Plant Omics Journal POJ

POJ 4(4):195-198 (2011) ISSN:1836-3644

Research Note

Comparative evaluation of PCR success with universal primers of maturase K (*matK*) and ribulose-1, 5-bisphosphate carboxylase oxygenase large subunit (*rbcL*) for barcoding of some arid plants

Sameera O. Bafeel[1], Ibrahim A. Arif[2,3], Mohammad A. Bakir[2,3], Haseeb A. Khan[2,3*], Ahmad H. Al Farhan[2,3], Ali A. Al Homaidan[2,3], Anis Ahamed[2,3], Jacob Thomas[3]

[1]Department of Biology, College of Science, King Abdulaziz University, Jeddah, Saudi Arabia
[2]Molecular Fingerprinting and Biodiversity Unit, Prince Sultan Research Chair for Environment and Wildlife, College of Sciences, King Saud University, Riyadh, Saudi Arabia
[3]Department of Botany and Microbiology, College of Sciences, King Saud University, Riyadh, Saudi Arabia

***Correspondence author: khan_haseeb@yahoo.com**

Abstract

DNA barcoding is the use of short DNA sequences (~650 bp) of the standard segment of the genome for large scale species identification. The Consortium for the Barcode of Life (CBOL) plant-working group recommended the 2-locus combination of *rbcL* and *matK* as the standard plant barcode. These two regions of chloroplast DNA were chosen due to efficient recovery of quality sequences and high levels of species discrimination. We evaluated the success rates of universal primers for amplification of *matK* and *rbcL* loci in 26 different plant species (covering 14 families) from Saudi Arabia. Success rate in PCR was higher for *rbcL* (88%) compared with *matK* (69%). The universal primers of both *matK* and *rbcL* failed to amplify the DNA form 3 plant species belonging to the family Asteraceae (*Anthemis deserti*, *Pulicaria undulate*, and *Sonchus oleraceus*). Two plant species *Malva parviflora* (Malvaceae) and *Salsola imbricate* (Chenopodiaceae) indicated different primer binding site (*matK*) as the amplified PCR products were of lower size than expected for these species. These findings indicate that although currently used universal primers of *rbcL* and *matK* are able to amplify many of the plant species they may fail in certain cases due to primer mismatch at the annealing site. Further studies are therefore needed for protocol development, particularly designing of novel universal primers, to extend the barcoding for a broader coverage of plant species.

Keywords: Barcoding, Plants, *matK*, *rbcL*, PCR.
Abbreviations: *matK*_maturase K; *rbcL*_ribulose-1, 5-bisphosphate carboxylase/oxygenase large subunit; ITS_Internal Transcribed Spacer; CBOL_Consortium for the Barcode of Life; BOLD_The Barcode of Life Data Systems; BLAST_Basic Local Alignment Search Tool; 16S rRNA_16S ribosomal RNA; PCR_polymerase chain reaction.

Introduction

DNA barcoding relies on the information encoded in the nucleotide sequences of a standard region of the genome as a tool for species identification. The Consortium for the Barcode of Life (CBOL) plant working group recommended the 2-locus combination of ribulose-1, 5-bisphosphate carboxylase oxygenase large subunit (*rbcL*) and maturase K (*matK*) as the standard plant barcode based on assessments of recoverability, sequence quality and levels of species discrimination (CBOL Plant Working Group, 2009). These two regions of chloroplast DNA were chosen based on two main criteria: efficient recovery of good-quality sequences and high levels of species discrimination (Burgess et al., 2011). Recently, several investigators have used *rbcL* and *matK* sequences for barcoding or species identification (Asahina et al., 2010, Starr et al., 2009) as well as for phylogenetic analysis (Kuo et al., 2011, Manen et al., 2004, Tamura et al., 2004). However, universal barcode markers may not work in all cases (Roy et al., 2010). Success in PCR is a pre-requisite for barcoding as this technique is exclusively used to amplify the target sequence. Thus achieving high PCR success rates continues to be important, particularly for environmental sampling studies where primer bias and non-universality dramatically skews results (Soininen et al., 2009). Primer-template mismatches at the 3′ end of the primer sequence is a major factor to negate the universality of a primer as a single mismatched base at the 3′ end is able to prevent PCR amplification (Bru et al., 2008). Even the pitfalls of universal primers have been noticed for the highly conserved region, 16S rRNA, resulting in differential amplification in a mixture of template DNA from a complex microbiota leading to a distorted view of microbial diversity (von Wintzingerode et al., 1997). Plants of the arid environment are adapted to enduring harsh climatic conditions, thus possessing different survival characteristics and molecular diversity (Bokhari et al., 1990, Kamal et al., 2010). Universal barcode markers need to be evaluated for a broader spectrum due to morphological/geographical variation and reticulate evolution in plant species. The survival characteristics and diversity of plant in Saudi Arabia are different as they grow under harsh environment; the barcodes of most of the arid plants remains to be discovered. It is imperative that molecular variations in arid plants may have some impact on the PCR amplification of barcoding markers. We therefore investigated the success rate of *matK* and *rbcL* universal primers for barcoding of different plant species of Saudi Arabia.

Results and discussion

We observed a high success rate of *rbcL* universal primers than *matK* primers for PCR amplification of 26 plant species representing a wide coverage of 14 families. Out of 26 specimens, 23 (88%) were successfully amplified using rbcL-A and rbcL-B primer-pairs; both these pairs were equally effective in amplification success however the former pair resulted more intense bands. In contrast, only 7 (27%) and 18 (69%) samples could be amplified by *matK-A* and *matK-B* primer-pairs respectively (Fig. 1). These findings corroborate with the previous report on evaluation of the seven main candidate plastid regions (*rbcL*, *matK*, *rpoC1*, *rpoB*, *trnH-psbA*, *atpF-atpH*, and *psbK-psbI*) in three divergent groups of land plants (angiosperm, gymnosperm, and liverwort) demonstrated that no single locus has high levels of universality and resolvability (Hollingsworth et al., 2009). The successful isolation of quality DNA plays an important role in molecular characterization using PCR-based methods. Numerous problems have been reported about the isolation of plant DNA. Isolated plant DNA may contain colored substances, polysaccharides and phenolic compounds (Aras et al., 1993; Temiesak et al., 1993; Vanijajiva et al., 2005). The presence of certain metabolites may interfere with the plant DNA isolation procedures and downstream reactions such as PCR amplification. The chemotypic heterogeneity among species may not permit optimal DNA yields with a single protocol, and thus, even closely related species may require different DNA isolation protocols (Khanuja et al., 1999). PCR conditions constitute one of the crucial factors for obtaining amplified products, especially for plants (Jones et al., 1997). In terms of absolute discriminatory power, promising results occurred in liverworts using *rbcL* alone (90% species discrimination) (Hollingsworth et al., 2009). Two plant species *Malva parviflora* (Malvaceae) and *Saldola imbricate* (Chenopodiaceae) resulted in smaller amplified PCR product than the expected size with primers matK-A, indicating a different primer binding site for these species (Fig. 1). Three plant species belonging to the family Asteraceae (*Anthemis deserti*, *Pulicaria undulate*, and *Sonchus oleraceus*) were not amplified using the universal *matK* and *rbcL* primers. The family Asteraceae (Compositae) is the largest family of the flowering plants and it seems to be anomalous as combined analyses of 10 chloroplast DNA loci showed 12 major lineages of this family (Panero and Funk, 2008). The central concept in species identification is to match the sequence of the evidence item to a reference sequence, either through DNA sequence similarity searches (Altschul et al., 1997) or by phylogenetic reconstruction (Baker, 1994). Except RFLP, all the molecular markers including the barcoding genes require PCR-based protocols (Kumar et al., 2009). The use of universal primers is important for retrospective species identification as they allow amplification across a wide taxonomic range governed by the PCR success rates, particularly for new or rarely studied plant species of arid region. PCR amplification of *matK* and *rbcL* gene segments is a prerequisite for subsequent sequencing of plant barcodes towards species identification. Primer bias and non-universality have been reported for dramatically skewing results (Soininen et al., 2009). Several studies have investigated the effects of primer-template mismatches and it has been demonstrated that PCR can be prevented by a single mismatched base (Bru et al., 2008). Variation in primer specificity also affects sequence recovery at low DNA concentrations (Dawnay et al., 2007). Although we observed appreciable success in amplification of *rbcL* and *matK* genes, failure in PCR amplification in for certain species may be attributed to primer mismatch at the annealing sites. The possibility of poor quality of extracted DNA in failed PCR is remote because we confirmed the quality of DNA on gel as well as spectrophotometrically. Moreover, in many cases, the same DNA specimen showed negative amplification for *matK* but positive PCR for *rbcL*, negating the role of poor DNA quality or quantity in the former case. It is more likely that high frequency of failure in PCR amplification of *matK* gene occurred due to variations in the respective binding sites of these particular plant species of arid origin. The generation of *matK* sequences for ferns is currently problematic, because this part of the chloroplast genome underwent a strong restructuring during the evolution of the fern clade (Duffy et al., 2009). Recently, *rbcL* and trnL-F were used (instead of *matK*) for the two locus barcoding of European ferns (de-Groot et al., 2011). Roy et al. (2010) have reported a poor success in PCR for the species of *Berberis*. The maximum success in PCR was observed with *rbcL* and ITS (97%), followed by trnH-psbA (92%) and *matK* (76%). The first practical problem for barcoding of plants is the acquirement of sufficiently clean DNA for multi-locus sequencing as isolated plant DNA contains PCR inhibitors (Aras et al., 1993; Temiesak et al., 1993; Vanijajiva et al., 2005). The second technical issues of primer universality and sequence quality and complexity remain arguable for barcoding of all the land plants of different region (Fox et al., 1992; Schneider and Schuettpelz, 2006). Traditionally, the identification and characterization of cultivars and species has been based on morphological and physiological traits which are sometimes difficult (Kadkhodaei et al., 2010). Molecular tools that have been developed for the characterization for biodiversity may allow classification of synonyms and detection of the origin of species and cultivars (Rahman et al., 2009). Ongoing research on the development of universal primers, PCR conditions, and high throughput sequencing techniques would certainly enhance the application of DNA barcoding of

Table 1. List of plant specimens used in this study.

Family	Species
Apocynaceae	*Rhazya stricta*
Asteraceae	*Achillea fragrantissima*
	Anthemis deserti
	Pulicaria undulate
	Scorzonera intricate
	Sonchus oleraceus
Brassicaceae	*Horwoodia dicksoniae*
	Moricandia sinaica
	Zilla spinosa
Caryophyllaceae	*Spergularia bocconii*
Chenopodiaceae	*Bassia eriophora*
	Chenopodium murale
	Salsola imbricate
Convolvulaceae	*Convolvulus oxyphyllus*
Euphorbiaceae	*Andrachene telephoides*
Geraniaceae	*Erodium glaucophyllum*
	E. laciniatum
Malvaceae	*Malva parviflora*
Papilionaceae	*Astragalus sieberi*
	Melilotus indica
Poaceae	*Panicum antidotale*
	Pennisetum setaceum
Resedaceae	*Caylusea hexagyna*
Solanaceae	*Lycium shawii*
	Withania somnifera
Zygophyllaceae	*Zygophyllum propinquum*

Table 2. Primers used for amplification of *matK* and *rbcL* gene segments.

Primer set	Primer name	Binding	Primer sequence (5´– 3´)
matK-A	matK-2.1F	Forward	CCTATCCATCTGGAAATCTTAG
	matK-5R	Reverse	GTTCTAGCACAAGAAAGTCG
matK-B	matK-390F	Forward	CGATCTATTCATTCAATATTTC
	matK-1326R	Reverse	TCTAGCACACGAAAGTCGAAGT
rbcL-A	rbcLaF	Forward	ATGTCACCACAAACAGAGACTAAAGC
	rbcLaR	Reverse	GTAAAATCAAGTCCACCRCG
rbcL-B	rbcL-1F	Forward	ATGTCACCACAAACAGAAAC
	rbcL-724R	Reverse	TCGCATGTACCTGCAGTAGC

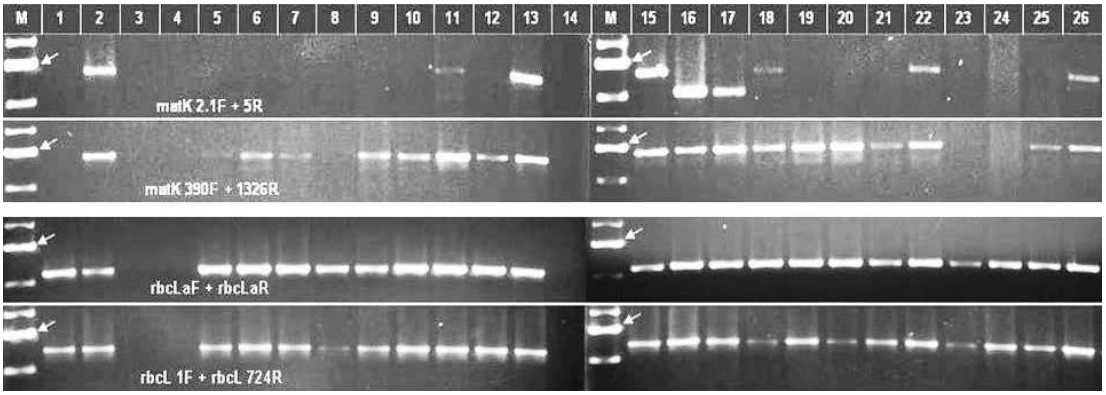

Fig 1. Agarose get electrophoretogram showing bands of PCR amplified products of matK and *rbcL* universal primers. M, marker 100 bp ladder; arrows indicate the 800 bp size of the marker; 1. *Horwoodia dicksoniae*, 2. *Rhazya stricta*, 3. *Pulicaria undulate*, 4. *Anthemis deserti*, 5. *Andrachene telephoides*, 6. *Zilla spinosa*, 7. *Caylusea hexagyna*, 8. *Achillea fragrantissima*, 9. *Lycium shawii*, 10. *Moricandia sinaica*, 11. *Bassia eriophora*, 12. *Zygophyllum propinquum*, 13. *Withania somnifera*, 14. *Sonchus oleraceus*, 15. *Chenopodium murale*, 16. *Malva parviflora*, 17. *Salsola imbricate*, 18. *Scorzonera intricate*, 19. *Pennisetum setaceum*, 20. *Panicum antidotale*, 21. *Erodium laciniatum*, 22. *Erodium glaucophyllum*, 23. *Convolvulus oxyphyllus*, 24. *Astragalus sieberi*, 25. *Spergularia bocconii*, 26. *Melilotus indica*.

plants and development of genetic databases for more efficient utilization of this technique.

Materials and methods

Plant material

This study comprised of specimens from 26 plant species covering 14 families (Table 1). The leaf samples were individually placed in plastic pouches and transported to lab where all the samples were stored at -80^{o}C until processed for DNA extraction.

DNA extraction

The leaf samples were immersed in liquid nitrogen and crushed using sterile mortal and pestle to get a fine powder. DNeasy plant mini kit (Qiagen) and an automated DNA extraction instrument (QIAcube, Qiagen) were used for DNA isolation. Quality of the extracted DNA was determined using gel electrophoresis and Nanodrop 8000 Spectrophotometer (Thermo Scientific, Wilmington, USA). Isolated plant genomic DNA was preserved at -20^{o}C.

PCR

A total volume of 30 µL of PCR reaction mixture contained the following: 15 µL of FideliTaq PCR Master Mix (USB Corporation, Cleveland, OH), giving a final concentration of 200 µM each deoxynucleotide and 1.5 mM $MgCl_2$, 1 µM each primer (Table 2, Eurofins MWG Operon, Germany}, 2 µL of genomic DNA and the rest was adjusted with sterile distilled water. PCR amplification was performed with a Veriti 96 well thermal cycler (Applied Biosystems) as follows: 95°C for 1 min, followed by 35 cycles of 95°C for 30 s, T°C for 30 s and 68°C for 1 min, followed by an elongation step at 68°C for 5 min. All the PCR conditions were the same for all the primer-pairs except the annealing temperature (T) for different primer pairs as follows: 50^{o}C for matK-A, 45^{o}C for matK-B, 51^{o}C for rbcL-A and 48^{o}C for rbcL-B (Table 2).

Agarose gel electrophoresis

A long (20 x 14 cm) 1% agarose gel using 1x TAE buffer containing 0.5 µg/mL ethidium bromide was used for electrophoresis of PCR-products. Gel images were obtained using Proxima C16 Phi+ (Isogen Life Science) UV transilluminator and Opticom (version 3.2.5; OptiGo) imaging system. The amplified PCR products were determined on gel for the presence or absence of the band. The size of PCR products resulting from the primer pairs of the specific barcoding gene were determined by using an Amersham 100-bp ladder (GE Healthcare) and the TotalLab TL100 1D software (version 2008.01).

Conclusion

The findings of this preliminary study indicate that currently used universal primers of *rbcL* and *matK* are able to amplify several of the plant species of arid origin with the amplification success rates of 88% and 69% respectively. All the primer pairs produced sharp bands that are suitable for

gene sequencing. Notwithstanding their failure in certain cases, the currently used primers for *rbcL* and *matK* are useful for plant barcoding until the discovery of more efficient and robust primers for a broader coverage of plant species. Thus, there is a need for protocol development to enhance the amplification strategies including the development of new primers or primer cocktails for enhanced success in barcoding of plant species of different regions.

Acknowledgments

The authors extend their appreciation to the Deanship of Scientific Research at King Saud University, Riyadh, Saudi Arabia for funding the work through the research group project No. RGP-VPP-009.

References

Altschul S, Madden T, Scha¨ffer A, Zhang, J, Zhang, Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402.

Aras S, Duran A, Yenilmez G (1993) Isolation of DNA for RAPD analysis from dry leaf material of some *Hesperis* L. specimens. Plant Mol Biol Rep 21: 461a-461f.

Asahina H, Shinozaki J, Masuda K, Morimitsu Y, Satake M (2010) Identification of medicinal Dendrobium species by phylogenetic analyses using *matK* and *rbcL* sequences. J Nat Med 64: 133-138.

Baker, CS, Palumbi SR (1994) Which whales are hunted—a molecular genetic approach to monitoring whaling. Science 265: 1538–1539.

Bokhari UG, Alyaeesh F, Al-Noori M (1990) Nutritional characteristics of important desert grasses in Saudi Arabia. J Range Manag 43: 202-204.

Burgess KS, Fazekas AJ, Kesanakurti PR, et al (2011) Discriminating plant species in a local temperate flora using the *rbcL+ matK* DNA barcode. Meth Ecol Evol doi: 10.1111/j.2041-210X.2011.00092.x.

Bru D, Laurent FM, Philippot L (2008) Quantification of the detrimental effect of a single primer-template mismatch by real-time PCR using the 16s rRNA gene as an example. Appl Environ Microbiol 74: 1660–1663.

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

Dawnay N, Ogden R, McEwing R, Carvalho GR, Thorpe, RS (2007) Validation of the barcoding gene COI for use in forensic genetic species identification. Forensic Sci Int 173: 1–6.

de-Groot GA , During HJ, Maas JW, Schneider H, Vogel JC, Erkens RHJ (2011) Use of *rbcL* and trnL-F as a two-locus DNA barcode for identification of NW European ferns: an ecological perspective. PLoS ONE 6: e16371. doi:10.1371/journal.pone.0016371.

Duffy AM, Kelchner SA, Wolf PG (2009) Conservation of selection on *matK* following an ancient loss of its flanking intron. Gene 438: 17–25.

Fox GE, Wisotzkey JD, Jurtshuk PJr (1992) How close is close: 16s rRNA sequence identity may not be sufficient to guarantee species identity. Int J Syst Bacteriol 42: 166-170. DOI 10.1099/00207713-42-1-166.

Hollingsworth ML, Clark AA, Forrest LL, 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.

Jones CJ, Edwards KJ, Castaglione S, et al (1997) Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Mol Breed 3: 381-390.

Kadkhodaei S, Elahy M, Nekouei MK, Imani A, Shahnazari M, Mardi M, Javanmard A, Ariff AB (2010) A panel of cultivate specific marker based on polymorphisms at microsatellite markers for Iranian cultivated Almonds (*Prunus dulcis*). Aust J Crop Sci 4: 730-736.

Kamal AHM, Kim KH, Shin KH, Choi JS, Baik BK, Tsujimoto H, Heo HY, Park CS, Woo SH (2010) Abiotic stress responsive proteins of wheat grain determined using proteomics technique. Aust J Crop Sci 4:196-208.

Khanuja SPS, Shasany AK, Darokar MP, Kumar S (1999) Rapid isolation of DNA from dry and fresh samples of plants producing large amounts of secondary metabolites and essential oils. Plant Mol Biol Rep 17: 1–7.

Kumar P, Gupta VK, Misra AK, Modi DR, Pandey BK (2009) Potential of molecular markers in plant biotechnology. Plant Omics J 2: 141-162.

Kuo LY, Li FW, Chiou WL, Wang CN (2011) First insights into fern *matK* phylogeny. Mol Phylogenet Evol 59: 556-566.

Manen JF, Habashi C, Jeanmonod D, Park JM, Schneeweiss GM (2004) Phylogeny and intraspecific variability of holoparasitic Orobanche (Orobanchaceae) inferred from plastid *rbcL* sequences. Mol Phylogenet Evol 33: 482-500.

Panero JL, Funk VA (2008) The value of sampling anomalous taxa in phylogenetic studies: major clades of the Asteraceae revealed. Mol Phylogenet Evol 47: 757-782.

Rahman MS, Molla MR, Alam MS, Rahman L (2009) DNA fingerprinting of rice (*Oryza sativa* L.) cultivars using microsatellite markers. Aust J Crop Sci 3: 122-128.

Roy S, Tyagi A, Shukla V, et al (2010) Universal plant DNA barcode loci may not work in complex groups: a case study with Indian *Berberis* species. PLoS One 5: e13674.

Schneider H, Schuettpelz E (2006) Identifying fern gametophytes using DNA sequences. Mol Ecol Notes 6: 989–991.

Soininen EM, Valentini A, Coissac E, et al (2009) Analyzing diet of small herbivores: the efficiency of DNA barcoding coupled with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures. Front Zool 6: 16.

Starr JR, Naczi RF, Chouinard BN (2009) Plant DNA barcodes and species resolution in sedges (Carex, Cyperaceae). Mol Ecol Resour 9 Suppl s1: 151-163.

Tamura MN, Fuse S, Azuma H, Hasebe M. Biosystematic studies on the family Tofieldiaceae I (2004) Phylogeny and circumscription of the family inferred from DNA sequences of *matK* and *rbcL*. Plant Biol 6: 562-567.

Temiesak P, Ponpim Y, Harada T (1993) RAPD analysis for varietal identification in *Brassica*. Kasetsart J Nat Sci 27: 37-42.

Vanijajiva O, Sirirugsa P, Suvachittanont W (2005) Confirmation of relationships among *Boesenbergia* (Zingiberaceae) and related genera by RAPD. Biochem Syst Ecol 33: 159-170.

von-Wintzingerode F, Gobel UB, Stackebrandt E (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS Microbiol Rev 21: 213–229.