DNA barcoding of selected desert medicinal species of Fabaceae and Poaceae

Aisha Tahir Corresp., [1] , **Fatma Hussain** [1] , **Nisar Ahmed** [2] , **Abdolbaset Ghorbani** [3] , **Amer Jamil** Corresp. [1]

[1] Department of Biochemistry, Faculty of Science, University of Agriculture, Faisalabad, Pakistan

[2] Centre of Agricultural Biochemistry and Biotechnology, University of Agriculture, Faisalabad, Pakistan

[3] Department of Organismal Biology, Uppsala Universitet, Uppsala, Sweden

Corresponding Authors: Aisha Tahir, Amer Jamil
Email address: aishatahir85@gmail.com, profamerjamil@gmail.com

In pursuit of developing fast and accurate species level molecular identification methods, we tested three DNA barcodes viz. ITS2, *matK* and *rbcLa* for their capacity to identify frequently consumed medicinal species of Fabaceae and Poaceae indigenous to the desert of Cholistan. Data were analysed by BLASTn sequence similarity, pairwise sequence divergence in TAXONDNA, and phylogenetic (neighbour-joining and maximum-likelihood trees) methods. Comparison of three barcode regions showed that ITS2 has the highest number of variable sites (209/360) for tested Fabaceae and (106/365) Poaceae species, the highest species level identification (40%) in BLASTn procedure, distinct DNA barcoding gap, 100% correct species identification in BM and BCM functions of TAXONDNA, and clear cladding pattern with high nodal support in phylogenetic trees in both families. ITS2 was followed by *matK* in its species level identification capacity (37%) and then *rbcLa* (30%). The study was concluded with advocating the DNA barcoding as an effective tool for species identification and ITS2 as the best and *matK* as the second best barcode region in identifying medicinal species of Fabaceae and Poaceae. Current research has practical implementation potential in the fields of pharmaco-vigilance, trade of medicinal plants and biodiversity conservation.

DNA Barcoding of Selected Desert Medicinal Species of Fabaceae and Poaceae

Aisha Tahir[1], Fatma Hussain[1], Nisar Ahmed[2], Abdolbaset Ghorbani[3], Amer Jamil[1]

[1] Department of Biochemistry, Faculty of Science, University of Agriculture, Faisalabad, Pakistan

[2] Centre of Agricultural Biochemistry and Biotechnology, University of Agriculture, Faisalabad, Pakistan

[3] Department of Organismal Biology, Uppsala University, Uppsala, Sweden

Corresponding Auhor:

Amer Jamil

Department of Biochemistry, Faculty of Science, University of Agriculture, Faisalabad, Pakistan

Email address: amerjamil@yahoo.com

Abstract: In pursuit of developing fast and accurate species level molecular identification methods, we tested three DNA barcodes viz. ITS2, *matK* and *rbcLa* for their capacity to identify frequently consumed medicinal species of Fabaceae and Poaceae indigenous to the desert of Cholistan. Data were analysed by BLASTn sequence similarity, pairwise sequence divergence in TAXONDNA, and phylogenetic (neighbour-joining and maximum-likelihood trees) methods. Comparison of three barcode regions showed that ITS2 has the highest number of variable sites (209/360) for tested Fabaceae and (106/365) Poaceae species, the highest species level identification (40%) in BLASTn procedure, distinct DNA barcoding gap, 100% correct species identification in BM and BCM functions of TAXONDNA, and clear cladding pattern with high nodal support in phylogenetic trees in both families. ITS2 was followed by *matK* in its species level identification capacity (37%) and then *rbcLa* (30%). The study was concluded with advocating the DNA barcoding as an effective tool for species identification and ITS2 as the best and *matK* as the second best barcode region in identifying medicinal species of Fabaceae and Poaceae. Current research has practical implementation potential in the fields of pharmaco-vigilance, trade of medicinal plants and biodiversity conservation.

Key words: DNA barcoding, medicinal plants, species identification, internal transcribed spacer region (ITS2), maturase K (*matK*), ribulose bisphosphate carboxylase large chain (*rbcLa*)

1. Introduction

Many species of plants belonging to multiple families are catalogued as medicinal plants on the basis of the presence of specific chemical constituents and their effects on the biological systems (Ahmad et al., 2014). Fabaceae and Poaceae are among the largest plant families having medically and therapeutically useful species all over the world (Gao et al., 2010; Dashora & Gosavi, 2013; Wariss et al., 2016). For instance, *Crotalaria burhia* of Fabaceae has antimicrobial, anti-inflammatory, wound healing, and antioxidant properties (Kataria et al., 2010). *Acacia*

nilotica of Fabaceae is used in tonics, and *Cenchrus ciliaris* of Poaceae has been reported anodyne, diuretic and emollient (Hameed et al., 2011; Wariss et al., 2013).

Excessive harvesting of medicinal plants is not only a threat to biodiversity but also leads to intentional and unintentional adulteration in herbal products due to unavailability of actual species and economical constraints (Sagar, 2014) as well as misidentifications due to superficial resemblance among species (Joharchi & Amiri, 2012). Conventional methods for species identification rely on the morphology only that prove inefficient when specimens are morphologically more similar but belong to entirely different taxa. In order to avoid the misidentification and adulteration, a simple, rapid and reliable identification method is inevitable. Methods of species identification from integrated specimens to processed products demand the incorporation of modern techniques and tools specifically if morphological characters are insufficient or unavailable for correct species assignment to unknown specimens (Newmaster & Ragupathy, 2009; Gathier et al., 2013; Mutanen et al., 2015; Ghorbani et al., 2017).

DNA barcoding is introduced in 2003 as a molecular based species identification tool by using a short, variable and standardized DNA region, the barcode (Hebert et al., 2003a; Hebert et al., 2003b; Hebert & Gregory, 2005). In order to meet the criteria of DNA barcode, a gene locus must possess enough species level genetic variability, short sequence length, and conserved flanking regions (Giudicelli et al., 2015). Common plant DNA barcodes proposed for plants are plastidial *matK*, *rbcL*, ITS, *rpoB* and *rpoC1*, the intergenic plastidial spacers (*trnH-psbA*, *atpF-atpH* and *psbK-psbI*) and the nuclear internal transcribed spacers that have been used singly or in combinations (Fazekas et al., 2008; Chen et al., 2010; De Mattia et al., 2011). Chloroplast DNA regions, *matK* and *rbcL,* are recommended as core barcode regions by Plant working group of CBOL (CBOL Plant Working Group, 2009) while nuclear DNA region, ITS2 which is a part of ITS, is also recommended as a potential marker for quick taxonomical classification in closely related species of wide range of taxa such as in Fabaceae, Lamiaceae, Asteraceae, Rutaceae,

Rosaceae and many more (Chen et al., 2010; Gao et al., 2010; Pang et al., 2011; Balachandran et al., 2015).

This study tests the hypothesis that molecular identification technique of DNA barcoding using three DNA barcodes named *matK*, *rbcLa*, and ITS2 can be applied for identification of selected medicinal species of Fabaceae and Poaceae that are indigenous to the desert of Cholistan. The main objective was to compare the universality and discrimination power of three standard barcode regions and to test the applicability of DNA barcoding for identification of selected medicinal species. The hypothesis is tested using evaluation criteria of Basic Local Alignment Search Tool (BLAST), pairwise genetic distance and sequence similarity test in TAXONDNA, and tree-based methods.

2. Materials and methods

2.1. Plant material

A total of 30 specimens belonging to 7 species of Fabaceae and 3 of Poaceae which are commonly used medicinal plants in herbal formulations were included in this study. Subfamilies of the species under consideration are not mentioned in this study. At least three individuals were sampled for each species from different locations of the Cholistan desert. All the specimens were identified taxonomically with the help of plant taxonomist Dr Mansoor Hameed at Department of Botany, University of Agriculture, Faisalabad using published flora and monographs (http://www.tropicos.org/Project/Pakistan). Voucher specimens are deposited at the Herbarium of Department of Botany, University of Agriculture, Faisalabad. Details of specimen and voucher numbers are given as supplemental information. The samples were collected from wild and locations did not include any park or protected area of land, neither the collection involved any endangered species.

2.2. DNA extraction, amplification and sequencing

Total genomic DNA was extracted from specimens by grinding silica-gel dried-leaf tissue in liquid nitrogen, and then using the CTAB procedure (White et al., 1990). Total genomic DNA was dissolved in TE buffer (10 mM Tris–HCl, pH 8.0, 1 mM EDTA) to a final concentration of 50 ng/µl.

Polymerase chain reaction (PCR) amplification of ITS2 and *rbcLa* regions was performed in 50 µl reactions containing 25 µl of 10% trehalose, 0.25 µl of Platinum Taq-polymerase (5 U/µl), 2.5 µl $MgCl_2$ (50 mM), 0.25 µl dNTPs (10 mM), 5.0 µl reaction buffer (10X), 0.5 µl of each primer (10 µM), 8.0 µl of ddH_2O and 8.0 µl of template DNA. PCR amplification of *matK* was performed in 50 µl reactions containing 14 µl of 20% trehalose, 1.2 µl Taq-polymerase (5 U/µl), 1.2 µl dNTPs (10 M), 5.5 µl reaction buffer (10X), 1.5 µl $MgCl_2$, 2.8 µl of each primer (10 µM), 1 µl of template DNA and 20.0 µl of ddH_2O. The primers' sequence information and optimal PCR conditions are displayed in Table 1 and Table 2 respectively. PCR products were examined by electrophoresis using 0.8% agarose gels. The PCR products were purified using FavorPrep™ PCR Clean-Up Mini kit and then were sequenced using the amplification primers.

All the DNA regions were sequenced by using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Inc., California, USA) according to the protocol provided in a GeneAmp PCR System 9700 thermal cycler. Quarter volume reactions were prepared with 0.5 µl sequencing premix and a 3.2 µM final concentration for the primers. The other components were 5X sequencing buffer and 3-20 ng PCR template. Standard cycling conditions were used [30 cycles of denaturation (30 sec @ 96°C); primer annealing (15 sec @ 58°C); extension (4 min @ 60°C)]. Cycle sequencing products were precipitated in ethanol and sodium acetate to remove excess dye terminators. Then they were again suspended into 10 µl HiDi formamide (ABI) before sequencing on an automated ABI 3130 *xl* Genetic Analyzer (ABI).

2.3. Data analysis

2.3.1. Editing and alignment of sequences

The software program Geneious R9.1 (www.geneious.com) was used to visualize, assemble and edit the sequence trace files. Consensus sequences were aligned with the MUSCLE (Edgar, 2004) plugin in Geneious R9.1. Alignments were then further refined by eye examination for resolving any gaps, insertions or deletion. Sequences were exported from Geneious R9.1 as aligned FASTA files for further analysis. The discriminatory power for all regions was assessed at genus and species-level by employing four analytical methods i.e., BLAST, the pairwise genetic distance method (PWG distance), the sequence similarity method (TAXONDNA) and phylogenetic-based method (Neighbor-Joining and Maximum Likelihood phylogenetic trees).

2.3.2. Analysis by BLAST procedure

All the newly acquired sequences were queried via BLASTn (http://blast.ncbi.nlm.nih.gov/Blast.cgi) against the online nucleotide database and further deposited in GenBank. BLAST was used to evaluate the species level identification power of three markers in the study. Aligned sequences were searched in National Centre for Biotechnology Information (NCBI) database through BLAST procedure (Altschul et al., 1990). Top matching hit having the highest (> 98%) maximal percent identity score was the criteria for successful conspecific/congeneric identification.

2.3.3. Pairwise genetic distance analysis

For the pairwise genetic-based method, average of inter-specific and intra-specific distances were calculated for both families separately in MEGA6 (Molecular Evolutionary Genetics Analysis Version 6.0) program (Tamura et al., 2013, http:// www.megasoftware.net) and TAXONDNA software using the Kimura-2-parameter (K2P) distance model to explore the intra- and interspecies variations. The pairwise intra- and interspecific distances were calculated for each species of both plant families. For each specie, the minimum interspecific distance was compared with its maximum intraspecific distance for the detection of barcoding gap (Meier et al., 2008; van Velzen et al., 2012).

124 **2.3.4. Sequence similarity analysis**

125 In the sequence similarity method, the species identification potential of all barcode regions was
126 assessed by calculating the proportion of correct identifications identified with the "Best Match"
127 (BM) and "Best Close Match" (BCM) tests in Species Identifier 1.8 program of TAXONDNA
128 software (Meier et al., 2006).

129 **2.3.5. Phylogenetic analysis**

130 In order to assess whether species are recovered as monophyletic groups, phylogenetic trees were
131 constructed in MEGA6 after appropriate model selection in the same software for each single
132 barcode region for all the species of both families. The species were categorized as successfully
133 identified if the monophyletic cluster of all the individuals of one species had the support nodes
134 above 75% for neighbor-joining (NJ) as well as for maximum-likelihood (ML) statistical
135 methods (Tang et al., 2015; Xu et al., 2015).

136 **3. Results**

137 **3.1. Amplification, sequence analysis, and genetic divergence**

138 The three commonly used barcoding loci performed differently in terms of universality for
139 amplification and sequencing in both families. Amplification success is 85%, 71% and 100% for
140 ITS2, *matK* and *rbcLa* respectively for Fabaceae and 100% for all regions for specimens of
141 Poaceae. Overall aligned length of the three regions ranged from 360 bp (ITS2) to 844 bp (*matK*)
142 for Fabaceae and from 365 bp (ITS2) to 772 bp (*matK*) for Poaceae. In this study, 18 sequences
143 of ITS2, 15 of *matK*, and 21 of *rbcLa* were generated from family Fabaceae and 27 sequences
144 (triplicate of each species with each region) from Poaceae. In addition, ITS2 had the highest
145 percentage of parsimony informative sites i.e. 56% (Fabaceae) and 29% (Poaceae), followed by
146 *matK* i.e. 12% (Fabaceae) and 3% (Poaceae) and *rbcLa* i.e. 8% (Fabaceae) and 3% (Poaceae)
147 (Table 3). Out of total seven medicinal species of Fabaceae, *Prosopis cineraria* was not amplified
148 with ITS2 while *Crotalaria burhia* and *Prosopis cineraria* both were not amplified with *matK*.

While comparing the markers in both families, *rbcLa* was the best at amplification and sequencing followed by ITS2 and *matK* while ITS2 had the highest percentage of variable and parsimony-informative sites and *rbcLa* had the lowest. The average intra- and interspecific divergence values in three barcoding markers in both families ranged from 0.00 to 0.02 and 0.02 to 0.35 respectively. *rbcLa* showed the lowest average intraspecific (0.00) and interspecific (0.02) divergence. While ITS2 showed the highest intraspecific (0.02%) as well as interspecific (0.35%) divergences. Average sequence divergence values for *matK* was slightly more than *rbcLa* but much less than ITS2 i.e. 0.00 for intraspecific and 0.07 for interspecific (Table 3).

In total, we generated 81 sequences (27 of ITS2, 24 of *matK*, and 30 of *rbcLa*) in this study. All of them are included in the analysis. Fifty six refined sequences and metadata of all the specimens are submitted to BOLD systems under the project named "DNA barcoding of medicinal plants of Pakistan (DBMPP)" as well as in GenBank (supplemental information).

3.2. DNA barcoding gap assessment

The relative distribution of the frequencies of K2P distances was calculated for the three candidate loci for the selected species of Fabaceae and Poaceae families included in the study using TAXONDNA software, thus barcoding gap was identified for three barcoding markers. Pairwise intra- and interspecific genetic distances showed similar overlapped pattern for *matK* and *rbcLa* while distances were not overlapping in case of ITS2. The discrimination power of a barcoding region was considered effective if the minimum interspecies distance was larger than its maximum intraspecies distance. Figure 1 is the illustration of the observed patterns in ITS2, *matK* and *rbcLa*.

3.3. Species identification using BLAST

rbcLa came up with the highest percentage of genus level identification while ITS2 leaded at species level identification. In this analysis, *Lasiurus scindicus* of Poaceae was an ambiguous sample among the collection because it did not match with expected genus or species with all

three markers while *Cymbopogon jwarancusa* of Poaceae did not match with expected genus/species with ITS2 but identified with other two markers. Overall, *rbcLa* was better at identifying unknown specimens up to genus level followed by *matK* and *ITS2* in both Fabaceae and Poaceae (Table 4).

3.4. Best match (BM) and best close match (BCM) analysis

The potential of the three barcoding regions for species identification accuracy was estimated by measuring the proportions of correct identifications using BM and BCM functions that evaluate the proportion of correct identifications through comparison of DNA sequences. One sequence is compared with all other sequences present in the dataset in the SpeciesIdentifier program of the TAXONDNA software package and then compared sequences were grouped on the basis of their pairwise genetic distances that ultimately determined the conspecificity of two sequences. K2P distance model was used in this analysis.

The closest match of a sequence was established by BM function. Identification is categorized as correct if compared sequences were from same species and incorrect if the closest sequences were from different species. If a sequence matches with both the sequences i.e. of same species and of different species with equally good similarity, then that sequence was considered ambiguous. The BCM function offered more stringent criteria by keeping a threshold of 95% pairwise distance in pairwise summary function. The queries above the threshold value were classified as “no match” and the others which are below the threshold value were analyzed according to the criteria established in “best match” analysis (Meier et al., 2006; Giudicelli et al., 2015).

The results of sequence similarity test performed in TAXONDNA software for three regions are presented in Table 5. Same proportion of success rate was observed for all markers in both TAXONDNA functions i.e. BM and BCM. Both families presented equal values (100%) of correct identifications with ITS2 and *matK* while *rbcLa* showed lower discriminatory power in

Fabaceae as six sample sequences were ambiguous. “Incorrect” and “no match” were 0% in both functions so they are not shown in Table 5. This analysis indicates that the ITS2 and *matK* both meet the rigorous standards for identifying the queries accurately among three selected markers.

3.5. Tree based analysis of barcoding regions

Kimura-2-parameter with 2.1 gamma distribution (K2G), Tamura-3-parameter with 0.35 evolutionary invariable sites (T92+I), and Tamura 3-parameter with 0.18 gamma distribution (T92+G) are the models having the lowest Bayesian Information Criterion (BIC) for the ITS2, *matK* and *rbcLa* sequences of all species of both families respectively and chosen for phylogeny reconstruction analysis for NJ and ML statistical methods (Fig 2). Three types of observations were made in analysis of clustering pattern in all phylogenetic trees i.e. value of nodal support, clustering of species, family wise branching pattern.

ITS2: NJ and ML statistical analyses resolved the sequences in distinct clusters of species with 75% and 100% nodal support respectively except for *Acacia nilotica* in both methods and *Indigofera heterantha* in ML phylogenetic tree. Family wise divergence pattern was not as clear as it was expected because species of both families were not branched together. Clustering pattern of ML phylogenetic tree was in accordance with BLAST analysis results i.e. *Lasiurus scindicus* and *Cymbopogon jwarancusa* were either taxonomically misidentified or potential cases of unintentional adulteration.

***matK*:** Species clustering pattern, nodal support values as well as family wise clustering was much distinct in both phylogenetic trees of *matK* sequences. The branch of *Cenchrus ciliaris* was less than 75% nodal support in NJ and ML trees.

***rbcLa*:** Nodal support values for both NJ and ML analyses was 100% while species resolution was not as strong as it was for ITS2 and *matK* e.g. closely related species, *Acacia jaquemontii* and *Acacia nilotica* were in the same clade. On the contrary, *rbcLa* represented the best divergence pattern at family and genus level among three markers.

4. Discussion

Floral biodiversity consists of a major category of medicinal plants that is important not only as a source of earning for local communities but also preserves traditional knowledge in the form of their medicinal uses (Shinwari & Qaisar, 2011). Our study approves the utility of DNA barcoding as species identification tool for the conservation of flora and safe use of medicinal plants of Fabaceae and Poaceae (Gao et al., 2010; Saadullah et al., 2016).

All of the barcoding regions included in this study are reasonably good regarding the universality in both families as reported earlier (Yan et al., 2015; Li et al., 2016). Since an ideal DNA barcode is expected to get amplified using standard PCR protocols in multiple species, we found that ITS2, *matK* and *rbcLa* fulfilled this criterion successfully with single pair of primers for each region. Comparatively, amplification success was slightly less for ITS2 and *matK* than *rbcLa* for Fabaceae that supports the opinion that barcodes are not consistent across the family Fabaceae but limited to a few genera (Hollingsworth et al., 2009; Newmaster & Ragupathy, 2009). On the contrary, Chen et al. (2010) and Han et al. (2013) stated that ITS2 was relatively easy to be amplified using one pair of universal primers as well as ITS2 has also been reported for having ability to overcome the amplification and sequencing problems being shorter in length and conserved than ITS1 (Yao et al., 2010; Gao et al., 2010; Pang et al., 2010).

Sequence statistics determined that ITS2 had the most number of variable sites as well as relatively larger interspecific distance, the properties that strengthen a marker as ideal barcode region for its species discrimination ability (Li et al., 2016) that's why ITS2 is recommended as taxonomic signatures in systematic evolution (Schultz et al., 2005; Coleman, 2007). Core barcoding regions, *matK* and *rbcLa* also had variable and species specific informative sites but much less as compared to that of ITS2, hence they are recommended to be used as multi-locus barcodes in earlier studies (CBOL Plant Working Group, 2009; China Plant BOL group, 2011; Li et al., 2016; Saadullah et al., 2016).

Sequence analysis through BLAST and TAXONDNA determined that ITS2 identified the most number of specimens of both families at species level. Performance of *matK* and *rbcLa* was relatively weak at species resolution ability similar to the study of Saadullah et al. (2016) on the DNA barcoding of Poaceae. *rbcLa* exhibited the highest genus level identification ability in both families. DNA barcoding gap also supported ITS2 region as a promising potential molecular marker to be used for species identification (Li et al., 2016).

Phylogenetic analysis provided a better species resolution than the nucleotide analysis (Kim et al., 2016) and has shown that despite of the fact that *matK* encountered amplification difficulty in few species of Fabaceae, its species resolution ability was considerably significant like ITS2 which is consistent to the study of Hilu & Liang (1997) and Hollingworth et al. (2011) who have declared *matK* as the best analogue of *CO1* animal barcode due to rapidly evolving plastid DNA region. Phylogenetic analysis strengthens the application of DNA barcoding as the biodiversity conservation tool (Harvig et al., 2015) and species authentication tool in quality control of herbal products (Seethapathy et al., 2014; Vassou et al., 2015).

Conclusion

Based on the sequence statistics, inter- and intraspecific distances, BLAST, TAXONDNA and phylogenetic analyses, it is concluded that DNA barcoding is a rapid and convenient species identification method as well as we suggest that ITS2 and *matK* both are the suitable barcode markers for medicinal species of Fabaceae and Poaceae.

Acknowledgement

We are thankful to Dr. Mansoor Hameed for critical morphological authentication of plant materials and preserving them as vouchers in Herbarium, Department of Botany, University of Agriculture, Faisalabad, Pakistan.

References

Ahmad S, Wariss HM, Alam K, Anjum S, Mukhtar M. 2014. Ethnobotanical Studies of Plant Resources of Cholistan Desert, Pakistan. *International Journal of Science and Research* 3:1782-1788.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of molecular biology* 215:403-10.

Balachandran KRS, Mohanasundaram S, Ramalingam S. 2015. DNA barcoding: a genomic-based tool for authentication of phytomedicinals and its products. *Botanics: Targets and Therapy* 5:77-84.

CBOL Plant Working Group. 2009. A DNA barcode for land plants. *Proceedings of National Academy of Sciences USA* 106:12794–12797.

Chen SL, Yao H, Han JP, Liu C, Song JY, Shi LC, Zhu YJ, Ma XY, Gao T, Pang XH et al. 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS One* 5:e8613. doi: 10.1371/journal.pone.0008613.

China Plant BOL group. 2011. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proceedings of National Academy of Sciences USA* 108:19641-19646. Doi: 10.1073/pnas.1104551108.

Coleman AW. 2007. Pan-eukaryote ITS2 homologies revealed by RNA secondary structure. *Nucleic Acids Research* 35:3322-3329.

Dashora K, Gosavi KVC. 2013. Grasses: An Underestimated Medicinal Repository. *Journal of Medicinal Plants Studies* 1:151-157.

de Mattia F, Bruni I, Galimberti A, Cattaneo F, Casiraghi M, Labra M. 2011. A comparative study of different DNA barcoding markers for the identification of some members of Lamiaceae. *Food Research International* 44:693–702.

297 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
298 throughput. *Nucleic Acids Research* 32:1792-7.
299 Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, Husband BC, Percy DM,
300 Hajibabaei M, Barrett SC. 2008. Multiple multilocus DNA barcodes from the plastid genome
301 discriminate plant species equally well. *PLoS ONE* 3:2802.
302 Gao T, Yao H, Song J, Liu C, Zhu Y, Ma X, Pang X, Xu H, Chen S. 2010. Identification of
303 medicinal plants in the family Fabaceae using a potential DNA barcode ITS2. *Journal of*
304 *Ethnopharmacology* 130:116-121.
305 Gathier G, vander Niet T, Peelen T, van Vugt RR, Eurlings MC, Gravendeel B. 2013. Forensic
306 identification of CITES protected slimming cactus (Hoodia) using DNA barcoding. *Journal of*
307 *Forensic Sciences* 58:1467-1471.
308 Ghorbani A, Saeedi Y, de Boer HJ. 2017. Unidentifiable by morphology: DNA barcoding of plant
309 material in local markets in Iran. *PLoS ONE* 12: e0175722.
310 Giudicelli GC, Mäder G, Brandão de Freitas L. 2015. Efficiency of ITS sequences for DNA
311 barcoding in Passiflora (Passifloraceae). *International Journal of Molecular Sciences* 16:7289-
312 303. doi:10.3390/ijms16047289.
313 Hameed M, Ashraf M, Al-Quriany F, Nawaz T, Ahmad MSA, Younis A, Naz N. 2011. Medicinal
314 flora of the Cholistan desert: A review. *Pakistan Journal of Botany* 43:39-50.
315 Hartvig I, Czako M, Kjær ED, Nielsen LR, Theilade I. 2015. The Use of DNA Barcoding in
316 Identification and Conservation of Rosewood (*Dalbergia* spp.). *PLoS ONE* 10:e0138231. DOI:
317 10.1371/journal.pone.0138231.
318 Hebert PDN, Cywinska A, Ball SL, DeWaard JR. 2003a. Biological identifications through DNA
319 barcodes. *Proceedings of The Royal Society B Biological Sciences* 270: 313-321.
320 DOI: 10.1098/rspb.2002.2218.

Hebert PDN, Gregory TR. 2005. The promise of DNA barcoding for taxonomy. *Systematic Biology* 54:852-859. DOI: 10.1080/10635150500354886.

Hebert PDN, Ratnasingham S, deWaard JR. 2003b. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of The Royal Society B Biological Sciences (Suppl.)* 270: S96–S99. DOI: 10.1098/rsbl.2003.0025.

Hilu KW, Liang H. 1997. The matK gene: sequence variation and application in plant systematics. *American Journal of Botany* 84:830-839.

Hollingsworth ML, Clark AA, Forrest LL, Richardson J, Pennington RT, Long DG, Cowan R, Chase MW, Gaudeul M, Hollingsworth PM. 2009. Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Molecular Ecology Resources* 9:439-457. doi:10.1111/j.1755-0998.2008.02439.x

Hollingsworth PM, Graham SW, Little DP. 2011. Choosing and using a plant DNA barcode. *PLoSONE* 6:e19254. doi:10.1371/journal.pone.0019254.

Joharchi MR, Amiri MS. 2012. Taxonomic evaluation of misidentification of crude herbal drugs marketed in Iran. *Avicenna Journal of Phytomedicine* 2:105-112.

Kataria S, Shrivastava B, Khajuria RK, Suri KA, Sharma P. 2010. Antimicrobial activity of *Crotalaria burhia* Buch.-Ham. roots. *Indian Journal of Natural Products and Resources* 1:481-484.

Kim WJ, Ji Y, Choi G, Kang YM, Yang S, Moon BC. 2016. Molecular identification and phylogenetic analysis of important medicinal plant species in genus Paeonia based on rDNA-ITS, matK, and rbcL DNA barcode sequences. *Genetics and Molecular Research: GMR* 5:15(3). DOI: 10.4238/gmr.15038472.

Kress WJ, Erickson DL. 2007. A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. *PLoS ONE* 2:e508. doi:10.1371/journal.pone.0000508.

346 Levin RA, Wagner WL, Hoch PC. 2003. Family-level relationships of Onagraceae based on
347 chloroplast *rbcL* and *ndhF* data. *American Journal of Botany* 90:107-115.
348 DOI: 10.3732/ajb.90.1.107.
349 Li Y, Tong Y, Xing F. 2016. DNA barcoding evaluation and its taxonomic implications in the
350 recently evolved genus Oberonia Lindl.(Orchidaceae) in China. *Frontiers in Plant Sciences*
351 7.1791. doi: 10.3389/fpls.2016.01791.
352 Meier R, Kwong S, Vaidya G, Ng Peter KL. 2006. DNA barcoding and taxonomy in diptera: a
353 tale of high intraspecific variability and low identification success. *Systematic Biology* 55:715-
354 728. DO1:10.1080/10635150600969864
355 Meier R, Zhang G, Ali F. 2008. The use of mean instead of smallest interspecific distances
356 exaggerates the size of the "barcoding gap" and leads to misidentification. *Systematic Biology* 57:
357 809-13. DOI: 10.1080/10635150802406343.
358 Mutanen M, Kekkonen M, Prosser SW, Hebert PD, Kaila L. 2015. One species in eight: DNA
359 barcodes from type specimens resolve a taxonomic quagmire. *Molecular Ecology Resources* 15:
360 967-984. doi: 10.1111/1755-0998.12361.
361 Pang X, Song J, Zhu Y, Xie C, Chen S. 2010. Using DNA barcoding to identify species within
362 Euphorbiaceae. *Planta Medica*. 76:1784-6. DOI: 10.1055/s-0030-1249806.
363 Pang XH, Song JY, Zhu YJ, Xu HX, Huang LF, Chen SL. 2011. Applying plant DNA barcodes
364 for Rosaceae species identification. Cladistics 27: 165–170. DOI: 10.1111/j.1096-
365 0031.2010.00328.x.
366 Saadullah, Khan ZUD, Ashfaq M, Zaib-u-Nisa. 2016. Identification of the grass family (Poaceae)
367 by using the plant dna barcodes rbcl and matK. *Journal of Biodiversity and Environmental*
368 *Sciences* 8:175-186.
369 Sagar PK. 2014. Adulteration and Substitution In Endangered, ASU Herbal Medicinal Plants of
370 India, Their Legal Status, Scientific Screening of Active Phytochemical Constituents. *Journal of*

371 *Pharmaceutical Sciences and Research* 5:4023-39. DOI: 10.13040/IJPSR.0975-8232.5(9).4023-
372 39.
373 Schultz J, Maisel S, Gerlach D, Müller T, Wolf M. 2005. A common core of secondary structure
374 of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. *RNA* 11:361-364. doi:
375 10.1261/rna.7204505.
376 Seethapathy GS, Ganesh D, Santhosh Kumar JU, Senthilkumar U, Newmaster SG, Ragupathy S,
377 Shaanker RU, Ravikanth G. 2014. Assessing product adulteration in natural health products for
378 laxative yielding plants, Cassia, Senna, and Chamaecrista in Southern India using DNA
379 barcoding. *International Journal of Legal Medicine* 129:693–700. DOI: 10.1007/s00414-014-
380 1120-z
381 Shinwari ZK, Qaisar M. 2011. Efforts on conservation and sustainable use of medicinal plants of
382 pakistan. *Pakistan Journal of Botany* 43:5-10.
383 Steven GN and Subramanyam R. 2009. Testing plant barcoding in a sister species complex of
384 pantropical Acacia (Mimosoideae, Fabaceae). *Molecular Ecology Resources* 9:172–180. DOI:
385 10.1111/j.1755-0998.2009.02642.x.
386 Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary
387 genetics analysis version 6.0. *Molecular Biology and Evolution* 30:2725-2729. DOI:
388 10.1093/molbev/mst197.
389 Tang GD, Zhang GQ, Hong WJ, Liu ZJ, Zhuang XY. 2015. Phylogenetic analysis of Malaxideae
390 (Orchidaceae: Epidendroideae): two new species based on the combined nrDNA ITS and
391 chloroplast matK sequences. *Journal of Guangxi Plant Science* 35:447-463. DOI: 10.11931 /
392 guihaia. gxzw201506015.
393 van Velzen R, Weitschek E, Felici G, Bakker FT. 2012. DNA barcoding of recently diverged
394 species: relative performance of matching methods. *PloS one* 7:e30490. doi:
395 10.1371/journal.pone.0030490.

Vassou SL, Kusuma G, Parani M. 2015. DNA barcoding for species identification from dried and powdered plant parts: a case study with authentication of the raw drug market samples of Sida cordifolia. *Gene* 559:86-93. DOI: 10.1016/j.gene.2015.01.025.

Wariss HM, Mukhtar M, Anjum S, Bhatti G, Pirzada S, Alam K. 2013. Floristic Composition of the Plants of the Cholistan Desert, Pakistan. *American Journal of Plant Sciences* 4:58-65. DOI: 10.4236/ajps.2013.412A1009.

Wariss HM, Wang H, Yi TS, Anjum S, Ahmad S, Alam K. 2016. Taxonomic Perspective of Grasses; a Potential Resource of Cholistan Desert, Pakistan. *Journal of Biodiversity and Environmental Sciences* 9:26-42.

White TJ, Bruns T, Lee S, Taylor J. 1990. Amplifcation and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: PCR Protocols: a guide to methods and applications. Academic Press, New York, USA: 315-322.

Xu S, Li D, Li J, Xiang X, Jin W, Huang W, Jin X, Huang L. 2015. Evaluation of the DNA barcodes in Dendrobium (Orchidaceae) from mainland Asia. *PloS one* 10:e0115168. doi: 10.1371/journal.pone.0115168.

Yan LJ, Liu J, Moller M, Zhang L, Zhang XM, Li DZ, Gao LM. 2015. DNA barcoding of Rhododendron (Ericaceae), the largest Chinese plant genus in biodiversity hotspots of the Himalaya-Hengduan Mountains. *Molecular Ecology Resources* 15:932-944. DOI: 10.1111/1755-0998.12353.

Yao H., Song JY, Liu C, Luo K, Han JP, Li Y, Pang X, Xu H, Zhu Y, Xiao P, Chen S. 2010. Use of ITS2 region as the universal DNA barcode for plants and animals. *PLoS ONE* 5:e13102. doi: 10.1371/journal.pone.0013102.

Table 1(on next page)

Primers

Name	Sequence 5′-3′	Source
matK		
KIM_3FKIM-r	CGTACAGTACTTTTGTGTTTACGAG	Ki-Joong Kim, Pers. comm.
KIM_1RKIM-f	ACCCAGTCCATCTGGAAATCTTGGTTC	Ki-Joong Kim, Pers. comm.
rbcL		
rbcLa-F	ATGTCACCACAAACAGAGACTAAAGC	(Levin, 2003)
rbcLa-R	GTAAAATCAAGTCCACCRCG	(Kress and Erickson, 2007)
ITS2		
ITS2-S2F	ATGCGATACTTGGTGTGAAT	(Chen et al., 2010)
ITS4	GTAAAATCAAGTCCACCRCG	(White et al., 1990)

Table 2(on next page)

Thermocycler programming for PCR of candidate barcodes

Cycle step and number of cycles		*matK*		*rbcLa*		ITS2	
		Block temp.	Hold time (mm:ss)	Block temp.	Hold time (mm:ss)	Block temp.	Hold time (mm:ss)
Init. denat.	1	98 ˚C	00:45	94 ˚C	04:00	94 ˚C	05:00
Denaturation Annealing Extension	35	98 ˚C 52 ˚C 72 ˚C	00:10 00:30 00:40	94 ˚C 55 ˚C 72 ˚C	00:30 00:30 01:00	94 ˚C 56 ˚C 72 ˚C	00:30 00:30 00:45
Final Ext.	1	72 ˚C	10:00	72 ˚C	10:00	72 ˚C	10:00

Table 3(on next page)

Sequence characteristics of ITS2, *matK* and *rbcLa* in selected medicinal species of Fabaceae and Poaceae

	Fabaceae			Poaceae		
	ITS2	*matK*	*rbcLa*	ITS2	*matK*	*rbcLa*
Universality of primers	Yes	Yes	Yes	Yes	Yes	Yes
Percentage PCR success (%)	85	71	100	100	100	100
Percentage sequencing success (%)	100	100	100	100	100	100
No. of species (No. of individuals)	7(21)	7(21)	7(21)	3(9)	3(9)	3(9)
No. of no sequence/singleton species	1	2	0	0	0	0
Aligned sequence length (bp)	360	844	553	365	772	553
Parsimony-Informative sites (bp)	200	98	43	106	27	16
Variable sites (bp)	209	99	44	106	27	17
Average interspecific distance (%)	0.35	0.07	0.03	0.26	0.02	0.02
Average intraspecific distance (%)	0.02	0.00	0.00	0.00	0.00	0.00

Figure 1

Relative abundance of intra- and interspecific K2P pairwise distance for ITS2, *matK*, and *rbcLa*

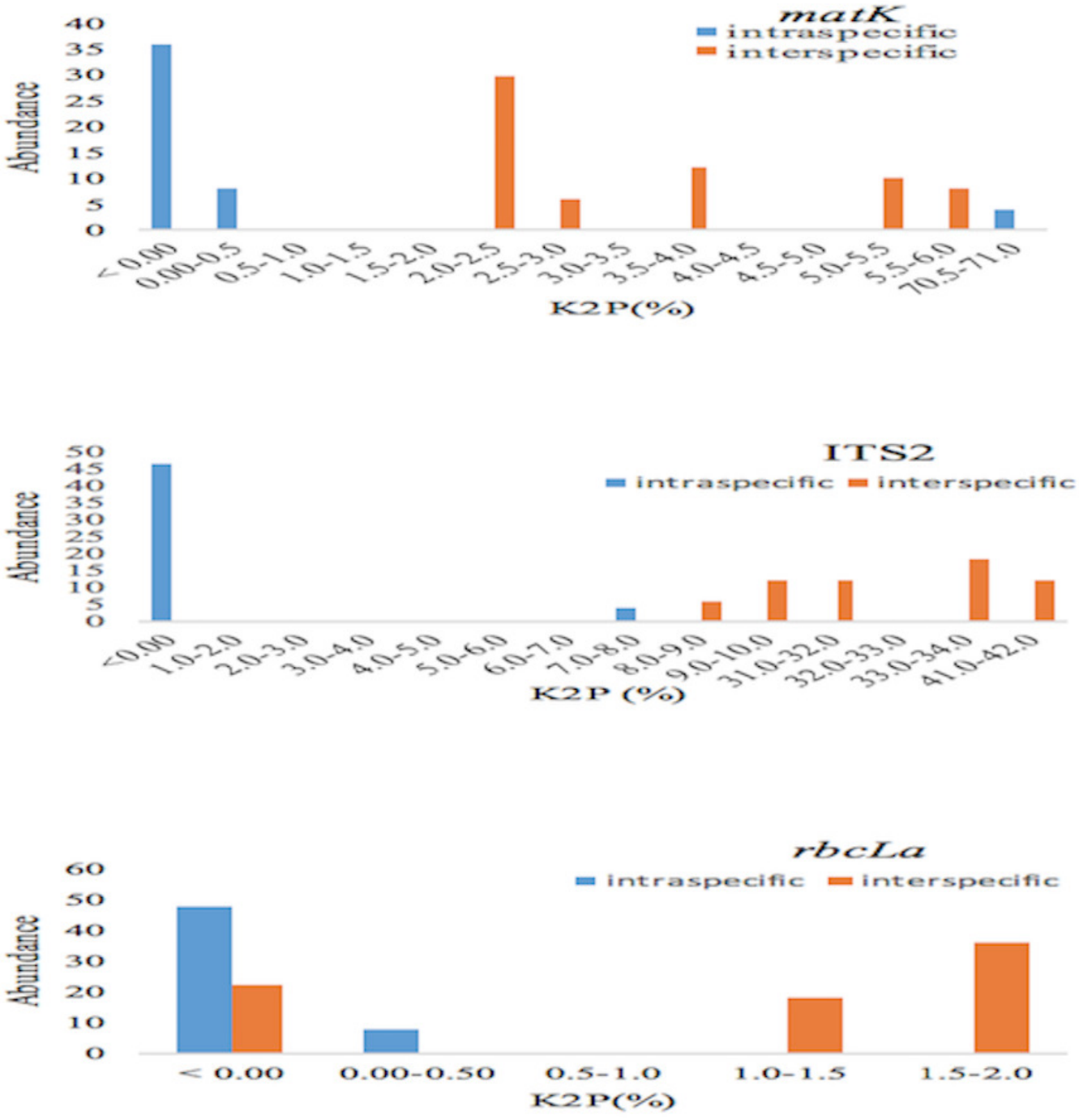


Table 4(on next page)

Genus and species level identification success of candidate barcodes by BLASTn analysis

Barcode region	Species-level identification rate	Genus-level identification rate
ITS2	40% (11/27)	74% (20/27)
matK	37% (9/24)	87% (21/24)
rbcLa	30% (9/30)	90% (27/30)

Table 5(on next page)

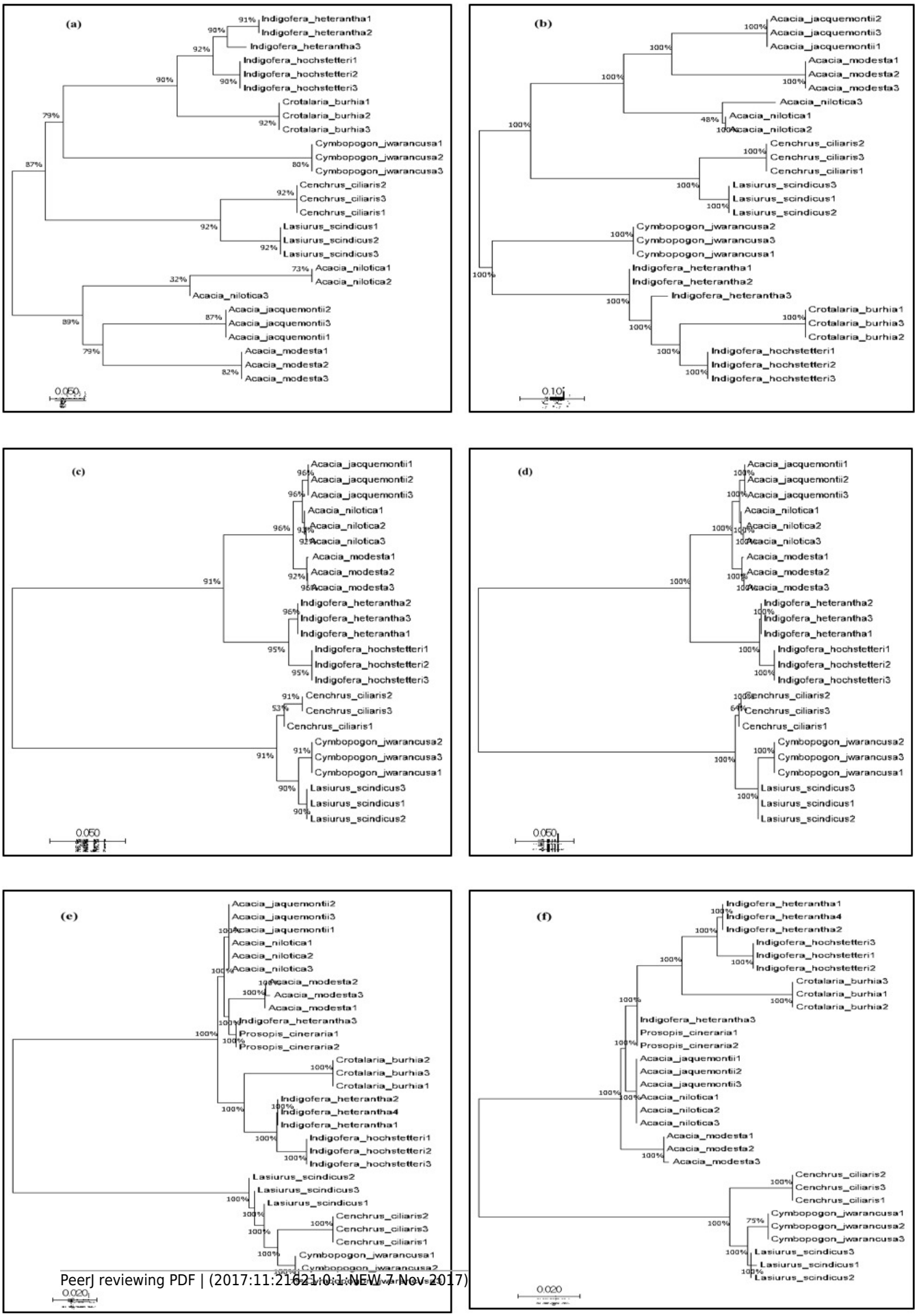
Species level discrimination ability of candidate barcodes by BM and BCM analysis

Family	Barcode region	Sample size	Best Match (%)		Best Close Match (%)	
			Correct	Ambiguous	Correct	Ambiguous
Fabaceae	**ITS2**	18	17 (100)	0.00 (0)	17 (100)	0.00 (0)
	matK	15	15 (100)	0.00 (0)	15 (100)	0.00 (0)
	rbcLa	21	12 (66.66)	6 (33.33)	12 (66.66)	6 (33.33)
Poaceae	**ITS2**	9	9 (100)	0.00 (0)	9 (100)	0.00 (0)
	matK	9	9 (100)	0.00 (0)	9 (100)	0.00 (0)
	rbcLa	9	9 (100)	0.00 (0)	9 (100)	0.00 (0)
Fabaceae + Poaceae	**ITS2**	27	26 (100)	0.00 (0)	26 (100)	0.00 (0)
	matK	24	24 (100)	0.00 (0)	24 (100)	0.00 (0)
	rbcLa	30	21 (77)	6 (22)	21 (77)	6 (22)

Figure 2

Phylogenetic analysis of candidate barcodes

(a) ITS2 NJ, (b) ITS2 ML, (c) *matK* NJ, (d) *matK* ML, (e) *rbcLa* NJ,(f) *rbcLa* ML


(a)
Indigofera_heterantha1
Indigofera_heterantha2
Indigofera_heterantha3
Indigofera_hochstetteri1
Indigofera_hochstetteri2
Indigofera_hochstetteri3
Crotalaria_burhia1
Crotalaria_burhia2
Crotalaria_burhia3
Cymbopogon_jwarancusa1
Cymbopogon_jwarancusa2
Cymbopogon_jwarancusa3
Cenchrus_ciliaris2
Cenchrus_ciliaris3
Cenchrus_ciliaris1
Lasiurus_scindicus1
Lasiurus_scindicus2
Lasiurus_scindicus3
Acacia_nilotica1
Acacia_nilotica2
Acacia_nilotica3
Acacia_jacquemontii2
Acacia_jacquemontii3
Acacia_jacquemontii1
Acacia_modesta1
Acacia_modesta2
Acacia_modesta3
0.050
(b)
Acacia_jacquemontii2
Acacia_jacquemontii3
Acacia_jacquemontii1
Acacia_modesta1
Acacia_modesta2
Acacia_modesta3
Acacia_nilotica3
Acacia_nilotica1
Acacia_nilotica2
Cenchrus_ciliaris2
Cenchrus_ciliaris3
Cenchrus_ciliaris1
Lasiurus_scindicus3
Lasiurus_scindicus1
Lasiurus_scindicus2
Cymbopogon_jwarancusa2
Cymbopogon_jwarancusa3
Cymbopogon_jwarancusa1
Indigofera_heterantha1
Indigofera_heterantha2
Indigofera_heterantha3
Crotalaria_burhia1
Crotalaria_burhia3
Crotalaria_burhia2
Indigofera_hochstetteri1
Indigofera_hochstetteri2
Indigofera_hochstetteri3
0.10
(c)
Acacia_jacquemontii1
Acacia_jacquemontii2
Acacia_jacquemontii3
Acacia_nilotica1
Acacia_nilotica2
Acacia_nilotica3
Acacia_modesta1
Acacia_modesta2
Acacia_modesta3
Indigofera_heterantha2
Indigofera_heterantha3
Indigofera_heterantha1
Indigofera_hochstetteri1
Indigofera_hochstetteri2
Indigofera_hochstetteri3
Cenchrus_ciliaris2
Cenchrus_ciliaris3
Cenchrus_ciliaris1
Cymbopogon_jwarancusa2
Cymbopogon_jwarancusa3
Cymbopogon_jwarancusa1
Lasiurus_scindicus3
Lasiurus_scindicus1
Lasiurus_scindicus2
0.050
(d)
Acacia_jacquemontii1
Acacia_jacquemontii2
Acacia_jacquemontii3
Acacia_nilotica1
Acacia_nilotica2
Acacia_nilotica3
Acacia_modesta1
Acacia_modesta2
Acacia_modesta3
Indigofera_heterantha2
Indigofera_heterantha3
Indigofera_heterantha1
Indigofera_hochstetteri1
Indigofera_hochstetteri2
Indigofera_hochstetteri3
Cenchrus_ciliaris2
Cenchrus_ciliaris3
Cenchrus_ciliaris1
Cymbopogon_jwarancusa2
Cymbopogon_jwarancusa3
Cymbopogon_jwarancusa1
Lasiurus_scindicus3
Lasiurus_scindicus1
Lasiurus_scindicus2
0.050
(e)
Acacia_jaquemontii2
Acacia_jaquemontii3
Acacia_jaquemontii1
Acacia_nilotica1
Acacia_nilotica2
Acacia_nilotica3
Acacia_modesta2
Acacia_modesta3
Acacia_modesta1
Indigofera_heterantha3
Prosopis_cineraria1
Prosopis_cineraria2
Crotalaria_burhia2
Crotalaria_burhia3
Crotalaria_burhia1
Indigofera_heterantha2
Indigofera_heterantha4
Indigofera_heterantha1
Indigofera_hochstetteri1
Indigofera_hochstetteri2
Indigofera_hochstetteri3
Lasiurus_scindicus2
Lasiurus_scindicus3
Lasiurus_scindicus1
Cenchrus_ciliaris2
Cenchrus_ciliaris3
Cenchrus_ciliaris1
Cymbopogon_jwarancusa1
Cymbopogon_jwarancusa2
0.020
(f)
Indigofera_heterantha1
Indigofera_heterantha4
Indigofera_heterantha2
Indigofera_hochstetteri3
Indigofera_hochstetteri1
Indigofera_hochstetteri2
Crotalaria_burhia3
Crotalaria_burhia1
Crotalaria_burhia2
Indigofera_heterantha3
Prosopis_cineraria1
Prosopis_cineraria2
Acacia_jaquemontii1
Acacia_jaquemontii2
Acacia_jaquemontii3
Acacia_nilotica1
Acacia_nilotica2
Acacia_nilotica3
Acacia_modesta1
Acacia_modesta2
Acacia_modesta3
Cenchrus_ciliaris2
Cenchrus_ciliaris3
Cenchrus_ciliaris1
Cymbopogon_jwarancusa1
Cymbopogon_jwarancusa2
Cymbopogon_jwarancusa3
Lasiurus_scindicus3
Lasiurus_scindicus1
Lasiurus_scindicus2
0.020