

Article

# Complete Chloroplast Genome of *Pinus densiflora* Siebold & Zucc. and Comparative Analysis with Five Pine Trees

Hye-In Kang <sup>1,†</sup>, Hyun Oh Lee <sup>2,†</sup>, Il Hwan Lee <sup>1</sup>, In Sik Kim <sup>1</sup>, Seok-Woo Lee <sup>1</sup>, Tae Jin Yang <sup>3</sup> and Donghwan Shim <sup>1,\*</sup>

<sup>1</sup> National Institute of Forest Science, Department of Forest Bio-Resources, Suwon 16631, Korea

<sup>2</sup> PHYZEN Genome Institute, 605, Baekgoong Plaza1, Seongnam-si, Gyeonggi-do 13558, Korea

<sup>3</sup> Department of Plant Science, Plant Genomics and Breeding Institute, and Research Institute of Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Korea

\* Correspondence: shim.donghwan@gmail.com

† These authors contributed equally to this work.

Received: 15 May 2019; Accepted: 16 July 2019; Published: 19 July 2019



**Abstract:** *Pinus densiflora* (Korean red pine) is widely distributed in East Asia and considered one of the most important species in Korea. In this study, the complete chloroplast genome of *P. densiflora* was sequenced by combining the advantages of Oxford Nanopore MinION and Illumina MiSeq. The sequenced genome was then compared with that of a previously published conifer plastome. The chloroplast genome was found to be circular and comprised of a quadripartite structure, including 113 genes encoding 73 proteins, 36 tRNAs and 4 rRNAs. It had short inverted repeat regions and lacked *ndh* gene family genes, which is consistent with other Pinaceae species. The gene content of *P. densiflora* was found to be most similar to that of *P. sylvestris*. The newly attempted sequencing method could be considered an alternative method for obtaining accurate genetic information, and the chloroplast genome sequence of *P. densiflora* revealed in this study can be used in the phylogenetic analysis of *Pinus* species.

**Keywords:** chloroplast; genome; pine tree; Oxford Nanopore Technology (ONT); MiSeq

## 1. Introduction

Pinaceae is the largest gymnosperm family, consisting of 10 genera and more than 230 species. Most of Pinaceae species are classified as forest and timber species, and they are mainly distributed in the northern hemisphere [1]. *Pinus densiflora* Siebold & Zucc., also called Korean red pine, is a species widely distributed in East Asia, including the Korean Peninsula, Japan and China [2]. This species occupies more than 23% of forest land in South Korea, and is the most important and popular coniferous species in Korea [3].

The chloroplast genome is a valuable resource in molecular phylogenetic studies [4,5]. It usually exhibits uniparental inheritance and contains conserved sequences as a result of its slower evolutionary rate of change compared to nuclear genomes. Specifically, the genome consists of multiple copies of a circular DNA molecule (110–210 kb) in a chloroplast [6]. The chloroplast DNA of seed plants have quadripartite structures that include a large single copy (LSC) region, a small single copy (SSC) region, and a pair of inverted repeats (IRs) [7]. Most land plants have 110–130 genes in their chloroplast genome [8].

The IR structures of a chloroplast genome typically range in length from 15 kb to 30 kb [6]. Because of the expansion or rearrangement of IRs, there are variations in gene number and order among species. Large IRs play an important role in stabilizing chloroplast genomes and, consequently, large IR loss

can lead to some variation in the genome structure and gene contents [9]. Reductions in IR sequences have been identified in species of Pinaceae, Taxaceae, Cephalotaxaceae and Legumes [10–12].

The development of next generation sequencing (NGS) technologies made it simple and inexpensive to obtain complete chloroplast sequences [13]. However, the technologies come with limitations, including the fact that they generate short reads less than several hundred bases in length. Short reads cause mis-mappings and mis-alignments, making heterozygous and repetitive regions of the genome inaccessible [14]. In addition, it is difficult to identify structural variation or haplotypic structure using short reads alone [15]. Oxford Nanopore Technology (ONT), a new third-generation sequencing platform, promises read lengths that are orders of magnitude longer than previous technologies, and a small, future-oriented, USB-powered sequencer [16]. Using the long reads produced by ONT is efficient as it is unnecessary to align short reads in order to complete a sequence. To overcome the high error rate of ONT reads, high quality MiSeq short reads can be used for error correction [17].

In this study, we characterized the complete single-molecule chloroplast genome of *Pinus densiflora* using Oxford Nanopore MinION (ONM) with long reads, and then Illumina MiSeq short reads were used for error correction. We described the gene contents of the chloroplast genome and compared them with related species. Also, phylogenetic analysis based on the chloroplast genomes of 12 conifers was performed.

## 2. Materials and Methods

### 2.1. Sampling, DNA Extraction and Sequencing

Fresh leaves of *P. densiflora* (MK285358) were collected from a designated cultural heritage site (333–360, Jungyeong-gil, Miro-myeon, Samcheok-si, Gangwon-do, Republic of Korea, N 37° 22' 2" E 129° 3' 32"), and total genomic DNA was extracted from ~100 mg of frozen leaves using Exgene<sup>TM</sup> Plant SV Kit (GeneAll Biotechnology Co., LTD: Seoul, Republic of Korea) following the manufacturer's instructions. Illumina MiSeq and Oxford Nanopore libraries were prepared—in accordance with the manufacturer's instructions—using a TruSeq Nano DNA Kit with a 670 bp average insert size and a rapid sequencing kit (SQK-RAD004, Oxford Nanopore Technologies: Oxford, UK) with a 55 kbp average insert size, respectively, to construct two types of DNA libraries. Then, the two genomic libraries were sequenced by Illumina MiSeq and the Oxford Nanopore MinION platform at PHYZEN (<http://phyzen.com>).

### 2.2. Assemblies of Chloroplast Genome Sequences and Annotation

ONM raw data was basecalled with the default option of the program Albacore (<https://github.com/JGI-Bioinformatics/albacore/blob/master/README.md>), and removal of adapter and chimeric sequences was performed using Porechop (<https://github.com/rrwick/Porechop>) with the default option. After removal of the adapter and chimeric sequences, MinION reads were de novo assembled using SMARTdenovo (<https://github.com/ruanjue/smartdenovo>). Assembled contigs were entered into the BLASTN program with the command line against the National Center for Biotechnology Information (NCBI) nucleotide database, after which only the chloroplast contigs were selected. Overlapping unitigs were used to assemble the complete chloroplast genome sequence (Figure S1). The Illumina MiSeq reads were mapped on the completed chloroplast genome using the `clc_ref_assemble` tool in the CLC Assembly Cell package (version 4.21, CLC Inc, Aarhus, Denmark). Error correction was conducted by manual curation. Gene annotation was conducted by the GeSeq program (<https://chlorobox.mpimp-golm.mpg.de/geseq-app.html>) with default options, which primarily determined protein coding and rRNA gene positions with a BLAT search, while tRNA gene positions were determined using ARAGORN (version 1.2.3, <http://bioinfo.thep.lu.se>) and tRNAscan-SE (version 2.0.3, <http://lowelab.ucsc.edu/tRNAscan-SE/>) [18–20]. The precise generic regions were determined by manual curation using the Artemis annotation tool, which is able to display and manipulate genome

sequences and gene features. The circular gene map was drawn using Organellar Genome DRAW software (ORDRAW, version 1.3.1, <https://chlorobox.mpimp-golm.mpg.de/OGDraw.html>).

### 2.3. Alignments and Construction of a Phylogenetic Tree

A BLASTN search was performed on the NCBI nucleotide database using the newly sequenced *P. densiflora* chloroplast genome sequence. From this, the entire chloroplast genome sequence of the species with the highest similarity was identified and downloaded. Assembled chloroplast DNA (cpDNA) sequences were compared to the chloroplast genome of the most similar species, *P. sylvestris*, by BLASTZ analysis (Figure S2). Sequence alignments of the sequences containing protein coding genes (CDSs) were conducted using mVISTA [21]. For phylogenetic analysis, 12 Pinaceae (*Pinus*, *Picea*, *Larix* and *Abies*) and a *Taxus* species were selected, as well as an outgroup: *T. baccata*. From the 13 species, 59 commonly conserved protein coding genes were extracted: *accD*, *atpA*, *atpB*, *atpE*, *atpF*, *atpH*, *atpI*, *chlB*, *chlL*, *chlN*, *infA*, *matK*, *petA*, *petB*, *petD*, *petG*, *psaA*, *psaB*, *psaC*, *psaI*, *psaJ*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *rbcL*, *rpl14*, *rpl16*, *rpl2*, *rpl20*, *rpl22*, *rpl23*, *rpl32*, *rpl33*, *rpl36*, *rpoA*, *rpoB*, *rpoC1*, *rpoC2*, *rps11*, *rps12*, *rps14*, *rps15*, *rps18*, *rps19*, *rps3*, *rps4*, *rps7*, *rps8*. Multiple sequence alignment was conducted using the MAFFT program (version 7, <https://mafft.cbrc.jp/alignment/software/>) [22]. The maximum likelihood phylogenetic tree was then constructed using the MEGA6 program with the general time reversible (GTR) model and 1000 bootstrap replications.

## 3. Results

### 3.1. The Structure of Chloroplast Genomes of *P. densiflora*

The complete chloroplast genome of *Pinus densiflora* was sequenced using an ONM sequencer and error corrected by Illumina MiSeq short reads. After mapping the ONM reads of the completed chloroplast genome sequence, the sequencing error rate for each base position was calculated to be about 9.29% (Figure S3). The total amount of raw data produced was about 14.6 Gb for MiSeq and 2.1 Gb for ONM (Table 1). The average coverage of the chloroplast genome was  $303.27 \times$  for MiSeq and  $28.46 \times$  for ONM. The completed chloroplast genome sequence of *P. densiflora* was submitted to GenBank (accession number MK285358).

**Table 1.** Information on next generation sequencing (NGS) data of *Pinus densiflora* sequenced in this study.

Sequencing Platform	Input Reads	Trimmed Reads	Raw Bases	Trimmed Bases
MiSeq	49,013,296	40,786,223 (83.21%)	14,563,262,097	10,101,761,675 (69.36%)
ONM	305,965	306,493 (100.21%)	2,116,530,768	2,089,930,503 (98.74%)

The chloroplast of *P. densiflora* had circular DNA molecules 119,875 bp in size, with the typical quadripartite structure composed of an LSC, SSC and two IRs (IRa and IRb). The total length of the genome (119,875 bp) consisted of the 65,654 bp LSC and the 53,231 bp SSC, as well as the two IRs, which contributed 495 bp each. The newly assembled *P. densiflora* chloroplast genome (119,875 bp) was a little larger than other reported *Pinus* species, except loblolly pine (*P. taeda*) (Table 2).

The *P. densiflora* chloroplast genome encoded 108 unique genes, including 72 protein-coding, 32 tRNA and 4 rRNA genes (Table 3). One of these genes (*trnI-CAU*) was repeated in the IR regions, two genes (*psaM* and *trnS-GCU*) had two copies in LSC, and another two genes (*trnH-GUG* and *trnT-GGU*) had one copy in each of LSC and SSC. The duplicated genes were repeated identically and the orientations of four pairs were inverted (except *trnT-GGU*). As a result of duplication, a total of 113 genes—encoding 73 proteins, 36 tRNAs and 4 rRNAs—were detected in the chloroplast genome. Among the 108 unique genes, 12 genes included one intron and two genes contained two introns. The introns ranged in size from 479 bp (for *trnL-UAA*) to 2,501 bp (for *trnK-UUU*). The *rps12* gene was

found to have 3 exons and was assumed to require trans-splicing as exon1 was found in the LSC region, while exon2 and exon3 were in the SSC region. The five *ndh* genes had pseudogenized as follows: *ndhB*, *ndhD*, *ndhE*, *ndhH* and *ndhI*. The gene map of the newly generated *P. densiflora* chloroplast genome is shown in Figure 1.

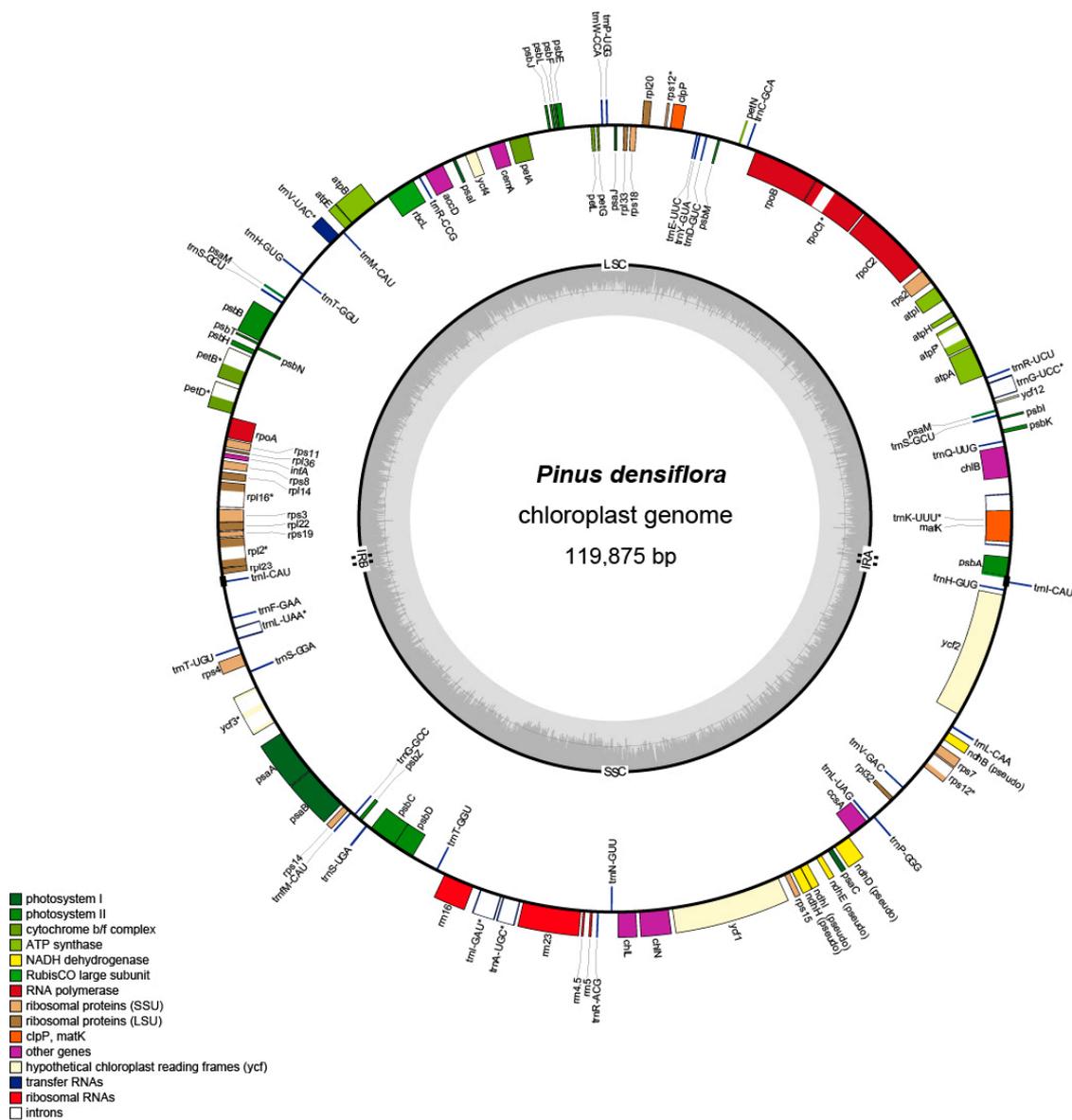
**Table 2.** Summary of *Pinus* chloroplast genome features.

	Genome Size (bp)	LSC Length (bp)	SSC Length (bp)	IR Length (bp)	Number of Genes
<i>P. densiflora</i> (MK285358)	119,875	65,654	53,231	495	113
<i>P. sylvestris</i> (KR476379)	119,758	65,559	53,209	495	112
<i>P. thunbergii</i> (D17510)	119,707	65,696	53,021	495	113
<i>P. tabuliformis</i> (KT740995)	119,646	65,618	53,038	495	114
<i>P. taeda</i> (KC427273)	121,530	66,272	54,288	485	110

**Table 3.** List of genes annotated in the chloroplast genome of *Pinus densiflora* sequenced in this study.

Function	Genes
<b>RNAs, ribosomal</b>	<i>rrn4.5, rrn5, rrn16, rrn23,</i>
<b>RNAs, transfer</b>	<i>trnA-UGC *</i> , <i>trnC-GCA</i> , <i>trnD-GUC</i> , <i>trnE-UUC</i> , <i>trnF-GAA</i> , <i>trnG-GCC</i> , <i>trnG-UCC *</i> , <b><i>trnH-GUG</i></b> , <b><i>trnI-CAU</i></b> , <i>trnI-GAU *</i> , <i>trnK-UUU *,T</i> , <i>trnL-CAA</i> , <i>trnL-UAA *</i> , <i>trnL-UAG</i> , <i>trnM-CAU</i> , <i>trnM-CAU</i> , <i>trnN-GUU</i> , <i>trnP-UGG</i> , <i>trnP-GGG</i> , <i>trnQ-UUG</i> , <i>trnR-ACG</i> , <i>trnR-CCG</i> , <i>trnR-UCU</i> , <b><i>trnS-GCU</i></b> , <i>trnS-GGA</i> , <i>trnS-UGA</i> , <b><i>trnT-GGU</i></b> , <i>trnT-UGU</i> , <i>trnV-GAC</i> , <i>trnV-UAC *</i> , <i>trnW-CCA</i> , <i>trnY-GUA</i>
<b>Transcription and splicing</b>	<i>rpoA, rpoB, rpoC1 *, rpoC2, matK</i>
<b>Translation, ribosomal proteins</b>	
Small subunit	<i>rps2, rps3, rps4, rps7, rps8, rps11, rps12 **,T, rps14, rps15, rps18, rps19</i>
Large subunit	<i>rpl2 *, rpl14, rpl16 *, rpl20, rpl22, rpl23, rpl32, rpl33, rpl36</i>
<b>Photosynthesis</b>	
ATP synthase	<i>atpA, atpB, atpE, atpF *, atpH, atpI</i>
Photosystem I	<i>psaA, psaB, psaC, psaI, psaJ, <b>psaM</b>, ycf3 **, ycf4</i>
Photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
Calvin cycle	<i>rbcl</i>
Cytochrome complex	<i>petA, petB *, petD *, petG, petL, petN</i>
Chlorophyll biosynthesis	<i>chlB, chlL, chlN</i>
<b>Others</b>	<i>clpP, accD, cemA, ccsA, infA, ycf1, ycf2, ycf12</i>

Genes containing one intron; \* genes containing one intron; \*\* genes containing two introns; <sup>T</sup>trans-splicing of the related gene. Genes in boldface type have two gene copies.



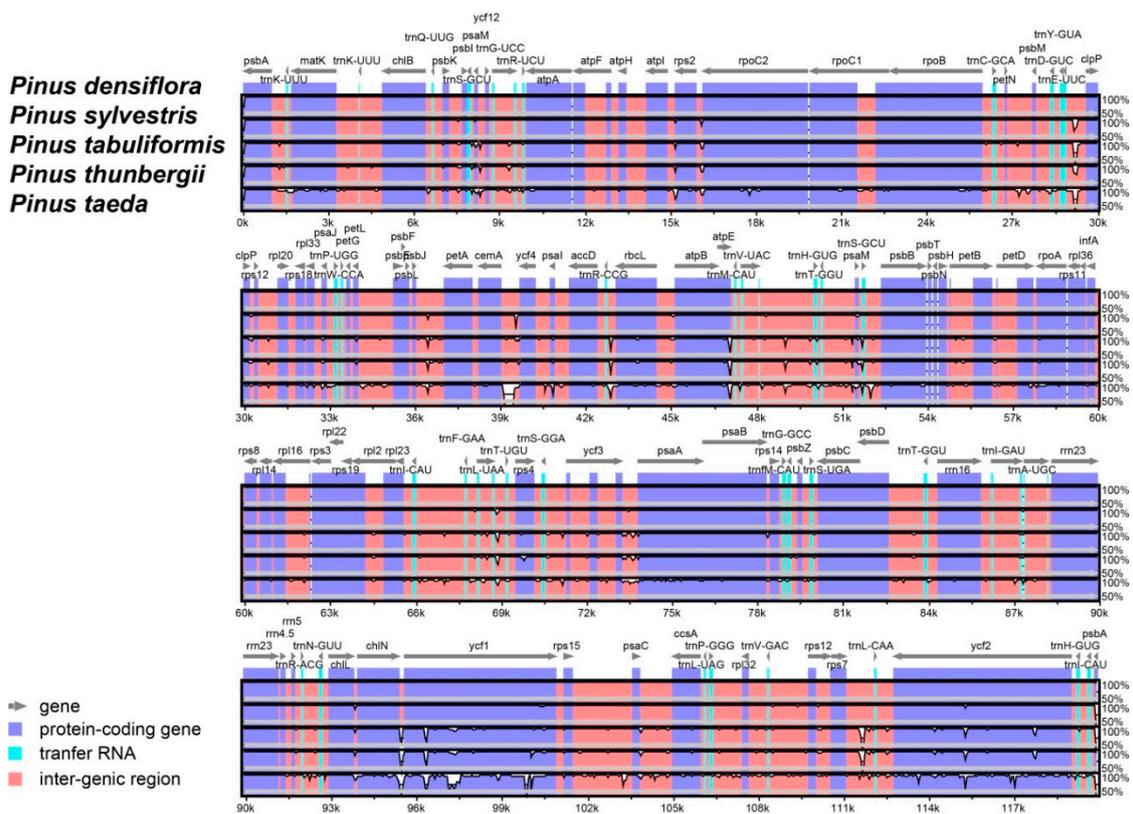
**Figure 1.** Gene map of the *P. densiflora* chloroplast genome. Genes drawn inside the circle are transcribed clockwise, while those drawn outside are transcribed counterclockwise. Different functional gene groups are color-coded. A GC-content graph is depicted within the inner circle. The circle inside the GC content graph marks the 50% threshold.

### 3.2. Comparative Analyses of the Chloroplast Genome with Other *Pinus* Species for the Identification of DNA Variation

The gene content and order of the *P. densiflora* chloroplast genome were nearly identical to the previously published chloroplast genomes of four *Pinus* species: *P. sylvestris* (KR476379), *P. thunbergii* (D17510), *P. tabuliformis* (KT740995), and *P. taeda* (KC427273), excepting the *ycf* genes and *psaM* gene duplication. Single nucleotide polymorphisms (SNPs) and insertion/deletion (InDel) variations among these species were typically located in the inter-genic regions (Figure 2). However, sequences showed low similarity to each other in the *ycf1* and *ycf2* regions. Also, a large insertion was found for *P. densiflora* between *trnE-UUC* and *clpP* in the LSC.

A detailed comparison of the border structure was performed in five *Pinus* chloroplast genomes (Figure 3). *P. densiflora* was most similar to *P. sylvestris*, which coincided with the pairwise alignment result. *P. taeda* had an IR structure that was significantly different from the other four species. For

*P. densiflora*, *P. sylvestris*, *P. thunbergii*, and *P. tabuliformis*, gene positioning at the IRs was stable, whereas SSC and LSC showed variable regions. IR regions had *trnI-CAU* and partial *psbA* in the IRa/LSC border, and both gene size and location were well-conserved. Genes adjacent to the border in the LSC and SSC regions were found to have 1–11 bp variations. The *trnK-UUU* gene, located in the LSC region, ranged from 1500 bp to 1511 bp away from the SSC/IRa border. The first exon of *rpl2* was located 1463–1468 bp upstream of the LSC/IRb border. *trnF-GAA* and *trnH-GUG*, in SSC, were located 1530–1537 bp and 144–169 bp away from the borders, respectively.

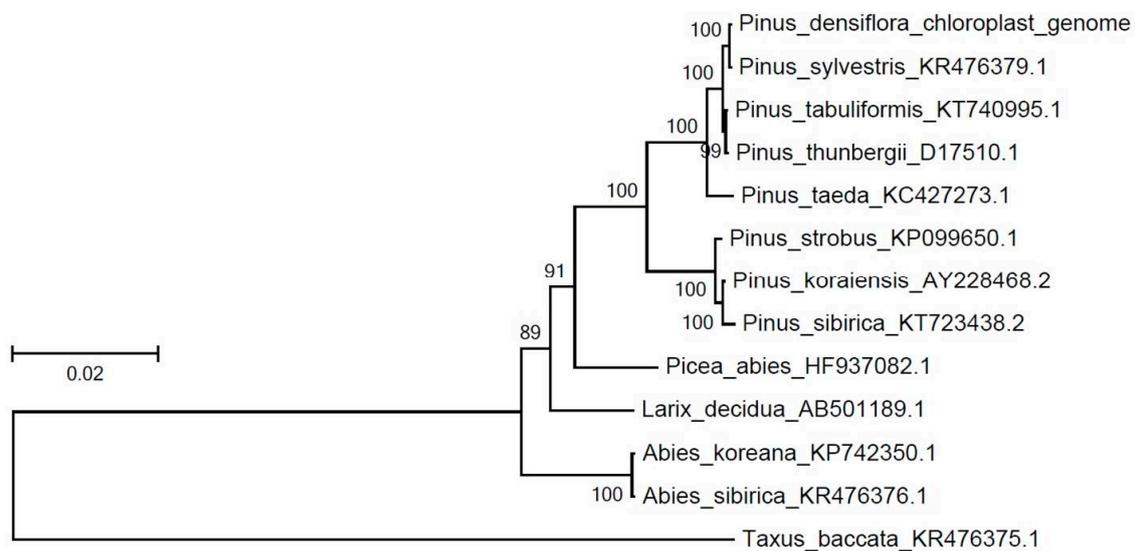


**Figure 2.** Pairwise alignments of chloroplast genome sequences of four *Pinus* species, each with that of *P. densiflora*.



**Figure 3.** Distance between adjacent genes and junctions of the small single copy (SSC), large single copy (LSC) and two inverted repeats (IRs) among plastid genomes from five *Pinus* species.

A phylogenetic tree based on the sequence alignment of the chloroplast genome of 12 Pinaceae and Taxaceae species showed that *P. densiflora* was most closely related to *P. sylvestris* (Figure 4, Table 4).



**Figure 4.** Phylogenetic tree based on protein coding genes (CDSs) of *P. densiflora* and 12 reference species. Bootstrap values (%) are shown above branches.

**Table 4.** Chloroplast genome comparison with 12 conifer species.

Species	Accession No.	No. of Protein Coding Genes	No. of Common Protein Coding Genes with <i>P. densiflora</i>
<i>Pinus sylvestris</i> Linn.	KR476379	73	73
<i>Pinus tabuliformis</i> Carr.	KT740995	74	73
<i>Pinus thunbergii</i> Parl.	D17510	69	69
<i>Pinus taeda</i> L.	KC427273	71	71
<i>Pinus strobus</i> Linn.	KP099650	-	-
<i>Pinus koraiensis</i> Sieb. et Zucc.	AY228468	74	72
<i>Pinus sibirica</i> (Loud.) Mayr	KT723438	77	73
<i>Picea abies</i> (L.) Karst.	HF937082	74	72
<i>Larix decidua</i> Mill	AB501189	72	71
<i>Abies koreana</i> Wils	KP742350	74	73
<i>Abies sibirica</i> Ledeb.	KR476376	74	73
<i>Taxus baccata</i> L.	KR476375	81	70

#### 4. Discussion and Conclusions

In this study, the complete chloroplast genome of *P. densiflora* was sequenced by ONT, annotated (Figure 1, Table 3) and compared with previously published conifer plastome sequences. We found gene content, order and intron structure were highly conserved among them.

We assembled a nearly complete chloroplast genome using only Oxford Nanopore sequencer data and error correction by MiSeq. The hybrid strategy of ONT combined with MiSeq was attempted for the first time in bacteria [23,24]. It is conventional to use the hybrid method for relatively short genomes, such as chloroplast DNA. This was well demonstrated in a previous study, wherein a 4.6 Mbp chromosome was sequenced essentially perfectly (>99.99% accuracy) [23]. The hybrid pipelines show higher accuracy than assembly using only MiSeq data [17,25]. Further, in the conventional method,

additional polymerase chain reaction (PCR) processes are typically performed to confirm short IRs. However, as sequencing has become more accurate with the combination of ONM and Illumina MiSeq, this additional experiment is no longer necessary [26].

The *ndh* genes encoding the NAD(P)H-dehydrogenase-like (NDH) are located in the nucleus, mitochondria, and chloroplast genomes. As is the case for Pinaceae species, the *ndh* genes were absent in the *P. densiflora* chloroplast genome (Table 3). Previous studies have suggested that the *ndh* genes have been transferred to the nuclear genome or left as pseudogenes in the chloroplast genome [27,28]. Thus, we suggest that the *ndh* genes missing from the chloroplast genome of *P. densiflora* have either been moved to the nuclear genome or remain in the chloroplast genome as one of five pseudogenes (*ndhB*, *ndhD*, *ndhE*, *ndhH*, and *ndhI*) (Figure 1).

There are extremely shortened IR regions in conifers, including *Pinus* [11,29]. In *P. densiflora*, we found large IRs had been replaced by two pairs of IRs highly reduced in size (495 bp and 738 bp), a finding which is consistent with previous studies [30]. We focused on shorter IRs for the comparative analysis. Rearrangements are frequently observed in species with loss of IRs, leading to interspecies variation [9]. Short IRs had some variation in size and sequence among the *Pinus* species (Table 1, Figure 2). It is presumed that loss of IRs occurred during the speciation of conifers from gymnosperm groups (cycads, conifers, *Ginkgo*, and Gnetales), as extant seed plants such as *Cycas taitungensis* (23 kbp), *Gnetum parvifolium* and *Ginkgo biloba* (17 kbp) have a large pair of IRs which are not found in Taxaceae, Pinaceae, and Cupressaceae [8,31]. With regard to the evolution of Cupressophytes, Li et al. argued that, after an inverted repeat was modified into a tandem repeat, the tandem repeat was divided by a rearrangement into two different parts to become short inverted repeats [32].

A lot of phylogenetic studies in land plants have used chloroplast genome sequences to analyze relatedness and classify the species [33–35]. Some CDSs (e.g., *matK*, *rbcL* and *rpoB*) and intergenic regions have been used as barcode markers in phylogenetic studies [36]. However, it is inappropriate to use a few markers when closely related species are classified because each marker only has a low amount of variability [37,38]. For example, the *matK* and *rpoB* genes of *P. densiflora* are identical to those of *P. sylvestris*. Moreover, we found there were only 151 and 72 parsimony-informative sites in the alignment of the 13 species investigated when using the *matK* and *rbcL* markers, respectively. Also, using CDSs was more exact than the whole genome sequence due to mis-alignments, such as structural variation and huge indels (Supplementary data1). In such a case, whole, commonly conserved CDSs that contain more information could be used in phylogenetic analyses [31]. The phylogenetic analysis of 59 CDSs was conducted in 12 Pinaceae species, using *Taxus baccata* (Taxaceae) as an outgroup. The number of parsimony-informative sites was 2,436 using 59 CDSs (Supplementary data2). Compared within the *Pinus* genus, *P. densiflora* was found to be most closely related to *P. sylvestris*. The composition and location of IR adjacent genes also supported the results (Figure 3). *P. sylvestris* and *P. densiflora* are known to be crossable [39]. Previous research has shown these species are very similar in their signal patterns in comparative analyses of their FISH karyotypes, which means they are closely related to each other [40]. *P. taeda* presented a different size IR structure, which is consistent with conventional plant taxonomic knowledge. The phylogenetic tree constructed in this study (depicted in Figure 4) revealed that the genetic distance of *P. taeda* is far from other *Pinus* species with two leaves [1,41]. The two groups (sect. *Pinus* and sect. *Cembra*) consisted of eight pine species, which supports the results of previous studies [1,42,43].

In this work, complete chloroplast sequencing using Oxford Nanopore Technology was newly attempted, assisting in the development of a method for obtaining more accurate sequences. The variations found in this study, which distinguish *P. densiflora* from other *Pinus* species, are useful for designing markers for specie identification. The complete chloroplast genome sequence of *P. densiflora* is also useful for understanding phylogenetic relationships among *Pinus* species.

**Supplementary Materials:** Supplementary materials can be found at <http://www.mdpi.com/1999-4907/10/7/600/s1>. Figure S1: Comparison of complete chloroplast genome of *P. densiflora* and initial assembled unitig2087 using BLASTZ program. Proximal regions of utg2087 were overlapped approximately 21.9 kb. (A) Initial assembled unitig2087 was used Nanopore reads. (B) Complete chloroplast genome of *P. densiflora* was corrected using Illumina paired-end reads. Figure S2: BLASTZ analysis of *P. densiflora* and reference species (*P. sylvestris*). Figure S3: The chloroplast genome sequencing error rate of Oxford Nanopore MinION reads. Supplementary data1: Multiple alignment sequence. Supplementary data2: Parsimony-informative sites for 59 protein coding genes.

**Author Contributions:** D.S., S.-W.L. and T.J.Y. conceived the conception and design of this study; H.-I.K. wrote the first draft of the manuscript; H.-I.K., H.O.L. performed the analysis and prepared figures and tables; I.H.L., I.S.K., S.-W.L., T.J.Y., and D.S. carefully checked and revised the manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** This work was supported by the National Institute of Forest Science, Republic of Korea (FG0400-2017-01).

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

## References

- Ni, Z.; Ye, Y.; Bai, T.; Xu, M.; Xu, L.A. Complete chloroplast genome of pinus massoniana (*Pinaceae*): Gene rearrangements, loss of *ndh* genes, and short inverted repeats contraction, expansion. *Molecules* **2017**, *22*, 1528. [[CrossRef](#)] [[PubMed](#)]
- Szmidt, A.E.; Wang, X.R. Molecular systematic and genetic differentiation of *Pinus sylvestris* (L.) and *P. densiflora* (sieb. et Zucc.). *Theor. Appl. Genet.* **1993**, *86*, 159–165. [[CrossRef](#)] [[PubMed](#)]
- Korea Forest Service. *Statistical Yearbook of Forestry*; Korea Forest Service: Daejeon, Korea, 2011.
- Duan, R.Y.; Yang, L.M.; Lv, T.; Wu, G.L.; Huang, M.Y. The complete chloroplast genome sequence of *Pinus dabeshanensis*. *Conserv. Genet. Resour.* **2016**, *8*, 395–397. [[CrossRef](#)]
- Asaf, S.; Waqas, M.; Khan, A.L.; Khan, M.A.; Kang, S.M.; Imran, Q.M.; Shahzad, R.; Bilal, S.; Yun, B.W.; Lee, I.J. The complete chloroplast genome of wild rice (*Oryza minuta*) and its comparison to related species. *Front. Plant Sci.* **2017**, *8*, 304. [[CrossRef](#)] [[PubMed](#)]
- Sugiura, M. The chloroplast genome. *Plant Mol. Biol.* **1992**, *19*, 149–168. [[CrossRef](#)] [[PubMed](#)]
- Palmer, J.D.; Thompson, W.F. Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. *Cell* **1982**, *29*, 537–550. [[CrossRef](#)]
- Lin, C.P.; Wu, C.S.; Huang, Y.Y.; Chaw, S.M. The complete chloroplast genome of *Ginkgo biloba* reveals the mechanism of inverted repeat contraction. *Genome Biol. Evol.* **2012**, *4*, 374–381. [[CrossRef](#)] [[PubMed](#)]
- Strauss, S.H.; Palmer, J.D.; Howe, G.T.; Doerksen, A.H. Chloroplast genomes of two conifers lack a large inverted repeat and are extensively rearranged. *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 3898–3902. [[CrossRef](#)]
- Perry, A.S.; Wolfe, K.H. Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. *J. Mol. Evol.* **2002**, *55*, 501–508. [[CrossRef](#)]
- Wu, C.S.; Wang, Y.N.; Hsu, C.Y.; Lin, C.P.; Chaw, S.M. Loss of different inverted repeat copies from the chloroplast genomes of *Pinaceae* and cupressophytes and influence of heterotachy on the evaluation of gymnosperm phylogeny. *Genome Biol. Evol.* **2011**, *3*, 1284–1295. [[CrossRef](#)] [[PubMed](#)]
- Zhang, Y.; Ma, J.; Yang, B.; Li, R.; Zhu, W.; Sun, L.; Tian, J.; Zhang, L. The complete chloroplast genome sequence of *Taxus chinensis* var. *mairei* (*Taxaceae*): Loss of an inverted repeat region and comparative analysis with related species. *Gene* **2014**, *540*, 201–209. [[CrossRef](#)] [[PubMed](#)]
- Heather, J.M.; Chain, B. The sequence of sequencers: The history of sequencing DNA. *Genomics* **2016**, *107*, 1–8. [[CrossRef](#)] [[PubMed](#)]
- McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernytzky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The genome analysis toolkit: A mapreduce framework for analyzing next-generation dna sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [[CrossRef](#)] [[PubMed](#)]
- Scarcelli, N.; Mariac, C.; Couvreur, T.L.; Faye, A.; Richard, D.; Sabot, F.; Berthouly-Salazar, C.; Vigouroux, Y. Intra-individual polymorphism in chloroplasts from NGS data: Where does it come from and how to handle it? *Mol. Ecol. Resour.* **2016**, *16*, 434–445. [[CrossRef](#)] [[PubMed](#)]
- Mikheyev, A.S.; Tin, M.M. A first look at the Oxford Nanopore MinION sequencer. *Mol. Ecol. Resour.* **2014**, *14*, 1097–1102. [[CrossRef](#)] [[PubMed](#)]

17. Lu, H.; Giordano, F.; Ning, Z. Oxford Nanopore MinION sequencing and genome assembly. *Genom. Proteom. Bioinform.* **2016**, *14*, 265–279. [[CrossRef](#)] [[PubMed](#)]
18. Lowe, T.M.; Eddy, S.R. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **1997**, *25*, 955–964. [[CrossRef](#)]
19. Laslett, D.; Canback, B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* **2004**, *32*, 11–16. [[CrossRef](#)]
20. Tillich, M.; Lehwark, P.; Pellizzer, T.; Ulbricht-Jones, E.S.; Fischer, A.; Bock, R.; Greiner, S. GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **2017**, *45*, W6–W11. [[CrossRef](#)]
21. Mayor, C.; Brudno, M.; Schwartz, J.R.; Poliakov, A.; Rubin, E.M.; Frazer, K.A.; Pachter, L.S.; Dubchak, I. VISTA: Visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* **2000**, *16*, 1046–1047. [[CrossRef](#)]
22. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [[CrossRef](#)]
23. Goodwin, S.; Gurtowski, J.; Ethe-Sayers, S.; Deshpande, P.; Schatz, M.C.; McCombie, W.R. Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. *Genome Res.* **2015**, *25*, 1750–1756. [[CrossRef](#)]
24. Laver, T.; Harrison, J.; O'Neill, P.A.; Moore, K.; Farbos, A.; Paszkiewicz, K.; Studholme, D.J. Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomol. Detect Quantif.* **2015**, *3*, 1–8. [[CrossRef](#)]
25. Giordano, F.; Aigrain, L.; Quail, M.A.; Coupland, P.; Bonfield, J.K.; Davies, R.M.; Tischler, G.; Jackson, D.K.; Keane, T.M.; Li, J.; et al. De novo yeast genome assemblies from MinION, PacBio and MiSeq platforms. *Sci. Rep.* **2017**, *7*, 3935. [[CrossRef](#)]
26. Ishizuka, W.; Tabata, A.; Ono, K.; Fukuda, Y.; Hara, T. Draft chloroplast genome of *Larix gmelinii* var. *japonica*: Insight into intraspecific divergence. *J. For. Res.* **2017**, *22*, 393–398. [[CrossRef](#)]
27. Ranade, S.S.; Garcia-Gil, M.R.; Rossello, J.A. Non-functional plastid *ndh* gene fragments are present in the nuclear genome of Norway spruce (*Picea abies* L. Karsch): Insights from in silico analysis of nuclear and organellar genomes. *Mol. Genet. Genom.* **2016**, *291*, 935–941. [[CrossRef](#)]
28. Lin, C.S.; Chen, J.J.; Chiu, C.C.; Hsiao, H.C.; Yang, C.J.; Jin, X.H.; Leebens-Mack, J.; de Pamphilis, C.W.; Huang, Y.T.; Yang, L.H.; et al. Concomitant loss of NDH complex-related genes within chloroplast and nuclear genome in some orchids. *Plant J.* **2017**, *90*, 994–1006. [[CrossRef](#)]
29. Raubeson, L.A.; Jansen, R.K. A rare chloroplast-DNA structural mutation is shared by all conifers. *Biochem. Syst. Ecol.* **1992**, *20*, 17–24. [[CrossRef](#)]
30. Wu, C.S.; Lin, C.P.; Hsu, C.Y.; Wang, R.J.; Chaw, S.M. Comparative chloroplast genomes of pinaceae: Insights into the mechanism of diversified genomic organizations. *Genome Biol. Evol.* **2011**, *3*, 309–319. [[CrossRef](#)]
31. Wu, C.S.; Wang, Y.N.; Liu, S.M.; Chaw, S.M. Chloroplast genome (cpDNA) of *Cycas taitungensis* and 56 cp protein-coding genes of *Gnetum parvifolium*: Insights into cpDNA evolution and phylogeny of extant seed plants. *Mol. Biol. Evol.* **2007**, *3*. [[CrossRef](#)]
32. Li, J.; Gao, L.; Chen, S.; Tao, K.; Su, Y.; Wang, T. Evolution of short inverted repeat in cupressophytes, transfer of *accD* to nucleus in *Sciadopitys verticillata* and phylogenetic position of *Sciadopityaceae*. *Sci. Rep.* **2016**, *6*, 20934. [[CrossRef](#)]
33. Moore, M.J.; Soltis, P.S.; Bell, C.D.; Burleigh, J.G.; Soltis, D.E. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 4623–4628. [[CrossRef](#)]
34. Olsson, S.; Grivet, D.; Vian, J.C. Species-diagnostic markers in the genus *Pinus*: Evaluation of the chloroplast regions *matK* and *ycf1*. *For. Syst.* **2018**, *27*, 2. [[CrossRef](#)]
35. Wu, F.; Li, M.; Liao, B.; Shi, X.; Xu, Y. DNA Barcoding Analysis and Phylogenetic Relation of Mangroves in Guangdong Province, China. *Forests* **2019**, *10*, 56. [[CrossRef](#)]
36. Cbol Plant Working Group. A DNA barcode for land plants. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 12794–12797. [[CrossRef](#)]
37. Snel, B.; Bork, P.; Huynen, M.A. Genome phylogeny based on gene content. *Nat. Genet.* **1999**, *21*, 108. [[CrossRef](#)]
38. Nishiyama, T.; Wolf, P.G.; Kugita, M.; Sinclair, R.B.; Sugita, M.; Sugiura, C.; Wakasugi, T.; Yamada, K.; Yoshinaga, K.; Yamaguchi, K.; et al. Chloroplast phylogeny indicates that bryophytes are monophyletic. *Mol. Biol. Evol.* **2004**, *21*, 1813–1819. [[CrossRef](#)]

39. Kosinski, G. Hybridization. In *Genetics of Scots Pine*; Giertych, M., Matyas, C., Eds.; Elsevier: Amsterdam, The Netherlands, 1991; pp. 183–189.
40. Hizume, M.; Shibata, F.; Matsusaki, Y.; Garajova, Z. Chromosome identification and comparative karyotypic analyses of four *Pinus* species. *Appl. Genet.* **2002**, *105*, 491–497. [[CrossRef](#)]
41. Eckert, A.J.; Benjamin, D.H. Phylogeny, historical biogeography, and patterns of diversification for *Pinus* (Pinaceae): Phylogenetic tests of fossil-based hypotheses. *Mol. Phylogenet. Evol.* **2006**, *40*, 166–182. [[CrossRef](#)]
42. Parks, M.; Cronn, R.; Liston, A. Separating the wheat from the chaff: Mitigating the effects of noise in a plastome phylogenomic data set from *Pinus* L. (*Pinaceae*). *BMC Evol. Biol.* **2012**, *12*, 100. [[CrossRef](#)]
43. Asaf, S.; Khan, A.L.; Khan, M.A.; Shahzad, R.; Lubna; Kang, S.M.; Al-Harrasi, A.; Al-Rawahi, A.; Lee, I.J. Complete chloroplast genome sequence and comparative analysis of loblolly pine (*Pinus taeda* L.) with related species. *PLoS ONE* **2018**, *13*, e0192966. [[CrossRef](#)]



© 2019 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 (<http://creativecommons.org/licenses/by/4.0/>).