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

DNA barcoding of arid wild plants using *rbcL* gene sequences

S.O. Bafeel[1], I.A. Arif[2,3], M.A. Bakir[2,3], A.A. Al Homaidan[2,3], A.H. Al Farhan[2,3] and H.A. Khan[2,3]

[1]Department of Biology, College of Science, King Abdulaziz University, Jeddah, Saudi Arabia
[2]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

Corresponding author: H.A. Khan
E-mail: khan_haseeb@yahoo.com

Genet. Mol. Res. 11 (3): 1934-1941 (2012)
Received February 6, 2012
Accepted June 22, 2012
Published July 19, 2012
DOI http://dx.doi.org/10.4238/2012.July.19.12

ABSTRACT. DNA barcoding is currently gaining popularity due to its simplicity and high accuracy as compared to the complexity and subjective biases associated with morphology-based identification of taxa. The standard chloroplast DNA barcode for land plants recommended by the Consortium for the Barcode of Life (CBOL) plant working group needs to be evaluated for a wide range of plant species. We therefore tested the potential of the *rbcL* marker for the identification of wild plants belonging to diverse families of arid regions. Maximum likelihood tree analysis was performed to evaluate the discriminatory power of the *rbcL* gene. Our findings showed that using *rbcL* gene sequences enabled identification of the majority of the samples (92%) to genus level and only 17% to species level.

Key words: DNA barcoding; *rbcL*; Wild plants; Identification; Phylogenetics

INTRODUCTION

Based on assessments of recoverability, sequence quality, and levels of species discrimination, the Consortium for the Barcode of Life (CBOL) plant working group has recommended a standard barcode comprising ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (*rbcL*) and/or maturase K (*matK*) for the barcoding of all land plants (CBOL Plant Working Group, 2009). However, the universality of barcode markers is hampered due to morphological/geographical variation and reticulate evolution in plant species (Roy et al., 2010). The ongoing research on plant barcoding suggests that the development of universal DNA barcoding markers for land plants is challenging; even the choice of the correct loci has been debated (Chase et al., 2005; Kress et al., 2005; Fazekas et al., 2008; de Groot et al., 2011). Arguments about the selected core loci for plant barcoding are related to the lack of discriminatory power and/or primer universality (Roy et al., 2010). Plant species of the desert are adapted to tolerate multiple stresses, including high extremes of drought, temperature, solar radiation, wind, and salinity (Batanouny, 2001). Constitution of seed bank of viable seeds available for potential germination and recruitment of new plants is important for plant conservation (Baker, 1989). During severe drought conditions or under severe disturbances, a persistent seed bank reduces the chance of extinction of a population at a site (Bakker and Berendse, 1999). Recently, it was determined that approximately 35% of the plant species that constitute the standing vegetation in the Red Sea area are not represented in the seed bank and are potentially vulnerable to elimination (Hegazy et al., 2009). Appropriate measures for the preservation of these desert plant species are therefore urgently needed.

Traditional methods based on morphological criteria are difficult to apply accurately due to subjective biases. Particularly, in the case of medicinal plants, the use of chromatographic profiles of marker compounds to standardize botanical preparations is also of limited value because the medicines have varied sources and chemical complexity, which is affected by growth, storage conditions, and harvest times (Joshi et al., 2004; Zhang et al., 2007). DNA-based identification (barcoding) is simple, does not require taxonomic expertise, and is free from subjective errors, which is not the case in morphological identification. Valid identification of unknown samples is the main goal of barcoding (Hebert and Gregory, 2005), despite ongoing criticism of the feasibility or even necessity of DNA barcoding for general taxonomic purposes (Will et al., 2005; Spooner, 2009). Nowadays, it is widely accepted that any valid plant barcode should be multi-locus, preferably comprising a conserved coding region such as *rbcL* and a more rapidly evolving region that is most likely non-coding (Kress et al., 2009). Sequences of the *rbcL* and *trnL-F* genes as a two-locus DNA barcode have recently been used successfully to identify NW-European ferns, whereas the selected *matK* locus was unsuccessful for barcoding (de Groot et al., 2011). However, whether *rbcL* exhibits sufficient variation to allow general identification of wild plants grown in arid environments below genus level remains unexplored. In continuation of our previous study on the PCR success rate (Bafeel et al., 2011) and molecular characterization of desert medicinal plants (Arif et al., 2010), we evaluated the barcoding performance of *rbcL* for the identification of Saudi Arabian wild plants and demonstrated genus- and species-level discriminations using this marker.

MATERIAL AND METHODS

DNA extraction

Plant leaf samples from 12 different species were used. The specimens were macerated using a sterile mortar and pestle under liquid nitrogen. A DNeasy Plant Mini Kit (Qiagen, Germany) and an automated DNA extraction instrument (QIAcube, Qiagen) were used to isolate the DNA. The concentration and quality of the extracted DNA were determined using gel electrophoresis and a NanoDrop 8000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA). The isolated genomic DNA was stored at -20°C until used.

PCR and gene sequencing

A total volume of 30 µL PCR master mixture contained the following: 15 µL 2X FideliTaq PCR Master Mix (USB Corporation, Cleveland, OH, USA), providing a final concentration of 200 µM of each deoxynucleoside triphosphate and 1.5 mM $MgCl_2$, 1 µM of each primer (Eurofins MWG Operon, Germany), and 25-500 ng genomic DNA of each plant sample, with the remaining volume topped-up with sterile distilled water. The primer pairs rbcLaF (5'-ATG TCA CCA CAA ACA GAG ACT AAA GC-3') and rbcLaR (5'-GTA AAA TCA AGT CCA CCR CG-3'), and rbcL 1F (5'-ATG TCA CCA CAA ACA GAA AC-3') and rbcL 724R (5'-TCG CAT GTA CCT GCA GTA GC-3') were used for the PCR.

The PCR 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, 51°C for rbcLaF-rbcLaR and 48°C for rbcL 1F-rbcL 724R for 30 s, and 68°C for 1 min, followed by an elongation step at 68°C for 5 min. A long (20 x 14 cm) 1% agarose gel using 1X TAE buffer containing 0.5 µg/mL ethidium bromide was used for PCR product electrophoresis. Gel images were obtained using a Proxima C16 Phi+ (Isogen Life Science) UV transilluminator and Opticom (version 3.2.5; OptiGo) imaging system. The PCR product sizes were determined using a 100-bp ladder (GE Healthcare) and the TotalLab TL100 1D software (version 2008.01).

PCR products were purified using a QIAquick PCR Purification Kit (Qiagen) before being sequenced using the dideoxynucleotide chain termination method with a DNA sequencer (ABI 3130XL, Applied Biosystems) and a BigDye Terminator version 3.1 Cycle Sequencing RR-100 Kit (Applied Biosystems). All sequences were submitted to GenBank, USA (accession Nos. JN375994 and JN376005).

Assignment of taxa

BLAST searches were applied to all produced sequences using the available online databases (DDBJ/EMBL/GenBank). BLAST was never intended to be used in this manner, but could provide valuable insights into how well we can expect the possibly more appropriate plastid *matK* and *rbcL* short sequence regions to perform as barcodes (Chase et al., 2005). There are very few *rbcL* records on the current BOLD (Barcode of Life Data) identification system (v 2.5) (Ratnasingham and Hebert, 2007); thus, queries might not return an authentic match. Identification at genus level was considered successful when all hits with maximal percent identity scores >95% involved a single genus. Species identification was considered successful only when the highest maximal percent identity included a single species and scored >95% (de Groot et al., 2011). The *rbcL* sequences were matched with the query sequences and available *rbcL* sequences of the examined plant species; if not available, then genera were retrieved from the DDBJ/EMBL/GenBank databases.

The sequences were aligned using CLUSTAL X version 1.81 (Thompson et al., 1997). Phylogenetic analyses were conducted in MEGA5 (Tamura et al., 2007), and the phylogenetic trees were inferred with the maximum likelihood method based on the Tamura-Nei model (Tamura and Nei, 1993). The topologies of the phylogenetic trees were evaluated using the bootstrap resampling method of Felsenstein (1985) with 1000 replicates. In phylogenetic analyses, genus identification was considered successful when the unknown sample formed a monophyletic group together with all members of a single genus, with a bootstrap support of >70%. An equal strategy was applied for species-level identification (de Groot et al., 2011).

RESULTS AND DISCUSSION

When the overall outputs of BLAST matching and tree analysis were compared, the latter strategy resulted in better taxonomic assignment. The use of *rbcL* sequences with BLAST searching yielded 50 and 8% genus- and species-level identifications, respectively (Table 1). Tree analyses with the *rbcL* gene sequences assigned the majority of samples (92%) up to genus level and 17% up to species level (Table 1; Figure 1). In phylogenetic analyses, we considered genus identification successful when the unknown sample formed a monophyletic group together with all members of a single genus, with a bootstrap support of >70%. An equal strategy was applied for species-level identification (de Groot et al., 2011). Tree analyses using *rbcL* sequences assigned 17% of the tested plant samples to known species. Our findings, notwithstanding *rbcL* is considered to possess less species-discriminating power than *matK*, are possibly due to its minimal sequence variation (Asahina et al., 2010). The estimated range of the total number of plant species worldwide is believed to be approximately 310,000-422,000 (Graham, 2002). When the data analyses of this experiment were carried out, the DDBJ/EMBL/GenBank databases contained only 8289 nucleotide sequences of the *matK* gene and 12,909 nucleotide sequences of the *rbcL* gene of plant species. The availability of the sequences of barcoding genes in the databases is expected to increase rapidly, and subsequently, their utilization in the identification of plant species.

Table 1. Database search match for similarities and phylogenetic relationship using *rbcL* gene sequences.

Sample ID	Morphological identification	BLAST search match	BLAST similarity (%)	Phylogenetic affinity
C6	*Rhazya stricta*	*Alstonia macrophylla*	98	*Rhazya* sp
		Rhazya stricta	98	
C14	*Lycium shawii*	*Lycium chinense*	100	*Lycium* sp
C15	*Moricandia sinaica*	*Brassica napus*	99	*Moricandia* sp
		Brassica oleracea	99	
C17	*Bassia eriophora*	*Bassia scoparia*	99	*Bassia* sp
C19	*Withania somnifera*	*Solanum tampicense*	99	*Withania somnifera*
		Solanum panduriforme	99	
C21	*Chenopodium murale*	*Chenopodium bonus-henricus*	99	*Chenopodium* sp
C23	*Salsola imbricata*	*Salsola vermiculata*	100	*Salsola* sp
		Salsola orientalis	100	
		Salsola dshungarica	100	
C25	*Scorzonera intricate*	*Hecastocleis shockleyi*	97	*Scorzonera* sp
		Scolymus maculatus	97	
C28	*Panicum antidotale*	*Pennisetum purpureum*	98	*Panicum* / *Cenchus* sp
		Panicum virgatum	98	
		Panicum stoloniferum	98	
C29	*Erodium laciniatum*	*Erodium malacoides*	99	*Erodium* sp
C30	*Erodium glaucophyllum*	*Erodium glaucophyllum*	99	*Erodium glaucophyllum*
C88	*Melilotus indica*	*Melilotus alba*	100	*Melilotus* sp

BLAST = Basic local alignment search tool.

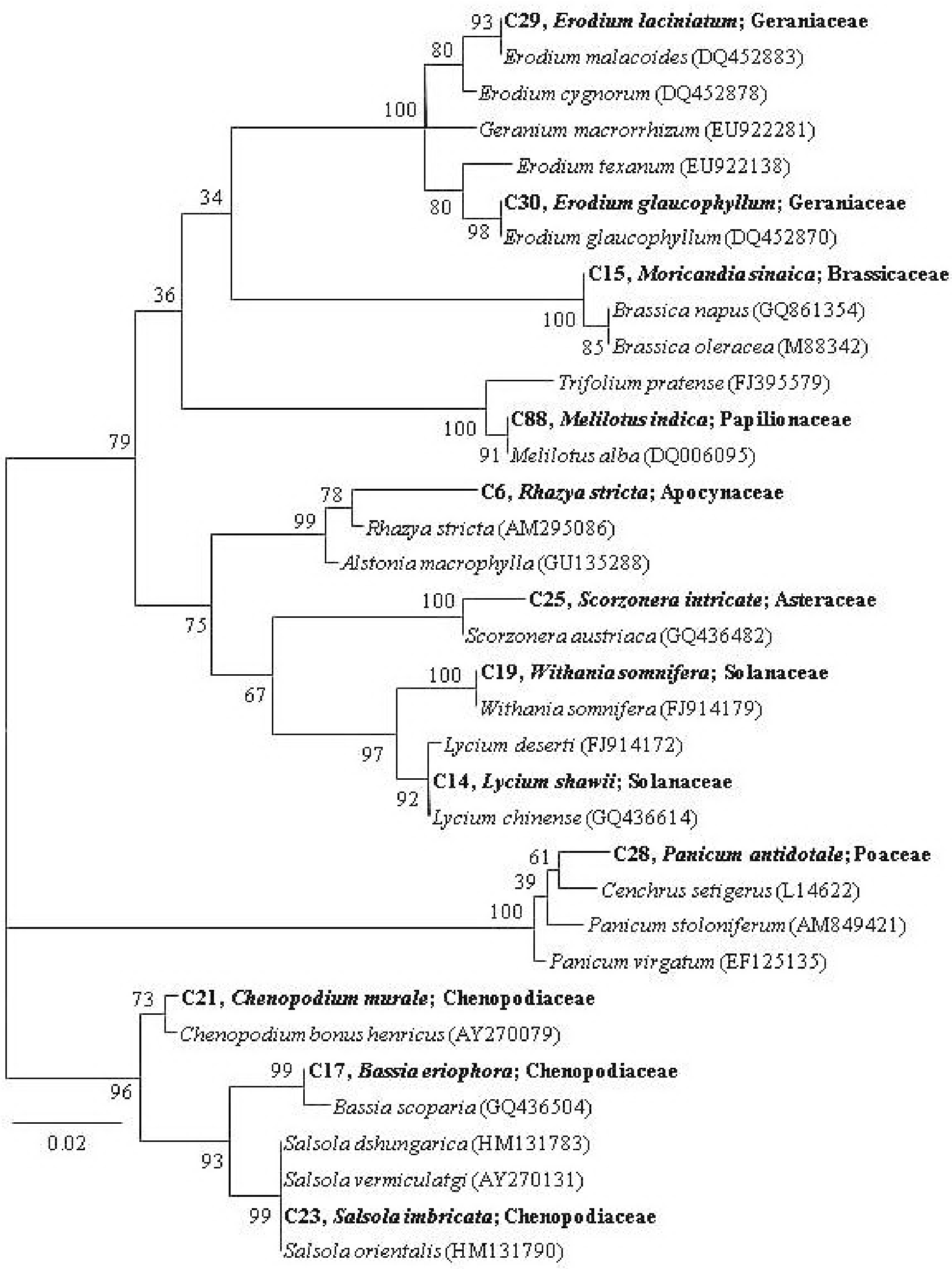

Figure 1. Phylogenetic affinities of *rbcL* gene sequences of the plant samples. Number of the samples, morphological identification and family names are typed in bold. The evolutionary history was inferred by using the maximum likelihood method based on the Tamura-Nei model (1993). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Bootstrap values based on 1000 replications are listed as percentages at branching points.

The clades in the tree constructed with *rbcL* gene sequences were supported by 89% (25/28) of >50% bootstrap values (Figure 1). In a recent study, the *rbcL* marker exhibited intermediate-level (80%) resolution among the evaluated regions (*matK* > *atpF*-*atpH* > *rbcL* > *trnH*-*psbA* > *rpoC1*) (Burgess et al., 2011). Phylogenetic methods were applied in a recently conducted study of barcoding species using each barcode locus taken alone and in combinations to evaluate species recovery (Roy et al., 2010). The NJ, MP, and UPGMA methods were used for both single- and multi-locus analyses with 500 bootstrap replicates. When all sequences for a given locus were considered, ITS, *matK*, and *trnH*-*psbA* were able to form a species-specific clade for only *Berberis pachyacantha*. Not a single species was recovered with *rbcL* using any of the three methods. The clades formed in the trees were mostly mixtures of several species. Therefore, establishing a local barcode database will be valuable for a broad range of potential ecological applications, including the building of community phylogenies (Kress et al. 2009).

Morphological identification is inapplicable when studying population biology. In such cases, barcoding is an efficient and valuable technique. Some ecologists have started using the barcoding approach to identify specific unknown plant samples for practical purposes (Li et al., 2009; Van de Wiel et al., 2009). Ongoing developments of new primers and improvements in sequencing techniques have facilitated the data-emergence process of plant barcoding (Soltis et al., 1996; Plunkett et al., 1997; Van de Wiel et al., 2009; Burgess et al., 2011). Recently, plant diversity belowground was determined using *rbcL* gene sequences as a core plant DNA barcoding marker (Kesanakurti et al., 2011). Tsukaya et al. (2011) described a new genus based on DNA sequences of the chloroplast *matK* pseudogene and ITS of the nuclear ribosomal DNA. The generation of *matK* sequences for some plant groups has been reported to be problematic, because this part of the chloroplast genome underwent a large-scale restructuring during evolution (Duffy et al., 2009; de Groot et al., 2011). None of the currently existing primer sets are likely suitable for all lineages of land plants (Hollingsworth et al., 2009; Li et al., 2009; Roy et al., 2010) and efforts are now focusing on the development of complex primer assays to achieve reliable amplification and sequencing of land plants.

In conclusion, this study provides preliminary assessment data that will be useful for wider application of DNA barcoding in ecological studies on desert plants. With the current development of primers, we found that *rbcL* will be very useful for the barcoding of plant species in Saudi Arabia. However, further protocol development to enhance clean DNA extraction, PCR amplification strategies, including the development of new primers, and local authenticated databases could play important roles in efficient utilization of plant barcoding.

ACKNOWLEDGMENTS

The authors extend their appreciation to the Deanship of Scientific Research at King Saud University for funding the research through project #RGP-VPP-009. We are thankful to Dr. Jacob Thomas for sample collection and identification and Anis Ahamed for DNA extraction.

REFERENCES

Arif IA, Bakir MA, Khan HA, Al Farhan AH, et al. (2010). Application of RAPD for molecular characterization of plant species of medicinal value from an arid environment. *Genet. Mol. Res.* 9: 2191-2198.

Asahina H, Shinozaki J, Masuda K, Morimitsu Y, et al. (2010). Identification of medicinal *Dendrobium* species by phylogenetic analyses using matK and rbcL sequences. *J. Nat. Med.* 64: 133-138.
Bafeel SO, Arif IA, Bakir MA, Khan HA, et al. (2011). 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. *Plant Omics J.* 4: 195-198.
Baker HG (1989). Some Aspects of the Natural History of Seed Banks. In: Ecology of Soil Seed Banks (Leck MA, Parker VT and Simpson RL, eds.). Academic Press, San Diego, 9-21.
Bakker JP and Berendse F (1999). Constraints in the restoration of ecological diversity in grassland and heathland communities. *Trends Ecol. Evol.* 14: 63-68.
Batanouny KH (2001). Plants in the Deserts of the Middle East. Springer, New York.
Burgess KS, Fazekas AJ, Kesanakurti PR, Graham SW, et al. (2011). Discriminating plant species in a local temperate flora using the rbcL+matK DNA barcode. *Method Ecol. Evol.* 2: 333-340.
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. Lond. B Biol. Sci.* 360: 1889-1895.
de Groot GA, During HJ, Maas JW, Schneider H, et al. (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.
Duffy AM, Kelchner SA and Wolf PG (2009). Conservation of selection on matK following an ancient loss of its flanking intron. *Gene* 438: 17-25.
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.
Felsenstein J (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783-791.
Graham S (2002). Global estimate of endangered plant species triples. Scientific American. Available at [http://www.scientificamerican.com/article.cfm?id=global-estimate-of-endang]. Accessed January 10, 2012.
Hebert PD and Gregory TR (2005). The promise of DNA barcoding for taxonomy. *Syst. Biol.* 54: 852-859.
Hegazy AK, Hammouda O, Lovettt-Doust J and Gomaa NH (2009). Variations of the germinable soil seed bank along the altitudinal gradient in the northwestern Red Sea region. *Acta Ecol. Sin.* 29: 20-29.
Hollingsworth ML, Andra CA, 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.
Joshi K, Chavan P, Warude D and atwardhan B (2004). Molecular markers in herbal drug technology. *Curr. Sci.* 87: 159-165.
Kesanakurti PR, Fazekas AJ, Burgess KS, Percy DM, et al. (2011). Spatial patterns of plant diversity below-ground as revealed by DNA barcoding. *Mol. Ecol.* 20: 1289-1302.
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.
Kress WJ, Erickson DL, Jones FA, Swenson NG, et al. (2009). Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proc. Natl. Acad. Sci. U. S. A.* 106: 18621-18626.
Li FW, Tan BC, Buchbender V, Moran RC, et al. (2009). Identifying a mysterious aquatic fern gametophyte. *Plant Syst. Evol.* 281: 77-86.
Plunkett G, Soltis D and Soltis P (1997). Clarification of the relationship beteen Apiaceae and Araliaceae based on matK and rbcL sequence data. *Am. J. Bot.* 84: 565.
Ratnasingham S and Hebert PDN (2007). BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Mol. Ecol. Note.* 7: 355-364.
Roy S, Tyagi A, Shukla V, Kumar A, 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.
Soltis DE, Kuzoff RK, Conti E, Gornall R, et al. (1996). matK and rbcL gene sequence data indicate that *Saxifraga* (Saxifragaceae) is polyphyletic. *Am. J. Bot.* 83: 371-382.
Spooner DM (2009). DNA barcoding will frequently fail in complicated groups: An example in wild potatoes. *Am. J. Bot.* 96: 1177-1189.
Tamura K and Nei M (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10: 512-526.
Tamura K, Dudley J, Nei M and Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24: 1596-1599.
Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, et al. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25: 4876-4882.
Tsukaya H, Nakajima M and Okada H (2011). *Kalimantanorchis*: A new genus of Mycotrophic orchid from West Kalimantan, Borneo. *Syst. Bot.* 36: 49-52.

Van de Wiel CC, Van Der Schoot J, Van Valkenburg JL, Duistermaat H, et al. (2009). DNA barcoding discriminates the noxious invasive plant species, floating pennywort (*Hydrocotyle ranunculoides* L.f.), from non-invasive relatives. *Mol. Ecol. Resour*. 9: 1086-1091.

Will KW, Mishler BD and Wheeler QD (2005). The perils of DNA barcoding and the need for integrative taxonomy. *Syst. Biol.* 54: 844-851.

Zhang YB, Shaw PC, Sze CW, Wang ZT, et al. (2007). Molecular authentification of Chinese herbal materials. *J. Food Drug Anal.* 15: 1-9.