Research Article

Molecular identification of species in *Prunus* sect. *Persica* (Rosaceae), with emphasis on evaluation of candidate barcodes for plants

[1,2]Xu QUAN   [1]Shi-Liang ZHOU*

[1](*State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences*, Beijing 100093, China)

[2](*Graduate School of the Chinese Academy of Sciences*, Beijing 100049, China)

Abstract   Species of *Prunus* L. sect. *Persica* are not only important fruit trees, but also popular ornamental and medicinal plants. Correct identification of seedlings, barks, or fruit kernels is sometimes required, but no reliable morphological characters are available. Nowadays, the technique of DNA barcoding has the potential to meet such requirements. In this study, we evaluated the suitability of 11 DNA loci (*atp*B-*rbc*L, *trn*H-*psb*A, *trn*L-F, *trn*S-G, *atp*F-H, *rbc*L, *mat*K, *rpo*B, *rpo*C1, *nad*1, and internal transcribed spacer [ITS]) as candidate DNA barcodes for peaches, using samples from 38 populations, covering all the species in sect. *Persica*. On the whole, the primers worked well in this group and sequencing difficulties were met only in the case of ITS locus. Five loci (*rbc*L, *mat*K, *rpo*B, *rpo*C, and *nad*1) have very low variation rates, whereas *atp*B-*rbc*L, *atp*F-H, *trn*H-*psb*A, *trn*L-F and *trn*S-G show more variability. The most variable loci, *atp*B-*rbc*L and *trn*H-*psb*A, can distinguish three of the five species. Two two-locus combinations, *atp*B-*rbc*L+*trn*L-F and *atp*B-*rbc*L+*atp*F-H, can resolve all five species. We also find that identification powers of the loci are method-dependent. The NeighborNet method shows higher species identification power than maximum parsimony, neighbor joining, and unweighted pair group method with arithmetic mean methods.

Key words   *atp*B-*rbc*L, DNA barcode, *Prunus* sect. *Persica*, *trn*L-F.

First proposed by Hebert et al. (2003), DNA barcoding is a technique that uses DNA sequence information to identify species. Recently, DNA barcoding has become a hotspot of studies on biodiversity and taxonomy of plants. However, plant DNA barcoding is still in its infant stage. No universally applicable locus has been found and no consensus has been reached on the final choice from the available candidates, although *mat*K and *rbc*L have been suggested as viable options (CBOL Plant Working Group, 2009). Great effort has been put into screening the recommended candidates (e.g. Lahaye et al., 2008; Liu et al., 2010; Newmaster et al., 2008; Shi et al., 2009). Some loci, such as internal transcribed spacer (ITS) and *trn*H-*psb*A, have been proposed more than once (Sass et al., 2007; Shi et al., 2009). Due to unsatisfactory performance of the majority of the candidates, combinations of loci were suggested, for example *rpo*C1+*rpo*B+*mat*K or *rpo*C1+*mat*K+*trn*H-*psb*A (Chase et al., 2007), *mat*K+*atp*F-H+*psb*K-I or *mat*K+*atp*F-H+*trn*H-*psb*A (summarized by Pennisi, 2007).

The early studies suffered from sparse and remote sampling of taxa. It is questionable if the candidates based on those early studies are applicable to the well-sampled cases. A final conclusion must be based on a lot of case studies. In this study, we present such a case study carried out on peaches, *Prunus* L. sect. *Persica* (L.) S. L. Zhou & X. Quan, stat. nov. [Basionym. *Prunus* L. subgen. *Persica* Linn., Sp. Pl. 472, 1753].

Prunus sect. *Persica* is an ideal group for evaluation of candidate DNA barcodes. The species in the section are well understood and reliably delimited. Peaches are common fruits. Peach flowers are early spring blooms, and peach kernels are a medicine. Thus, correct identification is of paramount importance in peach fruit production and peach kernel toxicology.

Prunus sect. *Persica* is one of several monophyletic clades in the genus of *Prunus*. According to the updated taxonomy by Quan (2010), there are five species in the section: *P. davidiana* (Carr.) Franch., *P. kansuensis* Rehd., *P. mira* Koehne, *P. potanini* (Batal.) S. L. Zhou & X. Quan, comb. nov. [Basionym. *P. persica* (L.) Batsch var. *potanini* Batal. in Acta Hort. Petrop. 12: 164, 1892], and *P. persica* (L.) Batsch. All these species are endemic to China. *Prunus persica* has been cultivated in China for approximately 3000 years. There are

two subspecies, subsp. *persica* and subsp. *ferganensis* Kost. & Riab., three varieties, var. *compressa* Loud., var. *densa* Makino, and var. *nucipersica* (Suckow) Schneid., and many forms such as f. *pendula* Dipp. and f. *duplex* Rehd.

Taxonomists use morphological features of flowers and drupes to identify the species in the section. To the lay mind those features are difficult to grasp. In peach agriculture, farmers need to know the species of seedling rootstocks used for grafting. Peach kernels are used both for medicinal purposes and food. The concentration of compounds containing the CN radical in the kernels differs considerably among species. Misuse or abuse of peach kernels may cause poisoning. Reliable methods, such as DNA barcoding, for identification of species would be very useful in such situations.

In this study, we evaluate the suitability of nine candidate chloroplast loci (*atp*B-*rbc*L, *trn*H-*psb*A, *trn*L-F, *trn*S-G, *atp*F-H, *rbc*L, *mat*K, *rpo*B, and *rpo*C1), one mitochondrial locus (*nad*1) and one nuclear locus (ITS) to be DNA barcodes of sect. *Persica*.

1   Material and methods

1.1   Sampling

All the species of sect. *Persica* and important subspecies, varieties, and forms of *P. persica* were sampled in this study. Fresh leaves were collected from wild populations in Hebei, Gansu, Sichuan, and Tibet, or from Zhengzhou Fruit Research Institute (the Chinese Academy of Agricultural Sciences), where they were introduced from the wild (Table 1). *Prunus armeniaca* L. and *P. mume* Sieb. [sect. *Armeniaca* (Mill.) Koch] were selected as outgroups when rooting of a gene tree was necessary. Voucher specimens were deposited in PE (Herbarium of Institute of Botany, Chinese Academy

Table 1   Taxa of *Prunus* L. sect. *Persica*, sources of materials sequenced, and voucher specimens (PE)

Taxon	Locality	Voucher
P. davidiana		
1	Xiangshan, Beijing, China	QX095
2	ZFRY†	QX016
3	ZFRY	QX017
4	ZFRY	QX018
5	Beijing Botanical Garden of Chinese Academy of Sciences, Beijing, China	QX092
P. potanini		
1	Chengkou, Chongqing, China	SL4805–80
2	Chengkou, Chongqing, China	SL4805–83
3	Chengkou, Chongqing, China	SL4805–84
P. kansuensis		
1	Maijishan, Tianshui, Gansu, China	WJ8013–88
2	ZFRY	QX026
P. mira		
1	Maerkang, Sichuan, China	QX079
2	Chayu, Tibet, China	QX138
3	Luobulinka, Tibet, China	QX093
4	Mangkang, Tibet, China	QX137
5	ZFRY	QX028
P. persica		
subsp. *persica*1	ZFRY	QX045
subsp. *persica*2	ZFRY	QX046
subsp. *persica*3	ZFRY	QX048
subsp. *persica*4	ZFRY	QX051
subsp. *persica*5	ZFRY	QX052
subsp. *ferganensis*1	Kashi, Xinjiang, China	QX023
subsp. *ferganensis*2	ZFRY	QX019
subsp. *ferganensis*3	ZFRY	QX020
f. *duplex*	ZFRY	QX029
f. *pendula*	ZFRY	QX032
var. *compressa*	ZFRY	QX044
var. *densa*	ZFRY	QX050
var. *nucipersica*	ZFRY	QX039
Outgroups		
P. armeniaca	Lüergou, Tianshui, Gansu, China	SL4802–69
P. mume	Beijing Botanical Garden of Chinese Academy of Sciences, Beijing, China	QXZ05

†Refer to Zhengzhou Fruit Research Institute, the Chinese Academy of Agricultural Sciences, Zhengzhou, China.

Table 2   Evaluated loci and primers used for amplification and sequencing in this study

Genome	Gene	Primer name	Primer sequence (5′ to 3′)
Chloroplast			
	*atp*B-*rbc*L	atpB-1	ACATCKARTACKGGACCAATAA
		rbcL-1	AACACCAGCTTTRAATCCAA
	*atp*F-H	atpF	ACTCGCACACACTCCCTTTCC
		atpH	GCTTTTATGGAAGCTTTAACAAT
	*mat*K	matK1f	CGATCTATTCATTCAATATTTC
		matK1r	TCTAGCACACGAAAGTCGAAGT
	*rbc*L	rbcL-af	ATGTCACCACAAACAGAGACTAAAGC
		rbcL-ar	TTGGAAATGGGAAGATCTAGG
	*rpo*B	rpoB1f	AAGTGCATTGTTGGAACTGG
		rpoB4r	GATCCCAGCATCACAATTCC
	*rpo*C1	rpoC1f	GTGGATACACTTCTTGATAATGG
		rpoC4r	CCATAAGCATATCTTGAGTTGG
	*trn*L-F	trn-c	CGAAATCGGTAGACGCTACG
		trn-f	ATTTGAACTGGTGACACGAG
	*trn*H-*psb*A	trnH^{GUG}	CGCGCATGGTGGATTCACAATCC
		psbA	GTTATGCATGAACGTAATGCTC
	*trn*S-G	trnS^{UUC}	AGATAGGGATTCGAACCCTCGGT
		trnG2S	TTTTACCACTAAACTATACCCGC
Mitochondrial			
	*nad*1	matRf	AAGCTCGGACTCGGTCA
		nad1r	TTGGAAATGGGAAGATCTAGG
Nuclear			
	ITS	ITS-1	AGAAGTCGTAACAAGGTTTCCGTAGG
		ITS-4	TCCTCCGCTTATTGATATGC
	ITS	C26A	GTTTCTTTTCCTCCGCT
		N-nc18s10	AGGAGAAGTCGTAACAAG

ITS, internal transcribed spacer.

of Sciences) and experiments were carried out at the State Key Laboratory of Systematic and Evolutionary Botany.

1.2   DNA isolation, fragment amplification, and sequencing

Total genomic DNA was isolated by the CTAB method (Doyle & Doyle, 1987). Eleven loci were amplified by polymerase chain reaction (PCR) using the primers given in Table 2. The amplifications were carried out in 20 μL reactions, containing 2.0 μL 10× buffer, 2.0 μL dNTPs (2 μmol/L), 1.0 μL each primer (5 μmol/L), 1.0 μL genomic DNA (∼5 ng), 0.2 μL Taq polymerase (5 U/μL), and 11.8 μL ddH$_2$O. The PCR program started at 94°C for 3 min, followed by 34 cycles of 94°C for 30 s, 50°C (55°C for ITS) for 30 s and 72°C for 2 min, and ended at 72°C for 5 min. These PCR amplifications were carried out on a C1000 Thermal Cycler (Bio-Rad Laboratories, Hercules, CA, USA). Both strands were sequenced on an ABI 3730xl DNA Analyzer (Applied Biosystems, Foster City, CAS, USA), following the manufacturer's instruction. The sequences were treated using Sequencer 4.7 (Gene Codes, Ann Arbor, MI, USA) for assemblage, ClustalX (Thompson et al., 1997) for alignment, and Se-Al 2.0 (Rambaut, 1996) for manual adjustment.

1.3   Data analysis

The successful rates of PCR amplifications and sequencing were calculated manually. DNA polymorphism was estimated using DnaSP v5 (Librado & Roza, 2009). The inter- and intraspecific variations of each locus were calculated using Mega 4.0 (Kumar et al., 2001). The significance of differences between inter- and intraspecific genetic distances was tested online (www.fon.hum.uva.nl/Service/Statistics.html). As not one of the known tree-building methods is substantially better than the others, a NeighborNet network was built by SplitsTree 4 (Huson & Bryant, 2006) and the widely used distance methods of neighbor joining (NJ) and unweighted pair group method with arithmetic mean (UPGMA), based on Kimura 2-parameter (K2P) distance, and the maximum parsimony (MP) method were tried using PAUP* 4.0b10 (Swofford, 2003). Bootstrap values were calculated with 1000 replicates for NJ and UPGMA. Gaps were encoded zero or one as new characters. Maximum parsimony analysis was carried out using heuristic search by treating all characters equally weighted and unordered, with tree bisection–reconnection branch swapping in effect and MulTrees on. Branch support for MP trees was assessed with 1000 bootstrap replicates, 100 random taxon addition replicates, and saving all trees at each step.

Table 3   Variability of 11 loci across *Prunus* sect. *Persica*

Locus	N	Ps (%)	Ss (%)	L (bp)	Indel	Pi	Vs		Is		D1	D2	R
							No.	%	No.	%			
*atp*B-*rbc*L	28	100%	100%	796	1	3.40×10^{-3}	10	1.26	9	1.13	0.508×10^{-3}	3.179×10^{-3}	6.3
*atp*F-H	28	100%	100%	543	0	5.24×10^{-3}	7	1.29	7	1.29	0	6.706×10^{-3}	†
*mat*K	17	100%	100%	773	0	1.03×10^{-3}	5	0.65	2	0.26	0.782×10^{-3}	5.817×10^{-3}	7.4
*nad*1	17	100%	100%	787	0	0.15×10^{-3}	1	0.13	1	0.13	0.254×10^{-3}	0.318×10^{-3}	1.3
*rbc*L	17	100%	100%	596	0	1.29×10^{-3}	2	0.34	2	0.34	0	2.021×10^{-3}	†
*rpo*B	17	100%	100%	323	0	2.91×10^{-3}	4	1.24	2	0.62	1.244×10^{-3}	4.354×10^{-3}	3.5
*rpo*C1	17	100%	100%	555	0	1.25×10^{-3}	5	0.90	1	0.18	1.449×10^{-3}	2.172×10^{-3}	1.5
*trn*H-*psb*A	28	94%	90%	337	1	3.24×10^{-3}	7	2.01	4	1.19	1.986×10^{-3}	3.806×10^{-3}	1.9
*trn*L-F	28	93%	95%	853	0	4.21×10^{-3}	16	1.88	9	1.06	1.293×10^{-3}	5.817×10^{-3}	4.5
*trn*S-G	17	88%	100%	716	5	3.92×10^{-3}	12	1.68	5	0.70	2.340×10^{-3}	5.149×10^{-3}	2.2
ITS‡	28	86%	37%	678	NA	NA	NA	NA	NA	NA	NA	NA	NA
ITS§	28	91%	48%	678	NA	NA	NA	NA	NA	NA	NA	NA	NA

†When intraspecies distance (D1) is zero, R (D2/D1, where D2 is interspecies distance) is not available. ‡Polymerase chain reaction (PCR) and sequencing by primer pair ITS-1 and ITS-4. §PCR and sequencing by primer pair C26A and N-nc18s10. Is, informative site; L, sequenced length of fragments; N, samples examined; NA, not applicable; Pi, nucleotide diversity per site; Ps, PCR success; Ss, sequencing success; Vs, variable site.

2   Results

2.1   Universality of primers

The universal primers listed in Table 2 are generally applicable to all plants in this study. Relatively low success rates were encountered with *trn*H-*psb*A, *trn*S-G, and *trn*L-F. We had some sequencing problems with ITS locus using both pairs of primers.

2.2   Variability and resolution of single locus

Loci *rbc*L, *rpo*B, *rpo*C, *mat*K, and *nad*1 were highly conserved, with very lower variable rates and nucleotide diversity (Table 3). They are inadequate for identification of the five species in the section. Loci *atp*B-*rbc*L, *trn*H-*psb*A, *atp*F-H, *trn*S-G and *trn*L-F have relatively higher variable rates and *Pi* values, with more than six variable sites and more than three parsimony-informative sites (Table 3). Among the 10 candidate loci, *trn*L-F is the most variable locus and *atp*F-H had the highest nucleotide diversity.

Based on the K2P distance, the inter- and intraspecies average genetic distances varied significantly among the loci (Table 3). No intraspecies distances were detected at loci *atp*F-H and *rbc*L, and very large interspecies distances exist at *atp*F-H, *trn*L-F, and *mat*K. The *R* value, a ratio between the average interspecific genetic distances to the average intraspecific genetic distance, measures the probability of correct identifications. The highest *R* values were found for *mat*K, followed by *atp*B-*rbc*L and *trn*L-F. Median test indicated *atp*B-*rbc*L, *atp*F-H, and *mat*K as the best ($P = 0.034$–0.062), whereas Wilcoxon two-sample tests suggested *atp*B-*rbc*L, *atp*F-H, and *trn*L-F ($P < 0.05$, Table 4).

The barcoding power was determined by the number of monophyletic clades of the same species. The barcoding powers vary significantly among loci and slightly among tree-building methods (Table 5). The loci *atp*B-*rbc*L and *trn*H-*psb*A showed the highest discriminatory power, in that three of the five species were resolved. The NeighborNet method resolved more species than other methods.

Table 4   Significance of interspecific *versus* intraspecific variations given by the median test and Wilcoxon test based on Kimura 2-parameter distances of each locus

Locus	Median test				Wilcoxon two-sample test			
	#A	#B	M	$P \leq$	#A	#B	W	$P \leq$
*atp*B-*rbc*L	5	10	0.0018	0.062	5	10	15.0	0.0007
*atp*F-H	5	10	0.0037	0.062	5	10	17.5	0.0040
*trn*H-*psb*A	5	10	0.0032	0.206	5	10	27.0	0.1219
*trn*L-F	5	10	0.0061	0.206	5	10	23.0	0.0273
*trn*S-G	5	10	0.0047	0.388	5	10	27.5	0.1299
*mat*K	5	10	0.0013	0.034	5	10	27.0	0.1272
*nad*1	5	8	0	†	5	8	31	0.6000
*rbc*L	5	10	0	†	5	10	25	0.1000
*rpo*B	5	10	0.0031	0.388	5	10	24.0	0.0506
*rpo*C1	5	10	0.0018	0.562	5	10	31.0	0.2704

†When median value (M) is zero, probability (p) is not available. #A, number of intraspecific comparisons; #B, number of interspecific comparisons; W, Wilcoxon value.

Table 5   Barcoding power of single locus and two-locus combinations using different dendrogram building methods. Five species of *Prunus* L. sect. *Persica* were used for analysis

Loci		MP	NJ	UPGMA	NeighborNet
Single locus	*atp*B-*rbc*L	3	3	3	3
	*trn*H-*psb*A	3	3	3	3
	*atp*F-H	2	2	2	3
	*trn*L-F	2	2	2	3
	*mat*K	1	1	2	2
	*rpo*B	1	1	1	1
	*rbc*L	0	1	0	0
	*trn*S-G	0	1	0	0
	*rpo*C1	0	0	1	1
	*nad*1	0	0	0	0
Two-locus combination	*atp*B-*rbc*L+*trn*L-F	5	5	5	5
	*atp*F-H+*atp*B-*rbc*L	4	5	5	5
	*mat*K+*atp*F-H	4	4	4	4
	*rbc*L+*atp*B-*rbc*L	3	4	4	4
	*mat*K+*trn*L-F	3	3	4	4
	*trn*L-F+*trn*H-*psb*A	3	3	4	4
	*trn*H-*psb*A+*atp*B-*rbc*L	3	3	3	4
	*atp*F-H+*trn*H-*psb*A	3	3	3	3
	*rbc*L+*atp*F-H	3	3	3	3
	*mat*K+*atp*B-*rbc*L	3	3	2	3
	*rbc*L+*trn*L-F	2	3	2	3
	*atp*F-H+*trn*L-F	2	2	2	3
	*mat*K+*rbc*L	2	2	2	3
	*mat*K+*trn*H-*psb*A	2	2	2	2
	*rbc*L+*trn*H-*psb*A	1	1	1	1

MP, maximum parsimony; NJ, neighbor joining; UPGMA, unweighted pair group method with arithmetic mean methods.

2.3   Resolution of the two-locus combinations

Four loci (*atp*B-*rbc*L, *atp*F-H, *mat*K, and *trn*L-F), selected according to *R* value, median test, and Wilcoxon two-sample test, together with *trn*H-*psb*A and *rbc*L suggested by CBOL Plant Working Group (2009), were combined for testing the discriminatory powers of two-locus combinations. Other invariable loci were ignored in the analyses of two-locus combinations. Among the 15 tested combinations, *atp*B-*rbc*L+*trn*L-F and *atp*B-*rbc*L+*atp*F-H showed the highest discriminatory power, resolving all five species (Table 5, Fig. 1). As only *atp*B-*rbc*L+*trn*L-F is capable of identifying all five species by all four methods, this combination is the best choice for barcoding the peach species. A phylogenetic NJ tree of this combination was built based on K2P distance (Fig. 1). The NJ tree shows that each of the five species in the section forms a well-supported monophyletic clade with bootstrap values larger than 70%.

2.4   Comparisons of different methods

Three phylogenetic methods (MP, NJ, and UP-GMA) and a network method (NeighborNet) were compared. Results (Table 5) showed only slight differences in the species discriminatory powers. The NeighborNet network method performs the best in species discrimination.

3   Discussion

Although many scientific teams are pursuing a standard, universal barcode for plants and some suggestions have been put forward (CBOL Plant Working Group, 2009), there are many different opinions (Lahaye et al., 2008; Newmaster et al., 2008; Shi et al., 2009; Zhang et al., 2009; Liu et al., 2010; Zuo et al., 2010) and no consensus has been reached. The early works were primarily based on extensive sampling of relatively remote taxa. This explains why *rbc*L and *mat*K were suggested as candidates: these two loci work well at a family level. It has been suggested that barcoding should be done both at family level and below family level (Zhang et al., 2009). Following this identification strategy we would pinpoint first a family, then a species in this family using more sensitive barcodes.

Screening the already-suggested loci to find both an effective, universal barcode for most plants and to obtain specific DNA markers for certain taxa are the tasks to be recommended. A universal barcode assessment might have to draw from many case studies, and a specific barcode resolving a high percentage of species is appealing to solve species identification problems.

The universality of the primers for each of the tested loci is high, except for ITS in this study. Internal
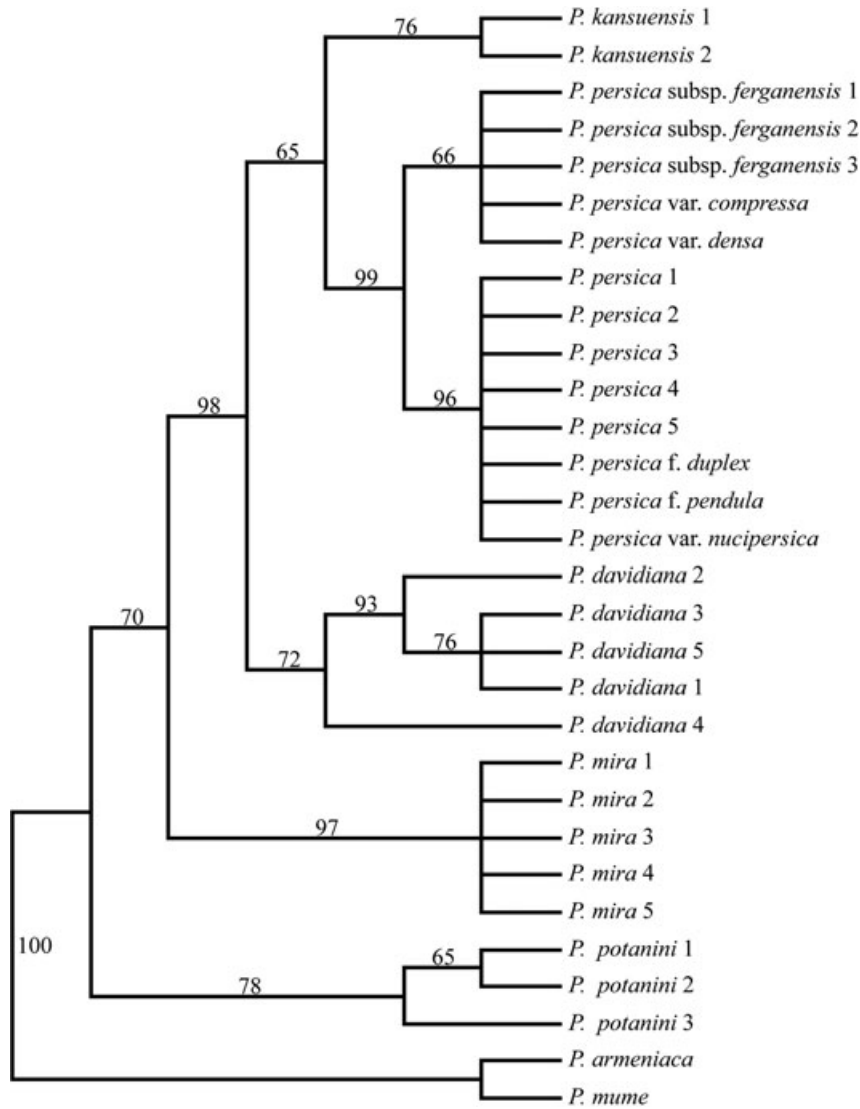
Fig. 1. Neighbor-joining tree based on the suggested barcode (a combination of *atp*B-*rbc*L and *trn*L-F) of *Prunus* sect. *Persica*. Kimura 2-parameter distances were used. Bootstrap values are shown above the branches.

transcribed spacer was suggested as DNA barcode for plants in several previous studies (Kress et al., 2005; Ren et al., 2010; Zuo et al., 2010), but the sequencing success rates were inadequate in some studies (Kress & Erickson, 2007; Sass et al., 2007). In this study, amplification of ITS using either pair of primers was satisfactory, although some weak bands were also produced. However, sequencing of ITS was a big problem, in that many sequences obtained were of very low quality. This is probably due to false priming or incomplete concerted evolution of different repeats.

Both *rbc*L and *mat*K were recommended by the CBOL Plant Working Group (2009) as the most promis-

ing candidates of the plant barcodes. However, *rbc*L is extremely conservative among peaches and *mat*K is only a little better (Table 3). They are perhaps suitable for identification of families instead of species (e.g. Ren et al., 2010). The loci *rpo*B and *rpo*C1 are even more conservative. These four loci (*rbc*L, *mat*K, *rpo*B, and *rpo*C1) are not good choices for plant barcodes at low taxonomic levels. In contrast, *trn*S-G, often used for plant phylogeny (Shaw & Small, 2004; Murdock, 2008; Zhang et al., 2009), showed considerable infra- and interspecific variations both within and among different species. In addition, a 277bp indel of *trn*S-G may be present or absent in different samples of the

same species. This locus is also unsuitable for DNA barcoding.

The non-coding regions have relatively higher evolution rates as well as more informative characters, and are more powerful in species identification. Hence studies have emphasized searching barcodes for plants at species level (Pennisi, 2007; Fazekas et al., 2008; Lahaye et al., 2008; Zuo et al., 2010). Four non-coding loci (*atp*B-*rbc*L, *trn*H-*psb*A, *trn*L-F, and *atp*F-H) are the potential candidates for sect. *Persica*. The loci *atp*B-*rbc*L and *trn*H-*psb*A showed the best performance, discriminating three species out of five. To identify all species in the section, at least two loci are needed. Among 15 combinations, *atp*B-*rbc*L+*trn*L-F and *atp*B-*rbc*L+*atp*F-H showed the best discriminatory ability for sect. *Persica,* identifying all five species. Compared with *atp*B-*rbc*L+*atp*F-H, the combination of *atp*B-*rbc*L+*trn*L-F can identify all five species using either MP, NJ, UPGMA, or NeighborNet methods. Therefore, *atp*B-*rbc*L+*trn*L-F is the best choice for DNA barcoding sect. *Persica*. Moreover, this combination is also workable to identify some varieties of hybrid origin with *P. persica* subsp. *ferganensis* being their maternal parent, such as *P. persica* var. *compressa* and *P. persica* var. *densa* (Fig. 1). However, it is likely to fail if the maternal parent is *P. persica* subsp. *persica*, such as *P. persica* f. *duplex* and *P. persica* f. *pendula*.

Assessment of a locus is also method-dependent (Austerlitz et al., 2009). The parsimony-based methods ignore singletons, but the distance-based methods use them. Therefore, parsimony-based methods are less sensitive but more reliable. If the phylogeny of the group has not been well established, parsimony-based methods are preferred for DNA barcoding research. The distance-based methods usually have high resolution, but they cannot distinguish sequencing problems from real differences. Such methods are useful in groups for which the methods have been already tested. In our study, the NeighborNet method, a kind of network method, seems to be more sensitive, with higher discriminatory power for closely related species (this also applies to *Panax* L. (Zuo et al., 2010), showing relatively high ability in resolving semispecies). However, this method's universality and suitability for DNA barcoding studies requires more in-depth investigation.

References

Austerlitz F, David O, Schaeffer B, Bleakley K, Olteanu M, Leblois R, Veuille M, Laredo C. 2009. DNA barcode analysis: A comparison of phylogenetic and statistical classification methods. BMC Bioinformatics 10 (Suppl 14): S10.

CBOL Plant Working Group. 2009. A DNA barcode for land plants. Proceedings of the National Academy of Sciences USA 106: 12794–12797.

Chase MW, Cowan RS, Hollingsworth PM, wan den Berg C, Madrinan S, Petersen G, Seberg O, Jorgsensen T, Cameron KM, Carine M, Pedersen N, Hedderson TAJ, Conrad F, Salazar GA, Richardson JE, Hollingsworth ML, Barraclough TG, Kelly L, Wilkinson M. 2007. A proposal for a standardized protocol to barcode all land plants. Taxon 56: 295–299.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of leaf tissue. Phytochemistry Bulletin 19: 11–15.

Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, Husband BC, Percy DM, Hajibabaei M, Barrett SCH. 2008. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS ONE 3: e2802. doi: 10.1371/journal.pone.0002802.

Hebert PDN, Cywinska A, Ball SL, de Waard JR. 2003. Biological identification through DNA barcodes. Proceedings of the Royal Society of London Series B: Biological Sciences 270: 313–321.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23: 254–267.

Kress WJ, Erickson DL. 2007. A two-locus global DNA barcode for land plants: The coding *rbcL* gene complements the non-coding *trn*H-*psb*A spacer region. PLoS ONE 2: e508.

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 USA 102: 8369–8374.

Kumar S, Tamura K, Jacobsen IB, Nei M. 2001. MEGA 2: Molecular evolutionary genetics analysis software. Bioinformatics 17: 1244–1245.

Lahaye R, Savolainen V, Duthoit S, Murin O, Michelle Van Der-Bank. 2008. A test of *psb*K-*psb*I and *atp*F-*atp*H as potential plant DNA barcodes using the flora of the Kruger National Park as a model system (South Africa). Nature 1896: 1.

Librado P, Roza J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451–1452.

Liu Y, Yan HF, Cao T, Ge XJ. 2010. Evaluation of 10 plant barcodes in Bryophyta (mosses). Journal of Systematics and Evolution 48: 36–46.

Murdock A. 2008. Phylogeny of marattioid ferns (Mararrriaceae): Inferring a root in the absence of a closely related outgroup. American Journal of Botany 95: 626–641.

Newmaster SG, Fazekas AJ, Steeves RAD, Janovec J. 2008. Testing candidate plant barcode regions in the Myristicaceae. Molecular Ecology Resources 8: 490–490.

Pennisi E. 2007. Wanted: A barcode for plants. Science 318: 190–191.

Quan X. 2010. Systematics of peaches and apricots. Masters Degree Dissertation. Beijing: Institute of Botany, Chinese Academy of Sciences. (In the dissertation, the generic name *Amygdalus* Linn. was used instead of *Prunus*.)

Rambaut A. 1996. Se-Al: Sequence alignment editor (Se-Al v2.0a11) [online]. Available from http://tree.bio.ed.ac.uk/software/seal.

Ren BQ, Xiang XG, Chen ZD. 2010. Species identification of *Alnus* (Betulaceae) using nrDNA and cpDNA genetic markers. Molecular Ecology Resources 10: 594–605.

Sass C, Little DP, Stevenson DW, Specht CD. 2007. DNA barcoding in the Cycadales: Testing the potential of proposed barcoding markers for species identification of cycads. PLoS ONE 11: e1154. doi: 10.1371/journal.pone.0001154.

Shaw J, Small RL. 2004. Addressing the "hardest puzzle in American pomology": Phylogeny of *Prunus* sect. *Prunocerasus* (Rosaceae) based on seven noncoding chloroplast DNA regions. American Journal of Botany 91: 985–996.

Shi LC, Liang ZS, Han JP, Song JY, Yao H, Zhu YJ, Chen SL. 2009. Screening potential DNA barcode regions in *Rhododendron*. World Science and Technology/Modernization of Traditional Chinese Medicine and Materia Medica 11 (1): 54–57.

Swofford DL. 2003. PAUP* 4.0b10 phylogenetic analysis using parsimony (*and other methods).Version 4. Sunderland: Sinauer Associates.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The ClustalX windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acid Research 25: 4876–4882.

Zhang JM, Wang JX, Xia T, Zhou SL. 2009. DNA barcoding: Species delimitation in tree peonies. Science in China Series C: Life Sciences 52 (6): 568–578.

Zuo YJ, Chen ZJ, Kondo K, Funamoto T, Wen J, Zhou SL. 2010. DNA barcoding of *Panax* species. Planta Medica 77: 180–185.