Variation and Evolution of the Whole Chloroplast Genomes of Fragaria spp. (Rosaceae)

Chenxin Li1,2, Chaonan Cai2,3, Yutian Tao3, Zhongshuai Sun2,3, Ming Jiang2, Luxi Chen2 and Junmin Li2,3*

1College of Life Sciences and Medicine, Zhejiang Sci-Tech University, Hangzhou, China, 2Zhejiang Provincial Key Laboratory of Plant Evolutionary Ecology and Conservation, Taizhou University, Taizhou, China, 3School of Advanced Study, Taizhou University, Taizhou, China

Species identification is vital for protecting species diversity and selecting high-quality germplasm resources. Wild Fragaria spp. comprise rich and excellent germplasm resources; however, the variation and evolution of the whole chloroplast (cp) genomes in the genus Fragaria have been ignored. In the present study, 27 complete chloroplast genomes of 11 wild Fragaria species were sequenced using the Illumina platform. Then, the variation among complete cp genomes of Fragaria was analyzed, and phylogenetic relationships were reconstructed from those genome sequences. There was an overall high similarity of sequences, with some divergence. According to analysis with mVISTA, non-coding regions were more variable than coding regions. Inverted repeats (IRs) were observed to contract or expand to different degrees, which resulted in different sizes of cp genomes. Additionally, five variable loci, trnS-trnG, trnR-atpA, trnC-petN, rbcL-accD, and psbE-petL, were identified that could be used to develop DNA barcoding for identification of Fragaria species. Phylogenetic analyses based on the whole cp genomes supported clustering all species into two groups (A and B). Group A species were mainly distributed in western China, while group B contained several species from Europe and Americas. These results support allopolyploid origins of the octoploid species F. chiloensis and F. virginiana and the tetraploid species F. moupinensis and F. tibetica. The complete cp genomes of these Fragaria spp. provide valuable information for selecting high-quality Fragaria germplasm resources in the future.

Keywords: Fragaria, chloroplast genome, comparative analysis, wild species, phylogenetic

INTRODUCTION

The genus Fragaria Linnaeus belongs to the family Rosaceae and is comprised of 25 species, including 13 diploids (2n), five tetraploids (4n), one pentaploid (5n), one hexaploid (6n), three octoploids (8n), and two decaploids (10n; Staudt, 1962, 1989, 2009; Davis et al., 2010; Hummer, 2012; Lei et al., 2017). Most Fragaria species are wild, except for F. × ananassa, which is a cultivated species and an economically important crop (Staudt, 1962; Potter et al., 2000; Cheng et al., 2017). China has been recognized as an important distribution center of wild strawberry resources in the world, as it has 14 wild Fragaria species including nine diploid species and
five tetraploid species (Staudt, 1989, 2006; Lei et al., 2017). Compared with cultivated species, wild species have several advantages, including stronger resistance (Diamanti et al., 2012; Guo et al., 2018), unique fruit aromas (Urrutia et al., 2017), and low nucleotide substitution rate (Palmer, 1985; Wolfe et al., 1987; Jansen et al., 2005; Wei et al., 2006; Wicke et al., 2011; Terakami et al., 2012; Liu et al., 2019). Furthermore, the cp genome is haploid and uniparentally inherited, so it is helpful for tracing source populations and conducting phylogenetic studies to resolve complex evolutionary relationships (Jansen et al., 2007; Parks et al., 2009; Ruhsam et al., 2015). To date, phylogenetic relationships based on complete cp genomes of many angiosperms have been fully studied in many genera (Hu et al., 2016; Zhang et al., 2016; Zhao et al., 2018). Njuguna et al. (2013) systematically studied the phylogenetic relationships of Fragaria based on cp genomes, but there were only 10 sequences assembled from total genomic data, and the coverage of PCR amplified sequences ranged from 60 to 90%, which may mean the assembled chloroplast sequences were incomplete. To the best of our knowledge, there is only one study that has focused on the molecular phylogenetic analysis of Fragaria genus based on whole complete chloroplast genome sequences (Sun et al., 2021), which revealed that 20 Fragaria species were clustered into northern group (eight species), southern group (11 species), and an oldest extant species (one species) based on whole cp genomes. However, this previous study focused on molecular clock analysis and ignored the variation and evolution of whole cp genomes of Fragaria, including, for example, the contraction and expansion of inverted repeats (IRs) regions (Plunkett and Downie, 2000; Kode et al., 2005; Ma et al., 2014; Lei et al., 2016).

In addition, although some Fragaria species can be identified by their morphological characteristics, some species that have similar morphological structures, such as F. mandshurica and F. orientalis (Staudt, 2003; Lei et al., 2017), are likely to be inaccurately identified if morphological indexes are not collected (Yang et al., 2020). Furthermore, Yang et al. (2020) found that most wild Fragaria in Yunnan were difficult to accurately identify without flowers and fruits of collected species. In total, owing to the stage of species development and subjective factors (such as differences in personal knowledge and experience), traditional morphology classification is often inconsistent and unreliable, which can also affect the results of species identification (Yang et al., 2020). Over the years, many molecular analyses have provided insights into the species taxonomy and identification. Currently, DNA barcoding such as rbcL, matK, psbA-trnH and ITS sequences (Potter et al., 2007; Fazekas et al., 2008; Chen et al., 2013; Xin et al., 2013; Tegally et al., 2019; Phi et al., 2020; Islam et al., 2021) has been used for species identification. In Fragaria, the study of DNA barcoding dates back to the phylogenetic analysis conducted by Potter et al. (2000) using ITS and trnL-trnF sequences, but the authors were unable to completely distinguish among species in this genus. Njuguna and Bassil (2011) analyzed psbA-trnH and ITS sequences and reported that the two DNA barcoding sequences are not suitable for species identification in Fragaria. Moreover, more and more studies demonstrated that the four universal barcoding sequences were problematic with low bootstrap support and inability to distinguish between species in land plants (Xin et al., 2013; Tegally et al., 2019; Phi et al., 2020; Islam et al., 2021). Therefore, it is urgent to excavate superior DNA barcoding for special land plants, including Fragaria species, utilizing complete cp genomes, which contain more important variation information for taxonomic and phylogenetic purposes (Huang et al., 2014).

In the present study, we sequenced 27 complete cp genomes of 11 wild Fragaria species from different collection sites in China and downloaded the whole cp genome sequences of seven more Fragaria species. This research had the following objectives: (1) to describe the characteristics of Fragaria cp genomes; (2) to infer the phylogenetic relationships among Fragaria spp.; (3) to detect the variations among Fragaria cp genomes and infer the evolution of the whole cp genomes of Fragaria species; and (4) to provide candidate DNA barcodes for Fragaria species identification. These results provide new insights into the interspecies relationships and evolution of Fragaria spp. as well as basic reference material for the application of Fragaria germplasm resources.

MATERIALS AND METHODS

Plant Material Collection and Genome Data Sources

Twenty-seven individuals belonging to 11 Fragaria species were included in the present study, as summarized in Table 1. All the sampled plants were cultured in the greenhouse facilities of Taizhou University under conditions of 70% relative humidity with temperatures of 25°C in the day and 20°C at night. The plants were identified by Professor Beifen Yang of Taizhou University. The specimens were stored in Zhejiang Provincial Key Laboratory of Plant Evolutionary Ecology and Conservation, Taizhou University, China.

The complete cp genome sequences of F. orientalis (NC_035501), Fragaria chiloensis (NC_019601), and F. virginiana (NC_019602) were downloaded from the National Center for Biotechnology Information (NCBI). Fragaria nipponica (KY769125) and F. iinumae (KC507759) were excluded for there is no supporting of publication reference. Fragaria × amanassia (KY358226) was also downloaded from NCBI to test the accuracy of the sequencing and de novo assembly of cp genome.

Raw sequence data from three species, including F. gracilis (BOP214815), F. tibetica (BOP214818), and F. moschata (BOP214819; Sun et al., 2021), were downloaded from NCBI
and de novo assembly was performed according to the following processes. The other seven species listed by Sun et al. (2021) were excluded for duplication of our own sequencing species.

### Genome Sequencing and Assembly

Fresh and clean leaves of each sampled species were collected and frozen in liquid nitrogen immediately. The samples were used to extract the total DNA by the modified CTAB method (Doyle and Doyle, 1987). Then, paired-end sequencing (Insert size: 350 bp) was performed using the Illumina Novaseq 6000 platform (Illumina, San Diego, CA, United States). Raw reads were filtered to obtain high-quality clean data. Then, de novo assembly was performed with the GetOrganelle software package (Jin et al., 2020).

### Chloroplast Genome Annotation

Thirty cp genomes were annotated using the online GeSeq tool (Tyagi et al., 2020)\(^1\) with default parameters and *F. chiloensis* (NC_019601) was used as the reference to predict protein-coding genes (PCGs), transfer RNA (tRNA) genes, and ribosomal RNA (rRNA) genes. Then, the cp genome sequences were manually modified using Geneious Prime 2021.1.1 (Biomatters Ltd., Auckland, New Zealand). A circular diagram of the cp genome sequences was generated using the online OrganellarGenome DRAW tool (OGDRAW; Lohse et al., 2013; Bai et al., 2017).

### Comparative Genome Analyses

The boundaries of the large single-copy region (LSC), short single-copy region (SSC), and IRs of 12 newly assembled complete cp genomes of *Fragaria* species (*F. nilgerrensis*\(_1\), *F. mandshurica*_JL, *F. corymbosa*\(_1\), *F. moupinensis*\(_XZ\), *F. pentaphylla*\(_1\), *F. viridis*, *F. moschata* (BOP214815), *F. virginiana* (BOP214818), *F. orientalis* (NC_035501), *F. chiloensis* (NC_019601), and *F. virginiana* (NC_019602) from NCBI were compared using IReScope software (Ali et al., 2018). The level of divergence among the 18 sequences was visualized with Shuffle-LAGAN mode (Cheng et al., 2017; Tyagi et al., 2020) in mVISTA software (Kawabe et al., 2018; Jeon and Kim, 2019) with the default settings, and *F. gracilis* was used as a reference sequence.

### Hypervariable Site Identification

All 18 sequences were aligned using the Geneious prime 2021.1.1 plugin MAFFT v.7.450 (Katoh and Standley, 2013), and the alignment was manually adjusted. Then, we performed a sliding window analysis using DnaSP v6 software (Rozas et al., 2017; Jeon and Kim, 2019) to analyze nucleotide diversity (\(\pi\)) in order to detect hypervariable sites among *Fragaria* cp genomes. The window length was set to 600bp, and the step size was set to 200bp (Sun et al., 2021).

---

\(^1\)https://chlorobox.mpimp-golm.mpg.de/geseq.html

---

**TABLE 1 | *Fragaria* species collection information.**

| Voucher      | Ploidy | Locality     | Latitude (N) | Longitude (E) | Altitude (m) | Genbank accession |
|--------------|--------|--------------|--------------|---------------|--------------|-------------------|
| *F. nilgerrensis*\(_1\) | 2      | Yunnan, China | 27.01°       | 100.12°       | 3,403.50     | MZ851761          |
| *F. nilgerrensis*\(_2\) | 2      | Yunnan, China | 25.32°       | 100.13°       | 2,159.26     | MZ851762          |
| *F. mandshurica* JL | 2      | Jilin, China  | 42.23°       | 128.17°       | 1,047.00     | MZ851758          |
| *F. mandshurica* HLJ | 2      | Heilongjiang, China | 52.41° | 125.14° | 350.00 | MZ851757 |
| *F. corymbosa* JL | 4      | Jilin, China  | 42.09°       | 128.00°       | 1,530.00     | MZ851750          |
| *F. corymbosa*\(_XZ\) | 4      | Tibet, China  | 29.61°       | 94.70°        | 4,082.81     | MZ851751          |
| *F. corymbosa* GS | 4      | Gansu, China  | 35.77°       | 103.96°       | 2,833.09     | MZ851749          |
| *F. moupinensis*\(_XZ\) | 4      | Tibet, China  | 29.76°       | 94.73°        | 3,381.00     | MZ851760          |
| *F. moupinensis* SC | 4      | Sichuan, China | 29.21° | 94.22° | 3,062.85 | MZ851759 |
| *F. pentaphylla*\(_1\) | 2      | Qinghai, China | 36.65° | 101.48° | 2,999.77 | MZ851764 |
| *F. pentaphylla*\(_2\) | 2      | Qinghai, China | 36.98° | 102.43° | 2,327.00 | MZ851765 |
| *F. pentaphylla*\(_3\) | 2      | Tibet, China  | 28.07°       | 86.00°        | 3,261.00     | MZ851766          |
| *F. pentaphylla*\(_4\) | 2      | Gansu, China  | 35.83°       | 104.12°       | 1,966.78     | MZ851767          |
| *F. nubicola*\(_2\) | 2      | Tibet, China  | 28.06°       | 85.99°        | 3,353.00     | MZ851768          |
| *F. daltoniana*\(_1\) | 2      | Tibet, China  | 28.07°       | 86.00°        | 3,261.00     | MZ851762          |
| *F. daltoniana*\(_2\) | 2      | Tibet, China  | 28.07°       | 86.00°        | 3,261.00     | MZ851752          |
| *F. daltoniana*\(_3\) | 2      | Tibet, China  | 28.02°       | 85.98°        | 2,724.00     | MZ851754          |
| *F. daltoniana*\(_4\) | 2      | Tibet, China  | 28.03°       | 85.98°        | 2,956.00     | MZ851755          |
| *F. daltoniana*\(_5\) | 2      | Tibet, China  | 28.03°       | 85.98°        | 2,956.00     | MZ851756          |
| *F. viridis*\(_2\) | 2      | Segovia, Spain | 41.42° | −3.76° | 1,170.00 | MZ851772 |
| *F. vesca* ssp. bracteata | 2      | California, United States | 38.77° | −120.45° | 1,044.00 | MZ851768 |
| *F. vesca* ssp. vesca\(_1\) | 2      | Valcea, Romania | 46.43° | 23.76° | 355.61 | MZ851769 |
| *F. vesca* ssp. vesca\(_2\) | 2      | Sichuan, China | 30.04° | 101.82° | 3,968.00 | MZ851770 |
| *F. vesca* ssp. vesca\(_3\) | 2      | Xinjiang, China | 43.87° | 85.38° | 1,468.00 | MZ851771 |
| *F. chinensis*\(_1\) | 2      | Shaanxi, China | 33.27° | 108.30° | 1,186.95 | MZ851747 |
| *F. chinensis*\(_2\) | 2      | Shaanxi, China | 33.27° | 108.30° | 1,186.95 | MZ851748 |
| *F. × ananassa* | 8      | Taizhou, China | 28.66° | 121.39° | 22.69 | MZ851773 |
**Phylogenetic Analysis**

Thirty-four complete cp genome sequences were used to reconstruct the phylogenetic relationships among Fragaria spp. based on maximum likelihood (ML) using RAxML 8.2.10 (Stamatakis, 2014), with 1,000 bootstrap replicates employed for estimating node support. *Potentilla fruticosa* (NC_036423) and *Drymocallis saviczii* (NC_050966) were downloaded from Genbank and used as outgroups (Eriksson et al., 1998; 2003; Potter et al., 2000; Feng et al., 2017; Dimeglio et al., 2014). In total, 36 cp genome sequences were aligned using a Geneious prime 2021.1.1 plugin MAFFT v.7.450 (Katoh and Standley, 2013), and the alignment was manually adjusted when necessary. Modeltest 3.7 (Liu et al., 2013) was used to select the best-fit evolutionary model of Fragaria cp genome sequence evolution for reconstruction of the phylogenetic relationships of Fragaria.

**RESULTS**

**General Chloroplast Genome Characteristics**

The total genome sequence lengths of Fragaria species ranged from 155,479 bp (F. viridis) to 155,832 bp (F. daltioniana_3). The cp genomes presented a typical quadripartite structure including a pair of IR regions with lengths of 51,872 bp (F. corymbosa_JL), separated by a LSC region from 85,471 bp (F. viridis) to 85,726 bp (F. daltioniana_3) and a SSC region from 18,116 bp (F. viridis) to 18,219 bp (F. moupinensis_ZX). The GC contents ranged from 37.2 to 37.3%. Overall, the cp genome of Fragaria encodes a total of 130 genes, including 85 PCGs, 37 tRNA genes, and eight rRNA genes (Figure 1; Table 2). Among these genes, 15 contained a single intron (ndhA, ndhB, petB, petD, rpl2, rpl16, rpoC1, rps12, rps16, trnA-UGC, trnG-GCC, trnI-CAU, trnK-UUU, trnL-UAA, and trnV-UAC), while two harbored two introns (ycf3 and clpP). The trnK-UUU gene had the longest intron (2,492–2,496 bp), which contained the matK gene, whereas trnL-UAA was smallest (420 bp; Table 3).

**Comparative Analyses**

The detailed comparisons of LSC, SSC, and IR boundaries in 19 Fragaria complete cp genomes are shown in Figure 2, revealing differences at boundary regions. There were five genes around borders of these regions, including rps19, rpl2, ycf1, ndhF, and trnH, in which rpl2 and ycf1 had two copies. Overall, F. moschata, F. mandshurica_JL, F. viridis, F. vesca ssp. vesca_1, F. vesca ssp. bracteata, and F. ×ananassa showed higher structural similarity, with rps19, rpl2, and trnH genes found at the same distances from the boundaries. The ycf1 gene in Fragaria species spanned 1,092 bp from the SSC to IRb region, resulting in a non-functional ycf1 fragment gene with the same length in IRa. Additionally, ycf1 extended to the SSC region, with different distances ranging from 3 bp (F. orientalis) to 37 bp (F. daltioniana_1), with the cp genomes of F. mandshurica_JL, F. viridis, F. vesca ssp. vesca_1, F. vesca ssp. bracteata, F. ×ananassa, F. chiloensis, and F. virginiana all showing the same gap of 13 bp. Moreover, ycf1 and ndhF had more length polymorphisms at the IRA/SSC border (Figure 2).

**Phylogenetic Analysis**

The GTR+I+G model was selected as the best-fit evolutionary model by using Modeltest 3.7. The phylogenetic tree was constructed using 34 complete cp sequences of Fragaria species, with *P. fruticosa* and *D. saviczii* as outgroups. As shown in Figure 5, Fragaria species can be clustered into two groups, A and B, with 100% bootstrap support. Group A included two subgroups, A1 (F. chilenis and F. daltioniana) and A2. Subgroup A2 contained six Fragaria species (F. nubicola, F. pentaphylla, F. corymbosa, F. moupinensis, F. gracilis and F. tibetica) that are mainly distributed in western China. Group B was composed of the remaining species, which included F. nilgerrensis, F. mandshurica, F. viridis, F. orientalis, F. moschata, F. ×ananassa, F. chiloensis, F. virginiana, F. vesca ssp. vesca and F. vesca ssp. bracteata. These latter species are from Europe and America, apart from F. nilgerrensis, which is mainly distributed in southeast Asia, F. mandshurica, which is distributed in northeast China, and F. orientalis, which is mainly found in northern Asia.

**DISCUSSION**

**Variations and Evolution of Whole Cp Genomes of Fragaria spp.**

In this study, 27 cp genomes of Fragaria species were sequenced and found to range in size from 155,479 to 155,832 bp, which falls within the cp genome size range for angiosperms but tends to be smaller than the cp genomes of other Rosaceae species (Palmer, 1985; Cheng et al., 2017). LSC regions showed the most difference in size, ranging from 85,471 to 85,726 bp. Additionally, the inferred structures and gene contents were in accordance with previous research (Sun et al., 2021).
Overall, *Fragaria* cp genomes were highly conservative, both in sequence and structure. Analysis with mVISTA showed that there is high similarity among *Fragaria* species apart from *F. chinensis*, *F. viridis*, and *F. orientalis*. We also observed that most variable regions were located in LSC, and non-coding regions were more variable than coding regions. This is a common phenomenon in the cp genomes of most angiosperms (Nazareno et al., 2015; Cheng et al., 2017; Asaf et al., 2018; Tyagi et al., 2020). Additionally, some of the most divergent regions of *ycf1*, *rps16-trnQ*, *petN-psbM*, and *rpl32-trnL*, as shown in Figure 5, were consistent with previous research (Cheng et al., 2017), indicating that these regions indeed evolve rapidly in *Fragaria*.
| Species                  | Size (bp) | Gene number | GC content (%) | References          |
|-------------------------|-----------|-------------|----------------|---------------------|
|                         | Total     | LSC         | SSC            | IR                  | Total | PCG | tRNA | rRNA |                     |
| F. nilgerrensis_1       | 155,783   | 85,712      | 18,165         | 25,963             | 130   | 85  | 37   | 8    | 37.3                |
| F. nilgerrensis_2       | 155,675   | 85,602      | 18,147         | 25,963             | 130   | 85  | 37   | 8    | 37.2                |
| F. mandshurica_JL       | 155,559   | 85,504      | 18,161         | 25,947             | 130   | 85  | 37   | 8    | 37.2                |
| F. mandshurica_HLJ      | 155,556   | 85,507      | 18,155         | 25,947             | 130   | 85  | 37   | 8    | 37.2                |
| F. corymbosa_JL         | 155,684   | 85,538      | 18,198         | 25,974             | 130   | 85  | 37   | 8    | 37.2                |
| F. corymbosa_XZ         | 155,683   | 85,544      | 18,217         | 25,961             | 130   | 85  | 37   | 8    | 37.2                |
| F. corymbosa_GS         | 155,696   | 85,557      | 18,217         | 25,961             | 130   | 85  | 37   | 8    | 37.2                |
| F. mucopinensis_XZ      | 155,630   | 85,487      | 18,219         | 25,962             | 130   | 85  | 37   | 8    | 37.2                |
| F. mucopinensis_SC      | 155,626   | 85,489      | 18,215         | 25,961             | 130   | 85  | 37   | 8    | 37.2                |
| F. pentaphylla_1        | 155,626   | 85,510      | 18,192         | 25,962             | 130   | 85  | 37   | 8    | 37.2                |
| F. pentaphylla_2        | 155,640   | 85,524      | 18,192         | 25,962             | 130   | 85  | 37   | 8    | 37.2                |
| F. pentaphylla_3        | 155,628   | 85,511      | 18,193         | 25,962             | 130   | 85  | 37   | 8    | 37.2                |
| F. pentaphylla_4        | 155,666   | 85,549      | 18,193         | 25,962             | 130   | 85  | 37   | 8    | 37.2                |
| F. nubicola             | 155,608   | 85,512      | 18,174         | 25,961             | 130   | 85  | 37   | 8    | 37.2                |
| F. daltioniana_1        | 155,829   | 85,723      | 18,170         | 25,968             | 130   | 85  | 37   | 8    | 37.2                |
| F. daltioniana_2        | 155,829   | 85,723      | 18,170         | 25,968             | 130   | 85  | 37   | 8    | 37.2                |
| F. daltioniana_3        | 155,832   | 85,726      | 18,170         | 25,968             | 130   | 85  | 37   | 8    | 37.2                |
| F. daltioniana_4        | 155,827   | 85,721      | 18,170         | 25,968             | 130   | 85  | 37   | 8    | 37.2                |
| F. daltioniana_5        | 155,827   | 85,721      | 18,170         | 25,968             | 130   | 85  | 37   | 8    | 37.2                |
| F. viridis              | 155,479   | 85,471      | 18,116         | 25,946             | 130   | 85  | 37   | 8    | 37.2                |
| F. vesca ssp. bracteata | 155,564   | 85,541      | 18,151         | 25,936             | 130   | 85  | 37   | 8    | 37.2                |
| F. vesca ssp. vesca_1   | 155,607   | 85,520      | 18,191         | 25,948             | 130   | 85  | 37   | 8    | 37.3                |
| F. vesca ssp. vesca_2   | 155,638   | 85,559      | 18,173         | 25,953             | 130   | 85  | 37   | 8    | 37.2                |
| F. vesca ssp. vesca_3   | 155,564   | 85,521      | 18,147         | 25,948             | 130   | 85  | 37   | 8    | 37.2                |
| F. chinensis_1          | 155,806   | 85,696      | 18,184         | 25,963             | 130   | 85  | 37   | 8    | 37.2                |
| F. chinensis_2          | 155,797   | 85,888      | 18,183         | 25,963             | 130   | 85  | 37   | 8    | 37.2                |
| F. x ananassa           | 155,549   | 85,332      | 18,145         | 25,936             | 130   | 85  | 37   | 8    | 37.2                |
| F. x ananassa           | 155,549   | 85,332      | 18,145         | 25,936             | 130   | 85  | 37   | 8    | 37.2                |
| F. orientalis           | 147,835   | 83,233      | 13,396         | 25,608             | 128   | 84  | 36   | 8    | 37.6                |
| F. chilensis            | 155,603   | 85,567      | 18,146         | 25,945             | 130   | 85  | 37   | 8    | 37.2                |
| F. virginiana           | 155,621   | 85,586      | 18,145         | 25,945             | 130   | 85  | 37   | 8    | 37.2                |
| F. gracilis             | 155,684   | 85,538      | 18,198         | 25,974             | 130   | 85  | 37   | 8    | 37.2                |
| F. tibetica             | 155,643   | 85,498      | 18,219         | 25,963             | 130   | 85  | 37   | 8    | 37.2                |
| F. moschata             | 155,601   | 85,572      | 18,127         | 25,951             | 130   | 85  | 37   | 8    | 37.2                |

LSC, large single-copy region; SSC, small single-copy region; IR, inverted repeat; PCG, Protein-coding gene.
TABLE 3 | Genes in the Fragaria chloroplast genome.

| Category            | Gene group          | Gene name                      |
|---------------------|---------------------|--------------------------------|
|                      |                     | Subunits of photosystem I      |
|                     |                     | psaA, psaB, psaC, psaJ, psaJ   |
|                     |                     | psbA, psbB, psbC, psbD         |
|                     |                     | psbE, psbF, psbH, psbI, psbI   |
|                     |                     | psbJ, psbK, psbL, psbM         |
|                     |                     | psbN, psbT, psbZ               |
|                     |                     | ndhA, ndhB, ndhC              |
|                     |                     | ndhD, ndhE, ndhF, ndhG         |
|                     |                     | ndhH, ndhJ, ndhK              |
| Photosynthesis      | Subunits of NADH    |
|                     | dehydrogenase       |
|                     |                      |
|                     |                      |
|                     | Subunits of ATP     |
|                     | synthase            |
|                     |                      |
|                     |                      |
|                     | Large subunit of    |
|                     | ribosome            |
|                     |                      |
|                     |                      |
|                     | Small subunit of    |
|                     | ribosome            |
|                     |                      |
|                     |                      |
|                     | Subunits of RNA     |
|                     | polymerase          |
|                     |                      |
|                     |                      |
|                     | Ribosomal RNA genes |
|                     |                      |
|                     |                      |
|                     | Self-replication    |
|                     | Transfer RNA genes  |
|                     |                      |
|                     |                      |
|                     | Matrerase           |
|                     | Protease            |
|                     | Envelope membrane   |
|                     | protein             |
|                     | Acetyl-CoA carboxylase|
|                     | c-type cytochrome   |
|                     | synthesis gene      |
|                     | Conserved open reading frames |

*Genes with one intron.
*Genes with two introns.
*Genes with two copies.

The size differences among cp genomes in angiosperms may be caused by both the contraction and expansion of IR regions (Raubeson et al., 2007; Zhao et al., 2015, 2018). To elucidate this mechanism in Fragaria, we compared IR/SC boundaries of Fragaria cp genomes. In general, the distribution of border genes is conserved, but the distances between genes and the borders do differ somewhat. The distances between genes and IR/SC borders of F. virginiana, F. orientalis, and F. × ananassa are in accordance with the prior report by Cheng et al. (2017). Fragaria mandshurica, F. viridis, F. moschata, F. × ananassa, F. vesca ssp. vesca and F. vesca ssp. bracteata showed the same gap between rps19, rpl2, ycf1, trnH, and IR/SC junctions, which may explain why these cp genomes are more conserved. Additionally, some species also exhibited the same distances between the rps19 and LSC/IRa border, including F. pentaphylla, F. mospinensis, and F. tibetica. These results also indicated a low level of molecular divergence in the genus Fragaria.

Additionally, rpl2 and rps19 differed in their distances from the LSC/IRa border, which may be owing to IR contraction and expansion. Compared with F. × ananassa, the IR regions of F. pentaphylla, F. daltoniana, F. nubicola, F. chinensis, F. corymbosa, F. mospinensis, F. orientalis, F. gracilis, F. tibetica, F. virginiana, and F. chiloensis expanded to different degrees. Therefore, IR regions of these species are longer than that of F. × ananassa. For the other species in the family Rosaceae, such as Malus (Terakami et al., 2012), Prunus (Wang et al., 2013), Pyrus (Li et al., 2018), and Prunus (Kim et al., 2019), rps19 crossed the LSC region and IR region. Additionally, rps19 extended to the IRa region, resulting in the presence of wtrps19 having the same length within IRb region. Notably, we found that rps19 was located inside the LSC region. The contraction of rps19 inside the LSC region would result in the size of the Fragaria cp genomes being smaller than that of other species in Rosaceae.

**Phylogenetic Analysis**

The similar morphology of most Fragaria ssp. and the geographical overlap in ranges may lead to taxonomic confusion among collected specimens (Johnson et al., 2014) and finally result in misjudgment of phylogenetic relationship. To explore the evolutionary relationships among Fragaria species, we constructed a phylogenetic tree of Fragaria species based on whole cp genomes with more than 99% bootstrap support across all nodes. Our results showed high consistency with previous results, especially for species clustering in group B (Njuguna et al., 2013; Sun et al., 2021). Additionally, our results include multiple individuals of each species from different collection sites, which increases the reliability of the phylogenetic analysis.

The phylogenetic analysis showed that F. × ananassa and two octoploids, F. chiloensis and F. virginiana, formed a group, which was in accordance with the inferred origin of F. × ananassa from a hybridization between F. virginiana and F. chiloensis (Staudt, 1962, 1989). In addition, the two octoploids were considered to share a common maternal ancestor that may be F. vesca or F. mandshurica (Harrison et al., 1997; Rousseau-Guettina et al., 2009; Dimeglio et al., 2014). Fortunately, we observed that F. vesca ssp. bracteata is evolutionarily closely related to the two octoploids. Similar results were also presented by Sun et al. (2021). Additionally, Njuguna et al. (2013) and Edgar et al. (2019) suggested that F. vesca ssp. bracteata was probably the last diploid progenitor to the octoploid species. Thus, the conclusion that F. vesca ssp. bracteata is an ancestor of octoploid species was strengthened. However, the other ancestors of the octoploid species remain uncertain.

The evolutionary relationships of F. nubicola, F. pentaphylla, F. chinensis, F. daltoniana, F. corymbosa, F. mospinensis, F. gracilis,
FIGURE 2 | Comparison of the LSC, SSC, and IR border regions among 18 Fragaria chloroplast genomes. 'p' is used to indicate pseudogenes.
FIGURE 3 | Continued
and *F. tibetica* have never been clear (Rousseau-Gueutina et al., 2009; Sun et al., 2021). These species are mainly distributed in Western China (Lei et al., 2017) and, in our results, they were clustered into group A. In group A2, diploid *F. pentaphylla* was sister to the tetraploids *F. moupinensis* and *F. tibetica* with 99.9% bootstrap support (Figure 5), which was consistent with previous research (Njuguna et al., 2013; Sun et al., 2021). In addition, *F. pentaphylla* had been suggested to be the diploid ancestor to *F. moupinensis* (Rousseau-Gueutina et al., 2009; Kanneva et al., 2017). Lei et al. (2017) hypothesized that *F. tibetica* is a descendant of *F. pentaphylla* based on their runner branching and number of leaflets. Thus, it can be hypothesized that the tetraploid species *F. moupinensis* and *F. tibetica* may share the same female parent of *F. pentaphylla*, which is supported by their more similar morphological characteristics (Lei et al., 2017) and overlapping distribution in Southwestern China (Staudt, 1989; Johnson et al., 2014; Lei et al., 2017).

In our study, we revealed a sister relationship between *F. corymbosa* and *F. gracilis*, which was consistent with Rousseau-Gueutina et al. (2009). And *F. corymbosa* and *F. gracilis* have some similar morphological characteristics, such as runners are filiform and monopodial, petioles and peduncles have spreading hairs, fruits are red and tasteless, calyx is reflexed, etc. (Lei et al., 2017). So our results strengthened the point that *F. corymbosa* and *F. gracilis* may have the same ancestor (Rousseau-Gueutina et al., 2009). In addition, *F. corymbosa* and *F. gracilis* may be the descendant of *F. chinensis* (Staudt, 2009; Yang and Davis, 2017). Notably, in this study, all accessions of *F. chinensis* and *F. daltoniana* were clustered into group A1, in contrast with the findings of previous studies (Yang and Davis, 2017; Sun et al., 2021). Therefore, future research should explore the relationship among *F. chinensis*, *F. corymbosa*, and *F. daltoniana*. In the future, multiple molecular markers, including cp and nuclear sequence data from more samples from different geographical populations, should be combined with geographical distribution data to analyse ancestral state reconstruction to clarify their phylogenetic ancestor relationships among *F. moupinensis* and *F. tibetica*; *F. chinensis*, *F. corymbosa*, and *F. daltoniana*; *F. vesca* ssp. *bracteata* and the other octoploid species.

Chloroplast capture is an important process of plant evolution (Okuyama et al., 2005). Hybridization and repeated backcross, the cytoplasm of one species is replaced by the cytoplasm of another species through gene flow infiltration, so that the genetic components of the species not only have nuclear genome components inherited from parents, but also capture new chloroplast gene components (Fehr et al., 2007). More and more studies have proved the phenomenon of organelle DNA introgression (Du et al., 2011), and the phenomenon of chloroplast introgression between plant species has also been observed in previous studies on hazelnut (Hu et al., 2020). In this study, the phylogeographical relationships among *Fragaria* species were not declared for the lack of geographical population collections.
However, clear geographical patterns (Western China, Southwestern China, and Northeastern China) have been clearly inferred in phylogenetic tree figure. Chloroplast capture could be another explanation why the chloroplast genome analysis does not appear to reflect the species phylogeographical relationship (Tsitrone et al., 2003). Further research should be conducted by combining multiple molecular tools (e.g., nuclear DNA sequences) together with more comprehensive sampling to clarify if chloroplast capture does occur in Fragaria genus.

**Candidate Barcoding Sequences for Fragaria**

Species identification based on morphology is affected by season, environment, and human factors, which may cause results to be unreliable (Yang et al., 2020). In recent years, DNA barcoding has been widely used to promote accurate species identification owing to its clear advantages (Tegally et al., 2019; Phi et al., 2020; Islam et al., 2021). The ideal DNA barcode would be a single locus that could be universally amplified and sequenced across a broad range of taxa and provide sufficient variation to reliably distinguish among closely related species (Song et al., 2017). Many introns, coding regions, and intergenic regions, such as trnL-trnF (Potter et al., 2000), accD-psal (Wang et al., 2017), ycf1-ndhF (Amar, 2020), matK, and trnK (Hilu et al., 2008), have been used as barcodes for constructing phylogenetic relationships. In this study, with the threshold of nucleotide diversity as 0.007, the five intergenic regions, trnS-atpA, rbcL-accD, and psbE-petL were found to be the most divergent and provide potential information for species identification and phylogenetic analyses of Fragaria. Among them, three intergenic regions, including trnS-atpA, rbcL-accD, and psbE-petL were also suggested in Sun et al. (2021). However,
the threshold of nucleotide diversity used in Sun et al. (2021) was only 0.006, resulting in a low nucleotide diversity of the selected candidate barcoding sequences. In addition, the amplified fragments of the selected intergenic regions trnS-atpA more than 2,000bp in length would reduce the success of the sequencing. In our study, trnS-trnG and trnR-atpA are located in trnS-atpA regions and the length of these two regions are about 700bp, which will result in the high success of sequencing. Therefore, trnS-trnG and trnR-atpA are more suitable than trnS-atpA to be the potential candidate barcoding sequence.

CONCLUSION

This study provides 27 complete cp genome sequences of 11 wild Fragaria species. Comparative analysis of cp genomes of Fragaria species revealed that their genome structure is highly conserved. However, IR expansion or contraction was observed among different Fragaria cp genomes, resulting in cp genomes of different sizes. Five identified highly variable gene regions (trnS-trnG, trnR-atpA, trnC-petN, rbcL-accD, and psbE-petL) showed strong potential for species identification and phylogenetic relationship construction in the genus Fragaria. Phylogenetic analysis indicated that F. vesca ssp. bracteata may be one of the progenitor species of octoploids. Similarly, we hypothesize that F. pentaphylla is one of the progenitors of F. corymbosa and F. tibetica. The analysis of multiple molecular markers combined with morphological characters would be helpful for future research to test this hypothesis.

REFERENCES

Ali, A., Isakko, H., and Peter, P. (2018). IRIsoce: an online program to visualize the junction sites of chloroplast genomes. Bioinformatics 34, 3030–3031. doi: 10.1093/bioinformatics/bty220

Amar, M. H. (2020). yef-N-nRhF genes, the most promising plastid genomic barcode, sheds light on phylogeny at low taxonomic levels in Prunus persica. J. Genet. Eng. Biotechnol. 18, 42. doi: 10.1016/j.jgeb.2020.02.005

Asaf, S., Khan, A. L., Khan, M. A., Shahzad, R., and Lee, I. J. (2018). Increasing strawberry fruit sensorial and nutritional quality using wild and cultivated germplasm. PLoS One 13:e0192966. doi: 10.1371/journal.pone.0192966

Bai, L., Ye, Y., Chen, Q., and Tang, H. R. (2017). The complete chloroplast genome sequence of the white strawberry Fragaria pentaphylla. Conserv. Genet. Resour. 9, 659–661. doi: 10.1007/s12686-017-0713-5

Capocasa, F., Diamanti, J., Tulipani, S., and Battino, M. (2008a). Breeding strawberry (Fragaria × ananassa Duch.) to increase fruit nutritional quality. Biofactors 34, 67–72. doi: 10.1002/biof.5520340107

Capocasa, F., Scalzo, J., Mezzetti, B., and Battino, M. (2008b). Breeding strawberry (Fragaria × ananassa Duch.) to increase fruit nutritional quality. Biofactors 34, 67–72. doi: 10.1002/biof.5520340107

Du, F. K., Peng, X. L., Liu, J. Q., Lascoux, M., Hu, F. S., and Petit, R. J. (2011). Direction and extent of organelle DNA introgression between two spruce species in the Qinghai-Tibetan plateau. New Phytol. 192, 1024–1033. doi: 10.1111/j.1469-8137.2011.03853.x

Edger, P. P., Poorten, T. J., VanBuren, R., Hardigan, M. A., Colle, M., McKain, M. R., et al. (2019). Origin and evolution of the octoploid strawberry genome. Nat. Genet. 51, 541–547. doi: 10.1038/s41588-019-0356-4

Eriksson, T., Donoghue, M. J., and Hibbs, M. S. (1998). Phylogenetic analysis of Potentilla using DNA sequences of nuclear ribosomal internal transcribed spacers (ITS), and implications for the classification of Rosoideae (Rosaceae). Plant Syst. Evol. 211, 155–179. doi: 10.1007/BF00985357

Fehr, J., Huang, W., and Chang, H. (2010). Chloroplast DNA inheritance, ancestry, and sequencing in Fragaria. Acta Hortic. 859, 221–228. doi: 10.17660/ActaHortic.2010.859.25

Fazekas, A. J., Burgess, K. S., Kesanakurti, P. R., Graham, S. W., Newmaster, S. G., Eriksson, T., Hibbs, M. S., Yoder, A. D., Delwiche, C. F., and Donoghue, M. J. (1998). Phylogenetic analysis of the genus Fragaria (strawberry) using intron-containing sequence from the ADH-1 gene. PLoS One 9:e102237. doi: 10.1371/journal.pone.0102237

Davis, T. M., Shields, M. E., Reinhard, A. E., Reavey, P. A., Lin, J., Zhang, H., et al. (2010). Chloroplast DNA inheritance, ancestry, and sequencing in Fragaria. Acta Hortic. 859, 221–228. doi: 10.17660/ActaHortic.2010.859.25

Davis, T. M., Shields, M. E., Reinhard, A. E., Reavey, P. A., Lin, J., Zhang, H., et al. (2010). Chloroplast DNA inheritance, ancestry, and sequencing in Fragaria. Acta Hortic. 859, 221–228. doi: 10.17660/ActaHortic.2010.859.25

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository and accession number(s) can be found below: NCBI repository, accession numbers MZ851747 and MZ851773.

AUTHOR CONTRIBUTIONS

JL designed the research. CL, CC, YT, ZS, and MJ performed the research and analyzed the data. CL wrote the first draft of the manuscript. All authors commented on previous versions of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was financially supported by the Ten Thousand Talent Program of Zhejiang Province (No. 2019R52043) and the National Natural Science Foundation of China (No. 31261120580).

ACKNOWLEDGMENTS

The authors would like to thank Beiwen Yang of Taizhou University (Taizhou, China) for the help of identification of Fragaria species.

FRONTIERS IN PLANT SCIENCE | www.frontiersin.org 12 October 2021 | Volume 12 | Article 754209
Chloroplast Genomes Variation of Fragaria

Wang, S., Shi, C., and Gao, L. Z. (2013). Plastid genome sequence of a wild woody oil species, Prinusexis utilis, provides insights into evolutionary and mutational patterns of Rosaceae chloroplast genomes. *PLoS One* 8:e73946. doi: 10.1371/journal.pone.0073946

Wu, C., Xu, J-L., Chen, S. Y., Li, D. Z., and Yi, T. S. (2017). Plastomes of *Mimosoideae*: structural and size variation, sequence divergence, and phylogenetic implication. *Tree Genet. Genomes* 13:41. doi: 10.1007/s11295-017-1124-1

Wei, W., Zheng, Y. L., Chen, L., Wei, Y. M., Yan, Z. H., and Yang, R. W. (2006). PCR-RFLP analysis of cpDNA and mtDNA in the genus *Houttuynia* in some areas of China. *Hereditas* 142, 24–32. doi: 10.1111/j.1601-5223.2005.01704.x

Wicke, S., Schneeweiss, G. M., dePamphilis, C. W., Müller, K. F., and Quandt, D. (2011). The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol. Biol.* 76, 273–297. doi: 10.1007/s11103-011-9762-4

Wolle, K. H., Li, W., and Sharp, P. M. (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci. U. S. A.* 84, 9504–9508. doi: 10.1073/pnas.84.24.9505

Xin, T. Y., Yao, H., Gao, H. H., Zhou, X. Z., Ma, X. C.; Xu, C. Q., et al. (2013). Super food *Lycium barbarum* (Solanaceae) traceability via an internal transcribed spacer 2 barcode. *Food Res. Int.* 54, 1699–1704. doi: 10.1016/j.foodres.2013.10.007

Yang, Y., and Davis, T. M. (2017). A new perspective on polyploid *Fragaria* (strawberry) genome composition based on large-scale, multi-locus phylogenetic analysis. *Genome Biol. Evol.* 9, 3433–3448. doi: 10.1093/gbe/evx214

Yang, J. Y., Wei, S. J., Su, D. B., Chen, S. Y., Luo, Z. W., Shen, X. M., et al. (2020). Molecular identification and evolutionary characteristics of *Fragaria nigerensis* in Yunnan based on nrdNA ITS and cpDNA plastid trnH sequence analysis. *J. South Agric.* 51, 748–757.

Zhang, Y., Du, L., Liu, A., Chen, J. J., Wu, L., Hu, W. M., et al. (2016). The complete chloroplast genome sequences of five *Epimedium* species: lights into phylogenetic and taxonomic analyses. *Front. Plant Sci.* 7:306. doi: 10.3389/fpls.2016.00306

Zhang, M. L., Song, Y., Ni, J., Yao, X., Tan, Y. H., and Xu, Z. F. (2018). Comparative chloroplast genomics and phylogenetics of nine *Lindera* species (Lauraceae). *Sci. Rep.* 8:8844. doi: 10.1038/s41598-018-27090-0

Zhao, Y. B., Yin, J. L., Guo, H. Y., Zhang, Y. Y., Xiao, W., Sun, C., et al. (2015). Plastomes of five woody oil species, *Magnolia*, *Ailanthus*, *Liriodendron*, and *Pinus* traceability via internal transcribed spacer 2 barcode. *Mol. Biol. Evol.* 32, 2421–2434. doi: 10.1093/molbev/msv079

Zhao, M. L., Song, Y., Ni, J., Yao, X., Tan, Y. H., and Xu, Z. F. (2018). Comparative chloroplast genomics and phylogenetics of nine *Lindera* species (Lauraceae). *Sci. Rep.* 8:8844. doi: 10.1038/s41598-018-27090-0

 Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Li, Cai, Tao, Sun, Jiang, Chen and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.