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

A tiered barcode authentication tool to differentiate medicinal *Cassia* species in India

N. Purushothaman[1], S.G. Newmaster[2], S. Ragupathy[2], N. Stalin[1], D. Suresh[1,3], D.R. Arunraj[1], G. Gnanasekaran[1,4], S.L. Vassou[1], D. Narasimhan[5] and M. Parani[1]

[1]Genomics Laboratory, Department of Genetic Engineering, SRM University, Chennai, India
[2]Centre for Biodiversity Genomics, University of Guelph, Guelph, Ontario, Canada
[3]Present address: Department of Plant Science, South Dakota State University, Brookings, SD, USA
[4]Present address: Botanical Survey of India, Southern Regional Center, Coimbatore, India
[5]Centre for Floristic Research, Department of Plant Biology and Plant Biotechnology, Madras Christian College, Chennai, India

Corresponding author: M. Parani
E-mail: hod.gene@ktr.srmuniv.ac.in

Genet. Mol. Res. 13 (2): 2959-2968 (2014)
Received March 27, 2013
Accepted October 3, 2013
Published April 16, 2014
DOI http://dx.doi.org/10.4238/2014.April.16.4

ABSTRACT. DNA barcoding is a desirable tool for medicinal product authentication. DNA barcoding is a method for species identification using short DNA sequences that are conserved within species, but variable between species. Unlike animals, there is no single universal DNA barcode locus for plants. Coding markers, *matK* and *rbcL*, and noncoding markers, *trnH-psbA* (chloroplast) and *ITS2* (nuclear), have been reported to be suitable for the DNA barcoding of plants with varying degree of success. Sixty-four accessions from 20 species of the medicinal plant *Cassia* were collected, and analyzed for these 4 DNA

barcoding markers. PCR amplification was 100% successful for all 4 markers, while intra-species divergence was 0 for all 4 *Cassia* species in which multiple accessions were studied. Assuming 1.0% divergence as the minimum requirement for discriminating 2 species, the 4 markers could only differentiate 15 to 65% of the species studied when used separately. Adding indels to the divergence increased the percentage of species discrimination by *trnH-psbA* to 90%. In 2-locus barcoding, while *matK*+*rbcL* (which is recommended by Consortium for the Barcoding of Life) discriminated 90% of the species, the other combinations of *matK*+ITS and *rbcL*+*trnH-psbA* showed 100% species discrimination. However, matK is plagued with primer issues. The combination of *rbcL*+*trnH-psbA* provided the most accurate (100% species ID) and efficient tiered DNA barcoding tool for the authentication of *Cassia* medicinal products.

Key words: *Cassia*; DNA barcoding; trnH-psbA; matK; rbcL; ITS2

INTRODUCTION

Cassia L. is a highly valued medicinal plant that has considerable commercial importance, but is jeopardized by product substitution. The traditional Indian system of medicine recognizes several *Cassia* species with medicinal value (Singh, 2001; Cho et al., 2007; Khare, 2007; Farswan et al., 2009), of which some are traded in the range of 100 to 10,000 metric tons per year as commercial products (Ved and Goraya, 2007). While seeds are commonly used for producing medicines, leaves, bark, root, and flower are also used. These plant parts are chiefly collected from wild-grown plants, with the possibility of accidental species admixtures, misidentification due to morphological resemblances, and exploitation due to its high commercial value. Species authentication using DNA barcoding presents a desirable tool, but has yet to be validated in its ability to authentic *Cassia* species used in medicinal products.

Cassia is mainly distributed in tropical to warm temperate regions of the world. There are about 645 species in this genus (Mabberley, 2005) belonging to the family Caesalpiniaceae. The diversity of life forms in *Cassia* is high, and includes herbs, shrubs, and small to large trees that colonize a wide range of habitats. Based on morphological traits, Irwin and Barneby (1981) have presented a revised classification, in which the genus *Cassia* was elevated to the subtribe level as Cassiinae and *Senna* Mill. based on the straightness or curvature of the filaments, presence or absence of bracteole, and dehiscence nature of fruits. The revised classification is widely accepted in many parts of the world, including India. A monographic study on Cassiinae in India by Singh (2001) documented considerable species diversity in *Senna* (43 species), *Chamaecrista* (11 species), and *Cassia* (8 species). However, other studies based on morphological and molecular data have reported that the monophyletic relationship among *Cassia*, *Chamaecrista*, and *Senna* remains unclear (Doyle et al., 2000; Bruneau et al., 2001; Kajita et al., 2001). Hence, in the present study, we have treated all the species under the broader sense, i.e., *Cassia* L.

At present, there is no standard DNA barcode region for herbal product authentication. In 2009, the Consortium for the Barcoding of Life (CBOL) evaluated 7 promising DNA

barcoding loci in plants, which was based on a large data set. Based on species discrimination, the trnH-psbA intergenic region was ranked the best single-locus barcode; however, it had problems in obtaining high-quality bidirectional sequences, and showed extensive size variations. Therefore, CBOL has recommended matK and rbcL as a 2-locus standard barcode for plants. However, the matK locus requires improvements in primers to enable universal amplification. In addition, the combined discriminatory power of the 2 selected loci was not as high as that of other combinations from different regions (CBOL Plant Working Group, 2009). As a result, CBOL proposed to review the research to assess the performance of these 2 markers after 18 months, and suggested to continue the study on other markers during this review period. Recently, the trnH-psbA intergenic region was reported to differentiate many medicinal plants, including 17 species of *Dendrobium* (Yao et al., 2009), 16 species of *Phyllanthus* (Srirama et al., 2010), and 8 species of *Panax* (Zuo et al., 2011). In addition, research in China has documented success in barcoding medicinal plants using ITS2 (Chen et al., 2010; China Plant BOL Group et al., 2011; Sui et al., 2011). Ultimately, to provide an effective authentication barcode, research and development for each medicinal plant of interest is required, which considers all of the candidate regions that have been proposed in the published literature including matK, rbcL, trnH-psbA, and ITS2.

Here, we conducted research and development of a barcode authentication tool to differentiate medicinal species of *Cassia* in India. We investigated different combinations of several candidate barcode regions (matK, rbcL, trnH-psbA, and ITS2) to produce the most accurate and efficient multi-locus DNA barcoding tool for the authentication of *Cassia* products.

MATERIAL AND METHODS

Leaf samples from 64 accessions belonging to 20 species of *Cassia* were collected from different parts of Tamil Nadu, India. This collection included one accession for each species, except *C. tora*, *C. occidentalis*, *C. senna*, and *C. italic*, for which 12 accessions were collected for each species. Total genomic DNA from all the samples was isolated according to Saghai-Maroof et al. (1984), and 30-40 ng DNA was used as template for PCR. Primers for *trnH-psbA* (Code1F and Code1R) were designed by modifying the universal primers (psbA3_f and trnHf_05) reported by Kress et al. (2005). Code1F (GTAAAACGACGGCCAGTGCGGC CGCGTTATGCATGAACGTAATGCTC) contains the M13F primer sequence (underlined) followed by a spacer with an *Not*I recognition sequence and psbA3_f primer sequence. Code1R (CAGGAAACAGCTATGACAGTCGACCGCGCATGGTGGATTCACAATCC) contains the M13R primer sequence (underlined) followed by a spacer with an *Sal*I recognition sequence and trnHf_05 primer sequence. Primers already reported for *matK* (*MatK*-1RKIM-f and *MatK*-3FKIM-r, Ki-Joong Kim, School of Life Sciences and Biotechnology, Korea University, Korea, unpublished results), *rbcL* (*rbcL*a-F; Levin et al., 2003), *rbcL* ajf634R (Fazekas et al., 2008), and *ITS2* (Chen et al., 2010) were used without any modifications. PCR amplification was carried out in GeneAmp 9700 (Applied Biosystems, CA, USA) for 35 cycles, with a program setting of: initial denaturation at 94°C for 3 min, 35 cycles at 94°C for 30 s, 55°C for 30 s, and 72°C for 1 min, and final extension at 72°C for 5 min. The same program was used for the PCR amplification of all 4 barcode markers. The PCR products were separated by 1.0% agarose gel electrophoresis, and purified from the gel by using EZ-10 Spin Column DNA Gel Extraction Kit (Bio Basic Inc., Ontario, Canada). About 30 ng of gel purified PCR product was collected

for automated DNA sequencing by a 3130xl Genetic Analyzer (Applied Biosystems). All PCR products were sequenced from both ends using M13F and M13R primers for *trnH-psbA* and the respective PCR primers all other markers.

The sequences were manually edited, and full-length sequences were obtained for all the PCR-amplified markers. Intra-species sequence variation was not found in any of the 4 species where 12 accessions per species were analyzed. Therefore, a DNA sequence from 1 accession per species was used for the pairwise alignments of the 20 species using the Needleman and Wunsch algorithm (http://www.ebi.ac.uk/Tools/psa/). Divergence was calculated as the percentage of mismatched nucleotides over the total number of aligned nucleotides. In the case of trnH-psbA, divergence, including indels, was calculated by adding the number of gaps to the number of mismatched nucleotides, and then calculating it as a percentage over the total number of aligned nucleotides. It has been suggested that 10 times the mean intra-species variation should be set as standard sequence threshold to differentiate congeneric species (Hebert et al., 2004). Because intra-species divergence was 0 when estimated by the 4 markers from the 4 species in which multiple accessions were studied, 1% pairwise divergence was empirically set as the cut-off to authenticate the differentiation between 2 species.

The relationship of the classification structure of the species data to the molecular characters was analyzed with nonmetric multidimensional scaling (NMS) using the "R" software (Kruskal, 1964; Newmaster et al., 2008; R Core Team, 2012). In NMS, the Bray-Curtis distance measure was used because of its robustness for both large and small scales on the axes. Data were standardized by species maxima and two-dimensional solutions were appropriately selected based on plotting a measure of fit ("stress") to the number of dimensions. Stress represents distortion in the data, whereby a stress value over 0.15 indicates that the results are invalidated (Primer Software, 2002). One thousand iterations were used for each NMS run, using random start coordinates. The first 2 ordination axes were rotated to enhance interpretation with the different axes.

RESULTS

Four DNA barcoding markers *matK*, *rbcL*, *trnH-psbA*, and *ITS2* were PCR amplified from 64 accessions that belong to 20 species of the genus *Cassia*, and full-length sequences were obtained by bidirectional sequencing. PCR amplification was 100% successful, confirming the suitability of the universal primers used in this study. As expected, size variation was not observed with the coding markers, rbcL and matK, whereas trnH-psbA showed the highest size variation. The size of the *matK* and *rbcL* markers was 836 and 607 bp, respectively. The size of the *ITS2* marker showed minimal variation between 456 and 462 bp, except for 2 species with 429 and 430 bp. In contrast, the size of the *trnH-psbA* marker varied widely between 279 and 481 bp. The *trnH-psbA* marker was analyzed for the presence of microsatellites containing a minimum of six 'A' or 'T' mononucleotide repeats. Among the 20 *Cassia* species that were analyzed, 17 and 71, 'A' and 'T' repeats were obtained, respectively. The maximum length of the mononucleotide and dinucleotide repeats observed was 9 and 5, respectively. In addition, the pairwise alignment of the sequences in this marker revealed the presence of 1 to 71 indels in different species pairs. Divergence was separately calculated for all pairwise combinations of the 4 markers, and in combinations of 2 and 3 markers (**Appendix S1**). Intra-species divergence was 0 in the 4 species for which multiple accessions were analyzed.

For the single-locus barcode, the minimum divergence ranged between 0.0 and 0.27%, and species discrimination ranged between 15 and 65% for the 4 markers. When indels in the

trnH-psbA marker were included, the minimum divergence increased from 0.27 to 0.55%, and species discrimination increased from 65 to 90%. Among the 6 2-locus barcode combinations that are possible with the 4 markers, the minimum divergence ranged between 0.2 and 2.7%, and species discrimination ranged between 90 and 100%. Among the 4 possible 3-locus barcode combinations, minimum divergence ranged between 1.84 and 3.37%, while species discrimination was 100% for all combinations (Table 1).

Table 1. Percentage of divergence and species discrimination based on pairwise alignment of *matK*, *rbcL*, *trnH-psbA* and *ITS2* markers.

Marker	Divergence (%)			Percentage of species discrimination*
	Minimum	Maximum	Mean	
matK	0	4.8	2.3	40
rbcL	0	3.5	1.3	15
trnH-psbA (excluding indels)	0.27	9.62	4.3	65
trnH-psbA (including indels)	0.55	19.47	11.3	90
ITS2	0	30.8	15.5	55
matK+rbcL	0.20	7.8	3.7	90
matK+trnH-psbA	0.87	13.35	6.6	90
matK+ITS2	2.7	35.9	17.87	100
rbcL+trnH-psbA	1.75	12.39	5.6	100
rbcL+ITS2	0.2	33.4	16.87	90
trnH-psbA+ITS2	0.87	36.35	19.85	90
matK+rbcL+trnH-psbA	1.84	15.85	8.0	100
matK+rbcL+ITS2	3.3	37.2	19.2	100
matK+trnH-psbA+ITS2	3.37	40.45	22.19	100
rbcL+trnH-psbA+ITS2	2.37	38.85	21.18	100

*Considering minimum 1.0% divergence for discriminating two species.

Although *rbcL* does not differentiate all of the *Cassia* species, it does provide an appropriate 1st tier that is complemented by other regions in a 2-region barcode. The NMS ordination of *rbcL* indicates that some of the *Cassia* species exhibit relatively invariant sequences (Figure 1). The NMS classification of the combined dataset (Figures 2, 3 and 4) showed that 2 regions are complementary, but that not all combinations of regions are equal. For instance, both the *rbcL+matK* and *rbcL+ITS2* combinations do not resolve all *Cassia* species (Figure 2). In fact, the poorest combination was *rbcL+ITS2* (Figure 4). The combination of *rbcL+trnH-psbA* provided considerable variation, which could be used to identify all species (Figure 3).

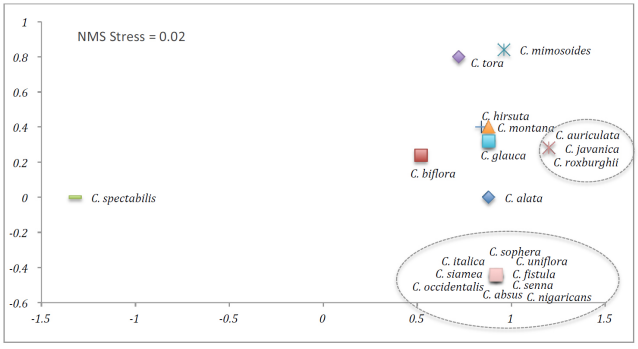

Figure 1. NMS ordination of 20 species of *Cassia* using *rbcL* sequence data. Gray circles represent species that exhibit relatively invariant sequences.

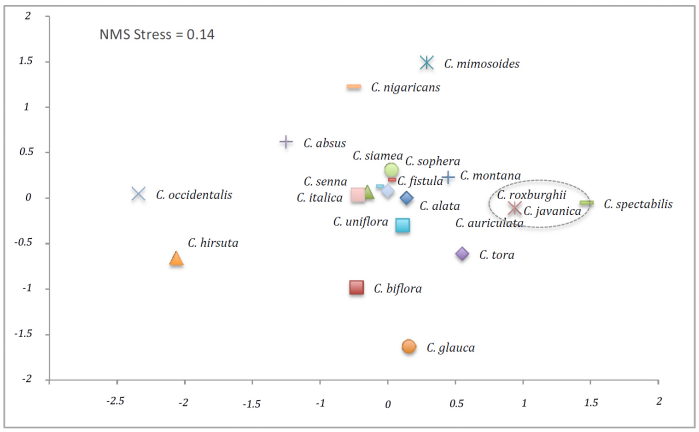

Figure 2. NMS ordination of 20 species of *Cassia* using *rbcL+matK* sequence data. Gray circle represents species that exhibit relatively invariant sequences.

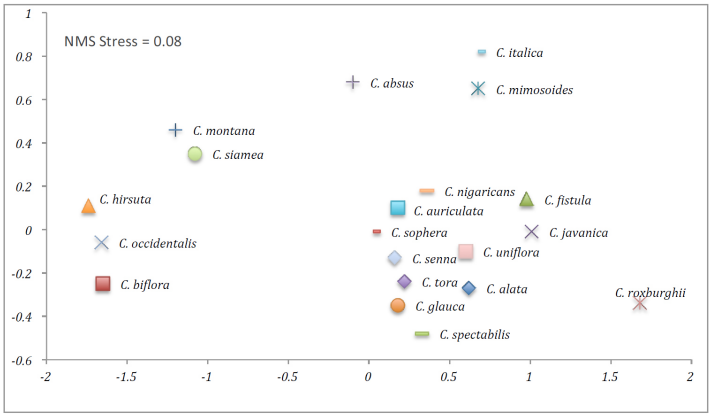

Figure 3. NMS ordination of 20 species of *Cassia* using *rbcL+trnH-psbA* sequence data.

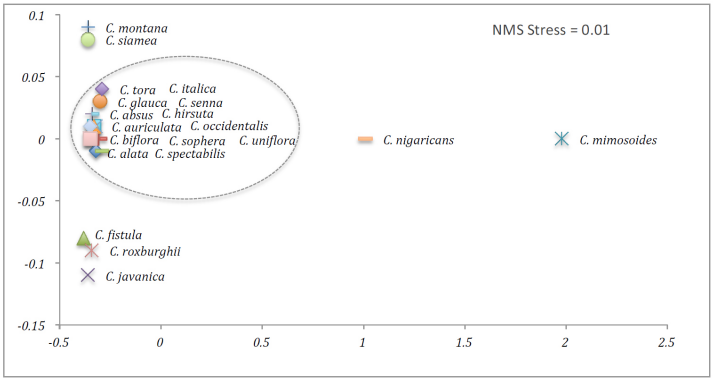

Figure 4. NMS ordination of 20 species of *Cassia* using *rbcL+ITS2* sequence data. Gray circle represents species that exhibit relatively invariant sequences.

DISCUSSION

Species identification after phylogenetic analysis requires correct multiple sequence alignment, which is time-consuming and highly complicated if multiple indels are present. Therefore, pairwise alignment was conducted for all possible combinations, and divergence between species pairs was calculated to quantify the ability of the markers to differentiate the species.

As separate single-locus markers, none of the 4 markers could differentiate all 20 *Cassia* species. Therefore, the possibility of improving the species discriminating ability of the *trnH-psbA* marker was explored by including the indels, which are *bonafide* heritable variations, and reliable phylogenetic characteristics (Lloyd and Calder, 1991; Van Dijk et al., 1999). Although, while minimum divergence did not increase significantly after including indels, mean divergence did, with species discrimination increasing from 65 to 90%, which is highly desirable for a single-locus barcode. However, the real concern with this marker is the formation of stutter PCR products, because of the presence of 'A' or 'T' mononucleotide repeats, which affect sequence quality in some samples (Devey et al., 2009). While mononucleotide repeats with 8 or more nucleotides are highly vulnerable to the formation of stutter products in PCR with *Taq* DNA polymerase (Shinde et al., 2003), their formation may be avoided in up to 15 mononucleotides by using Phusion DNA polymerase (Fazekas et al., 2010). In our samples, the maximum length of the mononucleotide repeats was 9 bases; therefore, we experienced no sequencing difficulties, even though PCR was performed using *Taq* DNA polymerase.

Another potential problem with *trnH-psbA* marker is that its size may be as large as 1094 bp (Kress et al., 2005), which exceeds the sequencing read length in Sanger's method. Therefore, M13F and M13R primers and the spacer sequences were added to enable us to read sequences that were closer to universal primers. We were able to read bases immediately after the 3'-end of the primers; however, this was not of added advantage because the maximum sized *trnH-psbA* marker in our samples was just 481 bp. While the additional sequences in the primers did not affect the success rate of PCR amplification in our samples, these sequences require systematic evaluation to identify any negative effect on PCR amplification.

Because the single-locus barcode did not provide sufficient discriminatory power, all possible 2-locus combinations were evaluated with the 4 markers to determine their ability to discriminate the species. The 2-locus barcode recommended by CBOL, using the 2 coding markers *matK* and *rbcL*, only showed 90% species discrimination. A similar level of species discrimination was observed when the 2 noncoding markers (*ITS2+trnH-psbA*) were combined. Two specific combinations of coding and noncoding markers, *matK+ITS2* and *rbcL+trnH-psbA*, showed 100% species discrimination, based on divergence in pairwise comparisons. However, multivariate analysis using NMS clearly showed that *rbcL+trnH-psbA* exhibits stronger discriminatory power than *matK+ITS2*.

The literature supports a multiregion approach to the barcoding of plants (Kress et al., 2005; Cowan et al., 2006; Newmaster et al., 2006; Chase et al., 2005, 2007; Kress and Erickson, 2007). Previously, Newmaster et al. (2006) proposed a tiered, or nested, approach to this type of analysis as one way of overcoming the issue of alignment with noncoding regions, such as the *trnH-psbA* spacer. The tiered approach is based on the use of a common, easily amplified, and aligned region, such as *rbcL*, which acts as scaffold on which to place data from a highly variable noncoding region, such as *trnH-psbA*. Evaluation of the use of *rbcL* as a primary tier barcode demonstrated that it is appropriate because of its universality, ease of amplification, ease of alignment, and because there is a significant body of data available for evaluation (New-

master et al., 2006; Chase et al., 2007); the analysis of these data has shown that *rbcL* is able to differentiate a large percentage of congeneric plant species (Newmaster et al., 2006). Our research suggests that the best combination for discriminating *Cassia* species is *rbcL+trnH-psbA*, which showed 100% species discrimination. Although we used a multivariate analysis to evaluate sequence variation, there are a number of searching algorithms that are able to utilize the aligned portion to nest or localize a search, in addition to combinations of clustering and similarity methods (Little and Stevenson, 2007) or the analysis of multi-locus plant barcodes with a composition vector method based on adjustable weighted distance (Li et al., 2012).

In conclusion, our research provides a tiered barcode authentication tool to differentiate the medicinal *Cassia* species in India. The combination of *rbcL+trnH-psbA* provides the most accurate (100% species ID) and efficient multi-locus DNA barcoding tool for the authentication of *Cassia* medicinal products. Single-locus barcoding did not differentiate the studied 20 *Cassia* species, with this limitation being supported by research on many other species of plants. Although the addition of indels to the pairwise divergence in *trnH-psbA* seems to increase the percentage of species discrimination, further confirmation is required after a more complete sampling of the genus. The multivariate analysis used here revealed considerable sequence variation that might easily differentiate all species of *Cassia*. Considering universal amplification and divergence, wherever necessary, *trnH-psbA* could serve as a supplementary marker to *rbcL* in a tiered barcode approach, in which other regions might be used as a second tier as needed. It is unlikely that more than 2 markers would be needed to sequence any specific group of plants.

Data accessibility

The nucleotide sequences *matK*, *rbcL*, *trnH-psbA*, and *ITS2* regions for 20 species of *Cassia* were submitted to the database of Genbank at the NCBI with the accession Nos. JQ301868 to JQ301887, JQ301848 to JQ301867, HQ161753 to HQ161772, and JQ301828 to JQ301847, respectively.

ACKNOWLEDGMENTS

We gratefully acknowledge the funding from the Central Council for Research on Ayurveda and Siddha, Department of AYUSH, Government of India. We also acknowledge Dr. P. Ravichandran, Sri Paramakalyani Centre for Environmental Science, Mamonmaniam Sundaranar University, Tamil Nadu, India and Dr. Chelladurai, Tiruvelveli, Tamil Nadu, India for their help with sample collection.

Supplementary material

REFERENCES

Bruneau A, Forest F, Herendeen PS, Klitgaard BB, et al. (2001). Phylogenetic relationships in Caesalpinioideae (Leguminosae) as inferred from chloroplast *trnL* intron sequence. *Syst. Bot.* 26: 487-514.

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

Chase MW, Salamin N, Wilkinson M, Dunwell JM, et al. (2005). Land plants and DNA barcodes: short-term and long-term goals. *Philos. Trans. R. Soc. Lond B Biol. Sci.* 360: 1889-1895.

Chase MW, Cowan RS, Hollingsworth PM, Berg CVD, et al. (2007). A proposal for a standardized protocol to barcode

all land plants. *Taxon* 56: 295-299.
Chen S, Yao H, Han J, Liu C, et al. (2010). Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS One* 5: e8613.
China Plant BOL Group, Gao LM, Li HT, Wang H, et al. (2011). Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proc. Natl. Acad. Sci. U. S. A.* 108: 19641-19646.
Cho IJ, Lee C and Ha TY (2007). Hypolipidemic effect of soluble fiber isolated from seeds of *Cassia tora* Linn. in rats fed a high-cholesterol diet. *J. Agric. Food Chem.* 55: 1592-1596.
Cowan RS, Chase MW, Kress JW and Savolainen V (2006). 300 000 species to identify: problems, progress and prospects in DNA barcoding of land plants. *Taxon* 55: 611-616.
Devey DS, Chase MW and Clarkson JJA (2009). A stuttering start to plant DNA barcoding: microsatellites present a previously overlooked problem in non-coding plastid regions. *Taxon* 58: 7-15.
Doyle JJ, Chappill JA, Bailey CD and Kajita T (2000). Towards a Comprehensive Phylogeny of Legume: Evidence from rbcL Sequence & Non-Molecular Data. In: Advances in Legume Systematics (Herendeen PS and Bruneau A, eds.). Royal Botanical Gardens, Kew.
Farswan M, Mazumder PM and Percha V (2009). Protective effect of *Cassia glauca* Linn. on the serum glucose and hepatic enzymes level in streptozotocin induced NIDDM in rats. *Indian J. Pharmacol.* 41: 19-22.
Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, et al. (2008). Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS One* 3: e2802.
Fazekas A, Steeves R and Newmaster S (2010). Improving sequencing quality from PCR products containing long mononucleotide repeats. *Biotechniques* 48: 277-285.
Hebert PD, Stoeckle MY, Zemlak TS and Francis CM (2004). Identification of Birds through DNA Barcodes. *PLoS Biol.* 2: e312.
Irwin HS and Barneby RC (1981). Tribe Cassieae Bronn. In: Recent Advances in Legume Systematics (Polhill RM and Raven PH, eds.). Royal Botanic Gardens, Kew.
Kajita T, Ohashi H, Tateishi Y, Bailey CD, et al. (2001). *RbcL* and legume phylogeny with particular reference to Phaseoleae, Millettieae & Allies. *Syst. Bot.* 26: 515-536.
Khare CP (2007). Indian Medicinal Plants - An Illustrated Dictionary. Springer-Verlag, Heidelberg.
Kress WJ and 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.
Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, et al. (2005). Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. U. S. A.* 102: 8369-8374.
Kruskal JB (1964). Non-metric multidimensional scaling: a numerical method. *Psychometrika* 29: 115-129.
Levin RA, Wagner WL, Hoch PC, Nepokroeff M, et al. (2003). Family-level relationships of Onagraceae based on chloroplast rbcL and ndhF data. *Am. J. Bot.* 90: 107-115.
Li CP, Yu ZG, Han GS and Chu KH (2012). Analyzing multi-locus plant barcoding datasets with a composition vector method based on adjustable weighted distance. *PLoS One* 7: e42154.
Little DP and Stevenson DW (2007). A comparison of algorithms for the identification of specimens using DNA barcodes: examples from gymnosperms. *Ladistics* 23: 1-21.
Lloyd DG and Calder VL (1991). Multi-residue gap, a class of molecular characters with exceptional reliability for phylogenetic analyses. *J. Evol. Biol.* 4: 9-21.
Mabberley DJ (2005). The Plant Book, A Portable Dictionary of Vascular Plants. Cambridge University Press, Cambridge.
Newmaster SG, Fazekas AJ and Ragupathy S (2006). DNA barcoding in the land plants: evaluation of *rbcL* in a multigene tiered approach. *Can. J. Bot.* 84: 335-341.
Newmaster SG, Fazekas AJ, Steeves RA and Janovec J (2008). Testing candidate plant barcode regions in the Myristicaceae. *Mol. Ecol. Resour.* 8: 480-490.
Primer Software (2002). Primer Multivariate Software Version 5.2.9. PRIMER-E Ltd. Hedingham Gardens, Roborough Plymouth.
R Core Team. R version 2.15.1. (2012). A Language and Environment for Statistical. The R Foundation for Statistical Computing.
Saghai-Maroof MA, Soliman KM, Jorgensen RA and Allard RW (1984). Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad. Sci. U. S. A.* 81: 8014-8018.
Shinde D, Lai Y, Sun F and Arnheim N (2003). Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)n and (A/T)n microsatellites. *Nucleic Acids Res.* 31: 974-980.
Singh V (2001). Monograph on Indian Subtribe Cassiinae (Caesalpiniaceae). Scientific Publisher, Jodhpur.

Srirama R, Senthilkumar U, Sreejayan N, Ravikanth G, et al. (2010). Assessing species admixtures in raw drug trade of *Phyllanthus*, a hepato-protective plant using molecular tools. *J. Ethnopharmacol.* 130: 208-215.

Sui XY, Huang Y, Tan Y, Guo Y, et al. (2011). Molecular authentication of the ethnomedicinal plant *Sabia parviflora* and its adulterants by DNA barcoding technique. *Planta Med.* 77: 492-496.

Van Dijk MA, Paradis E, Catzeflis F and de Jong WW (1999). The virtues of gaps: *Xenarthra* (Edentata) monophyly supported by a unique deletion in alpha A-crystallin. *Syst. Biol.* 48: 94-106.

Ved DK and Goraya GS (2007). Demand and supply of medicinal plants in India. National Medicinal Plants Board/ Foundation for Revitalisation of Local Health Traditions (FRLHT), Bangalore.

Yao H, Song JY, Ma XY, Liu C, et al. (2009). Identification of Dendrobium species by a candidate DNA barcode sequence: the chloroplast psbA-trnH intergenic region. *Planta Med.* 75: 667-669.

Zuo Y, Chen Z, Kondo K, Funamoto T, et al. (2011). DNA barcoding of *Panax* species. *Planta Med.* 77: 182-187.

