GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

DNA barcoding for species identification in the Palmae family

A. Naeem[1], A.A. Khan[1], H.M.N. Cheema[1], I.A. Khan[1] and A. Buerkert[2]

[1]Plant Genetic Resources Lab, University of Agriculture, Faisalabad, Pakistan
[2]Organic Plant Production and Agroecosystems Research, University of Kassel, Witzenhausen, Germany

Corresponding author: A.A. Khan
E-mail: asifpbg@uaf.edu.pk / masooma@uaf.edu.pk

Genet. Mol. Res. 13 (4): 10341-10348 (2014)
Received March 18, 2014
Accepted July 2, 2014
Published December 4, 2014
DOI http://dx.doi.org/10.4238/2014.December.4.29

ABSTRACT. DNA barcoding is a promising tool for species identification at the molecular level. The barcoding system is well established for species differentiation in animals, while it is less common in plants. We evaluated 2 barcoding regions, maturase K (*mat*K) and ribulose bisphosphate carboxylase (*rbc*L), to compare species of Palmae according to amplification success, discrimination power, and inter- and intra-specific divergence. Both regions appear to have potential to discriminate most species of Palmae, but 2 species, *Phoenix dactylifera* and *Phoenix sylvestris*, did not show variation in the nucleotides of the barcode genes. *P. sylvestris* is said to be the sister species of *P. dactilyfera* according to its morphological and genetic proximity to the cultivated date palm. Thus, the status of these 2 species needs to be re-evaluated considering more genes as barcodes. Furthermore, *rbc*L has a higher discrimination power (90%) than *mat*K (66.6%) and can thus be potentially used as a standard barcode to discriminate the species of Palmae.

Key words: Date palm; Ribulose bisphosphate carboxylase (*rbc*L); Maturase K (*mat*K); Sequencing; Cloning; Taxonomy

INTRODUCTION

DNA barcoding involves sequencing short DNA sequences from a particular region of the genome and comparing them between and within species to present a "barcode" for species identification (Hebert et al., 2003). This technique proved to be a promising tool for species discrimination in animals using the mitochondrial gene cytochrome c oxidase I (*COI*) (Cowan et al., 2006). In plants, this is not a straightforward technique because of the lower evolutionary rate of plant mitochondria, resulting in less variation in the *COI* gene among species. In some plants, the chloroplast genome showed potential for candidate barcoding genes with characteristic features such as conserved gene order, high copy number, and easy amplification by polymerase chain reaction (PCR) (Chase et al., 2005). A multi-locus approach using the chloroplast genome has been successfully used as an effective strategy to barcode various land plants (Kress and Erickson, 2007; Newmaster et al., 2008; CBOL Plant Working Group, 2009). Recently, standard barcodes for plants have been tested for this purpose (Chase et al., 2005; Cowan et al., 2006; Kress and Erickson, 2007; Lahaye et al., 2008; CBOL Plant Working Group, 2009). The maturase K (*mat*K) and ribulose bisphosphate carboxylase (*rbc*L) genes have been adopted as barcodes by the Consortium for the Barcode of Life (CBOL) because of the easy recovery of *rbc*L and the high discrimination power of *mat*K (CBOL Plant Working Group, 2009). These barcoding efforts resulted in *in silico* libraries of DNA barcode sequences that could be used as a standard for species identification such as the Barcodes of Life Data system (BOLD; Ratnasingham and Hebert, 2007).

Palmae (Arecaceae) is the third most important plant family for human use; it includes about 180 genera and 2000 species distributed across the tropics and subtropics (Hahn, 2002). Date palm (*Phoenix dactylifera* L.) is an important fruit crop in the Palmae family that covers 3% of the cultivated area of the world. The genus *Phoenix* contains 17 species, and 4 of them are present in Pakistan. The fruit of *P. dactylifera* is sweeter than other fruit crops of the family, and it is the only *Phoenix* species that is used as a staple by millions of people around the globe (Zaid and deWet, 2002).

Determining the genetic variation in palm cultivars using morphological traits is difficult because the required phenotypic data are often difficult to assess due to high environmental effects. Moreover, morphological discrimination identifies individuals at a specific life stage or gender in such a way that many individuals remain unidentified or are wrongly identified (Sedra et al., 1993, 1996). Under these conditions, a molecular approach seems particularly useful. Therefore, this study assessed the potential and efficacy of barcoding using 2 chloroplast loci (*mat*K and *rbc*L) as barcodes of Palmae for species discrimination.

MATERIAL AND METHODS

DNA extraction and PCR amplification

Leaf samples of 10 species of the Palmae family (Table 1) were collected for DNA extraction. To further evaluate the candidate barcodes, 34 date palm cultivars, morphologically all belonging to the same species (*P. dactylifera*), were also sampled (Table 2). Good quality DNA was isolated from the collected samples by a modified cetyltrimethylammonium bromide method (Sharma et al., 2003). A dilution of 20 ng/µL was prepared to optimize the DNA concentration for PCR amplification. Chloroplast markers (*mat*K and *rbc*L) were ampli-

fied using a set of primers (Table 3). The amplifications were performed in 25-µL reactions. PCR conditions were as follows: initial denaturation at 95°C for 4 min; 40 cycles of 1 min at 94°C, 1 min at 50°C, and 1 min at 72°C; and a final extension step of 10 min at 72°C. The PCR amplifications were separated on a 1% agarose gel with a 1-kb gene ruler (Fermentas, USA).

Table 1. Species of Palmae used for barcoding.

Species	Common name	Species	Common name
Chamaedorea elegans	Good luck palm	*Ravenea rivularis*	Majesty palm
Livistona chinensis	Common palm	*Caryota mitis*	Fishtail palm
Phoenix roebelenii	Pygmy date palm	*Phoenix sylvestris*	Silver date palm
Chamaerops humilis	European fan palm	*Archontophoenix alexandrae*	Alexandra
Dypsis decaryi	Triangle palm	*Phoenix dactylifera*	Date palm

Table 2. Varieties of *Phoenix dactylifera*.

1	Akhrot	10	Neelam	19	Begum-jangi	28	Peli sundar
2	Dhakki	11	Zerin	20	Deglet Noor	29	Khadrawi-I
3	Aseel	12	Jaman	21	Pela Dora	30	Khadrawi-II
4	Hillawi-I	13	Ko-Harba	22	Shamran-I	31	Wahn wali
5	Qantar	14	Kozan Abad	23	Shamran-II	32	Hillawi-II
6	Angor	15	Karblain	24	Rachna	33	Champa kali
7	Chohara	16	Jan-Sohar	25	Seib	34	Makran
8	Zaidi	17	Kokna	26	Zardo		
9	Be-Rehmi	18	Danda	27	Shado		

Table 3. Barcoding primers and sequences.

No.	Primer name	Primer sequence
1	*rbc*L-F	5'-ATGTCACCACAAACAGAGACTAAAGC-3'
2	*rbc*L-R	5'-GAAACGGTCTCTCCAACGCAT-3'
3	*mat*K-F	5'-CGTACAGTACTTTTGTGTTTACGAG-3'
4	*mat*K-R	5'-ACCCAGTCCATCTGGAAATCTTGGTTC-3'

Sequencing

The PCR-amplified products of the *mat*K and *rbc*L genes were eluted and ligated into a TA cloning vector (pTZ57R/T). The recombinant clones were confirmed by restriction digestion. Purified clones were sequenced using forward and reverse primers of the *mat*K and *rbc*L genes, which were also used for PCR amplification. The sequencing was carried out by Macrogen, Korea. Primer sequences were confirmed in the raw sequencing data, and all sequences were submitted to BOLD under the sequence submission tool (http://www.boldsystems.org). Subsequently, the success ratios of PCR amplification and sequencing were calculated.

Data analysis

The sequencing data acquired for all 10 Palmae species and 34 genotypes of date palm for the 2 genes (*mat*K and *rbc*L) were aligned separately using CLUSTAL W (http://www.ebi.ac.uk/Tools/msa/clustalw/). The interspecific and intraspecific divergence of each barcoding region was calculated by Kimura 2 parameter distances in the BOLD analysis tool

(http://www.boldsystems.org). The maximum likelihood substitution matrix was estimated by the MEGA (version 5.0) software, and the neighbor-joining method was used to construct a phylogenetic tree in the MEGA 5 software (Tamura et al., 2011).

RESULTS

Amplification

The amplified PCR products of *mat*K and *rbc*L were 873 and 650 bp, respectively. The amplification success rate of both regions (*mat*K and *rbc*L) was 100%. The cloning of PCR products in pTZ57R/T was confirmed by double digestion with *Hin*dIII and *Bam*HI. This also confirmed the inserted fragment size of 873 and 650 bp for the *mat*K and *rbc*L barcode genes, respectively. The barcodes for all species and genotypes were preserved as glycerol stocks for future analysis.

Sequencing

High-quality bidirectional sequencing data were obtained for both *rbc*L and *mat*K with a success rate of 93.02 and 81.3%, respectively (Table 4). Although the sequence success rate of *rbc*L was higher than that of *mat*K, the overall sequencing efficiency of both barcodes was above 80%. The sequence alignment showed prominent nucleotide polymorphism among species. This polymorphism was further evaluated to estimate the interspecific and intraspecific divergence. Higher interspecific variation than intraspecific variation represents the success of the study and may have wide scientific applications. For the *rbc*L gene, the analysis involved 9 nucleotide sequences. There were a total of 654 positions in the final dataset. The T to G and C to G transversion rate was 4.39, that of A to T and G to T was 5.51, that of T to A and C to A was 5.67, and that of A to C and G to C was 4.05. The A to G transition rate was 14.54, that of C to T was 26.74, that of T to C was 19.61, and that of A to G was 18.72. Frequencies of the nucleotide substitution were A = 28.88%, T/U = 28.10%, C = 20.36%, and G = 22.66% (Table 5).

Table 4. Success rate of *rbc*L and *mat*K sequencing.

Barcode markers	*rbcL*	*matK*
Number of samples	43	43
Sequencing success	40	35
Sequencing success rate (%)	93.02%	81.3%

Table 5. Transition and transversion rates of *rbc*L nucleotide sequences in Palmae.

	A	T/U	C	G
A	-	*5.51*	*4.05*	**8.51**
T/U	*5.67*	-	**17.48**	*4.39*
C	*5.67*	**23.77**	-	*4.39*
G	**10.99**	*5.51*	*4.05*	-

Transition rates are shown in bold, and transversion rates are shown in *italics*. For simplicity, the sum of r values is 100.

For the *mat*K gene, the analysis involved 6 nucleotide sequences. There were a total of 874 positions in the final dataset. The A to T and G to T transversion rate was 7.68, that of A to C was 4.09, that of T to A and C to A was 9.39, and that of A to C and G to C was 4.09. The A to G transition rate was 9.51, that of C to T was 12.50, that of T to C was 6.65, and that of A to G was 20.19. The frequencies of nucleotide substitution were A = 36.71%, T/U = 30.02%, C = 15.98%, and G = 17.30% (Table 6). Both genes in our study showed high interspecific divergence and low or no intraspecific divergence (Tables 7 and 8).

Table 6. Transition and transversion rates of *matK* nucleotide sequences in Palmae.

	A	T/U	C	G
A	-	*7.68*	*4.09*	**9.51**
T/U	*9.39*	-	**6.65**	*4.42*
C	*9.39*	**12.50**	-	*4.42*
G	**20.19**	*7.68*	*4.09*	-

Transition rates are shown in bold, and transversion rates are shown in *italics*. For simplicity, the sum of r values is 100.

Table 7. Interspecific and intraspecific distances for *rbcL* by the Kimura 2-parameter method.

	N	Taxa	Comparisons	Minimum distance (%)	Mean distance (%)	Maximum distance (%)	SE distance (%)
Within species	32	1	496	0	12.733	75	0.053
Within genus	33	1	32	0	6.661	75	0.629
Within family	40	1	252	0.306	16.593	75	0.119

N = total number of samples included in species, genus, and family.

Table 8. Interspecific and intraspecific distances for *mat*K by the Kimura 2-parameter method.

	N	Taxa	Comparisons	Minimum distance (%)	Mean distance (%)	Maximum distance (%)	SE distance (%)
Within species	30	1	435	0	18.326	75	0.073
Within genus	32	1	61	1.944	70.209	75	0.296
Within family	35	1	99	1.021	48.169	75	0.352

N = total number of samples included in species, genus, and family.

A neighbor-joining tree was constructed to examine the phylogenetic relationship between species. The tree constructed for *rbc*L sequences showed similarity between *Archontophoenix alexandrae* and *Chamaedorea elegans* because they were in the same node. The *Dypsis decaryi*, *Chamaerops humilis*, and *Livistona chinensis* were independent entries. Additionally, *Caryota mitis* had more diversification than other species (Figure 1). For *mat*K, *C. humilis* and *L. chinensis* were similar. *Phoenix roebelenii* was similar to *P. dactylifera* and *Phoenix sylvestris* (Figure 2).

P. dactylifera and *P. sylvestris* revealed a strong relation of 100% similarity for both genes. The nucleotide variation in *P. sylvestris* and *P. dactylifera* was found to be at the same positions although they are morphologically different from each other and exist as independent species.

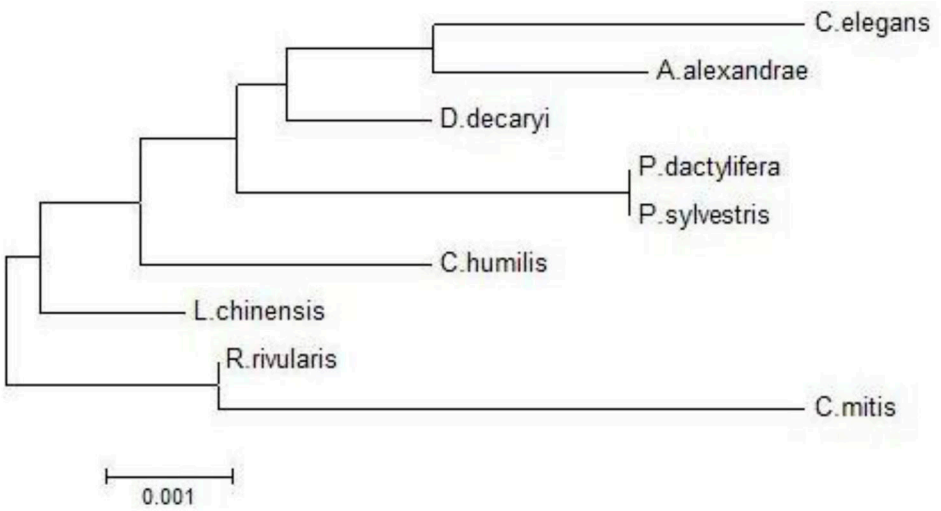

Figure 1. Neighbor-joining tree for *rbcL* sequences. Species are those listed in Table 1.

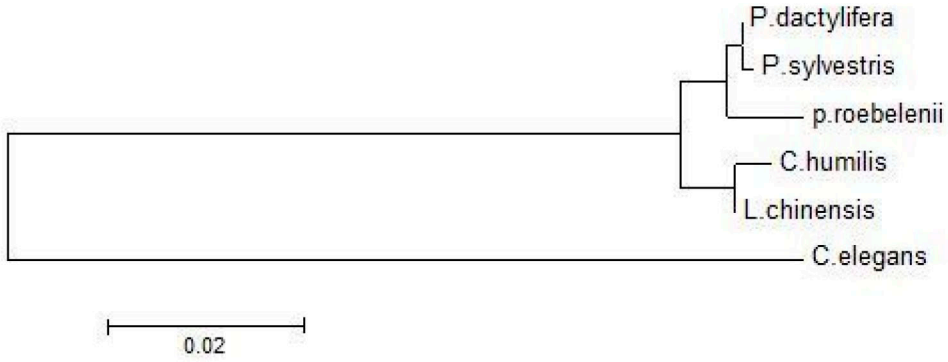

Figure 2. Neighbor-joining tree for *matK* sequences. Species are those listed in Table 1.

DISCUSSION

DNA barcoding is a valuable tool for taxonomists working with Palmae. It can be used to identify species efficiently and accurately on the basis of a standard region as a marker. To identify an ideal region in plants, which must be sufficiently variable to differentiate all the species and conserved enough to be minimally variable within species, is nevertheless a challenge (Kress et al., 2005; Liu et al., 2010). Various loci have been tested and evaluated as DNA barcodes, and different studies have tried to define a standard barcode for plants. CBOL has proposed *rbc*L and *mat*K as standard barcodes for land plants (CBOL Plant Working Group, 2009).

To evaluate a DNA barcode, universal PCR amplification and sequencing success are important parameters. In this study, 2 barcode genes (*mat*K and *rbc*L) showed 100% amplification success. The *mat*K locus could not be sequenced from some samples because of the inter-

ruption of sequencing reads by downstream mononucleotide repeats. This problem was also reported for *mat*K in previous studies (Fazekas et al., 2008; CBOL Plant Working Group, 2009).

The *mat*K gene is used as a universal barcode for flowering plants because of its easy alignment (Lahaye et al., 2008; CBOL Plant Working Group, 2009). In this study, the *mat*K and *rbc*L sequences from the varieties of *P. dactylifera* did not show any polymorphism. The sequences were 100% conserved within the species as suggested by Hebert et al. (2003). Nucleotide polymorphism was observed in alignments that included different species, and these polymorphisms were enough to discriminate these species. The power of a barcode to discriminate species is an essential criterion that determines its utility. The *mat*K sequence had a lower calculated discrimination power of 66.66% (discriminated 6 of 10 species) than *rbc*L, which was about 90% (discriminated 9 of 10 species).

The maximum likelihood substitution matrix using MEGA 5 shows the probability of substitution from one base to another. These changes include the substitution of a pyrimidine by a purine or a purine by a pyrimidine (transversion) and the substitution of a purine by a purine or a pyrimidine by a pyrimidine (transition). Our results (Tables 7 and 8) are in agreement with the fact that a favorable barcode must possess more interspecific divergence than intraspecific divergence to distinguish the species (Zhu et al., 2010). The lack of sequence variation in *P. dactylifera* and *P. sylvestris* may be due to low rates of sequence evolution and taxonomic misidentification (Kress and Erickson, 2007), which need to be further explored. *P. sylvestris* has morphological and genetic proximity with *P. dactilyfera* and is considered to be a progenitor of *P*. *dactilyfera* (Barrow, 1998). On the basis of these findings, it may be useful to include more genes and more species for a precise and comprehensive system of species identification in Palmae.

In conclusion, a barcoding system based on *rbc*L and *mat*K has the potential to differentiate most of the species of Palmae that were examined. To obtain a holistic assessment of the taxonomy of Palmae, a detailed morphological study combined with a DNA barcoding system using additional loci is necessary.

ACKNOWLEDGMENTS

Research supported by the DAAD-funded International Center for Development and Decent Work (ICDD, www.icdd.uni-kassel.de) at the University of Kassel, Germany.

REFERENCES

Barrow SC (1998). A monograph of *Phoenix* L. (Palmae: Coryphoideae). *Kew Bull.* 53: 513-575.

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

Chase MW, Salamin N, Wilkinson M, Dunwell JM, et al. (2005). Land plants and DNA barcodes: short-term and long-term goals. *Philos. Trans. R. Soc. Ser. B. Biol. Sci.* 360: 1889-1895.

Cowan RS, Chase MW, Kress WJ and Savolainen V (2006). 300,000 species to identify: problems, progress, and prospects in DNA barcoding of land plants. *Taxon* 55: 611-616.

Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, et al. (2008). Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS One* 3: e2802.

Hahn JW (2002). A molecular phylogenetic study of the Palmae (Arecaceae) based on *atp*B, *rbc*L, and 18S nrDNA sequences. *Syst. Biol*. 51: 92-112.

Hebert PD, Cywinska NA, Ball SL and deWaard JR (2003). Biological identifications through DNA barcodes. *Proc. Biol. Sci*. 270: 313-321.

Kress WJ and Erickson DL (2007). A two-locus global DNA barcode for land plants: the coding *rbc*L gene complements

the noncoding *trnH-psbA* spacer region. *PLoS One* 2: e508.
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 RM, 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.
Liu ZK, Chen K, Luo K, Pan H, et al. (2010). DNA barcoding in medicinal plants Caprifoliaceae. *Zhongguo Zhong Yao Za Zhi* 35: 2527-2532.
Newmaster SG, Fazekas AJ, Steeves RAD and Janovec J (2008). Testing candidate plant barcode regions in the Myristicaceae. *Mol. Ecol. Resour.* 8: 480-490.
Ratnasingham S and Hebert PDN (2007). BOLD: the Barcode of Life Data System (http://www.barcodinglife.org). *Mol. Ecol. Notes* 7: 355-364.
Sedra M, El Filali H and Frira D (1993). Observations sur quelques caracteristiques phenotypiques et agronomiques du fruit des varietés et clones du palmier dattier selectionnés. *Al Awamia* 82: 105-120.
Sedra M, El Filali H, Benzine A, Allaoui M, et al. (1996). La palmeraie dattiere marocaine: evaluation du patrimoine phenicicole. *Fruits* 1: 247-259.
Sharma R, Mahla HR, Mohapatra T, Bhargava SC, et al. (2003). Isolating plant genomic DNA without liquid nitrogen. *Plant. Mol. Biol. Rep.* 21: 43-50.
Tamura K, Peterson D, Peterson N, Stecher G, et al. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731-2739.
Zaid A and deWet PF (2002). Origin, Geographical Distribution and Nutritional Value of Date Palm In: Date Palm Cultivation (Zaid A, ed.). Food and Agricultural Organization of the United Nations, Rome.
Zhu YJ, Chen SL, Yao H, Tan R, et al. (2010). DNA barcoding the medicinal plants of the genus *Paris*. *Yao Xue Xue Bao* 45: 376-382.