

Analysis of the chloroplast genome and SNP detection in a salt tolerant breeding line in Korean ginseng

Ick-Hyun Jo · Kyong-Hwan Bang · Chi Eun Hong · Jang-Uk Kim · Jung-Woo Lee · Dong-Hwi Kim · Dong-Yun Hyun · Hojin Ryu · Young-Chang Kim

Received: 9 November 2016 / Revised: 14 November 2016 / Accepted: 15 November 2016

© Korean Society for Plant Biotechnology

Abstract The complete chloroplast genome sequence of *Panax ginseng* breeding line ‘G07006’, showing higher salt tolerance, was confirmed by *de novo* assembly using whole genome next-generation sequences. The complete chloroplast (CP) genome size is 156,356 bp, including two inverted repeats (IRs) of 52,060 bp, separated by the large single-copy (LSC 86,174 bp) and the small single-copy (SSC 18,122 bp) regions. One hundred fourteen genes were annotated, including 80 protein-coding genes, 30 tRNA genes, and 4 rRNA genes. Among them, 18 sites were duplicated in the inverted repeat regions. By comparative analyses of the previously identified CP genome sequences of nine cultivars of *P. ginseng* and that of G07006, five useful SNPs were defined in this study. Since three of the five SNPs were cultivar-specific to Chunpoong and Sunhyang, they could be easily used for distinguishing from other ginseng accessions. However, on arranging SNPs according to their gene location, the G07006 genotype was ‘GTGGA’, which was distinct from other accessions. This complete chloroplast DNA sequence could be conducive to discrimination of the line G07006 (salt-tolerant) and further enhancement of the genetic improvement program for this important medicinal plant.

Keywords *Panax ginseng*, Complete chloroplast genome, Single nucleotide polymorphisms

Introduction

Recent rapid changes in climate have resulted in rising temperature and salinity stresses, causing challenge for stable ginseng production (Kim et al. 2008; Jin et al. 2009). Particularly, salinity stress has decreased the annual ginseng yield up to 20% (Kim et al. 2015). The optimal soil salinity concentration for ginseng production is as low as 0.5 ds/m, and thus, it is a salinity-sensitive crop (Kim et al. 2014). Its growth is inhibited at high salinity concentrations and the plant withers in severe cases (Cho and Kim 2004). Thus, National Institute of Horticultural and Herbal Science (NIHHS) prepared an *in vitro* test system to develop a salinity-resistant ginseng cultivar and has selected salinity-resistant lines from the collected genetic resources. ‘G07006’ was selected for salt tolerance based on the changes of test markers, such as physiological responses and chlorophyll fluorescence, following immersion of petioles in 5 ds/m mixed salinity solution (21 mM KNO₃, 6.1 mM KH₂PO₄, 6.2 mM MgSO₄·7H₂O) at the NIHHS in 2014–2016 (Jo et al. 2015).

In order to develop this line into a new variety, it must be genetically stabilized as a pure line through repeated selections and segregations. To maintain the pure line during development, the genetic homogeneity of seeds must be tested, which requires a system that can distinguish the new variety from existing ones (Bang et al. 2015).

DNA barcoding is a standard method used to distinguish species based on single nucleotide polymorphisms (SNPs) in the same gene loci in the genomes of different organisms (Jo et al. 2014). No standardized gene locus is currently available to distinguish plant species; therefore, coding regions of chloroplast DNA, including *matK*, *rbcL*, *rpoB*, and *rpoC1*, and

I.-H. Jo · C.-E. Hong · J.-U. Kim · J.-W. Lee · D.-H. Kim · D.-Y. Hyun · Y.-C. Kim (✉)

Department of Herbal Crop Research, National Institute of Horticultural and Herbal Science, Rural Development Administration, Eumseong, 27709, Republic of Korea
e-mail: ycpiano@korea.kr

K.-H. Bang

Department of Planning and Coordination, National Institute of Horticultural and Herbal Science, Rural Development Administration, Wansan, 55365, Republic of Korea

H. Ryu (✉)

Department of Biology, Chungbuk National University, Cheongju 28644, Republic of Korea
e-mail: hjryu96@chungbuk.ac.kr

non-coding regions, including *atpF-atpH*, *psbK-psbI*, and *psbA-trnH* spacers, have been suggested (Kress et al. 2005). Chloroplast is a plant organelle containing its own circular DNA molecules that vary in size from 120 kb to 160 kb depending on the plant species. Chloroplast DNA (CpDNA) is largely composed of LSC, SSC, and they encode 4 rRNAs, 30 tRNAs, and 80 proteins on average, which are involved in photosynthesis or the expression of chromosomal genes (Palmer 1985; Jansen et al. 2005; Wang et al. 2008).

CpDNA, compared to genomic DNA, shows greater evolutionary conservation, facilitating phylogenetic analysis and species identification based on SNPs. In addition, CpDNA is inherited maternally, enabling evolutionary taxonomy analysis of plants, and is shorter than genomic DNA, thereby decreasing the cost of DNA sequencing and analysis by NGS (Nock et al. 2011; Straub et al. 2012; McPherson et al. 2013). The first full CpDNA sequence was reported in tobacco leaves in 1986 (Yamada et al. 1986). Following this report, numerous studies have been conducted on chloroplasts, and the full CpDNA sequences of approximately 300 species have been deposited in the NCBI database to date. Recently, Yang et al. (2015) analyzed the CpDNA sequence of nine varieties of ginseng and developed SNP- and indel-based DNA markers that can differentiate between the Chunpoong and Sunhyang varieties. Thus, this study was conducted to develop DNA markers capable of determining distinctness and homogeneity among the salt-resistant varieties of ginseng. For this, the full CpDNA sequence of ‘G07006’, a salt-resistant line, was analyzed using NGS and compared with the recently released CpDNA sequences of nine ginseng varieties (Chunpoong, Yunpoong, Gopoong, Gumpoong, Sunpoong, Sunun, Sunone, Cheongsun and Sunhyang).

Materials and Methods

Plant material

The *P. ginseng* breeding line ‘G07006’, was selected for salt tolerance through salinity screening of mature leaves at the NIHHS of the Rural Development Administration (RDA) in 2014–2016. Fourth years of its growing season, in July, the leaves were harvested, powdered using liquid nitrogen and stored in a deep freezer (-80°C) until further processing.

DNA extraction, Cp genome assembly and annotation

Total genomic DNA was extracted from fresh leaves of 4 year old plants of *P. ginseng* breeding line (G07006) by using Dneasy Plant Mini Kit (Qiagen, Hilden, Germany). The quantity

and quality of DNA samples were measured using a NanoDrop ND-1000 spectrophotometer (Nano-Drop Technologies, Wilmington, DE, USA). An Illumina paired-end (PE) library with 300-bp insert size was constructed using genomic DNA and sequenced using an Illumina HiSeq 2000 instrument (Illumina, San Diego, CA) according to the instructions of the manufacturer. High quality PE reads of 3.95 Gb were assembled using CLC genome assembler (v. beta 4.6, CLC Inc., Aarhus, Denmark), after trimming of reads (Phred scores of 20 or less). Afterward, representing chloroplast genome sequences were joined by comparison with reference sequences (*P. ginseng* cv. Yunpoong, NC_006290) (Kim and Lee 2004) and validated by PE read mapping. The chloroplast genome was annotated using DOGMA (<http://dogma.cccb.utexas.edu/>) and manual curation based on BLAST searches.

Identification of single nucleotide polymorphisms based on chloroplast genome sequences of nine *P. ginseng* cultivars

In this study, to identify SNP mutations in the CP genome of the decoded G07006 system based on the NCBI DB, CP genome sequences for nine cultivars of *P. ginseng* were downloaded for comparative analysis (KM067386, KM067387, KM067388, KM067390, KM067391, KM067392, KM067393, KM088019, KM088020). Base sequences were edited using the BioEdit program, saved as no gap, and arranged as Cluster W (ver. 1.83) to identify the location and type of common SNP.

Results and Discussion

Characterization of the chloroplast genome

NGS was used to obtain 3.95 Gb data from the decoded DNA base nucleotides of the G07006 system. Results showed that 492,591 contigs were within the cpDNA region and the coverage was identified to be 939× (Table 1). The total size of the completely mapped CP genome of the G07006 was 156,356 bp. This CP genome encodes 114 unigenes (80 protein-coding genes, four rRNA genes, and 30 tRNA genes), in which 18 are duplicated in the IR regions (Fig. 1). Comparative analyses of nine cultivars of *P. ginseng* with the CP genome showed that CP genome size of the G07006 was identical to that of the Gumpoong cultivar.

SNP detection of a breeding line in Korean ginseng

Comparative analyses of previously identified CP genome sequences of nine cultivars of *P. ginseng* and that of the G07006 showed five SNPs (Fig. 2). Sunhyang cultivar had a thymine

Table 1 Data regarding cpDNA of ‘G07006’ generated by the Illumina HiSeq sequencer

Species	Organism source	Raw data bases (bp)	Aligned reads to the cp genome (#)	Cp genome coverage (x)	Cp genome length (bp)
<i>P. ginseng</i>	G07006	3,955,213,989	492,591	939	156,356

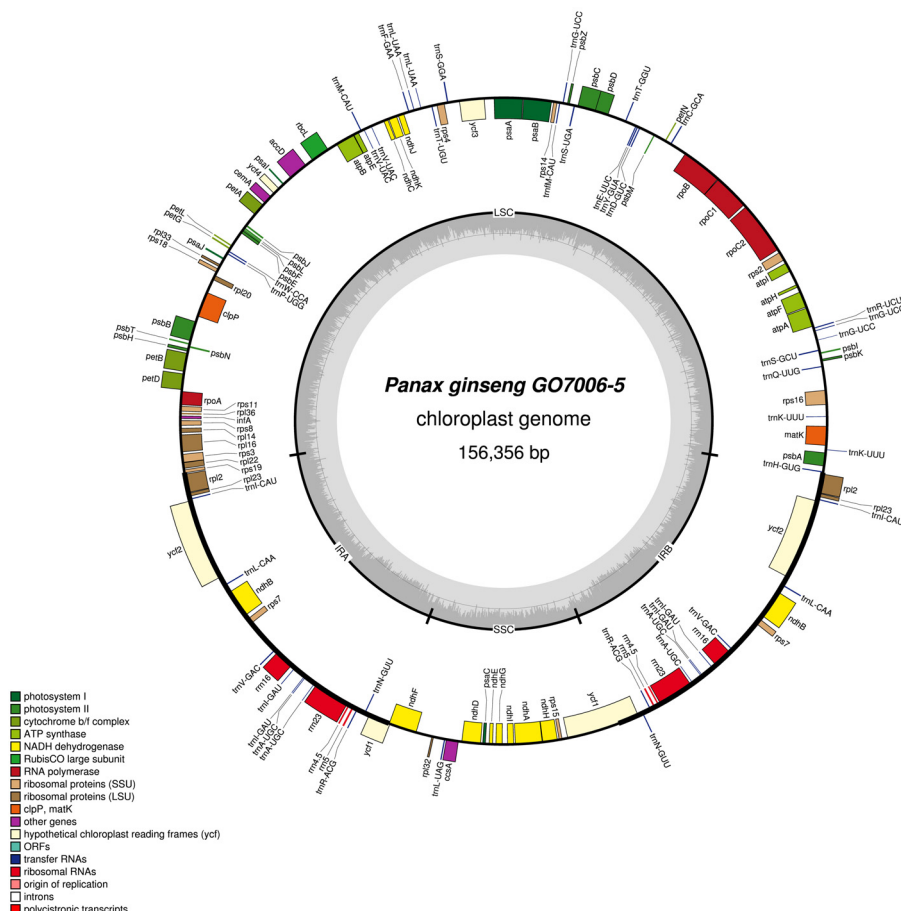


Fig. 1 Circular gene map of the breeding line of *Panax ginseng* (G07006) chloroplast genome. Genes shown inside the circle are transcribed clockwise, and those outside the circle are transcribed counterclockwise

(T) nucleotide at 7,159 bp in the *rps16*-*trnUUG* region, and other accessions had guanine (G) in that region. Further, Sunhyang had a T nucleotide in the *ndhF*-*rpl32* region, in contrast with that of other accessions, which had G. Therefore, base substitutions in *rps16*-*trnUUG* and *ndhF*-*rpl32* could be used as cultivar-specific markers to identify the Sunhyang. In addition, the *rpoC2* gene samples showed SNPs (located in the total CP genome of 21,344 bp): Gumpoong, Chungsun, and G07006 had T and other accessions had cytosine (C). The Chunpoong CP genome at 22,287 bp nucleotides (*rpoC1* gene) had T and the other accessions had G. Chunpoong had a unique root shape, and thus, was an appropriate strain for red ginseng production. Therefore, the specific SNP to Chunpoong at the *rpoC1* region could be used as a molecular marker to select suitable cultivar for red ginseng process production and the

management of product quality. Lastly, at the *ccsA* gene region, SNP types in accessions were divided into two groups. Gumpoong and Chungsun had G, and the others had adenine (A). Since three of five SNPs were cultivar-specific to Chunpoong and Sunhyang, they could be easily distinguished from other accessions. However, when arranging SNPs according to their gene locations, the G07006 genotype was ‘GTGGA’, which was distinct from other accessions (Table 2). G07006 is a salt-resistant line selected by a simplified test for salt resistance from 2014 to 2016. For such a cultivar to be cultivated, it is important to maintain a genetically uniform pure line. Therefore, the five selected SNPs from the completed CP genome sequence of the G07006 will provide valuable information to distinguish salt-resistant breeding line from other cultivars and maintain and preserve the pure line.

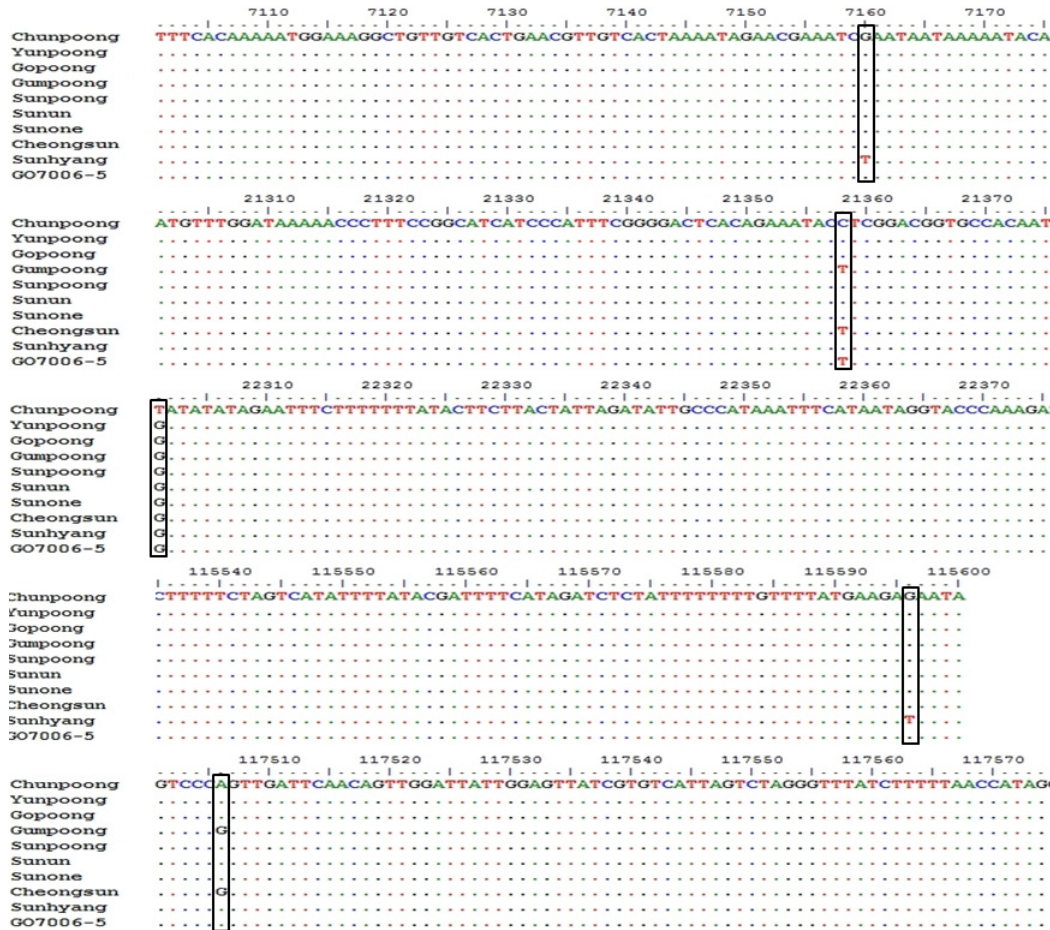


Fig. 2 Sequence alignment representing DNA variation detected in rps16-trnUUG, rpoC2, rpoC1, ndhF-rpl32 and ccsA gene regions of cpDNA of ten Korean ginseng accessions. The boxes indicate SNP positions among Korean ginseng accessions

Table 2 SNP combinations among Korean ginseng accessions by comparison of the cp genome

Type	Locus	SNP Position (bp)	<i>P. ginseng</i> cultivars ^a & breeding line ^b										
			CP	YP	GO	GP	SP	SU	SO	SH	CS	GO7006	
SNP	rps16-trnUUG	7159	G	G	G	G	G	G	G	G	T	G	G
	rpoC2	21344	C	C	C	T	C	C	C	C	T	T	T
	rpoC1	22287	T	G	G	G	G	G	G	G	G	G	G
	ndhF-rpl32	115594	G	G	G	G	G	G	G	G	T	G	G
	ccsA	117376	A	A	A	G	A	A	A	A	A	G	A
SNP combination			GCTGA	GCGGA	GCGGA	GTGGG	GCGGA	GCGGA	GCGGA	GCGGA	TCGTA	GTGGG	GTGGA

Underlined italic letters indicate a cultivar-specific SNP and allele combination

^aCP, ‘Chunpoong’; YP, ‘Yunpoong’; GO, ‘Gopoong’; GP, ‘Gumpoong’; SP, ‘Sunpoong’; SU, ‘Sunun’; SO, ‘Sunone’; SH, ‘Sunhyang’; CS, ‘Cheongseon’

^bG07006; salt-resistant breeding line

Acknowledgments

This work was carried out with the support of the “Cooperative Research Program for Agriculture Science & Technology Development (Project No. PJ010104012016),” Rural Development Administration, Republic of Korea.

References

Bang KH, Kim YC, Lim JY, Kim JU, Lee JW, Kim DH, Kim KH, Jo IH (2015) Internal transcribed spacer barcoding DNA region coupled with high resolution melting Analysis for authentication of *Panax* Species. Korean Journal of Medicinal Crop Science. 23:439-445

- Cho JW, Kim CS (2004) Photosynthetic response of Korean ginseng under saline condition. *Korean Journal of Crop Science*. 49:100-104
- Jansen RK, Raubeson LA, Boore JL, DePamphilis CW, Chumley TW, Haberle RC, Wyman SK, Alverson AJ, Peery R, Herman SJ, Fourcade HM, Kuehl JV, McNeal JR, Leebens-Mack J, Cui L (2005) Methods for obtaining and analyzing chloroplast genome sequences. *Methods in Enzymology*. 395:348-384
- Jin HO, Kim UJ, Yang DC (2009). Effect of nutritional environment in ginseng field on the plant growth of ginseng (*Panax ginseng* C. A. Meyer). *Journal of Ginseng Research*. 33:234-239
- Jo IH, Kim YC, Kim JU, Lee SH, Lim JY, Moon JY, Noh BS, Hyun DY, Kim DH, Kim KH, Bang KH (2014) A rapid identification of Korean ginseng cultivar, Cheonryang, using specific DNA markers. *Korean Journal of Medicinal Crop Science*. 22: 429-434
- Jo IH, Lee SH, Kim YC, Kim DH, Kim HS, Kim KH, Chung JW, Bang KH (2015) De novo transcriptome assembly and the identification of gene-associated single-nucleotide polymorphism markers in Asian and American ginseng roots. *Molecular Genetics and Genomics*. 290:1055-1065
- Kim JU, Hyun DY, Kim YC, Lee JW, Jo IH, Kim DH, Kim KH, Sohn JK (2015) Effects of salt in soil condition on chlorophyll fluorescence and physiological disorder in *Panax ginseng* C. A. Meyer. *Korean Journal of Medicinal Crop Science*. 23: 446-453
- Kim K, Lee SC, Lee J, Lee HO, Joh HJ, Kim N.H, Park HS, Yang TJ (2015) Comprehensive survey of genetic diversity in chloroplast genomes and 45S nrDNAs within *Panax ginseng* species. *PLoS One*. 10: e0117159
- Kim KJ, Lee HL (2004) Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA research*. 11:247-261
- Kim ST, Bae DW, Lee KH, Hwang JE, Bang KH, Kim YC, Kim OT, Yoo NH, Kang KY, Hyun DY, Lim CO (2008) Proteomic analysis of Korean ginseng (*Panax ginseng* C. A. Meyer) following exposure to salt stress. *Journal of Plant Biotechnology*. 35:185-193
- Kim SW, Min CW, Gupta R, Jo IH, Bang KH, Kim YC, Kim KH, Kim ST (2014) Proteomics analysis of early salt-responsive proteins in Ginseng (*Panax ginseng* C. A. Meyer) leaves. *Korean Journal of Medicinal Crop Science*. 22:398-404
- Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*. 102:8369-8374
- McPherson H, van der Merwe M, Delaney SK, Edwards MA, Henry RJ, McIntosh E, Rymer PD, Milner ML, Siow J and Rossetto M (2013) Capturing chloroplast variation for molecular ecology studies: a simple next generation sequencing approach applied to a rainforest tree. *BMC ecology*. 13:8
- Nock CJ, Waters DL, Edwards MA, Bowen SG, Rice N, Cordeiro GM, Henry RJ (2011) Chloroplast genome sequences from total DNA for plant identification. *Plant biotechnology journal*. 9:328-333
- Palmer JD (1985) Comparative organization of chloroplast genomes. *Annual Review of Genetics*. 19:325-354
- Straub SC, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A (2012) Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany*. 99:349-364
- Wang RJ, Cheng CL, Chang CC, Wu CL, Su TM, Chaw SM (2008) Dynamics and evolution of the inverted repeat-large single copy junctions in the chloroplast genomes of monocots. *BMC Evolutionary Biology*. 8:36
- Yamada K, Shinozaki K, Sugiura M (1986) DNA sequences of tobacco chloroplast genes for tRNASer (GGA), tRNAThr (UGU), tRNALeu (UAA), tRNAPhe (GAA): the tRNALeu gene contains a 503 bp intron. *Plant molecular biology*. 6: 193-199