

A peer-reviewed version of this preprint was published in PeerJ on 13 March 2018.

[View the peer-reviewed version](https://doi.org/10.7717/peerj.4499) (peerj.com/articles/4499), which is the preferred citable publication unless you specifically need to cite this preprint.

Tahir A, Hussain F, Ahmed N, Ghorbani A, Jamil A. 2018. Assessing universality of DNA barcoding in geographically isolated selected desert medicinal species of Fabaceae and Poaceae. PeerJ 6:e4499
<https://doi.org/10.7717/peerj.4499>

Assessing universality of DNA barcoding in geographically isolated 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 six DNA barcodes viz. ITS2, *matK*, *rbcLa*, ITS2+*matK*, ITS2+*rbcLa*, *matK+rbcLa*, ITS2+*matK+rbcLa* for their capacity to identify frequently consumed but geographically isolated 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 six 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+*matK+rbcLa* followed ITS2 in its species level identification capacity. The study was concluded with advocating the DNA barcoding as an effective tool for species identification and ITS2 as the 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.

Assessing Universality of DNA Barcoding in Geographically Isolated Selected Desert Medicinal Species of Fabaceae and Poaceae

Aisha Tahir¹, Fatma Hussain¹, Nisar Ahmed², Abdolbaset Ghorbani³, Amer Jamil¹

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 Author:

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 six DNA barcodes viz. ITS2, *matK*, *rbcLa*, ITS2+*matK*, ITS2+*rbcLa*, *matK+rbcLa*, ITS2+*matK+rbcLa* for their capacity to identify frequently consumed but geographically isolated 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 six 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+*matK+rbcLa* followed ITS2 in its species level identification capacity. The study was concluded with advocating the DNA barcoding as an effective tool for species identification and ITS2 as the 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.

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 (Herrera et al., 2016). 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). Ethnobotanical investigations revealed that *Crotalaria burhia* has antimicrobial, anti-inflammatory, wound healing, and antioxidant properties (Kataria et al., 2010). *Acacia sp.* are used in tonics and for the treatment of dysentery, asthma, constipation, fever and gastric problems (Ahmed et al., 2014). *Indigofera sp.* have antioxidant property which

is used in the treatment of infectious diseases, abdominal and spastic pain, and skin problems (Rahman et al., 2014). *Cenchrus ciliaris* has been reported anodyne, diuretic and emollient (Hameed et al., 2011; Wariss et al., 2013). *Cymbopogon jwarancusa* is reported as expectorant and used in treatment of flu, infections and epilepsy (Ahmed et al., 2014).

Local communities of far furlong areas of the country including Cholistan rely on herbal remedies without considering proper identification and documentation of valued medicinal species (Mahmood et al., 2013; Ahmed et al., 2014). 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 (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 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 (De Mattia et al., 2011; Saddhe & Kumar, 2017). Owing to the strengths and limitations

associated with each marker, *matK* and *rbcL* are recommended as core barcode regions, which worked well with many of the plant groups. The need of supplementary barcodes arose due to comparatively lower discrimination success rate of *matK+rbcL* than COI in plants and inefficient resolution in difficult plant taxa such as *Quercus* and *Salix*. Among supplementary markers, several constraints are reported in *trnH-psbA* such as premature sequence termination, presence of duplicated loci, and variable sequence lengths (100-1000 bp) thus paving the way for nuclear DNA region, ITS2 which is a part of ITS, either as individual barcoding marker or supplementary region with core barcode for quick taxonomical classification in closely related species of wide range of taxa such as in Fabaceae, Lamiaceae, Asteraceae, Rutaceae, Rosaceae and many more (CBOL Plant Working Group, 2009; Chen et al., 2010; Gao et al., 2010; Hollingsworth et al., 2011; Pang et al., 2011; Balachandran et al., 2015; Wu et al., 2017), hence tested in the current study as well.

The present study aims to re-evaluate the universality of commonly used DNA barcoding loci, ITS2, *matK*, *rbcLa*, ITS2+*matK*, ITS2+*rbcLa*, *matK+rbcLa*, and ITS2+*matK+rbcLa* by applying them on medicinal plants indigenous to harsh environment of Cholistan Desert for the first time. The objective was to barcode the species and to compare the discriminatory power of the standard barcode regions that will be an addition to the previous barcoding studies on Fabaceae and Poaceae, which were conducted on geographically different species and populations (Gao et al., 2010; Wu et al., 2017). Bioinformatics approach was practiced in the investigation for sequence analysis and barcode region evaluation.

2. Materials and methods

2.1. Plant material

A total of 30 specimens belonging to 7 species of Fabaceae and 3 of Poaceae were included in this study. According to ethnobotanical survey (Hameed et al., 2011; Ahmed et al., 2014), all of the collected species are commonly used as medicinal plants in herbal formulations, but they are

difficult to identify morphologically specifically in dried and processed form. 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. 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 MgCl₂ (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 ddH₂O 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 MgCl₂, 2.8 μl of each primer (10 μM), 1 μl of template DNA and 20.0 μl of ddH₂O. 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 single-barcode (ITS2, *matK*, *rbcLa*) and combination-barcode (ITS2+*matK*, ITS2+*rbcLa*, *matK+rbcLa*, ITS2+*matK+rbcLa*) analyses. Only those species were included in combination-barcode analyses that have triplets of sequences of each marker of combination. 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 and their combinations 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](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 single and multilocus barcode, 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).

2.3.4. Sequence similarity analysis

In the sequence similarity method, the species identification potential of all barcode regions was assessed by calculating the percentage of correct identifications identified with the “Best Match” (BM) and “Best Close Match” (BCM) tests in Species Identifier 1.8 program of TAXONDNA software (Meier et al., 2006). Three aligned datasets of sequences of Fabaceae, Poaceae, and Fabaceae+Poaceae were prepared to compare the candidate markers’ efficacy in closely and distantly related taxa. K2P distance model was used in this analysis.

2.3.5. Phylogenetic analysis

In order to assess whether species are recovered as monophyletic groups, phylogenetic trees were reconstructed in MEGA6 after appropriate model selection in the same software for each single and combination barcode for all the studied species of both families. The barcode markers were compared on the basis of conspecific monophyletic clusters and the nodal bootstrap support in

neighbor-joining (NJ) as well as in maximum-likelihood (ML) statistical methods (Tang et al., 2015; Xu et al., 2015; Zhang et al., 2015).

3. Results

3.1. Amplification, sequence analysis, and genetic divergence

The three commonly used barcoding loci performed differently in terms of universality for amplification and sequencing in both families. Amplification success is 85%, 71% and 100% for ITS2, *matK* and *rbcLa* respectively for Fabaceae and 100% for all regions for specimens of Poaceae. Overall aligned length of the three regions ranged from 360 bp (ITS2) to 844 bp (*matK*) for Fabaceae and from 365 bp (ITS2) to 772 bp (*matK*) for Poaceae. In this study, 18 sequences of ITS2, 15 of *matK*, and 21 of *rbcLa* were generated from family Fabaceae and 27 sequences (triplicate of each species with each region) from Poaceae. In addition, ITS2 had the highest percentage of parsimony informative sites i.e. 56% (Fabaceae) and 29% (Poaceae), followed by *matK* i.e. 12% (Fabaceae) and 3% (Poaceae) and *rbcLa* i.e. 8% (Fabaceae) and 3% (Poaceae) (Table 1). Out of total seven medicinal species of Fabaceae, *Prosopis cineraria* was not amplified 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 1). Multilocus barcodes were prepared by concatenation of single barcodes hence their characteristics corresponded to their counterparts with an altered species identification effect.

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 as single- and combination-barcodes. 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.

3.2. DNA barcoding gap assessment

The relative distribution of the frequencies of K2P distances was calculated for the three single and four combined loci for the selected species of Fabaceae and Poaceae families included in the study using TAXONDNA software, thus barcoding gap was identified for all the barcoding markers. Pairwise intra- and interspecific genetic distances showed similar overlapped pattern for *rbcLa*, ITS2+*matK* and *matK+rbcLa* while distances were narrow in case of *matK* and ITS2+*rbcLa*. ITS2 among single, and ITS2+*matK+rbcLa* among multilocus markers have distinct gap between pairwise intra- and interspecies genetic distance at 1% and 0.5% divergence respectively. 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*, *rbcLa*, ITS2+*matK*, ITS2+*rbcLa*, *matK+rbcLa* and ITS2+*matK+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 among all single and combination barcodes. 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 2).

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

The potential of all barcoding regions for species identification accuracy was estimated by measuring the proportions of correct identifications using BM and BCM functions. Both tools evaluate the proportion of correct identifications through different comparisons of input DNA sequences. In the SpeciesIdentifier program of the TAXONDNA software package, each sequence is compared with all other sequences present in the dataset and then compared sequences are grouped on the basis of their pairwise genetic distances that ultimately determines the conspecificity of two sequences.

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 significant similarity, then that sequence was considered ambiguous. The BCM function offered more stringent criteria by keeping a threshold of 0.1-0.5% pairwise distance in pairwise summary function. The queries above the threshold value were classified as “no match” and the others that are below the threshold value were analyzed according to the criteria established in “best match” analysis (Meier et al., 2006; Giudicelli et al., 2015; Hartvig et al., 2015; Mishra et al., 2017).

The results of sequence similarity test performed in TAXONDNA software for all single and combination barcodes are presented in Figure 2. With both functions (BM and BCM), ITS2 was consistent in achieving the highest percentage of correct identification and the lowest number of unidentified sequences in all datasets. *rbcLa*, showed the lowest discriminatory power for Fabaceae as six sample sequences were found ambiguous. An increase in identification power of *rbcLa* is observed when it is combined with ITS2 in all datasets. “Incorrect” and “no match” were 0% in both functions so they are not shown in Figure 2. This analysis indicates that the ITS2 met the rigorous standards for identifying the queries accurately among all single and combination barcodes.

225 3.5. Tree based analysis of barcoding regions

226 Before proceeding to reconstruct the phylogeny using NJ and ML statistical methods, appropriate
 227 models having the lowest Bayesian Information Criterion (BIC) for the ITS2, *matK*, *rbcLa*,
 228 ITS2+*matK*, ITS2+*rbcLa*, *matK+rbcLa* and ITS2+*matK+rbcLa* were chosen (Austerlitz et al.,
 229 2009). Three types of observations were made in analysis of clustering pattern in all phylogenetic
 230 trees i.e. value of nodal support, clustering of species, family wise branching pattern.

231 Both, NJ and ML, statistical methods consistently recovered monophyletic clades at species level
 232 using all the single and combination barcodes except that of *rbcLa* which could not discriminate
 233 between two species of genus *Acacia* of Fabaceae. Apart from tree topologies, bootstrap values
 234 were used as a criterion in this study, which was set at $\geq 99\%$ as threshold. ITS2 under ML,
 235 ITS2+*rbcLa* under NJ, and ITS2+*matK+rbcLa* under both phylogenetic methods worked equally
 236 well at species level for both families having higher percentage of nodes with $\geq 99\%$ support as
 237 compared to other barcoding markers (Table 3).

238 4. Discussion

239 Floral biodiversity consists of a major category of medicinal plants that is important not only as a
 240 source of earning for local communities but also preserves traditional knowledge in the form of
 241 their medicinal uses (Shinwari & Qaisar, 2011). Our study approves the utility of DNA barcoding
 242 as species identification tool for the conservation of flora and safe use of medicinal plants of
 243 Fabaceae and Poaceae (Gao et al., 2010; Saadullah et al., 2016). Though, environmental (desert,
 244 marshes, lime rocks etc.) and biological factors (poorly dispersed, salt tolerant and relatively
 245 isolated species) influence the universality and standardization of DNA barcoding technique (Yao
 246 et al., 2017).

247 All of the barcoding regions included in this study are reasonably good regarding the universality
 248 in both families as reported earlier (Yan et al., 2015; Li et al., 2016). Since an ideal DNA barcode
 249 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). 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, species specific informative sites but performed relatively poor than that of ITS2. In consistence with prior studies (China Plant BOL group, 2011; Zhang et al., 2012; Li et al., 2016; Saadullah et al., 2016; Mishra et al., 2017), *matK* and *rbcLa* are recommended to be used as multi-locus barcodes (ITS2+*matK*, ITS2+*rbcLa*, ITS2+*matK*+*rbcLa*) as evident in figures 1 and 2 and table 3.

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 (Clement and Donoghue 2012; Kim et al., 2016) and has shown that despite of the fact that all of the

275 barcoding regions except *rbcLa* resolved specimens into distinct monophyletic clades at family,
 276 genus and species levels but considerably differed with respect to nodal support values.
 277 Phylogenetic trees of ITS2, ITS2+*rbcLa*, and ITS2+*matK*+*rbcLa* had similar percentage of nodes
 278 having 99% or more bootstrap support hence keeping the cost and time effectiveness into
 279 account, single barcode is preferred on multi-locus barcode specifically for small dataset (Feng et
 280 al., 2015; Braukmann et al., 2017; Mishra et al., 2017). This is in contrast to the study of Hilu &
 281 Liang (1997) and Hollingworth et al. (2011) who have declared *matK* as the best analogue of
 282 *COI* animal barcode due to rapidly evolving plastid DNA region. Phylogenetic analysis
 283 strengthens the application of DNA barcoding as the biodiversity conservation tool (Harvig et al.,
 284 2015) and species authentication tool in quality control of herbal products (Seethapathy et al.,
 285 2014; Vassou et al., 2015).

286 Conclusion

287 Based on the sequence statistics, inter- and intraspecific distances, BLAST, TAXONDNA and
 288 phylogenetic analyses, it is concluded that DNA barcoding is a rapid, convenient and universal
 289 species identification method that has been refined enough that it can discriminate the relatively
 290 isolated desert species as well as we suggest that ITS2 is the most suitable barcode markers for
 291 identification of medicinal species of Fabaceae and Poaceae.

292 Acknowledgement

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

296 References

297 Ahmed N, Mahmood A, Tahir SS, Bano A, Malik RN, Hassan S, Ashraf A. 2014. Ethnomedicinal
 298 knowledge and relative importance of indigenous medicinal plants of Cholistan desert, Punjab

- 299 Province, Pakistan. *Journal of Ethnopharmacology* 155:1263-1275.
- 300 <http://dx.doi.org/10.1016/j.jep.2014.07.007>.
- 301 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool.
- 302 *Journal of molecular biology* 215:403-10.
- 303 Austerlitz F, David O, Schaeffer B, Bleakley K, Olteanu M, Leblois R, Veuille M, Laredo, C.
- 304 2009. DNA barcode analysis: a comparison of phylogenetic and statistical classification
- 305 methods. *BMC Bioinformatics* 10:S10. <http://doi.org/10.1186/1471-2105-10-S14-S10>.
- 306 Balachandran KRS, Mohanasundaram S, Ramalingam S. 2015. DNA barcoding: a genomic-
- 307 based tool for authentication of phytomedicinals and its products. *Botanics: Targets and Therapy*
- 308 5:77-84.
- 309 Braukmann TWA, Kuzmina ML, Sills J, Zakharov EV, Hebert PDN. 2017. Testing the Efficacy
- 310 of DNA Barcodes for Identifying the Vascular Plants of Canada. *PLoS ONE* 12:e0169515.
- 311 [doi:10.1371/journal.pone.0169515](https://doi.org/10.1371/journal.pone.0169515).
- 312 CBOL Plant Working Group. 2009. A DNA barcode for land plants. *Proceedings of National*
- 313 *Academy of Sciences USA* 106:12794–12797.
- 314 Chen SL, Yao H, Han JP, Liu C, Song JY, Shi LC, Zhu YJ, Ma XY, Gao T, Pang XH, Luo K, Li
- 315 Y, Li X, Jia X, Lin Y, Leon C. 2010. Validation of the ITS2 region as a novel DNA barcode for
- 316 identifying medicinal plant species. *PLoS ONE* 5:e8613. [doi: 10.1371/journal.pone.0008613](https://doi.org/10.1371/journal.pone.0008613).
- 317 China Plant BOL group. 2011. Comparative analysis of a large dataset indicates that internal
- 318 transcribed spacer (ITS) should be incorporated into the core barcode for seed plants.
- 319 *Proceedings of National Academy of Sciences USA* 108:19641-19646. [Doi: 10.1073/pnas.1104551108](https://doi.org/10.1073/pnas.1104551108).
- 320 [10.1073/pnas.1104551108](https://doi.org/10.1073/pnas.1104551108).
- 321 Clement WL, Donoghue MJ. 2012. Barcoding success as a function of phylogenetic relatedness
- 322 in *Viburnum*, a clade of woody angiosperms. *BMC Evolutionary Biology* 12:73.
- 323 [doi:10.1186/1471-2148-12-73](https://doi.org/10.1186/1471-2148-12-73).

- 324 Coleman AW. 2007. Pan-eukaryote ITS2 homologies revealed by RNA secondary structure.
325 *Nucleic Acids Research* 35:3322-3329.
- 326 Dashora K, Gosavi KVC. 2013. Grasses: An Underestimated Medicinal Repository. *Journal of*
327 *Medicinal Plants Studies* 1:151-157.
- 328 de Mattia F, Bruni I, Galimberti A, Cattaneo F, Casiraghi M, Labra M. 2011. A comparative study
329 of different DNA barcoding markers for the identification of some members of Lamiaceae. *Food*
330 *Research International* 44:693–702.
- 331 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
332 throughput. *Nucleic Acids Research* 32:1792-7.
- 333 Feng S, Jiang Y, Wang S, Jiang M, Chen Z, Ying Q, Wang H. 2015. Molecular Identification of
334 *Dendrobium* Species (Orchidaceae) Based on the DNA Barcode ITS2 Region and Its Application
335 for Phylogenetic Study. *International Journal of Molecular Sciences* 16:21975-21988.
336 doi:10.3390/ijms160921975.
- 337 Gao T, Yao H, Song J, Liu C, Zhu Y, Ma X, Pang X, Xu H, Chen S. 2010. Identification of
338 medicinal plants in the family Fabaceae using a potential DNA barcode ITS2. *Journal of*
339 *Ethnopharmacology* 130:116-121.
- 340 Gathier G, vander Niet T, Peelen T, van Vugt RR, Eurlings MC, Gravendeel B. 2013. Forensic
341 identification of CITES protected slimming cactus (Hoodia) using DNA barcoding. *Journal of*
342 *Forensic Sciences* 58:1467-1471.
- 343 Ghorbani A, Saeedi Y, de Boer HJ. 2017. Unidentifiable by morphology: DNA barcoding of plant
344 material in local markets in Iran. *PLoS ONE* 12: e0175722.
- 345 Giudicelli GC, Mäder G, Brandão de Freitas L. 2015. Efficiency of ITS sequences for DNA
346 barcoding in *Passiflora* (Passifloraceae). *International Journal of Molecular Sciences* 16:7289-
347 303. doi:10.3390/ijms16047289.

- 348 Hameed M, Ashraf M, Al-Quriany F, Nawaz T, Ahmad MSA, Younis A, Naz N. 2011. Medicinal
349 flora of the Cholistan desert: A review. *Pakistan Journal of Botany* 43:39-50.
- 350 Hartvig I, Czako M, Kjær ED, Nielsen LR, Theilade I. 2015. The Use of DNA Barcoding in
351 Identification and Conservation of Rosewood (*Dalbergia* spp.). *PLoS ONE* 10:e0138231. DOI:
352 10.1371/journal.pone.0138231.
- 353 Hebert PDN, Cywinska A, Ball SL, DeWaard JR. 2003a. Biological identifications through DNA
354 barcodes. *Proceedings of The Royal Society B Biological Sciences* 270: 313-321.
355 DOI: 10.1098/rspb.2002.2218.
- 356 Hebert PDN, Gregory TR. 2005. The promise of DNA barcoding for taxonomy. *Systematic*
357 *Biology* 54:852-859. DOI: 10.1080/10635150500354886.
- 358 Hebert PDN, Ratnasingham S, deWaard JR. 2003b. Barcoding animal life: cytochrome c oxidase
359 subunit 1 divergences among closely related species. *Proceedings of The Royal Society B*
360 *Biological Sciences (Suppl.)* 270: S96–S99. DOI: 10.1098/rsbl.2003.0025.
- 361 Herrera M, Rodríguez O, Torregrosa L, Vásquez M, Blanco E, Calderón L. 2016. Traditional
362 use of plants as antihypertensive in Jipijapa, Manabí; Comparison with the literature. *In*
363 *Proceedings of the MOL2NET, International Conference on Multidisciplinary Sciences*, 15
364 January–30 December 2016; Sciforum Electronic Conference Series, Vol. 2, 2016. DOI:
365 10.3390/mol2net-02-03890.
- 366 Hilu KW, Liang H. 1997. The *matK* gene: sequence variation and application in plant
367 systematics. *American Journal of Botany* 84:830-839.
- 368 Hollingsworth ML, Clark AA, Forrest LL, Richardson J, Pennington RT, Long DG, Cowan R,
369 Chase MW, Gaudeul M, Hollingsworth PM. 2009. Selecting barcoding loci for plants: evaluation
370 of seven candidate loci with species-level sampling in three divergent groups of land plants.
371 *Molecular Ecology Resources* 9:439-457. doi:10.1111/j.1755-0998.2008.02439.x

- 372 Hollingsworth PM, Graham SW, Little DP. 2011. Choosing and using a plant DNA barcode.
- 373 *PLoS ONE* 6:e19254. doi:10.1371/journal.pone.0019254.
- 374 Joharchi MR, Amiri MS. 2012. Taxonomic evaluation of misidentification of crude herbal drugs
- 375 marketed in Iran. *Avicenna Journal of Phytomedicine* 2:105-112.
- 376 Kataria S, Shrivastava B, Khajuria RK, Suri KA, Sharma P. 2010. Antimicrobial activity of
- 377 *Crotalaria burhia* Buch.-Ham. roots. *Indian Journal of Natural Products and Resources* 1:481-
- 378 484.
- 379 Kim WJ, Ji Y, Choi G, Kang YM, Yang S, Moon BC. 2016. Molecular identification and
- 380 phylogenetic analysis of important medicinal plant species in genus *Paeonia* based on rDNA-ITS,
- 381 *matK*, and *rbcL* DNA barcode sequences. *Genetics and Molecular Research: GMR* 5:15(3).
- 382 DOI: 10.4238/gmr.15038472.
- 383 Kress WJ, Erickson DL. 2007. A two-locus global DNA barcode for land plants: the coding *rbcL*
- 384 gene complements the non-coding *trnH-psbA* spacer region. *PLoS ONE* 2:e508.
- 385 doi:10.1371/journal.pone.0000508.
- 386 Levin RA, Wagner WL, Hoch PC. 2003. Family-level relationships of Onagraceae based on
- 387 chloroplast *rbcL* and *ndhF* data. *American Journal of Botany* 90:107-115.
- 388 DOI: 10.3732/ajb.90.1.107.
- 389 Li Y, Tong Y, Xing F. 2016. DNA barcoding evaluation and its taxonomic implications in the
- 390 recently evolved genus *Oberonia* Lindl. (Orchidaceae) in China. *Frontiers in Plant Sciences*
- 391 7:1791. DOI: 10.3389/fpls.2016.01791.
- 392 Mahmood A, Mahmood A, Malik RN, Shinwari ZK. 2013. Indigenous knowledge of medicinal
- 393 plants from Gujranwala district, Pakistan. *Journal of Ethnopharmacology* 148:714–723.
- 394 <https://doi.org/10.1016/j.jep.2013.05.035>.

- 395 Meier R, Kwong S, Vaidya G, Ng Peter KL. 2006. DNA barcoding and taxonomy in diptera: a
396 tale of high intraspecific variability and low identification success. *Systematic Biology* 55:715-
397 728. DOI:10.1080/10635150600969864
- 398 Meier R, Zhang G, Ali F. 2008. The use of mean instead of smallest interspecific distances
399 exaggerates the size of the “barcoding gap” and leads to misidentification. *Systematic Biology* 57:
400 809-13. DOI: 10.1080/10635150802406343.
- 401 Mishra P, Kumar A, Nagireddy A, Shukla AK, Sundaresan V. 2017. Evaluation of single and
402 multilocus DNA barcodes towards species delineation in complex tree genus *Terminalia*. *PLoS*
403 *ONE* 12: e0182836. <https://doi.org/10.1371/journal.pone.0182836>.
- 404 Mutanen M, Kekkonen M, Prosser SW, Hebert PD, Kaila L. 2015. One species in eight: DNA
405 barcodes from type specimens resolve a taxonomic quagmire. *Molecular Ecology Resources* 15:
406 967-984. doi: 10.1111/1755-0998.12361.
- 407 Pang X, Song J, Zhu Y, Xie C, Chen S. 2010. Using DNA barcoding to identify species within
408 Euphorbiaceae. *Planta Medica*. 76:1784-6. DOI: 10.1055/s-0030-1249806.
- 409 Pang XH, Song JY, Zhu YJ, Xu HX, Huang LF, Chen SL. 2011. Applying plant DNA barcodes
410 for Rosaceae species identification. *Cladistics* 27: 165–170. DOI: 10.1111/j.1096-
411 0031.2010.00328.x.
- 412 Saadullah, Khan ZUD, Ashfaq M, Zaib-u-Nisa. 2016. Identification of the grass family (Poaceae)
413 by using the plant dna barcodes *rbcL* and *matK*. *Journal of Biodiversity and Environmental*
414 *Sciences* 8:175-186.
- 415 Sagar PK. 2014. Adulteration and Substitution In Endangered, ASU Herbal Medicinal Plants of
416 India, Their Legal Status, Scientific Screening of Active Phytochemical Constituents. *Journal of*
417 *Pharmaceutical Sciences and Research* 5:4023-39. DOI: 10.13040/IJPSR.0975-8232.5(9).4023-
418 39.

- 419 Schultz J, Maisel S, Gerlach D, Müller T, Wolf M. 2005. A common core of secondary structure
420 of the internal transcribed spacer 2 (ITS2) throughout the Eukaryota. *RNA* 11:361-364. doi:
421 10.1261/rna.7204505.
- 422 Seethapathy GS, Ganesh D, Santhosh Kumar JU, Senthilkumar U, Newmaster SG, Ragupathy S,
423 Shaanker RU, Ravikanth G. 2014. Assessing product adulteration in natural health products for
424 laxative yielding plants, Cassia, Senna, and Chamaecrista in Southern India using DNA
425 barcoding. *International Journal of Legal Medicine* 129:693–700. DOI: 10.1007/s00414-014-
426 1120-z
- 427 Rahman TU, Liaqat W, Khattak KF, Choudhary MI, Kamil A, Zeb MA. 2017. Cytotoxicity of
428 aerial parts of *Indigofera heterantha*. *Scientific research and essays* 12:77-80. DOI:
429 10.5897/SRE2014.5814.
- 430 Saddhe AA, Kumar K. 2017. DNA barcoding of plants: Selection of core markers for taxonomic
431 groups. *Plant Science Today* 5:9-13. doi: <https://dx.doi.org/10.14719/pst.2018.5.1.356>.
- 432 Shinwari ZK, Qaisar M. 2011. Efforts on conservation and sustainable use of medicinal plants of
433 pakistan. *Pakistan Journal of Botany* 43:5-10.
- 434 Steven GN and Subramanyam R. 2009. Testing plant barcoding in a sister species complex of
435 pantropical Acacia (Mimosoideae, Fabaceae). *Molecular Ecology Resources* 9:172–180. DOI:
436 10.1111/j.1755-0998.2009.02642.x.
- 437 Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary
438 genetics analysis version 6.0. *Molecular Biology and Evolution* 30:2725-2729. DOI:
439 10.1093/molbev/mst197.
- 440 Tang GD, Zhang GQ, Hong WJ, Liu ZJ, Zhuang XY. 2015. Phylogenetic analysis of Malaxideae
441 (Orchidaceae: Epidendroideae): two new species based on the combined nrDNA ITS and
442 chloroplast *matK* sequences. *Journal of Guangxi Plant Science* 35:447-463. DOI: 10.11931 /
443 guihaia. gxzw201506015.

- 444 van Velzen R, Weitschek E, Felici G, Bakker FT. 2012. DNA barcoding of recently diverged
445 species: relative performance of matching methods. *PLoS ONE* 7:e30490. doi:
446 10.1371/journal.pone.0030490.
- 447 Vassou SL, Kusuma G, Parani M. 2015. DNA barcoding for species identification from dried and
448 powdered plant parts: a case study with authentication of the raw drug market samples of *Sida*
449 *cordifolia*. *Gene* 559:86-93. DOI: 10.1016/j.gene.2015.01.025.
- 450 Wariss HM, Mukhtar M, Anjum S, Bhatti G, Pirzada S, Alam K. 2013. Floristic Composition of
451 the Plants of the Cholistan Desert, Pakistan. *American Journal of Plant Sciences* 4:58-65. DOI:
452 10.4236/ajps.2013.412A1009.
- 453 Wariss HM, Wang H, Yi TS, Anjum S, Ahmad S, Alam K. 2016. Taxonomic Perspective of
454 Grasses; a Potential Resource of Cholistan Desert, Pakistan. *Journal of Biodiversity and*
455 *Environmental Sciences* 9:26-42.
- 456 White TJ, Bruns T, Lee S, Taylor J. 1990. Amplification and direct sequencing of fungal
457 ribosomal RNA genes for phylogenetics. In: PCR Protocols: a guide to methods and applications.
458 Academic Press, New York, USA: 315-322.
- 459 Wu F, Ma J, Meng Y, Zhang D, Pascal Muvunyi B, Luo K, Di H, Guo W, Wang Y, Feng B. 2017.
460 Potential DNA barcodes for *Melilotus* species based on five single loci and their combinations.
461 *PLoS ONE* 12:e0182693. <https://doi.org/10.1371/journal.pone.0182693>.
- 462 Xu S, Li D, Li J, Xiang X, Jin W, Huang W, Jin X, Huang L. 2015. Evaluation of the DNA
463 barcodes in *Dendrobium* (Orchidaceae) from mainland Asia. *PLoS ONE* 10:e0115168. doi:
464 10.1371/journal.pone.0115168.
- 465 Yan LJ, Liu J, Moller M, Zhang L, Zhang XM, Li DZ, Gao LM. 2015. DNA barcoding of
466 *Rhododendron* (Ericaceae), the largest Chinese plant genus in biodiversity hotspots of the
467 Himalaya-Hengduan Mountains. *Molecular Ecology Resources* 15:932-944. DOI: 10.1111/1755-
468 0998.12353.

- 469 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
470 ITS2 region as the universal DNA barcode for plants and animals. *PLoS ONE* 5:e13102. doi:
471 10.1371/journal.pone.0013102.
- 472 Yao PC, Gao HY, Wei YN, Zhang JH, Chen XY, Li HQ. 2017. Evaluating sampling strategy for
473 DNA barcoding study of coastal and inland halo-tolerant Poaceae and Chenopodiaceae: A case
474 study for increased sample size. *PLoS ONE* 12:e0185311.
475 <https://doi.org/10.1371/journal.pone.0185311>.
- 476 Zhang J, Chen M, Dong X, Lin R, Fan J, Chen Z. 2015. Evaluation of Four Commonly Used
477 DNA Barcoding Loci for Chinese Medicinal Plants of the Family Schisandraceae. *PLoS ONE*
478 10:e0125574. <https://doi.org/10.1371/journal.pone.0125574>.

Table 1 (on next page)

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

Table 1: Sequence characteristics of ITS2, *matK* and *rbcLa* in selected medicinal species of Fabaceae and Poaceae

	Fabaceae			Poaceae		
	ITS2	<i>matK</i>	<i>rbcLa</i>	ITS2	<i>matK</i>	<i>rbcLa</i>
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

Table 2 (on next page)

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

Table 2: 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)
<i>matK</i>	37% (9/24)	87% (21/24)
<i>rbcLa</i>	30% (9/30)	90% (27/30)
ITS2+ <i>matK</i>	37% (9/24)	87% (21/24)
ITS2+ <i>rbcLa</i>	22% (6/27)	89% (24/27)
<i>matK</i> + <i>rbcLa</i>	37% (9/24)	87% (21/24)
ITS2+ <i>matK</i> + <i>rbcLa</i>	37% (9/24)	87% (21/24)

Table 3(on next page)

Discriminatory power of single and combination barcodes based on NJ and ML phylogenetic trees

*Number of nucleotide sequences #Column I: The percentage of conspecific monophyletic clusters. Column II: The percentage of conspecific monophyletic clusters with $\geq 99\%$ bootstrap support value. NJ, Neighbor Joining; ML, Maximum Likelihood

Table 3: Discriminatory power of single and combination barcodes based on phylogenetic trees

DNA barcodes	N*	Ability to discriminate		Ability to discriminate	
		(NJ) [#] (%)		(ML) [#] (%)	
		I	II	I	II
ITS2	27	87.50	25.00	100.00	88.88
<i>matK</i>	24	100.00	0.00	100.00	25.00
<i>rbcLa</i>	30	88.88	44.44	88.88	55.55
ITS2+ <i>matK</i>	24	100.00	75.00	100.00	55.55
ITS2+ <i>rbcLa</i>	27	100.00	88.88	100.00	50.00
<i>matK</i> + <i>rbcLa</i>	24	100.00	50.00	100.00	75.00
ITS2+ <i>matK</i> + <i>rbcL</i>	24	100.00	87.50	100.00	87.50
a					

Figure 1

Relative abundance of intra- and interspecific K2P pairwise distance for single and combination barcodes

(A) ITS2. (B) *matK*. (C) *rbcLa*. (D) ITS2+*matK*. (E) ITS2+*rbcLa*. (F) *matK*+*rbcLa*. (G) ITS2+*matK*+*rbcLa*.

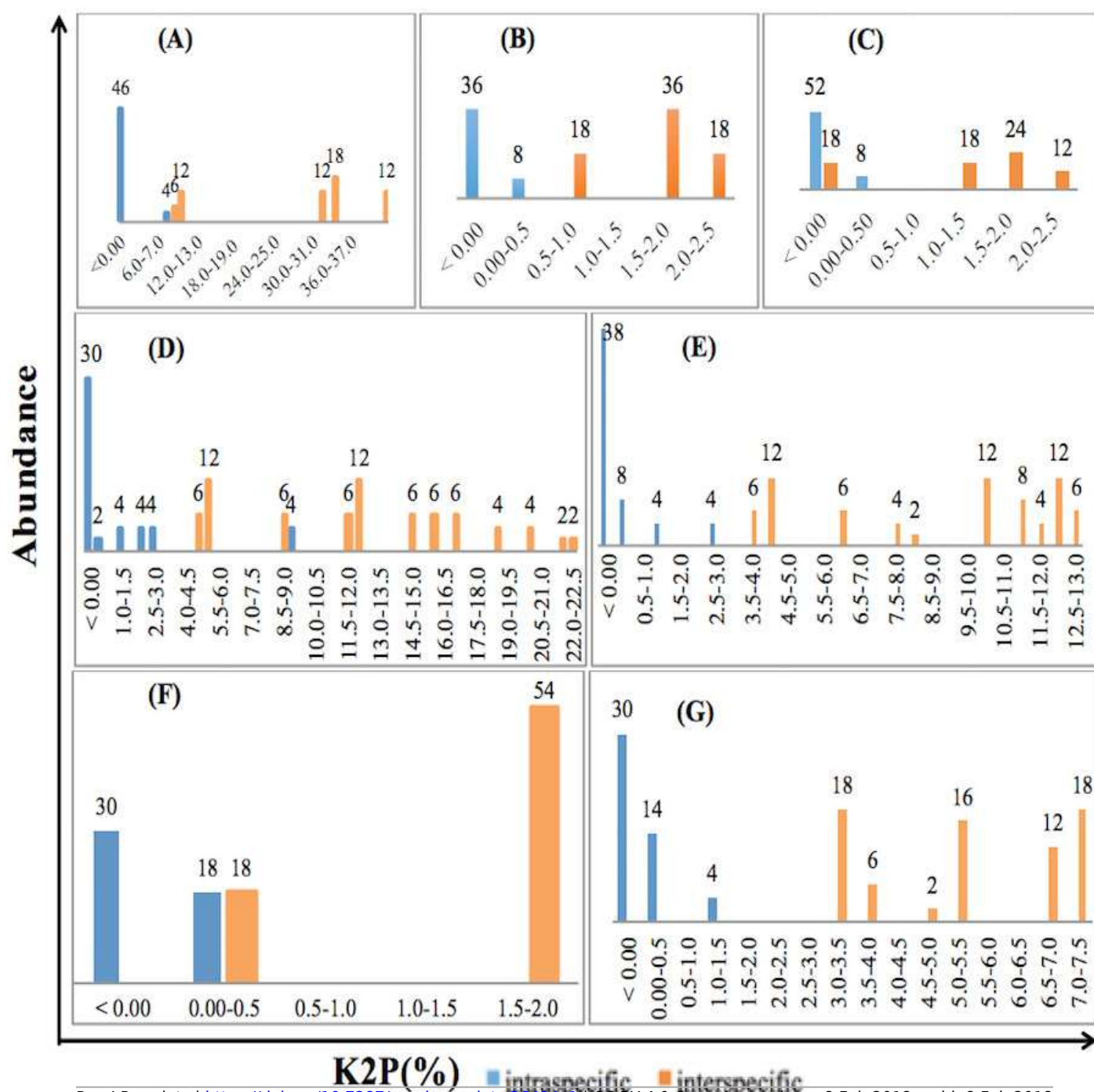


Figure 2

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

