	36.7	
P. avium	MH756631	157.987	85.975	19.121	26.445	82	35	130	35.72	*Chen et al. (2018)*
Cerasus humilis	MF405921	158.084	86.374	19.038	26.336	90	33	131	36.8	*Mu et al. (2018)*
P. serotina	MF374324	158.788	87.289	18.911	26.294	84	37	130	36.6	*Luan et al. (2018)*
P. mongolica	MG602256	158.039	86.173	19.084	26.391	84	37	131	36.8	*Duan et al. (2018)*
P. pedunculata	MG602257	157.851	86.052	19.029	26.385	85	36	131	36.8	*Duan et al. (2018)*, *Wang et al. (2018a)*, *Wang et al. (2018b)*
P. pseudocerasus	KX255667	157.834	86.954	19.084	26.398	86	37	131	36.7	*Feng et al. (2018)*
P. takesimensis	MG754959	157.948	85.959	19.117	26.436	83	37	128	36.7	*Cho, Yang & Kim (2018)*
P. davidiana	MH460864	158.055	86.248	19.047	26.380	86	37	131	36.8	*Zhang et al. (2018)*
P. yedoensis	KP732472	157.786	85.908	19.120	26.379	86	37	131	36.7	*Cho et al. (2016)*
P. mume	NC_023798	157.712	85.861	19.063	26.394	84	37	131	38.9	*Wang, Gao & Gao (2016)*
P. cerasoides	MF621234	157.685	85.792	19.061	26.416	84	37	129	36.7	*Xu et al. (2018)*
P. speciosa	NO Accession No.	157.916	85.927	19.123	26.433	84	37	129	36.7	*Sun, Katsuki & Liu (2019)*
P. persica	HQ336405	157.790	85.968	19.060	26.381	83	37	128	36. 8	*Jansen et al. (2011)*

(Table 1). The maximum (26,445 bp) and minimum (26,294 bp) length of IR regions were found in *P. avium* and *P. serotina*. No substantial differences were found in the sequence lengths of SSC among the *Prunus* species. The genome size variation can be explained mainly by differences in the length of LSC and IR regions. The gene number in cp genome covered the range of 128 to 131. Compared with other *Prunus* species, the GC content of *P. mume* was the highest (38.9%) (Table 1).

We further calculated sequence similarity for six species of cpDNA using mVISTA by aligning the cp genomes with *P. avium* 'Summit' (Fig. 3). Sequence comparison results revealed that the LSC and the SSC regions were more divergent than the IR regions as expected. The highly divergent regions among the six chloroplast genomes mainly occured in the intergenic spacers like *trnH-psbA*, *trnK-rps16*, *rps16-trnQ*, *trnS-trnG*, *trnR-atpA*, *atpH-atpI*, *rpoB-trnC*, *trnC-petN*, *petN-psbM*, *trnT-psbD*, *psbC-trnS*, *psbZ-trnG*, *ycf3-trnS*, *trnF-ndhJ*, *ndhC-trnV*, *psbE-petL*, *ndhF-rpl32*, *rpl32-trnL*, and *ndhG-ndhI*. The sequence similarity between Mazzard cherry and 'Summit' cherry was relatively high, but several non-coding regions, such as *ndhC-trnV*, *rps12-trnV* and *rpl32-trnL*, exhibited divergence.

IR expansion and contraction

The IR-SSC and IR-LSC boundaries, together with the adjacent genes, among the cp genomes of five *Prunus* species and *M. prunifolia* were aligned. From Fig. 4, *P. avium* 'Summit' contained nearly the same IR/SC structure with other congeneric species in which IRb/SC boundaries lay respectively in coding regions of a *rps19* and *ndhF*. For *P. avium* 'Summit', *P. avium*, *P. tomentosa* and *P. padus*, it was found to be 19 bp of *ndhF* extension into IRb while a shorter length of 10 bp extension into IRb in *P. persica*. Similarly, the IRb/ LSC junction was located in the complete *rps19* region in all six species cp genomes

Table 2 List of genes annotated in the cp genome of *P.avium* 'Summit' sequence.

Function	Family	Genes
Self-replication	tRNA genes	*trnA-UGC*[a](2); *trnC-GCA*; *trnD-GUC*; *trnE-UUC*; *trnF-GAA*; *trnG-GCC*; *trnH-GUG*; *trnG-UCC*[a]; *trnI-GAU*[a](2); *trnI-CAU* (2); *trnK-UUU*[a]; *trnL-CAA* (2); *trnL-UAA*[a]; *trnL-UAG*; *trnfM-CAU*; *trnM-CAU*; *trnN-GUU* (2);*trnP-UGG*; *trnQ-UUG*; *trnR-ACG* (2); *trnR-UCU*;*trnS-GCU*; *trnS-GGA*; *trnS-UGA*; *trnT-UGU*; *trnT-GGU*; *trnV-GAC* (2); *trnV-UAC*[a]; *trnW-CCA*; *trnY-GUA*
	rRNA genes	*rrn4.5S* (2); *rrn5S* (2); *rrn16S* (2); *rrn23S* (2)
	DNA-dependent RNA polymerase	*rpoA*; *rpoB*; *rpoC1*[a]; *rpoC2*
	Small subunit of ribosome	*rps2*; *rps3*; *rps4*; *rps7* (2); *rps8*; *rps11*; *rps12*[a](2); *rps14*;*rps15*; *rps16*[a]; *rps18*; *rps19*
	Large subunit of ribosome	*rpl2*[a](2); *rpl14*; *rpl16*; *rpl20*; *rpl22*; *rpl23* (2); *rpl32*; *rpl33*; *rpl36*
Photosynthesis	ATP synthase	*atpA*; *atpB*; *atpE*; *atpF*[a]; *atpH*; *atpI*
	Photosystem I	*psaA*; *psaB*; *psaC*; *psaI*; *psaJ*; *ycf3*[b]; *ycf4*
	Photosystem II	*psbA*; *psbB*; *psbC*; *psbD*; *psbE*; *psbF*; *psbH*; *psbI*; *psbJ*; *psbK*; *psbL*; *psbM*; *psbN*; *psbT*; *psbZ*
	Calvin cycle	*rbcL*
	Cytochrome complex	*petA*;*petB*; *petD*; *petG*; *petL*; *petN*
	NADH dehydrogenase	*ndhA*[a]; *ndhB*[a](2); *ndhC*; *ndhD*; *ndhE*; *ndhF*; *ndhG*; *ndhH*; *ndhI*; *ndhJ*;*ndhK*
Other genes	Others	*ycf1*; *ycf2* (2); *ccsA*; *pbf1*; *clpP*[b]; *cemA*; *accD*; *lhbA*; *matK*

Notes.
[a]Genes containing one intron.
[b]Genes containing two introns.

and extended into the LSC region by different lengths depending on the species, *P. avium* was 93 bp extension into LSC region while 240 bp in *P. padus.* A truncated *rps19* in IRa region was found, and only 1 bp away from the JLA junction in *P. avium*, *P. tometosa*, *P. padus*, and *M. prunifolia*, while 3 bp in *P. avium* 'Summit'. Also, the length of *rps19 of P. padus* in IRa region was only 39 bp, which was much shorter than that in other fiver species (180 bp, 186 bp, 183 bp, 187 bp, 120 bp, respectively).

SSR analysis

In our study, a total of 73 SSRs were identified in the cp genome of *P. avium* 'Summit', most of which were detected in the LSC region (Table 3). Among them, 54 (74.0%) were mononucleotide SSRs and most of them belonged to the A/T type, 13 (17.8%) were di-nucleotide SSRs, five (6.8%) were tetra-nucleotide SSRs, one (1.4%) was a penta-nucleotide SSR, there were no tri-nucleotide and hexa-nucleotide SSRs. Only 24 SSRs were located in genes and the others were in the intergenic regions.

Phylogenetic analysis

Six datasets of 20 *Prunus* cp genome sequences were used to build the phylogenetic tree. When the six phylogenetic trees were compared with each other, we found that the topological structures based on LSC region, CDS region, intergenic region and whole cp

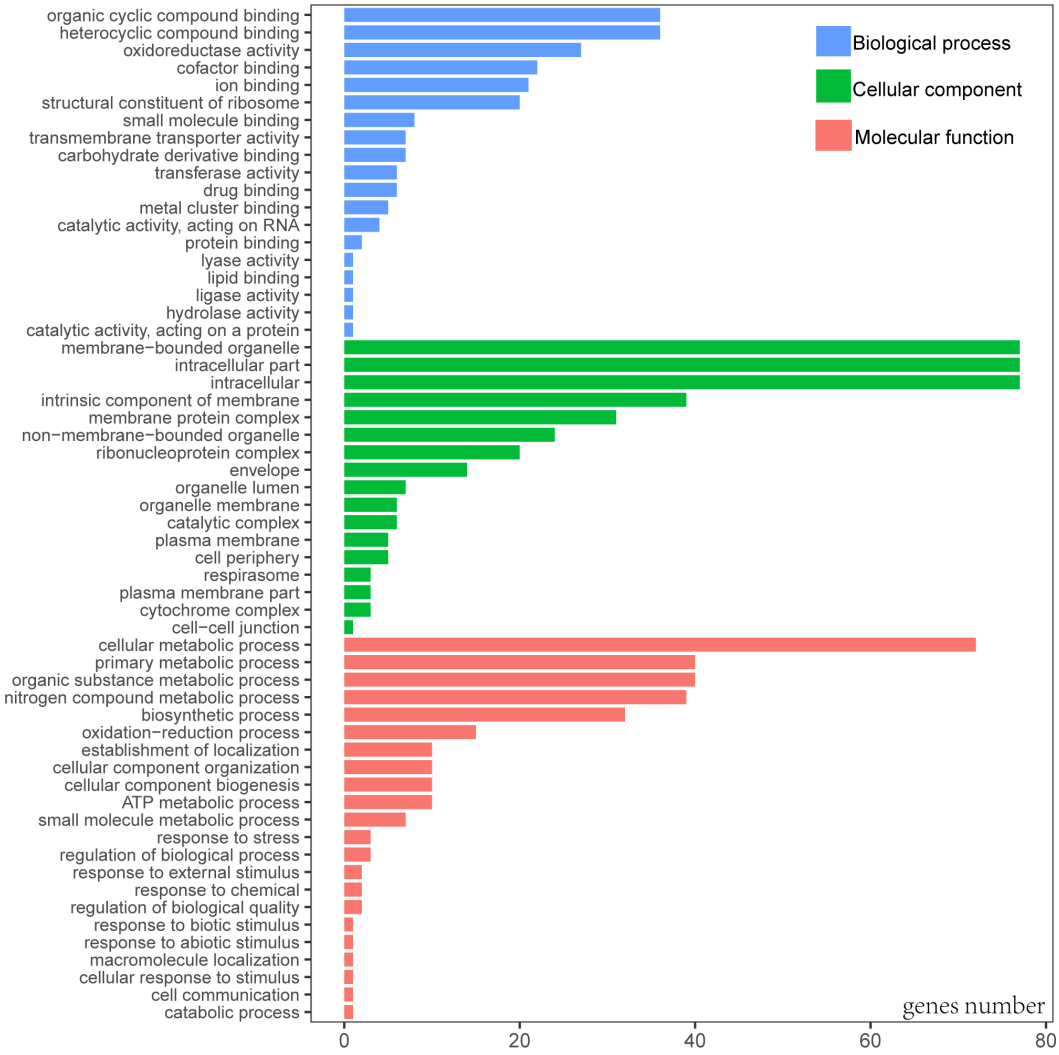

Figure 2 **Gene Ontology (GO) annotation of genes from *P. avim* 'Summit'.**
Full-size DOI: 10.7717/peerj.8210/fig-2

genome datasets were similar (Fig. 5). The four similar phylogenetic trees demonstrated that the monophyly of the genus *Prunus* was well-supported with a high bootstrap value. Four clades corresponding to subgenus *Amygdalus*, *Cerasus*, *Padus*, and *Prunus* were recovered. The ML trees suggested that the subgenus *Padus*, consisting of *P. padus* and *P. serotine*, was a farther lineage from *Amygdalus* and *Prunus* subgenus than *Cerasus*. The *Amygdalus* clade consisted mainly of peaches and almonds, while the *Prunus* clade consisted of plums, apricots and plum blossom. The subgenus *Cerasus* consisted of tree cherry species including cherry blossoms. Our results confirmed that the *P. avium* 'Summit' and Mazzard cherry was a member of *Cerasus* as expected. However, there were some inconsistent phylogenetic relationships among species based on the SSC region and IR region datasets. Based on both datasets, the two phylogenetic trees provided a different position of *P. peduculata*, *P. tomentosa*, *P. davidiana*, and *P. mongolica*.

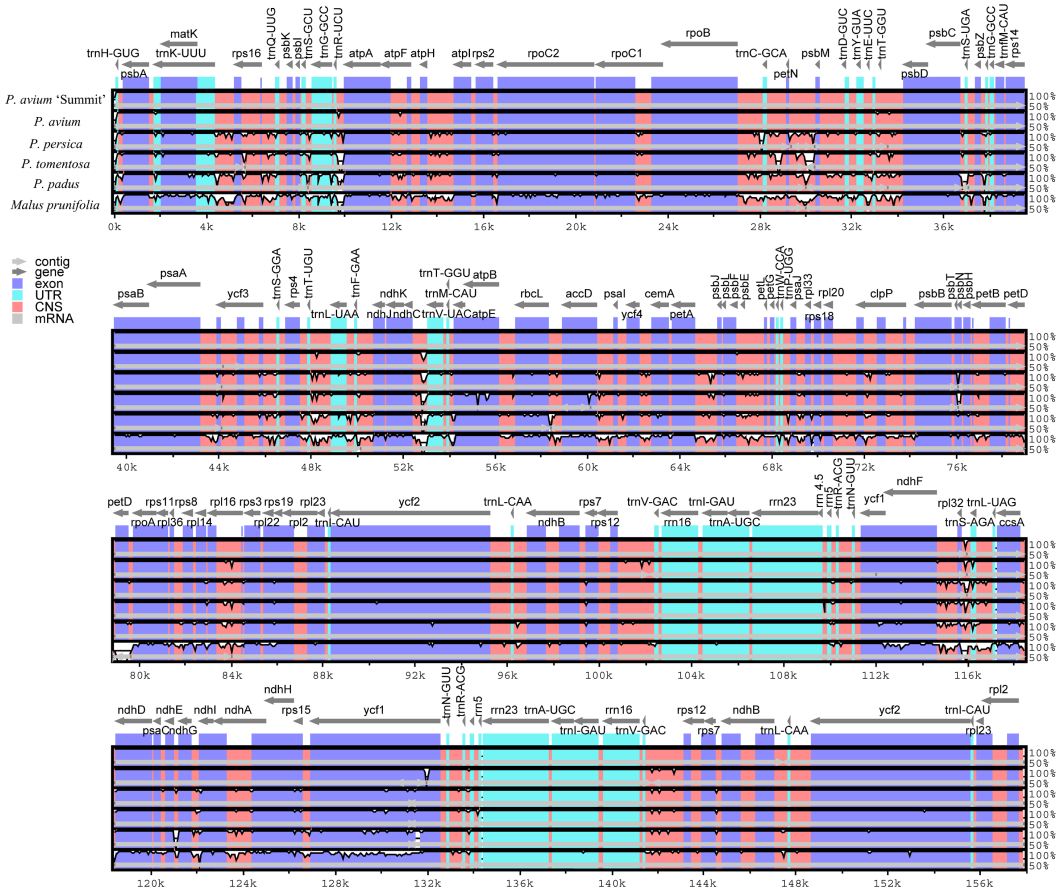

Figure 3 Visualization of genome alignment of the chloroplast genomes of five *Prunus* species and *M. prunifolia* using *P. avium* 'Summit' as reference. Y-scale stands for identity from 50 to 100%.
Full-size DOI: 10.7717/peerj.8210/fig-3

DISSCUSSION

The cp genome normally has a circular structure, and it is composed of a LSC region, a SSC region and two IR regions. From the results, the genome structure, gene order and GC content of *P. avium* 'Summit' were much similar to those reported *Prunus* cp genomes (*Cho et al., 2016*; *Feng et al., 2018*; *Luan et al., 2018*). Through comparative analysis of complete cp genome sequences, much genetic information could be discovered. Our results revealed that the sequence divergence of IR regions was lower than that in LSC and SSC regions, which was also reported in many land plants. In angiosperms cp genomes, the higher divergent intergenic regions, especially the *rpl32-trnL* region, has been used for phylogenetic and evolutionary studies even at the species level (*Dong et al., 2012*; *Zecca et al., 2012*; *Jara-Arancio, Vidal & Arroyo, 2018*). The highly divergent non-coding regions revealed by comparative analysis showed the potentiality for genetic analysis in *Prunus* genus.

Raubeson et al. (2007) pointed out that contraction and expansion at the borders of IR regions were common evolutionary events, and might be the main reason for size diversity

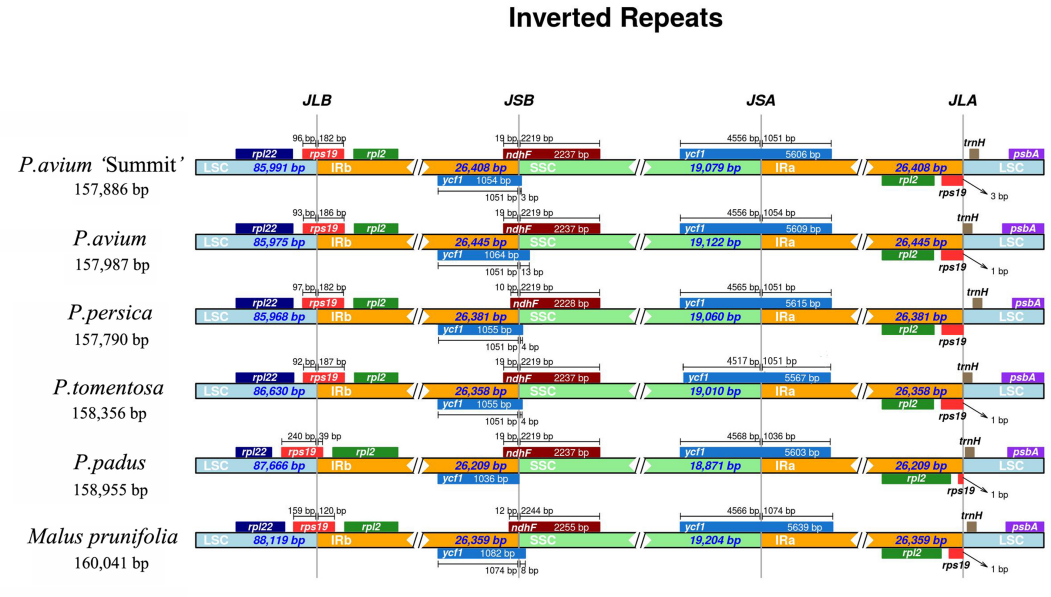

Figure 4 **The comparison of IR boundary among 5 *Prunus* species and *M. prunifolia*.**
Full-size DOI: 10.7717/peerj.8210/fig-4

of cp genomes. The contraction and expansion result in genes at or near the boundaries such as *rps19* and *ycf1* became truncated as incomplete duplications of the normal copy. In most higher plants of chloroplast genomes, *ycf1* was one of the giant ORFs and it usually spaned the boundary of the IR and SSC regions of the plastid genome (*Neubig et al., 2009*), but there were some exceptions. *Chang et al. (2006)* demonstrated that the entire *ycf1* gene in *Phalaenopsis aphrodite* was not across the IR/SSC boundary but within the SSC region. In addition, there were reports on the deletion of *ycf1* gene in IRb/SSC border region in *P. maximowiczii* and *Cerasus humilis* (*Mu et al., 2018*). The function of *ycf1* gene in the evolution of chloroplast genome requires further investigations.

In this study, two *rps19* genes in the IR/SC boundaries were found. In *Dianthus*, there was one copy of the *rps19* gene at the IRb/SSC junction and the other truncated one at IRA/LSC junction a pseudogene (*Raman & Park, 2015*). *Lu, Li & Qiu (2017)* also reported that in three *Cardiocrinum* (Liliaceae) species, the *rps19* gene located in the LSC/IRa boundary apparently lost its protein-coding ability due to partial gene duplication. In our study, pseudogene *rps19* gene located in LSC/IRa boundaries remained to be further elucidated, especially in *P. padus* which a much shorter *rps19* in the LSC/IRa boundary was found.

Further analysis of the cp genomes of wild Mazzard cherry and 'Summit' cherry revealed a relatively conserved structure, though there were some variations in both cp genomes. The contraction and expansion of IR regions resulted in minor variation of *rps19* and *ycf1* extension length in IR/SC boundaries. The sequence variations between Mazzard cherry and 'Summit' cherry were mostly restricted to the non-coding regions, such as *ndhc-trnV*, *rps12-trnV* and *rpl32-trnL*. *Wang et al. (2018a)* reported that the intraspecific variation among four peanut varieties cp genomes was also relatively limited. Owing to the conserved

Table 3 Simple sequence repeats (SSRs) in the *P. avium* 'Summit' cp genome.

Repeat unit	Length (bp)	Number of SSRs	Start position
A	10	6	3801(*trnK-UUU*);5663(*rps16*);16438;67241;79696;114707
	11	3	2964;45743;70816
	12	2	27440;65244
	13	3	69114;70112(*rps18*);110074
	14	2	6853;12395(*atpF*)
	15	1	122057
	16	1	130256(*ycf1*)
	18	2	7702;114619
T	10	14	8308;9211(*trnG-UCC*);14546;26359(*rpoB*);28684;44211(*ycf3*);49746;56081(*atpB*);69383;85316;115469;129653(*ycf1*);130643(*ycf1*)
	11	3	3310(*trnK-UUU*;*matK*);18619(*rpoC2*);122154(*ndhI*)
	12	7	1649;4073(*trnK-UUU*);9601;29278;58797;61300;72286(*clpP*)
	13	3	37286;69086;133792
	14	6	16419;29635;65897;76898(*petB*);84516;123717(*ndhA*)
G	11	1	66458
AT	5	9	6958;13274;19992(*rpoC2*);31097;50125;50370;50381;52952;52968
	6	4	6841;73963;76816(*petB*);115877
AAAT	3	2	5602(*rps16*);72130(*clpP*)
	4	1	1790(*trnK-UUU*)
AATT	3	1	85883
ATTT	3	1	4064(*trnK-UUU*)
AATTT	3	1	32774

properties of cpDNA, cp genome sequence variation was scarcely used below species level. However, the variations in these non-coding regions provides potentials for developing molecular markers in cultivar identification, which has been reported in Fig (*Baraket et al., 2008*) and olive (*Mariotti et al., 2010*).

Nuclear SSRs have been recognized as powerful and advantageous genetic markers due to its abundance in genomes, high degree of polymorphism, and co-dominance. A variety of SSR markers have been applied to the analysis of genetic variability, cultivar identification, parentage assessment, and quality control of rootstock in *P. avium* (*Guarino et al., 2009*; *Lacis et al., 2009*; *Turkoglu et al., 2012*; *DeRogatis et al., 2013*; *Ivanovych & Volkov, 2017*). Additionally, Molecular markers of cpDNA have been successfully used for assessment of genetic diversity in *P. avium* cultivars and populations. The haplotype diversity in sweet cherry populations or cultivars helped to understand the maternal inheritance of chloroplast genome in sweet cherry (*Mohanty, Martín & Aguinagalde, 2001a*; *Mohanty, Martín & Aguinagalde, 2001b*; *Panda et al., 2003*). *Khadivi-Khub et al. (2014)* revealed that intraspecific polymorphism was observed by cpSSR primers in *P. avium* and other related *Prunus* species. This intraspecific polymorphism revealed by cpSSR also had conformity with viewpoints of *Powell et al. (1995)* and *Provan et al. (1997)*. More recently, chloroplast SSRs in *P. salicina* had shown to be highly useful markers for phylogenetic studies in *Prunus*

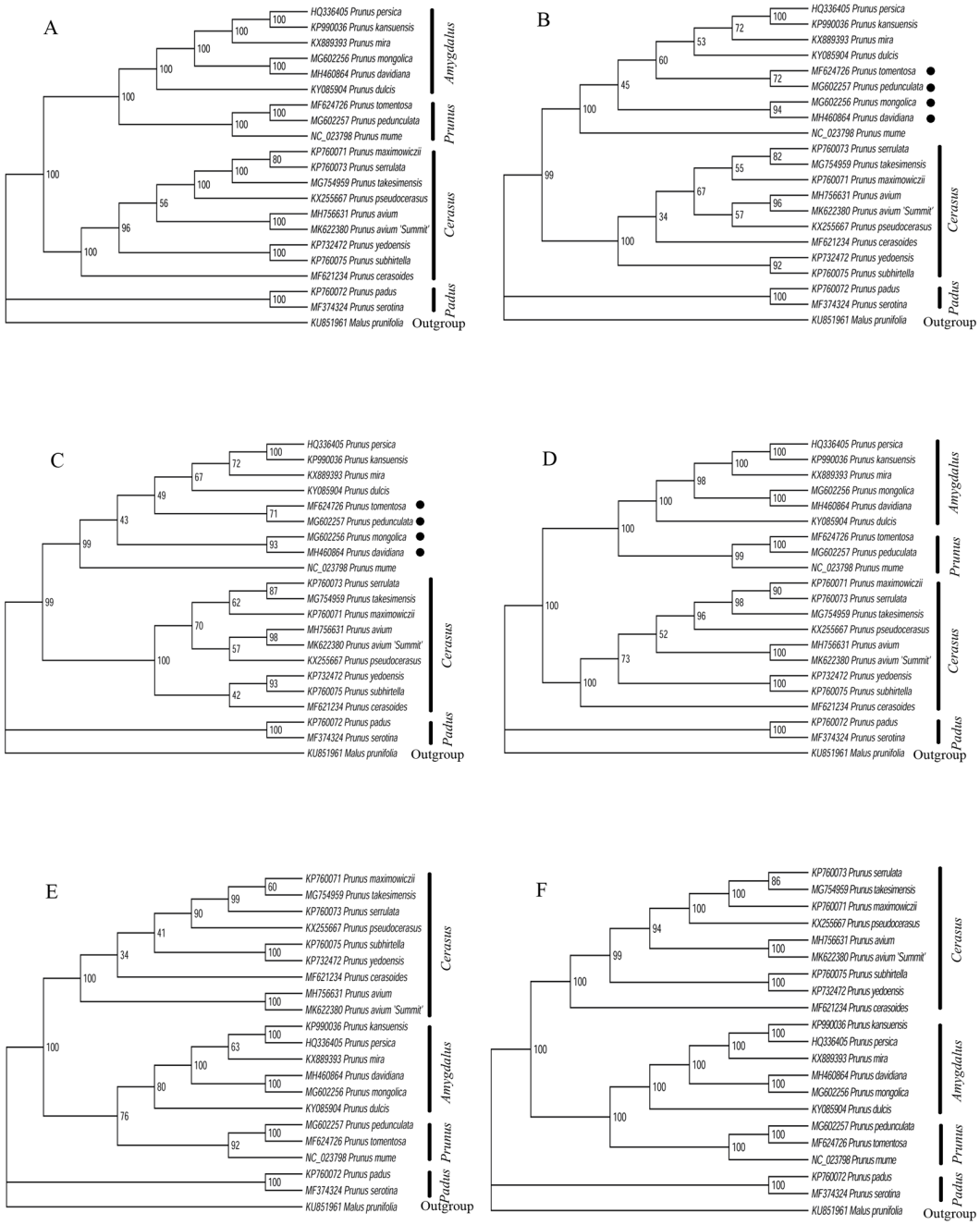

Figure 5 **Phylogenetic relationships of 20 *Prunus* species using Maximum likelihood (ML) analysis by six cp genome partition datasets.** (A) LSC region. (B) SSC region. (C) IR region. (D) CDS region. (E) intergenic region. (F) whole cp genome.

Full-size DOI: 10.7717/peerj.8210/fig-5

genus (*Ohta, Nishitani & Yamamoto, 2005*). Furthermore, *Turkec, Sayar & Heinze (2006)* suggested that cpDNA analysis was a straightforward way to classify cherry cultivars. The cpSSR markers in this study may be further developed for candidate markers to detect

genetic diversity among different cultivars and populations in sweet and sour cherry, which will help breeders select parental genotypes aiding to cherry breeding programmes.

Many studies attempted to construct a phylogenetic framework of *Prunus* from different aspects but interspecies relationships within the genus still remained ambiguous. Therefore, the relationships within *Prunus* species need further investigation. Linnaeus divided the *Prunus* into *Amygdalus*, *Padus*, and *Prunus*, and later recognized four genera: *Armeniaca*, *Cerasus*, *Padus* (including *Laurocerasus*) and *Prunus*. In Shi's analysis, *Amygdalus* and *Prunus* were merged into one subgenus *Prunus*, and three subgenera *Cerasus*, *Prunus* and *Padus* were constructed according to cp regions and nuclear genes data (*Shi et al., 2013*). Our phylogenetic results based on LSC region, CDS region, intergenic region and whole cp genome datasets recognized four subgenera: *Amygdalus*, *Cerasus*, *Padus*, and *Prunus*, which was in accordance with previous phenotype-based classification of Linnaeus in 1754 and Koehne in 1911 (*Lee & Wen, 2001*). Without *Laurocerasus* (laurel-cherries) clade in our results may be the limited cp genome datasets and additional cp genome data would be necessary to test the genetic relationship of *Laurocerasus* within *Prunus*. *P. padus* and *P. serotine* were assigned to subgenus *Padus* which was in line with previous results (*Wen et al., 2008*; *Shi et al., 2013*). The position of subgenus *Amygdalus* as sister to *Prunus* was in accordance with the results of *Yazbek & Oh (2013)* and *Yazbek & Al-Zein (2014)*. However, the monophyletic subgenus *Amygdalus* was contrary to the results of *Lee & Wen (2001)*, who reported *Amygdalus* to be paraphyletic. This difference between mono- and paraphyly of *Amygdalus* may be due to marker or sampling differences. *Bortiri, Heuvel & Potter (2006)* also deemed that molecular data alone do not support the monophyly of subgenus *Amygdalus*. More molecular data combined morphological data are needed to address this question thoroughly.

Two subgenera *Prunus* and *Amgydalus*, especially members of *P. tomentosa*, *P. pedunculata*, *P. mongolica* and *P. davidiana* were intermixed in SSC and IR trees (Fig. 5), which indicated a close tie between *Prunus* and *Amgydalus*. Previous results also demonstrated that subgenera *Prunus* and *Amgydalus* were more closely related to one another than either to subgenus *Cerasus* (*Badenes & Parfitt, 1995*; *Lersten & Horner, 2000*; *Lee & Wen, 2001*; *Wen et al., 2008*). According to Rehder's classification, *P. tomentosa* was classified in Subgenus *Cerasus*. Hybridization studies (*Kataoka, Sugiura & Tomana, 1988*) and isozyme results (*Mowrey & Werner, 1990*), together with cp regions and nuclear genes data, demonstrated that it was closer to subgenus *Prunus* rather than to *Cerasus* (*Bortiri et al., 2001*; *Bortiri et al., 2002*; *Shi et al., 2013*). In addition, *P. pedunculata* was traditionally classified as a member of genus *Amygdalus* (*Lu & Bartholomew, 2003*). However, *Yazbek & Oh (2013)*, *Yazbek & Al-Zein (2014)*, and *Duan et al. (2018)* suggested that *P. pendunculata* should be excluded from subgenus *Amygdalus*, and recovered in subgenus *Prunus*. *P. mongolica and P. davidiana* were closely related to species of peach (*P. persica*), and previous studies based on molecular and morphological analysis all supported the placement of subgenus *Amygdalus* (*Yazbek & Oh, 2013*; *Yazbek & Al-Zein, 2014*). Since previous assertions and results of LSC, CDS, intergenic region and whole cp genome trees in this study did not support the placement of *P. tomentosa*, *P. pedunculata*, *P. mongolica* and *P. davidiana*, thus, we maintained that chloroplast genome datasets, such as LSC,

CDS, intergenic region and whole cp genome could be employed to construct phylogenetic inferences in *Prunus*.

Numerous phylogenetic studies based on the cpDNA sequences have been carried out during the past years. The cp genome approaches together with nuclear and phenotypic data can provide complimentary information for genetic analysis in *Prunus*. Our incongruent phylogenetic relationship results among *Prunus* species illustrated that when phylogenetic analysis were conducted, the plastome data partitions should be prepared with meticulous care.

CONCLUSIONS

The study reported the first complete cp genome of a sweet cherry (*P. avium*) cultivar 'Summit'. Comparison with other *Prunus* species revealed that *P. avium* 'Summit' was quite conserved in structure as well as gene content. The cp SSRs and several intergenic regions compared with other *Prunus* species could be selected to develop into valuable DNA markers in further study. The phylogenetic analysis using SSC region and IR region datasets were not in accordance with the results using other cp data partitions and other published phylogenies. LSC, CDS, intergenic region and whole cp genome datasets could be employed to evaluate phylogenetic relationships in *Prunus*.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the Natural Science Foundation of Jiangsu Province (No. BK20180768), the Initiative Project for Talents of Nanjing Forestry University (No. GXL2014070, GXL2018032), and the Priority Academic Program Development of Jiangsu High Education Institutions (PAPD). 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:
Natural Science Foundation of Jiangsu Province: BK20180768.
Initiative Project for Talents of Nanjing Forestry University: GXL2014070, GXL2018032.
Priority Academic Program Development of Jiangsu High Education Institutions (PAPD).

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Xueqing Zhao performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Ming Yan analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, approved the final draft.

- Yu Ding performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, approved the final draft.
- Yan Huo analyzed the data, prepared figures and/or tables, approved the final draft.
- Zhaohe Yuan conceived and designed the experiments, authored or reviewed drafts of the paper, approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The raw data is available at NCBI: MK622380, and Genbank SRA: PRJNA579503.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.8210#supplemental-information.

REFERENCES

Amar MH, Magdy M, Zhou H, Wang L, Han Y. 2018. The complete chloroplast genome of *Amygdalus mira* (Rosaceae) a threatened wild Chinese peach. *Conservation Genetics Resources* **10(4)**:801–803 DOI 10.1007/s12686-017-0934-7.

Amiryousefi A, Hyvönen J, Poczai P. 2018. IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* **34(17)**:3030–3031 DOI 10.1093/bioinformatics/bty220.

Badenes ML, Parfitt DE. 1995. Phylogenetic relationships of cultivated *Prunus* species from an analysis of chloroplast DNA variation. *Theoretical and Applied Genetics* **90(7–8)**:1035–1041 DOI 10.1007/BF00222918.

Baraket G, Olfa S, Khaled C, Messaoud M, Mohamed M, Mokhtar T, Amel S. 2008. Chloroplast DNA analysis in Tunisian fig cultivars (*Ficus carica* L.): sequence variations of the *trnL-trnF* intergenic spacer. *Biochemical Systematics and Ecology* **36(11)**:828–835 DOI 10.1016/j.bse.2008.09.005.

Beaver JA, Iezzoni AF, Ramm CW. 1995. Isozyme diversity in sour, sweet, and ground cherry. *Theoretical and Applied Genetics* **90(6)**:847–852 DOI 10.1007/BF00222021.

Bortiri E, Heuvel BV, Potter D. 2006. Phylogenetic analysis of morphology in *Prunus* reveals extensive homoplasy. *Plant Systematics and Evolution* **259(1)**:53–71 DOI 10.1007/s00606-006-0427-8.

Bortiri E, Oh S, Gao F, Potter D. 2002. The phylogenetic utility of nucleotide sequences of sorbitol 6-phosphate dehydrogenase in *Prunus* (Rosaceae). *American Journal of Botany* **89(11)**:1697–1708 DOI 10.3732/ajb.89.10.1697.

Bortiri E, Oh S, Jiang J, Baggett A, Granger C, Weeks M, Buckingham D, Potter R, Parfitt DE. 2001. Phylogeny and systematics of *Prunus* (Rosaceae) as determined by sequence analysis of ITS and the chloroplast *trnL-trnF* spacer DNA. *Systematic Botany* **26(4)**:797–807.

Bouhadida M, Martín JP, Eremin G, Pinochet J, Moreno MÁ, Gogorcena Y. 2007. Chloroplast DNA diversity in *Prunus* and its implication on genetic relationships. *Journal of the American Society for Horticultural Science* **132(5)**:670–679 DOI 10.21273/JASHS.132.5.670.

Carrasco B, Meisel L, Gebauer M, Garcia-Gonzales R, Silva H. 2013. Breeding in peach, cherry and plum: from a tissue culture, genetic, transcriptomic and genomic perspective. *Biological Research* **46(3)**:219–230 DOI 10.4067/S0716-97602013000300001.

Chang CC, Lin HC, Lin I, Chow TY, Chen HH, Chen WH, Cheng CH, Lin CY, Liu SM, Chang CC, Chaw SM. 2006. The chloroplast genome of *Phalaenopsis aphrodite* (Orchidaceae): comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. *Molecular Biology and Evolution* **23(2)**:279–291 DOI 10.1093/molbev/msj029.

Chen C, Xia R, Chen H, He Y. 2018a. TBtools, a toolkit for biologists integrating various HTS-data handing tools with a user-friendly interface. *bioRxiv* DOI 10.1101/289660.

Chen S, Zhou Y, Chen Y, Gu J. 2018b. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34(17)**:i884–i890 DOI 10.1093/bioinformatics/bty560.

Chen T, Hu G, Wang Y, Chen Q, Wang L, Zhang J, Tang H, Wang X. 2018a. Characterization of complete chloroplast genome and phylogenetic analysis of sweet cherry *Cerasus avium* (L.) Moench (Prunoideae, Rosaceae). *Mitochondrial DNA Part B* **3(2)**:1274–1275 DOI 10.1080/23802359.2018.1532835.

Chen T, Wang Y, Wang L, Chen Q, Zhang J, Tang H, Wang X. 2018b. The complete chloroplast genome of Tomentosa cherry *Prunus tomentosa* (Prunoideae, Rosaceae). *Mitochondrial DNA Part B* **3(2)**:672–673 DOI 10.1080/23802359.2018.1476068.

Cho MS, Cho CH, Kim SY, Yoon HS, Kim SC. 2016. Complete chloroplast genome of *Prunus yedoensis* Matsum (Rosaceae), wild and endemic flowering cherry on Jeju Island, Korea. *Mitochondrial DNA Part A* **27(5)**:3652–3654 DOI 10.3109/19401736.2015.1079840.

Cho MS, Yang JY, Kim SC. 2018. Complete chloroplast genome of Ulleung Island endemic flowering cherry, *Prunus takesimensis* (Rosaceae), in Korea. *Mitochondrial DNA Part B* **3(1)**:274–275 DOI 10.1080/23802359.2018.1443034.

Daniell H, Lin CS, Yu M, Chang WJ. 2016. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biology* **17(1)**:134 DOI 10.1186/s13059-016-1004-2.

DeRogatis A, Ferrazzini D, Ducci F, Guerri S, Carnevale S, Belletti P. 2013. Genetic variation in Italian wild cherry (*Prunus avium* L.) as characterized by nSSR markers. *Forestry* **86(3)**:391–400 DOI 10.1093/forestry/cpt009.

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/gkw1060.

Dong W, J Liu., Yu J, Wang L, Zhou S. 2012. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLOS ONE* **7(4)**:e35071 DOI 10.1371/journal.pone.0035071.

Duan Y, Shen Y, Kang F, J Wang. 2018. Characterization of the complete chloroplast genomes of the endangered shrub species *Prunus mongolica* and *Prunus pedunculata* (Rosales: rosaceae). *Conservation Genetics Resources* **11(3)**:249–252 DOI 10.1007/s12686-017-0979-7.

Dubchak I. 2007. Comparative analysis and visualization of genomic sequences using VISTA browser and associated computational tools. *Methods in Molecular Biology* **395**:3–16 DOI 10.1007/978-1-59745-514-5_1.

Faircloth BC. 2008. MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources* **8**:92–94 DOI 10.1111/j.1471-8286.2007.01884.x.

Feng Y, Liu T, Wang XY, Li BB, Liang CL, Cai YL. 2018. Characterization of the complete chloroplast genome of the Chinese cherry *Prunus pseudocerasus* (Rosaceae). *Conservation Genetics Resources* **10(1)**:85–88 DOI 10.1007/s12686-017-0770-9.

Frascaria N, Santi F, Gouyon PH. 1993. Genetic differentiation within and among populations of chestnut (*Castanea sativa* Mill.) and wild cherry (*Prunus avium* L.). *Heredity* **70(6)**:634–641 DOI 10.1038/hdy.1993.91.

Guarino C, Santoro S, De Simone L, Cipriani G. 2009. *Prunus avium*: nuclear DNA study in wild populations and sweet cherry cultivars. *Genome* **52(4)**:320–337 DOI 10.1139/G09-007.

Ivanovych Y, Volkov R. 2017. Gnenetic relatedness of sweet cherry (*Prunus avium* L.) cultivars from Ukraine determined by microsatellite markers. *The Journal of Horticultural Science and Biotechnology* **93(1)**:64–72 DOI 10.1080/14620316.2017.1342568.

Jansen RK, Saski C, Lee SB, Hansen AK, Daniell H. 2011. Complete plastid genome sequences of three rosids (*Castanea*, *Prunus*, *Theobroma*): evidence for at least two independent transfers of rpl22 to the nucleus. *Molecular Biology and Evolution* **28(1)**:835–847 DOI 10.1093/molbev/msq261.

Jara-Arancio P, Vidal PM, Arroyo MTK. 2018. Phylogenetic reconstruction of the genus *Triptilion* (Asteraceae, Nassauvieae) based on nuclear and chloroplast DNA sequences. *Journal of Systematics and Evolution* **56(2)**:120–128 DOI 10.1111/jse.12294.

Kataoka I, Sugiura A, Tomana T. 1988. Interspecific hybridization between *Microcerasus* and other *Prunus* spp. *Journal of the Japanese Society for Horticultural Science* **56(4)**:398–407 DOI 10.2503/jjshs.56.398.

Kazutaka K, Standley DM. 2013. MAFFT multiple sequence alignment software Version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30(4)**:772–780 DOI 10.1093/molbev/mst010.

Kearse M, Moir R, Wilson A, Stoneshavas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**:1647–1649 DOI 10.1093/bioinformatics/bts199.

Khadivi-Khub A, Zamani Z, Fattahi R, Powell W. 2014. Genetic variation in wild *Prunus*, L. subgen. *Cerasus*, germplasm from Iran characterized by nuclear and chloroplast SSR markers. *Trees* **28(2)**:471–485 DOI 10.1007/s00468-013-0964-z.

Lacis G, Rashal I, Ruisa S, Trajkovski V, Iezzoni AF. 2009. Assessment of genetic diversity of Latvian and Swedish sweet cherry (*Prunus avium* L.) genetic resources collections by using SSR (microsatellite) markers. *Scientia Horticulturae* **121**:451–457 DOI 10.1016/j.scienta.2009.03.016.

Lee S, Wen J. 2001. A phylogenetic analysis of *Prunus* and the Amygdaloideae (Rosaceae) using ITS sequences of nuclear ribosomal DNA. *American Journal of Botany* **88(1)**:150–160 DOI 10.2307/2657135.

Lersten NR, Horner HT. 2000. Calcium oxalate crystal types and trends in their distribution patterns in leaves of *Prunus* (Rosaceae: Prunoideae). *Plant Systematics and Evolution* **224(1-2)**:83–96 DOI 10.1007/BF00985267.

Li J, Wang S, Yu J, Wang L, Zhou S. 2013. A modified CTAB protocol for plant DNA extraction. *Chinese Bulletin of Botany* **48(1)**:72–78 DOI 10.3724/SP.J.1259.2013.00072.

Lohse M, Drechsel O, Bock R. 2007. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Current Genetics* **52(5-6)**:267–274 DOI 10.1007/s00294-007-0161-y.

Lu LT, Bartholomew B. 2003. Amygdalus. In: *Flora of China, vol 9*. St. Louis: Botanical Garden Press.

Lu RS, Li P, Qiu YX. 2017. The complete chloroplast genomes of three *Cardiocrinum* (Liliaceae) species: comparative genomic and phylogenetic analyses. *Frontiers in Plant Science* **7**:2054.

Luan A, Gao A, He J, Bi G, He Y. 2018. Characterization of the complete chloroplast genome of black cherry (*Prunus serotina* Ehrh.). *Conservation Genetics Resources* **10(3)**:367–370 DOI 10.1007/s12686-017-0826-x.

Mariotti R, Cultrera NGM, Díez CM, Baldoni L, Rubini A. 2010. Identification of new polymorphic regions and differentiation of cultivated olives (*Olea europaea* L.) through plastome sequence comparison. *BMC Plant Biology* **10(1)**:211 DOI 10.1186/1471-2229-10-211.

Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, Pachter LS, Dubchak I. 2000. VISTA: visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* **16(11)**:1046–1047 DOI 10.1093/bioinformatics/16.11.1046.

Mohanty A, Martín JP, Aguinagalde I. 2001a. A population genetic analysis of chloroplast DNA in wild populations of *Prunus avium* L. in Europe. *Heredity* **87(4)**:421–427 DOI 10.1046/j.1365-2540.2001.00922.x.

Mohanty A, Martín JP, Aguinagalde I. 2001b. Chloroplast DNA study in wild populations and some cultivars of *Prunus avium* L. *Theoretical and Applied Genetics* **103(1)**:112–117 DOI 10.1007/s001220000532.

Mowrey BD, Werner DJ. 1990. Phylogenetic relationships among species of *Prunus* as inferred by isozyme markers. *Theoretical and Applied Genetics* **80(1)**:129–133 DOI 10.1007/BF00224026.

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 DOI 10.1371/journal.pone.0196473.

Neubig KM, Whitten WM, Carlsward BS, Blanco MA, Endara L, Williams NH. 2009. Phylogenetic utility of *ycf1* in orchids: a plastid gene more variable than *matK*. *Plant Systematics and Evolution* **277(1–2)**:75–84 DOI 10.1007/s00606-008-0105-0.

Nguyen L, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Molecular Biology and Evolution* **32**:268–274 DOI 10.1093/molbev/msu300.

Ohta S, Nishitani C, Yamamoto T. 2005. Chloroplast microsatellites in *Prunus*, Rosaceae. *Molecular Ecology Notes* **5(4)**:837–840 DOI 10.1111/j.1471-8286.2005.01080.x.

Panda S, Martin JP, Aguinagalde I, Mohanty A. 2003. Chloroplast DNA variation in cultivated and wild *Prunus avium* L: a comparative study. *Plant Breeding* **122(1)**:92–94 DOI 10.1046/j.1439-0523.2003.00768.x.

Potter D. 2011. *Prunus*. In: Kole C, ed. *Wild crop relatives: genomic and breeding resources, temperate fruits*. New York: Springer-Verlag Berlin Heidelberg, 130–136.

Powell W, Morgante M, McDevitt R, Vendramin GG, Rafalski JA. 1995. Polymorphic simple sequencer repeat regions in chloroplast genomes: applications to the population genetics of pines. *Proceedings of the National Academy of Sciences of the United States of America* **92(17)**:7759–7763 DOI 10.1073/pnas.92.17.7759.

Provan J, Corbett G, Mcnicol JW, Powell W. 1997. Chloroplast DNA variability in wild and cultivated rice (*Oryza* spp.) revealed by polymorphic chloroplast simple sequence repeats. *Genome* **40(1)**:104–110 DOI 10.1139/g97-014.

Raman G, Park SJ. 2015. Analysis of the complete chloroplast genome of a medicinal plant, *Dianthus superbus* var. longicalyncinus, from comparative genomics perspective. *PLOS ONE* **10(10)**:e0141329 DOI 10.1371/journal.pone.0141329.

Raubeson LA, Peery R, Chumley TW, Dziubek C, Fourcade HM, Boore JL, Jansen RK. 2007. Comparative chloroplast genomics: analyses including new sequences from the angiosperms *Nupharadvena* and *Ranunculus macranthus*. *BMC Genomics* **8**:174.

Sansavini S, Lugli S. 2008. Sweet cherry breeding programs in Europe and Asia. *Acta Horticulturae* **795**:41–58.

Shi S, Li J, Sun J, Yu J, Zhou S. 2013. Phylogeny and classification of *Prunus sensulato* (Rosaceae). *Journal of Integrative Plant Biology* **55(11)**:1069–1079.

Sun Z, Katsuki T, Liu X. 2019. Complete chloroplast genome of the wild Oshima Cherry (*Prunus speciosa*, Rosaceae) in Izu islands, Japan. *Mitochondrial DNA Part B* **4(1)**:509–510 DOI 10.1080/23802359.2018.1551080.

Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. 2017. GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic Acids Research* **45(W1)**:W6–W11 DOI 10.1093/nar/gkx391.

Turkec A, Sayar M, Heinze B. 2006. Identification of sweet cherry cultivars (*Prunus avium* L.) and analysis of their genetic relationships by chloroplast sequence-characterised amplified regions (cpSCAR). *Genetic Resources and Crop Evolution* **53(8)**:1635–1641 DOI 10.1007/s10722-005-2285-6.

Turkoglu Z, Koc A, Ercisli S, Bilgener S, Akbulut M, Yildirim N, Gercekcioglu R, Esitken A, Gunes M. 2012. Genetic relationships among *Prunus* rootstocks for

sweet cherry (*Prunus avium* L.) cultivars. *Plant Genetic Resources* **10(2)**:101–107 DOI 10.1017/S147926211200007X.

Wang J, Li C, Yan C, Zhao X, Shan S. 2018a. A comparative analysis of the complete chloroplast genome sequences of four peanut botanical varieties. *PeerJ* **6**:e5349 DOI 10.7717/peerj.5349.

Wang S, Gao C, Gao L. 2016. Plastid genome sequence of an ornamental and editable fruit tree of Rosaceae, *Prunus mume*. *Mitochondrial DNA Part A* **27(6)**:4407–4408 DOI 10.3109/19401736.2015.1089546.

Wang W, Wang H, Xiao X, Xu X. 2018b. Characterization of the complete chloroplast genome of longstalk almond (*Prunus pedunculata* (Pall.) Maxim.), an important sand-fixation shrub plant endemic to Northern China. *Conservation Genetics Resources* **11(4)**:419–421 DOI 10.1007/s12686-018-1039-7.

Wen J, Berggren ST, Lee C, Ickert-Bond S, Yi T, Yoo K, Xie L, Shaw J, Potter D. 2008. Phylogenetic inferences in *Prunus* (Rosaceae) using chloroplast *ndh*F and nuclear ribosomal ITS sequences. *Journal of Systematics and Evolution* **46(3)**:322–332.

Wicke S, Schneeweiss GM, depamphilis CW, Müller KF, Quandt D. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology* **76(3-5)**:273–297 DOI 10.1007/s11103-011-9762-4.

Xu X, Wen J, Wang W, Zheng W. 2018. The complete chloroplast genome of the threatened *Prunus cerasoides*, a rare winter blooming cherry in the Himalayan region. *Conservation Genetics Resources* **10(3)**:499–502 DOI 10.1007/s12686-017-0859-1.

Yazbek MM, Al-Zein MS. 2014. Wild almonds gone wild: revisiting Darwin's statement on the origin of peaches. *Genetic Resources and Crop Evolution* DOI 10.1007/s10722-014-0113-6.

Yazbek M, Oh SH. 2013. Peaches and almonds: phylogeny of *Prunus* subg. *Amygdalus* (Rosaceae) based on DNA sequences and morphology. *Plant Systematics and Evolution* **299(8)**:1403–1418 DOI 10.1007/s00606-013-0802-1.

Zecca G, J.Abbott R, Sun W, Spada A, Sala F, Grassi F. 2012. The timing and the mode of evolution of wild grapes (Vitis). *Molecular Phylogenetics and Evolution* **62(2)**:736–747 DOI 10.1016/j.ympev.2011.11.015.

Zhang X, Yan J, Ling Q, Fan L, Zhang M. 2018. Complete chloroplast genome sequence of *Prunus davidiana* (Rosaceae). *Mitochondrial DNA Part B* **3(2)**:888–889 DOI 10.1080/23802359.2018.1501325.