DNA Barcode Reference Library for Indian Medicinal Plants of High Trade Volume

Saloni Malik and Shashi B. Babbar
Department of Botany, University of Delhi, Delhi 110007, India

Background

- Medicinal plants constitute a numerically large group of economically important resource which are of high commercial value. According to National Medicinal Plant Board of India, 6000-7000 species are used as medicinal plants in various alternative systems of medicine, such as, Ayurveda, Siddha, Unani and Homoeopathy and/or as folk medicine. Of these, 960 are traded with 178 species having annual consumption in excess of 100 metric tons.
- Traditional methods of identification based on macroscopic and microscopic morphological characters fail to establish botanical identity of herbal samples, if these are presented in powdered/fragmented/vegetative form.
- Therefore, for identification of herbals to the species level fool proof identification methods are required to check the problems of substitution and adulteration, quite prevalent in medicinal plant trade.
- DNA barcoding, if standardized beforehand, could be an effective tool for the rapid identification of species, authenticating the herbals and for discriminating the adulterants from the actual medicinal plant.
- A number of loci from the chloroplast genome and one from the nuclear genome have been tested as possible barcodes for plants. However, no single locus has been found to be an effective barcode for the plants. Based on the comparison of species discrimination rates of seven chloroplast loci, a combination of *matK* and *rbc*L was proposed as the core barcode for land plants by the Consortium for the Barcode of Life (CBOL).
- A number of studies on diverse group of plants have demonstrated ITS/ITS2 to be a highly effective locus for the discrimination of species (Chen et al. 2010, Li et al. 2011, Singh et al. 2012).

Objectives

- *In silico* analysis to gain a prior insight into the efficacy of the above-mentioned locus/loci as DNA barcodes for the medicinal plants that belong to diverse taxonomic groups, before testing of these locus/loci through 'wet' research.
- To validate the results obtained through *in silico* approach through "wet' research on selected medicinal plants.

Materials And Methods

In silico Analysis

- The sequences of three most potential barcode loci, nrITS, *matK* and *rbcL*, of 500 medicinal plants belonging to 442 genera and 117 families, available on NCBI GenBank were downloaded. These were checked for their uniqueness for the species by BLAST analysis on NCBI.
- If the query sequence matched with its own with 100% similarity, it was considered to be unique for the species. However, if it had 100% similarity with the sequence(s) of other species/genus also, the sequence was not considered to be a possible recognition tag for the species.
- Species identification rate was calculated according to the formula:

$$\frac{\text{Number of species uniquely identified} \times 100}{\text{Total number of species}}$$

Experimental Analysis

- The three loci (ITS, *matK*,and *rbcL*), which in combination provided 100% species recognition through *In silico* analysis, along with *rpoC1* were used to develop DNA barcodes of 244 individuals, belonging to 88 species of the medicinal plants (Table 1, Figure 1).
- Whole plants or twigs of these medicinal plants were collected from Pachmarhi (Madhya Pradesh), Dehradun, Mussoorie, Dhanaulti, and adjoining areas (Uttarakhand) and Shillong (Meghalaya). The botanical identity of the plants was confirmed by matching the collected plants with the Herbarium specimens available at Botanical Survey of India (BSI), Dehradun.
- The herbarium specimens of the collected plants were prepared and deposited in Delhi University Herbarium (DUH). The accession numbers obtained are DUH 13556-13587, DUH 13677-13722, DUH 13693-13738, DUH 13751-13864, DUH 13870-13926, DUH 14169-14215.
- Genomic DNAs from all the samples were extracted using CTAB method (Doyle and Doyle, 1987). The selected loci were amplified and sequenced using the primer pairs listed in Table 2 following standard procedure.
- Some of the sequences have been submitted to NCBI Genbank and accession numbers obtained (KJ667606-KJ667679, KJ49865-KJ749960, KM887355-KM887433, KJ499918-KJ499986). Rest of the sequences would be submitted.
- Species identification success rate for each locus was determined on the basis of BLAST search performed on NCBI.

Figure 1. Some of the investigated medicinal plants. (a) *Acorus calamus*, (b) Picrorhiza kurroa, (c) *Bergenia ligulata*, (d) *Digitalis lanata*, (e) *Anacyclus pyrethrum*, (f) *Taxus wallichiana*, (g) *Aconitum ferox*, (h) *Polygonatum multiflorum*, (i) *Podophyllum hexandrum*

Table 2. List of primers used for the amplification/ Sequencing of DNA barcode loci in the present study

S.No	Locus	Primer Name	Primer Sequence
1	*rpoC1*	rpoC1 2F	5′-GGCAAAGAGGGAAGATTTC
		rpoC1 4R	5′-CCATAAGCATATCTTGAGTTGG
		rpoC1 3R	5′-TGAGAAAACATAAGTAAACCGGC
2	*rbcL*	rbcLa for	5′-ATGTCACCACAAACAGAGACTAAAGC
		rbcLa rev	5′-GTAAAATCAAGTCCACCCRCG
		rbcL 1F	5′-ATGTCACCACAAACAGAAAC
		rbcL 724R	5′-TCGCATGTACCTGCAGTAGC
3	*matK*	3F KIM A	5′-CGTACAGTACTTTTGTGTTTACGAG
		1R KIM	5′-ACCCAGTCCATCTGGAAATCTTGGTTC
4	ITS	ITS 4F	5′-TCCTCCGCTTATTGATATGC
		ITS 5R	5′-GGAAGTAAAAGTCGTAACAAGG

Results

In silico analysis

- Out of 500 ITS sequences of medicinal plant species downloaded from NCBI, 481 (96.2%) were unique for the plant, implying that they have 100 percent identity only with its own sequence. Of the rest 19 sequences, 15 showed cent percent identity with ITS sequences of other species of the same genus.
- Of the five hundred species, *matK* sequences (>900 bp in length) of only 278 and *rbcL* sequences (>600 bp in length) of 302 species were available on NCBI GenBank. Of these 254 (91.3%) of *matK* and 243 (80.4%) of *rbcL* were unique. Among the rest *matK* sequences of 24 species, four showed 100% similarity with *matK* sequences of species belonging to the other genera. Remaining 20 were identical to sequences of other species of the same genera. Of the *rbcL* sequences of 43 species that did not yield correct identification, 27 had 100% similarity with other species of the same genera (Fig. 2).
- *matK* + *rbcL* combination available for 203 species provided a species resolution of 98% while, ITS and *matK* combination yielded 99.6% species resolution with only one species, *Aquilegia vulgaris*, not being assigned correctly. Combination of ITS + *rbcL* resolved 98.6% of the species. All the species could be identified correctly if the combination of all the three loci (ITS + *matK* + *rbcL*) was used (fig.2).

Experimental Analysis

- The amplification and sequencing success rates of different loci of 244 accessions of 88 species are presented in **Table 3.** BLAST search on NCBI revealed that the ITS sequences of 85.9% species had 100% similarity only with its own species, whereas this value for *matK*, *rbcL* and *rpoC1* were 76.9, 59.4 and 61.7%, respectively.
- The highest success of 97% correct species identification was obtained with the combination of ITS+*matK*+*rpoC1*, closely followed by ITS+*matK*+*rbcL* with corresponding value of 96.8%.
- The species identification percentage by CBOL suggested barcode, *matK*+*rbcL* was 91.4%, slightly higher than 88.5 afforded by *matK*+*rpoC1* (Fig. 3). A three locus barcode comprising *matK*+*rpoC1*+*rbcL*, provided species identification success of 94.1%.

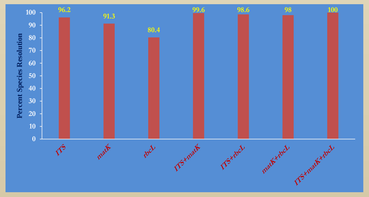

Figure 2. Percent species identification based on single as well as multi-locus combinations of the tested loci based on *in silico* analysis.

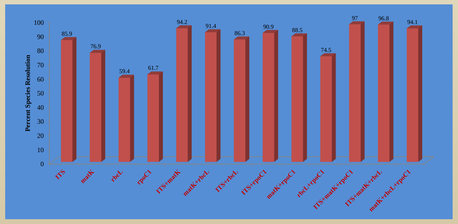

Figure 3. Percent species identification based on single as well as multi-locus combinations of the tested loci for the 88 medicinal plant species.

Table. 1. List of the investigated medicinal plants

S.No	Plant Name	S.No	Plant Name	S.No	Plant Name
1	*Aconitum ferox*	31	*Dioscorea deltoidea*	61	*Pterocarpus santalinus*
2	*Aconitum heterophyllum*	32	*Elaeodendron glaucum*	62	*Radar machera*
3	*Acorus calamus*	33	*Elettaria cardamomum*	63	*Randia dumetorum*
4	*Anacyclus pyrethrum*	34	*Embelia ribes*	64	*Ranunculus sceleratus*
6	*Aquilaria malaccensis*	36	*Flemingia macrophylla*	65	*Rauvolfia serpentina*
7	*Artemisia annua*	37	*Gardenia latifolia*	66	*Rauvolfia tetraphylla*
8	*Bacopa monnieri*	38	*Gentiana kurroo*	67	*Roscoea purpurea*
9	*Barleria cristata*	39	*Gloriosa superba*	68	*Rubia cordifolia*
10	*Barleria prionitis*	40	*Glycyrrhiza glabra*	79	*Saussurea costus*
11	*Berberis aristata*	41	*Gymnema sylvestre*	70	*Sida cordifolia*
12	*Bergenia ligulata*	42	*Hedychium coronarium*	71	*Skimmia anquetilia*
13	*Bixa orellana*	43	*Hedychium spicatum*	72	*Spilanthes acemella*
14	*Boerhavia diffusa*	44	*Helicteres isora*	73	*Sterculia villosa*
15	*Buchanania lanzan*	45	*Hemidesmus indicus*	74	*Swertia chirayita*
16	*Carum carvi*	46	*Hiptage benghalensis*	75	*Taraxacum officinalis*
17	*Celastrus paniculatus*	47	*Hollarhena antidysentrica*	76	*Taxus wallichiana*
18	*Centella asiatica*	48	*Hydrocotyle sibthorpioides*	77	*Terminalia bellerica*
19	*Cinnamomum camphora*	49	*Ichncarpus frutescens*	78	*Terminalia chebula*
20	*Cinnamomum tamala*	50	*Kaempferia galanga*	79	*Thalictrum foliolosum*
21	*Cissus quadrangularis*	51	*Litsea glutinosa*	80	*Tylophora indica*
22	*Clematis gouriana*	52	*Mallotus philippensis*	81	*Valeriana wallichii*
23	*Coptis teeta*	53	*Murraya paniculata*	82	*Vanda coerulea*
24	*Desmodium bracteata*	54	*Panax pseudoginseng*	83	*Ventilago madraspatana*
25	*Desmodium gangeticum*	55	*Picrorhiza kurroa*	84	*Woodfordia fruticosa*
26	*Desmodium gyrans*	56	*Plantago major*	85	*Wrightia arborea*
27	*Digitalis lanata*	57	*Plumbago zeylanica*	86	*Wrightia tinctorea*
28	*Digitalis purpurea*	58	*Podophyllum hexandrum*	87	*Wrightia tomentosa*
29	*Dillenia pentagyna*	59	*Polygonatum multiflorum*	88	*Zanthoxylum armatum*
30	*Dioscorea bulbifera*	60	*Polygonatum verticillatum*		

Table 3. Amplification, sequencing success and species identification rates for the four candidate loci based on 244 accessions.

Locus	No. of amplicons obtained	Amplification success	No. of finished sequences generated	Sequencing success	Species identification on the basis of BLAST (%)
ITS	232	95%	161	69.4%	85.9
matK	126	51.6%	83	65.9%	76.9
rbcL	238	97.5%	183	76.9%	59.4
rpoC1	203	83%	152	74.9	61.7

Conclusions

1. Demonstrates effectiveness of *in silico* approach in gaining a prior insight into the possible barcodes for a group of plants belonging to diverse taxonomic groups. This approach revealed relative efficacy of ITS, *matK* and *rbcL* in species identification in the same order, **ITS+*matK*+*rbcL*** providing **100%** species identification success.
2. The results of *in silico* analysis were also validated by 'wet' research, where the relative efficacy of the three loci remained same. However, *rpoC1* not included in *in silico* analysis proved to be slightly better than *rbcL*.
3. The maximum species identification success of **97%** was by the three-locus barcode comprising **ITS+*matK*+*rpoC1*,** closely followed by the combination where *rpoC1* is replaced by *rbcL*. A three locus barcode comprising *matK*+*rbcL*+*rpoC1* too provided more than 90% species identification success.
4. Among the two locus barcodes comprising loci from the chloroplast genome only, the combination of *matK*+*rbcL*, suggested by the CBOL Plant Working Group as the possible barcode for plants, proved to be the best. However, among two-locus barcodes, ITS+*matK* was the best.
5. The need for a multi-locus approach and **the inclusion of ITS**, wherever needed and available, in the core barcode is highlighted by the study, a global consensus seems to emerging by other studies too, in recent past.

References

1. CBOL Plant Working Group (2009) A DNA barcode for land plants. Proceedings of the National Academy of Sciences USA 106: 12794 – 12797.
2. Chen S, Yao H, Han J, Liu C, Song J, Shi L, Zhu Y, Ma X, Gao T, Pang X, Luo K, Li Y, Li X, Jia X, Lin Y, Leon C (2010) Validation of ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS One 5:e8613
3. China Plant BOL Group, Li DZ, Gao LM, Li HT, Wang H, Ge XJ, Liu JQ, Chen ZD, Zhou SL, Chen SL, Yang JB, Fu CX, Zeng CX, Yan HF, Zhu YJ, Sun YS, Chen SY, Zhao L, Wang K, Yang T, Duan GW (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 the National Academy of Sciences USA 108:19641 – 19646.
4. Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19: 11–15.
5. Singh HK, Parveen I, Raghuvanshi S, Babbar SB (2012) The loci recommended as universal barcodes for plants on the basis of floristic studies may not work with congeneric species as exemplified by DNA barcoding of *Dendrobium* species. BMC Research Notes 5:42.

Acknowledgement

The research work was supported by a research project by the Indian Council of Medical Research , New Delhi (ICMR) and a Research & Development grant by the University of Delhi awarded to SBB. SM thankfully acknowledges the receipt of Junior and Senior Fellowships by ICMR.