Research Article

The Complete Chloroplast Genome of *Arabidopsis thaliana* Isolated in Korea (Brassicaceae): An Investigation of Intraspecific Variations of the Chloroplast Genome of Korean *A. thaliana*

Jongsun Park,†,1,2 Hong Xi,1,2 and Yongsung Kim1,2

1InfoBoss Inc., 301 Room, 670, Seolleung-ro, Gangnam-gu, Seoul, Republic of Korea
2InfoBoss Research Center, 301 Room, 670, Seolleung-ro, Gangnam-gu, Seoul, Republic of Korea

Correspondence should be addressed to Jongsun Park; starflr@infoboss.co.kr

Received 16 June 2020; Revised 2 August 2020; Accepted 17 August 2020; Published 7 September 2020

Copyright © 2020 Jongsun Park et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Arabidopsis thaliana* (L.) Heynh. is a model organism of plant molecular biology. More than 1,700 whole genome sequences have been sequenced, but no Korean isolate genomes have been sequenced thus far despite the fact that many *A. thaliana* isolated in Japan and China have been sequenced. To understand the genetic background of Korean natural *A. thaliana* (named as 180404IB4), we presented its complete chloroplast genome, which is 154,464 bp long and has four subregions: 85,164 bp of large single copy (LSC) and 17,781 bp of small single copy (SSC) regions are separated by 26,257 bp of inverted repeat (IRs) regions including 130 genes (85 protein-coding genes, eight rRNAs, and 37 tRNAs). Fifty single nucleotide polymorphisms (SNPs) and 14 insertion and deletions (INDELs) are identified between 180404IB4 and Col0. In addition, 101 SSRs and 42 extended SSRs were identified on the Korean *A. thaliana* chloroplast genome, indicating a similar number of SSRs on the rest five chloroplast genomes with a preference of sequence variations toward the SSR region. A nucleotide diversity analysis revealed two highly variable regions on *A. thaliana* chloroplast genomes. Phylogenetic trees with three more chloroplast genomes of East Asian natural isolates show that Korean and Chinese natural isolates are clustered together, whereas two Japanese isolates are not clustered, suggesting the need for additional investigations of the chloroplast genomes of East Asian isolates.

1. Introduction

*Arabidopsis thaliana* (L.) Heynh. is a well-known species familiar to those who study plant molecular biology as well as genetic engineering. It was considered to be a weedy species before being used as a model organism [1], representing a good example of the usefulness of weeds. Owing to its importance as a model plant organism, its complete chloroplast genome of Col0 strain was deciphered in 1999 [2]. Its length is 154,478 bp, with a large single copy (LSC) region of 84,170 bp and a small single copy (SSC) region of 17,780 bp separated by inverted repeat (IR; 26,264 bp) regions. It was also found to have 128 genes consisting of 87 protein-coding, 37 tRNA, and four rRNA genes.

Consequently, the whole genome sequences of *A. thaliana* were released in 2000, presenting a 115.4Mb genome with 25,498 genes [3]. Owing to the rapid development of sequencing technologies, including next-generation sequencing (NGS) technologies, more than 1,750 *A. thaliana* whole genome sequences have been sequenced and analyzed; whole genomes of the Bur-0 and Tsu-1 strains were sequenced with an early version of the Illumina sequencer [4]. The genomes of ebi-1 and Ws-2, which are clock mutants, were also sequenced [5]. Whole genomes of 80 strains isolated from eight regions were also sequenced [6]. In addition, whole genomes of 18 *Arabidopsis* ecotypes were sequenced along with providing assembled sequences for each ecotype, which can be used in further comparative genomic analyses [7]. However, the organelle genomes of the 8 ecotype genomes were not correctly assembled, though these can be rectified using raw sequences. With additional improvements in NGS technologies that have lowered the costs of sequencing, the number of sequenced *A. thaliana* isolates has been increased over time. Specifically, genomes of 180 *Arabidopsis*
lines isolated from Sweden were sequenced [8]. 217 Arabidopsis individual genomes to uncover genome-wide methylation patterns were also sequenced [9], and the genomes of 118 Chinese Arabidopsis strains were sequenced showing that the Yangzte River population in China can be considered as an independent lineage on the same level of Central Asian and European isolates [10], and 1,135 natural isolates from all over the world were sequenced and analyzed [11]. Interestingly, except for one study on the sequencing Chinese Arabidopsis strains [10], only two Arabidopsis strains originated from East Asia have been sequenced (Kyoto and Tsu0 from Japan) [11]. Therefore, once any genome sequences of Arabidopsis isolated in Korea are available, they can serve as a bridge connecting between China and Japan as part of the effort to construct the evolution history of natural isolates of A. thaliana in East Asia.

Although more than 1,750 Arabidopsis genomes have been sequenced, chloroplast genomes from these strains have not been assembled to ascertain the sequence variations of the chloroplast genome except for Ler-0 [12]. This is likely to be due to the fact that the chloroplast genome does not contain enough information to attract researchers comparing with whole genome sequences. However, chloroplast genomes are occasionally useful for unravelling corresponding phylogenetic relationships based on the maternal lineage: e.g., Lindera genus [13], Fagopyrum genus [14], Potentilla genus [15–21], Pseudostellaria genus [22–26], and Dysphania genus [27–30]. Moreover, several studies that described assembled organelle genomes from NGS raw sequences generated with different purposes have been conducted [31, 32].

To understand characteristics of A. thaliana isolated in Korea (termed 1804041B4) based on chloroplast genome sequences, we completed its chloroplast genome, presenting the third chloroplast genome of A. thaliana based on the NCBI database and related publications [2, 12]. The chloroplast genome of A. thaliana 1804041B4 presented the shortest total length and the shortest inverted repeat (IR) region length, caused by a 6 bp deletion on ycf2. Due to the limitation of available chloroplast genomes of A. thaliana (only two are available in NCBI: Col0 and Ler-0), we assembled three additional chloroplast genomes of East Asian isolates of A. thaliana from raw sequences downloaded from the Short Read Archive (SRA) in NCBI, indicating that Tsu0 (Japanese isolate) had the longest length due to an approximately 500 bp insertion. Numbers of sequence variations calculated based on the 1804041B4 chloroplast genome are in the middle of the numbers of intraspecific variations of many plant chloroplast genomes. Phylogenetic analyses of these chloroplast genomes indicate that 1804041B4 (Korea) and 11-15 (Chinese) are clustered, whereas two Japanese isolates (Kyoto and Tsu0) are scattered. These results will provide a glimpse of the evolutionary history of A. thaliana in the East Asian area region together with upcoming research results.

2. Materials and Methods

2.1. DNA Extraction of Korean A. thaliana Natural Isolate. The Korean A. thaliana natural isolate was collected in Yeonggwang-gun, Junranam Province, South Korea (35.242862N, 126.508987E; 1804041B4; Y. Kim, IB-00583, InfoBoss Cyber Herbarium (IN)) in 2018. Its total DNA was extracted from fresh leaves using a DNeasy Plant Mini Kit (QIAGEN, Hilden, Germany).

2.2. Genome Sequencing and De Novo Assembly of the Korean Natural Isolate A. thaliana Chloroplast Genome. Genome sequencing was performed using HiSeqX at Macrogen Inc., Korea, from the extracted DNA of the Korean A. thaliana de novo assembly, with confirmation accomplished with Velvet 1.2.10 [33] after filtering raw reads using Trimmomatic 0.33 [34]. After obtaining the first draft of the chloroplast genome sequences, gaps were filled with GapCloser 1.12 [35] and all bases from the assembled sequences were confirmed by checking each base in the alignment (view mode in SAMtools 1.9 [36]) against the assembled chloroplast genome generated with BWA 0.7.17 [37]. All these bioinformatic processes were conducted under the environment of Genome Information System (GeiS; http://geis.infoboss.co.kr/; Park et al., in preparation).

2.3. De Novo Assembly of the Japanese and Chinese Natural Isolates A. thaliana Chloroplast Genomes. Raw sequences downloaded from NCBI SRA (SRR492307; Kyoto (Japan), ERR031555; Tsu0 (Japan) and SRR2204166; 15-11 (China)) [7, 9, 10] were used for chloroplast de novo genome assembly with Velvet 1.2.10 [33] after filtering raw reads using Trimmomatic 0.33 [34] under the environment of Genome Information System (GeiS; http://geis.infoboss.co.kr/; Park et al., in preparation). The remaining steps for finalizing the chloroplast genomes are identical to those used in the assembly process of the Korean A. thaliana.

2.4. Chloroplast Genome Annotation. Geneious R11 11.0.5 (Biomatters Ltd., Auckland, New Zealand) was used for chloroplast genome annotation based on the A. thaliana chloroplast genome (NC_000921) [2] by transferring annotations while correcting exceptional cases, including missing start or stop codons. tRNA was predicted and confirmed using tRNAscan-SE [38].

2.5. Identification of Sequence Variations on Arabidopsis Complete Chloroplast Genomes. Single nucleotide polymorphisms (SNPs) and insertions and deletions (INDELs) were identified from the pairwise alignment of two chloroplast genomes conducted by MAFFT 7.450 [39] in the environment of the Plant Chloroplast Database (PCD; http://www.cp-genome.net/).

2.6. Identification of Simple Sequence Repeats (SSRs). Simple sequence repeats (SSRs) were identified on the chloroplast genome sequence using the pipeline of the SSR database (SSRDB; http://ssr.pe.kr/; Park et al., in preparation). Based on the conventional definition of a SSR on a chloroplast genome, monoSSR (1 bp) to hexaSSR (6 bp), the total length of SSRs on the chloroplast genome exceeds 10 bp. Owing to the different criteria of SSRs on chloroplast genomes [26, 40–45], we adopted the criteria used in chloroplast genome
of *Dysphania ambrosioides* [30], as follows: the monoSSR (unit sequence length of 1 bp) to hexaSSR (6 bp) are used as normal SSRs, and heptaSSR (7 bp) to decaSSR (10 bp) are defined as extendedSSRs. Among the normal SSRs, pentaSSRs and hexaSSRs for which the number of unit sequences is 2 are classified as potentialSSRs.

2.7. Comparison of SSRs Identified from Six *A. thaliana* Chloroplast Genomes. SSRs identified from six *A. thaliana* chloroplast genomes were compared based on their flanking sequences under the environment of the SSRDB (http://ssr.pe.kr; Park et al., in preparation). The pipeline of the SSR comparison implemented in the SSRDB was used with the following conditions: a cut-off e value of $1 \times 10^{-10}$ and a maximum flanking sequence for the comparison of 60 bp. This comparison was utilized in a comparative analysis of *Chenopodium* chloroplast genomes (Park et al., in preparation) and *Stegobium paniceum* (Park et al., under revision) and *Ficus binodulus* (Lee et al., [46]) mitochondrial genomes.

2.8. Nucleotide Diversity Analysis. Nucleotide diversity was calculated using the method proposed by Nei and Li [47] based on the multiple sequence alignment of *Arabidopsis* chloroplast genomes using a Perl script. The window size was set to 500 bp and the step size was 200 bp when using the sliding window method. Genomic coordination of each
window was compared to the gene annotation of the chloroplast genome under the environment of PCD for further analyses.

2.9. Construction of Phylogenetic Trees. Whole chloroplast genomes of seventeen Arabidopsis genomes and one Arabis chloroplast genome were aligned by MAFFT 7.450 [39], and alignment quality was checked manually. The neighbor-joining (N) and maximum likelihood (ML) trees were reconstructed in MEGA X [48]. In the ML analysis, a heuristic search was used with nearest-neighbor interchange (NNI) branch swapping, Tamura-Nei model, and uniform rates among sites. All other options used the default settings. Bootstrap analyses with 1,000 pseudoreplicates were conducted with the same options. The posterior probability of each node was estimated by Bayesian inference (BI) using the Mr. Bayes 3.2.6 [49] plug-in implemented in Geneious R11 11.0.5. The HKY85 model with gamma rates was used as a molecular model. A Markov chain Monte Carlo (MCMC) algorithm was employed for 1,100,000 generations, with four chains run-ning simultaneously. Trees from the first 100,000 generations were discarded as burn-in.

3. Results and Discussions

3.1. Complete Chloroplast Genome of the First Korean Isolate of A. thaliana and Comparison with A. thaliana Chloroplast Genomes Assembled from NGS Raw Reads. The chloroplast genome of A. thaliana 180404IB4 (GenBank accession number is MK353213) is 154,464 bp and has four subregions with 85,164 bp of large single copy (LSC) and 17,781 bp of small single copy (SSC) regions separated by 26,257 bp of inverted repeat (IR; Figure 1). It contains 130 genes (85 protein-coding genes, eight rRNAs, and 37 tRNAs), with 19 genes (8 protein-coding genes, 4 rRNAs, and 7 tRNAs) that are duplicated in IR regions (Figure 1). The overall GC content is 36.3%, and those contents in the LSC, SSC, and IR regions are 34.0%, 29.3%, and 42.3%, respectively.

To determine the characteristics of A. thaliana chloroplast genomes from East Asia, we also completed three chloroplast genomes of A. thaliana, two from Japan (Tokyo and Tsu0) and one from China (11-15; Table 1). In addition, two available chloroplast genomes (Col0 and Ler-0) were used for comparative analyses. Their lengths range from 154,464 bp (180404IB4) to 154,938 bp (Tsu0) and the IR length ranges from 26,257 bp to 26,264 bp (Table 1). The chloroplast genome of the Korean isolate, 180404IB4, is the shortest, and its IR is also the shortest among four East Asian A. thaliana chloroplasts (Table 1). It is caused by 6 bp deletion on ycf2 located in the IR region compared to the rest five A. thaliana chloroplast genomes. Interestingly, Tsu0 shows the longest length of chloroplast genome, caused by an insertion of approximately 500 bp between trnL and trnF, which is similar to the cases of Coffea arabica with one continuous insertion region [50], Duchesnea chrysantha showing three continuous insertion regions [21], Viburnum amphilacticum showing two continuous insertion regions [32], and mitochondrial genomes of Populus tremula x Populus glandulosa and Liriodendron tulipifera with four and thirty-three continuous insertion regions, respectively [51, 52].

The GC content of the six complete chloroplast genomes of the Arabidopsis isolates is 36.3% and those of LSC, SSC, and IR are 34.0%, 29.3%, and 42.0%, respectively. An exception is GC content of the LSC of Tsu0, which is 34.1% (Table 1). This is also caused by the inserted region of the Tsu0 chloroplast genome. Other plant chloroplasts of which intraspecific variations of the GC contents are identical to those of A. thaliana are Goodyera schlechtendaliana (37.1% and 37.2%) [53–55] and Gastrodia elata (26.7% and 26.8%) [56–58] which are same to those of Arabidopsis thaliana, while Coffea arabica [50, 59–63], Viburnum erosum [64, 65], Duchesnea chrysanthà [20, 21], Salix koryanaei [66, 67], Pseudostellaria palibiniana [23, 25], and Puerus ussuriensis [68, 69] present no difference in the intraspecific GC contents.

| Strain name | Country | GenBank accession | Whole Length (bp) | LSC Length (bp) | SSC Length (bp) | IR Length (bp) | Whole GC contents | LSC GC contents | SSC GC contents | IR GC contents |
|-------------|---------|-------------------|------------------|-----------------|-----------------|----------------|------------------|----------------|----------------|---------------|
| 180404IB4   | Korea   | MK353213          | 154,464          | 84,170          | 17,781          | 26,257         | 36.3%            | 34.0%          | 29.3%          | 42.3%         |
| Kyoto       | Japan   | MK380720          | 154,470          | 84,162          | 17,781          | 26,264         | 36.3%            | 34.0%          | 29.3%          | 42.3%         |
| Tsu0        | Japan   | MK380721          | 154,938          | 84,630          | 17,781          | 26,264         | 36.3%            | 34.1%          | 29.3%          | 42.3%         |
| 11-15       | China   | MK380719          | 154,487          | 84,179          | 17,781          | 26,264         | 36.3%            | 34.0%          | 29.3%          | 42.3%         |
| Col0        | USA     | NC_000932         | 154,478          | 84,170          | 17,781          | 26,264         | 36.3%            | 34.0%          | 29.3%          | 42.3%         |
| Ler-0       | Germany | KX551970          | 154,515          | 84,213          | 17,780          | 26,261         | 36.3%            | 34.0%          | 29.3%          | 42.3%         |

Table 1: List of chloroplast genomes of East Asian Arabidopsis strains.
| No | Type | Genomic coordination | Gene name | Base changes | Amino acid changes |
|----|------|----------------------|-----------|--------------|-------------------|
| 1  | SNP  | 4,370                | Intergenic space | T to A | - |
| 2  | SNP  | 5,705                | Intron of rps16 | C to A | - |
| 3  | SNP  | 5,943                | Intron of rps16 | G to A | - |
| 4  | SNP  | 8,502                | Intergenic | T to A | - |
| 5  | SNP  | 12,673               | Exon of atpF  | G to A | Synonymous SNPs |
| 6  | SNP  | 14,825               | Intergenic | T to A | |
| 7  | SNP  | 16,380               | Exon of rpoC2 | G to C | C to S |
| 8  | SNP  | 17,801               | Exon of rpoC2 | T to C | Synonymous SNP |
| 9  | SNP  | 17,823               | Exon of rpoC2 | T to G | K to T |
| 10 | SNP  | 23,915               | Exon of rpoB  | T to G | Synonymous SNP |
| 11 | SNP  | 26,817               | Exon of rpoB  | G to C | Synonymous SNP |
| 12 | SNP  | 26,999               | Intergenic | T to G | |
| 13 | SNP  | 27,538               | Intergenic | C to A | |
| 14 | SNP  | 28,360               | Intergenic | T to G | |
| 15 | SNP  | 31,912               | Intergenic | A to C | |
| 16 | SNP  | 35,167               | Intergenic | G to A | |
| 17 | SNP  | 36,798               | Intergenic | T to C | |
| 18 | SNP  | 45,107               | Intergenic | G to A | |
| 19 | SNP  | 47,760               | Intergenic | T to G | |
| 20 | SNP  | 49,938               | Intergenic | T to A | |
| 21 | SNP  | 50,606               | Intergenic | T to A | |
| 22 | SNP  | 50,688               | Intergenic | T to A | |
| 23 | SNP  | 56,373               | Exon of rbcL1 | C to T | Synonymous SNP |
| 24 | SNP  | 58,860               | Intergenic | G to C | |
| 25 | SNP  | 65,250               | Intergenic | T to C | |
| 26 | SNP  | 65,545               | Intergenic | T to G | |
| 27 | SNP  | 66,114               | Intergenic | C to A | |
| 28 | SNP  | 66,810               | Intergenic | T to C | |
| 29 | SNP  | 67,142               | Intergenic | T to A | |
| 30 | SNP  | 67,143               | Intergenic | G to A | |
| 31 | SNP  | 67,145               | Intergenic | T to A | |
| 32 | SNP  | 67,146               | Intergenic | G to C | |
| 33 | SNP  | 67,147               | Intergenic | T to A | |
| 34 | SNP  | 67,149               | Intergenic | T to C | |
| 35 | SNP  | 67,150               | Intergenic | T to A | |
| 36 | SNP  | 68,722               | Exon of rpl20 | T to C | Synonymous SNP |
| 37 | SNP  | 69,347               | Intergenic | G to A | |
| 38 | SNP  | 73,507               | Exon of psbB  | T to G | Synonymous SNP |
| 39 | SNP  | 75,386               | Intergenic | G to A | |
| 40 | SNP  | 77,830               | Intergenic | G to A | |
| 41 | SNP  | 80,688               | Intergenic | T to G | |
| 42 | SNP  | 91,824               | Exon of ycf2  | G to C | R to T |
| 43 | SNP  | 96,023               | Intron of ndhB | T to G | |
| 44 | SNP  | 111,048              | Exon of ndhF  | T to C | G to R |
| 45 | SNP  | 113,690              | Intergenic | T to G | |
| 46 | SNP  | 113,770              | Intergenic | T to G | |
| 47 | SNP  | 113,957              | Intergenic | C to A | |
| 48 | SNP  | 115,487              | Intergenic | T to G | |
Table 2: Continued.

| No | Type | Genomic coordination | Gene name         | Base changes | Amino acid changes |
|----|------|----------------------|-------------------|--------------|--------------------|
| 49 | SNP  | 142,626              | Intron of ndhB    | T to G       |                    |
| 50 | SNP  | 146,825              | Exon of ndhF      | T to C       | G to R             |
| 51 | INDEL| 92,548               | Exon of ycf2      | T to -       |                    |
| 52 | INDEL| 92,549               | Exon of ycf2      | G to -       |                    |
| 53 | INDEL| 92,550               | Exon of ycf2      | A to -       |                    |
| 54 | INDEL| 92,551               | Exon of ycf2      | T to -       |                    |
| 55 | INDEL| 92,552               | Exon of ycf2      | A to -       |                    |
| 56 | INDEL| 92,553               | Exon of ycf2      | A to -       |                    |
| 57 | INDEL| 108,115              | Intergenic        | A to -       |                    |
| 58 | INDEL| 130,525              | Intergenic        | T to -       |                    |
| 59 | INDEL| 146,092              | Exon of ycf2      | T to -       |                    |
| 60 | INDEL| 146,093              | Exon of ycf2      | T to -       |                    |
| 61 | INDEL| 146,094              | Exon of ycf2      | C to -       |                    |
| 62 | INDEL| 146,095              | Exon of ycf2      | A to -       |                    |
| 63 | INDEL| 146,096              | Exon of ycf2      | T to -       |                    |
| 64 | INDEL| 146,097              | Exon of ycf2      | T to -       |                    |

To evaluate the degree of these sequence variations including SNPs and INDELs, we investigated studies of intraspecific variations on chloroplast genomes that checked numbers of SNPs and INDELs (Table 3). Some of these studies are compared chloroplast genomes of natural isolates (e.g., Duchesnea chrysantha [21]) and some compared among cultivars to find useful molecular markers (e.g., Che- npodium quinoa [70]; Table 3). These studies cover 23 families showing relatively large coverage, so that we expected that some characteristics of these sequence variations on chloroplast genomes can be rescued. In addition, we used number of SNPs and INDELs directly during comparison of sequence variations for better understanding intuitively because their complete chloroplast genome lengths are around 150 kb except genera Marchantia, Selaginella, Gastrodia, Illicium, Pseudostellaria, and Daphne (Table 3).

The numbers of SNPs and INDELs, 50 and 14, respectively, between 180404IB4 and Col0 are smaller than those of Marchantia polymorpha between Korea and Poland (69 SNPs and 660 INDELs) [71], Camellia japonica (78 SNPs and 643 INDELs) [72], Rehmannia glutinosa (147 SNPs and 87 INDELs) [73], and Selaginella tamariscina (1,213 SNPs and 1,641 INDELs) [74] between Chinese and Korean isolates, Pseudostellaria palibiniana (84 SNPs and 125 INDELs) [25] and Pyrus ussuriensis inside Korea (121 SNPs and 781 INDELs) [68], Eucommia ulmoides inside China (75 SNPs and 80 INDELs) [75], some cases of Cucumis melo [76] and Chenopodium quinoa [70], all of Dioscorea polystachya [77], Oryza sativa among cultivars [78], G. schlechten- daliana [53], and G. elata [56] (Table 2). Some of species, such as Potentilla (49 SNPs and 17 INDELs are identified from Potentilla stolonifera var. quelpeaensis and Potentilla stolonifera var. chejuensis) [16, 18], present numbers of sequence variations similar to that of A. thaliana between 180404IB4 and Col0. Considering other cases involving the number of intraspecific variations on the chloroplast genome, including Salix (40 SNPs and 139 INDELs between Salix koriyamai and Salix gracilistyla) [66, 79], Ilex (55 SNPs and 429 INDELs between Ilex cornuta and Ilex integra) [15, 80], and Nymphaea (586 SNPs and 1,150 INDELs between Nymphaea capensis and Nymphaea ampla) [21], no clear levels pertaining to the number of intraspecific or interspecific variations exist. However, the numbers of SNPs and INDELs between 180404IB4 and Col0 are relatively small considering the intercontinental distance between two samples of the same plant species.

3.3. Comparison and Evaluation of Sequence Variations of Chloroplast Genomes of the Six A. thaliana in East Asia. Based on 15 pairwise alignments of the six A. thaliana chloroplast genomes, the numbers of SNPs and INDELs between two A. thaliana chloroplast genomes range from 10 to 116 and from 22 to 570, respectively (Figure 2). The Tsu0, Japanese natural isolate, chloroplast genome contains large insertions compared to the remaining five chloroplast genomes of A. thaliana, supported by the largest Tsu0 chloroplast genome (Table 1). The number of INDELs compared to the Tsu0 chloroplast genome (GenBank accession number is MK380721) ranges from 470 to 570, much higher than those of other combinations (Figure 2). This case is similar to those of C. arabica, showing one 84 bp insertion region [50] and D. chrysantha, presenting three insertion regions [21]. In terms of the number of INDELs, it is also in relation to high intraspecific variations that only P. ussuriensis [68], G. schlechten- daliana [53], and G. elata [56] present higher numbers of INDELs (Table 3). In addition, two out of the three Orchidaceae species shows high rates of divergence in terms of flower morphologies as well as the number of species [81–83]. This indicates that the Tsu0 insertion is an exceptional case of intraspecific variation. Consequently, Kyoto (GenBank accession number is MK380720), which was also isolated in Japan, and Tsu0 correspondingly present 97 SNPs and 482
| Family            | Species                          | Total length (bp) | # of SNPs | # of INDELs | Comparison samples                                                                 | Ref.     |
|-------------------|----------------------------------|-------------------|-----------|-------------|-----------------------------------------------------------------------------------|----------|
| Marchantiaceae    | Marchantia polymorpha            | 120,288–120,304   | 4         | 0           | Samples isolated in Korea                                                          | [71]     |
|                   |                                  |                   | 69        | 120,304     | Takaragaite-1 (Japan)                                                              |          |
|                   |                                  |                   | 660       |             | Kitashirakawa-2 (Japan)                                                            |          |
| Selaginellaceae   | Selaginella tamariscina           | 126,368–126,700   | 1,213     | 1,641       | Sample isolate in Korea                                                            | [74]     |
|                   | Selaginella uncinata             | 144,161–144,710   | 525       | 602         | Sample isolate in Korea                                                            |          |
|                   |                                  |                   |           |             | Sample isolate in Japan                                                             |          |
| Magnoliaceae      | Magnolia kobus                   | 159,411–159,443   | 50        | 50          | Sample isolate in Korea (Jeju Island)                                              | In prep. |
|                   | Liriodendron tulipifera          | 159,886           | 12        | 0           | Sample isolate in Korea                                                            | [52]     |
| Nymphaeaceae      | Nymphaea alba                    | 159,250–159,930   | 11        | 6           | Sample imported in Korea                                                           | [16]     |
| Schisandraceae    | Illicium anisatum                | 142,723–142,747   | 21        | 114         | Sample isolated in Korea                                                            | [102]    |
|                   |                                  |                   |           |             | Cold hardiness coffee tree (CH3) planted in Jeju Island                            |          |
|                   |                                  |                   | 2         | 2           | Coffee tree showing high productivity (HP1) in Jeju Island                          | NC_008535 |
|                   |                                  |                   | 0         | 2           | Coffee tree imported to Korea from Indonesia in near to 30 years ago (IN1)         | NC_008535 |
|                   |                                  |                   | 0         | 84          |                                                                                  | [50]     |
|                   |                                  |                   | 3         | 8           |                                                                                  |          |
|                   |                                  |                   | 0         | 3           |                                                                                  |          |
|                   |                                  |                   | 0         | 4           | MK342634, CH3 isolate                                                              | Coffea arabica “Typica” |
|                   |                                  |                   | 0         | 90          | HP1 isolate                                                                       | NC_008535 |
|                   |                                  |                   | 0         | 6           |                                                   | KY085909   |
|                   |                                  |                   | 0         | 1           |                                                   | NC_008535 |
|                   |                                  |                   | 0         | 6           |                                                   | KY085909   |
|                   |                                  |                   | 0         | 6           |                                                   | NC_008535 |
|                   |                                  |                   | 0         | 88          |                                                   | KY085909   |
|                   |                                  |                   | 0         | 2           |                                                   | CH1 isolate |
|                   |                                  |                   | 25        | 2           | Sample isolated in Soyeonpyeong-do, Incheon, Korea                                 | CH2 isolate |
|                   |                                  |                   |           |             | Sample isolated in Seogwang-ri, Jeju, Korea                                         | [72]     |
|                   |                                  |                   |           |             | Sample isolated in Soyeonpyeong-do, Incheon, Korea                                 |          |
|                   |                                  |                   |           |             | Sample isolated in China                                                            |          |
|                   |                                  |                   |           |             | Sample isolated in Soyeonpyeong-do, Incheon, Korea                                 | [105]    |
|                   |                                  |                   |           |             | Sample isolated in Mt. Taebak, Korea                                                |          |
|                   |                                  |                   |           |             | Sample isolated in Mt. Taebak, Korea                                                | Unpub.   |
| Theaceae          | Camellia japonica                | 156,606–157,047   | 78        | 643         | Sample isolated in Soyeonpyeong-do, Incheon, Korea                                 | [72]     |
|                   |                                  |                   |           |             | Sample isolated in Seogwang-ri, Jeju, Korea                                         |          |
|                   |                                  |                   |           |             | Sample isolated in Soyeonpyeong-do, Incheon, Korea                                 |          |
|                   |                                  |                   |           |             | Sample isolated in China                                                            |          |
| Caryophyllaceae   | Pseudostellaria longipedicellata  | 149,668           | 0         | 0           | Sample isolate in Mt. Taebak, Korea                                                | [25]     |
|                   | Pseudostellaria palibiniana      | 149,639–149,668   | 84        | 125         | Sample isolated in Mt. Taebak, Korea                                                |          |
|                   |                                  |                   |           |             | Sample isolated in Mt. Gwangdeok, Korea                                            |          |
| Salicaceae        | Salix koriyanagi                 | 155,548           | 0         | 0           | Male sample in Seoul, Korea                                                         | Female sample in Seoul, Korea |
|                   |                                  |                   |           |             |                                                                                  | [67]     |
| Family      | Species               | Total length (bp) | # of SNPs | # of INDELs | Comparison samples | Ref.  |
|------------|-----------------------|-------------------|----------|-------------|--------------------|------|
| Rosaceae   | Agrimonia pilosa *    | 155,125           | 258      | 542         | Samples isolated in Korea | [110] |
|            | Duchesnea indica *    | 156,050           | 45       | 273         | Samples isolated in Korea | [19]  |
|            | Duchesnea chrysantha  | 156,297–156,387   | 48       | 58          | Sample isolated in Korea | [21]  |
|            | Pyrus asduriensis     | 159,986–160,157   | 121      | 781         | Sample isolated in Bonghwa-gun, Korea | [68] |
| Fagaceae   | Fagus multiflora      | 158,349           | 2        | 2           | Sample isolated in Ulleungdo Island, Korea | [111] |
| Orobancheae| Rehmannia glutinosa   | 153,622–153,680   | 147      | 87          | Sample isolated in Korea | [73]  |
| Cucurbitaceae | Cucumis melo   | 155,815–156,017   | 1        | 0           | Cultivar Topmark | [76]  |
|             |                      |                   | 2        | 0           | Cultivar MR-1 | Cultivar DHL92 |
|             |                      |                   | 1        | 0           | Cultivar M4-  | Cultivar DHL92 |
|             |                      |                   | 2        | 0           | Cultivar M1-  | Cultivar DHL92 |
|             |                      |                   | 55       | 0           | Cultivar M4    | Cultivar DHL92 |
|             |                      |                   | 56       | 0           | Cultivar M1-3 | Cultivar DHL92 |
|             |                      |                   | 60       | 0           | Cultivar M1-7 | Cultivar DHL92 |
|             |                      |                   | 69       | 28          | Cultivar 0654 (Peru) | Cultivar PI 614886 (Chile) |
|             | Chenopodium quinoa   | 161,960–161,962   | 25       | 2           | Sample isolated in Anyang, Korea | [28]  |
|             |                      |                   | 25       | 13          | Sample isolated in Ganghwa, Korea | [112] |
|             |                      |                   | 17       | 36          | Sample isolated in China | [75]  |
|             |                      |                   | 37       | 24          | Sample isolated in China | [75]  |
|             | Eucommia ulmoides    | 163,341–163,586   | 75       | 80          | Sample isolated in China | [75]  |
|             |                      |                   | 99       | 0           | FLW genotype | BJXS genotype |
|             |                      |                   | 98       | 0           | TSW genotype | BJXS genotype |
|             |                      |                   | 110      | 0           | YTW genotype | BJXS genotype |
|             |                      |                   | 99       | 0           | XSW genotype | BJXS genotype |
|             |                      |                   | 106      | 0           | NJW genotype | BJXS genotype |
|             |                      |                   | 99       | 0           | MHW genotype | BJXS genotype |
|             | Dioscorea polystachya| 153,243–153,292   | 110      | 0           | Sample isolated in China | [77]  |
| Family      | Species                              | Total length (bp) | # of SNPs | # of INDELs | Comparison samples                                                                 | Ref.                     |
|-------------|--------------------------------------|-------------------|-----------|-------------|-----------------------------------------------------------------------------------|-------------------------|
| Asteraceae  | *Artemisia fukudo*                   | 151,011–151,021   | 7         | 11          | Sample isolated in Aphaedo, Korea Sample isolated in Jeungdo, Korea                | [113]                   |
|             |                                      |                   | 93        | 56          | Cultivar “Okhwang 1-ho” Cultivar candidate “Dae Ryun”                              | [90]                    |
|             |                                      |                   | 9         | 11          | Cultivar candidate “Dae Ryun”                                                      | [91]                    |
|             |                                      |                   | 102       | 63          | Sample isolated in Jincheon-gun, Korea Culivar candidate “Dae Ryun”                |                        |
|             |                                      |                   | 9         | 11          | Cultivar candidate “Sang Jae”                                                      | [92]                    |
|             |                                      |                   | 1         | 0           | MF407183                                                                            |                        |
|              |                                      | 200               | 511       |             | Sample isolated in Korea (MK144665) Sample isolated in Korea with different morphology (MK134679) |                        |
|              |                                      | 282               | 1,060     |             | Sample isolated in Korea (MK144665) Sample isolated in China (AB893949)            |                        |
|              |                                      | 827               | 1,794     |             | Sample isolated in Korea (MK144665) Sample isolated in China (NC_029364)          |                        |
|              |                                      | 700               | 1,066     |             | Sample isolated in Korea (MK144665) Sample isolated in China (LC085346)            |                        |
|              |                                      | 163               | 651       |             | Sample isolated in Korea with different morphology (MK1434679) Sample isolated in China (AB893949) | [53]                   |
|              | *Goodyera schlechtendaliana*         | 153,661–154,438   | 740       | 1,470       | Sample isolated in Korea with different morphology (MK1434679) Sample isolated in China (NC_029364) |                        |
| Orchidaceae |                                      | 597               | 1,065     |             | Sample isolated in Korea with different morphology (MK1434679) Sample isolated in China (LC085346) |                        |
|             |                                      | 844               | 2,045     |             | Sample isolated in China (AB893949) Sample isolated in China (NC_029364)          |                        |
|             |                                      | 445               | 414       |             | Sample isolated in China (AB893949) Sample isolated in China (LC085346)            |                        |
|             |                                      | 700               | 2,133     |             | Sample isolated in China (LC085346) Sample isolated in China (LC085346)            |                        |
|             |                                      | 457               | 670       |             | Sample isolated in Korea (MN296709) Sample isolated in China                      | [56]                   |
|             | *Gastrodia elata*                    | 35,066–35,304     | 493       | 650         | Sample isolated in Korea Sample isolated in China                                |                        |
| Adoxaceae   | *Viburnum erosum*                    | 158,587–158,624   | 16        | 49          | Sample isolated in Korea Sample isolated in Korea                                | [64]                   |
|             |                                      | 324               | 630       |             | Sample isolated in Korea Sample isolated in Korea                                | [57]                   |
| Ramunculaceae| *Aconitum coreanum*                  | 157,024–157,040   | 16        | 49          | Sample isolated in Korea Sample isolated in Korea                                | [64]                   |
|             |                                      | 157               | 17        |             | Sample isolated in Korea Sample isolated in Korea                                | [89]                   |
|             |                                      | 23                | 62        |             | Sample isolated in Korea Sample isolated in Korea                                |                        |
| Staphyleaceae| *Euscaphis japonica*                 | 160,467–160,606   | 424       | 809         | Sample isolated in Korea Sample isolated in China                                | [Oh et al., under review]|
| Thymelaeaceae| *Daphne genkwa*                     | 132,741–132,869   | 69        | 650         | Sample isolated in Korea Sample isolated in China                                | [Oh et al., under review]|

* indicates that the numbers of SNPs and INDELs were inferred only from LSC, SSC, and IRb regions because one of chloroplast genomes used for comparison is partial genome.
INDELs (Figure 2), suggesting that Tsu0 has a different genomic configuration compared to the remaining five strains.

The numbers of sequence variations on six Arabidopsis chloroplast genomes were plotted together with the numbers of intraspecific variations identified from 90 comparisons of 31 species (Table 3), resulting in three groups; one shows that the number of SNPs is less than 80 and that the number of INDELs is less than 100, the second indicates that the number of SNPs is less than 80 and number of INDELs is between 100 and 200, and the third shows that the number of SNPs exceeds 80 and that the number of INDELs is approximately 500 (Figure 3). The third group is caused by the long
insertion of the Tsu0 chloroplast genome. The third group is positioned with a relatively high number of variations areas, while the remaining groups are similar to the most of intra-specific variations on chloroplast genomes (green thick dotted circles in Figure 3).

3.4. Comparative Analysis of Simple Sequence Repeats (SSRs) Polymorphisms on Chloroplast Genomes inside East Asian A. thaliana. One hundred and one simple sequence repeats (SSRs) and 42 extendedSSRs on the chloroplast genome sequences of the Korean isolate of A. thaliana were identified (Supplementary Table 1). One hundred and four (72.72%), 18 (12.59%), and 21 (14.69%) SSRs and extendedSSRs were found in the LSC, IR, and SSC regions, respectively. This distribution is similar to that of Dysphania ambrosioides, but not to those of Dysphania pumilio or Dysphania botrys [30].

Eighteen SSRs and four extendedSSRs (15.38%) are located in the exonic regions of ten protein-coding genes, matK, trnK, trnK, rpoC2, rpoB, atpB, accD, psbB, rps12, rpoA, ycf1, and ndhF, and two tRNA genes, which is higher proportion than that of D. ambrosioides [30]. In addition, the number of genes on the A. thaliana chloroplast genome exceeds that of D. ambrosioides by one, while five out of ten protein-coding genes are shared between two species. 25 SSRs and 13 extendedSSRs (26.57%) are in intronic regions of five protein-coding genes and three tRNAs: ycfβ, rps12, cplP, rps16, and ndhA and trnK, trnR, and trnA, respectively. Compared to previous findings that identified SSRs in 12 chloroplast genomes of Brassicaceae, the numbers of SSRs found on the genes are similar to each other, ranging from 40 to 60 [45], which is similar to that of the A. thaliana Korean isolate.

We also applied the same method to identify SSRs of the other five chloroplast genomes of A. thaliana used in this study (Table 4). The total numbers of SSRs and extendedSSRs range from 143 to 145, showing that the Korean isolate of A. thaliana has the fewest, at 143 (Table 4). Based on the number of sequence variations among the six chloroplast genomes (Figure 2), the numbers of SSRs and extendedSSRs along with the motif length are expected to be nearly identical; however, only the triSSRs, nonaSSRs, and decaSSRs show identical numbers across the six chloroplast genomes (Table 4).

Using the SSR comparison pipeline implemented in SSRDB, 117 groups of SSRs or extendedSSRs containing six SSRs from the six A. thaliana chloroplast genomes are identified, accounting for 702 out of 864 SSRs or extendedSSRs (81.25%; Figure 4). There is one interesting SSR group (named as SSR Group 2) containing six SSRs from the six A. thaliana chloroplasts: two are octaSSRs (TATCTATA∗2) and four are diSSRs (TA∗5). Twenty-one SSR groups contain five SSRs or extendedSSRs from five chloroplast genomes, explaining 105 out of 864 SSRs or extendedSSRs (12.15%; Figure 4). Five SSR groups containing four SSRs or extendedSSRs from four chloroplast genomes and three SSR groups covering three SSRs or extendedSSRs from three chloroplasts, four SSR groups having two SSRs or extendedSSRs from two chloroplast, and 20 singletons, indicating unique SSRs, among the six chloroplast genomes are identified (Figure 4). Considering the coverage of the SSRs and extendedSSRs on Korean isolate of the A. thaliana chloroplast genome (in total, 1,825 bp out of 154,464 bp; 1.18%), the expected number of sequence variations of the SSR and extendedSSR regions is 0.75; however, the number of common SSRs or extendedSSRs is 117 (81.82%), indicating that the numbers of sequence variations located in SSR or extendedSSR regions are lower than expected number (107.25 sequence variations for SSR or extendedSSR regions). These variations can be used to develop molecular markers [41].

3.5. Comparison of Nucleotide Diversity among the Six A. thaliana Chloroplast Genomes. Nucleotide diversity among six Arabidopsis thaliana chloroplast genomes was calculated, indicating that the average nucleotide diversity is 0.00017 (Figure 5(a)), which is at least ten times lower than those of Dysphania (0.0068; Chenopodiaceae) [27] and Viburnum

| SSR type | 180404 | 15-11 | Kyoto | Tsu0 | Ler0 | Col0 |
|----------|--------|-------|-------|------|------|------|
| MonoSSR  | 67     | 69    | 68    | 70   | 67   | 69   |
| DiSSR    | 17     | 17    | 18    | 16   | 18   | 18   |
| TriSSR   | 6      | 6     | 6     | 6    | 6    | 6    |
| TetraSSR | 8      | 8     | 9     | 8    | 9    | 9    |
| PentaSSR | 2      | 3     | 2     | 1    | 3    | 2    |
| HexaSSR  | 1      | 1     | 1     | 1    | 0    | 1    |
| HeptaSSR | 27     | 25    | 25    | 26   | 26   | 25   |
| OctaSSR  | 8      | 8     | 9     | 10   | 8    | 7    |
| NonaSSR  | 6      | 6     | 6     | 6    | 6    | 6    |
| DecaSSR  | 1      | 1     | 1     | 1    | 1    | 1    |
| Total    | 143    | 144   | 145   | 144  | 144  | 144  |

![Figure 4: Distribution of the SSR groups identified from six chloroplast genomes of A. thaliana. X-axis indicates the types of the SSR groups and Y-axis means the number of the SSR groups or SSRs/extendedSSRs. Blue graph means the number of the SSR groups and orange bars mean the # of SSRs from the SSR groups.](image-url)
This is a justifiable result because sequence diversity within species is usually lower than interspecific nucleotide diversity.

There are two significant peaks identified in the sliding window analysis of nucleotide diversity: one is \textit{trnL/trnF} (pi value is 0.0147) and the other is \textit{trnP/psaJ} (pi value is 0.00441). There are fewer peaks than in other studies, including those focusing on \textit{Dysphania} [27] and \textit{Viburnum} [32], stemming from the low level of nucleotide diversity throughout the chloroplast genome. The first peak, \textit{trnL/trnF}, appeared due to one large insertion of the Tsu0 chloroplast genome (Figure 5(b)). The second peak, \textit{trnP/psaJ}, reflects the sequence variations occurring in 180404IB4 (Korea), 15-11 (China), Tsu0 (Japan), and Ler0 (Germany; Figure 5(c)). Specifically, SNPs located between 67,670 and 67,680 in both 180404IB4 and the 15-10 isolates mainly contribute to this peak (Figure 5(c)).

3.6. Comparison of the IR Junction among Arabidopsis thaliana Chloroplast Genomes. The IR region on the plant chloroplast genome is the major origin at which to expand or to shrink the chloroplast genome sequences [84–88]. An investigation of the IR junctions of \textit{A. thaliana} chloroplast genomes shows that there are no differences among six \textit{A. thaliana} chloroplast genomes, in agreement with the finding of no structural variations (Figure 6), identical to the case of \textit{D. ambrosioides} [30]. In addition, all \textit{Arabidopsis} chloroplast genomes used in this study present the same structure in the IR junctions.

3.7. Phylogenetic Analysis of Korean \textit{A. thaliana} Chloroplast Genome Sequence. Bootstrapped neighbor-joining (NJ), maximum parsimony (ML), and Bayesian inference (BI) phylogenetic trees of seventeen \textit{Arabidopsis} chloroplast genomes including six \textit{A. thaliana} chloroplasts and one \textit{Arabis} chloroplast genome as outgroup species indicate that 180404IB4 is clustered with the 15-11 (Chinese natural isolate) with high bootstrap support, while two Japanese isolates are not clustered together in contrast to what was expected here (Figure 7), as Tsu0 has an approximate 500 bp insertion compared to all other \textit{A. thaliana} chloroplast genomes. This indicates that more chloroplast genomes of East Asian \textit{A. thaliana} natural isolates should be investigated to find exceptional sequence variations, such as a Tsu0 insertion.
Practically, it is possible to utilize currently available NGS raw read datasets of *A. thaliana* natural isolates by adding an effort to assemble them. In addition, we must consider the possibility of leaking Col0 strains from many molecular laboratories in Korea, which will affect their genetic diversity in some ways. Based on the phylogenetic trees, there appears to be no contamination in Korea.

Several intraspecific phylogenetic relations of plant species using whole chloroplast genomes have been studied, including *Aconitum coreanum*, showing small branches of three individuals with high bootstrap values from both ML and NJ methods from mid-level of sequence variations [89]; *G. schlechtendaliana*, displaying branches of each samples caused by a sufficient number of sequence variations with high bootstraps in both methods [53, 54]; *Abeliophyllum distichum*, indicating partial support of intraspecific individuals from both methods [90–92]; and *Coﬀea arabica*, showing high bootstrap values from both methods with no branch of either individual sequences due to the low level of sequence variations [50, 59–63]. All these results differ from that of *A. thaliana*, presenting a different clade structure from the phylogenetic trees constructed by three methods (Figure 7). Instead, this phenomenon was found in genome studies focusing on intraspecific variations of insect, fungal, and marine invertebrate mitochondria. These include *Laodelphax striatellus* [93, 94] and *Nilaparvata lugens* [95–97] belonging to the Delphacidae family; *Fusarium oxysporum* which is a fungal plant pathogen [98, 99] and *Apostichopus japonicus* [100]. Because *A. thaliana* has a sufficient amount of sequencing data to construct chloroplast genomes, additional studies with more complete chloroplast genomes will provide a clear answer as to whether or not this phenomenon remains.

In addition, *A. thaliana* and *Arabidopsis suecica* are clustered because the maternal origin of *A. suecica* is *A. thaliana* [101] (Figure 7). Moreover, the topology of BI tree for the remaining *Arabidopsis* species, except for *A. thaliana* and *A. suecica*, is somewhat inconsistent to those of NJ and ML trees (Figure 7), which has also been found in various plant chloroplast genomes [15, 20, 28, 51, 67, 69, 72, 79, 91, 92, 102–109]. This suggests that detailed investigations of phylogenetic relationships among *Arabidopsis* species with various methods should be done in near future.

### 4. Conclusions

We sequenced and assembled the chloroplast genome of the Korean isolate of *A. thaliana* and compared this with the other East Asian *A. thaliana* chloroplast genomes assembled from NGS raw reads available to the public. Based on the numbers of sequence variations of the six *A. thaliana* chloroplast genomes, three groups with low, medium, and high levels of sequence variations were found, particularly due to the large insertion identified on the Tsu0 chloroplast genome. Here, 101 SSRs and 42 extendedSSRs were identified on the Korean *A. thaliana* chloroplast genome, with similar numbers of SSRs on the remaining five chloroplast genomes with a preference of sequence variations of the SSR region. Nucleotide diversity on the six *A. thaliana* chloroplast genomes indicates only two regions that are highly variable,
an outcome that is less dynamic than those of interspecific comparisons of chloroplast genomes. As expected, the IR borders of the six chloroplast genomes are conserved. Phylogenetic analyses of the six *A. thaliana* chloroplast genomes with those of other *Arabidopsis* species revealed that the geographical distribution is not congruent with the phylogenetic relationships; however, more complete chloroplast genomes are required for further analysis. Additional whole chloroplast genomes of *A. thaliana* strains based on a large amount of genomic resources of *A. thaliana* can describe the detailed evolutionary history of the natural isolates of *A. thaliana* in East Asia, especially for Korea, China, and Japan.

### Data Availability

Chloroplast genome sequence of Korean *A. thaliana* can be accessed via accession number MK353213 in NCBI GenBank. In addition, three more chloroplast genomes of *A. thaliana*, Kyoto, Tsu0, and 11-15, based on SRA datasets are accessible through MK380720, MK380721, and MK380719, respectively.

### Conflicts of Interest

The authors declare that they have no competing interests.

### Acknowledgments

This work was fully supported by the InfoBoss Research Grant (IBG-0023).

### Supplementary Materials

List of SSRs identified from *A. thaliana* 180404IB4 chloroplast genome. (Supplementary Materials)

### References

[1] E. M. Meyerowitz, “*Arabidopsis*, a useful weed,” *Cell*, vol. 56, no. 2, pp. 263–269, 1989.

[2] S. Sato, Y. Nakamura, T. Kaneko, E. Asamizu, and S. Tabata, “Complete structure of the chloroplast genome of *Arabidopsis thaliana*,” *DNA Research*, vol. 6, no. 5, pp. 283–290, 1999.
read RNA-Seq data: a comparative study," *BMC Bioinformatics*, vol. 12, Suppl 14, p. S2, 2011.

[36] H. Li, B. Handsaker, A. Wysoker et al., "The sequence alignment/map format and SAMtools," *Bioinformatics*, vol. 25, no. 16, pp. 2078-2079, 2009.

[37] H. Li, "Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM," 2013, http://arxiv.org/abs/1303.3997.

[38] P. Schattner, A. N. Brooks, and T. M. Lowe, "The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs," *Nucleic Acids Research*, vol. 33, no. Web Server, suppl_2, pp. W686-W689, 2005.

[39] K. Katoh and D. M. Standley, "MAFFT multiple sequence alignment software version 7: improvements in performance and usability," *Molecular Biology and Evolution*, vol. 30, no. 4, pp. 772–780, 2013.

[40] J. Chen, Z. Hao, H. Xu et al., "The complete chloroplast genome sequence of the relict woody plant *Metasequoia glyptostroboides* Hu et Cheng," *Frontiers in Plant Science*, vol. 6, 2015.

[41] N. Shukla, H. Kuntal, A. Shanker, and S. N. Sharma, "Mining and analysis of simple sequence repeats in the chloroplast genomes of genus *Vigna*," *Biotechnology Research and Innovation*, vol. 2, no. 1, pp. 9–18, 2018.

[42] W. Li, C. Zhang, X. Guo, Q. Liu, and K. Wang, "Complete chloroplast genome of *Camellia japonica* genome structures, comparative and phylogenetic analysis," *PLOS One*, vol. 14, no. 5, article e0216645, 2019.

[43] J.-H. Jeon and S.-C. Kim, "Comparative analysis of the complete chloroplast genome sequences of three closely related East-Asian wild roses (Rosa sect. Synstyelae; Rosaceae)," *Genes*, vol. 10, no. 1, p. 23, 2019.

[44] J. Cheng, Z. Zhao, B. Li et al., "A comprehensive characterization of simple sequence repeats in pepper genomes provides valuable resources for marker development in *Capsicum*," *Scientific Reports*, vol. 6, no. 1, 2016.

[45] S. G. Gandhi, P. Awaithi, and Y. S. Bedi, "Analysis of SSR dynamics in chloroplast genomes of Brassicaceae family," *Bioinformation*, vol. 5, no. 1, pp. 16–20, 2010.

[46] J. Lee, J. Park, W. Kwon, H. Xi, and J. Park, "The complete mitochondrial genome of *Ficusulus binodulus* Waterhouse, 1873 (Coleoptera: Lucanidae)," *Journal of Insect Science*, 2020.

[47] M. Nei and W.-H. Li, "Mathematical model for studying genetic variation in terms of restriction endonucleases," *Proceedings of the National Academy of Sciences*, vol. 76, no. 10, pp. 5269–5273, 1979.

[48] S. Kumar, G. Stecher, M. Li, C. Knyaz, and K. Tamura, "MEGA X: molecular evolutionary genetics analysis across computing platforms," *Molecular Biology and Evolution*, vol. 35, no. 6, pp. 1547–1549, 2018.

[49] J. P. Huelsenbeck and F. Ronquist, "MRBAYES: Bayesian inference of phylogenetic trees," *Bioinformatics*, vol. 17, no. 8, pp. 754-755, 2001.

[50] J. Park, Y. Kim, H. Xi, and K.-L. Heo, "The complete chloroplast genome of ornamental coffee tree, *Coffea arabica L.* (Rubiaceae)," *Mitochondrial DNA Part B*, vol. 4, no. 1, pp. 1059-1060, 2019.

[51] J. Park, Y. Kim, H. Xi, W. Kwon, and M. Kwon, "The complete chloroplast and mitochondrial genomes of Hyunsasi tree, *Populus alba* x *Populus glandulosa* (Salicaceae)," *Mitochondrial DNA Part B*, vol. 4, no. 2, pp. 2521-2522, 2019.

[52] J. Park, Y. Kim, W. Kwon, H. Xi, and M. Kwon, "The complete chloroplast genome of tulip tree, *Liriodendron tulipifera L.* (Magnoliaceae): investigation of intra-species chloroplast variations," *Mitochondrial DNA Part B*, vol. 4, no. 2, pp. 2523-2524, 2019.

[53] S.-H. Oh, H. J. Suh, J. Park, Y. Kim, and S. Kim, "The complete chloroplast genome sequence of *Goodrya schlechten- daliana* in Korea (Orchidaceae)," *Mitochondrial DNA Part B*, vol. 4, no. 2, pp. 2692–2693, 2019.

[54] S.-H. Oh, H. J. Suh, J. Park, Y. Kim, and S. Kim, "The complete chloroplast genome sequence of a morphotype of *Goodrya schlechten-daliana* (Orchidaceae) with the column appendages," *Mitochondrial DNA Part B*, vol. 4, no. 1, pp. 626–627, 2019.

[55] Z. Niu, Q. Xue, S. Zhu, J. Sun, W. Liu, and X. Ding, "The complete plastome sequences of four orchid species: insights into the evolution of the Orchidaceae and the utility of plastomic mutational hotspots," *Frontiers in Plant Science*, vol. 8, p. 715, 2017.

[56] J. Park, Y. Suh, and S. Kim, "A complete chloroplast genome sequence of *Gastrodia elata* (Orchidaceae) represents high sequence variation in the species," *Mitochondrial DNA Part B*, vol. 5, no. 1, pp. 517–519, 2020.

[57] M.-J. Kang, S.-C. Kim, H.-R. Lee et al., "The complete chloroplast genome of Korean *Gastrodia elata* Blume," *Mitochondrial DNA Part B*, vol. 5, no. 1, pp. 1015-1016, 2020.

[58] Y. Yuan, X. Jin, J. Liu et al., "The *Gastrodia elata* genome provides insights into plant adaptation to heterotrophy," *Nature Communications*, vol. 9, no. 1, p. 1615, 2018.

[59] J. Min, Y. Kim, H. Xi, K.-I. Heo, and J. Park, "The complete chloroplast genome of coffee tree, *Coffea arabica L.* ‘Typica’ (Rubiaceae)," *Mitochondrial DNA Part B*, vol. 4, no. 2, pp. 2240-2241, 2019.

[60] J. Park, Y. Kim, H. Xi, and K.-I. Heo, "The complete chloroplast genome of coffee tree, *Coffea arabica L.* ‘Blue Mountain’ (Rubiaceae)," *Mitochondrial DNA Part B*, vol. 4, no. 2, pp. 2436-2437, 2019.

[61] J. Park, Y. Kim, H. Xi, M. Nho, J. Woo, and Y. Seo, "The complete chloroplast genome of high production individual tree of *Coffea arabica L.* (Rubiaceae)," *Mitochondrial DNA Part B*, vol. 4, no. 1, pp. 1541-1542, 2019.

[62] J. Park, H. Xi, Y. Kim et al., "The complete chloroplast genome of cold hardiness individual of *Coffea arabica* L. (Rubiaceae)," *Mitochondrial DNA Part B*, vol. 4, no. 1, pp. 1083-1084, 2019.

[63] J. Park, Y. Kim, H. Xi et al., "The complete chloroplast genomes of two cold hardiness coffee trees, *Coffea arabica* L. (Rubiaceae)," *Mitochondrial DNA Part B*, vol. 5, no. 2, pp. 1619–1621, 2020.

[64] Y. G. Choi, N. Yun, J. Park et al., "The second complete chloroplast genome sequence of the *Viburnum erosum* (Adoxaceae) showed a low level of intra-species variations," *Mitochondrial DNA Part B*, vol. 5, no. 1, pp. 271-273, 2020.

[65] J. Park, Y. G. Choi, N. Yun et al., "The complete chloroplast genome sequence of *Viburnum erosum* (Adoxaceae)," *Mitochondrial DNA Part B*, vol. 4, no. 2, pp. 3278-3279, 2019.

[66] J. Kim, Y. Kim, and J. Park, "Complete chloroplast genome sequence of the *Salix koriyanagi* Kimura ex Goerz (Salica-ceae)," *Mitochondrial DNA Part B*, vol. 4, no. 1, pp. 549-550, 2019.
[67] J. Park, Y. Kim, and H. Xi, "The complete chloroplast genome sequence of male individual of Korean endemic willow, Salix koryanae Kimura ex Goerz (Salicaceae)," Mitochondrial DNA Part B, vol. 4, no. 1, pp. 1619–1621, 2019.

[68] M.-S. Cho, Y. Kim, S.-C. Kim, and J. Park, "The complete chloroplast genome of Korean Pyrus ussuriensis Maxim. (Rosaceae): providing genetic background of two types of P. ussuriensis," Mitochondrial DNA Part B, vol. 4, no. 2, pp. 2424–2425, 2019.

[69] H.-Y. Gil, Y. Kim, S.-H. Kim et al.,"The complete chloroplast genome of Pyrus ussuriensis Maxim. (Rosaceae), Mitochondrial DNA Part B, vol. 4, no. 1, pp. 1000-1001, 2019.

[70] P. J. Maughan, L. Chaney, D. J. Lightfoot et al., "Mitochondrial and chloroplast genomes provide insights into the evolutionary origins of quinoa (Chenopodium quinoa Willd.)," Scientific Reports, vol. 9, no. 1, p. 185, 2019.

[71] W. Kwon, Y. Kim, and J. Park, "The complete mitochondrial genome of Korean Marchantia polymorpha subsp. raderalis Bischl. & Boisselier: inverted repeats on mitochondrial genome between Korean and Japanese isolates," Mitochondrial DNA Part B, vol. 4, no. 1, pp. 769-770, 2019.

[72] J. Park, Y. Kim, H. Xi, Y. J. Oh, K. M. Hamh, and J. Ko, "The complete chloroplast genome of common camellia tree, Camellia japonica L. (Theaceae), adapted to cold environment in Korea," Mitochondrial DNA Part B, vol. 4, no. 1, pp. 1038–1040, 2019.

[73] J.-H. Jeon, H.-S. Park, J. Y. Park et al., "Two complete chloroplast genome sequences and intra-species diversity for Rheumannia glutinosa (Orobanchaceae)," Mitochondrial DNA Part B, vol. 4, no. 1, pp. 176-177, 2019.

[74] J. Park, Y. Kim, G.-H. Lee, and C.-H. Park, "The complete chloroplast genome of Selaginella tamariscina (Beauv.) Spring (Selaginellaceae) isolated in Korea," Mitochondrial DNA Part B, vol. 5, no. 2, pp. 1654–1656, 2020.

[75] W. Wang, S. Chen, and X. Zhang, "Whole-genome comparison reveals heterogeneous divergence and mutation hotspots in chloroplast genome of Eucommia ulmoides Oliver," International Journal of Molecular Sciences, vol. 19, no. 4, p. 1037, 2018.

[76] Q. Zhu, P. Gao, S. Liu, S. Amanullah, and F. Luan, "Comparative analysis of single nucleotide polymorphisms in the nuclear, chloroplast, and mitochondrial genomes in identification of phylogenetic association among seven melon (Cucumis melo L.) cultivars," Breeding Science, vol. 66, no. 5, pp. 711–719, 2016.

[77] J. Cao, D. Jiang, Z. Zhao et al., "Development of chloroplast genomic resources in Chinese Yam ( Dioscorea polystachya)," BioMed Research International, vol. 2018, 11 pages, 2018.

[78] J. Tang, H.-X. Xia, M. Cao et al., "A comparison of rice chloroplast genomes," Plant Physiology, vol. 135, no. 1, pp. 412–420, 2004.

[79] H. Xi, J. Park, and Y. Kim, "The complete chloroplast genome sequence of rose-gold pussy willow, Salix gracilistyla Miq. (Salicaceae)," Mitochondrial DNA Part B, vol. 4, no. 2, pp. 2118–2120, 2019.

[80] J. Park, Y. Kim, W. Kwon, S. Nam, and H. Xi, "The complete chloroplast genome of Nepal Holly, Ilex integra Thunb. (Aquifoliaceae)," Mitochondrial DNA Part B, vol. 4, no. 1, pp. 1257-1258, 2019.

[81] D. L. Roberts and K. W. Dixon, "Orchids," Current Biology, vol. 18, no. 8, pp. R325–R329, 2008.

[82] T. J. Givnish, D. Spalink, M. Ames et al., "Orchid phylogenomics and multiple drivers of their extraordinary diversification," Proceedings of the Royal Society B: Biological Sciences, vol. 282, no. 1814, p. 20151553, 2015.

[83] T. J. Givnish, D. Spalink, M. Ames et al., "Orchid historical biogeography, diversification, Antarctica and the paradox of orchid dispersal," Journal of Biogeography, vol. 43, no. 10, pp. 1905–1916, 2016.

[84] W. Dong, C. Xu, D. Li et al., "Comparative analysis of the complete chloroplast genome sequences in psammophytic Haloxylon species (Amaranthaceae)," PeerJ, vol. 4, article e2699, 2016.

[85] Y. Yang, T. Zhou, D. Duan, J. Yang, L. Feng, and G. Zhao, "Comparative analysis of the complete chloroplast genomes of five Quercus species," Frontiers in Plant Science, vol. 7, 2016.

[86] S. Asaf, A. L. Khan, A. R. Khan et al., "Complete chloroplast genome of Nicotiana otophora and its comparison with related species," Frontiers in Plant Science, vol. 7, 2016.

[87] Z.-Z. Li, J. Saina, A. Gichira, C. Kyalo, Q.-F. Wang, and J.-M. Chen, "Comparative genomics of the balsaminaceae sister genera Hydrocera trifora and Impatiens pinfanensis," International Journal of Molecular Sciences, vol. 19, no. 1, p. 319, 2018.

[88] S.-H. Kim, J. Y. Yang, J. Park, T. Yamada, M. Maki, and S.-C. Kim, "Comparison of whole plastome sequences between thermogenic skunk cabbage Symlocarpus renifolius and nonthermogenic S. nipponicus (Orontioidae; Araceae) in East Asia," International Journal of Molecular Sciences, vol. 20, no. 19, p. 4678, 2019.

[89] Y. Kim, J.-S. Yi, J. Min et al., "The complete chloroplast genome of Aconitum corеanum (H. Lév.) Rapaics (Ranunculaceae)," Mitochondrial DNA Part B, vol. 4, no. 2, pp. 3404–3406, 2019.

[90] J. Park, Y. Kim, H. Xi, T. Jang, and J.-H. Park, "The complete chloroplast genome of Abelophyllum distichum Nakai (Oleaceae), cultivar Ok Hwang 1ho: insights of cultivar specific variations of A. distichum," Mitochondrial DNA Part B, vol. 4, no. 1, pp. 1640–1642, 2019.

[91] J. Park, J. Min, Y. Kim et al., "The complete chloroplast genome of a new candidate cultivar, Dae Ryun, of Abelophyllum distichum Nakai (Oleaceae): initial step of A. distichum intraspecies variations atlas," Mitochondrial DNA Part B, vol. 4, no. 2, pp. 3713–3715, 2019.

[92] J. Min, Y. Kim, H. Xi et al., "The complete chloroplast genome of a new candidate cultivar, Sang Jae, of Abelophyllum distichum Nakai (Oleaceae): initial step of A. distichum intraspecies variations atlas," Mitochondrial DNA Part B, vol. 4, no. 2, pp. 3716–3718, 2019.

[93] J. Park, J. K. Jung, Y. Ho Koh, J. Park, and B. Y. Seo, "The complete mitochondrial genome of Laodelphax striatellus (Fallén, 1826) (Hemiptera: Delphacidae) collected in a mid-Western part of Korean peninsula," Mitochondrial DNA Part B, vol. 4, no. 2, pp. 2229-2230, 2019.

[94] B. Y. Seo, J. K. Jung, Y. Ho Koh, and J. Park, "The complete mitochondrial genome of Laodelphax striatellus (Fallén, 1826) (Hemiptera: Delphacidae) collected in a southern part of Korean peninsula," Mitochondrial DNA Part B, vol. 4, no. 2, pp. 2242-2243, 2019.

[95] N. J. Choi, B.-C. Lee, J. Park, and J. Park, "The complete mitochondrial genome of Nilaparvata lugens (Stål, 1854) captured in China (Hemiptera: Delphacidae): investigation
of intraspecific variations between countries,” *Mitochondrial DNA Part B*, vol. 4, no. 1, pp. 1677-1678, 2019.

[96] J. Park, W. Kwon, J. Park et al., “The complete mitochondrial genome of *Nilaparvata lugens* (Stål, 1854) captured in Korea (Hemiptera: Delphacidae),” *Mitochondrial DNA Part B*, vol. 4, no. 1, pp. 1674–1676, 2019.

[97] N. J. Choi, B.-C. Lee, J. Park, and J. Park, “The complete mitochondrial genome of *Nilaparvata lugens* (Stål, 1854) captured in Guangxi Province, China (Hemiptera: Delphacidae): identification of the origin of *N. lugens* migrated to Korea,” *Mitochondrial DNA Part B*, vol. 5, no. 2, pp. 1960-1961, 2020.

[98] W. Kwon, J. Park, J.-B. Kim, M.-J. Park, and T.-S. Kim, “Complete mitochondrial genome sequence of lettuce pathogenic fungus, *Fusarium oxysporum* f. sp. *lactucae* 16-086,” *Mitochondrial DNA Part B*, vol. 4, no. 2, pp. 3227-3228, 2019.

[99] J. Park, W. Kwon, J.-B. Kim, M.-J. Park, and T.-S. Kim, “Complete mitochondrial genome sequence of lettuce pathogenic fungus, *Fusarium oxysporum* f. sp. *lactucae* 09-002,” *Mitochondrial DNA Part B*, vol. 4, no. 2, pp. 3434–3436, 2019.

[100] J. Park, H. Xi, and C. Park, “The complete mitochondrial genomes from three body color variants of sea cucumbers, *Apostichopus japonicus* (Selenka, 1867),” *Mitochondrial DNA Part B*, vol. 4, no. 1, pp. 836-837, 2019.

[101] K. Munnenhoff and H. Hurka, “Allopolyploid origin of *Arabidopsis suecica* (Fries) Nordr: evidence from chloroplast and nuclear genome markers,” *Botanica Acta*, vol. 108, no. 5, pp. 449-456, 1995.

[102] J. Park, Y. Kim, and H. Xi, “The complete chloroplast genome of aniseed tree, *Ilicium anisatum* L. (Schisandraceae),” *Mitochondrial DNA Part B*, vol. 4, no. 1, pp. 1023-1024, 2019.

[103] W. Kwon, J. Min, Y. Kim, and J. Park, “The complete chloroplast genome of *Rehoulia hemisphaerica* (L.) Raddi (Aytoniaceae, Marchantiophyta),” *Mitochondrial DNA Part B*, vol. 4, no. 1, pp. 1459-1460, 2019.

[104] N. Yun, J. Park, and S.-H. Oh, “The complete chloroplast genome of the traditional medicinal plant *Stellera chamaejasme* L. (Thymelaeaceae),” *Mitochondrial DNA Part B*, vol. 4, no. 1, pp. 1796-1797, 2019.

[105] J. Park, Y. Kim, H. Xi, Y. J. Oh, K. M. Hahn, and J. Ko, “The complete chloroplast genome of common camellia tree in Jeju Island, Korea, *Camellia japonica* L. (Theaceae): intraspecific variations on common camellia chloroplast genomes,” *Mitochondrial DNA Part B*, vol. 4, no. 1, pp. 1292-1293, 2019.

[106] W. Kwon, J. Min, H. Xi, and J. Park, “The complete chloroplast genome of *Riccia fluviatils* L. (Ricciaceae, Marchantiophyta),” *Mitochondrial DNA Part B*, vol. 4, no. 1, pp. 1895-1896, 2019.

[107] Y. Kim, J. Min, W. Kwon, M. J. Song, S. Nam, and J. Park, “The complete chloroplast genome sequence of the *Nymphoides capensis* Thunb. (Nymphaeaceae),” *Mitochondrial DNA Part B*, vol. 4, no. 1, pp. 401-402, 2019.

[108] Y. Kim, J. Park, and Y. Chung, “The complete chloroplast genome of *Suaeda japonica* Makino (Amaranthaceae),” *Mitochondrial DNA Part B*, vol. 4, no. 1, pp. 1505–1507, 2019.

[109] S. S. Choi, W. Kwon, and J. Park, “The complete chloroplast genome of *Scapania ampliata* Steph., 1897 (Scapaniaceae, Jungermanniales),” *Mitochondrial DNA Part B*, vol. 5, no. 3, pp. 2908–2910, 2020.

[110] K.-I. Heo, J. Park, H. Xi, and J. Min, “The complete chloroplast genome of *Agrimonia pilosa* Ledeb. isolated in Korea (Rosaceae): investigation of intraspecific variations on its chloroplast genomes,” *Mitochondrial DNA Part B*, vol. 5, no. 3, pp. 2264–2266, 2020.

[111] J. Park and S.-H. Oh, “A second complete chloroplast genome sequence of *Fagus multinervis* Nakai (Fagaceae): intraspecific variations on chloroplast genome,” *Mitochondrial DNA Part B*, vol. 5, no. 2, pp. 1868-1869, 2020.

[112] Y. Kim, J. Park, and Y. Chung, “The comparison of the complete chloroplast genome of *Suaeda japonica* Makino presenting different external morphology (Amaranthaceae),” *Mitochondrial DNA Part B*, vol. 5, no. 2, pp. 1616–1618, 2020.

[113] J. Min, J. Park, Y. Kim, and W. Kwon, “The complete chloroplast genome of *Artemisia fukudo* Makino (Asteraceae): providing insight of intraspecies variations,” *Mitochondrial DNA Part B*, vol. 4, no. 1, pp. 1510–1512, 2019.