The Complete Chloroplast Genome Sequence and Intra-Species Diversity of *Rhus chinensis*

**Inseo Kim**, **Jee Young Park**, **Yun Sun Lee**, **Ho Jun Joh**, **Shin Jae Kang**, **Jayakodi Murukarthick**, **Hyun Oh Lee**, **Young-Jin Hur**, **Yong Kim**, **Kyung Hoon Kim**, **Sang-Choon Lee**, **Tae-Jin Yang**

1Department 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
2Phyzen Genomics Institute, Seongnam 13558, Korea
3Il Song ERT. Co. LTD, Yongin 16950, Korea

**ABSTRACT**  *Rhus chinensis* is a shrub widely distributed in Asia. It has been used for traditional medicine and ecological restoration. Here, we report the complete chloroplast genome sequence of two *R. chinensis* genotypes collected from China and Korea. The assembled chloroplast genome of Chinese *R. chinensis* is 149,094 bp long, consisting of a large single copy (97,246 bp), a small single copy (18,644 bp) and a pair of inverted repeats (16,602 bp). Gene annotation revealed 77 protein coding genes, 30 tRNA genes, and 4 rRNA genes. A phylogenomic analysis of the chloroplast genomes with 11 known complete chloroplast genomes clarified the relationship of *R. chinensis* with the other plant species in the Sapindales order. A comparative chloroplast genome analysis identified 170 SNPs and 85 InDels at intra-species level of *R. chinensis* between Chinese and Korean collections. Based on the sequence diversity between Korea and Chinese *R. chinensis* plants, we developed three DNA markers useful for genetic diversity and authentication system. The chloroplast genome information obtained in this study will contribute to enriching genetic resources and conservation of endemic *Rhus* species.

**Keywords**  *Rhus chinensis*, Medicinal plant, Ecological restoration, DMZ, Chloroplast genome sequence, Molecular marker

**INTRODUCTION**  *Rhus chinensis* is a deciduous shrub belonging to the family Anacardiaceae and distributed widely in Asia including India, Vietnam, China, Korea, and Japan (Min and Barfod 2008). It contains various pharmacologically active constituents and insect-induced galls, which have been used for medicinal purposes (Min and Barfod 2008; Djakpo and Yao 2010). In addition, *R. chinensis* is used as revegetation plant for ecological restoration owing to its cold tolerance and easy multiplication by both seed and clonal propagation (Nam et al. 2004; Lim and Oh 2015). The diverse utilization of this shrub led to international seed trading which in turn resulted in contamination and destruction of endemic population of *R. chinensis*. Although the generic and infrageneric delimitation of *Rhus* species remains controversial, very limited genetic resources are available for *R. chinensis* (Young 1978; Miller et al. 2001; Yi et al. 2004; Ma et al. 2013; Lee et al. 2016). In this study, we have characterized the complete chloroplast genome of Chinese *R. chinensis* and compared with Korean *R. chinensis* chloroplast genome for estimating genetic diversity and developing marker for authentication. Chloroplast DNA has a characteristic of exhibiting variations at intra-species level and thus chloroplast DNA has widely been used for classification among populations.
Chloroplast genomes are uni-parentally inherited and contain rare recombination and mutation events (Cheng et al. 2005; Dong et al. 2012). These features make chloroplast genomes good candidate for studying evolution at inter- or intra-species level. For instance, intra-species researches had been conducted using chloroplast DNA in various plant species including *Panax ginseng*, *Pedicularis chamissonis*, *Primula cuneifolia*, *Tellima grandiflora*, and *Tiarella trifoliata* (Soltis et al. 1991; Soltis et al. 1992; Fujii et al. 1995; Sewell et al. 1996; Fujii et al. 1997; Kim et al. 2015a; Joh et al. 2017). Chloroplast genome was also applied to taxonomic studies of *Rhus* and marker development to classify species (Miller et al. 2001; Lee et al. 2004; Yi et al. 2004; Yi et al. 2007). However, high resolution markers have not been developed for distinguishing *Rhus* species at both inter- and intra-species level.

In this study, we assembled the complete chloroplast genome of two *R. chinensis* genotypes from China and Korea. In addition, we have conducted a comparative phylogenomic analysis of Chinese and Korean *R. chinensis* along with 11 species belonged in the Sapindales order. We also developed polymorphic DNA markers derived from chloroplast genomes to practically apply for authentication and genetic diversity among *R. chinensis* collections.

**MATERIALS AND METHODS**

**Plant materials and genome sequencing**

Seeds of wild Chinese *R. chinensis* were obtained from Shandong and Henan provinces in China. We also collected seven *R. chinensis* samples in Korea (Table 1): two samples from demilitarized zone (DMZ) between South and North Korea and five samples from other locations in South Korea. We sequenced plants collected from Yang-gu province in Korea and Shandong and Henan provinces in China as representatives of the collections of Korea and China. The genomic DNAs were extracted from the leaf tissues or seeds of collected plants using a modified cetyltrimethylammonium bromide (CTAB) method (Allen et al. 2006) and quantified using Nanodrop ND-1000 (Nanodrop Technologies, Inc., Wilmington, DE, USA). Paired-end (PE) sequencing was conducted using Illumina MiSeq platform by LabGenomics (www.labgenomics.co.kr, Seongnam, Korea).

**Chloroplast genome assembly**

Raw PE reads were trimmed and de novo assembled using the method as described by Kim et al. (2015a, 2015b). From initial assembly, contigs representing chloroplast genome sequences were extracted, ordered, and merged to generate a single contig sequence using the reference chloroplast genome *Acer buergerianum* ssp. *ningpoense* (KF753631, Yang et al. 2014). The assembled sequence was manually corrected and gap-filled by a series of PE read mapping. The assembled chloroplast genome was annotated using GeSeq (https://chlorobox.mpimp-golm.mpg.de/geseq-app.html) and manually curated using the Artemis annotation tool (Rutherford et al. 2000).

**Phylogenetic analysis**

Phylogenetic analysis was carried out using multiple sequence alignments of 13 complete chloroplast genome

| Accession nos. | Countries | Collected locations | Collected year |
|---------------|-----------|---------------------|----------------|
| 1             | China     | Shandong and Henan provinces | 2014 |
| 2             | Korea     | Yang-gu, Gangwon-do | 2014 |
| 3             | Korea     | Hwacheon, Gangwon-do | 2014 |
| 4             | Korea     | Mt.Jiri, Sancheong, Kyungsangnam-do, | 2014 |
| 5             | Korea     | Han-taek Botanical Garden, Yong-in, Kyungki-do | 2015 |
| 6             | Korea     | Mt. Kariwang, Jeong-seon, Gangwon-do | 2015 |
| 7             | Korea     | Gwangreung Botanical Garden, Korea National Arboretum, Po-cheon, Gyeonggi-do | 2015 |
| 8             | Korea     | Gwangreung Botanical Garden, Korea National Arboretum, Po-cheon, Gyeonggi-do | 2015 |
sequences (Acer burgerianum ssp. ningpoense, KF753631; Acer davidii, NC_030331; Acer miaotaïense, NC_030433; Azadirachta indica, NC_023792; Boswellia sacra, KT934315; Citrus aurantifolia, KJ865401; Citrus sinensis, DQ864733; Dipteronia sinensis, NC_029338; R. chinensis (China), MF351625; R. chinensis (Korea), NC_033535; Sapindus mukorossi, KM454982; Spondias bahiensis, KU756561; Spondias tuberosa, KU756562) that belong to the Sapindales order. Phylogenetic tree was constructed using MEGA6.0 (Tamura et al. 2013) with the parameters of neighbor-joining method and 1000 bootstrap replicates.

Comparison of intra-species level and development of molecular marker

Sequence variations were identified by mVISTA program (http://genome.lbl.gov/vista/mvista/submit.shtml) and the polymorphic sites were arranged by MAFFT program (http://mafft.cbrc.jp/alignment/software). The two types of molecular markers, InDel and SNP, were developed based on the polymorphic sites in chloroplast genomes of two R. chinensis. The primers were designed using Primer-blast tool in NCBI (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). PCR reactions were performed in 25 µL final volume which is composed of 20 ng of template DNA, 1x Taq buffer, 2.5 mM dNTP, 10 pmol of each primer, and 2 unit/µL Taq DNA polymerase (Vivagen, Korea). PCR conditions were as follows: 5 minutes at 95°C, 35 cycles of 30 seconds at 95°C, 30 seconds at 56°C, and 30 seconds at 72°C, and 5 minutes at 72°C as final extension. PCR amplicons were inspected using 3% agarose gel including Inclone™ Safe Gel stain. Then they were visualized under UV trans-illuminator and a gel documentation system.

PCR reaction for high resolution melting (HRM) analysis against SNP target was performed in a 20 µL final volume which consists of 20 ng of template DNA, 1x Taq buffer, 2.5 mM dNTP, 10 pmol of each primer, 2 unit/µL Taq DNA polymerase (Vivagen, Korea), and fluorescent dye SYTO 9 (Roche Diagnostics). PCR conditions were as follows: 5 minutes at 95°C, 45 cycles of 30 seconds at 95°C, 30 seconds at 56°C, and 30 seconds at 72°C, and 5 minutes at 72°C as final extension. PCR amplicons were inspected using 3% agarose gel including Inclone™ Safe Gel stain. Then they were visualized under UV trans-illuminator and a gel documentation system.

RESULTS

Complete chloroplast genome and intra-species polymorphism between Korean and Chinese R. chinensis

The assembled chloroplast genome of Chinese R. chinensis is a circular molecule of 149,094 bp long and has typical quadripartite structure consisting of a large single copy (LSC) region of 97,246 bp small single copy (SSC) region of 18,644 bp and a pair of inverted repeats (IRa and IRb) of 16,602 bp (Fig. 1, Table 2). A total of 111 genes, including 77 protein coding genes, 30 tRNA genes, and 4 rRNA genes were annotated (Table 3) and the GC content of the chloroplast genome was 37.86%.

Comparison of the chloroplast genome of Korean and Chinese R. chinensis revealed a total length 83 bp differences. Chinese R. chinensis is 364 bp longer than Korean R. chinensis in LSC region, but 139 bp and 3 bp shorter in IR and SSC regions, respectively. We identified 170 SNPs and 85 InDels between two R. chinensis chloroplast genome sequences derived from Chinese and Korean collections (NC_033535, Lee et al. 2016). More intra-species polymorphic sites were found at intergenic regions than genic regions: 99 SNPs and 78 InDels were found at intergenic regions; 71 SNPs and 7 InDels were identified from genic regions.

Phylogenomic analysis of R. chinensis based on chloroplast genome

Phylogenomic analysis was conducted using two complete chloroplast genomes R. chinensis and 11 complete chloroplast genome sequences in the Sapindales order. This results showed, as expected, that R. chinensis was grouped with Spondias species in the Anacardiaceae family (Fig. 2). In addition, a group including R. chinensis showed a close sister relationship with species belonging to the Burseraceae and Sapindaceae family in the Sapindales order.
**Table 2.** Summary of NGS data and chloroplast genomes of two *Rhus chinensis* collections.

| Collected locations | Raw data bases (bp) | Cp genome coverage (x) | Cp length (bp) | LSC length (bp) | IR length (bp) | SSC length (bp) |
|---------------------|---------------------|------------------------|----------------|-----------------|----------------|-----------------|
| China               | 4,199,252,923       | 162                    | 149,094        | 97,246          | 16,602         | 18,644          |
| Korea (Yang-gu)     | 6,251,344,649       | 96                     | 149,011        | 96,882          | 16,741         | 18,647          |

\[^{2}\] reported by Lee *et al.* (2016).
Table 3. Genes annotated in *R. chinensis* chloroplast genome.

| Gene types                        | Gene names                          |
|------------------------------------|-------------------------------------|
| Photosystem I                      | psaA, B, C, I, J                    |
| Photosystem II                     | psbA, B, C, D, E, F, H, I, J, K, L, M, N, T, Z |
| Cytochrome b/f complex             | petA, B, D, G, N, L                 |
| ATP synthase                       | atpA, B, E, F, H, I                 |
| NADH dehydrogenase                 | ndhA, B, C, D, E, F, G, H, I, J, K  |
| RubisCO large subunit              | rbcL                                |
| RNA polymerase                     | rpoA, B, C1, C2                     |
| Ribosomal proteins (SSU)           | rps2, 3, 4, 7, 8, 11, 12, 14, 15, 16, 18 |
| Ribosomal proteins (LSU)           | rpl2, 14, 16, 20, 23, 32, 33, 36    |
| clpP, matK                         | clpP, matK                          |
| Other genes                        | ccsA, cemA, accD                     |
| Hypothetical chloroplast reading frames | trn(M-CUA, A-UGC, C-GCA, D-GUC, E-UUC, F-GAA, G-GCC, UCC, H-GUC, I-CAU, GAU, K-UUU, L-CAA, CAU, UAG, UAA, M-CAU, N-GUU, P-UGG, Q-UUG, R-UCU, ACG, S-UGA, GCU, GGA, T-GGU, V-GAC, UAC, W-CCA, Y-GUA |
| Transfer RNAs                      | rRNA4.5, 5, 16, 23                  |
| Ribosomal RNAs                     |                                      |

Fig. 2. Phylogenetic analysis of *R. chinensis*. The tree was generated with complete chloroplast genome sequences of *R. chinensis* and species belonging to the Sapindales order by multiple alignment using MAFFT (http://mafft.cbrc.jp/alignment/server/index.html) and a maximum likelihood (ML) analysis using MEGA 6.0 (Tamura et al. 2013). Numbers in the nodes are bootstrap support values (>50%) from 1000 replicates. Chloroplast genome sequences used for this tree are: *Acer burgerianum* ssp. ningpoense, KF753631; *Acer davidii*, NC_030331; *Acer miaotaense*, NC_030343; *Azadirachta indica*, NC_023792; *Boswellia sacra*, KT934315; *Citrus aurantiifolia*, KJ865401; *Citrus sinensis*, DQ864733; *Dipteronia sinensis*, NC_029338; *R. chinensis* (China), MF351625; *R. chinensis* (Korea), NC_033535; *Sapindus mukorossi*, KM454982; *Spondias bahtiensis*, KU756561; *Spondias tuberosa*, KU756562.
Molecular markers to discriminate Korean and Chinese collections of *R. chinensis* species

Three intra-species polymorphic markers were designed from polymorphic regions (Table 4). The PCR amplicons showed 288/249 bp and 221/196 bp difference (Chinese/Korean) for two InDel markers, rh_InDel_02 and rh_InDel_03 markers, respectively. The rh_InDel_02 and rh_InDel_03 showed InDel polymorphism between Korean *R. chinensis* (Yang-gu) and Chinese collection as expected (Fig. 3a, b). However other six Korean collections were identical with Chinese collections (Fig. 3c, d). One SNP marker (rh_hrm_11) distinguished Chinese collection from all the Korean collections based on HRM analysis although the size of PCR amplicons are same (Fig. 4).

**DISCUSSION**

Intra-species diversity of *R. chinensis*

*R. chinensis* is known as a valuable medicinal plants with remedial components (Min and Barfod 2008; Djakpo and Yao 2010). In addition, *R. chinensis* is used for ecological restoration (Nam et al. 2004; Lim and Oh 2015). Due to its wide utility, it is necessary to explore this plant for genetic diversity and marker development to prevent false trading.
Fig. 4. SNP-based marker to validate intra-species diversity in *R. chinensis*. (a) Alignment of nucleotide sequences to show the SNP (arrow) in *ycf1* gene of chloroplast of Korean and Chinese *R. chinensis* plants. (b) Agarose gel electrophoresis of the PCR products amplified by Marker rh_hrm_11 (Table 4). (c) HRM analysis of the PCR products to detect the SNP. *R. chinensis* collections 1: China (Shandong and Henan), 2: Korea (Yang-gu), 3: Korea (Hwacheon), 4: Korea (Mt. Jiri), 5: Korea (Han-taek Botanical Garden), 6: Korea (Mt. Kariwang), 7, 8: Korea (Gwangreung Botanical Garden), M: 100-bp DNA ladder.

(Khairallh and Salama 2009). Here, we generated the complete chloroplast genome of two *R. chinensis* plants collected from Korea and China. Comparative analysis of chloroplast genome of *R. chinensis* revealed 255 intra-species polymorphic sites, 170 SNPs and 85 InDels. The abundant genetic diversity would be found in the wild collections of *R. chinensis* and could be applied to phylogenetic analysis and development of molecular markers for verifying genetic diversity of *R. chinensis*.

**Discrimination of *R. chinensis* collections with three molecular markers**

Two InDel markers were unique for one Korean collection of *R. chinensis* (Yang-gu) from the others (Fig. 3a, b). The Yang-gu collection is derived from the DMZ region between South Korea and North Korea. The eco-system was well maintained in DMZ regions for longer than 60 recent years with less artificial interruption. Two InDel markers revealed unique genotype of Yang-gu collection among the eight collections, suggesting the Yang-gu collection are relatively isolated in the DMZ region from others. However, it is required to expand our study by evaluating more population for solid conclusion. Meanwhile, the SNP markers could efficiently divide the melting patterns for Chinese *R. chinensis* from Korean *R. chinensis* through HRM analysis (Fig. 4).

Chloroplast genomes are valuable genomic resources for practical application of DNA markers because of the uni-parentally inherited features and conserved structure of the chloroplast genomes (Cheng *et al.* 2005; Dong *et al.* 2012, Kim *et al.* 2015a, 2015b, 2015c; Joh *et al.* 2017). Here, we report three DNA markers which can be applied for genetic diversity and practical application for authentication of the plant collections of *Rhus* species.

**ACKNOWLEDGEMENTS**

This work was supported by the Korea Environmental Industry & Technology Institute (Project No.: 2014000130001) and the Bio & Medical Technology Development Program of the NRF funded by the Korean government, MSIP.
REFERENCES

Allen GC, Flores-Vergara MA, Krasnianski S, Kumar S, Thompson WF. 2006. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. Nat. Protoc. 1: 2320-2325.

Cheng Y, de Vicente MC, Meng H, Guo W, Tao N, Deng X. 2005. A set of primers for analyzing chloroplast DNA diversity in Citrus and related genera. Tree Physiol. 25: 661-672.

Djakpo O, Yao W. 2010. Rhus chinensis and Galla Chinensis--folklore to modern evidence: review. Phytother. Res. 24: 1739-1747.

Dong W, Liu J, Yu J, Wang L, Zhou S. 2012. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. PLoS One 7: e35071.

Fujii N, Ueda K, Watano Y, Shimizu T. 1995. Intraspecific sequence variation in chloroplast DNA of Primula cuneifolia Ledeb. (Primulaceae). J. Phytogeogr. Taxon. 43: 15-24.

Fujii N, Ueda K, Watano Y, Shimizu T. 1997. Intraspecific sequence variation of chloroplast DNA in Pedicularis chamissonis Steven (Scrophulariaceae) and geographic structuring of the Japanese “Alpine” plants. Journal of Plant Research 110: 195-207.

Joh HJ, Kim NH, Jayakodzi M, Jang WJ, Park JY, Kim YC, et al. 2017. Authentication of golden-berry P. ginseng cultivar ‘Gumpoong’ from a landrace ‘Hwangsook’ based on pooling method using chloroplast-derived markers. Plant Breed. Biotech. 5: 16-24.

Khairallh ARAT, Salama B. 2009. Genetic variation among sumac (Rhus Coriaria L.) samples collected from three locations in Jordan as revealed by AFLP markers. Advances in Environmental Biology 3: 107-112.

Kim K, Lee SC, Lee J, Lee HO, Joh HJ, Kim NH, et al. 2015a. Comprehensive survey of genetic diversity in chloroplast genomes and 45S nrDNAs within Panax ginseng species. PLoS One 10: e0117159.

Kim K, Lee SC, Lee J, Yu Y, Yang K, Choi BS, et al. 2015b. Complete chloroplast and ribosomal sequences for 30 accessions elucidate evolution of Oryza AA genome species. Sci Rep 5: 15655.

Kim KH, Kim YS, Kim MR, Lee HY, Lee KH, Kim JH, et al. 2015c. Development of primer sets for the detection of Polygonum multiflorum, Cynanchum wilfordii and C. auriculatum. Journal of Food Hygiene and Safety 30: 289-294.

Lee YS, Kim IS, Kim JK, Park JY, Joh HJ, Park HS, et al. 2016. The complete chloroplast genome sequence of Rhus chinensis Mill (Anacardiaceae). Mitochondrial DNA Part B 1: 696-697.

Lee WK, Kim MJ, Heo K. 2004. Phylogeny of Korean Rhus spp. based on ITS and rbcL sequences. Korean Journal of Medicinal Crop Science 12: 60-66.

Lim CY, Oh CH. 2015. Growth and spatial distribution characteristics of Rhus javanica populations sowed on cut-slopes. J. Korean Env. Res. Tech. 18: 39-49.

Ma WL, Wu M, Wu Y, Ren ZM, Zhong Y. 2013. Cloning and characterisation of a phenylalanine ammonia-lyase gene from Rhus chinensis. Plant Cell Rep. 32: 1179-1190.

Miller AJ, Young DA, Wen J. 2001. Phylogeny and biogeography of Rhus (Anacardiaceae) based on ITS sequences data. Int. J. Plant Sci. 162: 1401-1407.

Min T, Barfod A. 2008. Anacardiaceae (http://www.efloras.org/florataxon.aspx?flora_id=2&taxon_id=10038). In: ZY. Wu, PH. Raven, DY. Hong (eds). Flora of China. Vol. 11. Science Press and Missouri Botanical Garden Press, Beijing, China and St. Louis, Missouri, (MO), USA.

Nam SJ, Yeo WJ, Choi JH, Kim NC. 2004. Development of revegetation method using forest topsoils for ecological restoration of the slopes (I). J. of Korean Env. Res. & Reveg. Tech. 7: 110-119.

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. 2000. Artemis: sequence visualization and annotation. Bioinformatics 16: 944-945.

Sewell MM, Parks CR, Chase MW. 1996. Intraspecific chloroplast DNA variation and biogeography of North American Liriodendron L. (Magnoliaceae). Evolution 50: 1147-1154.

Soltis DE, Mayer MS, Soltis PS, Edgerton M. 1991. Chloroplast-DNA variation in Tiarella trifoliata (Saxifragaceae). Plant Systematics and Evolution 181: 203-216.

Soltis DE, Mayer MS, Soltis PS, Edgerton M. 1991. Chloroplast-DNA variation in Tiarella trifoliata (Saxifragaceae). American Journal of Botany 78: 1379-1390.
Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30: 2725-2729.
Yang JB, Li DZ, Li HT. 2014. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. Mol. Ecol. Resour. 14: 1024-1031.
Yi T, Miller AJ, Wen J. 2004. Phylogenetic and biogeographic diversification of Rhus (Anacardiaceae) in the Northern Hemisphere. Mol. Phylogenet. Evol. 33: 861-879.
Yi T, Miller AJ, Wen J. 2007. Phylogeny of Rhus (Anacardiaceae) based on sequences of nuclear Nia-i3 intron and chloroplast trnC-trnD. Systematic Botany 32: 379-391.
Young DA. 1978. Re-evaluation of the section of Rhus L. subgenus Lobadium (Raf.) T. and G. (Anacardiaceae). Brittonia 30: 411-415.