


Article

Comparisons of Chloroplast Genome Mutations among 13 Samples of Oil-Tea *Camellia* from South China

Jing Chen ^{1,*}, Kaibing Zhou ^{1,*}  and Xinwen Hu ^{2,*}¹ Sanya Nanfan Research Institute of Hainan University, Sanya 572025, China; kimchen111@163.com² School of Life Science, Hainan University, Haikou 570228, China

* Correspondence: zkb@hainanu.edu.cn (K.Z.); huxinwen@hainanu.edu.cn (X.H.)

Abstract: The differences in cpDNA SNPs and InDels of 13 samples from single trees of different species or populations of oil-tea camellia in South China were examined in this study, and phylogenetic trees were reconstructed based on CDSs and non-CDSs of cpDNAs to research the evolutionary relationships among all samples. The SNPs of all samples included all kinds of substitutions, and the frequency of the transition from AT to GC was highest; meanwhile, the frequencies of all kinds of transversions differed among the samples, and the SNPs exhibited polymorphism. The SNPs were distributed in all the different functional regions of cpDNAs, and approximately half of all SNPs in exons led to missense mutations and the gain or loss of termination codons. There were no InDels in the exons of any cpDNA samples, except those retrieved from *Camellia gigantocarpa*, although this InDel did not lead to a frame shift. The InDels of all cpDNA samples were unevenly distributed in the intergenic region and upstream and downstream of genes. The genes, regions of the same gene, sites and mutation types in the same region related to the distributions of SNPs, and InDels were inconsistent among samples. The 13 samples were divided into 2 clades and 7 or 6 subclades, and the samples of species from the same sections of the *Camellia* genus did not belong to the same subclades. Meanwhile, the genetic relationship between the samples of *Camellia vietnamensis* and the undetermined species from Hainan Province or the population of *C. gauchowensis* in Xuwen was closer than that between *C. vietnamensis* and the population of *C. gauchowensis* in Luchuan, and the genetic relationship among *C. osmantha*, *C. vietnamensis* and *C. gauchowensis* was very close. In sum, SNPs and InDels in the different cpDNAs resulted in variable phenotypes among the different species or populations, and they could be developed into molecular markers for studies on species and population identification and phylogenetic relationships. The conclusion from the identification of undetermined species from Hainan Province and the phylogenetic relationships among 13 oil-tea camellia samples based on cpCDS and cpnon-CDS sequences were the same as those from the former report.



Citation: Chen, J.; Zhou, K.; Hu, X. Comparisons of Chloroplast Genome Mutations among 13 Samples of Oil-Tea *Camellia* from South China. *Genes* **2023**, *14*, 1083. <https://doi.org/10.3390/genes14051083>

Academic Editors: Chao Shi, Lassaâd Belbahri and Shuo Wang

Received: 8 April 2023

Revised: 10 May 2023

Accepted: 12 May 2023

Published: 14 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: oil-tea camellia; cpDNA; SNPs; InDels; phylogenetic tree

1. Introduction

Oil-tea camellia is a special oil tree belonging to the *Camellia* genus whose seeds contain rich oil and that was first planted in China long ago. This group of plants includes approximately 100 species, and there are 30 common species among them, such as *C. oleifera*, *C. meiocarpa* and *C. vietnamensis* [1]. *Camellia* seed oil contains abundant tea polyphenols and tea saponin, whereas olive oil does not have both components. Olive oil contains harmful components such as cholesterol and sinapic acid but camellia seed oil does not. Therefore, camellia seed oil has better nutritional and healthcare value, and the oil-tea camellia industry is expected to develop broadly [2]. Oil-tea camellia plants strongly resist many kinds of stress and widely adapt to many kinds of environments, showing extremely strong resistance to typhoons [3], so they can be planted in low-yielding or abandoned land and produce high yields, resulting in better ecological and economic benefits.

Oil-tea camellia was planted in Hainan Province approximately 2000 years ago. Approximately 6000 hm² of oil-tea camellia forest was constructed in the 1960s–1970s, and 4600-year-old trees were found in Qionghai city. People in Hainan Province are accustomed to edible oil-tea camellia seed oil and view it as mysterious. As a result, the supply of oil-tea camellia seed oil cannot meet the market demand, so high-yield cultivars for afforestation need to be selected immediately to promote large-area afforestation at a high rate [4]. The indigenous species of oil-tea camellia trees in Hainan Province has been identified as *C. vietnamensis* based on full cpDNA sequences [5], but the absence of indigenous cultivars has become the first bottleneck in the development of the oil-tea camellia industry in Hainan Province [6].

The indigenous species of oil-tea camellia in Hainan Province is commonly mistaken for *C. oleifera* [7,8], and this is confirmed by the fact that the cultivars of *C. oleifera* outside the Hainan Island grow abnormally. Therefore, although the indigenous species of oil-tea camellia in Hainan Province has been identified as *C. vietnamensis* [5], it is still necessary to study the evolutionary relationships among *C. vietnamensis* from Hainan Province and the other oil-tea camellia species elsewhere in South China, which will aid in the selection of *C. vietnamensis* cultivars for afforestation and the collection of optimal germplasm of oil-tea camellia. Such research is also essential to establish a market for oil-tea camellia seed oil based on *C. vietnamensis* from Hainan Province and innovate and reasonably develop oil-tea camellia germplasm.

The identification of plants based on morphological characteristics is complicated by factors such as different environmental conditions and plant age, and identification based on DNA barcoding is limited in resolution among closely related species, which is not enough to differentiate them from each other [9,10]. The chloroplast genome (cpDNA) is called a “superbarcode” because it has some advantages compared with a single DNA barcode, including a smaller sequence, uniparental maternal inheritance without recombination, and much more abundant variation information, so it has begun to be used for the identification of different species and even different populations of the same species [11]. In cpDNA, the coding sequence (CDS) differs greatly in its evolution rate from the noncoding sequence (non-CDS), and there are low sequencing costs and small splicing errors. cpDNA is used to explore the systematic evolution, classification and identification of plant species more conveniently, accurately and cheaply than other materials [12]. Therefore, there have been many successful studies on the identification, classification and evolutionary relationship inference of plant species based on the method of cpDNA comparison in the past 30 years [13].

It is difficult to identify the species of *Camellia* plants because of hybridization and autopolyploidy [14]. Therefore, it is necessary to identify the different species and discuss their evolutionary relationships from multiple perspectives, including cpDNA comparative analysis. The NCBI genome database (<https://www.ncbi.nlm.nih.gov/genome/organelle/>, accessed on 8 February 2022) contains 7556 full cpDNA sequences, including 45 from *Camellia* (with 10 cpDNA sequences published by our group), providing a set of resources for analysing the evolutionary relationships and identifying species of *Camellia* and other plants. In this paper, based on the CDSs and non-CDSs of cpDNA, the indigenous oil-tea camellia species from Hainan Province were re-identified, and the significance of SNPs and InDels in cpDNA in the evolution of species was uncovered. In addition, the feasibility of using cpDNA SNPs and InDels as molecular markers of genetic evolution is discussed.

2. Materials and Methods

2.1. Experimental Materials

Leaf samples for cpDNA sequencing were collected from various plants, including *C. gauchowensis* from Gaozhou city, Guangdong Province (HD01); *C. gauchowensis* from Luchuan County, Guangxi Zhuang National Autonomous Region (HD02); *C. gigantocarpa* (HD03), *C. polyodonta* (HD04), *C. meiocarpa* (HD05), *C. semiserrata* (HD06), *C. oleifera* (HD07), *C. osmantha* (HD08), *C. vietnamensis* (HD09) and an undetermined species from Qionghai city,

Hainan Province (HD10); an undetermined species from Danzhou city, Hainan Province (HD11); an undetermined species from Wuzhishan city, Hainan Province (HD12); and *C. gauchowensis* from Xuwen County, Guangdong Province (HD13). According to assembly and comparison after cpDNA sequencing, the cpDNA sequences of the HD10–HD13 samples (i.e., 3 undetermined species from Hainan Province and *C. gauchowensis* from Xuwen County, Guangdong Province) were identical, and four emerged from the cpDNA of HD10 [5]. Therefore, 10 cpDNAs were used to analyse gene mutations in this study, whose data can be obtained from <https://www.ncbi.nlm.nih.gov/genbank/>, accessed on 8 February 2022.

2.2. SNP and InDel Analysis

The chloroplast genome of *C. oleifera* mentioned above was used as the reference genome, and the reads were mapped back to the reference genome with the BWA software. Then, the reads from PCR duplication were deleted with the Picard Tools software. According to the mapping results and comprehensively considering factors such as data features, sequencing quality and experiment operation, the probability of every possible genotype was calculated based on the actual data with the software GATK UnifiedGenotyper. The genotype with the highest probability was selected and regarded as the genotype at a specific site in the sequencing unit. Then, a quality value was assigned to reflect the accuracy of the gene, and the sequence identity was determined. According to the sequence identity, the polymorphic sites in the reference sequence were screened and filtered. The main steps were as follows: the SAM file was converted into a BAM file, the BAM file was sequenced, PCR duplicates were identified, reads of PCR duplicates were deleted, reads whose mapping Q values were no more than 10 were filtered and the indexes were calculated. The realignments of the sequences near the InDels were mapped, and then the variations with high accuracy were gained by filtering the results of variant calling based on SNPs and InDels GATK. SNP and small InDel annotation was carried out according to the off-annotation information of the reference genome with the Hannover procedure. With regard to the SNP and small InDel sites in the CDS region, their influences on the translation of the protein were identified.

2.3. Construction of a cpDNA Phylogenetic Map

Sixty-five complete cpDNA sequences were obtained according to the method in a previous report [5], and 10 sample sequences were added for analysis. CDSs and non-CDSs of the 75 full cpDNA sequences were extracted for phylogenetic analysis. The phylogenetic tree was constructed according to the method in the same former report [5]. Every group was analysed using the software MAFFT (default parameters), and sequence pruning was performed using Gblocks (with the following parameters: -t, d, -b5, h). The phylogenetic tree was constructed using the software IQTREE and the outgroup was set as *Hartia laotica* (NC_041509.1).

3. Results and Analysis

3.1. Identification and Preference Analysis of SNP Mutation Types

The cpDNA genome of *C. oleifera* (HD07) was used as the reference sequence, and the frequency distribution of the different SNPs from the different samples is shown in Figure 1. All samples had all the kinds of base substitution mutations, and the sites of every kind of substitution mutation were different from each other among the samples. However, transition mutations from AT to CG were the most common and transversion mutations from CG to GC and from TA to AT were the rarest and second rarest, ranking 6th and 5th, respectively. The other kinds of substitution mutations, such as transitions between CG and AT or transversions between TA and GC, ranked in the middle, and their rankings changed among samples. In sum, the frequency of transition mutations was higher than that of transversion mutations, and the ratio of the former to the latter was less than 2:1, so the substitution mutations of oil-tea camellia are biased towards transition mutations.

Meanwhile, the substitution mutations appeared to be unique and polymorphic among the oil-tea camellia species or populations.

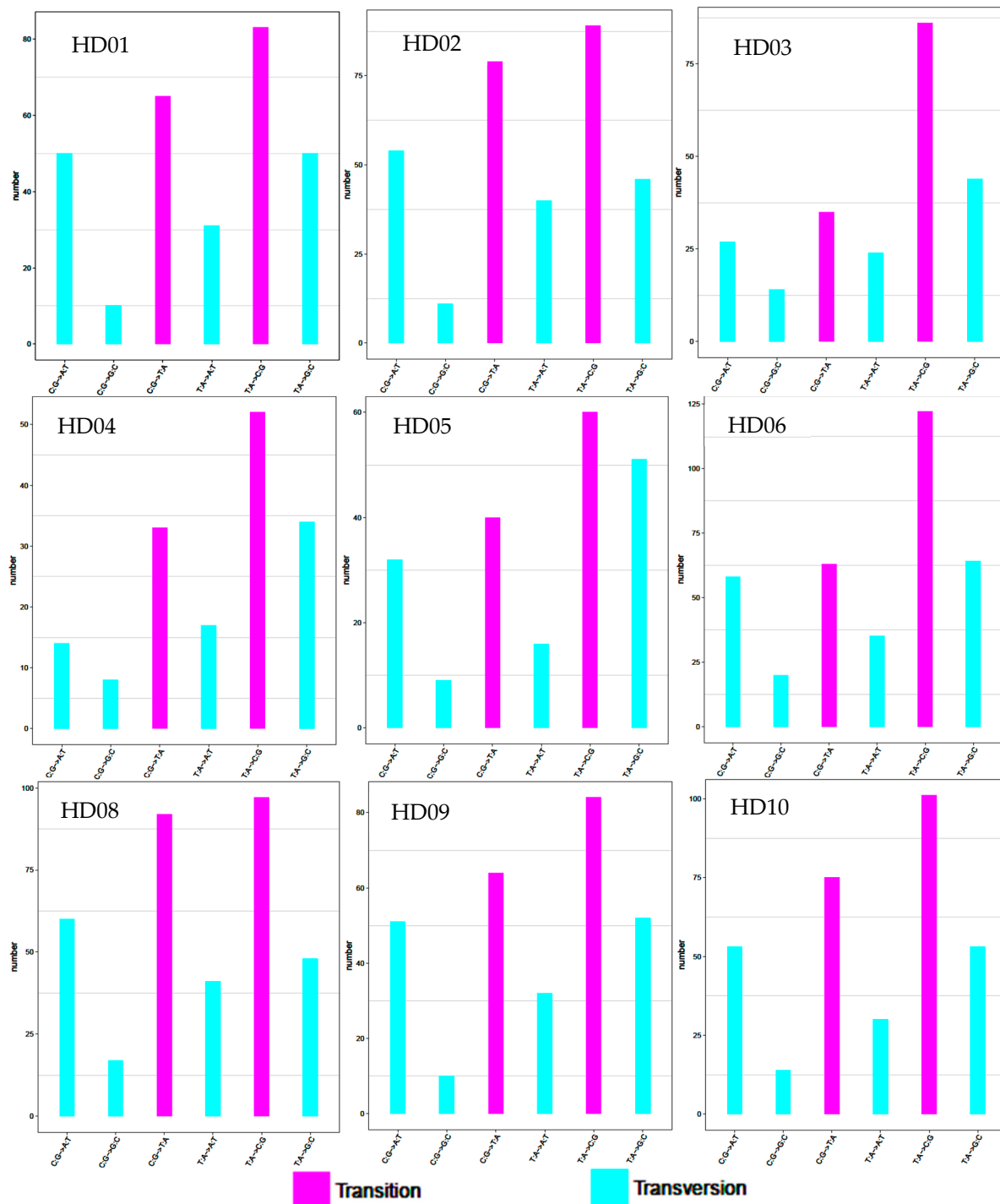


Figure 1. SNP mutations.

3.2. SNP Genotyping

The identification and statistical results of SNP genotyping are shown in Table 1. The total homozygous and heterozygous SNP sites of the different samples differed from each other, which meant that either the polymorphisms existed in the numbers of different genotypes of cpSNPs among the different species or they existed in the different populations of the same species. In other words, every sample had unique cpSNP sites. All samples

showed more homozygous SNP sites than heterozygous SNP sites, which indicated that the variation in cpSNPs was relatively stable. Generally, the more total SNPs there were, the greater the sequence difference from *C. oleifera* (HD07), so the similarity to *C. oleifera* (HD07) exhibited the following order: *C. polyodonta* (HD04), *C. meocarpa* (HD05), *C. gauchowensis* from Gaozhou (HD01), *C. gigantocarpa* (HD03), *C. vietnamensis* (HD09), *C. gauchowensis* from Luchuan (HD02), *C. gauchowensis* from Xuwen (HD13) and the undetermined species from Hainan Province (HD10–HD12), *C. osmantha* (HD08) and *C. semiserrata* (HD06). Generally, the more heterozygous SNP sites there were, the greater the difference in SNP complexity from *C. oleifera* (HD07), such that complexity exhibited the following order: *C. semiserrata* (HD06), *C. osmantha* (HD08), *C. gauchowensis* from Xuwen (HD13) and the undetermined species from Hainan Province (HD10–HD12), *C. gauchowensis* from Luchuan (HD02), *C. vietnamensis* (HD09), *C. gigantocarpa* (HD03), *C. gauchowensis* from Gaozhou (HD01), *C. meocarpa* (HD05) and *C. polyodonta* (HD04). In sum, all the above results indicated that cpSNPs could be used as molecular markers to identify the different species and populations of oil-tea camellia.

Table 1. The results of the SNP genotyping.

Samples	HD01	HD02	HD03	HD04	HD05	HD06	HD08	HD09	HD10
Homozygous SNP	143	152	80	57	77	129	156	142	128
Heterozygous SNP	3	15	70	44	54	104	43	9	70
Total	146	167	150	101	131	233	199	151	198

3.3. SNP Annotation

The cpSNP annotations of 13 samples are shown in Table 2. The specific sequence regions and the site numbers of SNP locations are shown. Although they are not listed individually in this paper, this cpSNP location information would be essential to guide further research on the regulation of gene expression, molecular breeding and cpDNA genetic engineering.

Table 2. The distributions of the SNPs in the genes of the populations of *Camellia gauchowensis*.

Samples	Regions of Gene	Gene (Numbers of the SNPs)	Total
HD01	Downstream	<i>psbA</i> (3), <i>rps16</i> (1), <i>psbI</i> (15), <i>atpA</i> (2), <i>atpI</i> (2), <i>petN</i> (4), <i>ndhJ</i> (5)	32
	Exon	<i>psbA</i> (1), <i>Matk</i> (2), <i>psbK</i> (2), <i>rps2</i> (1), <i>rpoC2</i> (5), <i>rpoB</i> (5), <i>psaA</i> (1), <i>ndhJ</i> (1), <i>atpB</i> (3), <i>rbcL</i> (1), <i>accD</i> (1), <i>accD</i> (3), <i>ycf4</i> (2)	28
	Intergenic space	<i>rpoB</i> (1), <i>psbM</i> (2), <i>rpl20</i> (43)	46
	Intron	<i>rps16</i> (2), <i>atpF</i> (1), <i>rpoC1</i> (2), <i>ycf3</i> (1)	6
	Upstream	<i>rps16</i> (3), <i>psbK</i> (1), <i>psbI</i> (2), <i>rpoB</i> (3), <i>psbM</i> (1), <i>psbD</i> (1), <i>rps4</i> (1), <i>ndhC</i> (2), <i>atpB</i> (2), <i>rpl20</i> (3)	19
	Upstream; downstream	<i>pabA</i> ; <i>matK</i> (1)/ <i>psbI</i> ; <i>psbK</i> (5)/ <i>atpH</i> ; <i>atpI</i> (2)/ <i>psaI</i> ; <i>accD</i> (1)/ <i>cemA</i> ; <i>ycf4</i> (2)/ <i>petG</i> ; <i>petL</i> (1)/ <i>psaJ</i> ; <i>petG</i> (1)/ <i>psaJ</i> (1)/ <i>rpl20</i> ; <i>rps18</i> (1)	15
	Total		146
HD02	Downstream	<i>psbA</i> (3), <i>rps16</i> (1), <i>psbI</i> (15), <i>atpA</i> (2), <i>atpI</i> (2), <i>petN</i> (3), <i>ndhJ</i> (6), <i>psaB</i> (2), <i>rpl20</i> (1)	35
	Exon	<i>matk</i> (1), <i>psbK</i> (2), <i>rps2</i> (1), <i>rpoC2</i> (3), <i>rpoC2</i> (1), <i>rpoB</i> (4), <i>rpoB</i> (2), <i>psaA</i> (1), <i>ndhJ</i> (1), <i>ndhJ</i> (3), <i>atpE</i> (1), <i>atpE</i> (3), <i>atpB</i> (3), <i>rbcL</i> (1), <i>accD</i> (1), <i>accD</i> (1), <i>accD</i> (2), <i>ycf3</i> (1), <i>psbF</i> (1), <i>psbF</i> (1), <i>psbF</i> (1)	34

Table 2. Cont.

Samples	Regions of Gene	Gene (Numbers of the SNPs)	Total
HD13, HD10 –HD12	Intergenic space	<i>psbI</i> (1), <i>psbM</i> (2), <i>rpl20</i> (52)	55
	Intron	<i>rps16</i> (2), <i>atpF</i> (1), <i>rpoC1</i> (2), <i>ycf3</i> (1)	6
	Upstream	<i>rps16</i> (2), <i>psbK</i> (1), <i>psbI</i> (2), <i>rpoB</i> (5), <i>petN</i> (1), <i>psbM</i> (2), <i>rps4</i> (1), <i>ndhC</i> (2), <i>atpB</i> (2), <i>petG</i> (1), <i>rpl20</i> (3)	22
	Upstream; downstream	<i>pabA</i> ; <i>matK</i> (1)/ <i>psbI</i> ; <i>psbK</i> (4)/ <i>atpA</i> ; <i>atpF</i> (1)/ <i>atpH</i> ; <i>atpI</i> (2)/ <i>psaI</i> ; <i>accD</i> (1)/ <i>cemA</i> ; <i>ycf4</i> (2)/ <i>petG</i> ; <i>petL</i> (1)/ <i>psaJ</i> ; <i>petG</i> (1)/ <i>psaJ</i> (1)/ <i>rpl20</i> ; <i>rps18</i> (1)	15
	Total		167
	Downstream	<i>psbA</i> (3), <i>rps16</i> (1), <i>psbI</i> (12), <i>atpA</i> (2), <i>atpI</i> (2), <i>petN</i> (4), <i>ndhJ</i> (5)	29
	Exon	<i>psbA</i> (1), <i>matk</i> (2), <i>psbK</i> (2), <i>rps2</i> (1), <i>rpoC2</i> (4), <i>rpoC2</i> (1), <i>rpoB</i> (3), <i>rpoB</i> (2), <i>psbC</i> (1), <i>psbC</i> (4), <i>psaA</i> (1), <i>ndhJ</i> (2), <i>ndhJ</i> (6), <i>atpE</i> (1), <i>atpE</i> (4), <i>atpB</i> (3), <i>rbcL</i> (1), <i>accD</i> (1), <i>accD</i> (1), <i>accD</i> (1), <i>ycf4</i> (1), <i>ycf4</i> (1), <i>psbJ</i> (2), <i>psbJ</i> (1), <i>psbL</i> (3), <i>psbF</i> (5), <i>psbF</i> (1), <i>psbE</i> (3), <i>psbE</i> (1), <i>petL</i> (4), <i>petL</i> (1), <i>petG</i> (1)	66
	Intergenic space	<i>rpoB</i> (1), <i>psbM</i> (2), <i>rpl20</i> (57)	60
	Intron	<i>rps16</i> (2), <i>atpF</i> (1), <i>rpoC1</i> (2), <i>ycf3</i> (1)	6
	Upstream	<i>rps16</i> (3), <i>psbK</i> (1), <i>psbI</i> (2), <i>rpoB</i> (3), <i>psbM</i> (4), <i>psbD</i> (1), <i>rps4</i> (1), <i>ndhC</i> (1), <i>atpB</i> (1), <i>rpl20</i> (3)	20
	Upstream; downstream	<i>pabA</i> ; <i>matK</i> (1)/ <i>psbI</i> ; <i>psbK</i> (5)/ <i>atpH</i> ; <i>atpI</i> (2)/ <i>psaI</i> ; <i>accD</i> (1)/ <i>cemA</i> ; <i>ycf4</i> (2)/ <i>psaJ</i> ; <i>petG</i> (4)/ <i>psaJ</i> (1)/ <i>rpl20</i> ; <i>rps18</i> (1)	17
	Total		198
	Downstream	<i>psbA</i> (3), <i>rps16</i> (1), <i>psbI</i> (15), <i>atpA</i> (2), <i>atpI</i> (2), <i>petN</i> (4), <i>ndhJ</i> (5)	32
	HD09	Exon	<i>psbA</i> (1), <i>Matk</i> (2), <i>psbK</i> (2), <i>rps2</i> (1), <i>rpoC2</i> (4), <i>rpoC2</i> (1), <i>rpoB</i> (3), <i>rpoB</i> (2), <i>psaA</i> (1), <i>ndhJ</i> (2), <i>atpB</i> (3), <i>rbcL</i> (1), <i>accD</i> (1), <i>accD</i> (1), <i>accD</i> (2), <i>ycf4</i> (1), <i>ycf4</i> (1)
Intergenic space		<i>rpoB</i> (1), <i>psbM</i> (2), <i>rpl20</i> (49)	52
Intron		<i>rps16</i> (2), <i>atpF</i> (1), <i>rpoC1</i> (2), <i>ycf3</i> (1)	6
Upstream		<i>rps16</i> (3), <i>psbK</i> (1), <i>psbI</i> (2), <i>rpoB</i> (3), <i>psbM</i> (1), <i>psbD</i> (1), <i>rps4</i> (1), <i>ndhC</i> (1), <i>atpB</i> (1), <i>rpl20</i> (3)	17
Upstream; downstream		<i>pabA</i> ; <i>matK</i> (1)/ <i>psbI</i> ; <i>psbK</i> (5)/ <i>atpH</i> ; <i>atpI</i> (2)/ <i>psaI</i> ; <i>accD</i> (1)/ <i>cemA</i> ; <i>ycf4</i> (2)/ <i>petG</i> ; <i>petL</i> (1)/ <i>psaJ</i> ; <i>petG</i> (1)/ <i>psaJ</i> (1)/ <i>rpl20</i> ; <i>rps18</i> (1)	15
Total			151
HD08	Downstream	<i>psbA</i> (3), <i>rps16</i> (2), <i>psbI</i> (20), <i>atpA</i> (3), <i>atpI</i> (2), <i>petN</i> (7), <i>ndhJ</i> (5), <i>petA</i> (1)	43
	Exon	<i>matk</i> (1), <i>psbK</i> (2), <i>rps2</i> (1), <i>rpoC2</i> (3), <i>rpoC2</i> (2), <i>rpoB</i> (3), <i>rpoB</i> (2), <i>psaA</i> (1), <i>psaA</i> (2), <i>ndhJ</i> (2), <i>atpE</i> (4), <i>atpE</i> (5), <i>atpB</i> (3), <i>rbcL</i> (1), <i>accD</i> (1), <i>accD</i> (1), <i>accD</i> (2), <i>ycf4</i> (1), <i>psbF</i> (1), <i>psbF</i> (1), <i>petG</i> (1), <i>rpl20</i> (1)	40
	Intergenic space	<i>psbM</i> (4), <i>rps4</i> (1), <i>rpl20</i> (60)	65
	Intron	<i>rps16</i> (3), <i>atpF</i> (1), <i>rpoC1</i> (3), <i>ycf3</i> (3)	10
	Upstream	<i>rps16</i> (3), <i>psbK</i> (1), <i>psbI</i> (2), <i>rpoB</i> (3), <i>psbM</i> (2), <i>rps4</i> (1), <i>ndhC</i> (1), <i>atpB</i> (2), <i>rpl20</i> (5)	20
Upstream; downstream	<i>pabA</i> ; <i>matK</i> (1)/ <i>psbI</i> ; <i>psbK</i> (4)/ <i>atpH</i> ; <i>atpI</i> (2)/ <i>rpoC2</i> ; <i>rpoC1</i> (1) / <i>ndhJ</i> ; <i>ndhC</i> (3)/ <i>psaI</i> ; <i>accD</i> (1)/ <i>cemA</i> ; <i>ycf4</i> (2)/ <i>petA</i> ; <i>cemA</i> (2)/ <i>petG</i> ; <i>petL</i> (1)/ <i>psaJ</i> ; <i>petG</i> (1)/ <i>psaJ</i> (1)/ <i>rpl33</i> (1)/ <i>rpl20</i> ; <i>rps18</i> (1)	21	
Total		199	

Table 2. Cont.

Samples	Regions of Gene	Gene (Numbers of the SNPs)	Total	
HD03	Downstream	<i>psbA</i> (1), <i>psbI</i> (12), <i>atpA</i> (1), <i>atpI</i> (1), <i>petN</i> (7), <i>petA</i> (3)	25	
	Exon	<i>matK</i> (1), <i>psbK</i> (1), <i>atpF</i> (1), <i>rps2</i> (1), <i>rpoC2</i> (1), <i>rpoC1</i> (1), <i>rpoB</i> (2), <i>psbC</i> (1), <i>psbC</i> (3), <i>ndhJ</i> (1), <i>ndhJ</i> (2), <i>ndhJ</i> (8), <i>atpE</i> (3), <i>atpE</i> (5), <i>atpB</i> (2), <i>atpB</i> (1), <i>rbcL</i> (1), <i>accD</i> (2), <i>ycf4</i> (1), <i>psbJ</i> (3), <i>psbJ</i> (2), <i>psbL</i> (3), <i>psbF</i> (1), <i>psbF</i> (1), <i>petL</i> (4), <i>petL</i> (2), <i>petG</i> (2)	56	
		Intergenic space	<i>rps4</i> (2), <i>psbM</i> (2), <i>rpl20</i> (31)	35
		Intron	<i>rps16</i> (2), <i>atpF</i> (2), <i>ycf3</i> (2)	6
	Upstream	<i>rps16</i> (1), <i>psbK</i> (1), <i>psbI</i> (2), <i>rpoB</i> (1), <i>psbM</i> (1), <i>ndhC</i> (1), <i>rpl20</i> (1)	8	
	Upstream; downstream	<i>psbI</i> ; <i>psbK</i> (1)/ <i>ndhJ</i> ; <i>ndhC</i> (2)/ <i>psaI</i> ; <i>accD</i> (1)/ <i>psbJ</i> ; <i>psbE</i> (2)/ <i>psaJ</i> ; <i>petG</i> (13)/ <i>psaJ</i> (1)	20	
	Total		150	
HD04	Downstream	<i>psbA</i> (1), <i>psbI</i> (12), <i>atpI</i> (1), <i>petN</i> (2), <i>atpB</i> (1)	17	
	Exon	<i>psbK</i> (1), <i>rpoC1</i> (1), <i>rpoB</i> (3), <i>rpoB</i> (1), <i>ndhJ</i> (4), <i>ndhJ</i> (1), <i>ndhJ</i> (7), <i>atpE</i> (2), <i>atpE</i> (6), <i>atpB</i> (2), <i>ycf4</i> (1), <i>psbJ</i> (3), <i>psbJ</i> (2), <i>psbL</i> (3), <i>psbF</i> (1), <i>psbF</i> (1), <i>psbF</i> (1)	40	
		Intergenic space	<i>psbM</i> (1), <i>rpl20</i> (19)	20
	Intron	<i>rps16</i> (2), <i>atpF</i> (1), <i>ycf3</i> (1)	4	
	Upstream	<i>rps16</i> (1), <i>psbI</i> (1), <i>atpF</i> (1), <i>rpoB</i> (1), <i>psbM</i> (5), <i>rps4</i> (1), <i>petG</i> (1), <i>rpl20</i> (1)	12	
	Upstream; downstream	<i>psbI</i> ; <i>psbK</i> (1)/ <i>ndhJ</i> ; <i>ndhC</i> (2)/ <i>ndhC</i> ; <i>atpE</i> (1)/ <i>psaJ</i> ; <i>accD</i> (1)/ <i>psaJ</i> ; <i>petG</i> (2)/ <i>psaJ</i> (1)	8	
Total		101		
HD05	Downstream	<i>psbA</i> (1), <i>psbI</i> (12), <i>atpI</i> (1), <i>petN</i> (2), <i>atpB</i> (1)	17	
	Exon	<i>matK</i> (2), <i>atpF</i> (1), <i>rps2</i> (1), <i>rpoC2</i> (2), <i>rpoB</i> (2), <i>rpoB</i> (1), <i>psbC</i> (2), <i>psaB</i> (1), <i>ndhJ</i> (4), <i>ndhJ</i> (6), <i>atpE</i> (1), <i>atpE</i> (4), <i>atpB</i> (1), <i>ycf4</i> (1), <i>psbJ</i> (2), <i>psbJ</i> (1), <i>psbL</i> (3), <i>psbF</i> (1), <i>psbF</i> (1), <i>psbF</i> (1), <i>petL</i> (4), <i>petL</i> (2), <i>petG</i> (2)	47	
		Intergenic space	<i>psbI</i> (1), <i>psbM</i> (1), <i>rpl20</i> (39)	41
		Intron	<i>rps16</i> (2), <i>atpF</i> (1), <i>ycf3</i> (1)	4
	Upstream	<i>rps16</i> (1), <i>psbK</i> (1), <i>psbI</i> (2), <i>rpoB</i> (1), <i>ndhC</i> (3), <i>atpB</i> (1), <i>rpl20</i> (1)	10	
	Upstream; downstream	<i>psbA</i> ; <i>matK</i> (1)/ <i>psbI</i> ; <i>psbK</i> (2)/ <i>atpF</i> (2)/ <i>psaI</i> ; <i>accD</i> (1)/ <i>psaJ</i> ; <i>petG</i> (5)/ <i>psaJ</i> (1)	12	
Total		131		
HD06	Downstream	<i>psbA</i> (3), <i>psbI</i> (15), <i>atpA</i> (2), <i>atpI</i> (1), <i>petN</i> (7), <i>ndhJ</i> (4)	32	
	Exon	<i>matK</i> (1), <i>psbK</i> (2), <i>atpH</i> (1), <i>rps20</i> (1), <i>rpoC 2</i> (1), <i>rpoC2</i> (1), <i>rpoB</i> (3), <i>rpoB</i> (3), <i>psbC</i> (1), <i>psbC</i> (4), <i>ndhJ</i> (1), <i>ndhJ</i> (4), <i>ndhJ</i> (7), <i>atpE</i> (6), <i>atpE</i> (9), <i>atpB</i> (5), <i>atpB</i> (6), <i>rbcL</i> (1), <i>accD</i> (1), <i>accD</i> (1), <i>ycf4</i> (1), <i>psbJ</i> (2), <i>psbJ</i> (1), <i>psbL</i> (1), <i>psbF</i> (1), <i>psbF</i> (1), <i>psbF</i> (1), <i>psbE</i> (2), <i>psbE</i> (1), <i>petL</i> (4), <i>petL</i> (1), <i>petL</i> (2), <i>petG</i> (2), <i>rps18</i> (1)	80	
		Intergenic space	<i>psbM</i> (3), <i>rpl20</i> (50)	53
		Intron	<i>rps16</i> (2), <i>atpF</i> (1), <i>rpoC1</i> (1), <i>ycf3</i> (1)	5
	Upstream	<i>matK</i> (1), <i>rps16</i> (3), <i>psbK</i> (1), <i>psbI</i> (2), <i>rpoB</i> (3), <i>psbM</i> (6), <i>rps4</i> (1), <i>ndhC</i> (1), <i>atpB</i> (2), <i>petG</i> (1), <i>rpl20</i> (2)	23	

Table 2. Cont.

Samples	Regions of Gene	Gene (Numbers of the SNPs)	Total
	Upstream; downstream	<i>psbI</i> ; <i>psbK</i> (4)/ <i>atpH</i> ; <i>atpI</i> (2)/ <i>psaA</i> ; <i>ycf3</i> (1)/ <i>ndhJ</i> ; <i>ndhC</i> (2)/ <i>psaJ</i> ; <i>accD</i> (1)/ <i>cemA</i> ; <i>ycf4</i> (2)/ <i>psaI</i> (1)/ <i>petG</i> ; <i>petL</i> (4)/ <i>petG</i> (1)/ <i>psaJ</i> ; <i>petG</i> (20)/ <i>psaJ</i> (1)/ <i>rpl20</i> ; <i>rps18</i> (1)	40
		Total	233

Note: The terms “upstream” and “downstream” are defined as 1 kb away from the transcription start site and transcription end site, respectively, taking into account the strand of the mRNA. If a variant is located in both downstream and upstream regions (possibly for 2 different genes), then “upstream; downstream” will be printed as the output. The intergenic space is >1 kb away from any gene of both. Gene names shown in the bold font and green, blue and red indicate a nonsynonymous mutation, stop-gain mutation and stoploss mutation, respectively. The chloroplast genome of *C. oleifera* was used as the reference genome. The same as below.

The SNPs in the different functional regions of cpDNAs were distributed unevenly. The sequences of the functional regions, according to the numbers of genes containing cpSNPs, were as follows: exons, upstream and downstream overlapping regions, upstream, downstream, introns and intergenic spaces. However, the kinds and numbers of specific related genes in the different regions led to different magnitudes of differences among the samples.

The samples could be divided into five groups on the basis of the number of cp-SNPs in the sequences of functional regions. The first group included the samples of the undetermined species from Hainan Province (HD10–HD12), the population of *C. gauchowensis* from Xuwen (HD13) and *C. osmantha* (HD08), with the sequence types showing the following order: exons, intergenic spaces, downstream, upstream, downstream and upstream overlapping regions and introns. The second group included *C. vietnamensis* (HD09) and the populations of *C. gauchowensis* from Gaozhou and Luchuan (HD01 and HD02, respectively), and the sequences showed the following order: intergenic spaces, downstream, exons, upstream, downstream and upstream overlapping regions and introns. The third group included *C. osmantha* (HD08), and the order was intergenic spaces, upstream, exons, downstream, downstream and upstream overlapping regions and introns. The fourth group included samples other than those of *C. semiserrata* (HD06), with the retrieved order being exons, intergenic spaces, downstream, downstream and upstream overlapping regions, upstream and introns. The last group included *C. semiserrata* (HD06), and the order retrieved for the sample was exons, intergenic spaces, downstream and upstream overlapping regions, downstream, upstream and introns. The samples of each group had similar sequence type orders based on the numbers of SNPs, but the differences in the numbers of SNPs among the specific functional regions varied in magnitude among the samples.

The genes and the sites of the cpSNPs in exons differed among samples, and these different genes and sites led to missense mutations and the gain and loss of termination codons. Additionally, all samples, except those of *C. semiserrata* (HD06), the undetermined species (HD10–HD12) and the Xuwen population of *C. gauchowensis*, which showed loss of the termination codon in the *petL* gene, had no mutations leading to loss of the termination codon but did show missense mutations and gain of the termination codon. All samples had 1–2 gains of termination codons, and the different genes and sites showing missense mutations and gain of termination codons led to variation in the magnitudes of differences, and the majority of these mutations were synonymous mutations. All of these results are summarized in Table 3.

Table 3. The results of the annotations of SNPs in exons.

Samples	HD01	HD02	HD03	HD04	HD05	HD06	HD08	HD09	HD10
Total	28	34	56	40	47	80	40	29	66
Nonsynonymous SNV	11	12	27	18	21	32	16	11	31
Stopgain SNV	1	2	2	2	1	2	1	1	2
Stopiloss SNV	-	-	-	-	-	1	-	-	1
Synonymous SNV	16	20	27	20	25	45	23	17	32

In general, the SNPs of the different samples had some common characteristics, but they still showed differences among the samples.

3.4. SNP Comparisons among Samples

The samples were divided into different groups according to their evolutionary relationships, geographical relationships, and the characteristics of differences between important traits. Then, the samples were compared within each group, and the results are shown in Table 4.

Table 4. The distributions of the same SNPs in the genes of the different sample groups.

	Regions of Gene	Gene (Numbers of the SNPs)	Total
HD01 HD02 HD13	Downstream	<i>psbA</i> (3), <i>rps16</i> (1), <i>atpA</i> (2), <i>atpI</i> (2)	18
	Exon	<i>matk</i> (2), <i>rps2</i> (1), <i>psaA</i> (1), <i>atpB</i> (3), <i>rbcl</i> (1), <i>accD</i> (1)	9
	Intergenic space	<i>rpoB</i> (1), <i>psbM</i> (2)	3
	Intron	<i>rps16</i> (2), <i>atpF</i> (1), <i>rpoC1</i> (2), <i>ycf3</i> (1)	6
	Upstream	<i>psbK</i> (1), <i>psbI</i> (2), <i>rps4</i> (1), <i>rpl20</i> (3)	7
	Upstream; downstream	<i>pabA</i> ; <i>matK</i> (1)/ <i>atpH</i> ; <i>atpI</i> (2)/ <i>psaI</i> ; <i>accD</i> (1)/ <i>cemA</i> ; <i>ycf4</i> (2)/ <i>psaJ</i> (1)/ <i>rpl20</i> ; <i>rps18</i> (1)	8
	Total		51
HD01 HD09 HD10– HD13	Downstream	<i>psbA</i> (3), <i>rps16</i> (1), <i>atpA</i> (2), <i>atpI</i> (2), <i>petN</i> (4), <i>ndhJ</i> (5)	17
	Exon	<i>psbA</i> (1), <i>matk</i> (2), <i>psbK</i> (2), <i>rps2</i> (1), <i>rpoC2</i> (5), <i>rpoB</i> (5), <i>atpB</i> (3), <i>rbcl</i> (1), <i>accD</i> (1), <i>ycf4</i> (2)	23
	Intergenic space	<i>rpoB</i> (1), <i>psbM</i> (2)	3
	Intron	<i>rps16</i> (2), <i>atpF</i> (1), <i>rpoC1</i> (2), <i>ycf3</i> (1)	6
	Upstream	<i>rps16</i> (3), <i>psbK</i> (1), <i>psbI</i> (2), <i>rpoB</i> (3), <i>psbD</i> (1), <i>rps4</i> (1), <i>rpl20</i> (3)	14
	Upstream; downstream	<i>pabA</i> ; <i>matK</i> (1)/ <i>psbI</i> ; <i>psbK</i> (5)/ <i>atpH</i> ; <i>atpI</i> (2)/ <i>psaI</i> ; <i>accD</i> (1)/ <i>cemA</i> ; <i>ycf4</i> (2)/ <i>psaJ</i> (1)/ <i>rpl20</i> ; <i>rps18</i> (1)	13
	Total		76
HD01 HD08 HD09	Downstream	<i>psbA</i> (3), <i>atpI</i> (2), <i>ndhJ</i> (5)	10
	Exon	<i>rps2</i> (1), <i>rpoC2</i> (5), <i>rpoB</i> (5), <i>atpB</i> (3), <i>rbcl</i> (1), <i>accD</i> (1), <i>ycf4</i> (2)	18
	Intron	<i>atpF</i> (1), <i>ycf3</i> (1)	2
	Upstream	<i>rps16</i> (3), <i>psbK</i> (1), <i>psbI</i> (2), <i>rpoB</i> (3), <i>rps4</i> (1), <i>atpB</i> (1)	11
	Upstream; downstream	<i>psaI</i> ; <i>accD</i> (1)/ <i>cemA</i> ; <i>ycf4</i> (2)/ <i>petG</i> ; <i>petL</i> (1)/ <i>psaJ</i> ; <i>petG</i> (1)/ <i>psaJ</i> (1)/ <i>rpl20</i> ; <i>rps18</i> (1)	7
		Total	
HD01 HD02 HD08 HD09 HD10– HD13	Downstream	<i>psbA</i> (3), <i>atpI</i> (2)	5
	Exon	<i>rps2</i> (1), <i>atpB</i> (3), <i>rbcl</i> (1), <i>accD</i> (1)	6
	Intron	<i>atpF</i> (1), <i>ycf3</i> (1)	2
	Upstream	<i>psbK</i> (1), <i>psbI</i> (2), <i>rps4</i> (1)	4
	Upstream; downstream	<i>cemA</i> ; <i>ycf4</i> (2)/ <i>psaJ</i> (1)/ <i>rpl20</i> ; <i>rps18</i> (1)	4
		Total	

Table 4. Cont.

	Regions of Gene	Gene (Numbers of the SNPs)	Total
HD03 HD05	Downstream	<i>psbA</i> (1), <i>psbI</i> (12), <i>atpA</i> (1)	14
	Exon	matK (1), atpF (1), <i>rps2</i> (1), <i>ycf4</i> (1), <i>petL</i> (6), <i>petG</i> (2), psbF (1), psbF (1)	14
	Intron	<i>rps16</i> (2)	2
	Upstream	<i>rps16</i> (1), <i>psbK</i> (1), <i>psbI</i> (2), <i>rpoB</i> (1), <i>rpl20</i> (1)	6
	Upstream; downstream	<i>psaI</i> ; <i>accD</i> (1)/ <i>psaJ</i> (1)	2
	Total		38
HD04 HD06	Downstream	<i>atpI</i> (1)	1
	Exon	rpoB (3), ndhJ (1), ndhJ (4), <i>ndhJ</i> (7), psbF (1), psbF (1), <i>psbF</i> (1)	18
	Intron	<i>rps16</i> (2), <i>atpF</i> (1), <i>ycf3</i> (1)	4
	Upstream	<i>rps4</i> (1), <i>petG</i> (1)	2
	Upstream; downstream	<i>psaJ</i> ; <i>accD</i> (1)/ <i>psaJ</i> (1)	2
	Total		27
	Intron	<i>rps16</i> (2)	2
	Upstream; downstream	<i>accD</i> (1)/ <i>psaJ</i> (1)	1
All samples	Total		3
	Total		0

Note: Gene names shown in the bold font and green, blue and red indicate a nonsynonymous mutation, stop-gain mutation and stoploss mutation, respectively.

3.4.1. Comparisons among the Different Populations of *C. gauchowensis* and among *C. gauchowensis*, *C. vietnamensis* and the Undetermined Species from Hainan Province

A total of 51 SNPs were consistent among the samples of the three populations of *C. gauchowensis* (HD01, HD02 and HD13), and nine in exons included a missense mutation of the *rbcl* gene and a gain of the termination codon of the *accD* gene. Seventy-six SNPs were consistent among the samples of the three populations of *C. gauchowensis* (HD01, HD02 and HD13), *C. vietnamensis* (HD09) and the undetermined species from Hainan Province (HD10–HD12), and twenty-three of those distributed among exons included the missense mutation of the *rbcl* gene and a gain of the termination codon of the *accD* gene. The results of further comparison between the two groups indicated that all of the SNPs in the former were also present in the latter. The results of the analysis combined with Table 2 revealed that each sample of the different populations had unique SNPs beyond those displayed in Table 3. The SNPs in the intergenic space of the *rpl20* gene differed significantly among the populations, and the samples of the undetermined species from Hainan Province (HD10–HD12) and the population of *C. gauchowensis* from Xuwen (HD13) had more SNPs than those of the populations of *C. gauchowensis* from Gaozhou and Luchuan (HD01 and HD02). The first two had more regions of unique SNPs, especially the exons of the genes *psbC*, *psbJ*, *psbL*, *psbF*, *psbE*, and *petL*, which all exhibited unique missense mutations.

3.4.2. Comparison among *C. osmantha*, *C. gauchowensis* and *C. vietnamensis*

C. osmantha, *C. gauchowensis* and *C. vietnamensis* were distributed in the same production area, and the differences in the SNPs among them are worth noting. There were 48 SNPs consistent with each other, and they included the same missense mutations of the *rbcl* gene and the same gain of the termination codon of the *accD* gene. The SNPs in the intergenic spaces were completely different, and there were few SNPs in the other functional regions of genes. Forty-five SNPs were identified in the comparisons of the samples of *C. gauchowensis*, *C. vietnamensis* and the undetermined species from Hainan Province, so there were three different SNPs upstream of the *atpB* gene and the overlapping regions of the *petG* and *petL* genes and *psaJ* and *petG* genes, respectively. The results of the analysis combined with Table 2 revealed that the genetic relationship between *C.*

osmantha (HD08) and the population of *C. gauchowensis* from Gaozhou (HD01) was farther than that between *C. vietnamensis* (HD09) and the population of *C. gauchowensis* from Gaozhou (HD01), and the genetic relationship between the population of *C. gauchowensis* from Gaozhou (HD01) and *C. osmantha* (HD08) was similar to that between *C. osmantha* (HD08) and *C. vietnamensis* (HD09).

3.4.3. Comparisons among *C. gauchowensis*, *C. vietnamensis*, Undetermined Species from Hainan Province and *C. osmantha*

Twenty-one SNPs were consistent among the samples of the three populations of *C. gauchowensis* (HD01, HD02 and HD13), *C. vietnamensis* (HD09), the undetermined species from Hainan Province (HD10–HD12) and *C. osmantha* (HD08). In addition, these SNPs were also observed in the comparisons of samples from *C. osmantha*, *C. gauchowensis*, and *C. vietnamensis* and those from *C. gauchowensis*, *C. vietnamensis* and the undetermined species from Hainan Province. These comparisons also revealed the same missense mutation of the *rbcL* gene and gain of the termination codon of the *accD* gene. There were no identical SNPs in the intergenic spaces, and there were few identical SNPs in the other functional regions of genes, so the cpSNPs of *C. osmantha* were different from those of the undetermined species from Hainan Province.

3.4.4. Comparison between *C. giganticarpa* and *C. meiocarpa*

The fruits of *C. giganticarpa* (HD03) and *C. meiocarpa* (HD05) are the largest and smallest among the afforestation species, respectively, examined in this paper, so it was worth comparing their SNPs. There were 38 consistent SNPs, no identical SNPs in the introns and far fewer identical SNPs in the other functional regions of genes than in the comparisons of the samples from the three populations of *C. gauchowensis* (HD01, HD02 and HD13) and the three populations of *C. gauchowensis* (HD01, HD02 and HD13), *C. vietnamensis* (HD09), *C. osmantha* (HD08) and the undetermined species from Hainan Province, respectively. The same missense mutations were observed in the exons of the *matk*, *atpF* and *psbF* genes, and the same gain of the termination codon of the *psbF* gene were observed. Additionally, there were numerous missense mutations that differed from each other, revealing large differences in the cpSNPs. Moreover, the SNPs in the intergenic space of the *rpl20* gene were different.

3.4.5. Comparison between *C. polyodonta* and *C. semiserrata*

The flowers of *C. polyodonta* (HD04) and *C. semiserrata* (HD06) are both red, so it was worth comparing the SNPs of these two species. There were 27 shared SNPs, fewer than in the comparisons mentioned above, but there were more identical SNPs in the exons (18), more than in the comparisons of the samples from the three populations of *C. gauchowensis* (HD01, HD02 and HD13). Moreover, the same missense mutations of the genes *rpoB*, *ndhJ* and *psbF* and gains of the termination codons of the genes *ndhJ* and *psbF* were observed. There were no identical SNPs in the intergenic spaces, and there were fewer identical SNPs in the other functional regions of genes.

3.4.6. Integrative Comparisons

The results of the comparisons of SNPs among the samples of *C. giganticarpa* (HD03), *C. polyodonta* (HD04), *C. meiocarpa* (HD05) and *C. semiserrata* (HD06) revealed only 3 shared SNPs, namely, 2 in the intron of the *rps16* gene and 1 in the overlapping downstream region of the *accD* gene and upstream region of the *psaJ* gene. Yet, there were no identical SNPs in the other functional regions of the genes. The results of the comparisons among all 13 samples revealed no identical SNPs.

In general, the similarity among the samples of *C. vietnamensis* (HD09), the undetermined species from Hainan Province (HD10–HD12) and the populations of *C. gauchowensis* from Gaozhou and Xuwen (HD01 and HD13) was higher than that among the samples of the three populations of *C. gauchowensis* (HD01, HD02 and HD13), which indicated that

C. vietnamensis (HD09), *C. gauchowensis* (HD01, HD02 and HD13) and the undetermined species from Hainan Province (HD10–HD12) might belong to different populations of the same species. The similarity of the SNPs among the different populations of the same species was higher than that among the different species, the difference among the SNPs in the intergenic spaces was higher than that in the regions within genes, the abundance and differences of the SNPs in the exons were higher than those in the introns and there were no identical SNPs among all samples.

3.5. Identification of the Kinds of InDels and Calculation of Statistics at the InDel Sites

The results for InDels of all samples are shown in Table 5. The numbers of total deletion and insertion sites of the different samples differed from each other, meaning that polymorphisms existed in the numbers of the different types of InDels among the species or populations of the same species. In other words, every sample had unique InDel sites. The ratios of the number of deletions to the number of insertions were all greater than one, suggesting that the frequency of deletions was greater than that of insertions. The ratios of the number of deletions to the number of insertions of the samples from the populations of *C. gauchowensis* from Gaozhou and Xuwen (HD01 and HD13), *C. vietnamensis* (HD09) and the undetermined species (HD10–HD12) were almost equivalent, and the ratios of the number of deletions to the number of insertions of the samples from *C. osmantha* (HD08) and *C. semiserrata* (HD06) were slightly higher than those from the populations of *C. gauchowensis* from Gaozhou and Xuwen (HD01 and HD13), *C. vietnamensis* (HD09) and the undetermined species from Hainan Province (HD10–HD12). However, those of the other three species (HD03, HD04 and HD05) and the population of *C. gauchowensis* from Luchuan (HD02) were much higher. Therefore, the cpDNA InDels tended to be deletions, and the relative abundance of deletions depended on the species or population of oil-tea camellia. In summary, all the above results indicated that cpDNA InDels could be used as molecular markers to identify different species and populations of oil-tea camellia.

Table 5. The results of the identification of InDels.

Samples	HD01	HD02	HD03	HD04	HD05	HD06	HD08	HD09	HD10
Deletion InDels	22	28	25	13	17	22	25	21	20
Insertion InDels	17	17	13	8	7	16	18	17	16
Total	39	45	38	21	24	38	43	38	36
The ratio of deletion to insertion	1.29	1.65	1.92	1.63	2.43	1.38	1.39	1.24	1.25

3.6. InDel Annotation

The cpDNA InDel annotations of 13 samples are shown in Table 6. The specific sequence regions and the site numbers of InDels were mainly distributed in the downstream, upstream and intergenic regions, whereas fewer were distributed in the introns, and almost none were distributed in the exons. Although the InDels are not introduced one by one in this paper, their location information would be essential for further research on the regulation of gene expression, molecular breeding and cpDNA-based genetic engineering.

Table 6. The distributions of the InDels in the genes.

Samples	Regions of Gene	Gene (Numbers of the SNPs)	Total
HD01	Downstream	<i>psbA</i> (1), <i>rps16</i> (1), <i>psbI</i> (3), <i>atpA</i> (2), <i>psbZ</i> (1), <i>psaB</i> (1), <i>petA</i> (1)	10
	Intergenic space	<i>psbM</i> (1), <i>rps4</i> (1), <i>rpl20</i> (4)	6
	Intron	<i>rps16</i> (2), <i>ycf3</i> (1)	3
	Upstream	<i>matK</i> (2), <i>psbK</i> (1), <i>psbD</i> (1), <i>ycf3</i> (1), <i>rps4</i> (1), <i>ndhC</i> (1)	7
	Upstream; downstream	<i>psbI</i> ; <i>psbK</i> (1)/ <i>atpI</i> (1)/ <i>psaI</i> ; <i>accD</i> (1)/ <i>psaI</i> (1)/ <i>ycf4</i> ; <i>accD</i> (1)/ <i>cemA</i> ; <i>ycf4</i> (3)/ <i>petA</i> ; <i>cemA</i> (1)/ <i>psaJ</i> (1)/ <i>rpl33</i> (1)/ <i>rpl20</i> ; <i>rps18</i> (2)	13
	Total		39
HD02	Downstream	<i>psbA</i> (1), <i>rps16</i> (1), <i>psbI</i> (3), <i>atpA</i> (4), <i>psbZ</i> (1), <i>psaB</i> (3), <i>petA</i> (2)	15
	Intergenic space	<i>psbM</i> (1), <i>rps4</i> (1), <i>rpl20</i> (6)	8
	Intron	<i>rps16</i> (2), <i>ycf3</i> (1)	3
	Upstream	<i>matK</i> (2), <i>rps16</i> (1), <i>psbK</i> (1), <i>psbD</i> (1), <i>ycf3</i> (1), <i>rps4</i> (1), <i>ndhC</i> (1)	8
	Upstream; downstream	<i>psbA</i> ; <i>matK</i> (1)/ <i>psbI</i> ; <i>psbK</i> (1)/ <i>atpI</i> (1)/ <i>ndhJ</i> ; <i>ndhC</i> (1)/ <i>psaI</i> ; <i>accD</i> (1)/ <i>ycf4</i> ; <i>accD</i> (1)/ <i>cemA</i> ; <i>ycf4</i> (1)/ <i>petA</i> ; <i>cemA</i> (1)/ <i>psaJ</i> (1)/ <i>rpl33</i> (1)/ <i>rpl20</i> ; <i>rps18</i> (1)	11
	Total		45
HD03	Downstream	<i>psbA</i> (1), <i>psbI</i> (2), <i>atpA</i> (3), <i>atpI</i> (1), <i>psaB</i> (1), <i>ndhJ</i> (1), <i>petA</i> (3)	12
	Exon	<i>rpoC2</i> (1)	1
	Intergenic space	<i>psbM</i> (2), <i>rps4</i> (1), <i>rpl20</i> (4)	7
	Upstream	<i>matK</i> (1), <i>rps16</i> (2), <i>psbK</i> (1), <i>ycf3</i> (1), <i>rps4</i> (2), <i>ndhC</i> (2), <i>petG</i> (1)	10
	Upstream; downstream	<i>psbI</i> ; <i>psbK</i> (1)/ <i>psaI</i> ; <i>accD</i> (1)/ <i>cemA</i> ; <i>ycf4</i> (2)/ <i>petA</i> ; <i>cemA</i> (1)/ <i>psaJ</i> (1)/ <i>rpl33</i> (1)/ <i>rpl20</i> ; <i>rps18</i> (1)	8
	Total		38
HD04	Downstream	<i>psbI</i> (2), <i>atpA</i> (3), <i>psaB</i> (1)	6
	Intergenic space	<i>rpl20</i> (3)	3
	Upstream	<i>ndhC</i> (2)	2
	Upstream; downstream	<i>atpI</i> (1)/ <i>psaA</i> ; <i>ycf3</i> (1)/ <i>ycf3</i> ; <i>rps4</i> (1)/ <i>psaI</i> ; <i>accD</i> (1)/ <i>cemA</i> ; <i>ycf4</i> (2)/ <i>petA</i> ; <i>cemA</i> (1)/ <i>psaJ</i> (1)/ <i>rpl33</i> (1)/ <i>rpl20</i> ; <i>rps18</i> (1)	10
	Total		21
HD05	Downstream	<i>psbA</i> (1), <i>psbI</i> (2), <i>atpA</i> (3), <i>psaB</i> (1)	7
	Intergenic space	<i>rpl20</i> (5)	5
	Upstream	<i>rps16</i> (2), <i>ndhC</i> (1)	3
	Upstream; downstream	<i>psbI</i> ; <i>psbK</i> (1)/ <i>atpI</i> ; <i>rps2</i> (1)/ <i>psaI</i> ; <i>accD</i> (1)/ <i>cemA</i> ; <i>ycf4</i> (2)/ <i>petA</i> ; <i>cemA</i> (1)/ <i>psaJ</i> (1)/ <i>rpl33</i> (1)/ <i>rpl20</i> ; <i>rps18</i> (1)	9
	Total		24
HD06	Downstream	<i>psbA</i> (1), <i>psbI</i> (3), <i>atpA</i> (2), <i>psbZ</i> (1), <i>psaB</i> (2), <i>petA</i> (1)	10
	Intergenic space	<i>psbM</i> (1), <i>rps4</i> (2), <i>rpl20</i> (7)	10
	Intron	<i>rps16</i> (2), <i>ycf3</i> (1)	3
	Upstream	<i>matK</i> (4), <i>ycf3</i> (1), <i>ndhC</i> (2)	7
	Upstream; downstream	<i>psaI</i> ; <i>accD</i> (1)/ <i>ycf4</i> ; <i>accD</i> (1)/ <i>cemA</i> ; <i>ycf4</i> (2)/ <i>petA</i> ; <i>cemA</i> (1)/ <i>psaJ</i> (1)/ <i>rpl33</i> (1)/ <i>rpl20</i> ; <i>rps18</i> (1)	8
	Total		38

Table 6. Cont.

Samples	Regions of Gene	Gene (Numbers of the SNPs)	Total
HD08	Downstream	<i>psbA</i> (1), <i>rps16</i> (2), <i>psbI</i> (3), <i>atpA</i> (2), <i>psbZ</i> (1), <i>psaB</i> (2), <i>petA</i> (2)	13
	Intergenic space	<i>psbM</i> (1), <i>rps4</i> (1), <i>rpl20</i> (7)	9
	Intron	<i>rps16</i> (2), <i>ycf3</i> (1)	3
	Upstream	<i>matK</i> (2), <i>rps16</i> (1), <i>psbK</i> (1), <i>petN</i> (1), <i>psbM</i> (1), <i>psbD</i> (1), <i>ycf3</i> (1), <i>ndhC</i> (1)	9
	Upstream; downstream	<i>psbI</i> ; <i>psbK</i> (1)/ <i>atpH</i> ; <i>atpI</i> (1)/ <i>atpI</i> (1)/ <i>psaI</i> ; <i>accD</i> (1)/ <i>ycf4</i> ; <i>accD</i> (1)/ <i>cemA</i> ; <i>ycf4</i> (2)/ <i>petA</i> ; <i>cemA</i> (1)/ <i>rpl33</i> (1)/ <i>rpl20</i> ; <i>rps18</i> (1)	9
	Total		43
HD09	Downstream	<i>psbA</i> (1), <i>rps16</i> (1), <i>atpI</i> (3), <i>atpA</i> (1), <i>psbZ</i> (1), <i>psaB</i> (1), <i>petA</i> (1)	9
	Intergenic space	<i>psbM</i> (1), <i>rps4</i> (1), <i>rpl20</i> (4)	6
	Intron	<i>rps16</i> (2), <i>ycf3</i> (1)	3
	Upstream	<i>matK</i> (2), <i>psbK</i> (1), <i>psbD</i> (1), <i>ycf3</i> (1), <i>rps4</i> (1), <i>ndhC</i> (1)	7
	Upstream; downstream	<i>psbI</i> ; <i>psbK</i> (1)/ <i>atpI</i> (1)/ <i>psaI</i> ; <i>accD</i> (1)/ <i>psaI</i> (1)/ <i>ycf4</i> ; <i>accD</i> (1)/ <i>cemA</i> ; <i>ycf4</i> (3)/ <i>petA</i> ; <i>cemA</i> (1)/ <i>psaI</i> (1)/ <i>rpl33</i> (1)/ <i>rpl20</i> ; <i>rps18</i> (2)	13
	Total		38
HD10	Downstream	<i>psbI</i> (1), <i>rps16</i> (1), <i>psbI</i> (1), <i>atpA</i> (2), <i>psbZ</i> (1), <i>psaB</i> (1), <i>petA</i> (1)	8
	Intergenic space	<i>psbZ</i> (1), <i>rps4</i> (1), <i>rpl20</i> (4)	6
	Intron	<i>rps16</i> (1), <i>ycf3</i> (1)	2
	Upstream	<i>matK</i> (2), <i>psbK</i> (1), <i>psbD</i> (1), <i>ycf3</i> (1), <i>rps4</i> (1), <i>ndhC</i> (1)	7
	Upstream; downstream	<i>psbI</i> ; <i>psbK</i> (1)/ <i>atpI</i> (1)/ <i>psaI</i> ; <i>accD</i> (1)/ <i>psaI</i> (1)/ <i>ycf4</i> ; <i>accD</i> (1)/ <i>cemA</i> ; <i>ycf4</i> (3)/ <i>petA</i> ; <i>cemA</i> (1)/ <i>psaI</i> (1)/ <i>rpl33</i> (1)/ <i>rpl20</i> ; <i>rps18</i> (2)	13
	Total		36

Note: The bolding of a gene's name indicates a non-frameshift deletion.

3.6.1. Exon Regions

The *rpoC2* exons of all samples, except the sample of *C. gigantocarpa* (HD03), had a six-base TTAAAA deletion. The InDel in the exon of the *rpoC2* gene of *C. gigantocarpa* (HD03) did not lead to a frameshift, but it might lead to the loss of the codon of phenylalanine and the loss of the termination codon. This meant that the InDels in the exon would be eliminated by mortality induction; yet, a few InDels not causing a frameshift, and thus having no effect on the phenotype, could occur occasionally.

3.6.2. Intron Regions

There were no InDels in the introns of the samples retrieved from *C. gigantocarpa* (HD03), *C. polyodonta* (HD04) and *C. meiocarpa* (HD05), but the introns of the *rps16* and *ycf3* genes contained InDels; the intron of the *ycf3* gene contained one insertion, i.e., T; and that of the *rps16* gene in the other samples, except those from the population of *C. gauchowensis* from Xuwen (HD13) and the undetermined species from Hainan Province (HD10–HD12), contained two deletions, i.e., CTTTTTC, and one InDel, i.e., TTTTCC. However, the samples from the population of *C. gauchowensis* from Xuwen (HD13) and the undetermined species from Hainan Province (HD10–HD12) contained only one deletion, i.e., CTTTTTC. As a result, there were fewer InDels in the introns of different oil-tea camellia species or different populations of the same oil-tea camellia species, but there were some differences in the numbers and kinds of InDels in the introns among the different species and among the populations of the same species.

3.6.3. Upstream, Downstream and Intergenic Regions

The InDels were mainly distributed in the upstream, downstream and intergenic regions of some genes in all samples, and the sizes of the InDels of all samples ranged from 1 to 14 bps. Sizes from 1 to 7 were common; sizes of 11, 12 or 14 bps were uncommon; and sizes of 9, 10 and 13 bps were rare. Importantly, there were two insertions (AACTTTAAATTGAA and AATTGAAAACCTTTC) made up of 14 bases in the downstream sequence of the gene *matk* and one insertion (ATTCCAGTAAAATG) made up of 14 bases in the overlapping regions upstream of the gene *petA* and downstream of the gene *cemA*. There was one deletion (TTTATTCAATCA) made up of 12 bps downstream of the gene *rps16*, and one insertion (ATATAAATAAA) of 11 bps and one deletion (TATGGTAATCCA) of 12 bps in the gene *rpl20*. Additionally, the intergenic spaces of the *rpl20* gene of all samples contained many insertions and deletions of different sizes. Although the genes, the functional regions of the genes in which the InDels were distributed and the number of InDels were not consistent among all samples, there were differences among species or populations. The similarities of the InDels in the samples from the three populations of *C. gauchowensis* (HD01, HD02 and HD13), *C. vietnamensis* (HD09) and the undetermined species from Hainan Province (HD10–HD12) were the highest, and the similarities of the samples mentioned above and *C. osmantha* (HD08) were also high.

Table 6 shows only the genes, the regions of the genes containing InDels and the numbers of InDels; there were also differences in the sites, type of mutation (insertion or deletion), size and base composition, among other features. For example, there was one deletion (AAATAGAAATGAAATTTTTTTT TATTTTATTAATAA) of 35 bps upstream of gene *rps16* in the sample of *C. meiocarpa* (HD05) and one insertion (TCTAAATAGAATTAGTATT) of 19 bps in the overlapping region upstream of the gene *ycf3* and downstream of the gene *rps4* in the sample of *C. polyodonta* (HD04). Only one of all samples contained both InDels, so both might be specific to the two species.

3.6.4. Example Analysis of InDels

The results placing InDels in the intergenic space of the *rpl20* gene are shown as an example in Table 7. All samples contained the deletion TATAATA composed of 7 bp. The samples from the populations of *C. gauchowensis* from Gaozhou (HD01) and Xuwen (HD13), the undetermined species from Hainan Province (HD10–HD12) and *C. vietnamensis* (HD09) had the same InDels, and the AATAT and T deletions were the most common InDels among them; yet, the samples from *C. gauchowensis* from Luchuan (HD02) and *C. osmantha* (HD08) had more InDels than the four mentioned above, such as the insertion of AT and the deletion of TCATGAT in HD02 and the two insertions of T and the deletion of TATGGTAATCCA in HD08. Moreover, the additional InDels in HD02 and HD08 distinguished the two; the additional InDels of HD02 could separate HD02 from the populations of *C. gauchowensis*, and the additional InDels of HD08 could separate HD08 from *C. gauchowensis* and *C. vietnamensis*. However, each of the samples from HD02 to HD08 had a unique combination of the different InDels, such as the insertion of AT and the deletion of TCATGAT in HD02; the insertion of A and the deletion of A in HD03; the insertion of AATAG and A and the deletions of CTTAC and TCTTTT in HD05; the insertions of AT and T and the deletion of TATGGTAATCCA in HD06; and the two insertions of T and the deletion of TATGGTAATCCA in HD08. In summary, the insertion of A in HD03 and the insertion of AATAG and the deletions of CTTAC and TCTTTT in HD05, were the most important because they were the unique InDels of each sample, and the deletion of TATGGTAATCCA in HD06 and HD08 was more important because it did not exist in the other samples and could separate HD06 and HD08 from each other and the other samples when used in combination with the other unique InDels. Additionally, the deletion of AATAT and the two deletions of T were more important InDels because they could be used to distinguish *C. gauchowensis*, *C. vietnamensis*, *C. semiserrata* and *C. osmantha* from the other species.

Table 7. The comparisons of the InDels in the intergenic space of *rp120* genes of the different samples.

Samples	InDels (Insertion/Deletion)	Total	Samples	InDels (Insertion/Deletion)	Total
HD01 HD09– HD13	TATAATA (Deletion) AATAT (Deletion) T (Deletion) T (Deletion)	4	HD05	A (Insertion) TATAATA (Deletion) CTTAC (Deletion) T (Deletion) TCTTTT (Deletion)	5
HD02	TATAATA (Deletion) TCATGAT (Deletion) AT (Insertion) AATAT (Deletion) T (Deletion) T (Deletion)	6	HD06	TATAATA (Deletion) AATAT (Deletion) AT (Insertion) T (Deletion) T (Insertion) TATGGTAATCCA (Deletion) T (Deletion)	7
HD03	TATAATA (Deletion) A (Insertion) A (Deletion) T (Deletion)	4	HD08	TATAATA (Deletion) T (Insertion) AATAT (Deletion) T (Deletion)	7
HD04	TATAATA (Deletion) AATAG (Insertion) T (Deletion)	3		T (Insertion) TATGGTAATCCA (Deletion) T (Deletion)	

The other InDels in the cpDNAs of all samples are not introduced individually here, but the characteristics of the differences in the InDels could be summarised based on results such as those presented above. The cpInDels were mainly distributed in the non-CDSs of cpDNAs, and the genes, functional regions, number, sites, type (insertion or deletion), size and base construction of the cpDNA InDels changed depending on the oil-tea camellia species or population. In general, the cpInDels of different samples shared some common characteristics but they still exhibited polymorphism.

3.7. Phylogenetic Inference

H. laotica was taken as the outgroup, and cpDNAs of all samples and seven other *Camellia* species were applied for phylogenetic reconstruction by maximum likelihood (ML) based on the CDSs. The results are shown in Figure 2. It divided all samples and the sequences from the NCBI into 2 clades; one included 2 subclades, and the other included 5 subclades. The subclade of *C. granthamiana* included samples of *C. vietnamensis* (HD09), *C. gauchowensis* (HD01, HD02 and HD13), the undetermined species from Hainan Province (HD10–HD12), *C. osmantha* (HD08) and *C. granthamiana*. The subclade of *C. azalea* included the samples of *C. semiserrata* (HD06) and *C. azalea*. The sample of *C. gigantocarpa* (HD03) formed an independent subclade. The subclade of *C. japonica* or *C. chekiangoleosa* included the samples of *C. polyodonta* (HD04), *C. japonica* and *C. chekiangoleosa*. The subclade of *C. oleifera* included the samples of *C. oleifera* (HD07) and *C. japonica* (the serial number from the NCBI differed from the one in the above subclade). The subclade of *C. sasanque* included the samples of *C. sasanque* and *C. meiocarpa* (HD05). *C. crapaelliana* formed an independent subclade. The samples of *C. gaochauensis* and *C. vietnamensis* are notable in that they showed the cluster nodes of the samples from the population of *C. gauchowensis* from Gaozhou (HD01) and *C. vietnamensis* (HD09) in the outermost position. Then, the cluster nodes of these two and the samples of the undetermined species from Hainan Province (HD10–HD12) or the population of *C. gauchowensis* from Xuwen (HD13) were located in the second outermost position, with the next node associated with the samples from the population of *C. gauchowensis* from Luchuan (HD02) and *C. osmantha* (HD08). Therefore, the phylogenetic relationships among *C. vietnamensis*, *C. gaochauensis* and the undetermined species from Hainan Province were closer than that between the populations of *C. gauchowensis* from Gaozhou and Luchuan, meaning that *C. vietnamensis*, *C. gauchowen-*

sis and the undetermined species from Hainan Province could be merged into the same species, with *C. osmantha* very closely associated with them. Through further analysis involving combining the above results, the different SNPs from the different samples were used as molecular markers for the identification of species, populations and phylogenetic relationships based on CDSs because there were more SNPs and almost no InDels in exons of cpDNAs.

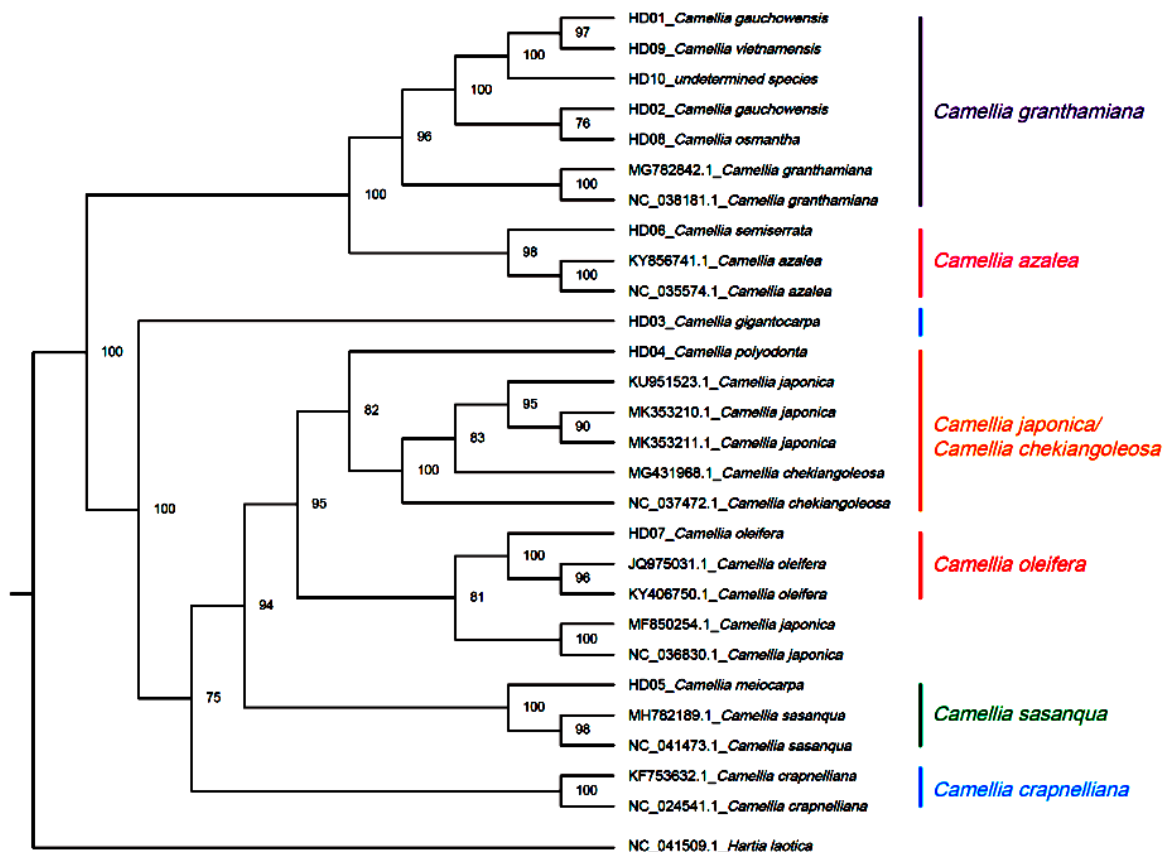


Figure 2. Phylogenetic tree of 13 samples based on CDS using the IQTREE method (ML).

H. laotica was taken as the outgroup, and the cpDNAs of all samples and the seven other *Camellia* species were applied for phylogenetic reconstruction via ML methods based on the non-CDSs. The results are shown in Figure 3. The results were similar to those based on CDSs, and generally showed the same phylogenetic relationships. The difference from the phylogenetic trees based on CDSs was that the sample of *C. gigantocarpa* (HD03) was merged with the subclade of *C. crapaelliana* rather than an independent subclade, so the second clade included 4 subclades instead of 5 clades, making the phylogenetic relationship between the sample of *C. gigantocarpa* and the sequence of *C. crapaelliana* closer. In regard to the subclade of *C. granthamiana*, the phylogenetic trees were extremely similar to those based on CDSs, proving that *C. vietnamensis*, *C. gaochauensis* and the undetermined species from Hainan Province belonged to the same species and that *C. osmantha* was closer than other species to them in terms of genetics. Through further analysis involving combining the above results, the SNPs and InDels from the different samples could be used as molecular markers for the identification of species, populations and phylogenetic relationships based on non-CDSs because there were some SNPs and InDels in the intergenic, upstream, downstream and intron regions of cpDNAs, and these SNPs and InDels were polymorphic among species and populations.

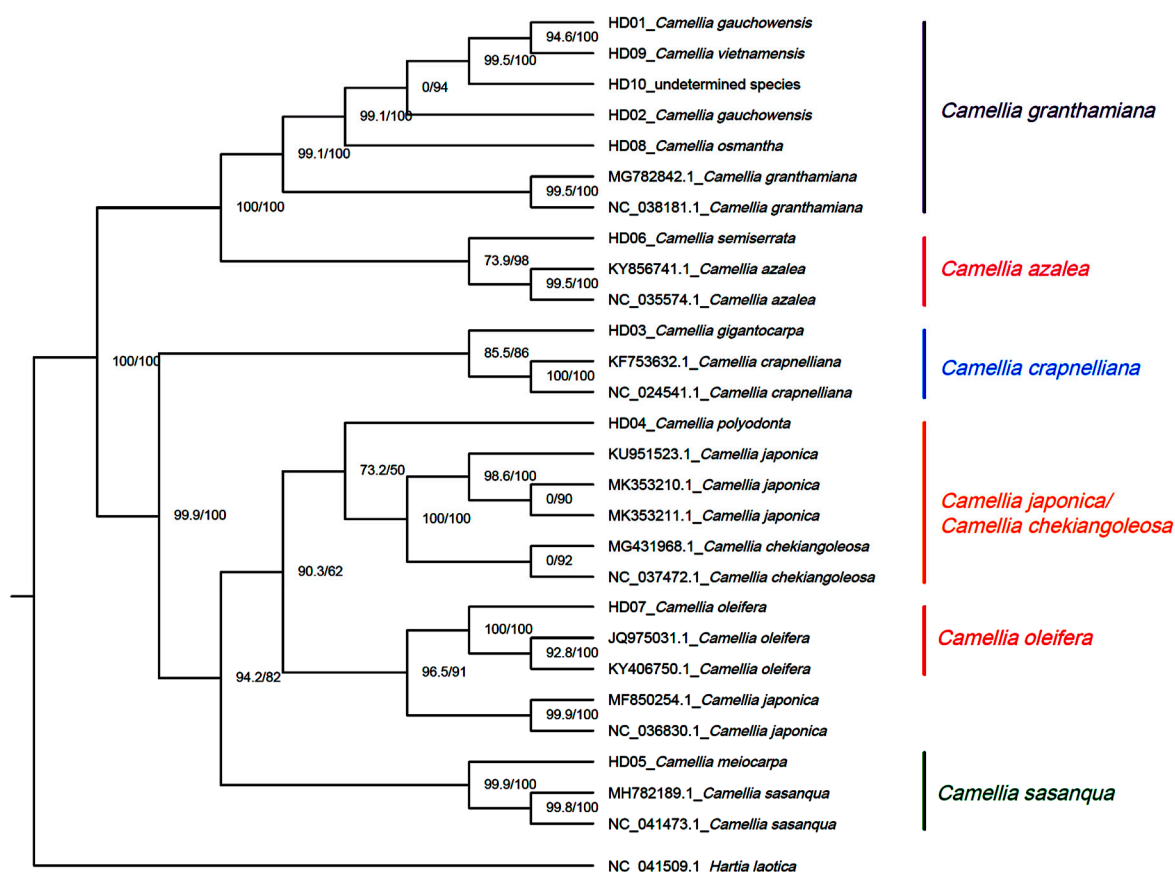


Figure 3. Phylogenetic tree of 13 samples based on non-CDS using the IQTREE method (ML).

4. Discussion

4.1. Primary Analysis of the Genetic Mechanism Underlying the Phylogenetic Relationships within the Camellia Genus

As shown in this paper, some exons in the cpDNA of all samples contained SNPs, and the SNPs appeared to be polymorphic among the different samples. There were inconsistent synonymous mutations, missense mutations and stop-gain mutations among the different samples, and a few samples, such as that of *C. semiserrata* (HD06), the undetermined species from Hainan Province (HD10–HD12) and the population of *C. gauchowensis* (HD13), even contained one stop-loss mutation. Because the SNPs in the exons or the regions for gene expression regulation of all samples could lead to variation in the expression of some traits, the phenotypes varied among the different populations of *C. gauchowensis* and the different oil-tea camellia species. This phenomenon has been reported for nuclear genomes and chloroplast genomes. Synonymous codon usage bias results from synonymous SNPs [15], and the phenotype varies because of differences in gene expression depending on usage frequency and expression efficiency of the different synonymous codons [16]. The *Brassica pekinensis* mutation *cer1*, associated with the absence of wax, is a transition SNP of CT in the fourth exon of the *BraA09g066480.3C* gene and it leads to a missense mutation. Cabbage without wax resulted, so the phenotype changed greatly [17]. Two SNPs in the granule-bound starch synthase I gene of *Oryza sativa* resulted in great changes in the structure and the function of the protein, causing the rice quality to change significantly [18]. SNPs in the tetrahydrocannabinolic acid (THCA) synthase gene could influence the strength of toxicity of *Cannabis sativa*, which allowed hemp products to be divided into those with strong toxicity and weak toxicity based on the difference in SNPs and to be distinguished from other plants' products [19]. In sum, cpSNPs had certain effects on the evolution of the different populations of *C. gauchowensis* and the different oil-tea camellia species.

According to the results in this paper, all samples, except for the sample of *C. gigantocarpa* (HD03), had no InDels in the exons of the cpDNAs, and *C. gigantocarpa* (HD03) had one deletion of 6 bp without a frameshift mutation. This InDel led to the loss of the amino acid and the gain of a termination codon, so it might have significant effects on the phenotype, making *C. gigantocarpa* different from the other plants of the *Camellia* genus. This result indicated that the InDels in the exons of cpDNA were responsible for variation among the populations and species if they led to missense mutations or nonsense mutations. However, cases of InDels in exons contributing to the phylogenetic relationships of organisms have not been reported, possibly because InDels cause frameshift mutations and result in serious harm and even mortality. The results in this paper showed that the InDels of the different samples were distributed in the regions of gene expression regulation, including the upstream, downstream and intergenic regions, even in the introns. Given that the mutation in the upstream promoter of the granule-bound starch synthase I gene of *Solanum tuberosum* could influence the expression of the gene [20], the InDels in the non-CDSs of cpDNA uncovered in this paper may also influence the phenotype by affecting gene expression, causing them to have certain effects on the evolution of *Camellia* plants.

4.2. Application Value Analysis of SNPs and InDels in the cpDNAs of Oil-Tea *Camellia*

The results in this paper reveal the following main points. The SNPs and InDels of cpDNAs were mainly distributed in the intergenic regions and the sequences upstream and downstream of genes and minorly in the introns, and the SNPs were still mainly distributed in the exons. Meanwhile, the CDSs of the samples contained different SNPs, and the non-CDSs of the samples contained different SNPs and InDels, so the cpSNPs and cpInDels could be used for the identification of oil-tea camellia species and their phylogenetic relationships, and they could also be developed into molecular markers.

The phylogenetic relationships of some plants, including 45 species of the *Taxodiaceae* family [11], 20 species of 20 genera of the *Schisandraceae* family [21], 63 species of the *Populus* genus [22] and 42 species of the *Camellia* genus [23], have been researched, and hypothetical results have been obtained. The phylogenetic trees of the different oil-tea camellia samples were constructed based on the CDSs and non-CDSs of the chloroplast genome in this paper, and the different species of oil-tea camellia and different populations of *C. gauchowensis* were clearly distinguished from each other. Furthermore, the phylogenetic relationships of all samples were illustrated, and the phylogenetic trees were consistent with those previously constructed based on full cpDNAs by our research group [5]. The following main conclusions were reached once again. *C. gauchowensis* and *C. vietnamensis* were merged into the same species, and the undetermined species from Hainan Province was *C. vietnamensis*. Additionally, verifying that *C. osmantha* is an independent species will require additional genetic evidence [24,25]. Moreover, because the samples of species from the same sections of the *Camellia* genus did not belong to the same subclades, divisions of oil-tea camellia species into different sections based on morphological characteristics might be unreasonable [26]. All these results proved that the cpSNPs and cpInDels could be used for the identification of species, populations and phylogenetic relationships in this genus.

Studies analysing the phylogenetic relationships among species or cultivars, constructing genetic maps and identifying the types of germplasm based on SNP or InDel molecular markers from the full chloroplast genomes or nuclear genomes of other plants, such as avocado (*Persea americana*), apple (*Malus pumila*), grape (*Vitis vinifera*), peach (*Prunus persica*) and yellow lupin (*Lupinus luteus*), have had desirable outcomes [27–31]. Therefore, the cpSNPs and cpInDels screened in this paper could be developed into molecular markers for the identification of species and populations, the analysis of genetic diversity, the construction of matrilineages and for research on phylogenetic relationships.

5. Conclusions

There were SNPs and InDels in the cpDNAs of all samples. They showed a preference for transition and deletion mutations, respectively. The cpSNPs in the CDSs caused synony-

mous mutations, missense mutations and gain or loss of the termination codon, and the cpSNPs and cpInDels in the non-CDSs caused variations in gene expression. Both cause the phenotypes of different oil-tea camellia species or different populations of *C. gaochauensis* to differ through mutations in cpDNA genes, so both have certain effects on the evolution of oil-tea camellia species or populations. The cpSNPs and cpInDels were unique to species, and can be used for research on the detection and identification of genetic diversity, species and phylogenetic relationships. These differences can also be developed into molecular markers. The phylogenetic trees based on the CDSs and non-CDSs of cpDNAs indicate again that the undetermined species from Hainan Province is *C. vietnamensis*, and that *C. gauchowenesis* and *C. vietnamensis* must be merged into the same species. Additionally, the genetic relationship among *C. osmantha*, *C. gauchowenesis* and *C. vietnamensis* is much closer than the others, and determining whether *C. osmantha* is an independent species needs more genetic evidence. Meanwhile, the sectional division of the *Camellia* genus based on morphological characteristics may need to be readjusted on the basis of differences in cpDNAs.

Author Contributions: Conceptualization, K.Z. and X.H.; Formal analysis, J.C.; Resources, J.C.; Data curation, J.C.; Writing—original draft preparation, J.C.; Writing—review and editing, K.Z.; Supervision, X.H. and K.Z.; Project administration, X.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key R&D Program of Hainan Province, China (ZDYF2022S HFZ020).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. National Forest Farm and Forest Seedling Work Station of State Forestry Administration. *Cultivars of Chinese Oil-Tea Camellia*; Press of Chinese Forestry: Beijing, China, 2016; pp. 11–15. (In Chinese)
2. Zhu, B.; Zhong, H.Y.; Cao, Q.M.; Long, Q.Z. Advance in research on bioactive compounds in *Camellia* spp. *Nonwood For. Res.* **2010**, *28*, 140–145. (In Chinese)
3. Chen, L.Q.; Yang, W.B.; Li, Y.; Wang, Y.; Fu, D.Q.; Niu, X.Q.; Wang, X.S. The development course and prospect of camellia oil industry in Hainan Island. *Mod. Agric. Sci. Technol.* **2012**, *1*, 374. (In Chinese)
4. Dai, J.; Zhong, S.J. Development Status of Camellia Oil Industry in Hainan and Suggestions. *Trop. Agric. Eng.* **2017**, *41*, 61–64. (In Chinese)
5. Chen, J.; Guo, Y.J.; Hu, X.W.; Zhou, K.B. Comparison of the Chloroplast Genome Sequences of 13 Oil-Tea Camellia Samples and Identification of an Undetermined Oil-Tea Camellia Species From Hainan Province. *Front. Plant Sci.* **2022**, *12*, 798581. [[CrossRef](#)] [[PubMed](#)]
6. Chen, Y.Z.; Chen, L.S.; Li, R.F.; Ma, L. Survey and the industry development proposals of *Camellia* sp. resources in Hainan. *Trop. For.* **2017**, *45*, 49–52. (In Chinese)
7. Ye, Z.C.; Wu, Y.G.; Dai, J.; Zhou, K.B.; Hu, X.W. Distribution of Duckweed Community and the Environmental Factors that Affect the Community Structure. *J. Trop. Biol.* **2015**, *3*, 310–314. (In Chinese)
8. Yuan, J.; Han, Z.Q.; He, S.Y.; Huang, L.Y.; Zhou, N.F. Investigation and Cluster Analysis of Main Morphological and Economical Characters for Oil tea Resource in Hainan Province. *J. Plant Genet. Resour.* **2014**, *15*, 1380–1384. (In Chinese)
9. Hebert, P.D.; Cywinska, A.; Ball, S.L.; De Waard, J.R. Biological identifications through DNA barcodes. *Proc. R. Soc. B-Biol. Sci.* **2003**, *270*, 313–321. [[CrossRef](#)]
10. Zhang, J.; Chen, M.; Dong, X.; Lin, R.Z.; Fan, J.H.; Chen, Z.D. Evaluation of four commonly used DNA barcoding loci for Chinese medicinal plants of the family Schisandraceae. *PLoS ONE* **2015**, *10*, e0125574. [[CrossRef](#)]
11. Li, X.; Yang, Y.; Henry, R.J.; Rossetto, M.; Wang, Y.T.; Chen, S. Plant DNA barcoding: From genes to genome. *Biol. Rev.* **2015**, *90*, 157–166. [[CrossRef](#)]
12. Semerikova, S.A.; Semerikov, V.L. Molecular phylogenetic analysis of the genus *Abies* (*Pinaceae*) based on the nucleotide sequence of chloroplast DNA. *Russ. J. Genet* **2014**, *50*, 7–19. [[CrossRef](#)]

13. Leigh, F.J.; Mackay, I.; Oliveira, H.R.; Gosman, N.E.; Horsnell, R.A.; Jones, H.; White, J.; Powell, W.; Brown, T.A. Using diversity of the chloroplast genome to examine evolutionary history of wheat species. *Genet. Resour. Crop Evol.* **2013**, *60*, 1831–1842. [[CrossRef](#)]
14. Liu, Y.; Yang, S.X.; Ji, P.Z.; Gao, L.Z. Phylogeography of *Camellia taliensis* (Theaceae) inferred from chloroplast and nuclear DNA: Insights into evolutionary history and conservation. *BMC Evol. Biol.* **2012**, *12*, 92. [[CrossRef](#)]
15. Zhao, C.L.; Peng, L.Y.; Wang, X.; Chen, J.L.; Wang, L.; Chen, H.; Lai, Z.X.; Liu, S.C. Codon bias and evolution analysis of *AtGAI* in *Amaranthus tricolor* L. *J. China Agric. Univ.* **2019**, *24*, 10–22. (In Chinese)
16. Kwak, S.Y.; Lew, T.T.S.; Sweeney, C.J.; Koman, V.B.; Wong, M.H.; Bohmert-Tatarev, K.; Snell, K.D.; Seo, J.S.; Chua, N.H.; Strano, M.S. Chloroplast-selective gene delivery and expression in planta using chitosan-complexed single-walled carbon nanotube carriers. *Nat. Nanotechnol.* **2019**, *14*, 447–455. [[CrossRef](#)] [[PubMed](#)]
17. Liu, C.; Song, G.; Wang, N.; Huang, S.; Gao, Y.; Fu, W.; Zhang, M.; Feng, H. A single SNP in *Brcr1* results in wax deficiency in Chinese cabbage (*Brassica campestris* L. ssp. *pekinensis*). *Sci. Hortic.* **2021**, *282*, 110019. [[CrossRef](#)]
18. Ardashir, K. An efficient computational method for screening functional SNPs in plants. *J. Theor. Biol.* **2010**, *265*, 55–62.
19. Kojoma, M.; Seki, H.; Yoshida, S.; Muranaka, T. DNA polymorphisms in the tetrahydrocannabinolic acid (THCA) synthase gene in “drug type” and “fiber-type” *Cannabis sativa* L. *Forensic Sci. Int.* **2006**, *159*, 132–140. [[CrossRef](#)]
20. Miao, H.X.; Sun, P.G.; Zhang, K.X.; Jin, Z.Q.; Xu, B.Y. Research progress on expression regulation mechanism of genes encoding granule-bound starch synthase in plants. *Biotechnol. Bull.* **2016**, *32*, 18–23. (In Chinese)
21. Qi, Y.D.; Zhang, B.G.; Liu, H.T.; Xiao, P.G. Complete chloroplast genome sequences of *Schisandra chinensis*: Genome structure, comparative analysis, and phylogenetic relationship of basal angiosperms. *Sci. China Life Sci.* **2017**, *7*, 728–739. (In Chinese)
22. Yun, T.; Li, J.M.; Zhou, A.P.; Yan, L.X.; Zong, D.; Li, D.; He, C.Z. Analysis of phylogenetic relationship of populus based on sequence data of chloroplast regions. *Plant Physiol. J. Plant Physiol. J.* **2015**, *51*, 1339–1346. (In Chinese)
23. Chen, C.; Ma, C.; Ma, J.; Liu, S.; Chen, L. Sequencing of Chloroplast Genome of *Camellia sinensis* and Genetic Relationship for *Camellia* Plants Based on Chloroplast DNA Sequences. *J. Tea Sci.* **2014**, *34*, 371–380. (In Chinese)
24. Liang, G.X.; Liu, K.; Ma, J.L.; Chen, G.C.; Ye, H.; Jiang, Z.P. Molecular taxonomy and identification of *Camellia osmantha*. *Nonwood For. Res.* **2017**, *35*, 26–29. (In Chinese)
25. Wang, D.X.; Ye, H.; Ma, J.L.; Zhou, Z.D. Evaluation and selection of *Camellia osmantha* germplasm resources. *Nonwood For. Res.* **2014**, *32*, 159–162. (In Chinese)
26. Luo, C.Q.; Tan, X.F.; Qi, L.L. A classification summary on plant of Genus *Camellia*. *J. Cent. South For. Coll.* **1999**, *19*, 78–81. (In Chinese)
27. Stahl, P.; Mirom, Y.L.; Stern, R.A.; Goldway, M. Comparing ‘Iriet’ and ‘Ettinger’ avocado cultivars as pollinators of ‘Hass’ using SNPs for paternal identification. *Sci. Hortic.* **2019**, *248*, 50–57. [[CrossRef](#)]
28. Chagné, D.; Gasic, K.; Crowhurst, R.N.; Han, Y.; Bassett, H.C.; Bowatte, D.R.; Lawrence, T.J.; Rikkerink, E.H.A.; Gardiner, S.E.; Korban, S.S. Development of a set of SNP markers present in expressed genes of the apple. *Genomics* **2008**, *92*, 353–358. [[CrossRef](#)]
29. Dong, Q.H.; Cao, X.; Yang, G.; Yu, H.P.; Nicholas, K.K.; Wang, C.; Fang, J.G. Discovery and characterization of SNPs in *Vitis vinifera* and genetic assessment of some grapevine cultivars. *Sci. Hortic.* **2010**, *125*, 233–238. [[CrossRef](#)]
30. Bielsa, B.; Jiwan, D.; i Marti, A.F.; Dhingra, A.; Rubio-Cabetas, M.J. Detection of SNP and validation of a SFP InDel (deletion) in inverted repeat region of the *Prunus* species chloroplast genome. *Sci. Hortic.* **2014**, *168*, 108–112. [[CrossRef](#)]
31. Osorio, C.E.; Udall, J.A.; Salvo-Garrido, H.; Maureira-Butler, I.J. Development and characterization of InDel markers for *Lupinus luteus* L. (Fabaceae) and cross-species amplification in other *Lupinus* species. *Electron. J. Biotechnol.* **2018**, *31*, 44–47. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.