Genetics and Molecular Research

Applying DNA barcodes for identification of economically important species in Brassicaceae

X.Q. Sun[1]*, Y.Q. Qu[1]*, H. Yao[2], Y.M. Zhang[1], Q.Q. Yan[1] and Y.Y. Hang[1]

[1]Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing, China
[2]Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, China

*These authors contributed equally to this study.
Corresponding author: Y.Y. Hang
E-mail: hangyueyu@cnbg.net

Genet. Mol. Res. 14 (4): 15050-15061 (2015)
Received June 11, 2015
Accepted August 11, 2015
Published November 24, 2015
DOI http://dx.doi.org/10.4238/2015.November.24.13

ABSTRACT. Brassicaceae is a large plant family of special interest; it includes many economically important crops, herbs, and ornamentals, as well as model organisms. The taxonomy of the Brassicaceae has long been controversial because of the poorly delimited generic boundaries and artificially circumscribed tribes. Despite great effort to delimitate species and reconstruct the phylogeny of Brassicaceae, little research has been carried out to investigate the applicability and effectiveness of different DNA regions as barcodes - a recent aid for taxonomic identification - to identify economically important species in Brassicaceae. In this study, we evaluated the feasibility of five intensively recommended regions [*rbc*L, *mat*K, *trn*H-*psb*A, internal transcribed spacer (ITS), ITS2] as candidate DNA barcodes to discriminate economic species of Brassicaceae in China and try to establish a new digital identification method for economic plants of Brassicaceae. All sequences of 58 samples from 27 economic species (Brassicaceae) in China were assessed in the success rates of

PCR amplifications, intra- and inter-specific divergence, DNA barcoding gaps, and efficiency of identification. Compared with other markers, ITS showed superiority in species discrimination with an accurate identification of 67.2% at the species level. Consequently, as one of the most popular phylogenetic markers, our study indicated that ITS was a powerful but not perfect barcode for Brassicaceae identification. We further discuss the discrimination power of different loci due to inheritance pattern, polyploidization and hybridization in species-specific evolution. Further screening of other nuclear genes related to species isolation as plant barcode candidates is also proposed.

Key words: *mat*K; *rbc*L; *trn*H-*psb*A; ITS; ITS2; Brassicaceae

INTRODUCTION

Brassicaceae (Cruciferae or mustard family) is a large plant family with approximately 338 genera and 3,709 species widely distributed globally. It is of special interest, as it includes many economically important crops, herbs, ornamentals, and model organisms. The most important edible oil crop is canola or rapeseed (*Brassica napus* L.), while mustard condiment crops include *Brassica juncea* (L.) Czern. et Coss. and *Sinapis alba* L. Many species are also important vegetable crops, e.g. *Brassica oleracea* L. Several species, e.g., *Camelina sativa* (L.) Crantz, *Crambe abyssinica* Hochst. ex Fries, and *Eruca sativa* Mill., have potential as new edible/industrial oil crops. Many crucifers are grown as ornamentals, e.g., *Orychophragmus violaceus* (L.) O.E. Schulz, *Matthiola incana* (L.) R. Br. and others. Over 100 genera have been used for medical purposes in virtue of biological constituents such as sinapine, cardiac glycoside, alkaloids, flavonol, and phenols. The Chinese Pharmacopeia (2010) has admitted several botanical origins, e.g. *Raphanus sativus* L. (Raphanus Semen), *Lepidium apetalum* Willd. or *Descurainia Sophia* (L.) Webb ex Prantl. (Descurainiae Semen or Lepidii Semen), *Isatis indigotica* Fort. (Isatidis Radix) and so on. Several representatives of the family, including *Arabidopsis thaliana* (L.) Heynh. and *Brassica* spp., have achieved the well-accepted status of “model organisms” for genomic studies.

Brassicaceae is a natural family and can be easily distinguished morphologically from species of other flowering families based on its highly conserved and fairly uniform flower architecture. However, the taxonomy of the Brassicaceae has long been controversial because of the often poorly delimited generic boundaries and artificially circumscribed tribes. Several authors have tried to provide a natural system to divide the family of Brassicaceae into tribes or genera (Schulz, 1936; Janchen, 1942; Al-Shehbaz, 1984). The characters traditionally used in these studies are few; they include orientation of the radicle in relation to the cotyledons in the embryo, fruit length-to-width ratio, fruit compression and dehiscence, number of rows of seeds in each locule, trichome type, and features of the nectarines. However, most of the characters considered are subject to convergent evolution, at least on the tribal and subtribal level (Hedge, 1976; Al-Shehbaz, 1984). Within the past two decades, several molecular phylogenetic studies on Brassicaceae (Bailey et al., 2006; Koch et al., 2007; Beilstein et al., 2008; German et al., 2009; Khosravi et al., 2009; Warwick and Hall, 2009; Couvreur et al., 2010; Warwick et al., 2010; German et al., 2011; Goodson et al., 2011) have refined the tribal classification, resurrected several tribes previously misrecognized, added the newly established ones, and adjusted limits of many genera. Despite the substantial progress achieved during the past 20 years along the phylogenetic and

systematic fronts of the family, many unresolved problems, especially the limits of tribes and the discrimination of species, remain unaddressed.

Plant DNA barcoding has recently emerged as a tool for global species identification and has proven extremely useful for numerous applications such as ecological forensics, identification of traded materials, undertaking identifications where there is a shortage of taxonomic expertise available, and assisting species discovery in some plant groups (reviewed in Hollingsworth et al., 2011). In animals, the mitochondrial cytochrome *c* oxidase I gene (*COI*) has been favored in species identification; however, this gene has been precluded as a universal plant barcode because of its generally low rate of nucleotide substitution in plant mitochondrial genomes. Therefore, several candidate markers have been proposed for use in plants, including coding plastid markers (*mat*K, *rbc*L, *rpo*B and *rpo*C1) (Kress et al., 2005; Chase et al., 2007; Lahaye et al., 2008), noncoding spacers (*psb*A-*trn*H, *atpF*-*atp*H, ITS and ITS2) (Kress and Erickson, 2008; Chen et al., 2010; China Plant BOL Group et al., 2011), or various combinations of several markers (Chase et al., 2007).

Despite efforts to delimitate species and reconstruct the phylogeny of Brassicaceae, little research has been carried out to investigate the applicability and effectiveness of different DNA regions as barcodes to identify species within Brassicaceae. This is especially true for economically important species including edible and industrial oilseed, vegetable, herb, ornamental and fodder crop species. In this study, we utilized five intensively recommended regions (*rbc*L, *mat*K, *trn*H-*psb*A, ITS, ITS2) to evaluate their feasibility as candidate DNA barcodes to discriminate economically important Brassicaceae species in China and to establish a new digital identification method.

MATERIAL AND METHODS

Plant materials

Plant samples were collected from different locations in China and identified by one of our authors, Prof. Yueyu Hang. In total, 58 individual samples belonging to 27 species, representing a majority of economic species, were collected for further analysis. Fresh leaves were dried in silica gel at the time of collection. Voucher specimens were deposited in the herbarium at the Kunming Institute of Botany, Chinese Academy of Sciences (KIM) (Table 1).

DNA extraction, amplification, and sequencing

Genomic DNA was extracted following a cetyl trimethylammonium bromide (CTAB) protocol modified from Paterson et al. (2011). The universal primers rbcLa-f and rbcLa-rev (the Consortium for the Barcode of Life (CBOL) recommended), 3F_KIM and 1R_KIM (CBOL recommended), trnH and psbA (Sang et al., 1997), and ITS1 and ITS4 (White et al., 1990) were used in the amplification of *rbc*L, *mat*K, *psb*A-*trn*H, and ITS regions respectively.

Polymerase chain reaction (PCR) amplification of the four candidate barcodes was carried out using the following program: a premelt of 3 min at 94ºC, followed by 35 cycles of 45 s denaturation at 94ºC, 30 s annealing reaction at 53º-58ºC, and finally a 30 s extension at 72ºC. Each 20-µL reaction mixture contained 30 ng genomic DNA template, 2.5 mM $MgCl_2$, 1X Mg-free DNA polymerase buffer, 0.12 mM dNTPs, 0.3 µM of each primer, and 1 U *Taq* DNA polymerase. PCR products were examined electrophoretically on 0.8-1.2% agarose gels. Purification and bidirectional sequencing were completed by Beijing Genomics Institute (BGI) using the amplification primers.

Table 1. Collection sites and GenBank accession Nos. of 58 samples of 27 Brassicaceae species.

Species name	Uses	Collection sites	Voucher number	GenBank accession No.			
				*rbc*L	*mat*K	*tm*H-*psb*A	ITS
Arabidopsis thaliana (Linnaeus) Heynhold	model plant	Nanshan, Urumqi, Xinjiang, China	PS3001MT01	KM892707	KM892769	KM892553	KM892660
Arabidopsis thaliana (Linnaeus) Heynhold	model plant	Biwa Lake, Nanjing, Jiangsu, China	PS3001MT02	KM892696	KM892758	KM892554	KM892649
Arabis flagellosa Miquel	herb	Hangzhou Botanical Garden, Zhejiang, China	PS3023MT01	KM892688	-[a]	KM892605	KM892641
Brassica juncea (Linnaeus) Czernajew	vegetable; oil plant	Yongfeng, Urumqi, Xinjiang, China	PS3002MT01	KM892708	KM892770	KM892555	KM892661
Brassica juncea (Linnaeus) Czernajew	vegetable; oil plant	Huzhu, Xining, Qinghai, China	PS3002MT02	KM892709	KM892771	KM892556	KM892662
Brassica juncea (Linnaeus) Czernajew	vegetable; oil plant	Langya Mountain, Chuzhou, Anhui, China	PS3002MT03	KM892676	KM892743	KM892557	KM892629
Brassica juncea (Linnaeus) Czernajew	vegetable; oil plant	Tianzhu, Hangzhou, Zhejiang, China	PS3002MT04	KM892693	KM892755	KM892558	KM892646
Brassica juncea (Linnaeus) Czernajew	vegetable; oil plant	Nanjing Botanical Garden, Jiangsu, China	PS3002MT05	KM892704	KM892766	KM892559	KM892657
Brassica napus Linnaeus	vegetable; oil plant	Hangzhou Botanical Garden, Zhejiang, China	PS3024MT01	KM892692	-	KM892606	KM892645
Capsella bursa-pastoris (Linnaeus) Medikus	vegetable	Hougang, Dongtai, Jiangsu, China	PS3003MT01	KM892710	KM892772	KM892560	KM892663
Capsella bursa-pastoris (Linnaeus) Medikus	vegetable	Nanshan, Urumqi, Xinjiang, China	PS3003MT02	KM892711	KM892773	KM892561	KM892664
Capsella bursa-pastoris (Linnaeus) Medikus	vegetable	Qinghai Lake, Qinghai, China	PS3003MT03	KM892712	KM892774	KM892562	KM892665
Capsella bursa-pastoris (Linnaeus) Medikus	vegetable	Langya Mountain, Chuzhou, Anhui, China	PS3003MT04	KM892678	KM892745	KM892563	KM892631
Capsella bursa-pastoris (Linnaeus) Medikus	vegetable	Hangzhou Botanical Garden, Zhejiang, China	PS3003MT05	KM892684	KM892759	-	KM892637
Capsella bursa-pastoris (Linnaeus) Medikus	vegetable	Biwa Lake, Nanjing, Jiangsu, China	PS3003MT06	KM892697	-	KM892564	KM892650
Cardamine flexuosa Withering	herb	Zhangzhu, Yixing, Jiangsu, China	PS3004MT01	KM892713	KM892775	KM892565	KM892666
Cardamine flexuosa Withering	herb	Langya Mountain, Chuzhou, Anhui, China	PS3004MT02	KM892680	-	KM892566	KM892633
Cardamine flexuosa Withering	herb	Hangzhou Botanical Garden, Zhejiang, China	PS3004MT03	KM892686	KM892751	KM892567	KM892639
Cardamine flexuosa Withering	herb	Biwa Lake, Nanjing, Jiangsu, China	PS3004MT04	KM892699	KM892761	KM892568	KM892652
Cardamine impatiens Linnaeus	herb	Langya Mountain, Chuzhou, Anhui, China	PS3020MT01	KM892677	KM892744	KM892599	KM892630
Cardamine impatiens Linnaeus	herb	Hangzhou Botanical Garden, Zhejiang, China	PS3020MT02	KM892683	KM892749	KM892600	KM892636
Cardamine impatiens Linnaeus	herb	Hangzhou Botanical Garden, Zhejiang, China	PS3020MT03	KM892690	KM892753	KM892601	KM892643
Cardamine impatiens Linnaeus	herb	Biwa Lake, Nanjing, Jiangsu, China	PS3020MT04	KM892700	KM892762	KM892602	KM892653
Cardamine limprichtiana Pax	herb	Hangzhou Botanical Garden, Zhejiang, China	PS3022MT01	KM892685	KM892750	KM892604	KM892638
Cardamine flexuosa With. var. ovatifolia T.Y.Cheo et R.C.Fang	herb	Hangzhou Botanical Garden, Zhejiang, China	PS3004MT05	KM892687	KM892752	KM892569	KM892640
Chorispora tenella (Pallas) de Candolle	herb	Nanshan, Urumqi, Xinjiang	PS3005MT01	KM892714	KM892776	KM892570	KM892609
Coronopus didymus (Linnaeus) Smith	herb	Zhangzhu, Yixing, Jiangsu, China	PS3006MT01	KM892715	-	KM892571	KM892610
Coronopus didymus (Linnaeus) Smith	herb	Langya Mountain, Chuzhou, Anhui, China	PS3006MT02	KM892679	KM892746	KM892572	KM892632
Coronopus didymus (Linnaeus) Smith	herb	Tianzhu, Hangzhou, Zhejiang, China	PS3006MT03	KM892694	KM892756	KM892573	KM892647
Coronopus didymus (Linnaeus) Smith	herb	Biwa Lake, Nanjing, Jiangsu, China	PS3006MT04	KM892695	KM892757	KM892574	KM892648
Descurainia sophia (Linnaeus) Webb ex Prantl	herb	Tianchi Lake, Fukang, Xinjiang, China	PS3007MT01	KM892716	KM892725	KM892575	KM892611
Descurainia sophia (Linnaeus) Webb ex Prantl	herb	Qingshi Tsui, Mengyuan, Qinghai, China	PS3007MT02	KM892717	KM892726	KM892576	KM892612
Descurainia sophia (Linnaeus) Webb ex Prantl	herb	Hangzhou Botanical Garden, Zhejiang, China	PS3007MT03	KM892682	KM892748	KM892577	KM892635
Descurainia sophia (Linnaeus) Webb ex Prantl	herb	Hougang, Dongtai, Jiangsu, China	PS3007MT04	KM892706	KM892768	KM892578	KM892659
Draba nemorosa Linnaeus	herb	Biwa Lake, Nanjing, Jiangsu, China	PS3025MT01	KM892698	KM892760	KM892607	KM892651
Lepidium apetalum Willdenow	herb	Xining Botanical Garden, Qinghai, China	PS3008MT01	KM892718	KM892727	KM892579	KM892613
Lepidium ferganense Korshinsky	herb	Urumqi, Xinjiang, China	PS3009MT01	KM892719	KM892728	KM892580	KM892614
Lepidium virginicum Linnaeus	herb	Muxuyuan Street, Nanjing, Jiangsu, China	PS3027MT01	KM892705	KM892767	KM892552	KM892658
Lobularia maritima (Linnaeus) Desvaux	ornamental plant	Urumqi Botanical Garden, Xinjiang, China	PS3010MT01	KM892720	KM892729	KM892581	KM892615

Continued on next page

Table 1. Continued.

Species name	Uses	Collection sites	Voucher number	GenBank accession No.			
				*rbc*L	*mat*K	*trn*H-*psb*A	ITS
Malcolmia africana (Linnaeus) R. Brown	forage grass	Wild Walnut Valley, Gongliu, Xinjiang, China	PS3011MT01	KM892721	KM892730	KM892582	KM892616
Malcolmia africana (Linnaeus) R. Brown	forage grass	Nanshan, Urumqi, Xinjiang, China	PS3011MT02	KM892722	KM892731	KM892583	KM892617
Malcolmia africana (Linnaeus) R. Brown	forage grass	Heka, Xinghai, Qinghai, China	PS3011MT03	KM892723	KM892732	KM892584	KM892618
Nasturtium officinale R. Brown	vegetable	China Germplasm Bank of Wild Species, China	PS3012MT01	KM892724	KM892733	KM892585	KM892619
Neotorularia humilis (C. A. Meyer) Hedge & J. Léonard	herb	Qinghai Lake, Qinghai, China	PS3018MT01	KM892673	KM892740	KM892596	KM892626
Orychophragmus violaceus (Linnaeus) O. E. Schulz	vegetable; ornamental plant	Zhangzhu, Yixing, Jiangsu, China	PS3013MT01	KM892667	KM892734	KM892586	KM892620
Orychophragmus violaceus (Linnaeus) O. E. Schulz	vegetable; ornamental plant	Hangzhou Botanical Garden, Zhejiang, China	PS3013MT02	KM892691	KM892754	KM892587	KM892644
Orychophragmus violaceus (Linnaeus) O. E. Schulz	vegetable; ornamental plant	Nanjing Botanical Garden, Jiangsu, China	PS3013MT03	KM892702	KM892764	KM892588	KM892655
Raphanus sativus Linnaeus	vegetable; oil plant	Qinghai Lake, Qinghai, China	PS3014MT01	KM892668	KM892735	KM892589	KM892621
Rorippa dubia (Persoon) H. Hara	herb	Nanjing Botanical Garden, Jiangsu, China	PS3016MT01	KM892671	KM892738	KM892593	KM892624
Rorippa dubia (Persoon) H. Hara	herb	Hangzhou Botanical Garden, Zhejiang, China	PS3016MT02	KM892689	-	KM892594	KM892642
Rorippa indica (Linnæus) Hiern	herb	Zhangzhu, Yixing, Jiangsu, China	PS3015MT01	KM892669	KM892736	KM892590	KM892622
Rorippa indica (Linnæus) Hiern	herb	Langya Mountain, Chuzhou, Anhui, China	PS3015MT02	KM892670	KM892737	KM892591	KM892623
Rorippa indica (Linnæus) Hiern	herb	Biwa Lake, Nanjing, Jiangsu, China	PS3015MT03	KM892701	KM892763	KM892592	KM892654
Rorippa palustris (Linnaeus) Besser	herb	Chansi, Gongliu, Xinjiang, China	PS3017MT01	KM892672	KM892739	KM892595	KM892625
Rorippa palustris (Linnaeus) Besser	herb	Langya Mountain, Chuzhou, Anhui, China	PS3021MT01	KM892681	KM892747	KM892603	KM892634
Thlaspi arvense Linnaeus	vegetable; herb; oil plant	Tianshan Grand Canyon, Urumqi, Xinjiang, China	PS3019MT01	KM892674	KM892741	KM892597	KM892627
Thlaspi arvense Linnaeus	vegetable; herb; oil plant	Qinghai Lake, Qinghai, China	PS3019MT02	KM892675	KM892742	KM892598	KM892628
Thlaspi arvense Linnaeus	vegetable; herb; oil plant	Nanjing Botanical Garden, Jiangsu, China	PS3026MT01	KM892703	KM892765	KM892608	KM892656

[a]denotes failure of PCR amplification.

Data analysis

As the recent intensively recommended DNA barcode candidate, the internal transcribed spacer 2 (ITS2) was also adopted as the fifth region for species discrimination, and the sequences of ITS2 were retrieved according to Keller et al. (2009) and GenBank annotations.

Sequences were aligned and adjusted manually using Sequencer v.4.5 software (GeneCodes, Ann Arbor, MI, USA). The nucleotide sequence data of the four regions were deposited in the GenBank database (Table 1). All genetic distances were calculated using MEGA (5.0 Version) software. Average intra-specific distance, mean theta and coalescent depth were calculated to determine intra-specific variation and average interspecific distance. Theta prime and the minimum interspecific distance were calculated to determine interspecific divergence (Meier et al., 2008; Chen et al., 2010). The distribution of intra-specific versus interspecific variability was evaluated by assessment of the presence of DNA barcoding gaps. Moreover, BLAST 1 and the nearest distance method were used to test the power of species identification as described previously (Sun et al., 2012).

RESULTS

PCR amplification efficiency and the success rate of sequencing

The sequence information of the five DNA barcode candidates, *rbc*L, *mat*K*, trn*H-*psb*A, ITS and ITS2, is provided in Table 2. The lengths of alignable sequences ranged from 209 bp for ITS2 to 747 bp for *matK*. *rbc*L was the most conserved region (529/577 nucleotides), based on both sequence length and number of conserved sites. *trn*H-*psb*A had the greatest nucleotide variation (233/351), followed by ITS2 (104/209) and ITS (195/517), based on sequence length and number of variable sites. *trn*H-*psb*A had the richest parsimony (parsim)-informative sites (182/351), followed by ITS2 (84/209), ITS (145/517) and *mat*K (120/747), with *rbc*L being the lowest. It could be inferred that *trn*H-*psb*A, ITS and ITS2 are the best regions for use as DNA barcodes for phylogenetic reconstruction, whereas *rbc*L is the least suitable marker for Brassicaceae.

Table 2. Sequence information of five candidate barcodes.

Marker	Sequence length (bp)	Alignment length (bp)	Conserved sites (bp)	Variable sites (bp)	Parsim-informative sites (bp)
*rbc*L	587-612	577	529	48	33
*mat*K	773-882	747	566	181	120
*trn*H- *psb*A	253-447	351	109	233	182
ITS	474-625	517	287	195	145
ITS2	181-200	209	75	104	84

*rbc*L, *mat*K, *trn*H-*psb*A and ITS were all successfully amplified using one pair of universal primers per locus and were compared in the success rates of PCR amplification. As shown in Table 3, *rbc*L and ITS displayed the highest efficiency of PCR amplification, followed by *trn*H-*psb*A, with *mat*K being the lowest.

Table 3. Analysis of inter-specific divergence between species and intra-specific variation.

	*mat*K	*rbc*L	*trn*H-*psb*A	ITS	ITS2
All inter-specific distance	0.0183 ± 0.0170	0.0028 ± 0.0043	0.0388 ± 0.0524	0.0846 ± 0.0700	0.0878 ± 0.0702
Theta prime	0.0262 ± 0.0156	0.0034 ± 0.0042	0.0441 ± 0.0456	0.0730 ± 0.0429	0.0860 ± 0.0437
Minimum inter-specific distance	0.0181 ± 0.0203	0.0025 ± 0.0033	0.0245 ± 0.0385	0.0355 ± 0.0430	0.0419 ± 0.0541
All intra-specific distance	0.0135 ± 0.0120	0.0012 ± 0.0035	0.0325 ± 0.0567	0.0607 ± 0.0797	0.0706 ± 0.0972
Theta	0.0180 ± 0.0171	0.0027 ± 0.0050	0.0403 ± 0.0583	0.0718 ± 0.0557	0.0876 ± 0.0798
Coalescent depth	0.0250 ± 0.0220	0.0035 ± 0.0058	0.0595 ± 0.0776	0.1125 ± 0.0865	0.1346 ± 0.1124
Success rate of PCR amplification /%	89.66	100	98.28	100	-[a]

[a]ITS2 sequences were retrieved from ITS sequences using methods developed by Keller et al., 2009.

Intra-specific variation and inter-specific divergence

A favorable barcode should possess a high inter-specific divergence to distinguish different species (Gao et al., 2010). Six metrics were used to characterize inter- versus intra-specific variations (Lahaye et al., 2008). ITS2 and ITS exhibited significantly higher inter-specific discrimination than *rbc*L, *mat*K and *trn*H-*psb*A. The intra-specific variations were similar, with ITS2 and ITS contributing the largest, and *rbc*L the smallest variations (Table 3). ITS2 and ITS were found to have high inter-specific divergence and high intra-specific variation, which indicated that ITS2 and ITS could be proposed as the most suitable DNA barcodes to distinguish the species of economic importance in Brassicaceae.

Barcoding gap assessment

A robust DNA barcode should have separate and non-overlapping genetic variations between intra- and inter-specific samples. The distributions of intra-specific versus inter-specific divergence were examined in the seven barcodes at a scale of 0.001 distance units. Although no distinct barcoding gaps, as typical of *CO1,* were found in the distributions of all the loci, the distributions of intra-specific versus inter-specific divergence does suggest a clearly defined range, where the intra-specific variation is considerably lower than the inter-specific divergence (Figure 1). Among them, ITS revealed a relatively well separated distribution, indicating significantly higher inter-specific divergences than their corresponding intra-specific variations, whereas the other four candidate barcodes displayed a distinct overlap without gaps between intra-specific variation and inter-specific divergence.

Identification efficiency of the DNA barcodes

BLAST 1 and the nearest genetic distance were utilized to assess correct discrimination using different barcodes. The results based on BLAST 1 method indicated that ITS and *trn*H-*psb*A had the list highest identification efficiency (67.2 and 63.2%) at the species level, followed by ITS2, *mat*K and *rbc*L. At the genus level, both *mat*K and *rbc*L had the highest success rate (78.9 and 78.6%), meanwhile *trn*H-*psb*A, ITS and ITS2 also performed well with 76.4, 73.2 and 64.3% successful identification rates, respectively. Similar results could be obtained by the nearest genetic distance method, while identification efficiency by the nearest genetic distance method was much lower than BLAST 1 (Table 4).

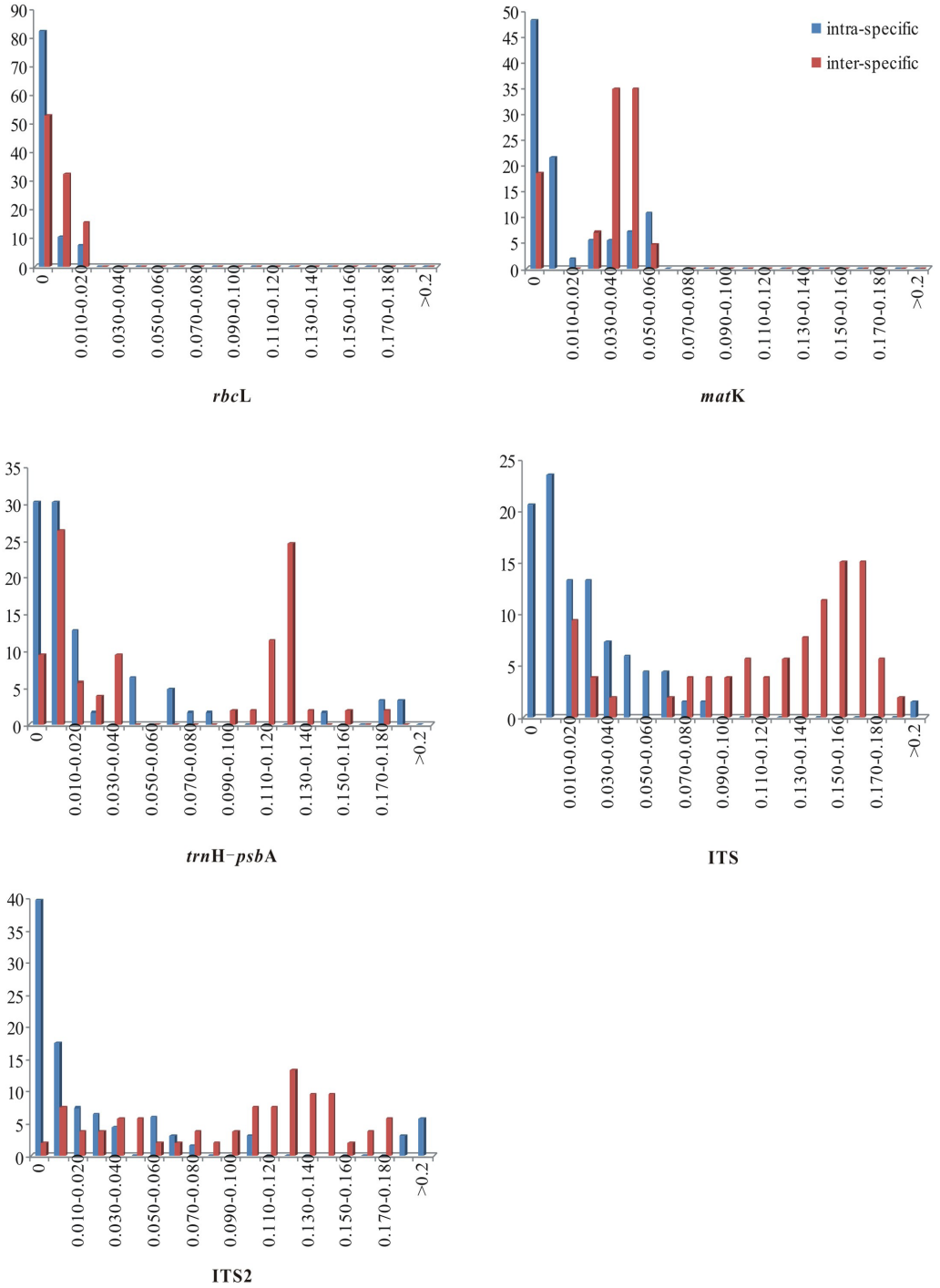

Figure 1. Relative distribution of inter-specific divergence between congenic species and intra-specific variation.

Table 4. Comparison of identification efficiency for candidate barcodes using different methods of species identification.

	Method	No. of species	No. of samples	Successful identification		Incorrect identification		Ambiguous identification	
				Species	Genus	Species	Genus	Species	Genus
*rbc*L	BLAST 1	27	58	56.9	78.6	0	0	43.1	21.4
	Distance	27	58	56.9	67.9	0	0	43.1	32.1
*mat*K	BLAST 1	25	52	61.1	78.9	0	0	38.9	21.1
	Distance	25	52	61.1	71.2	0	0	38.9	28.8
*trn*H-*psb*A	BLAST 1	27	57	63.2	76.4	0	0	36.8	23.6
	Distance	27	57	56.1	65.5	0	0	43.9	34.5
ITS	BLAST 1	27	58	67.2	73.2	0	0	32.8	26.8
	Distance	27	58	60.3	67.9	0	0	39.7	32.1
ITS2	BLAST 1	27	58	60.3	64.3	0	0	39.7	35.7
	Distance	27	58	53.4	62.5	0	0	46.6	37.5

DISCUSSION

Assessment of the applicability of the candidate barcodes

Several DNA regions, the majority taken from the plastid genome, have been tested for universality and discriminatory power in plants (Kress et al., 2005; Hollingsworth et al., 2009). The two-marker combination of *rbc*L + *mat*K was proposed as the core barcode for land plants in 2009 (CBOL Plant Working Group, 2009). The Third International Barcoding of Life Conference in Mexico City suggested that a third chloroplast DNA region (cpDNA) (*trn*H-*psb*A), and the nuclear ribosomal internal transcribed spacer (ITS) regions, should be treated as complementary loci. Despite the fact that cpDNA regions have been proposed repeatedly for plant barcoding, our results suggest these cpDNA regions cannot establish genetic delimitations between closely related species. This may be attributable to the maternal inheritance of cpDNA in most angiosperms. Thus, plastid variants are only dispersed by seed and cannot travel as far as nuclear alleles, which are dispersed by both pollen and seed (Petit et al., 2005). The limited dispersal of the plastid plant barcodes consequently have a built-in limitation to tracking species boundaries in some cases, which may provide a satisfactory explanation for the low discrimination power of plastid plant barcodes.

As summarized by Hollingsworth (2011), hybridization or polyploid speciation can lead to incongruence between barcode sequences and taxon concepts. Past hybridization or allopolyploidization can lead to shared haplotypes among species (Fazekas et al., 2009). As many as three polyploidization events have occurred in Brassicaceae, with the last one pinpointed asspecific for “core Brassicaceae” (Schranz and Mitchell-Olds, 2006). Consequently, new polyploid species could be generated by autopolyploidy or through inter-species hybridization; for example, *Brassica napus* (N = 19) was formed by hybridization of *B. rapa* L. (N = 10) and *B. oleracea* (N = 9). Taxonomic treatment of polyploid derivatives and their respective progenitors is problematic. Moreover, in some cases, the taxonomic groups were poorly defined according to limited morphological characters under convergent evolution, which makes Brassicaceae barcoding an even greater necessity. An ideal DNA barcode should be universal, reliable, cost-effective and show good discriminatory power (CBOL Plant Working Group, 2009). Despite the highest list species-level identification efficiency (67.2%) of ITS, in this study, none of the five DNA barcode candidates met all these criteria.

ITS and ITS2, which is better as an ideal DNA barcode?

The ITS or its part sequence, ITS2, have been suggested repeatedly as barcodes for plants (Kress et al., 2005; Chen et al., 2010). The ITS generally has a greater discriminatory power over plastid regions at low taxonomic levels; however, three primary concerns have prevented it from being a core plant barcode. Key concerns regarding ITS are 1) incomplete concerted evolution can lead to divergent paralogous copies within individuals, 2) fungal contamination, and 3) difficulties in amplifying and sequencing this region from diverse sample sets. In our analysis, direct sequencing of single-copy ITS sequences was successful in all sampled species (representatives from all three lineages in Brassicaceae), and no fungal contamination was detected. It seems the extent of problems concerning ITS as a core plant DNA barcode is not as pervasive, at least in Brassicaceae, as previously estimated. ITS2 was proposed as an alternative plant barcode with the advantage of easily amplifying and sequencing (Chen et al., 2010; Yao et al., 2010), and it has proven useful in several studies (Gao et al., 2010). Our study revealed the discriminatory power of ITS2 is generally 7% lower than ITS. Previous studies have suggested that ITS1 and ITS exhibit higher inter-specific divergence relative to ITS2 (Kress et al., 2005). Thus, the use of ITS2 involves a trade-off between using a shorter region of ITS to make recovery and sequencing easier, while sacrificing the number of available characters. Moreover, the preclusion of ITS from being a core plant barcode needs rigorous consideration instead of few formal empirical estimates.

As suggested by Chase and Fay (2009), nuclear genes can provide more information than organelle DNA, which is inherited from only one parent. Multiple attempts have been made to shed light on the way to the ideal plant barcode. Five nuclear low-copy loci (*CHS*, *DET1*, *COP1*, *PGIC1*, and *RPS2*) were investigated to discriminate two species of *Pugionium* (Brassicaceae), while only one locus (*DET1*) related to flowering regulations was able to delimitate species (Wang et al., 2011). It is proposed that genes related to species isolation (“speciation genes”) or linked genes such as mitochondrial DNA (mtDNA) for animals may be more effective in discriminating between closely related species. However, until now, it has been difficult to find such a locus that is universally linked to the speciation of the different plant groups. Recently, 59 low-copy nuclear genes were carefully selected in analysis of angiosperm phylogeny and resulted in highly supported relationships (Zeng et al., 2014), and also highlighted the feasibility of low-copy nuclear loci in the barcoding of plants.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by the Ministry of Science and Technology, China, Basic Research Project (#2013FY112600), the National Natural Science Foundation of China (#31200177; #31470448) and Funds of Jiangsu Province Key Laboratory for Plant *Ex Situ* Conservation (#QIAN201202) granted to X.Q. Sun.

REFERENCES

Al-Shehbaz IA (1984). The tribes of Cruciferae (Brassicaceae) in the southeastern United States. *J. Arnold Arbor.* 65: 343-373.

Bailey CD, Koch MA, Mayer M, Mummenhoff K, et al. (2006). Toward a global phylogeny of the Brassicaceae. *Mol. Biol. Evol.* 23: 2142-2160.

Beilstein MA, Al-Shehbaz IA and Kellogg EA (2006). Brassicaceae phylogeny and trichome evolution. *Am. J. Bot.* 93: 607-619.

Beilstein MA, Al-Shehbaz IA, Mathews S and Kellogg EA (2008). Brassicaceae phylogeny inferred from phytochrome A and ndhF sequence data: tribes and trichomes revisited. *Am. J. Bot.* 95: 1307-1327.

CBOL Plant Working Group (2009). A DNA barcode for land plants. *Proc. Natl. Acad. Sci. U.S.A.* 106: 12794-12797.

Chase MW, Cowan RS, Hollingsworth PM, van den Berg C, et al. (2007). A proposal for a standardised protocol to barcode all land plants. *Taxon* 56: 295-299.

Chase MW and Fay MF (2009). Barcoding of plants and fungi. *Science* 325: 682-683.

Chen S, Yao H, Han J, Liu C, et al. (2010). Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS ONE* 5: e8613.

China Plant BOL Group, Li DZ, Gao LM, Li HT, et al. (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. U.S.A.* 108: 19641-19646.

Couvreur TL, Franzke A, Al-Shehbaz IA, Bakker F, et al. (2010). Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). *Mol. Biol. Evol.* 1: 55-71.

Fazekas AJ, Kesanakurti PR, Burgess KS, Percy DM, et al. (2009). Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? *Mol. Ecol. Resour.* 9: 130-139.

Gao T, Yao H, Song J, Zhu Y, et al. (2010). Evaluating the feasibility of using candidate DNA barcodes in discriminating species of the large Asteraceae family. *BMC Evol. Biol.* 10: 324.

German DA, Friesen N, Neuffer B, Al-Shehbaz IA, et al. (2009). Contribution to ITS phylogeny of the Brassicaceae, with special reference to some Asian taxa. *Plant Syst. Evol.* 283: 33-56.

German DA, Grant JR, Lysák MA and Al-Shehbaz IA (2011). Molecular phylogeny and systematics of the tribe Chorispora (Brassicaceae). *Plant Syst. Evol.* 294: 65-86.

Hedge IC (1976). A systematic and geographical survey of the Old World Cruciferae. In: The biology and chemistry of the Cruciferae (Vaughan JG, MacLeod AJ and Jones BMG, eds.). Academic Press, London, 1-145.

Hollingsworth ML, Andra Clark A, Forrest LL, Richardson J, et al. (2009). Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Mol. Ecol. Resour.* 9: 439-457.

Hollingsworth PM, Graham SW and Little DP (2011). Choosing and using a plant DNA barcode. *PLoS ONE* 6: e19254.

Janchen E (1942). Das System der Cruciferen. *Oesterreichische Botanische Zeitschrift* 91:1-28.

Keller A, Schleicher T, Schultz J, Müller T, et al. (2009). 5.8S-28S rRNA interaction and HMM-based ITS2 annotation. *Gene* 430: 50-57.

Khosravi AR, Mohsenzadeh S and Mummenhoff K (2009). Phylogenetic relationships of Old World Brassicaceae from Iran based on nuclear ribosomal DNA sequences. *Biochem. Syst. Ecol.* 37: 106-115.

Koch MA, Dobes C, Kiefer C, Schmickl R, et al. (2007). Supernetwork identifies multiple events of plastid trnF(GAA) pseudogene evolution in the Brassicaceae. *Mol. Biol. Evol.* 24: 63-73.

Kress WJ and Erickson DL (2008). DNA barcodes: genes, genomics, and bioinformatics. *Proc. Natl. Acad. Sci. U.S.A.* 105: 2761-2762.

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, et al. (2005). Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. U.S.A.* 102: 8369-8374.

Lahaye R, van der Bank M, Bogarin D, Warner J, et al. (2008). DNA barcoding the floras of biodiversity hotspots. *Proc. Natl. Acad. Sci. U.S.A.* 105: 2923-2928.

Meier R, Zhang GY and Ali F (2008). The use of mean instead of smallest interspecific distances exaggerates the size of the "barcoding gap" and leads to misidentification. *Syst. Biol.* 57: 809-813.

Paterson AH, Brubaker CL and Wendel JF (1993). A rapid method for extraction of cotton (Gossypium spp) genomic DNA suitable for RFLP of PCR analysis. *Plant Mol. Biol. Rep.* 11: 122-127.

Petit RJ, Duminil J, Fineschi S, Hampe A, et al. (2005). Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Mol. Ecol.*14: 689-701.

Pharmacopoeia Commission of People's Republic of China. (2010) Pharmacopoeia of People's Republic of China. Part 1. Chemical Industry Press, Beijing, 191: 255-314.

Ran JH, Wang PP, Zhao HJ and Wang XQ (2010). A test of seven candidate barcode regions from the plastome in Picea (Pinaceae). *J. Integr. Plant Biol.* 52: 1109-1126.

Sang T, Crawford DJ and Stuessy TF (1997). Chloroplast DNA phylogeny, reticulate evolution, and biogeography of Paeonia (Paeoniaceae). *Am. J. Bot.* 84: 1120.

Schranz ME and Mitchell-Olds T (2006). Independent ancient polyploidy events in the sister families Brassicaceae and Cleomaceae. *Plant Cell* 18: 1152-1165.

Schulz OE (1936). Cruciferae. In: Die Natu¨rlichen Pflanzenfamilien, vol 17B, 2nd edn. (Engler A, Harms H, eds.) Verlag von Wilhelm Engelmann, Leipzig, 227-658.

Sun XQ, Zhu YJ, Guo JL, Peng B, et al. (2012). DNA barcoding the Dioscorea in China, a vital group in the evolution of monocotyledon: use of matK gene for species discrimination. *PLoS One* 7: e32057.

Wang Q, Yu QS and Liu JQ (2011). Are nuclear loci ideal for barcoding plants? A case study of genetic delimitation of two sister species using multiple loci and multiple intraspecific individuals. *J. Syst. Evol.* 49: 182-188.

Warwick SI and Hall JC (2009). Phylogeny of Brassica and wild relatives. In: Biology and breeding of crucifers (Gupta S, eds.). CRC Press, New York, 19-36.

Warwick SI, Mummenhoff K, Sauder CA, Koch MA, et al. (2010). Closing the gaps: phylogenetic relationships in the Brassicaceae based on DNA sequence data of nuclear ribosomal ITS region. *Plant Syst. Evol.* 285: 209-232.

White TJ, Bruns TD, Lee SB and Taylor JW (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: PCR protocols: a guide to methods and applications (Innis MA, Gelfand DH, Sninsky JJ and White TJ, eds.). Academic Press, London, 315-322.

Yao H, Song J, Liu C, Luo K, et al. (2010). Use of ITS2 region as the universal DNA barcode for plants and animals. *PLoS One* 5: e13102.

Zeng LP, Zhang Q, Sun RR, Kong HZ, et al. (2014). Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. *Nat. Commun.* 5: 4956.