Gene 530 (2013) 39–43

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

A fast SNP identification and analysis of intraspecific variation in the medicinal *Panax* species based on DNA barcoding

Xiaochen Chen [a], Baosheng Liao [a], Jingyuan Song [a], Xiaohui Pang [a], Jianping Han [a,*,1], Shilin Chen [a,b,**,1]

[a] Center for Computational Biology and Bioinformatics, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100193, PR China
[b] Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, PR China

ARTICLE INFO

Article history:
Accepted 29 July 2013
Available online 9 August 2013

Keywords:
Panax
SNPs
ITS2
Intraspecific variation
Intragenomic divergence

ABSTRACT

Medicinal plants of the *Panax* genus belonging to Araliaceae family are well-known, rare plants used as tonics in traditional Chinese medicine and have been described in the Chinese Pharmacopoeia. Because of the high price and the huge human demand, these commercial products often contain adulterants. In this study, 377 sequences from four species were analyzed. Single nucleotide polymorphisms (SNPs) were detected and patterns of intragenomic variation in internal transcribed spacer 2 (ITS2) from the four *Panax* species were studied. Intraspecific variations were analyzed based on three typical DNA barcodings (ITS2, *matK* and *psbA-trnH*). Results from this study revealed that intraspecific genetic distances in *Panax ginseng* and *Panax quinquefolius* were quite low (0–0.002) and the multi-copy ITS2 could be considered a single locus in the genomes of these two species. Five stable SNPs were detected in ITS2 region to identify the *Panax* medicinal species. Considering the mixed powder of *P. ginseng* and *P. quinquefolius*, double peaks could be clearly examined at SNP positions and the height of the peaks could indicate the mixed ratio roughly. Our findings indicate that SNP-based molecular barcodes could be developed as a routine method for the identification of the *Panax* genus with closely related species and the mixed powder *P. ginseng* and *P. quinquefolius*.

1. Introduction

Species of the *Panax* genus are relic plants. Medicinal *Panax* species (*Panax ginseng*, *Panax notoginseng*, *Panax quinquefolius* and *Panax japonicus*) are widely used as traditional medicines around the world. *P. ginseng*, also known as Oriental, Chinese or Korean ginseng, has a strong tonic efficacy. *P. quinquefolius*, commonly termed American ginseng, in spite of its similarity in appearance to Chinese ginseng, is recognized as sweet and slightly bitter in flavor but cold in nature (Chen et al., 2008). The morphological characteristics of those two medicines are very similar, especially in the form of shredded slice and powder, which makes the identification of crude plants, seeds, seedlings, flowers and fibers very difficult with traditional methods alone. *P. notoginseng*, commonly termed Sanchi, is famous for its hemostatic effects; it also protects against cerebral ischemia and has cardiovascular benefits (Ng, 2006).

Every year, large amounts of these four medicinal plants are used in various types of products to meet human demand. Because of the high price of ginseng, these commercial products often contain adulterants. Methods to identify *Panax* species have been developed from the traditionally morphological and histological characteristics to chemical analysis, and now methods at molecular level have been more and more put in use. e. g. DNA barcoding identification method will be indexed in the Chinese Pharmacopoeia (Updated edition 2010); Random amplified polymorphic DNA (RAPD) analysis (Shaw and But, 2007), Amplified fragment length polymorphism (AFLP) (ChangHo and YongEui, 2009; Ha et al., 2002), EST-SSR Markers (Yang, 2011) and gradient PCR method (Shim et al., 2005). Besides that PCR-based authentication, many other molecular identification methods, e.g. Restriction fragment length polymorphism (RFLP) and DNA fingerprinting, have also been reported (Fushimi et al., 1997; Ho and Leung, 2002; Leung, 1998). Compared with those methods above, DNA sequencing always brings us more information. Researchers compared ten candidate of DNA barcodes for the identification of *Panax* species and found that internal transcribed spacer (ITS) exhibited the highest sequence divergence (Zuo et al., 2011). Chen et al. first validated the internal transcribed spacer 2 (ITS2) region as a novel DNA barcode for identifying medicinal plant species in 2010 (Chen et al., 2010), and the China Plant BOL Group recommended the ITS as a core barcode for seed plants (China Plant BOL Group et al., 2011). Thus, ITS/ITS2 is a hotspot of recent DNA barcode

Abbreviations: SNP, Single nucleotide polymorphisms; ITS, internal transcribed spacer; ITS2, internal transcribed spacer 2; RAPD, Random amplified polymorphic DNA; AFLP, Amplified fragment length polymorphism; RFLP, Restriction fragment length polymorphism; K2P, Kimura-two-parameter; HMM, Hidden Markov Model.

* Correspondence to: J. Han, No. 151 Malianwa North Road, Haidian District, Beijing 100094, PR China. Tel.: +86 10 57833198; fax: +86 10 57833197.
** Correspondence to: S. Chen, Center for Computational Biology and Bioinformatics, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100193, PR China.
E-mail addresses: jphan@implad.ac.cn (J. Han), slchen@implad.ac.cn (S. Chen).
[1] Contributed equally.

0378-1119/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2013.07.097

research, and the analysis of intraspecific diversity is important for DNA barcode studies. However, the intraspecific variation of *Panax* species has not yet been revealed by DNA barcoding. Additionally, while ITS2 exhibits intragenomic variation in organisms ranging from plants to vertebrates (Sánchez and Dorado, 2008), this variation has been ignored in DNA barcoding analysis, which might lead to invalid diversity estimates. Therefore, the evaluation of intragenomic and intraspecific variation among ITS2 sequences is essential.

In the present study, we focused on four highly valued medicinal *Panax* species to develop a rapid SNP-based identification method for distinguishing medicinal species of the *Panax* genus from each other based on core DNA barcode sequences and reveal intraspecific variation based on DNA barcoding analysis using three regions (ITS2, *matK* and *psbA-trnH*) including 377 sequences. Additional details are discussed in subsequent sections.

2. Materials and methods

2.1. Plant materials and data source

In this study, 129 samples of *P. ginseng*, *P. quinquefolius* and *P. notoginseng* were collected, respectively, from geographical regions in Korea, Canada, China (Jilin, Liaoning and Heilongjiang provinces), and drugstores in Hong Kong and Shanghai to develop SNP and intraspecific variation based on an analysis with a total of 377 ITS2, *psbA-trnH* and *matK* sequences (Table S.1). For all newly collected materials, voucher specimens were deposited in the herbarium of the Institute of Medicinal Plant Development. Other published data were collected from GenBank (Table S.2).

2.2. DNA extraction, amplification and sequencing

DNA extraction, PCR amplification and sequencing were performed as described in previous studies (Chen et al., 2010; Kress and Erickson, 2007; Kress et al., 2005). For the mixed samples, powders of *P. ginseng* and *P. quinquefolius* were mixed in different ratios of 13:7, 10:10, 6:14 and 1:19 before extraction; the following steps were as above.

2.3. Data analysis

All the sequences obtained in our experiments and from GenBank were screened using BLAST to eliminate false sequences and then subjected to Hidden Markov Model (HMM) (Keller et al., 2009) analysis to remove the conserved 5.8S and 26S rRNA genes. Sequence alignment and SNP detection were performed using CodonCode Aligner V 3.7.1. Intraspecific genetic distances were computed and gaps were treated as complete deletion by Mega5.1. Intragenomic divergence data were obtained from our group (Song et al., 2012).

3. Results

3.1. Analysis of SNPs in ITS2

SNP analysis is a valuable tool for analyzing genetic diversity in plants, particularly in very closely related species (Arif et al., 2010). All the ITS2 sequences here (including our experiments and GenBank data) from the *Panax* species were examined for SNPs at the interspecies level (Fig. S.1). Five stable bases at position 28 bp, 32 bp, 43 bp, 140 bp and 207 bp were found, which could be used as a unique marker to distinguish between species (Table 1).

Table 1
Five stable SNPs in ITS2 as a unique marker of each species. The length of the alignment of the four species was approximately 230 bp.

		SNP location			
Species name	28 bp	32 bp	43 bp	140 bp	207 bp
P. ginseng	C	C	T	A	C
P. quinquefolius	C	T	C	A	C
P. notoginseng	T	C	C	T	T
P. japonicus	C	C	$C_{(A)}$	A	C

Because *P. ginseng* and *P. quinquefolius* are so morphologically similar, these two species are commonly confused with each other. Thus, SNPs to distinguish between these two species will be especially useful. The interspecific nucleotide diversity in the ITS2 of *P. ginseng* and *P. quinquefolius* is represented by two SNPs (32 bp and 43 bp). Based on the analysis of all the experimental and the published GenBank ITS2 sequences of *P. ginseng* and *P. quinquefolius*, the two SNPs in ITS2 were found to exist stably between them, and all the GenBank sequences of *P. ginseng* and *P. quinquefolius* proved the results. A combination of C and T at positions 32 and 43, respectively, indicates *P. ginseng*, while a combination of T and C indicates *P. quinquefolius*.

Additionally, the nucleotide at position 32 of *P. quinquefolius* is unique (T) and all the other *Panax* species have C. Thus *P. quinquefolius* could be discriminated from the related species of *Panax* genus by this SNP site.

Furthermore, each of the three SNPs at positions 28,140 and 207 could be used to distinguish *P. notoginseng* from all the other *Panax* species. At these three positions, *P. notoginseng* carries T and others carry A or C.

Owing to the fact that ITS2 sequences of *P. ginseng* and *P. quinquefolius* shows a very high degree of homology and only two SNPs exist, we mixed powder of *P. ginseng* and *P. quinquefolius* aiming to detect contamination and adulteration in the market. Mixing powder of four different *P. ginseng*–*P. quinquefolius* ratios were detected (13:7, 10:10, 6:14 and 1:19). The trace data from PCR products of the mixed powder indicates that in the impurity samples, as low as 5% *P. ginseng* contamination can be detected from *P. quinquefolius* (Fig. 1). Double peaks could be clearly examined at SNP positions and the height of the peaks could indicate the ratio roughly, e.g. at the ratio of 1:19, the main peaks G/A from reverse sequencing (equals to T/C in forward sequencing) attributed to *P. quinquefolius* were much higher than the lower peaks A/G (equals to C/T in forward sequencing) attributed to *P. ginseng*.

3.2. Intraspecific genetic distance and variation among the ITS2, matK and psbA-trnH regions

To uncover intraspecific polymorphisms in the four medicinal plants, besides ITS2, another two regions were amplified and sequenced, *psbA-trnH* and *matK*. The Kimura-two-parameter (K2P) model was used to calculate the intraspecific genetic distance, and variable bases were detected using Mega 5.1. The results below indicated that the intraspecific variation among *P. ginseng*, *P. quinquefolius* and *P. notoginseng* was quite low for the three selected regions (Table 2).

3.2.1. ITS2

The intraspecific divergence revealed by ITS2 sequences of *P. ginseng*, *P. quinquefolius* and *P. notoginseng* was surprisingly low, and no variation in ITS2 was observed in 54 samples of *P. ginseng* or 18 samples of *P. quinquefolius* from different locations, including Canada, Korea and three provinces in North China. The same result arose upon examination of *P. ginseng* and *P. quinquefolius* published ITS2 sequences obtained from GenBank. ITS2 sequences (some of which were derived from whole ITS sequences) submitted by nine different institutions were identical to those obtained via our own experiments by direct sequencing of PCR products. Significantly, 17 ITS2 sequences from ginseng submitted by a Korean institution (GenBank accession nos. AB043871, AB043872, HM446498-504, AY548192, AF274532-534, DQ284918 and DQ339097-099), including 5 different cultivars with considerable morphological variation, exhibited little molecular diversity, with no intraspecific variation of ITS2. A higher level of sequence