Research Article

Comprehensive comparative analysis of chloroplast genomes from seven Panax species and development of an authentication system based on species–unique single nucleotide polymorphism markers

Van Binh Nguyen1, Vo Ngoc Linh Giang1, Nomar Espinosa Waminal1, Hyun-Seung Park1, Nam-Hoon Kim1, Woojong Jang1, Junki Lee1, Tae-Jin Yang1,2,*

1Department of Plant Science, Plant Genomics and Breeding Institute, Research Institute of Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University, Seoul, Republic of Korea
2Crop Biotechnology Institute/GreenBio Science and Technology, Seoul National University, Pyeongchang, Republic of Korea

Abstract

Background: Panax species are important herbal medicinal plants in the Araliaceae family. Recently, we reported the complete chloroplast genomes and 45S nuclear ribosomal DNA sequences from seven Panax species, two (P. quinquefolius and P. trifolius) from North America and five (P. ginseng, P. notoginseng, P. japonicus, P. vietnamensis, and P. stipuleanatus) from Asia.

Methods: We conducted phylogenetic analysis of these chloroplast sequences with 12 other Araliaceae species and comprehensive comparative analysis among the seven Panax whole chloroplast genomes. Results: We identified 1,128 single nucleotide polymorphisms (SNP) in coding gene sequences, distributed among 72 of the 79 protein-coding genes in the chloroplast genomes of the seven Panax species. The other seven genes (including psaJ, psbN, rpl23, psbF, psbL, rps18, and rps5) were identical among the Panax species. We also discovered that 12 large chloroplast genome fragments were transferred into the mitochondrial genome based on sharing of more than 90% sequence similarity. The total size of transferred fragments was 60,331 bp, corresponding to approximately 38.6% of chloroplast genome. We developed 18 SNP markers from the chloroplast genic coding sequence regions that were not similar to regions in the mitochondrial genome. These markers included two or three species-specific markers for each species and can be used to authenticate all the seven Panax species from the others.

Conclusion: The comparative analysis of chloroplast genomes from seven Panax species elucidated their genetic diversity and evolutionary relationships, and 18 species-specific markers were able to discriminate among these species, thereby furthering efforts to protect the ginseng industry from economically motivated adulteration.

© 2018 The Korean Society of Ginseng. Publishing services by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Introduction

Panax (ginseng) species are widely distributed from high altitude freeze-free regions including the Eastern Himalayas, the Hoang Lien Son, and the Annamite mountain range to the freezing winter regions of Northeastern Asia and North America. Ginseng contains many important pharmaceuticals that have been used in traditional medicine for thousands of years. Ginseng is also becoming one of the most important national agricultural commodities not only in Asian countries such as Korea, China, and Vietnam but also in Russia, Canada, and United States. Of the 14 known species in the Panax genus, five species (Panax ginseng, P. quinquefolius, P. notoginseng, P. japonicus, and P. vietnamensis) are used as expensive herbal medicines in Korea, United States, China, Japan, and Vietnam. However, limited genetic information is available on other species such as P. stipuleanatus and P. trifolius.

* Corresponding author. Department of Plant Science, Plant Genomics and Breeding Institute, Research Institute of Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Republic of Korea.
E-mail address: tjiang@snu.ac.kr (T.-J. Yang).

These authors contributed equally to this work.
Notable therapeutic effects of ginseng on life-threatening diseases such as neurodegenerative diseases [1,2], cardiovascular diseases [3], diabetes [4], and cancer [5,6] are well documented. Owing to the high pharmacological and economical value of ginseng, many economically motivated adulterations (EMAs) of ginseng products have occurred [7]. Traditional methods for authentication of herb plants primarily depend on morphological and histological characteristics. However, morphological and histological authentication methods are not precise enough to distinguish among ginseng species because of their similar morphological appearances and intraspecies morphological differences caused by variation in growing conditions. Moreover, almost all commercial ginseng products are sold in various forms such as dried root, powder, liquid extracts, or other processed products, which are impossible to authenticate based on morphology. Methods of ginsenoside profiling have been developed for authentication of ginseng [8–11]; however, their applications are limited because ginsenosides are secondary metabolites and their accumulation varies among different tissues (such as roots, leaves, stems, flower buds, and berries) [12,13], cultivars [14], age [12,15], environmental conditions [16,17], storage conditions, and manufacturing processes [7].

Chloroplasts are multifunctional organelles required for photosynthesis and carbon fixation that contain their own genetic material. Chloroplast genomes are highly conserved in plants, with a quadruplicate structure comprising two copies of inverted repeat (IR) regions that separate the large and small single-copy (LSC and SSC, respectively) regions. The chloroplast genome size in angiosperms ranges from 115 to 165 kb [18]. Since the emergence of next-generation sequencing, the number of completely sequenced chloroplast genomes rapidly increased. As of September 2017, more than 1541 complete chloroplast genomes from land plants are available in the GenBank Organelle Genome Resources. Of these, five chloroplast genomes from the Panax genus have been sequenced [19].

Sequence-based DNA markers are advantageous and powerful tools used in species identification with high accuracy, simplicity, and time and cost efficiency [7]. Various types of DNA markers have been applied to the authentication of Panax species including nuclear genome-derived random amplified polymorphic DNA [20], microsatellites [21], and expressed sequence tag–simple sequence repeats [22,23]. However, these nuclear genome-derived DNA markers are usually used to analyze intraspecies level diversity. DNA markers derived from the chloroplast genome have been widely used and are considered to be the best barcoding targets for plant species identification [24] because of their highly conserved structure and high copy numbers that are easily detected. Chloroplast genome divergence is lower at the intraspecies specific level and higher at the interspecies specific level. Recently, chloroplast-derived DNA markers were developed to authenticate ginseng, including markers of single nucleotide polymorphisms (SNPs) and insertions or deletions (InDels) [25–27]. However, these markers are still of limited use due to the lack of genomic information for intra-species and interspecies variations.

Recently, we obtained complete chloroplast genome and nuclear ribosomal DNA sequences from five major Panax species [19,25] and two basal Panax species [28] from either Asia or North America by de novo assembly using low-coverage whole-genome shotgun next-generation sequencing (dnapLCW) [29]. Using this information, we previously developed InDel-based authentication markers among the five species [7]. Although these markers are easy to apply, their usefulness is somewhat limited by the relatively rich intraspecies polymorphism at the InDel regions. In this study, we conducted a comprehensive comparative genomics study of the chloroplast genomes from the seven Panax species and identified 18 chloroplast CDS-derived SNP markers that can be used to authenticate each of the seven species. This study provides valuable genetic information as well as a practical marker system for authentication of each Panax species that will be very helpful for regulating the ginseng industry.

Materials and methods

Plant materials and genomic DNA extraction

P. ginseng cultivars and P. quinquefolius plants were collected from the ginseng farm at Seoul National University in Suwon, Korea. P. notoginseng and P. japonicus plants were collected from Dafang County, Guizhou Province, and Enshi County, Hubei Province, China, respectively. P. vietnamensis and P. stipuleanatus plants were collected from Kon Tum and Lao Cai Province, Vietnam, respectively. P. trifolius plants were collected from North Eastern America. DNA was extracted from leaves and roots using a modified cetyltrimethylammonium bromide method [30]. The quality and quantity of extracted genomic DNA was measured using a UV-spectrophotometer and agarose gel electrophoresis.

Phylogenetic analysis

Phylogenetic tree construction and the reliability assessment of internal branches were conducted using the maximum likelihood method with 1,000 bootstrap replicates using MEGA 6.0 [31].

Comparative analysis of 79 protein-coding genes between seven Panax species

The chloroplast genome sequences of 11 P. ginseng cultivars (ChP_KM088019, YP_KM088020, GU_KM067388, GO_KM067387, SP_KM067391, SO_KM067390, SU_KM067392, SH_KM067393, CS_KM067386, HS_KM067394, JK_KM067389), two P. quinquefolius (KM088018, KT028714), four P. notoginseng (KP036468, KP036469, KP036471, KP036470), one P. japonicus (KP036469), two P. vietnamensis (KP036471, KP036470), one P. stipuleanatus (KX247147), and one P. trifolius (MF100782) were obtained from our previous studies [19,25,28] and GenBank. Chloroplast protein-coding gene sequences were extracted using Artemis [32] and manually curated. These chloroplast CDS regions were concatenated and aligned using the MAFFT program (http://mafft.cbrc.jp/alignment/server/). The SNPs from 79 CDSs were

| Species          | Whole genome sequencing data (Mb) | Sequence reads used Amounts (Mb) | Chloroplast coverage (x) | Chloroplast genome length (bp) |
|------------------|---------------------------------|---------------------------------|--------------------------|-----------------------------|
| P. ginseng       | 10,418                          | 505                             | 97                       | 156,248                     |
|                  |                                  |                                 |                          | (KM088019)                  |
| P. quinquefolius | 3,557                           | 1,010                           | 127                      | 156,088                     |
|                  |                                  |                                 |                          | (KM088018)                  |
| P. notoginseng   | 5,619                           | 2,811                           | 246                      | 156,466                     |
|                  |                                  |                                 |                          | (KP036468)                  |
| P. japonicus     | 5,738                           | 2,870                           | 237                      | 156,188                     |
|                  |                                  |                                 |                          | (KP036469)                  |
| P. vietnamensis  | 7,541                           | 4,586                           | 1,005                    | 155,993                     |
|                  |                                  |                                 |                          | (KP036470)                  |
| P. stipuleanatus | 2,218                           | 599                             | 154                      | 156,064                     |
|                  |                                  |                                 |                          | (KX247147)                  |
| P. trifolius     | 14,657                          | 2,300                           | 993                      | 156,157                     |
|                  |                                  |                                 |                          | (MF100782)                  |
identified using MEGA 6 [31]; then the SNPs that were located on
the chloroplast CDS maps from seven Panax species were identified
using Circos v.0.67 [33].

Identification of chloroplast gene insertion in mitochondria

The mitochondrial genome of P. ginseng was retrieved from
GenBank (KF735063) and mapped to chloroplast genomes to
eliminate BLAST hits of transferred genes between chloroplast and
mitochondrial genomes. The maps of chloroplast and mitochon-
drial genomes from Panax species as well as the fragments of gene
transfers were drawn using Circos v.0.67 [33].

Development and validation of derived cleaved amplified polymorphic sequence markers

To discriminate among the seven Panax species, we used poly-
morphisms in the chloroplast CDSs. For this, derived cleaved
amplified polymorphic sequence (dCAPS) primers were designed
based on SNP polymorphic sites after eliminating intraspecies
polymorphic sites and chloroplast gene transfer regions. The
dCAPSs were designed to create restriction enzyme cut sites using
dCAPS Finder 2.0 (http://helix.wustl.edu/dcaps/dcaps.html), and
the specific primers were designed using the Primer3 program
(http://bioinfo.ut.ee/primer3-0.4.0/).

Fig. 1. Complete chloroplast genomes from seven Panax species. Colored boxes represent conserved chloroplast genes that were classified based on product function. Genes shown inside the circle are transcribed clockwise, and those outside the circle are transcribed counterclockwise. Genes belonging to different functional groups are color-coded. The dashed area in the inner circle indicates the GC content.
Polymerase chain reaction (PCR) was carried out in a 25 μl reaction mixture containing 2.5 μl of 10× reaction buffer, 1.25 mM deoxynucleotide triphosphate, 5 pmol of each primer, 1.25 units of *Taq* DNA polymerase (Inclone, Korea), and 20 ng of DNA template. The PCR reaction was performed in thermocyclers using the following cycling parameters: 94°C (5 min); 35 cycles of 94°C (30 s), 56–62°C (30 s), then 72°C (7 min). PCR products were visualized on agarose gels (2.0–3.0%) containing safe gel stain (Inclone, Korea).

Analytical restriction enzyme reactions were performed in a volume of 10 μl containing 5 μl of PCR product, 1 μl of restriction enzyme buffer, and 0.3 μl (10 units) of restriction enzyme. The reaction mixtures were incubated at the optimum temperature for 3 hours or overnight, then visualized on agarose gels (2.0–3.0%) containing safe gel stain.

Results

Characteristics of the complete chloroplast genomes from seven *Panax* species

Complete chloroplast genome length from the seven *Panax* species ranged from 155,993 bp to 156,466 bp (Table 1). These chloroplast genomes had a typical quadripartite structure, consisting of a pair of IRs separated by the LSC and SSC regions (Fig. 1). There were no structural variations except for small InDels and SNPs. Each genome contained 113 functional genes, including 79 protein-coding genes, 30 transfer RNA genes, and four ribosomal RNA genes. The gene map for the seven *Panax* chloroplast genomes is shown in Fig. 1.

Phylogenomic analysis of 19 complete chloroplast genomes from Araliaceae species

Phylogenetic relationships were inferred using the entire chloroplast genome sequences from 19 species in the Araliaceae family. The results indicate that the nine genera in Araliaceae were divided into two typical monophyletic lineages consisting of the *Aralia–Panax* group and the other group with the seven remaining genera (Fig. 2). Species from each genus, *Panax*, *Aralia*, *Schefflera*, *Dendropanax*, *Eleutherococcus*, *Brassaiopsis*, *Fatsia*, *Kalopanax*, and *Metapanax* were grouped accordingly. Based on the phylogenetic tree, the seven *Panax* species were divided among a few subgroups, in which *P. stipuleanatus* and *P. trifolius* diverged from the common ancestor earlier than the other five *Panax* species (Fig. 2).

SNPs in chloroplast genomes of seven *Panax* species

SNPs were identified in the chloroplast genomes from seven *Panax* species and were compared among them to develop SNP-derived markers for authentication. In total, 1,783 SNP sites were identified in the whole chloroplast genome sequences of seven species, and of these, 1,128 sites were in protein-coding regions, i.e., CDSs. Despite having more SNP sites, the total number of SNP types in CDS regions accounted for less than half of all SNPs in whole chloroplast genome sequences because multiple SNP types were often found in a given site in the noncoding regions (Table 2). The two closest tetraploid species (*P. ginseng* and *P. quinquefolius*) had a lower number of SNPs in both CDSs and whole chloroplast genome sequences than any other pair (Table 2). *P. trifolius* had the highest numbers of SNPs in both CDS and whole chloroplast sequences in...
SNPs were distributed in 72 of the 79 protein-coding gene sequences of seven *Panax* species, the exceptions being seven highly conserved genes including *psaJ*, *psbN*, *rpl23*, *psbf*, *psbl*, *rps18*, and *rps7* (Fig. 3). SNP density was lower in IR regions than in LSC and SSC regions (Fig. 3).

**Table 2**  
Number of SNPs among seven *Panax* chloroplast genomes.

| CDS regions | PG     | PQ     | PN     | PJ     | PV     | PS     | PT     |
|-------------|--------|--------|--------|--------|--------|--------|--------|
| PG          | /      | 131    | 460    | 495    | 531    | 1157   | 1485   |
| PQ          | 59     | /      | 493    | 496    | 518    | 1145   | 1479   |
| PN          | 171    | 210    | /      | 476    | 535    | 1159   | 1514   |
| PJ          | 183    | 220    | 183    | /      | 316    | 1150   | 1513   |
| PV          | 246    | 245    | 243    | 157    | /      | 1196   | 1555   |
| PS          | 497    | 534    | 522    | 524    | 566    | 664    | 639    |
| PT          | 594    | 610    | 621    | 624    | 664    | 639    |        |

Number of SNPs from the entire chloroplast genome sequences and number of SNPs from 79 protein coding sequences are shown above and below the self-comparison diagonal, respectively.

*Characterization of chloroplast genome transfer into the mitochondrial genome*

The mitochondrial genome sequence of *P. ginseng* retrieved from GenBank is 464,680 bp, which is approximately 3 times larger than...
the chloroplast genome and consists of 94 functional genes (Fig. 4). We identified 12 large chloroplast genomes fragments in the mitochondrial genome. The fragments ranged from 2,297 to 8,250 bp and retained ≥90% sequence identity with their original chloroplast counterparts (Fig. 4). The combined total size of these fragments was 60,331 bp, which corresponds to approximately 38.6% of chloroplast genome (Fig. 4). Collectively, the gene transfer regions spanned almost 49 chloroplast genes as well as intergenic regions (Fig. 4).

Identification of species-specific SNP markers for authentication of the seven Panax species

A total of 18 dCAPS markers were selected from the species-specific SNP targets among the seven Panax species. Each of these SNP targets was derived from CDS regions and showed a unique polymorphism in one of the seven Panax species. At least two unique dCAPS markers were selected for each species, for a total of 18 (Table 3). Each of these markers resulted in the expected band sizes before and after restriction enzyme digestion (Fig. 5). Markers Pgd1—3 that were derived from the rpl20, ndhK, and rps15 gene sequences, respectively, were specific to P. ginseng and resulted in different band sizes when digested compared to other species (Fig. 5). Markers Pqd4—6 were derived from rpoC1, ndhA, and ndhK sequences, respectively, and resulted in a unique digestion pattern for P. quinquefolius (Fig. 5). Markers Pnd7—9 were derived from rpoC1, rpoC2, and ndhK sequences, respectively, and resulted in a unique digestion pattern for P. notoginseng (Fig. 5). Markers Pjd10 and 11 were derived from the rpoC2 and rpoB sequences, respectively, and their digestion pattern was

Fig. 4. Schematic representation of gene transfer between the chloroplast and mitochondrial genomes from Panax species. Each gray line within the circle shows the regions of chloroplast genome that has been inserted into the indicated location in the mitochondrial genome. Colored boxes show conserved chloroplast genes, classified based on product function. Genes shown inside the circle are transcribed clockwise, and those outside the circle are transcribed counterclockwise.
unique for \textit{P. japonicus}, while markers PvdM12 and 13 were derived from \textit{rpoC2} and \textit{ndhH} genes, respectively, and resulted in a digestion pattern that was unique for \textit{P. vietnamensis} (Fig. 5). Markers PvdM14–16 were derived from \textit{psbB}, \textit{rpoC1}, and \textit{rpoB}, respectively, and resulted in a digestion pattern that was unique for \textit{P. stipuleanatus} (Fig. 5). Two markers, PvdM17 and 18 were derived from \textit{ndhA} and \textit{rpoC1}, respectively, and resulted in a unique digestion pattern for \textit{P. trifolius} (Fig. 5). All 18 markers were practical and successful for distinguishing among the seven \textit{Panax} species and can therefore be applied to ginseng species authentication.

### Discussion

**Characterization of complete chloroplast genome structures provides valuable genetic information for \textit{Panax} species**

Chloroplast DNA sequences are useful in genetic engineering [34], DNA barcoding [35], and studying evolutionary relationships among plants [36,37]. With recent technical advances in DNA sequencing, the number of completely sequenced chloroplast genomes has grown rapidly. However, the complete chloroplast genome sequences for many high-value plant species are not available yet because of the high cost of sequencing [38]. In our previous studies, we applied a \textit{de novo} assembly method using dnaLCW [29] to obtain complete chloroplast genomes of five \textit{Panax} species [19]. Here, we added the complete chloroplast genomes of two more basal \textit{Panax} species [28] for comparative structure analysis. All seven chloroplast genome sequences were supported by an average read-mapping coverage of 97–1,005x (Table 1). Among the seven \textit{Panax} species examined here, the chloroplast genome structures are identical except for small InDels and different numbers of SNPs. These seven complete chloroplast genomes will provide more valuable genetic information for the study of the evolutionary relationships, breeding, and authentication of ginseng species.

**Phylogenetic analysis**

The Araliaceae is a family of flowering plants that consists of about 70 genera and approximately 750 species that vary in type from trees and shrubs to lianas and perennial herbs [39]. Araliaceae speciation is predicted to have occurred in two particular regions of North America and South East Asia [39]. Furthermore, the diversification and speciation were associated with whole genome duplication (WGD) or polyploidy events [40–42]. Previous studies indicated that two tetraploid \textit{Panax} species, \textit{P. ginseng} and \textit{P. quinquefolius}, have undergone two rounds of WGD [43,44]. These WGD events, along with geographic and ecological isolation, have contributed to the diversification of \textit{Panax} species [45]. Taxonomy of \textit{Panax} that is based on the morphological characteristics is considered controversial due to the complicated morphological
variation between intra-species and inter-species according to geographic and ecological environment.

Our phylogenetic tree based on whole-chloroplast genome sequences clearly showed the evolutionary relationship between Panax species and between genera in the Araliaceae family. In particular, our results indicated that the diploid species *P. trifolius*, which diverged from a common ancestor earlier and migrated to North America, was not involved in the tetraploidization of *P. ginseng* and *P. quinquefolius*. Another diploid species (*P. stipuleanatus*) which diverged earlier than the five remaining species had an overlapping distribution with the three diploid species group in Southeast Asia (*P. notoginseng*, *P. vietnamensis*, and *P. japonicus*). Two tetraploid species, *P. ginseng* and *P. quinquefolius*, which are involved in the recent second WGD, had diverged from the group of three diploid species and located in Northeastern Asia and North America due to geographic isolation (Fig. 2).

Comparative analysis of Panax chloroplast genomes

The number of SNPs at the intraspecies level is very low compared to that at the interspecies level. SNPs within the whole chloroplast genomes from 12 *P. ginseng* cultivars are rare, with only six SNPs identified in 12 *P. ginseng* cultivars [25]. By contrast, a total of 1,783 and 1,128 interspecies SNP sites were identified among seven Panax species in the whole chloroplast genome and protein-coding gene sequences, respectively (Table 2). Nevertheless, chloroplast genomes are highly conserved within the *Panax* genus, displaying high similarity (≥97.6%) at the nucleotide sequence level. In our previous study, we found that some chloroplast protein-coding genes are highly divergent while others are highly conserved among different Araliaceae species. Four genes, *infA*, *rpl22*, *rps19*, and *ndhE*, were more divergent and displayed large numbers of SNPs between different species. By contrast, *atpF*, *atpE*, *ycf2*, and *rps15*, had a high number of nonsynonymous mutations which might be related to evolution under positive selection [19]. However, some genes were highly conserved at the family (Araliaceae) level, such as *petN*, *psaJ*, *psbN*, and *rpl23*, or even at the order (Apiales) level, such as *psbF* [19]. The current study is consistent with these findings except *petN* gene, and in addition to four of the five chloroplast-encoded highly conserved genes (*psaJ*, *psbN*, *rpl23*, and *psbF*), we found three more among the seven Panax species (*psbL*, *rps18*, and *rps7*) (Fig. 3).

Chloroplast genome fragments were found in mitochondrial genomes

The sequencing of different genomes (nuclear, chloroplast, and mitochondrial) has uncovered staggering amounts of intracellular gene transfer between them [46,47]. Studies have shown that there is a high frequency of organelle DNA transfer to the nucleus in angiosperms [48–50]. Interorganelle genome transfer from chloroplast to mitochondrial genomes is also reported recently as a common phenomenon in higher plants in the course of evolution [48,51]. We identified 12 large fragments of chloroplast genome (representing 38.6% of the chloroplast genome) in mitochondrial genomes from *Panax* species, including both genes and intergenic
regions (Fig. 4). Genome transfer can result in assembly errors in chloroplast or mitochondrial genomes due to the high sequence similarity between the original chloroplast genome and the transferred chloroplast genome segments in mitochondrial genome. Moreover, the study of evolution or the development of molecular markers within gene transfer regions can generate confusing or biased results [7]. To counter this limitation, we examined all the gene transfer regions and removed all SNPs in these regions from our analysis before developing SNP-derived markers for authentication.

Use of dCAPS markers for ginseng species authentication

DNA barcoding may be defined as the use of short DNA sequences from either nuclear or organelle genomes to identify a species. DNA barcoding is a new technique that is widely used as a biological tool for species identification, breeding, and evolutionary research [52]. Identification of plant species is important for standardizing the food and herbal medicine industries and for preventing EMAs. Since ginseng has a high pharmacological and economical value, there is ample potential for EMA of ginseng products. Therefore, easy, reliable, and practical methods that accurately identify the origins of ginseng products play an important role in the development and protection of the ginseng industry.

Chloroplast genomes are endemic to plants, smaller in size, and have hundreds of copies in a cell as compared to the nuclear genome. Furthermore, since the chloroplast genome has sufficient interspecific divergence coupled with low intraspecific variation, chloroplast genome–based DNA barcodes are the best targets for methods of species authentication [24]. Recently, chloroplast genome sequences have been used to develop markers for ginseng authentication [7,26,72,27,53]; however, this method can be applied only to certain species because of a lack of information about variation at the intra-species and interspecies levels.

In this study, we developed 18 CDS-derived, species-specific, SNP markers from chloroplast genomes for the authentication of seven Panax species including five representative Panax species and two basal Panax species from Asia and North America. Recently we developed cultivar-unique markers for P. ginseng based on a comprehensive comparative genomic analysis of the chloroplast genome sequences from 12 ginseng cultivars [25]. We excluded those intraspecies polymorphic markers in this study because the aim of this study is to distinguish among different species. We also excluded the chloroplast genome targets that were transferred into mitochondrial genomes. All 18 dCAPS markers presented in this study are unique for one of the seven species and can be practically applied toward species authentication and breeding.

Conflicts of interest

The authors have no conflicts of interest to declare.

Acknowledgments

This work was supported by the Cooperative Research Program for Agriculture Science & Technology Development of the Rural Development Administration [PJ01103001], Republic of Korea.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jgr.2018.06.003.

References

[1] Radad K, Gille G, Liu L, Rausch WD. Use of ginseng in medicine with emphasis on neurodegenerative disorders. J Pharmaco Sci 2006;100:175–86.
[2] Cho IH. Effects of Panax ginseng in neurodegenerative diseases. J Ginseng Res 2012;36:342.
[3] Zheng SD, Wu HJ, Wu DL. Roles and mechanisms of ginseng in protecting human health. Chin J Integr Med 2012;18:548–55.
[4] Xie JT, Mehdendale SR, Li X, Quigg R, Wang X, Wang CZ, Wu JY, Aung HH, A-Rue P, Bell GI, et al. Anti-diabetic effect of ginsenoside Re in oob/o mice. Biochim Biophys Acta 2005;1740:319–25.
[5] Jung HJ, Choi H, Lim BW, Shin D, Kim H, Kwon B, Lee JE, Park IH, Lim CJ. Enhancement of anti-inflammatory and antiinseptic actions of red ginseng extract by fermentation. J Pharm Pharmacol 2012;64:756–62.
[6] Wong AS, Che CM, Leung KW. Recent advances in ginseng as cancer therapeutics: a functional and mechanistic overview. Nat Prod Rep 2015;32:256–72.
[7] Nguyen VB, Park HS, Lee SC, Lee J, Park JY, Yang TJ. Authentication markers for five major Panax species developed via comparative analysis of complete chloroplast genome sequences. J Agric Food Chem 2017;65:6298–306.
[8] Chan T, But P, Cheng S, Kwook I, Lau F, Xu H. Differentiation and authentication of Panax ginseng, Panax quinquefolius, and ginseng products by using HPLC/MS. Anal Chem 2000;72:1281–7.
[9] Yuj K, McIntyre KL, Fischer C, Hicks J, Colson KL, Lui E, Brown D, Arnason JT. Distinguishing Ontario ginseng landraces and ginseng species using NMR-based metabolomics. Bioanal Chem 2013;405:4409–509.
[10] Wang X, Sakuma T, Asafu-Adjey E, Shiu GK. Determination of ginsenosides in plant extracts from Panax ginseng and Panax quinquefolius L. by LC/MS/MS. Anal Chem 1999;71:R69–72.
[11] Yang W, Qiao X, Li K, Fan J, Bo T, Guo DA, Ye M. Differentiation and identification of Panax ginseng, Panax quinquefolium, and Panax notoginseng by monitoring multiple diagnostic chemical markers. Acta Pharm Sin B 2016;6:421–75.
[12] Shi W, Wang Y, Li J, Zhang H, Ding L. Investigation of ginsenosides in different parts and ages of Panax ginseng. Food Chem 2007;102:664–8.
[13] Oh JY, Kim YJ, Jung MC, Joo SC, Kwon WS, Kim SY, Jung SK. Yang DC. Investigation of ginsenosides in different tissues after elector treatment in Panax ginseng. J Ginseng Res 2014;38:270–7.
[14] Lee YS, Park HS, Lee DK, Jayakodi M, Kim NH, Koo HJ, Lee SC, Kim YJ, Kwon SW, Yang TJ. Integrated transcriptomic and metabolomic analysis of five Panax ginseng cultivars reveals the dynamics of ginsenoside biosynthesis. Front Plant Sci 2017;8:1048.
[15] Xiao D, Yue H, Xiu Y, Sun X, Wang Y, Liu S. Accumulation characteristics and correlation analysis of five ginsenosides with different cultivation ages from different regions. J Ginseng Res 2015;39:338–44.
[16] Kim YJ, Jeon JN, Jung MG, Oh JY, Kwon WS, Jung SK, Yang DC. Ginsenoside profiles and related gene expression during foliation in Panax ginseng Meyer. J Ginseng Res 2014;38:66–72.
[17] Jiang M, Liu J, Quan X, Quan L, Wu S. Different chilling stresses stimulates the accumulation of different types of ginsenosides in Panax ginseng cells. Acta Physiol Plant 2016;38:210.
[18] Leebees-Mack J, Raubenheimer LA, Cui L, Kuehl JV, Fourcade MH, Chumley TW, Boone JL, Jansen RK, Depamphilis CW. Identifying the basal angiosperm node in chloroplast genome phylogenies: sampling one's way out of the Felsenstein zone. Mol Biol Evol 2005;22:1948–63.
[19] Kwon SW, Nguyen VB, Park JY, Lee SC, Yang TJ. Evolution of the Ariaeaceae family inferred from complete chloroplast genomes and 45S nrDNAs of 10 Panax-related species. Sci Rep 2017;7:4917.
[20] Artuyukova E, Kozynenko M, Reunova G, Muzarok T, Zhuravlev YN. RAPD analysis of genome variability of planted ginseng, Panax ginseng. Mol Biol 2000;34:297–302.
[21] Ma KH, Dixit A, Kim YC, Lee DY, Kim TS, Cho EG, Park YJ. Development and characterization of new microsatellite markers for ginseng (Panax ginseng CA Meyer). Conserv Genet 2007;8:1507–9.
[22] Kim NH, Jayakodi M, Lee SC, Choi BS, Jang W, Lee J, Kim HH, Waminal NE, Lahavita M, Binh NW, Tal C. Genome and evolution of the ginseng family of medicinal herb Panax ginseng. Plant Biotechnol J 2018. https://doi.org/10.1111/pbi.12926.
Kim K, Lee SC, Lee J, Yu Y, Yang K, Choi BS, Koh HJ, Waminal NE, Choi HI, Kim NH, et al. Complete chloroplast and ribosomal sequences for 30 accessions elucidate evolution of Oryza AA genome species. Sci Rep 2015;5:15055.

Allen G, Flores-Vergara M, Krasynanski S, Kumar S, Thompson W. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. Nat Protoc 2006;1:2320–5.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 2013;30:2725–9.

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. Artemis: sequence visualization and annotation. Bioinformatics 2000;16:944–5.

Krzyszynski M, Schein J, Birol I, Connor J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. Genome Res 2009;19:1639–45.

Cui Y, Qin S, Jiang P. Chloroplast transformation of Platymonas (Tetraselmis) subcordiformis with the bar gene as selectable marker. PloS One 2014;9:e98607.

Dong W, Liu H, Xu C, Zuo Y, Chen Z, Zhou S. A chloroplast genomic strategy for designing taxon specific DNA mini-barcodes: a case study on ginsengs. BMC Genetics 2014;15:138.

Li R, Ma PF, Wen J, Yi TS. Complete sequencing of five Araliaceae chloroplast genomes and the phylogenetic implications. PloS One 2013;8:e78568.

Huang H, Shi C, Liu Y, Mao SY, Gao LZ. Thirteen Camellia chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships. BMC Evol Biol 2014;14:151.

Cho KS, Yun BK, Yoon YH, Hong SY, Mekapogu M, Kim KH, Yang TJ. Complete chloroplast genome sequence of Tartary Buckwheat (Fagopyrum tataricum) and comparative analysis with Common Buckwheat (F. esculentum). PloS One 2015;10.e0125332.

Court WE. Ginseng: the genus Panax. Taylor & Francis E-Library; 2006. p. 15.

Freeing M, Thomas BC. Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. Genome Res 2006;16:805–14.

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, Wall PK, Soltis PS. Polyplody and angiosperm diversification. Am J Bot 2009;96:336–48.

Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. The frequency of polyploid speciation in vascular plants. Proc Natl Acad Sci 2009;106:13875–9.

Kim NH, Choi HI, Kim KH, Jang W, Yang TJ. Evidence of genome duplication revealed by sequence analysis of multi-loci expressed sequence tag—simple sequence repeat bands in Panax ginseng Meyer. J Ginseng Res 2014;38:130–5.

Choi HI, Kim NH, Lee J, Choi BS, Do Kim K, Park JY, Lee SC, Yang TJ. Evolutionary relationship of Panax ginseng and P. quinquefolius inferred from sequencing and comparative analysis of expressed sequence tags. Genet Resour Crop Evol 2013;60:1377–87.

Shi PX, Li MX, Li YL, Jiang P, Zhang C, Pan YZ, Liu B, Xiao HX, Li LF. The impacts of polyploidy, geographic and ecological isolations on the diversification of Panax (Araliaceae). BMC Plant Biol 2015;15:297.

Kleine T, Maier UG, Leister D. DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. Annu Rev Plant Biol 2009;60:115–38.

Park S, Ruhlman TA, Sabir JS, Murwakil MH, Baeshen MN, Sabir MJ, Baeshen NA, Jansen RK. Complete sequences of organelle genomes from the medicinal plant Rhazya stricta (Apocynaceae) and contrasting patterns of mitochondrial genome evolution across asterids. BMC Genomics 2014;15:405.

Hakizimana-Covo E, Zeller RM, Martin W. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLoS Genet 2010;6:e1000834.

Smith DR, Crosby K, Lee RW. Correlation between nuclear plastid DNA abundance and plastid number supports the limited transfer window hypothesis. Genome Biol Evol 2011;3:365–71.

Gui S, Wu Z, Zhang H, Zheng Y, Zhu Z, Liang D, Ding Y. The mitochondrial genome map of Nelumbo nucifera reveals ancient evolutionary features. Sci Rep 2016;6:30158.

Zuo Y, Chen Z, Rondo K, Funamoto T, Wen J, Zhou S. DNA barcoding of Panax species. Planta Med 2011;77:182.

Nigan F, Shaw P, But P, Wang J. Molecular authentication of Panax species. Phytochem 1999;50:787–91.