HortScience 54(8):1319–1323. 2019. https://doi.org/10.21273/HORTSCI14006-19

Molecular Distinguishment of *Trapa natans* L. Varieties in Taihu Lake Region of China and Development of a RAPD-SCAR Marker for Authentication of 'Heshangling'

Yanmei Zhang
Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing 210014, China

Xuelin Shen
Seed Administrative Station of Suzhou, Suzhou 215011, China

Xiaoqin Sun and Jia Liu
Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing 210014, China

Yifeng Xia and Xin Zou
Seed Administrative Station of Suzhou, Suzhou 215011, China

Yueyu Hang[1]
Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing 210014, China

Additional index words. *Trapa natans*, molecular distinguishment, RAPD, SCAR, Heshangling

Abstract*. Water chestnut (*Trapa natans* L.) is a group of annual, floating-leaved aquatic plants that serves as food and medical resources in many countries. However, the molecular method for distinguishing different *T. natans* L. resources is lacking. In this study, we detected genetic diversity of several chloroplast and nuclear genic or intergenic sequences in four varieties of *T. natans* and one wild type of *Trapa incisa* Siebold & Zuccarini to evaluate their potential as molecular markers. Our data revealed that the three chloroplast fragments (*rbcL*, *matK*, and *pbsA-trnH*) show no sequence difference among all tested samples. Only one nucleotide substitution is detected for the nuclear ribosomal internal transcribed spacer (ITS) in the *T. natans* variety Shuihongling. Four nucleotide substitutions are detected for the nuclear *carotenoid isomerase* (*CRTISO*) gene in the variety Hongxiuxie. In contrast, a total of 29 polymorphic sites are detected for a *Toll and interleukin-1 receptor-nucleotide binding site–leucine rich repeat* (*TNL*) gene in the five samples, among which six are nucleotide substitutions and the rest are insertions/deletions. The five samples could be fully distinguished from each other based on the *TNL* gene. To specifically authenticate 'Heshangling', 33 randomly amplified polymorphic DNA (RAPD) markers were adopted to amplify genomic sequences from the five samples. A pair of sequence characterized amplified region (SCAR) primers were designed based on the results of RAPD markers, which could specifically amplify one target band from all eight individuals of 'Heshangling', but none from any individuals of other *T. natans* varieties or one *T. incisa*. Taken together, a *TNL* sequence was provided in this study to distinguish four *T. natans* varieties and one *T. incisa*. Furthermore, a RAPD-SCAR marker was developed for efficient authentication of 'Heshangling'.

Water chestnut (*Trapa natans* L.) is an aquatic floating herb that belongs to the family Trapaceae. It is an important aquatic economic plant, mainly distributed in the tropical, subtropical, and temperate regions of Eurasia and Africa. In China, it is widely distributed, especially in the middle and lower reaches of the Yangtze River (Wang, 2012). The cultivation of water chestnut has a long history in China, which could be traced back to the Zhou Dynasty ≈3000 years ago. The Taihu Lake area is probably the origin center of the domestication and cultivation of water chestnut in China (Hui and Cao, 2015). In the Tang Dynasty (618–907 A.D.), water caltrop was an important food for worship as prayer offerings.

Traditionally, the classification of *Trapa* species has largely relied on the morphological diversity; however, the complex variation of morphological traits of the genus has raised a great controversy about the species division of the *Trapa* genus. For example, the *Trapa* genus has been classified into 70 species, 30 species, 20 species, and 13 species by different studies (Cook, 1990; Kak, 1988; Tutin et al., 1968; Vassiljev, 1965). The *Chinses Advanced Aquatic Plant Illustration* describes 30 species in the genus, and 11 are in China (Yan, 1983). *Flora Reipublicae Popularis Sinicae* records 15 species of the genus, whereas *Flora of China* proposes that *Trapa* consists of only two species, *T. natans* and *T. incisa* (Chen et al., 2007; Wan, 2000). In the past decades, quantitative taxonomy, cell taxonomy, pollen morphology, and other methods have been used for the systematics of the genus. Taking quantitative classification study as an example, this method divides the *Trapa* genus in the Hubei province of China into five species and eight varieties (Xiong et al., 1985). Another study, according to the morphology of pollens, flowers, and fruits, reassigns nine previously defined *Trapa* species in the Zhejiang province of China to three species, namely *T. incisa* var. Sieb, *Trapa bicornis* Osbeck, and *T. incisa* (Ding et al., 1999). By measuring the size of the fruit, Wang et al. (2006) divides *Trapa* species in China into three groups. The horticultural classification and standard used by Peng et al. (1998) divides *Trapa* into *Tapa quadrispinosa* Roxb, *Tapa bispinosa* Roxb, and *Tapa acornis* Nakano. Therefore, the authentication of a particular *Trapa* species or variety has been greatly hampered by the undistinguishable morphologies and the uncertainty of classification criteria of the genus.

In recent years, the rapid development of molecular biological technology has enabled researchers to efficiently distinguish specific species or varieties with characteristic sequence features. Many nuclear genes, chloroplast genes and mitochondrial genes are used for species-level identification. Different molecular markers have also been used to study the genetic relationship and species identification of the *Trapa* genus. For example, several RAPD molecular markers, a nuclear gene *APETALA2* (*AP2*), and the chloroplast gene *trnL-F* have proved to be capable of distinguishing *T. natans* and *T. incisa* (Jiang and Ding, 2004; Kim et al., 2010). Comparing with the interspecies differences, identification at lower taxonomic levels usually has relatively fewer genetic differences due to shorter differentiation time. The application of conserved interspecies molecular markers among lower taxonomic units usually has limitations in providing adequate genetic signals. Therefore, genes with rapid evolutionary rate are desirable for below species-level analysis. One type of well-known fast-evolving gene is plant *disease resistance gene* (*R* gene), the evolution rate of which could be several

Received for publication 28 Feb. 2018. Accepted for publication 6 May 2019.
This work was supported financially by the earmarked fund for Jiangsu Agricultural Industry Technology System to X. Shen, grant no. JATS (2018) 064.
[1]Corresponding author. E-mail: hangyueyu@cnbg.net.

times faster than that of many other genes (Zhang et al., 2011).

NBS-LRR genes compose the largest group of plant *R* genes, which occupied more than 60% of known functional *R* genes (Kourelis and van der Hoorn, 2018). Three *NBS-LRR* subclasses have been characterized, termed *RESISTANCE TO POWDERY MILDEW8 (RPW8)-NBS-LRR* (*RNL*), *coiled-coil-NBS-LRR* (*CNL*), and *Toll and interleukin-1 receptor-NBS-LRR* (*TNL*) (Shao et al., 2016, 2019). These genes show distinct evolutionary pattern due to their functional difference. In a previous study, 21 different haplotypes from 39 *Glycine max* varieties were identified using a single *TNL* gene *Rpp1* (Kim, et al., 2012). Another study of Li (2014) showed that the *CNL* gene *RPS2* can be used to identify different populations of wild mustard in China. These studies suggested that plant *R* genes have the potential to be candidates of molecular markers below species level.

Heshangling is a local variety of *T. natans* in Wuxian, Suzhou (a city located in Taihu Lake region) and has a long cultivation history. Its trivial name comes from its semicircle fruit, with one flat side and bulging side. The shell of 'Heshangling' is relatively thin. It is well known for being juicy, crispy, sweet, and tasty. Although its fruit morphology is characteristic, it is hard to distinguish this particular variety from other water chestnut varieties cultivated in the same region before maturity. It is also difficult to distinguish commercial products made from 'Heshangling' with those from other *T. natans* varieties. As a result, development of efficient and specific molecular markers is of a great scientific and commercial significance for the protection of varieties of traditional crops.

Materials and Methods

Plant materials. Four varieties of *T. natans* commonly cultivated in the middle and lower reaches of the Yangtze River were collected from Suzhou Vegetable Research Institute. One wild type of *T. incisa* Siebold & Zuccarini species was collected outside the Menan village of Meili town, Suzhou city (Table 1; Supplemental Fig. 1). Fresh leaves from eight individuals for each variety were collected, dried, and kept for DNA extraction.

Total DNA extraction. A total of 100 mg of collected leaves were crushed with tissue Lyser LT (QIAGEN, Hilden, Germany). Genomic DNA was extracted using the Easy Pure Plant Genomic DNA Kit (Beijing Full Gold Biotechnology Co., Ltd., Beijing, China) according to the manufacturer's instructions. Then, the concentration and purity of DNA was detected using the nucleic acid protein detector (Eppendorf, Hauppauge, NY) and saved in –20 °C following adjusted to 20 ng/μL.

Identification and analysis of NBS-LRR genes from T. bispinosa transcriptome. The assembled unigenes for the transcriptome of *T. bispinosa* (Li et al., 2017) were downloaded from the National Center for Biotechnology Information Web site (https://www.ncbi.nlm.nih.gov/nuccore/GEZE00000000.1/). The identification of *NBS-LRR* genes from *T. bispinosa* transcriptome was carried out according to our previous studies with slight modifications (Shao et al., 2014, 2016; Zhang et al., 2016). Briefly, all assembled transcripts of the transcriptome were translated into protein sequences in six frames. Then, both hidden Markov model (HMM) and BLAST searches were performed against the translated proteins using the HMM profile or amino acid sequence of NB-ARC domain (Pfam no. PF00931) as a query. All hits obtained using HMM or BLAST searches were then merged together to remove redundant hits. The remaining sequences were further subjected to the online Pfam analysis (http://pfam.sanger.ac.uk/) to verify whether they indeed possessed the NBS domain, with the E-value setting to 10^{-4}.

Phylogenetic analysis was performed for *NBS-LRR* genes identified from *T. bispinosa* transcriptome with those from *Arabidopsis thaliana* using the ML method as described in our previous studies (Shao et al., 2014, 2016; Zhang et al., 2016). The *NBS-LRR* genes from *T. bispinosa* transcriptome were classified into *TNL*, *CNL*, and *RNL* according their phylogenetic relationship to those form *A. thaliana*.

Target sequence amplification and sequence analysis. Primers for the nuclear *ITS* sequence and three chloroplast genes (*rbcL*, *matK*, and *psbA-trnH*) were retrieved from a previous study (China Plant BOL Group et al., 2011). Specific primers against four nuclear fragments [*carotenoid isomerase* (*CRTISO*), GEZE01020818 (*RNL*), GEZE01070617 (*CNL*), and GEZE01034532 (*TNL*)] were designed using the Oligo 7.0 software basing the transcriptome of *T. bispinosa*. The primer is synthesized by Nanjing Genscript Biotech Co., Ltd. (Nanjing, China). Polymerase chain reaction (PCR) amplification of the target genes were carried out using the following program: pre-degeneration for 5 min at 94 °C; degeneration for 45 s at 94 °C, annealing for 45 s at 52 °C, extension for 1 min at 72 °C, 30 cycles and extension 10 min at 72 °C. The reaction system is 20 μL and its components and final concentrations are as follows: DNA template (20 ng/μL) 1.0 μL, 2 × Reaction Mix (including 20 mM Tris-HCl, 100 mM KCl, 33 mM MgCl2, 4400 μM dNTPs, bromine blue) 10.0 μL, upstream and downstream primers (10 mM) 1.0 μL each, taq DNA polymerase (2.5 U/μL) 0.4 μL, and finally double distilled water up to 20 μL. The PCR reaction is carried out in Biometra (Göttingen, Germany) T1 PCR instrument. PCR products were examined electrophoretically using 1.0% agarose gels under 80 V for 0.5 h and visualized via gel imaging system. Bidirectional sequencing of purified PCR products was completed by Beijing Genomics Institute (Beijing, China) using the amplification primers. The sequencing results were processed via Sequencher 4.5 to remove low-mass sequences and subjected to MEGA6 for further analysis.

Screening of RAPD primers. DNA samples pooled by eight individuals for the four varieties of *T. natans* and *T. incisa* were used as templates for PCR. RAPD primers were retrieved from a previous study (Jiang and Ding, 2004; Kim et al., 2010) and synthesized by the Nanjing Genscript Biotech Co., Ltd. PCR analysis was performed as described

Table 1. Plant materials used in this study.

Plant name	Species name	Variety type	Main morphological description
Laowuling	*Trapa natans*	Local cultivar, Suzhou, Jiangsu	Leaf blade deltoid-rhombic, green; 2-horned, horns recurved; tender peel turquoise and dark after maturity
Hongxiuxie	*T. natans*	Local cultivar, Chaohu, Anhui	Leaf blade deltoid-rhombic; 2-horned, horns horizontal; peel amaranth, thin; fruit tender, soft quality
Shuihongling	*T. natans*	Local cultivar, Suzhou, Jiangsu	Leaf blade deltoid-rhombic, turquoise, petiole amaranth; fruit bright red, 4-horned, shell thin and fruit crispy
Heshangling	*T. natans*	Local cultivar, Suzhou, Jiangsu	Leaf blade deltoid-rhombic, light green; fruit semicircular, 0-horned, tender peel light green and aged peel yellowish white, shell thin, fruit juicy, crispy
Yeling	*Trapa incisa*	Wild variety, Changshu, Jiangsu	Leaf blade rhombic-triangular; fruit narrowly rhombic, 4-horned, horns conic; shell thin, fruit glutinous

previously. The PCR products were examined by 2.0% agarose gel electrophoresis under 90 V for 1 h and visualized via gel imaging system.

Design and validation of a RAPD-SCAR marker. A band specific to 'Heshangling' was excised and extracted from the gel. The obtained DNA fragment was cloned into pMD19-T vector and transformed into *Escherichia coli* DH5α competent cells (Takara Biomedical Technology (Beijing) Co., Ltd., Beijing, China). Plasmids containing target sequence were extracted from the transformed bacterial cells and sequenced by Nanjing Genscript Biotech Co. A pair of specific SCAR primers targeting the sequence of the cloned band was designed. The SCAR primers were used to validate specificity for the 40 individuals of the four varieties of *T. natans* and *T. incisa* by PCR using a procedure the same as that described previously.

Results

Amplification and sequence analysis of conserved nuclear and chloroplast genes. Three chloroplast markers (*rbcl*, *matK*, and *pbsA-trnH*) and two nuclear marker (*ITS* and *CRTISO*) genes were successfully amplified from the four varieties of *T. natans* and *T. incisa*. The obtained PCR products were subjected to direct sequence. The results showed that the obtained sequences are 695, 867, 354, 788–789, and 2186 in length (Table 2), respectively. Analysis of these sequences revealed that the three chloroplast genes showed no sequence difference among the four varieties of *T. natans* and one of *T. incisa*. In contrast, the two nuclear genes show different extents of sequence diversity, with only one nucleotide substitution observed for *ITS* sequence isolating 'Shuihongling' from the other four samples, and four polymorphic sites detected for *CRTISO* to 'Hongxiuxie' separate it from the other four samples (Table 3).

Amplification and sequence analysis of NBS-LRR genes. The transcriptome of *T. bispinosa* was subjected to analysis for identification of *NBS-LRR* genes based on the previously published protocol (Shao et al., 2014, 2016; Zhang et al., 2016). A total of 71 of 73,284 assembled unigenes were found to encode *NBS-LRR* genes (Supplemental Table 1). After removing unigenes encoding a partial NBS domain (shorter than two-thirds of the full length), the remaining 53 sequences were subjected to phylogenetic analysis with the *NBS-LRR* genes from *A. thaliana*. The phylogeny (Supplemental Fig. 2) showed that the 53 *NBS-LRR* transcripts from *T. natans* could be assigned to three *NBS-LRR* subclasses (*RNL*, *CNL*, and *TNL*) basing their relationship to *NBS-LRR* genes from *A. thaliana*. There are 3, 22, and 28 genes in *RNL*, *CNL*, and *TNL* subclasses, respectively.

One unigene from each class (GEZE01020818 from *RNL*, GEZE01070617 from *CNL*, and GEZE01034532 from *TNL*) were selected to design specific primers. The primers were designed to target the LRR domain of each gene to obtain maximal sequence diversity, although partial NBS domain was also covered by the primers for the *TNL* gene (Supplemental Table 2). All of the three primer pairs successfully amplified orthologous genes from the four *T. natans* varieties and *T. incisa*. Sequence analysis revealed that there was no difference for the selected *RNL* and *CNL* genes among the five samples. In contrast, the *TNL* sequences from the five samples have 29 detected polymorphism sites in the coding region, of which six are nucleotide substitutions, 23 are insertions/deletions. 'Heshangling' has a four-base deletion from the position 1071 to1074, whereas *T. incisa* has a 19 base pair (bp) insertion at the position of ≈926 to 944. These *TNL* gene fragments amplified from the five samples are all different from each other; therefore, may serve as a molecular marker to distinguish different *T. natans* varieties and *T. incisa*.

RAPD molecular marker screening for Heshangling varieties. PCR amplification was carried out on four varieties of *T. natans* and *T. incisa* with 33 RAPD primers. The

Table 2. Primers used in this study.

Primers	Primer sequence	Amplified sequence length
rbcL-F	ATGTCACCACAAACAGAAAC	695
rbcL-F	TCGCATGTACCTGCAGTAGC	
matK-F	CGTACAGTACTTTTGTGTTTACGA	867
matK-R	ACCCAGTCCATCTGGAAATCTTGGTTC	
psbA-trnH-F	GTTATGCATGAACGTAATGCTC	354
psbA-trnH-R	CGCGCATGGTGGATTCACAATCC	
ITS-F	AACAAGGTTTCCGTAGGTGA	788–790
ITS-R	TGAGGACGCTTCTCCAGAC	
CRTISO-F	CTTACTAGCTTCAAATGCACCA	2186
CRTISO-R	GCTAAGTATCTCCTGTGCGTCT	
GEZE01034532-F	TAACAACTAGTAGCCGTAGAGT	1523–1546
GEZE01034532-R	CTTACTAGCTTCAAATGCACCA	
GEZE01070617-F	CTGCCTTACTACCTCAAGTCG	816
GEZE01070617-R	GTCAACTTCCCAATACCCCTC	
GEZE01020818-F	TCTGATCGACCATTCAATGCC	1373
GEZE01020818-R	CCAGAGATCATTGCACCGAGA	

Table 3. Variable sites of internal transcribed spacer (*ITS*), *carotenoid isomerase* (*CRTISO*), and *Toll and interleukin-1 receptor-nucleotide binding site–leucine rich repeat* (*TNL*) genes in four varieties of *Trapa natans* and one *Traps incisa*.

	ITS	*CRTISO*				*TNL*											
Plant name	706	618	1756	1846	2019	441	516	699	851	916	918	926	927	928	929	930	931
Laowuling	C	T	A	G	A	G	G	G	C	G	T	–	–	–	–	–	–
Hongxiuxie	C	C	G	A	G	A	G	G	G	G	T	–	–	–	–	–	–
Shuihongling	A	T	A	G	A	G	T	G	C	G	G	–	–	–	–	–	–
Heshangling	C	T	A	G	A	A	G	G	G	G	T	–	–	–	–	–	–
Yeling	C	T	A	G	A	G	G	C	G	G	T	G	A	A	G	C	A

	TNL																
Plant name	932	933	934	935	936	937	938	939	940	941	942	943	944	1071	1072	1073	1074
Laowuling	–	–	–	–	–	–	–	–	–	–	–	–	–	G	T	G	T
Hongxiuxie	–	–	–	–	–	–	–	–	–	–	–	–	–	G	T	G	T
Shuihongling	–	–	–	–	–	–	–	–	–	–	–	–	–	G	T	G	T
Heshangling	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Yeling	A	C	A	G	G	T	G	T	A	T	T	C	C	G	T	G	T

"–" indicates base deletions.

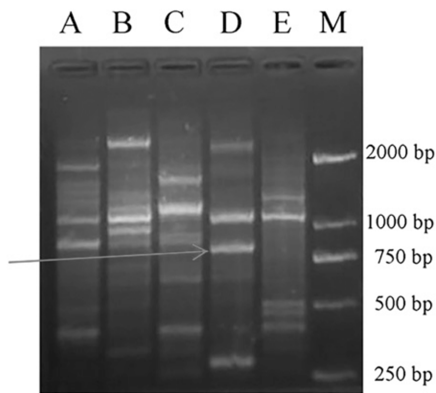


Fig. 1. Amplification profiles of four varieties of Trapa natans and one Trapa incisa using RAPD26 marker. Lanes A to E in order are ‘Laowuling’, ‘Hongxiuxie’, ‘Shuihongling’, ‘Heshangling’, and ‘Yeling’. M, DL2000 DNA marker; the arrow represents a specific band of ‘Heshangling’. bp = base pairs.

ATTGCTTCGATACGCGTCTTGCAAACCTTATAGGCCACCATCGCACGATCAA
CCATCGGAATTAGCGCGTCGATATCTGGCTCTTGTTGCGTCCTCAAGGTTTC
TGCTATTTTTTTCAGCGTTGCATAGTTTTCTTTAAAAGAATCAGACTCTGCG
TTCATATTATTCTCCAACTTTTAAAAATTAGTAATTCACTGCCGGTTTAACTT
TGATATCGCCATCCCTAAATTCGATATTCAGAAGTTCAGCTTGCTTCGCTTGT
TCTGCCGTGACGACCGGTTTACCACTGAGATCGCGGACAAGCGCAAAGCC
TCGTCGTAAGGTCGCCTTGGGGCCTATGCCTAGTATCTCCCTGGCCAAGGA
CTCGATTTGAATCGCGGCATCGTCAATTTTCTGTTGTCCCAACTCAGTTACT
CTCCGCAAGTAAAACTCAATTTCCCGTTCGGCGTCTAATATTTGGATGTCGG
CAAGGAAGGTCACTTTTTCCCATTCTTTATCGGCCAATTTTTCTGCGTCCGC
CAAAACCGAAAATGATTTTGTCTCGATCCAGCGCATCCCGTGTTCGAGCCT
GTTTTCGGCGGTGGACAGGAGCATCTGAGCAGCCGGTTGTAGGCCAGCCG
CAAGTTGTTCGATTTCATTATCCACCAGGTTTAATTGCTCAGAAACCTGCGT
CTGGATTTTTCCGATAAAGATCCCACACGCCTTTTCAGCACTCGATAACTGC
TTTTCGGTATGTAACAAGAATTCCTGAAAATGCGCCTCGGCCGTCGTCGCA
TTGCTGCAAATGACATCCCTGATGTGCCCGATGACTTTTGATGGCGTATCGA
AGCAATC

Fig. 2. The nucleotide sequence of ‘Heshangling’-specific fragment amplified using primer RAPD26. The underlined sequences are designed sequence characterized amplified region primers.

template of each DNA sample is composed of eight individuals. The products amplified from ‘Heshangling’ by the primer RAPD26 (5′-GAGAGAGAGAGAGAGAYT-3′) showed a unique band with a size ≈800 bp in addition to other bands amplified from all varieties (Fig. 1). This ‘Heshangling’-specific fragment was then cloned and sequenced. The unique band finally turned out to be a fragment of 834 bp in length; the nucleotide sequence is shown in Fig. 2.

Design and validation of a pair of RAPD-SCAR primers for specific identification of ‘Heshangling’. A pair of SCAR primers is designed to target the ‘Heshangling’-specific fragment (as shown in Fig. 2): SCAR-F (5′-CTCTAACCGATTCTCACCTCC-3′) and SCAR-R (5′- TATAGGCTTTTTATCAA TCGTA-3′), corresponding to the 419–444 and the 713–736 positions. PCR amplification was performed on 40 individuals from four varieties of *T. natans* and *T. incisa* by the SCAR primers. The result showed that a specific band was amplified from all eight individuals of ‘Heshangling’, which is in accordance with its expected size of 318 bp. In contrast, no bands were amplified from any individuals from the remaining three varieties of *T. natans* or *T. incisa* (Fig. 3).

Discussion

The morphological variation of *Trapa* genus is extremely complicated in the number of angles, color of peel, and the size of fruit. Investigators around the world have a great controversy regarding delimitation of *Trapa* species. Some proposed that there are dozens of *Trapa* species, and others combine them into one (Cook, 1990; Kak, 1988; Tutin et al., 1968; Vassiljev, 1965; Yan, 1983). In our country, *Flora of China* combined the 15 *Trapa* species/varieties recorded in *Flora Reipublicae Popularis Sinicae* into two different species, *T. natans* and *T. incisa*. It seems to indicate that researchers are uncertain about the classification inside the genus *Trapa*, and the previous taxonomic algorithm based on morphology does not work well. China has a long history in *Trapa* cultivation, and *Trapa* is widely distributed in the middle and lower reaches of the Yangtze River. This study explored molecular markers that potentially distinguish different *T. natans* resources.

With the rapid development of sequencing technology, the molecular method has been frequently used for phylogenetic analysis and barcoding of plants. The chloroplast genes and the nuclear *ITS* sequence have been widely used as barcoding markers at different taxonomic levels (CBOL Plant Working Group, 2009). Previous studies have shown that the chloroplast DNA sequences are more conservative than the nuclear *ITS* region. They usually could distinguish taxa at or above genus level, but may not suit for the classification among closely related species within the same genus (Li et al., 2018). By comparison, the *ITS* fits better at species-level resolution than chloroplast genes. In the present study, none of the three chloroplast markers is capable of distinguishing varieties of *T. natans*. *ITS* is not much better, showing only one single nucleotide polymorphism. As for the other chosen nuclear marker *CRTISO*, only one variety could be identified based on the four nucleotide substitutions.

To find more effective molecular markers, we examined one of the most rapid-evolving plant gene families, *NBS-LRR* genes. The *NBS-LRR* gene family is composed of hundreds of members with highly diverged sequences due to their active evolutionary patterns (Shao et al., 2016). According to the results of previous studies (Kim et al., 2012; Li, 2014), some *NBS-LRR* genes could be highly variable even among individuals within the same species. Inspired by these findings, this study selected three *NBS-LRR* genes to evaluate their sequence diversity among the four varieties of *T. natans* and *T. incisa*. Two criteria were used to select candidate *NBS-LRR* genes for amplification in this study. First, the assembled transcript of a selected gene has a relatively long coding region and covers the LRR domain. The LRR domain is the most variable region of *NBS-LRR* genes. This criterion would help us to amplify a relatively long sequence with high polymorphism. Second, the selected gene has no detected paralogs with high sequence similarity. This criterion would avoid nonspecific amplification.

Although no sequence differences were observed for the two of them, the selected *TNL* gene showed high sequence diversity among the five samples. Polymorphic sites of both nucleotide substitutions and insertions/deletions were observed. Nearly all polymorphic sites were detected in the LRR-encoding region, whereas none of them was

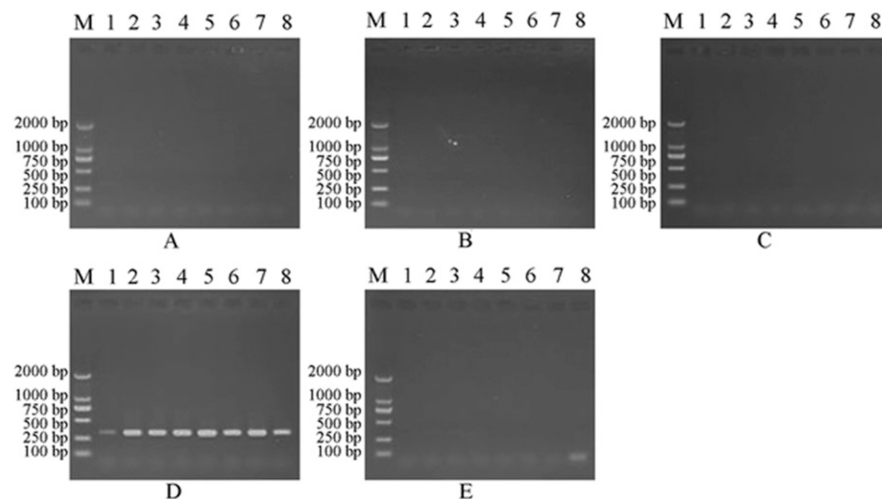


Fig. 3. Verification of the ‘Heshangling’-specific randomly amplified polymorphic DNA–sequence characterized amplified region by polymerase chain reaction (PCR) in eight individuals from each of four varieties of *Trapa natans* and one *Trapa incisa*. A to E are PCR results from eight individuals of ‘Laowuling’, ‘Hongxiuxie’, ‘Shuihongling’, ‘Heshangling’, and ‘Yeling’, respectively. M = DL2000 DNA marker. Lanes 1 to 8 are PCR products from eight individuals. bp = base pairs.

detected in the amplified partial NBS domain region. The higher diversity of the *TNL* gene than that of the *CNL* and the *RNL* genes may due to the different evolutionary pattern among the three *NBS-LRR* subclasses (Shao et al., 2016). *RNL* is the most conserved *NBS-LRR* subclass and evolves slower than the other two subclasses due to its specific function in signal transduction. *CNL* and *TNL* genes are usually involved in pathogen detection and therefore show fast evolutionary rate to cope with the rapidly evolving pathogens (Shao et al., 2016). Furthermore, several studies have reported that *TNL* genes evolve faster than *CNL* genes; we speculate that this may be partially because of its ancient intron-rich structure compared with the intron-less structure of *CNL* genes (Shao et al., 2016; Zhang et al., 2011), and introns are much less subject to evolutionary constraints so that allowed for higher sequence diversity. Overall, the results from our study suggested that the *TNL* gene amplified in this study may serve as a candidate molecular marker for *T. natans* and *T. incisa*.

Although the *TNL* gene is capable of distinguishing the tested four varieties of *T. natans* and *T. incisa* effectively, this method requires a professional background for molecular experiments and data analysis, as well as understanding of the theory of evolution, which might not be so appropriate in practical application. RAPD markers are DNA fragments obtained by PCR amplification of random segments from the genomic DNA. The SCAR marker is derived from RAPD that aims to specifically and efficiently amplify target bands from specified species or varieties. In the present study, a RAPD-SCAR marker for specific authentication of 'Heshangling' has been developed. Using this method, 'Heshangling' can be quickly and intuitively identified with an efficiency of 100%.

Taken together, we developed two different approaches to distinguish four varieties of *T. natans* and *T. incisa*. The sequence character of the *TNL* gene could distinguish all four varieties of *T. natans* and *T. incisa* simultaneously, whereas the RAPD-SCAR marker could specifically authenticate 'Heshangling'. Our study may serve as a foundation for further application of molecular markers to identify and protect the high quality of local *Trapa* resources, as well as commercialized products of *Trapa*.

Literature Cited

CBOL Plant Working Group. 2009. A DNA barcode for land plants. Proc. Natl. Acad. Sci. USA 106(31):12794–12797.

Chen, C., B.Y. Ding, and A.M. Funston. 2007. Trapaceae, 290–291. In: Z.Y. Wu and P.H. Raven (eds.). Fl. Science Press, Bejing; Missouri Botanical Garden, St. Louis.

China Plant BOL Group, D.Z. Li, L. M. Gao, H.T. Li, H. Wang, X.J. G, J.Q. Liu, Z.D. Chen, S.L. Zhou, S.L. Chen, Y.B. Yang, C.X. Fu, C.X. Zeng, H.F. Yan, Y.J. Zhu, Y.S. Sun, S.Y. Chen, L. Zhao, K. Wang, T. Yang, and G.W. Duan. 2011. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. Proc. Natl. Acad. Sci. USA 108(49):19641–19646.

Cook, C.D.K. 1990. Aquatic plant book. SPB Academic publishing, The Hague, The Netherlands.

Ding, B.Y., T. Huang, W.M. Jiang, L.A. Zhao, and C.Z. Zheng. 1999. The seedling morphology of *Trapa* and its systematic significance. J. Zhejiang Univ. (Sci. Ed.). 26(3):92–98.

Hui, F.P. and Y. Cao. 2015. Water caltrop cultivation around Taihu lake in the Ming-Qing dynasties. Agr. Hist. China. (5):24–33.

Jiang, W.M. and B.Y. Ding. 2004. Genetic relationship among *Trapa* species assessed by RAPD markers. J. Zhejiang Univ. (Agr. & Life Sci.). 30(2):191–196.

Kak, A.M. 1988. Aquatic and wetland vegetation of western Himalayas. J. Econ. Taxon. Bot. 12(2):447–451.

Kim, C., H.R. Na, and H.K. Choi. 2010. Molecular genotyping of *Trapa bispinosa* and *T. japonica* (Trapaceae) based on nuclear AP2 and chloroplast DNA *trnL-F* region. Amer. J. Bot. 97(12):E149–E152.

Kim, K.S., J.R. Unfried, D.L. Hyten, R.D. Frederick, G.L. Hartman, R.L. Nelson, Q. Song, and B.W. Diers. 2012. Molecular mapping of soybean rust resistance in soybean accession PI 561356 and SNP haplotype analysis of the *Rpp1* region in diverse germplasm. Theor. Appl. Genet. 125(6):1339–1352.

Kourelis, J. and R.A.L. van der Hoorn. 2018. Defended to the nines: 25 years of resistance gene cloning identifies nine mechanisms for R protein function. Plant Cell 30(2):285–299.

Li, J., Y. Yang, S. Yang, Z. Zhang, S. Chen, C. Zhong, R. Zhou, and S. Shi. 2017. Comparative transcriptome analyses of a mangrove tree *Sonneratia caseolaris* and its non-mangrove relatives, *Trapa bispinosa* and *Duabanga grandiflora*. Mar. Genom. 31:13–15.

Li, S., X. Qian, Z. Zheng, M. Shi, X. Chang, X. Li, J. Liu, T. Tu, and D. Zhang. 2018. DNA barcoding the flowering plants from the tropical coral islands of Xisha (China). Ecol. Evol. 8(21):10587–10593.

Li, Y. 2014. Genetic diversity and genetic structure of wild *Brassica juncea* in China. Nanjing Agr. Univ., Nanjing, China, M.S. Thesis.

Peng, J., Q.D. Kong, W.D. Ke, and Y.M. Liu. 1998. Horticultural classification of the germplasm resources of water caltrap (*Trapa* spp.). Adv. Hort. 2:645–649.

Shao, Z.Q., J.Y. Xue, P. Wu, Y.M. Zhang, Y. Wu, Y.Y. Hang, B. Wang, and J.Q. Chen. 2016. Large-scale analyses of angiosperm nucleotide-binding site-leucine-rich repeat genes reveal three anciently diverged classes with distinct evolutionary patterns. Plant Physiol. 170(4):2095–2109.

Shao, Z.Q., J.Y. Xue, Q. Wang, B. Wang, and J.Q. Chen. 2019. Revisiting the origin of plant NBS-LRR genes. Trends Plant Sci. 24(1):9–12.

Shao, Z.Q., Y.M. Zhang, Y.Y. Hang, J.Y. Xue, G.C. Zhou, P. Wu, X.Y. Wu, X.Z. Wu, Q. Wang, B. Wang, and J.Q. Wang. 2014. Long-term evolution of nucleotide-binding site-leucine-rich repeat genes: Understanding gained from and beyond the legume family. Plant Physiol. 166(1):217–234.

Tutin, T.G., V.H. Heywood, N.A. Burges, D.M. Moore, D.H. Valentine, S.M. Walters, D.A. Webb, P.W. Ball, A.O. Chater, and I.K. Ferguson. 1968. Fl. Europaea, Vol. 2: Rosaceae to Umbelliferae. Cambridge University Press, Cambridge.

Vassiljev, V. 1965. Species novae Africanicae generis *Trapa* L. Nov. Sist. Vyss. Rast. (Leningrad) 1965:175–179.

Wan, W. 2000. Trapaceae, p. 1–26. In: C. Chen (ed.). Fl. Reipubl. Popularis Sin. Science Press, Beijing.

Wang, Q. 2012. Fruits of Hemitrapa (Trapaceae) from the Miocene of eastern China, their correlation with *Sporotrapoidites erdtmanii* pollen and paleobiogeographic implications. J. Paleontol. 86(1):156–166.

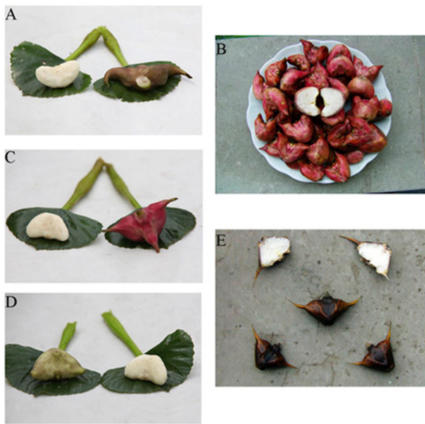
Wang, Y.F., B.Y. Ding, R.Y. Hui, and M.L. Jin. 2006. Analysis of morphological plasticity of *Trapa* from China and its taxonomic signification. J. Zhejiang Univ. (Sci. Ed.). 33(5):567–572.

Xiong, Z.T., H.Q. Wang, and X.Z. Sun. 1985. Numerical taxonomic studies in Trapaceae in Hubei (I). J. Plant Sci. 3(1):45–53.

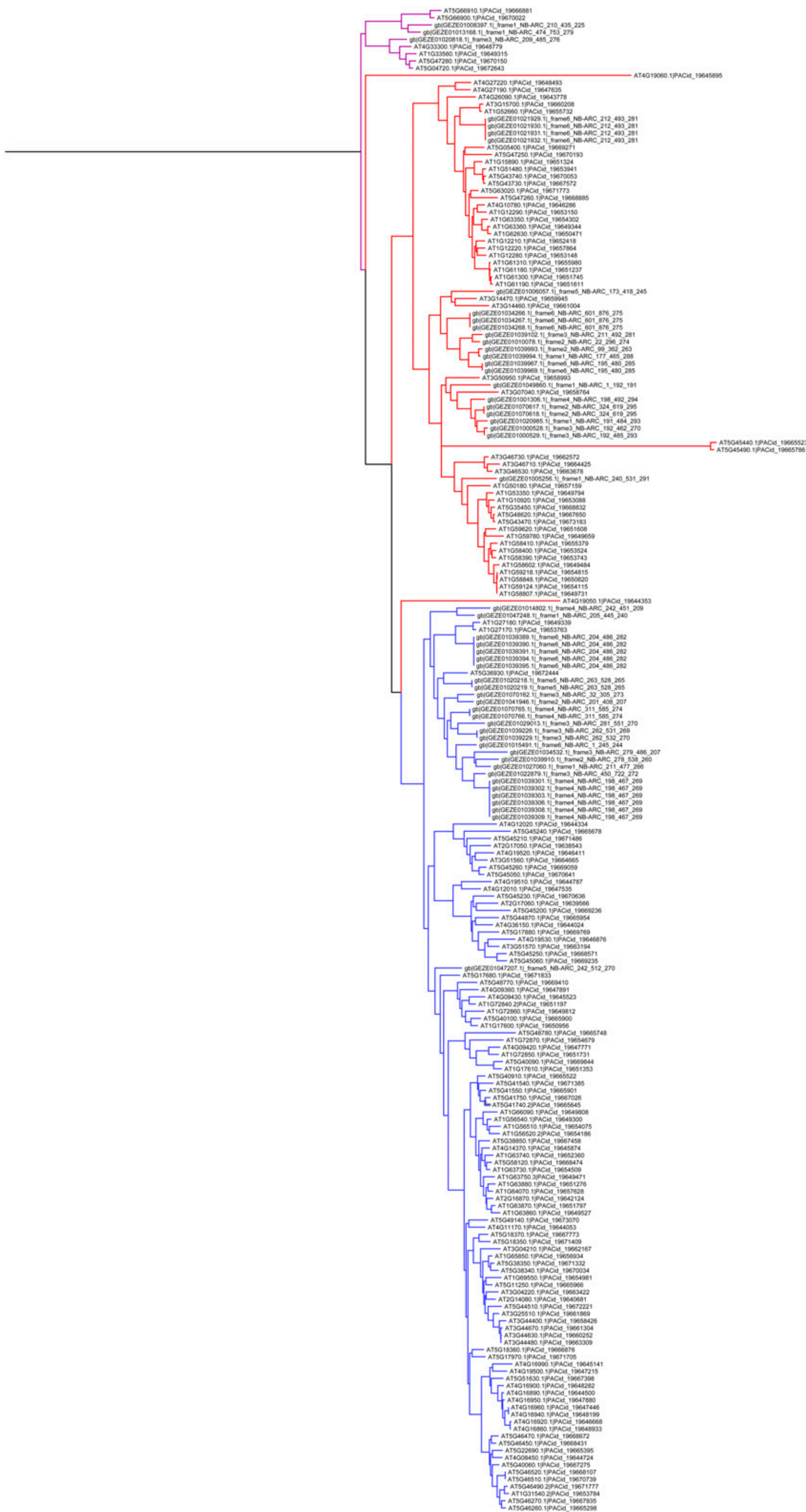
Yan, S.Z. 1983. Chinses advanced aquatic plant illustration. Science Press, Beijing, China.

Zhang, X.H., Y. Feng, H. Cheng, D. Tina, S. Yang, and J.Q. Chen. 2011. Relative evolutionary rates of NBS-encoding genes revealed by soybean segmental duplication. Mol. Genet. Genomics 285(1):79–90.

Zhang, Y.M., Z.Q. Shao, Q. Wang, Y.Y. Hang, J.Y. Xue, B. Wang, and J.Q. Wang. 2016. Uncovering the dynamic evolution of nucleotide-binding site-leucine-rich repeat (NBS-LRR) genes in Brassicaceae. J. Integr. Plant Biol. 58(2):165–177.



Supplemental Fig. 1. The morphological features of four varieties of *Trapa natans* and one wild type of *Trapa incisa* in this study. A to E in order are 'Laowuling', 'Hongxiuxie', 'Shuihongling', 'Heshangling', and 'Yeling'.



Supplemental Fig. 2. Phylogeny of *Nucleotide binding site–leucine rich repeat* (*NBS-LRR*) genes from *Trapa bispinosa* and *Arabidopsis thaliana*.

Supplemental Table 1. *Nucleotide binding site–leucine rich repeat* (*NBS-LRR*) genes identified from *Trapa bispinosa* transcriptome.

Number	Gene ID	*NBS-LRR* subclasses
1	GEZE01013168.1	*RNL*
2	GEZE01020818.1	*RNL*
3	GEZE01008397.1	*RNL*
4	GEZE01000528.1	*CNL*
5	GEZE01000529.1	*CNL*
6	GEZE01070617.1	*CNL*
7	GEZE01070618.1	*CNL*
8	GEZE01001306.1	*CNL*
9	GEZE01020985.1	*CNL*
10	GEZE01005256.1	*CNL*
11	GEZE01039994.1	*CNL*
12	GEZE01039967.1	*CNL*
13	GEZE01039969.1	*CNL*
14	GEZE01021929.1	*CNL*
15	GEZE01021930.1	*CNL*
16	GEZE01021931.1	*CNL*
17	GEZE01021932.1	*CNL*
18	GEZE01039102.1	*CNL*
19	GEZE01034266.1	*CNL*
20	GEZE01034267.1	*CNL*
21	GEZE01034268.1	*CNL*
22	GEZE01010078.1	*CNL*
23	GEZE01039993.1	*CNL*
24	GEZE01006057.1	*CNL*
25	GEZE01049860.1	*CNL*
26	GEZE01039389.1	*TNL*
27	GEZE01039390.1	*TNL*
28	GEZE01039391.1	*TNL*
29	GEZE01039394.1	*TNL*
30	GEZE01039395.1	*TNL*
31	GEZE01070765.1	*TNL*
32	GEZE01070766.1	*TNL*
33	GEZE01070162.1	*TNL*
34	GEZE01022879.1	*TNL*
35	GEZE01029013.1	*TNL*
36	GEZE01039229.1	*TNL*
37	GEZE01047207.1	*TNL*
38	GEZE01039226.1	*TNL*
39	GEZE01039301.1	*TNL*
40	GEZE01039302.1	*TNL*
41	GEZE01039303.1	*TNL*
42	GEZE01039306.1	*TNL*
43	GEZE01039308.1	*TNL*
44	GEZE01039309.1	*TNL*
45	GEZE01027060.1	*TNL*
46	GEZE01020218.1	*TNL*
47	GEZE01020219.1	*TNL*
48	GEZE01039910.1	*TNL*
49	GEZE01015491.1	*TNL*
50	GEZE01047248.1	*TNL*
51	GEZE01014802.1	*TNL*
52	GEZE01034532.1	*TNL*
53	GEZE01041946.1	*TNL*
54	GEZE01020819.1	*RNL*
55	GEZE01005764.1	*RNL*
56	GEZE01005765.1	*RNL*
57	GEZE01004861.1	*CNL*
58	GEZE01039103.1	*CNL*
59	GEZE01012995.1	*CNL*
60	GEZE01006173.1	*CNL*
61	GEZE01056022.1	*CNL*
62	GEZE01006569.1	*CNL*
63	GEZE01039970.1	*CNL*
64	GEZE01064925.1	*CNL*
65	GEZE01002219.1	*CNL*
66	GEZE01030308.1	*CNL*
67	GEZE01039968.1	*CNL*
68	GEZE01039227.1	*TNL*
69	GEZE01071760.1	*TNL*
70	GEZE01039909.1	*TNL*
71	GEZE01007953.1	Undetermined

Note: Genes indicated in gray are not included in phylogenetic analysis due to short nucleotide binding site domain. *RNL = RESISTANCE TO POWDERY MILDEW8 (RPW8)-NBS-LRR; CNL = coiled-coil-NBS-LRR; TNL = Toll and interleukin-1 receptor-nucleotide binding site–leucine rich repeat.*

Supplemental Table 2. Amplification regions of the *Toll and interleukin-1 receptor-nucleotide binding site–leucine rich repeat* (*TNL*), *coiled-coil-NBS-LRR* (*CNL*), and *RESISTANCE TO POWDERY MILDEW8 (RPW8)-NBS-LRR* (*RNL*) genes.

Gene name	Domain name	Domain region (amino acid)	Amplification regions (amino acid)
GEZE01034532	TIR	2–119	
	NB-ARC	154–409	
	LRR	433–928	268–680
GEZE01070617	CC	2–126	
	NB-ARC	168–463	
	LRR	601–893	580–860
GEZE01020818	RPW8	9–126	
	NB-ARC	185–452	
	LRR	611–781	576–875