

ORIGINAL RESEARCH

Development of high-resolution DNA barcodes for *Dioscorea* species discrimination and phylogenetic analysis

Wei Xia¹  | Bo Zhang¹ | Dan Xing¹ | Ying Li¹ | Wenqiang Wu¹ | Yong Xiao² | Jinhua Sun³ | Yajing Dou¹ | Wenqi Tang¹ | Jinlan Zhang¹ | Xiaolong Huang¹ | Yun Xu¹ | Jun Xie¹ | Jihua Wang⁴ | Dongyi Huang¹

¹Hainan Key Laboratory for Sustainable Utilization of Tropical Bioresources, College of Tropical Crops, Hainan University, Haikou, China

²Coconut Research Institute, CATAS, Wenchang, China

³Environment and Plant Protection Institute, CATAS, Haikou, China

⁴Guangdong Key Laboratory for Crops Genetic Improvement, Crops Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou, China

Correspondence

Wei Xia and Dongyi Huang, Hainan Key Laboratory for Sustainable Utilization of Tropical Bioresources, College of Tropical Crops, Hainan University, Haikou 570228, China. Emails: saizxiaawei@hainanu.edu.cn (W. X.); hdongyi@hainanu.edu.cn (D. H.)

Jihua Wang, Guangdong Key Laboratory for Crops Genetic Improvement, Crops Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou 510640, China. Email: wangjihua@daas.cn (J. W.)

Funding information

The Natural Science Foundation of China, Grant/Award Number: 31760406; The Opening Project Fund of Key Laboratory of Rubber Biology and Genetic Resource Utilization, Ministry of Agriculture/State Key Laboratory Breeding Base of Cultivation & Physiology for Tropical Crops/Danzhou Investigation & Experiment Station of Tropical Crops, Ministry of Agriculture, Grant/Award Number: RRI-KLOF201601; The Crop Science Postgraduate Innovation Project of Hainan University Tropical Agriculture and Forestry College, Grant/Award Number: ZWCX2018033

Abstract

The genus *Dioscorea* is widely distributed in tropical and subtropical regions, and is economically important in terms of food supply and pharmaceutical applications. However, DNA barcodes are relatively unsuccessful in discriminating between *Dioscorea* species, with the highest discrimination rate (23.26%) derived from *matK* sequences. In this study, we compared genic and intergenic regions of three *Dioscorea* chloroplast genomes and found that the density of SNPs and indels in intergenic sites was about twice and seven times higher than that of SNPs and indels in the genic regions, respectively. A total of 52 primer pairs covering highly variable regions were designed and seven pairs of primers had 80%–100% PCR success rate. PCR amplicons of 73 *Dioscorea* individuals and assembled sequences of 47 *Dioscorea* SRAs were used for estimating intraspecific and interspecific divergence for the seven loci: The *rpoB-trnC* locus had the highest interspecific divergence. Automatic barcoding gap discovery (ABGD), Poisson tree processes (PTP), and generalized mixed Yule coalescence (GMYC) analysis were applied for species delimitation based on the seven loci and successfully identified the majority of species, except for species in the *Enantiophyllum* section. Phylogenetic analysis of 51 *Dioscorea* individuals (28 species) showed that most individuals belonging to the same species tended to cluster in the same group. Our results suggest that the variable loci derived from comparative analysis of plastid genome sequences could be good DNA barcode candidates for taxonomic analysis and species delimitation.

KEYWORDS

chloroplast genome, *Dioscorea*, DNA barcode, intergenic variation

Wei Xia and Bo Zhang contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

The genus *Dioscorea* (family Dioscoreaceae) is comprised of approximately 630 species which are distributed across Southeast Asia, Africa, Central America, South America, and other tropical and subtropical regions. This genus is economically important for their tubers, which provide starch as a dietary staple as well as cortisone and other steroid hormones, such as dioscin (Aumsuwan et al., 2016; Cho et al., 2013; Jeon et al., 2006). However, *Dioscorea* species are hard to identify due to high morphological diversity, dioecy, small flowers, and morphological similarities between various species in this genus (Raman et al., 2014; Wilkin et al., 2005). Distinguishing *Dioscorea* species based on morphological traits is unreliable, while using DNA barcodes (*matK*, *rbcl*, *psbA-trnH*, *trnL-F*) for *Dioscorea* species, identification has previously showed relatively low discrimination success, with the highest rate of 23.26% derived from use of the *matK* sequences (Gao et al., 2008; Mukherjee & Bhat, 2013; Sun et al., 2012). Currently, chloroplast genome sequences of four species in the *Dioscorea* genus are available (Mariac et al., 2014; Wu et al., 2016; Zhou, Chen, Hua, & Wang, 2016), and thorough sequence comparison between these genomes could perhaps provide candidate regions for developing useful barcodes.

In the past decade, seven plastid DNA regions (*atpF-atpH* spacer, *matK* gene, *rbcl* gene, *rpoB* gene, *rpoC1* gene, *psbK-psbI* spacer, and *trnH-psbA* spacer) and 2-locus combinations were frequently used to distinguish the land plants (Hollingsworth, Graham, & Little, 2011). To date, these DNA barcodes along with other barcodes such

as *ycf5*, *psbK-I*, *psbM*, *trnD*, and *rps16* are still widely used for the identification of varieties and analysis of the provenances of varieties (Lee, Wang, Yen, & Chang, 2017; Techen, Parveen, Pan, & Khan, 2014). Moreover, new barcodes have been developed based on increasingly available sequence data and on deep mining for highly variable regions. Dong et al. (2015) were analyzed available plastid genomes and designed suitable primers for the most variable regions, and finally found that *ycf1b* generally performed better than any of the *matK*, *rbcl*, and *trnH-psbA* barcodes. Among 18 *Oryza* chloroplast genomes, five variable regions (*rps16-trnQ*, *trnTEYD*, *psbE-petL*, *rpoC2*, and *rbcl-accD*) were analyzed for species discrimination (Song et al., 2017). However, systematic comparisons for plastid genome sequences have not been conducted between *Dioscorea* species and would provide useful information for identifying better-performing DNA barcodes for *Dioscorea* species.

In this study, we downloaded plastid genome sequences for three *Dioscorea* species—*D. rotundata*, *D. elephantipes*, and *D. zingiberensis*—and made a comprehensive comparison of the genic and intergenic regions to characterize conserved regions and variable regions. Top variable regions were selected and covered by 52 pairs of primers, and we tested primer universality in 10 *Dioscorea* species. Moreover, 47 sequence read archives (SRAs) for 18 *Dioscorea* species were also downloaded and assembled for the corresponding plastid sequences for our selected variable regions. This study aimed to develop efficient DNA barcodes for *Dioscorea* species discrimination and to provide useful information for further DNA barcode development.

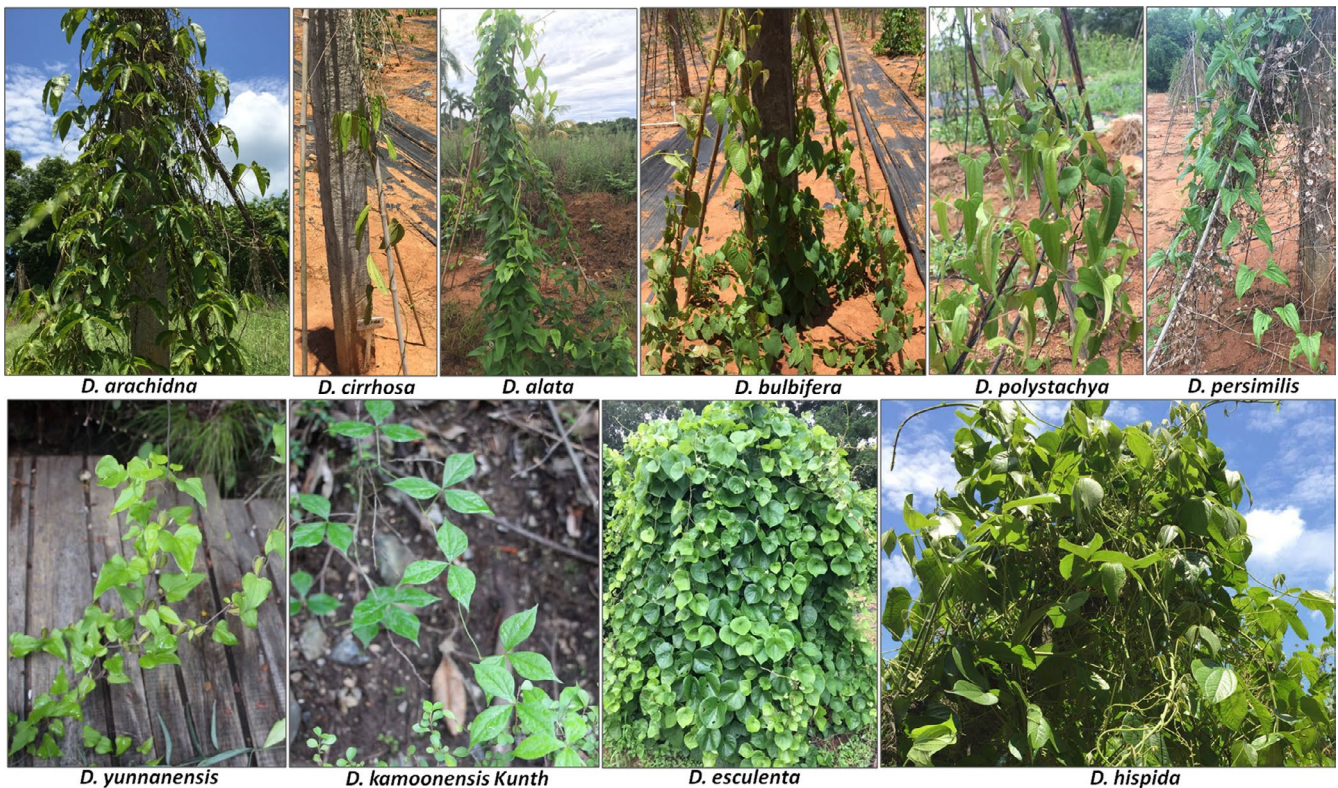


FIGURE 1 Representative plant individuals of the 10 *Dioscorea* species used in this study

TABLE 1 Genic and intergenic variation between *Dioscorea elephantipes* (Del), *D. rotundata* (Dro), and *D. zingiberensis* (Dzi) based on nucleotide BLAST

Compared species	Genic region				Intergenic region				Length ratio (Intergenic/genic)		
	Aligned length (bp)	Similarity	Variable sites		Aligned length (bp)	Similarity	Variable sites		Total %	SNP	Indel
			SNP (bp)	Indel (bp)			SNP (bp)	Indel (bp)			
Del	109,121	96%–100%	890	94	44,717	79%–100%	901	287	2.66	2.5	7.5
Dro	109,295	87%–100%	2,387	175	40,055	77%–100%	1,802	504	5.76	2.1	7.9
Dzi	109,989	87%–100%	2,049	183	44,614	76%–100%	1,693	495	4.90	2.0	6.7

2 | MATERIALS AND METHODS

2.1 | Plant materials

Plant samples were collected from different provinces in China and were kept in the *Dioscorea* germplasm nursery of Danzhou, Hainan, China. A total of 74 individuals belonging to 10 species and representing a high number of economically useful plants were used for this study. The individual images of the 10 *Dioscorea* species (*D. alata*, *D. polystachya*, *D. esculenta*, *D. persimilis*, *D. bulbifera*, *D. cirrhosa*, *D. hispida*, *D. arachidna*, *D. kamoensis* Kunth, and *D. yunnanensis*) are shown in Figure 1. Detailed information for these analyzed *Dioscorea* species is listed in Table S1. Fresh leaves were used to extract DNA.

2.2 | Sequence analysis for plastid genome of *Dioscorea* species and primer design

The chloroplast genome sequences for *D. rotundata*, *D. elephantipes*, and *D. zingiberensis* were downloaded from the National Center for Biotechnology Information (NCBI; <https://www.ncbi.nlm.nih.gov/>). We used BLAST to align the genic and intergenic regions of the three plastid sequences. Divergent hot regions were identified, and a set of primers were designed to cover these plastid regions (Table S2). The primer design was using the software—Primer Premier 5.

A total of 47 SRAs for 18 *Dioscorea* species (*D. baya*, *D. burkilliana*, *D. cayennensis*, *D. dumetorum*, *D. hirtiflora*, *D. minutiflora*, *D. preussii*, *D. quartiniana*, *D. sagittifolia*, *D. sansibarensis*, *D. schimperiana*, *D. smilacifolia*, *D. togoensis*, *D. villosa*, *D. bulbifera*, *D. rotundata*, *D. abyssinica*, and *D. praehensilis*) locating in Liberia, Cameroon, Republic of the Congo, Gabon, Ethiopia, Benin, Senegal, Guinea, Ghana, Malawi, Cote d'Ivoire, Togo, USA, and Benin were downloaded from the NCBI, and detailed information for the datasets and the species were listed in Table S3. Based on the plastid genome sequences for the above three *Dioscorea* species, the mapping software bowtie2 (Langmead & Salzberg, 2012) was used to identify plastid-related sequences. The mapped reads were assembled with CAP3 (Huang & Madan, 1999). We applied the NCBI Primer-BLAST to test the efficiency of primers.

2.3 | DNA extraction, amplification, and sequencing

DNA extraction was following a cetyl trimethylammonium bromide (CTAB) protocol modified from Paterson, Brubaker, and Wendel (1993). One individual for each of the 10 *Dioscorea* species we sampled was used to select primers and test the amplification efficiency. The polymerase chain reaction (PCR) mixture contained 4 μ l diluted DNA (50 ng/ μ l), 10 μ l of 2 \times Mix (Yugong Biolab), and 1 μ l of each forward and reverse primer (10 μ M) in a final volume of 20 μ l. PCR amplification was carried out under following conditions: 5 min at 94°C, and 32 cycles of 30 s at 94°C, 30 s at 55°C, 1 min at 72°C, and a final step of 7 min at 72°C. PCR products were examined electrophoretically on 2% agarose gels. Purification and sequencing were done by Guangzhou Tianyi Huiyuan Biological Technology Company

TABLE 2 Candidate DNA barcode regions with high variations between *Dioscorea elephantipes* (Del), *D. rotundata* (Dro), and *D. zingiberensis* (Dzi)

Regions with high variation	Length	Variable sites					
		Del-Dro		Dzi-Dro		Del-Dzi	
		Numbers	%	Numbers	%	Numbers	%
Genic							
rbcL	1,434	23	1.60	37	2.6	25	1.74
ndhF	2,250	29	1.29	116	5.2	102	4.60
matK	1,560	30	1.92	76	4.9	69	4.43
atpF	1,576	30	1.94	91	5.8	76	4.82
rpoB	3,213	30	0.93	76	2.4	67	2.09
rpl16	1,463	31	2.15	89	6.1	82	5.60
rpoC1	2,089	37	1.32	87	3.1	86	3.06
clpP	2,025	39	1.96	92	4.6	86	4.25
ndhA	2,190	40	1.84	101	4.6	78	3.56
rpoC2	4,153	49	1.18	126	3.0	109	2.63
ycf1	5,629	118	2.11	316	5.6	288	5.12
Intergenic							
trnH-psbA	295	3	1.02	13	4.6	15	5.38
psbK-psbI	417	9	2.19	30	7.2	31	7.43
atpF-atpH	261	9	3.45	7	6.1	10	5.95
psaA-ycf3	638	19	3.11	38	6.0	33	5.33
clpP-psbB	508	21	4.27	24	4.7	27	5.33
trnE-trnT	838	23	2.96	54	6.4	50	6.01
psbE-petL	685	28	4.09	52	8.3	43	6.95
trnT-psbD	1,114	34	3.85	48	7.4	77	6.91
trnL-rpl32	915	35	4.00	116	12.7	113	12.39
trnT-trnL	927	39	4.68	64	6.9	57	6.22
trnC-petN	1,056	40	4.01	71	6.8	66	6.25
ycf4-cemA	695	41	6.56	34	6.8	62	8.92
rpoB-trnC	1,378	57	4.47	116	8.4	81	5.92
ndhC-trnV	1,019	62	6.14	90	8.8	56	5.58
trnK-trnQ	1,653	68	4.11	77	7.4	77	7.35
trnS-trnG	1,091	76	7.20	147	13.8	125	11.46
rpl32-ndhF	790	96	12.15	16	6.1	31	6.97

Note: The classic DNA barcodes are in bold font.

using the amplification primers. The nucleotide sequence data were deposited in the European nucleotide Archive database (Table S1).

2.4 | Sequence alignment and data analysis for DNA barcode

All sequences were aligned and adjusted manually by MEGA 7.0 (Kumar, Stecher, & Tamura, 2016), and all variable sites for these sequences obtained by sequencing in this study were rechecked on the original trace files for final confirmation. Both concatenated dataset and single locus sequences were applied for phylogenetic tree construction. The phylogenetic trees were constructed using maximum

likelihood (ML), and node support was assessed by a bootstrap test (1,000 pseudoreplicates of run with K2P distance as a model of substitution). All genetic distances were calculated in MEGA 7.0. Average intraspecific distance and interspecific distance were calculated to determine interspecific and intraspecific divergence, respectively. Wilcoxon signed-rank tests for the interspecific divergences among the selected barcode loci were performed by SPSS (IBM Corp, 2017).

We used three independent species delimitation approaches, automatic barcoding gap discovery (ABGD, Puillandre, Lambert, Brouillet, & Achaz, 2012), Poisson tree processes (PTP, Zhang, Kapli, Pavlidis, & Stamatakis, 2013), and generalized mixed Yule coalescence (GMYC, Suchard et al., 2018), to determine putative molecular species

TABLE 3 Variability of the seven new markers and the DNA barcodes in *Dioscorea* species

Markers	Length (bp)		Conserved sites (bp)	Variable sites (bp)			Amplification efficiency ^a	ePCR efficiency ^b
	Sequence	Aligned		Indels	SNPs	Total		
<i>atpF</i>	478–636	678	408	223	47	270	9/10	14/18
<i>rpoB-trnC</i>	567–664	710	459	112	139	251	10/10	15/18
<i>trnD-trnT</i>	860–873	895	778	60	57	117	8/10	17/18
<i>psaA-ycf3</i>	774–927	968	664	218	86	304	10/10	18/18
<i>ycf4-cemA</i>	293–916	946	260	660	26	686	10/10	18/18
<i>clpP-psbB</i>	843–880	906	753	96	57	153	10/10	17/18
<i>rpl14-rpl16</i>	858–889	912	677	93	142	235	10/10	16/18

^aPCR amplification conducted in 10 *Dioscorea* species (*D. alata*, *D. polystachya*, *D. esculenta*, *D. persimilis*, *D. bulbifera*, *D. cirrhosa*, *D. hispida*, *D. arachidna*, *D. kamoensis* Kunth, and *D. yunnanensis*).

^bPrimer-BLAST analysis conducted in assembled sequences for 18 *Dioscorea* species (*D. baya*, *D. burkilliana*, *D. cayennensis*, *D. dumetorum*, *D. hirtiflora*, *D. minutiflora*, *D. preussii*, *D. quartiniana*, *D. sagittifolia*, *D. sansibarensis*, *D. schimperiana*, *D. smilacifolia*, *D. togoensis*, *D. villosa*, *D. bulbifera*, *D. rotundata*, *D. abyssinica*, and *D. praehensilis*).

in our dataset and evaluate the performance of the selected barcode loci. Matrices of pairwise genetic distances using the p-distance, the Kimura 2-parameter (K2P), and the Jukes-Cantor (JC69) models were computed by MEGA 7.0 and used as input files on the ABGD webpage (<http://www.wabi.snv.jussieu.fr/public/abgd/abgdweb.html>). We set parameters as follows: $P_{\min} = 0.001$, $P_{\max} = 0.01$, Steps = 50, $X = 1.0$, and Nb bins = 20. We performed PTP analyses on the bPTP web server (<http://species.h-its.org/ptp/>) with the RAXML topology (Kozlov, Darriba, Flouri, Morel, & Stamatakis, 2019) and used the 50% majority-rule consensus topology resulting from the BI analysis as output files. We ran PTP analyses for 400,000 MCMC generations, set the thinning value = 100 and burn-in = 0.25. We visually confirmed the convergence of the MCMC chain as recommended by Zhang et al. (2013). We used ultrametric trees generated with BEAST 1.10.4 for GMYC analyses (Suchard et al., 2018). The ultrametric trees were constructed as follows: Coalescent tree prior and the heterogeneity of the mutation rate across lineages were set under an uncorrelated log-normal relaxed clock. The analysis was run for 100 million generations with a sampling frequency of 10,000. After checking adequate mixing and convergence of all runs with Tracer 1.7.1 (Rambaut, Drummond, Xie, Baele, & Suchard, 2018), the first 25% trees were discarded as burn-in. The maximum clade credibility tree was computed using TreeAnnotator 1.10.4 (Suchard et al., 2018). The resulting ultrametric tree was imported into R 3.6.0 (R Core Team, 2018), and GMYC analyses were run using the Splits (Ezard, Fujisawa, & Barraclough, 2009) and Ape (Paradis, Claude, & Strimmer, 2004) libraries.

3 | RESULTS

3.1 | Chloroplast genome sequence divergence in three *Dioscorea* species

To identify suitable sequences for species discrimination in *Dioscorea* species, chloroplast genome sequences for *D. elephantipes*, *D. rotundata*, and *D. zingiberensis* were downloaded from

the NCBI website and analyzed. The three chloroplast genomes ranged from 152,609 bp (*D. elephantipes*) to 155,406 bp (*D. rotundata*), consisting of a pair of inverted repeats (25,476–25,509 bp) separated by the long single copy section (80,777–85,601 bp) and short single copy section (18,814–19,038 bp) regions (Table S4). All three chloroplast genomes had the same gene number (140), including 94 protein-coding genes, 38 tRNA genes, and eight rRNA genes.

Alignments for 140 genic sequences from the three chloroplast genomes showed that the three *Dioscorea* species have 109,121–109,989 aligned genic sequence and shared high sequence similarities (87%–100%) (Table 1). Moreover, genic sequences between *D. elephantipes* and *D. rotundata* showed higher similarity (96%–100%), with fewer nucleotide polymorphisms (SNPs) and indels than the other two interspecies comparisons. More than 20/1,000 nucleotide variations were detected between *D. zingiberensis* and either of the other two species, while <10/1,000 nucleotide variations were detected between *D. elephantipes* and *D. rotundata* (Table 1). The total length of indel sites was 10 times less than that of the SNPs, with 94–183 bp of indels identified between the three species. The intergenic sequence alignments between the three *Dioscorea* species indicated lower sequence similarity for the most variable regions (76%–79%) than the genic sequence alignments (87%–96%). The densities of SNPs and indels in intergenic sites were about twice and seven times higher, respectively, than the densities of SNPs and indels in the genic regions. About 49–57 per 1,000 nucleotides variations were detected between *D. zingiberensis* and either of the other two species, while <26 per 1,000 nucleotide variations were detected between *D. elephantipes* and *D. rotundata*.

Top variable genic and intergenic regions were selected mainly based on the number of variable sites and listed in Table 2 with seven widely used DNA barcode markers. The seven DNA barcodes including four genic regions (*matK*, *rbcL*, *rpoB*, and *rpoC1* genes) and three intergenic regions (*atpF-atpH* spacer, *psbK-psbI* spacer, and *trnH-psbA* spacer) showed a lower frequency of variation between the

TABLE 4 The pairwise intraspecific and interspecific distances for seven variable loci in *Dioscorea* species and species delimitation analysis through automatic barcoding gap discovery (ABGD), Poisson tree processes (PTP), and generalized mixed Yule coalescence (GMYC) analysis

Loci	Intraspecific distances			Interspecific distances			No. of species ^a	ABGD analysis ^b	PTP analysis ^c	GMYC analysis ^d
	Minimum	Maximum	Mean	Minimum	Maximum	Mean				
atpF	0.000	0.002	0.001	0.000	0.093	0.027	25	22 (0.1%)	13 (0.3%)	16 (0.05)
rpoB-trnC	0.000	0.013	0.002	0.002	0.104	0.037	21	27 (0.1%)	17 (0.3%)	27 (0.001)
trnD-trnT	0.000	0.014	0.001	0.000	0.025	0.013	25	22 (0.1%)	20 (0.2%)	22 (9.2E-06)
psaA-ycf3	0.000	0.012	0.001	0.000	0.063	0.018	29	39 (0.1%)	20 (0.3%)	45 (0.01)
ycf4-cemA	0.000	0.012	0.001	0.000	0.080	0.021	30	31 (0.1%)	15 (0.6%)	29 (0.001)
clpP-psbB	0.000	0.005	0.001	0.000	0.041	0.013	28	26 (0.1%)	2 (2.5%)	26 (0.01)
rp14-rp16	0.000	0.044	0.004	0.000	0.110	0.024	28	32 (0.1%)	8 (1.7%)	26 (0.002)

^aNumber of *Dioscorea* species with full-length corresponding sequences used in the analysis.

^bThe values outside the brackets represent the numbers of estimated species; barcode gap values are inside the brackets. Two barcode gap thresholds are displayed.

^cNumber of estimated species with support values higher than 0.5.

^dThe values outside the brackets represent the number of estimated species, and the *p* value is inside the brackets.

three *Dioscorea* species. The number of variable sites for *rbcl* (23), *matK* (30), and *rpoB* (30) genes was less or close to those of the seven other variant genic regions (30–118), and the frequency of variants in the *rpoB* (0.93%) and *rpoC1* (1.32%) genes was close to the average variation frequency (0.93%) between *D. elephantipes* and *D. rotundata* (Table 2). The frequency of the variable sites for *rbcl*, *rpoB*, and *rpoC1* was lower than that of most other variable genic regions between *D. zingiberensis* and the two other species. However, the *matK* gene showed similar variant frequency when compared to other variable genic regions. The *atpF-atpH* spacer, *psbK-psbI* spacer, and *trnH-psbA* showed less variable sites and lower variable frequency than most other variable intergenic regions (Table 2).

3.2 | Seven highly variable regions for candidate DNA barcodes

To develop DNA barcode for *Dioscorea* species discrimination, 52 primer pairs covering highly variable regions were designed (Table S2), including the top variable regions in Table 2. One individual from each of the 10 *Dioscorea* species (*D. alata*, *D. polystachya*, *D. esculenta*, *D. persimilis*, *D. bulbifera*, *D. cirrhosa*, *D. hispida*, *D. arachidna*, *D. kamoensis* Kunth, and *D. yunnanensis*), which belong to five *Dioscorea* sections—Enantiophyllum Uline (4), Combilium Prain et Burkill (1), Opsophyton Mine (1), Lasiophyton Uline (3), and Shannicorea Prain et Burkill (1) (Table S1), was used for selecting primers with high universality. The PCR results showed that 11/52 pairs of primers had positive amplification products. Moreover, seven pairs of primers covering one genic sequence—*atpF*, four intergenic sequences—*rpoB-trnC*, *ycf4-cemA*, *clpP-psbB*, and *rp14-rp16*, and two containing both genic and intergenic sequences—*trnD-trnT* and *psaA-ycf3* had 80%–100% PCR success rate, while the other four primer pairs primer successfully amplified only in one to two species.

We also tested the primers for the assembled sequences from 47 individuals, which belong to 18 *Dioscorea* species: *D. baya*, *D. burkiliana*, *D. cayennensis*, *D. dumetorum*, *D. hirtiflora*, *D. minutiflora*, *D. preussii*, *D. quartiniana*, *D. sagittifolia*, *D. sansibarensis*, *D. schimperiana*, *D. smilacifolia*, *D. togoensis*, *D. villosa*, *D. bulbifera*, *D. rotundata*, *D. abyssinica*, and *D. praehensilis*. These species belong to seven *Dioscorea* sections: Enantiophyllum Uline (10), Lasiophyton Uline (1), Asterotricha (2), Macrocarpaea (1), Botryosicyos (1), Macrourea (1), and Opsophyton Mine (1) (Table S3). The ePCR results showed that the seven pairs of primers have ePCR success rates as 77%–100%, which were similar to the PCR results for the 10 *Dioscorea* species (Table 3). The primers for *psaA-ycf3* and *ycf4-cemA* have the top ePCR success rate, followed by *trnD-trnT* and *clpP-psbB* with ePCR success rate as 94% (17/18). Combined with the above PCR results, *psaA-ycf3* and *ycf4-cemA* were still the top primers with 100% ePCR success rate (Table 3).

We aligned sequences from PCR amplification and the assembled sequences for the seven regions with high PCR success rate. The PCR amplicons ranged from 293 to 927 bp in size after trimming flanking sequences with low quality. A total of 29 *Dioscorea* species belonging to eleven sections were analyzed,

TABLE 5 Wilcoxon signed-rank test of interspecific divergence between the seven loci

W+	W-	Relative ranks, n, p value ^a	Result
<i>atpF</i>	<i>clpP-psbB</i>	W+ = 37,033, W- = 2,588, 300, $p \leq .000$	<i>atpF</i> > <i>clpP-psbB</i>
<i>atpF</i>	<i>rpl14-rpl16</i>	W+ = 25,879, W- = 16,899, 300, $p \leq .002$	<i>atpF</i> > <i>rpl14-rpl16</i>
<i>atpF</i>	<i>ycf4-cemA</i>	W+ = 21,888, W- = 11,265, 276, $p \leq .000$	<i>atpF</i> > <i>ycf4-cemA</i>
<i>atpF</i>	<i>rpoB-trnC</i>	W+ = 4,612.5, W- = 9,922.5, 171, $p \leq .000$	<i>atpF</i> < <i>rpoB-trnC</i>
<i>atpF</i>	<i>trnD-trnT</i>	W+ = 18,908.5, W- = 4,527.5, 231, $p \leq .000$	<i>atpF</i> > <i>trnD-trnT</i>
<i>atpF</i>	<i>psaA-ycf3</i>	W+ = 29,502.5, W- = 10,683.5, 300, $p \leq .000$	<i>atpF</i> > <i>psaA-ycf3</i>
<i>rpl14-rpl16</i>	<i>clpP-psbB</i>	W+ = 47,683.5, W- = 13,741.5, 378, $p \leq .000$	<i>rpl14-rpl16</i> > <i>clpP-psbB</i>
<i>rpl14-rpl16</i>	<i>ycf4-cemA</i>	W+ = 32,662, W- = 32,662, 378, $p = .31$	<i>rpl14-rpl16</i> = <i>ycf4-cemA</i>
<i>rpl14-rpl16</i>	<i>rpoB-trnC</i>	W+ = 5,575, W- = 12,570, 190, $p \leq .000$	<i>rpl14-rpl16</i> < <i>rpoB-trnC</i>
<i>rpl14-rpl16</i>	<i>trnD-trnT</i>	W+ = 26,095, W- = 18,158, 300, $p \leq .000$	<i>rpl14-rpl16</i> > <i>trnD-trnT</i>
<i>rpl14-rpl16</i>	<i>psaA-ycf3</i>	W+ = 24,304, W- = 27,699, 325, $p = .31$	<i>rpl14-rpl16</i> = <i>psaA-ycf3</i>
<i>rpoB-trnC</i>	<i>clpP-psbB</i>	W+ = 17,966, W- = 179, 190, $p \leq .000$	<i>rpoB-trnC</i> > <i>clpP-psbB</i>
<i>rpoB-trnC</i>	<i>ycf4-cemA</i>	W+ = 13,736, W- = 4,409, 190, $p \leq .000$	<i>rpoB-trnC</i> > <i>ycf4-cemA</i>
<i>rpoB-trnC</i>	<i>trnD-trnT</i>	W+ = 8,706, W- = 610, 136, $p \leq .000$	<i>rpoB-trnC</i> > <i>trnD-trnT</i>
<i>rpoB-trnC</i>	<i>psaA-ycf3</i>	W+ = 12,783, W- = 1,923, 171, $p \leq .000$	<i>rpoB-trnC</i> > <i>psaA-ycf3</i>
<i>ycf4-cemA</i>	<i>clpP-psbB</i>	W+ = 56,159, W- = 12,847, 378, $p \leq .000$	<i>ycf4-cemA</i> > <i>clpP-psbB</i>
<i>ycf4-cemA</i>	<i>trnD-trnT</i>	W+ = 25,155, W- = 17,916, 300, $p = .013$	<i>ycf4-cemA</i> > <i>trnD-trnT</i>
<i>ycf4-cemA</i>	<i>psaA-ycf3</i>	W+ = 25,050, W- = 25,990, 300, $p = .776$	<i>ycf4-cemA</i> = <i>psaA-ycf3</i>
<i>clpP-psbB</i>	<i>trnD-trnT</i>	W+ = 9,684, W- = 33,387, 300, $p \leq .000$	<i>clpP-psbB</i> < <i>trnD-trnT</i>
<i>clpP-psbB</i>	<i>psaA-ycf3</i>	W+ = 3,243, W- = 47,478, 325, $p \leq .000$	<i>clpP-psbB</i> < <i>psaA-ycf3</i>
<i>trnD-trnT</i>	<i>psaA-ycf3</i>	W+ = 11,536, W- = 244,779, 276, $p \leq .000$	<i>trnD-trnT</i> < <i>psaA-ycf3</i>

^aThe symbols “W+” and “W-” represent the sum of all of the positive values and the sum of all of the negative values in the signed-rank column, respectively.

including plastid sequences from *D. elephantipes* and *D. zingiberensis*. Multiple sequence alignments showed that these sequences have ample indels covering 60–660 bp and 26–142 SNPs for these amplicons between the *Dioscorea* species (Table 3 and Figure S1). The indels identified in these amplicons were consistent within species. The sequences for *atpF*, *psaA-ycf3*, and *ycf4-cemA* contained an indel site more than 100 bp and distinct sequence divergences between *Dioscorea* species were detected (Figure S1). *D. yunnanensis* (Dy1, Dy2 and Dy3) had the longest deletion: 174 bp in this indel region of *atpF* and eleven distinct types of sequences existed, while eight types of sequences were observed in the indel region of *psaA-ycf3*. However, the large indel regions of *atpF* and *psaA-ycf3* for *D. hispida* and *D. arachidna* were indistinguishable (Figure S1). A 645 bp deletion was detected for the *ycf4-cemA* region in *D. cirrhosa*, followed with a 182 bp deletion in *D. arachidna*. Eleven distinct types of sequences existed in this indel region of *ycf4-cemA*, but *D. alata* and *D. hispida* were indistinguishable. Besides, the *rpoB-trnC* (139) and *rpl14-rpl16* (142) contained more SNPs than the other five regions.

Analysis of intraspecific and interspecific distances showed that *rpl14-rpl16* locus had the highest intraspecific distance (maximum = 0.044; mean = 0.004), and the remaining loci presented an average intraspecific distances as 0.001–0.002 (Table 4). The intergenic region (*rpoB-trnC*) showed the greatest mean interspecific

divergence (0.037), followed by *atpF* (0.027) and *rpl14-rpl16* (0.024). The noncoding region (*trnD-trnT*) had the smallest average interspecific divergence (0.013). Wilcoxon signed-rank tests demonstrated that the *rpoB-trnC* had significantly higher interspecific divergence than that of other species, and the locus with the second highest interspecific divergence is *atpF*, while *rpl14-rpl16*, *ycf4-cemA*, and *psaA-ycf3* had similar interspecific divergences (Table 5).

3.3 | Applicability for species discrimination

A total of 73 individuals belonging to 10 *Dioscorea* species (Table S1), a set of 18 *Dioscorea* species with available SRAs (Table S3), and the three *Dioscorea* species with complete plastid genomes were used for estimation of species discrimination efficiency of the above loci. A total of 11 sections of *Dioscorea* species were included in this analysis, including *Enantiophyllum*, *Shannicorea*, *Asterotracha*, *Lasiophyton*, *Macrocarpaea*, *Lasiophyton*, *Testudinana*, *Combilium*, *Stenophora*, *Macrourea*, and *Botryosicyos* (Tables S1 and S3).

Measuring the intraspecific variation and interspecific divergence showed intraspecific variations for the seven loci were much lower than interspecific divergences (Figure 2). The majority of intraspecific variation for *atpF* (92%), *psaA-ycf3* (92%), and *ycf4-cemA* (98%) were in sections 0–0.002, while *clpP-psbB* (77%), *trnD-trnT* (76%), *rpl14-rpl16* (64%), and *rpoB-trnC* (63%) had

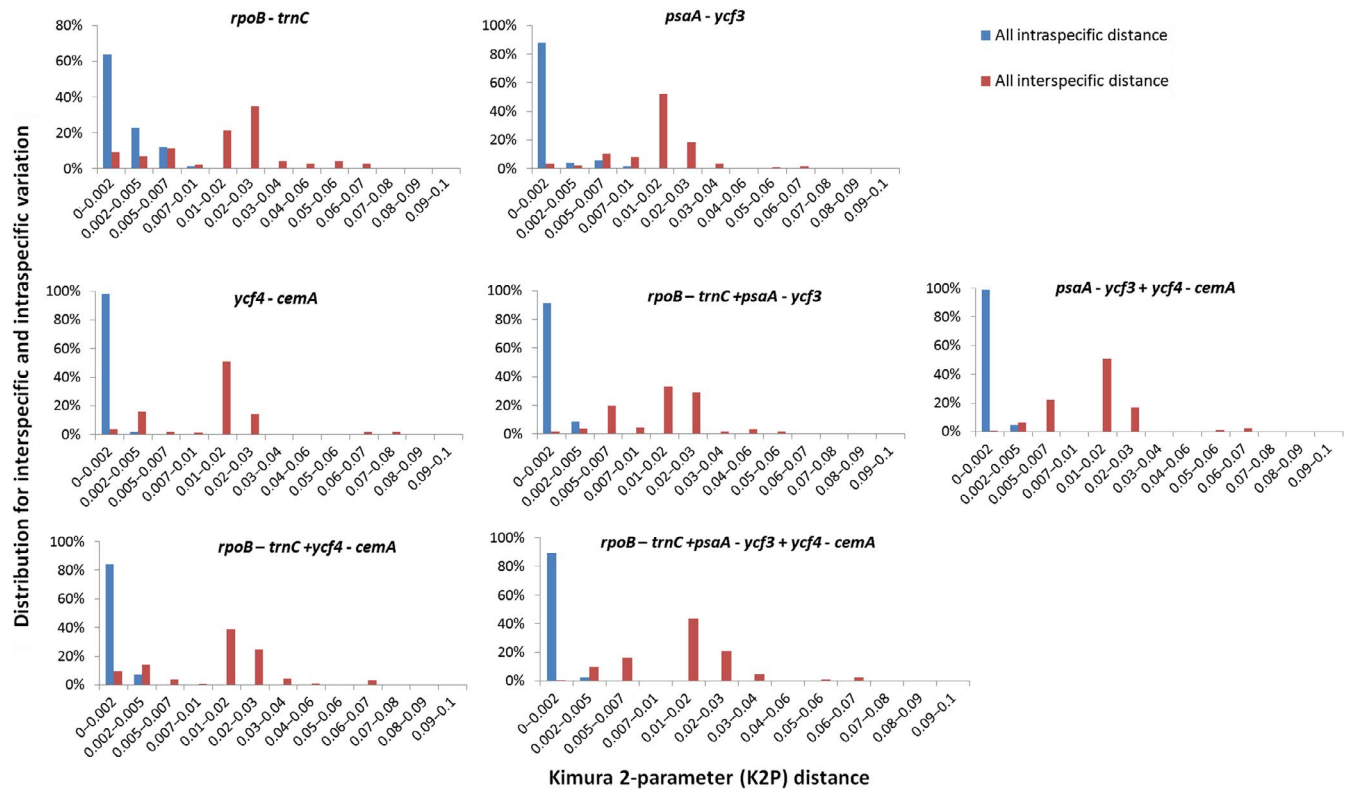


FIGURE 2 Relative distribution of interspecific divergence between congenic species and intraspecific variation

relatively smaller proportions of intraspecific variation in sections 0–0.002. Interspecific variations were mainly >0.01, including for *rpoB-trnC* (93%), *atpF* (79%), *ycf4-cemA* (72%), *psaA-ycf3* (70%), *trnD-trnT* (68%), *rpl14-rpl16* (62%), and *clpP-psbB* (56%). Based on ABGD analysis, *psaA-ycf3* had the highest estimated number of species discriminated (39) with a low threshold value (0.1%), followed by *rpl14-rpl16* (31) and *ycf4-cemA* (32) (Table 4). When the barcode gap was set as 0.3%–0.6%, both *psaA-ycf3* and *trnD-trnT* discriminated 20 species. The PTP analysis indicated that *rpoB-trnC* (36), *psaA-ycf3* (36), and *ycf4-cemA* (35) had the highest estimated species numbers (Table 4). The GMYC analysis showed similar species numbers as the ABGD analysis with low threshold values, and *rpoB-trnC* (27), *psaA-ycf3* (45), and *ycf4-cemA* (29) had the highest estimated species numbers.

Since *ycf4-cemA*, *psaA-ycf3*, *clpP-psbB*, and *rpl14-rpl16* had sequences for 28 *Dioscorea* species, the combined sequences were used for phylogenetic tree construction via maximum-likelihood analysis (Figure 3). The phylogenetic analysis showed that most individuals belonging to the same species tended to cluster together (node value >0.8), such as *D. bulbifera*, *D. arachidna*, *D. esculenta*, and *D. sansibarensis*. Moreover, species belonging to the same sections tended to group together, and the three sections—*Enantiophyllum*, *Shannicorea*, *Asterotracha*—have closer evolutionary relationships than the other sections. In the ABGD analysis, the pairwise genetic distance distribution showed two modes with a barcoding gap located between 0.1% and 1% (Table 4). The higher threshold levels suggested two species (1%), while low threshold values

(0.2%) identified 23 species (Figure 3). The PTP analysis suggested 21 species with support value higher than 0.5. The GMYC analysis has predicted 24 effective candidate species, according to the lineage-through-time plot and the likelihood function estimated by the software R ($L_0 = 390.15$, $L_{\text{multiple}} = 399.27$, $L_{\text{ratio}} = 18.23$, $p\text{-value} = .00011$). Majority PTP and GMYC analysis shared similar species delimitation, except for species belonging to the *Lasiophyton* and *Enantiophyllum* sections.

4 | DISCUSSION

An ideal DNA barcode should have high PCR amplification efficiency and cover regions with enough variability for species identification. Sun et al. (2012) applied *rbcl*, *matK*, and *psbA-trnH* to identify species within the *Dioscorea* genus and found *matK* was the best DNA barcoding candidate, with a species discrimination rate of 23.26%. In this study, we made a thorough comparison of genic and intergenic regions of three *Dioscorea* chloroplast genome sequences to identify highly variable regions for DNA barcode development. We used 74 *Dioscorea* individuals from China and 47 *Dioscorea* SRAs from the NCBI database to estimate primer universality and species discrimination efficiency. A total of 29 species belonging to 11 sections were included in the analysis, among which *Enantiophyllum* Uline (14), Sect. *Opsophyton* Mine (2), *Lasiophyton* Uline (4), and *Asterotracha* (2) have more than one species. We selected seven pairs of primers (7/52) for further analysis which had high PCR amplification rates

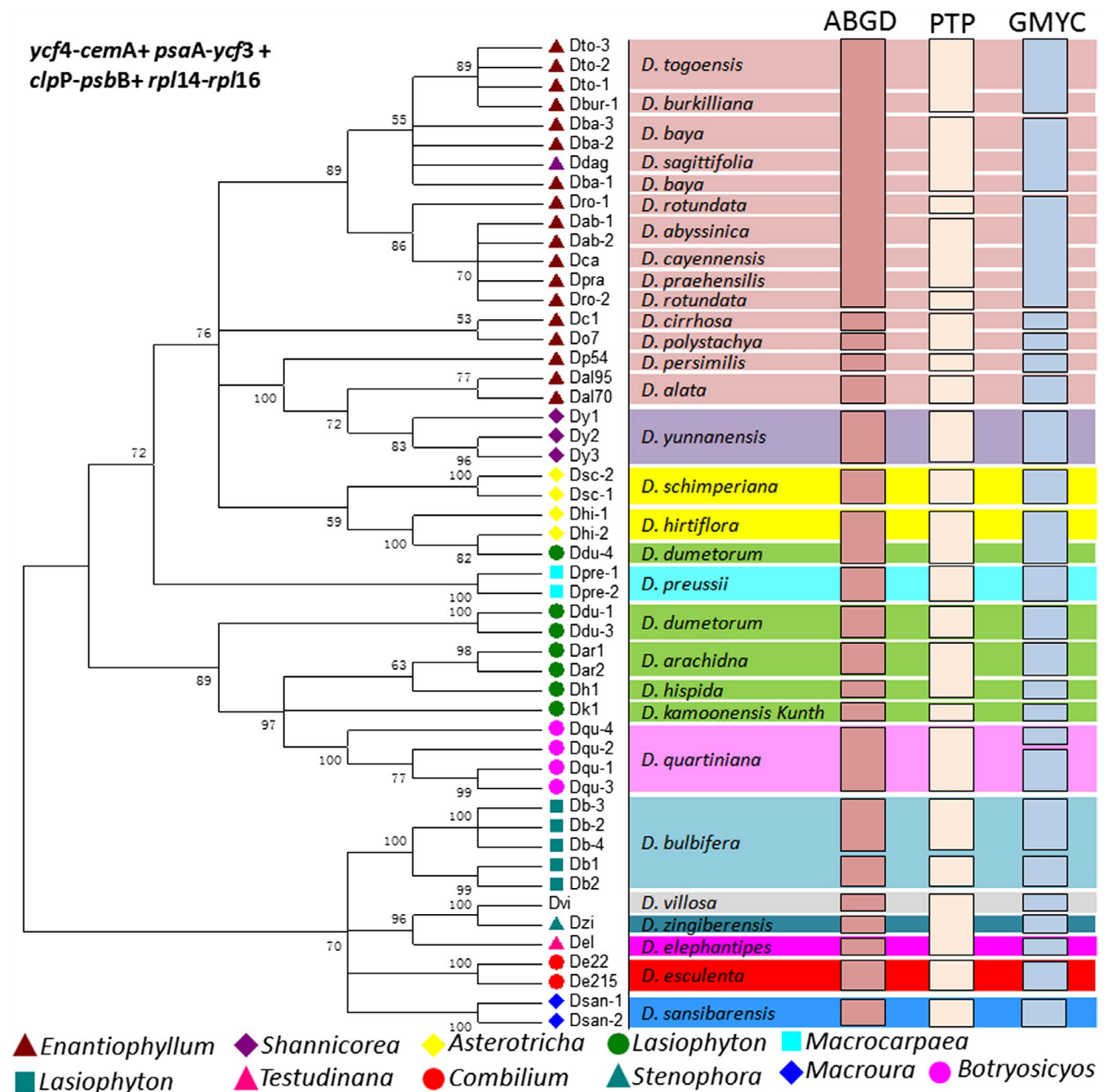


FIGURE 3 The phylogenetic tree constructed using maximum likelihood for *Dioscorea* species based on *ycf4-cemA + psaA-ycf3 + clpP-psbB + rpl14-rpl16* (on the left) and summary of putative species delimitation drawn by BLAST, ABGD, PTP, and GMYC (on the right, one column per method)

and distinct sequence variations between species. The intraspecific and interspecific variation analysis, along with different methods of species discrimination, indicated that these loci have divergent species discrimination efficiency.

DNA barcodes show a relatively variable species discrimination efficiency in different plants (Gogoi & Bhau, 2018; Hollingsworth et al., 2011; Liu et al., 2017; Sun et al., 2015, 2012), and more DNA barcodes for species-level resolution have been developed and tested (Dong et al., 2015; Song et al., 2017). At present, the development of universal primers for highly variable regions relies on the

availability of sequences for different species. New primers of ITS regions of plants with improved universality and specificity were designed based on 1,264,929 sequences of 18S, 5.8S, and 26S from the plant and fungus kingdoms (Cheng et al., 2016). The comparison of chloroplast genomes for genic and intergenic region between three *Dioscorea* species indicated that intergenic regions had more variable loci than genic regions and that conserved genic regions were suitable for primer design (Table 1). However, the primer sequences conserved between three *Dioscorea* species still have low ratios of universal amplification success across different species (Table S2).

The low available numbers of *Dioscorea* chloroplast genomes sequences may limit the efficiency of primer design. With the growing of available chloroplast genome sequences, more efficient primers could be designed in silico.

Through analysis of a set of PCR amplicons from 73 *Dioscorea* individuals and 47 DNA SRA datasets of *Dioscorea* species, the seven selected loci showed significant variation for their interspecies distances. The *rpoB-trnC* locus has the greatest average interspecies distances, and the Wilcoxon signed-rank test indicated the same result (Tables 4 and 5). The *Dioscorea* genus contains more than 600 species, while *Dioscorea* spp. is used for unnamed wild *Dioscorea* species. Abundant efforts have been made to reveal the diversity and evolutionary relationship between *Dioscorea* species (Chair et al., 2005; Girma, Spillane, & Gedil, 2016; Hsu, Tsai, Chen, Ku, & Liu, 2013; Magwetindo, Zapfack, & Sonke, 2016; Mukherjee & Bhat, 2013; Ngwe, Omokolo, & Joly, 2015). Eleven *Dioscorea* sections were included in analysis in this study (Figure 3 and Tables S1 and S3). Phylogenetic analysis based on *ycf4-cemA*, *psaA-ycf3*, *clpP-psbB*, and *rpl14-rpl16* loci produced clear clustering of most species to the sections, but species discrimination for species belonging to *Lasiophyton* and *Enantiophyllum* sections was not very accurate (Figure 3). This may be caused by the close evolutionary relationships between *Dioscorea* species in these sections.

With the growing availability of sequence information, species discrimination through molecular evidence is becoming both feasible and reliable. Plastid markers, such as *rbcL*, *matK*, and *trnH-psbA*, have been widely used with high amplification success in these regions (Hollingsworth et al., 2011). The internal transcribed spacers from nuclear ribosomal DNA, complete plastid genomes, and single copy nuclear genes have also been used in species discrimination (Cheng et al., 2016; Duarte et al., 2010; Song et al., 2017). In this study, we selected primers covering highly variable regions in the *Dioscorea* chloroplast genome. Although only seven pairs of primers had good amplification success, the success rates for species discrimination using these primers were high. Along with other research, in which primers for DNA barcodes have been designed based on available sequences, our results suggest that the growing amount of sequence information will greatly enhance the development of suitable DNA barcodes for taxonomy analysis and species delimitation.

ACKNOWLEDGMENTS

Thank you very much for Annaliese S. Mason from the Justus Liebig University Giessen, who made crucial suggestions and revisions for grammars correcting. This work was supported by grants from the Natural Science Foundation of China (No. 31760406), the Opening Project Fund of Key Laboratory of Rubber Biology and Genetic Resource Utilization, Ministry of Agriculture/State Key Laboratory Breeding Base of Cultivation & Physiology for Tropical Crops/Danzhou Investigation & Experiment Station of Tropical Crops, Ministry of Agriculture (RRI-KLOF201601), and the Crop Science Postgraduate Innovation Project of Hainan University Tropical Agriculture and Forestry College (ZWCX2018033).

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

W.X., J.W., and D.H. conceived the presented idea. B.Z., W.X., and J.S. performed the data analysis and did the experiments. Y.D., W.W., W.T., and J.Z. collected the samples and did the DNA extraction. X.H., Y.X., and J.X. revised the manuscript. B.Z. and W.X. wrote the manuscript.

DATA AVAILABILITY STATEMENT

Three chloroplast genome sequences for *Dioscorea* species are available in NCBI with accession numbers as KJ490011.1, EF380353.1 and KP899622.1. The sequences for *atpF*, *rpoB-trnC*, *trnD-trnT*, *psaA-ycf3*, *ycf4-cemA*, *clpP-psbB* and *rpl14-rpl16* are listed in Table S1 and deposited in ENA (Accession: PRJEB29179, ERZ773993). The 47 SRAs for *Dioscorea* species are available in NCBI and accession numbers are listed in Table S3.

ORCID

Wei Xia  <https://orcid.org/0000-0002-6070-9308>

REFERENCES

- Aumsuwan, P., Khan, S. I., Khan, I. A., Ali, Z., Avula, B., Walker, L. A., & Dasmahapatra, A. K. (2016). The anticancer potential of steroidal saponin, dioscin, isolated from wild yam (*Dioscorea villosa*) root extract in invasive human breast cancer cell line MDA-MB-231 in vitro. *Archives of Biochemistry and Biophysics*, 591, 98–110.
- Chair, H., Perrier, X., Agbangla, C., Marchand, J. L., Dainou, O., & Noyer, J. L. (2005). Use of cpSSRs for the characterisation of yam phylogeny in Benin. *Genome*, 48(4), 674–684. <https://doi.org/10.1139/g05-018>
- Cheng, T., Xu, C., Lei, L., Li, C., Zhang, Y., & Zhou, S. (2016). Barcoding the kingdom Plantae: New PCR primers for ITS regions of plants with improved universality and specificity. *Molecular Ecology Resources*, 16(1), 138–149.
- Cho, J., Choi, H., Lee, J., Kim, M., Sohn, H., & Lee, D. G. (2013). The antifungal activity and membrane-disruptive action of dioscin extracted from *Dioscorea nipponica*. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1828(3), 1153–1158. <https://doi.org/10.1016/j.bbamem.2012.12.010>
- Dong, W., Xu, C., Li, C., Sun, J., Zuo, Y., Shi, S., ... Zhou, S. (2015). *ycf1*, the most promising plastid DNA barcode of land plants. *Scientific Reports*, 5(1), 8348. <https://doi.org/10.1038/srep08348>
- Duarte, J. M., Wall, P. K., Edger, P. P., Landherr, L. L., Ma, H., Pires, J. C., ... Depamphilis, C. W. (2010). Identification of shared single copy nuclear genes in Arabidopsis, Populus, Vitis and Oryza and their phylogenetic utility across various taxonomic levels. *BMC Evolutionary Biology*, 10(1), 61. <https://doi.org/10.1186/1471-2148-10-61>
- Ezard, T., Fujisawa, T., & Barraclough, T. G. (2009). *SPLITS: Species limits by threshold statistics. R package version 1.0-11*. Retrieved from <http://r-forge.r-project.org/projects/splits/>
- Gao, X., Zhu, Y. P., Wu, B. C., Zhao, Y. M., Chen, J. Q., & Hang, Y. Y. (2008). *Phylogeny of Dioscorea sect. Stenophora based on chloroplast matK, rbcL and trnL-F sequences*. Paper presented at the International Symposium on the Tree of Life.

- Girma, G., Spillane, C., & Gedil, M. (2016). DNA barcoding of the main cultivated yams and selected wild species in the genus *Dioscorea*. *Journal of Systematics and Evolution*, 54(3), 228–237.
- Gogoi, B., & Bhau, B. S. (2018). DNA barcoding of the genus *Nepenthes* (Pitcher plant): A preliminary assessment towards its identification. *BMC Plant Biology*, 18(1), 153. <https://doi.org/10.1186/s12870-018-1375-5>
- Hollingsworth, P. M., Graham, S. W., & Little, D. P. (2011). Choosing and using a plant DNA barcode. *PLoS ONE*, 6(5), e19254. <https://doi.org/10.1371/journal.pone.0019254>
- Hsu, K. M., Tsai, J. L., Chen, M. Y., Ku, H. M., & Liu, S. C. (2013). Molecular phylogeny of *Dioscorea* (Dioscoreaceae) in East and Southeast Asia. *Blumea*, 58(1), 21–27.
- Huang, X., & Madan, A. (1999). CAP3: A DNA sequence assembly program. *Genome Research*, 9(9), 868–877. <https://doi.org/10.1101/gr.9.9.868>
- IBM Corp (2017). *IBM SPSS statistics for windows, Version 25.0*. Armonk, NY: IBM Corp.
- Jeon, J. R., Lee, J. S., Lee, C. H., Kim, J. Y., Kim, S. D., & Nam, D. H. (2006). Effect of ethanol extract of dried Chinese yam (*Dioscorea batatas*) flour containing dioscin on gastrointestinal function in rat model. *Archives of Pharmacological Research*, 29(5), 348–353. <https://doi.org/10.1007/BF02968583>
- Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, btz305. <https://doi.org/10.1093/bioinformatics/btz305>
- Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33(7), 1870–1874. <https://doi.org/10.1093/molbev/msw054>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
- Lee, S.-C., Wang, C.-H., Yen, C.-E., & Chang, C. (2017). DNA barcode and identification of the varieties and provenances of Taiwan's domestic and imported made teas using ribosomal internal transcribed spacer 2 sequences. *Journal of Food and Drug Analysis*, 25(2), 260–274. <https://doi.org/10.1016/j.jfda.2016.06.008>
- Liu, Z. F., Ci, X. Q., Li, L., Li, H. W., Conran, J. G., & Li, J. (2017). DNA barcoding evaluation and implications for phylogenetic relationships in Lauraceae from China. *PLoS ONE*, 12(4), e0175788. <https://doi.org/10.1371/journal.pone.0175788>
- Magwetindo, J., Zapfack, L., & Sonke, B. (2016). Diversity of wild yams (*Dioscorea* spp., Dioscoreaceae) collected in continental Africa. *Biodiversity and Conservation*, 25(1), 77–91. <https://doi.org/10.1007/s10531-015-1031-4>
- Mariac, C., Scarcelli, N., Pouzadou, J., Barnaud, A., Billot, C., Faye, A., ... Couvreur, T. L. P. (2014). Cost-effective enrichment hybridization capture of chloroplast genomes at deep multiplexing levels for population genetics and phylogeography studies. *Molecular Ecology Resources*, 14(6), 1103. <https://doi.org/10.1111/1755-0998.12258>
- Mukherjee, P., & Bhat, K. V. (2013). Phylogenetic relationship of wild and cultivated yam species (*Dioscorea* spp.) of India inferred from PCR-RFLP analysis of two cpDNA loci. *Plant Systematics & Evolution*, 299(8), 1587–1597.
- Ngwe, M. F. S. N., Omokolo, D. N., & Joly, S. (2015). Evolution and phylogenetic diversity of yam species (*Dioscorea* spp.): Implication for conservation and agricultural practices. *PLoS ONE*, 10(12), e0145364.
- Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, 20(2), 289–290. <https://doi.org/10.1093/bioinformatics/btg412>
- Paterson, A. H., Brubaker, C. L., & Wendel, J. F. (1993). A rapid method for extraction of cotton (*Gossypium* spp.) genomic DNA suitable for RFLP or PCR analysis. *Plant Molecular Biology Reporter*, 11(2), 122–127.
- Puillandre, N., Lambert, A., Brouillet, S., & Achaz, G. (2012). ABGD, automatic barcode gap discovery for primary species delimitation. *Molecular Ecology*, 21(8), 1864–1877. <https://doi.org/10.1111/j.1365-294X.2011.05239.x>
- R Core Team (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- Raman, V., Galal, A., Avula, B., Sagi, S., Smillie, T. J., & Khan, I. A. (2014). Application of anatomy and HPTLC in characterizing species of *Dioscorea* (Dioscoreaceae). *Journal of Natural Medicines*, 68(4), 686–698. <https://doi.org/10.1007/s11418-014-0849-5>
- Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, 67(5), 901–904. <https://doi.org/10.1093/sysbio/syy032>
- Song, Y., Chen, Y., Lv, J., Xu, J., Zhu, S., Li, M., & Chen, N. (2017). Development of chloroplast genomic resources for *Oryza* species discrimination. *Frontiers Plant Science*, 8, 1854. <https://doi.org/10.3389/fpls.2017.01854>
- Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., & Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*, 4(1), vey016. <https://doi.org/10.1093/ve/vey016>
- Sun, X. Q., Qu, Y. Q., Yao, H., Zhang, Y. M., Yan, Q. Q., & Hang, Y. Y. (2015). Applying DNA barcodes for identification of economically important species in Brassicaceae. *Genetics and Molecular Research*, 14(4), 15050–15061. <https://doi.org/10.4238/2015.November.24.13>
- Sun, X., Zhu, Y., Guo, J., Peng, B., Bai, M., & Hang, Y. (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(2), e32057. <https://doi.org/10.1371/journal.pone.0032057>
- Techen, N., Parveen, I., Pan, Z., & Khan, I. A. (2014). DNA barcoding of medicinal plant material for identification. *Current Opinion in Biotechnology*, 25(2), 103. <https://doi.org/10.1016/j.copbio.2013.09.010>
- Wilkin, P., Schols, P., Chase, M. W., Chayamarit, K., Furness, C. A., Huysmans, S., ... Thapayai, C. (2005). A plastid gene phylogeny of the yam genus, *Dioscorea*: Roots, fruits and madagascar. *Systematic Botany*, 30(4), 736–749.
- Wu, L., Wang, B., Yang, J., Song, C., Wang, P., Chen, S., & Sun, W. (2016). The chloroplast genome sequence of an important medicinal plant *Dioscorea nipponica*. *Mitochondrial DNA Part A*, 27(4), 2559–2560. <https://doi.org/10.3109/19401736.2015.1038803>
- Zhang, J., Kapli, P., Pavlidis, P., & Stamatakis, A. (2013). A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, 29(22), 2869–2876. <https://doi.org/10.1093/bioinformatics/btt499>
- Zhou, W., Chen, C., Hua, W.-P., & Wang, Z.-Z. (2016). The complete chloroplast genome sequence of *Dioscorea zingiberensis* (Dioscoreaceae). *Mitochondrial DNA Part A*, 27(4), 2730–2731. <https://doi.org/10.3109/19401736.2015.1046168>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Xia W, Zhang B, Xing D, et al.

Development of high-resolution DNA barcodes for *Dioscorea* species discrimination and phylogenetic analysis. *Ecol Evol*. 2019;9:10843–10853. <https://doi.org/10.1002/ece3.5605>