3 Biotech (2021) 11:463
https://doi.org/10.1007/s13205-021-03001-5

ORIGINAL ARTICLE

Quantification of adulteration in traded ayurvedic raw drugs employing machine learning approaches with DNA barcode database

Suma Arun Dev[1] · Remya Unnikrishnan[1,5] · R. Jayaraj[3] · P. Sujanapal[2] · V. Anitha[4]

Received: 1 March 2021 / Accepted: 26 September 2021 / Published online: 18 October 2021

Abstract

Adulteration of expensive raw drugs with inferior taxa has become a routine practice, conceding the quality and safety of derived herbal products. In this regard, the study addresses the development of an integrated approach encompassing DNA barcode and HPTLC fingerprinting to authenticate chiefly traded ayurvedic raw drugs in south India [viz. *Saraca asoca* (Roxb.) Willd., *Terminalia arjuna* (Roxb. ex DC.) Wight and Arn., *Sida alnifolia* L. and *Desmodium gangeticum* (L.) DC.] from its adulterants. Consortium of Barcode of Life (CBOL) recommended DNA barcode gene regions viz. nuclear ribosomal—Internal Transcribed Spacer (nrDNA-ITS), maturase K (*mat*K), ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (*rbc*L) and *psb*A-*trn*H spacer regions along with HPTLC profiling were experimented and a reference database was created. Further, an integrated analytical approach employing genetic distance-based Maximum Likelihood phylogenetic tree and Artificial Intelligence (AI)based Machine Learning Algorithms (MLA)-Waikato Environment for Knowledge Analysis (WEKA) and Barcoding with Logic (BLOG) were employed to prove efficacy of DNA barcode tool. Even though, among the four barcodes, *psb*A-*trn*H (*S. alnifolia* and its adulterants, *T. arjuna* and its adulterants) or ITS region (*S. asoca* and its adulterants, *D. gangeticum* and its adulterants) showed highest inter specific divergences in the selected Biological Reference Materials (BRMs), *rbc*L or *mat*K barcode regions alone were successful for authentication of traded samples. The automated species identification techniques, WEKA and BLOG, experimented for the first time in India for raw drug validation, could achieve rapid and precise identification. A national certification agency for raw drug authentication employing an integrated approach involving a DNA barcoding tool along with standard organoleptic and analytical methods can strengthen and ensure safety and quality of herbal medicines in India.

Keywords Machine Learning Algorithm · Artificial Intelligence · DNA barcoding · Ayurvedic raw drug

Introduction

India is renowned for its traditional medicine systems (Ayurveda, Siddha, Unani, Homeo, etc.) and is the largest producer of medicinal herbs in the world (Ganesan et al. 2016). As per the recent report of Food and Agriculture Organization (FAO) and United Nations Environmental Programme (UNEP), more than 28,000 plant species are being utilized worldwide for various medicinal purposes (FAO and UNEP 2020). In India, 8000 plant species have been utilized in various systems of medicine with approximately 25,000 effective herbal formulations and 1200 actively traded species (Goraya and Ved 2017; Gautam et al. 2020). This huge demand for herbal drugs invariably leads to the upsurge of herbal industries in India. Commercial demand of herbal drugs has been recently valued at INR 300 billion in 2018, which is more than double as compared to preceding years

Suma Arun Dev and Remya Unnikrishnan have contributed equally.

✉ Suma Arun Dev
sumadev@kfri.res.in

1 Forest Genetics and Biotechnology Division, Kerala Forest Research Institute, Peechi, Thrissur 680653, Kerala, India

2 Sustainable Forest Management Division, Kerala Forest Research Institute, Peechi, Thrissur 680653, Kerala, India

3 Forest Ecology and Biodiversity Conservation Division, Kerala Forest Research Institute, Peechi, Thrissur 680653, Kerala, India

4 Forestry and Human Dimensions Division, Kerala Forest Research Institute, Peechi, Thrissur 680653, Kerala, India

5 Cochin University of Science and Technology, Kochi, Kerala, India

(CISION PR News Wire 2020). Although the use of herbal medicines has become commercialised, majority of plants are still harvested from the wild. The growing demand and limited supply exert a heavy pressure on the available natural resources (Goraya and Ved 2017; Joshi et al. 2017). Consequently, overexploitation, unscientific extraction and limited supply have instigated adulteration and substitution of potential raw drugs (Newmaster et al. 2013; Srirama et al. 2017; Simmler et al. 2018). Recently, raw drug adulteration has become a pressing concern in herbal industries (Walker and AppleQuist 2012; Ouarghidi et al. 2013; Techen et al. 2014; Mishra et al. 2016). The consumer's faith on herbal medicines is in the phase of decline due to extremities in adulteration/substitution and ensuing consequences. The percentage of adulterated products varies among countries and the highest percentage of herbal product adulteration was reported from Brazil, followed by India (Ichim 2019).

To ensure the quality of herbal medicines, WHO pharmacopeia and Ayurvedic Pharmacopoeia of India (API) have implemented certain criteria for proper identification of plant species and assessment using potent chemical markers to guarantee the quality of herbal medicines (WHO 2011; Joshi et al. 2017). API is a legally binding document describing the quality, purity, and strength of selected drugs that are manufactured, distributed and sold by the licensed manufacturers in India (Joshi et al. 2017). The quality assurance is mandatory for the extensive usage and acceptability of herbal formulations in complementary and alternative medicine practicing countries (WHO 2011). Although, there are several recommended methods such as morphological, chemical/biochemical and organoleptic, it is extremely difficult to confirm authenticity of traded dried form of raw drugs (Coghlan et al. 2012). The lack of taxonomic expertise and phenotypic plasticity also complicate the identification of raw drug species from the wild. In many instances, the quality and quantity estimation of potent herbal medicines are performed using chemical finger printing methods and Thin Layer Chromatography (TLC) is the basic identification test in most of the Pharmacopeia. Recently, TLC is replaced by High Performance Thin Layer Chromatography (HPTLC) owing to its accuracy and reproducibility (Upton 2010). HPTLC finger printing has been successfully employed for the rapid identification of Apiaceae species, adulteration in *Aristolochia* herbal materials, profiling of *Tinospora* species, chemical standardisation of *Sida* species, authentication of *Taraxacum officinale* and authentication of *Echniacea* herbal products, among others (Khatoon et al. 2005; Cortes et al. 2014; Dechbumroong et al. 2018; Raclariu et al. 2018; Shawky and Abou 2018; Parveen et al. 2020). However, various external environmental factors such as age of plant, storage conditions and also type of the plant parts used can significantly influence the chemical fingerprints (Liu et al. 2011; Kaur et al. 2016). Recently, DNA barcode based plant identification has gained profound consideration as a powerful method in herbal pharmacovigilance research, mostly for quality assurance in conjunction with analytical/hyphenated methods (WHO 2011; De Boer et al. 2015; Palhares et al. 2015; Kumar et al. 2016a, 2016b; Seethapathy et al. 2018). DNA barcode analysis has revealed a wide range of discrepancies in the claimed composition of various herbal products and authenticity of raw market samples *viz.* saw palmetto dietary supplements, ginkgo products, black cohosh, herbal teas, ginseng and raw drugs like Aswagandha, *Cinnamomum verum*, *Cassia*, *Myristica fragrans*, *Phyllanthus*, *Sida* and *Santalum album* (Stoeckle et al. 2011; Baker et al. 2012; Wallace et al. 2012; Dev et al. 2013; Little and Jeanson 2013; Newmaster et al. 2013). Computer aided Artificial Intelligence (AI) based Machine Learning Algorithm (MLA) has been reported as an accurate means for species authentication (More et al. 2016; He et al. 2018). Recently, supervised Machine Learning Algorithm, WEKA and BLOG were utilised for the authentication of recently diverged, morphologically similar timber species (*Dalbergia* and *Pterocarpus*) (Hartvig et al. 2015; He et al. 2018) and raw drug authentication of traded medicinal plants in herbal industries (Hartvig et al. 2015; He et al. 2018, 2019; Unnikrishnan et al. 2021). Ability to analyse huge data is one of the major advantages of Artificial Intelligence, which offers quite a large platform for rapid authentication of raw drug samples at the industrial scale, which subsequently reduces the human labour (Hartvig et al. 2015).

In this backdrop, the present study addresses the development and validation of an integrated approach involving DNA barcoding as well as High Performance Thin Layer Chromatography (HPTLC) fingerprinting to authenticate majorly traded selected ayurvedic raw drugs [*Saraca asoca* (Roxb.) Willd., *Terminalia arjuna* (Roxb. ex DC.) Wight & Arn., *Sida alnifolia* L. and *Desmodium gangeticum* (L.) DC.], identified through market and industrial surveys in south India as part of earlier research projects. The study also compares the efficacy of Machine Learning Algorithm (WEKA and BLOG) over the genetic distance based approaches to analyse the barcode sequence data and hierarchical clustering for chemical finger prints to ascertain the extent of adulteration.

Materials and methods

Collection of authenticated biological reference material (BRM)

The highly traded raw drugs and its market adulterants were identified through a preliminary survey. Mature stem, leaf and flower samples of *Saraca asoca* and its adulterants [*Polyalthia longifolia* (Sonn.) Thwaites, *Polyalthia*

coffeoides (Thwaites) Hook.f. and Thomson, *Shorea roxburghii* G.Don], *Terminalia arjuna* and its adulterants [*Lagerstroemia speciosa* (L.) Pers., *Lagestroemia microcarpa* Hance], *Sida alnifolia* and its adulterants (*Sida acuta* Burm.f., *Sida cordifolia* L., *Sida rhombifolia* L., *Urena lobata* L.), *Desmodium gangeticum* and its adulterants [*Desmodium triangulare* (Retz.) Merr., *Desmodium triquetrum* (L.) DC., *Desmodium pulchellum* (L.) Benth., *Desmodium velutinum* (Willd.) DC.] ($n = 68$, 17 species × 4 accessions) were collected from different geographic locations of its distribution zones in south India for development of the DNA barcode database (Fig. 1). Medicinally important parts (bark of *S. asoca* and its adulterants as well as *T. arjuna* and its adulterants, whole plants of *S. alnifolia* and its adulterants as well as *D. gangeticum* and its adulterants) were collected with two individuals each ($n = 34$) for HPTLC analysis. GPS coordinates of the locations and details of the collected plant samples are provided in Supplementary table S1. Multiple accessions were collected and specimens were stored in silica gel for further use. The voucher specimens were deposited in the KFRI herbarium.

Fig. 1 Four selected taxa (*Desmodium gangeticum*, *Saraca asoca*, *Terminalia arjuna*, *Sida alnifolia*) of ayurvedic raw drugs

Collection of traded samples

Traded samples such as dried plant parts such as leaf, stem, root, and bark were collected from the selected authorised dealers of raw drugs and major herbal industries in south India ($n = 60$) (Supplementary table S2). About 100 g of each raw drug (available in extremely dried and shredded form) was purchased from authorized shops to check adulteration in the traded samples. Each collected raw drug sample was given Herbal Authentication Service Code (HAS) with details of location. To avoid the chances of mixing up, care was taken from collection to final data analysis. Most of the procured raw drugs had not retained any morphological features of the original plant species. Raw drugs of all the selected species were able to purchase from herbal markets except *T. arjuna.*

DNA extraction

Genomic DNA extraction was performed using modified Cetyl Trimethyl Ammonium Bromide (CTAB) method (Doebley et al. 1990) as well as DNeasy Plant Mini Kit for difficult samples according to manufacturer's protocol (Qiagen, USA). Total genomic DNA was also extracted from the collected raw drug samples. DNA samples were stored in the deep freezer at − 20 °C until further use. The samples were separated on 1.5 per cent agarose gel and stained in ethidium bromide and visualised under UV transilluminator to check the quality. It was quantified using a nanospectrophotometer (Nanodrop Fisher Thermo., USA).

PCR amplification

The taxonomically authenticated Biological Reference Materials (BRM) were used to develop species specific barcodes for four standard barcode gene regions *(rbc*L*, mat*K, ITS and *psb*A-*trn*H). The barcode regions, primer sequences and PCR conditions used are provided in Supplementary table S3.

Working concentration of genomic DNA was prepared by diluting the stock solution at a concentration 25 ng/µL. 25 µL of PCR reaction mixture comprised of 2.5 µL PCR buffer at 1× (supplied with 10× concentration), 1 µL each of forward and reverse primers (5 pmol), 2.5 µL of dNTPs from 10 mM stock, 2 U/25 µL of Taq-polymerase, 1 µL template DNA with the concentration of 25 ng/µL and the final volume of the PCR reaction mixture was made up to 25 µL with sterile distilled water. PCR reaction was performed with the following conditions, initial denaturation of 5 min at 94 °C, cycle denaturation of 1 min at 94 °C, cycle annealing of 1 min at 60 °C (*rbc*L and ITS), 58 °C (*psb*A-*trn*H) and 50 °C (*mat*K) and cycle extension of 1 min at 72 °C for 35 cycles and a final extension at 72 °C for 10 min. PCR products were resolved by 2 per cent agarose. Electrophoresis was performed on agarose gel by applying constant voltage to resolve the products and documented with Alpha Imager (Alpha Innotech, USA).

Sequencing of PCR products

PCR reaction was scaled up to 50 µL for the elution of PCR products by Nucleospin gel and PCR clean up kit as per the manufacturer's protocol (Machery-Nagel, U.S.A.). DNA sequencing was performed for the eluted PCR products in both forward and reverse directions employing Sanger's dideoxy chemistry.

HPTLC analysis

CAMAG Linomat 5 with twin plate chamber and CAMAG TLC scanner instrument programmed through Win CATS software were used for HPTLC finger printing. Medicinally important parts (bark of *S. asoca* and its adulterants as well as *T. arjuna* and its adulterants, whole plants of *S. alnifolia* and its adulterants as well as *D. gangeticum* and its adulterants) with two individuals each ($n = 34$) were finely powdered. Ten gram of powder accurately weighed from each sample, was extracted using solvents. Methanol extraction was carried out for *S. alnifolia*, *S. asoca*, *D. gangeticum* and its adulterants whereas chloroform extraction was performed for *T. arjuna*. Extracts were filtered and concentrated under reduced pressure and made upto 10 mL. Stationary phase of aluminium TLC plates was pre-coated with silica gel 60 F_{254} of 0.2 mm thickness and mobile phase for each set was standardised (Supplementary table 4). Samples were visualised at 366 nm. The plates were derivatized with anisaldehyde sulphuric acid reagent for band visualization of phenolic components at 366 nm.

HPTLC banding profile of BRM was documented at 366 nm under UV light. Chemical profiles of samples were analysed according to their RF values (Retention factor). Dendrogram was constructed using SPSS v.16.0 (SPSS Inc 2007) with nearest neighbour, adopting euclidean distances, which revealed the relationships between each species as per their phytochemical constituents. Owing to the variations in HPTLC fingerprints among geographical accessions of species, the technique could not be further considered for market sample validation.

Sequence data analysis

Raw chromatograms were edited and trimmed using BioEdit software (Hall 1999). The edited sequences were aligned using Clustal W (Thompson et al. 1997) and submitted to GenBank (http://www.ncbi.nlm.nih.gov/genbank/) (Supplementary table S5). Homology searches were performed

using BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to confirm the identity of the sequences.

For pair-wise genetic distance (PWG) method, genetic pair-wise distances (interspecific as well as intraspecific distances) were determined by MEGA v.6.0 using Kimura two-parameter distance model (K2P) adopting complete deletion option (Tamura et al. 2013). The interspecific divergence between species was calculated using three parameters; (i) average interspecific distance, (ii) average theta prime (θ') and (iii) minimum interspecific distances. Intraspecific parameters; (iv) average intraspecific distance, (v) theta (θ) and (vi) coalescent depth were also calculated to characterize intraspecific divergences (Meyer and Paulay 2005). DNA barcoding gaps were calculated by comparing intra and interspecific genetic distances (Meyer and Paulay 2005; Meier et al. 2006). Significance of barcoding gap was assessed using Wilcoxon matched pairs signed rank test in SPSS v.16.0 (SPSS Inc 2007). Dendrogram was constructed using developed barcode sequences for authentication of traded samples based on BRM, with 1000 bootstrap using MEGAv.0.7 adopting Kimura 2 model (Kumar et al. 2016a, 2016b).

Machine learning algorithm (MLA)-based analysis (WEKA and BLOG)

MLA's, WEKA is a GUI tool, which allows loading a data set, running an algorithm and designing an experiment. In WEKA, DNA barcoding analysis was performed with a reference data set composed of DNA sequences of known species (BRM) and query data set comprising sequences of unknown species (market samples), the function-based method Support Vector Machines (SMO) (Suykens and Vandewalle 1999), the rule-based RIPPER (Jrip) (Shahzad et al. 2013), the decision tree C4.5 (J48) (Quinlan 1996) and the Bayesian-based method Naive Bayes (Lewis 1998) were tested on DNA barcodes with ten-fold cross validation. The ".fasta" files of barcode sequences were converted to ".arff " format using "Fasta2Weka" programme for analysis in WEKA (Weitschek et al. 2014). All four classification methods in WEKA were run with four barcode primer sequences of BRM. Best classifier was selected according to their efficiency in species discrimination. Using the best classifier, sequences from traded market samples were further analysed along with BRM sequence database.

BLOG 2.0 (Barcoding with logic) is specifically developed to handle the DNA barcode data, which provides a character-based method, to identify the species using key diagnostic nucleotides. Input parameters for feature selection were given as, a maximum number of 35 features chosen ('BETA = 35'), a maximum of 200 iterations ('GRASPITER = 200'), and a maximum time of 500 min for analysis ('GRASPSECS = 30,000'). BLOG uses reference data set as training data to compute the classification rules, which was applied to test data of unknown market samples. BLOG produces a set of logic rules for the data set (BRM sequence) which can be utilized to identify the test species (market samples).

Results

DNA barcode analysis

All the analysed barcode regions (ITS, *psb*A-*trn*H, *mat*K and *rbc*L) showed 100 per cent PCR amplification in all the Biological Reference Materials (BRM) selected (Supplementary Fig. 1). Intra- and inter-specific genetic divergences of the selected BRM sequences analysed using the four barcode regions showed interspecific as well as intra-specific divergences (Table 1). Among these four barcode regions, either *psb*A-*trn*H (*S. alnifolia* and its adulterants, *T. arjuna* and its adulterants) or ITS region (*S. asoca* and its adulterants, *D. gangeticum* and its adulterants) showed highest inter and

Table 1 Evaluation of the four DNA barcode regions used in this study

Species	Average intraspecific divergence				Average interspecific divergence			
	*rbc*L	*mat*K	*psb*A-*trn*H	ITS	*rbc*L	*mat*K	*psb*A-*trn*H	ITS
Desmodium gangeticum + adulterants	0	0	0.0013 ± 0.0008	0	0.00463 ± 0.001	0.0151 ± 0.0028	0.0278 ± 0.0049	0.1722 ± 0.1061
Saraca asoca + adulterants	0	0	0	0	0.2064 ± 0.0761	0.0541 ± 0.104	0.1944 ± 0.208	0.321 ± 0.1297
Sida alnifolia + adulterants	0	0	0.0006 ± .0003	0.0008 ± 0.002	0.00083 ± 0.00054	0.0033 ± 0.00096	0.0812 ± 0.0196	0.0207 ± 0.0028
Terminalia arjuna + adulterants	0	0	0	0	0.066 ± 0.0024	0.040 ± 0.00487	2.65 ± 1.99	0.068 ± 0.0109

intra-specific divergences in the selected BRMs. Wilcoxon's signed rank test performed to test the significance of inter-specific divergences in the barcode regions showed significant values (p value ≤ 0.05).

DNA barcode analysis of the extremely shredded traded specimens showed successful PCR amplification for *rbc*L barcode region except for *D. gangeticum*. Thus, *rbc*L barcode region alone was adopted for authentication of traded samples of *Saraca asoca* and *Sida alnifolia* and *mat*K barcode region was opted for *D. gangeticum*. The barcode sequences developed from the traded samples clustered along with the respective sequences of BRM samples in the dendrogram. The phylogenetic tree generated based on these sequences showed a clear clustering of traded samples with those of BRM [*D. gangeticum* (Fig. 2), *S. asoca* (Fig. 3), *S. alnifolia* (Fig. 4)], thereby, presence of adulterants could be clearly identified. Twenty traded samples of *S. asoca* were clustered along with the BRM of *P. longifolia*. Out of 20 traded samples of *S. alnifolia*, 15 samples clustered with *S. cordifolia* and 5 with *S. acuta*. Traded samples of *D. gangeticum* showed least similarity with any of the BRM and formed a clade sister to *D. gangeticum* in the ML tree.

Machine learning algorithm (WEKA and BLOG)

Results similar to the distance base approach were generated in the MLA analysis. The identification rates of four machine learning algorithms in the selected raw drugs are provided in Table 2**.** Performance success of different machine learning algorithms also varied for BRM and market samples (Table 3). Best performance was shown by SMO and J48 with 100 per cent accuracy in authenticating the test data of all traded samples, though Jrip and Naive bayes were also successful in combinations with SMO and J48 (Fig. 5). MLA results substantiated the phylogenetic tree-based analysis.

BLOG successfully identified the test data and generated logic formula for the identification of unknown market samples (Table 2). The identification rate of train data (BRM) and test data (market samples) are provided in the table (Table 3). BLOG showed higher identification rate in *rbc*L and *mat*K barcode regions to authenticate the market samples or unknown samples. Logic rules derived from the train data (BRM samples) were utilized for the identification of test data (Market samples). BLOG generated logic formula was used to identify the presence of adulteration in market samples of selected raw drugs.

HPTLC analysis

Dendrogram generated using RF values was used to analyse the banding pattern of BRM samples (Supplementary Fig. 2). Accessions of a species belonging to different geographical locations showed intraspecific variations in their banding patterns. For example, chemical fingerprints of *D. velutinum* collected from two different geographical locations clustered separately. Interspecific similarities in HPLTC profiles were also evident as in the case of *D. velutinum* which was collected from Aralam, showed similarity to that of *D. gangeticum* and *D. velutinum* samples from Athirapilly showed similarity to *D. triquetrum.* Similarly, *P. longifolia* collected from KFRI showed similarity to *P. coffeoides* and *P. longifolia* from Palakkad showed similarity to *S. roxburghii.* Further, *S. alnifolia* showed similarity to *S. rhombifolia,* while *S. cordifolia* showed similarity to *U. lobata* (Figs. 6, 7). Therefore, HPTLC fingerprinting could not be further used as a species authentication tool because of the inconsistencies in the fingerprints generated.

Discussion

Authentic herbal drugs play a crucial role in determining the quality, safety and efficacy of herbal formulations. Therefore, to guarantee the quality of herbal medicines, WHO pharmacopeia or international pharmacopeia implemented certain criteria for appropriate identification of the plant species and quality valuation of potent phytochemical principles (Palhares et al. 2015). Yet adulteration/substitution is a prevailing issue in herbal industries. Though, herbal medicine has gained much popularity in India and all around the world, proper certification is yet to be in place. Herbal medicine once formulated is easily available to the public through various portals, where no mention of any clinical trials or authenticity is cited. This demands a proper industry standard and organized public BRM library for herbal products. The development of voucher Biological Reference Material (BRM) is a critical part of raw drug authentication process.

Even though vast array of physical, chemical (analytical), biochemical, anatomical and organoleptic methods are available, DNA barcode-based method of plant identification has comparatively been considered as the most powerful and potential standard in herbal pharmacovigilance research (Palhares et al. 2014). In India, 31 per cent of herbal drug adulteration and wide range of incongruences between claimed and identified species composition have been reported through DNA barcode-based authentication studies (Moustafa et al. 2016). Recently, an integrated approach of two or more advanced techniques was employed for species authentication. This multi-tier approach of DNA barcode along with HPTLC, NMR or HPLC was utilized for quality assurance and species authentication of popular species such as *Hamamelis virginiana*, *Matricaria recutita*, *Maytenus ilicifolia*, *Mikania glomerata*, *Panax ginseng*, *Passiflora incarnata*, *Peumus boldus* and *Valeriana officinalis*

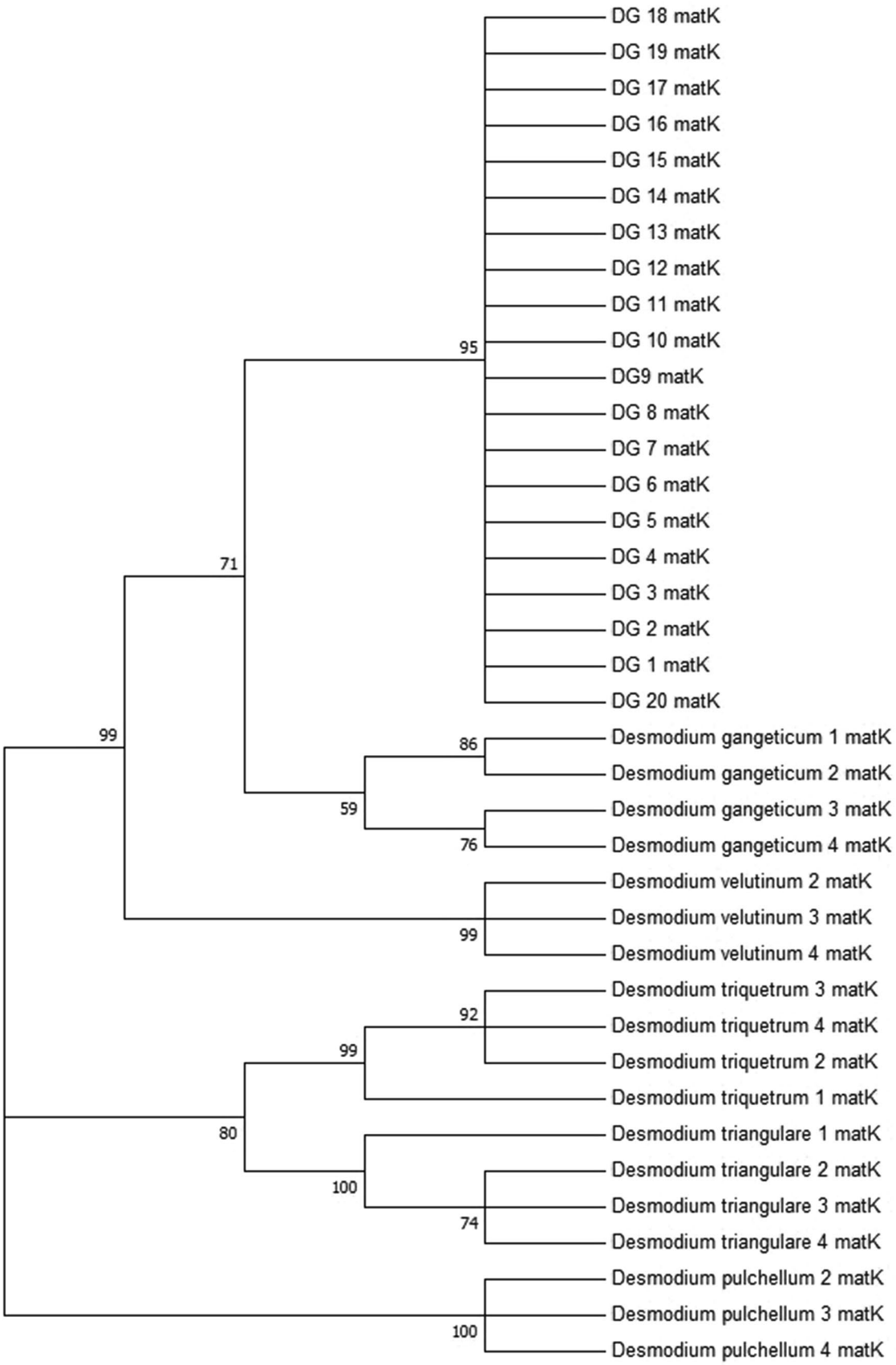

Fig. 2 Maximum Likelihood tree (ML) of market samples along with BRM (*D. gangeticum*)

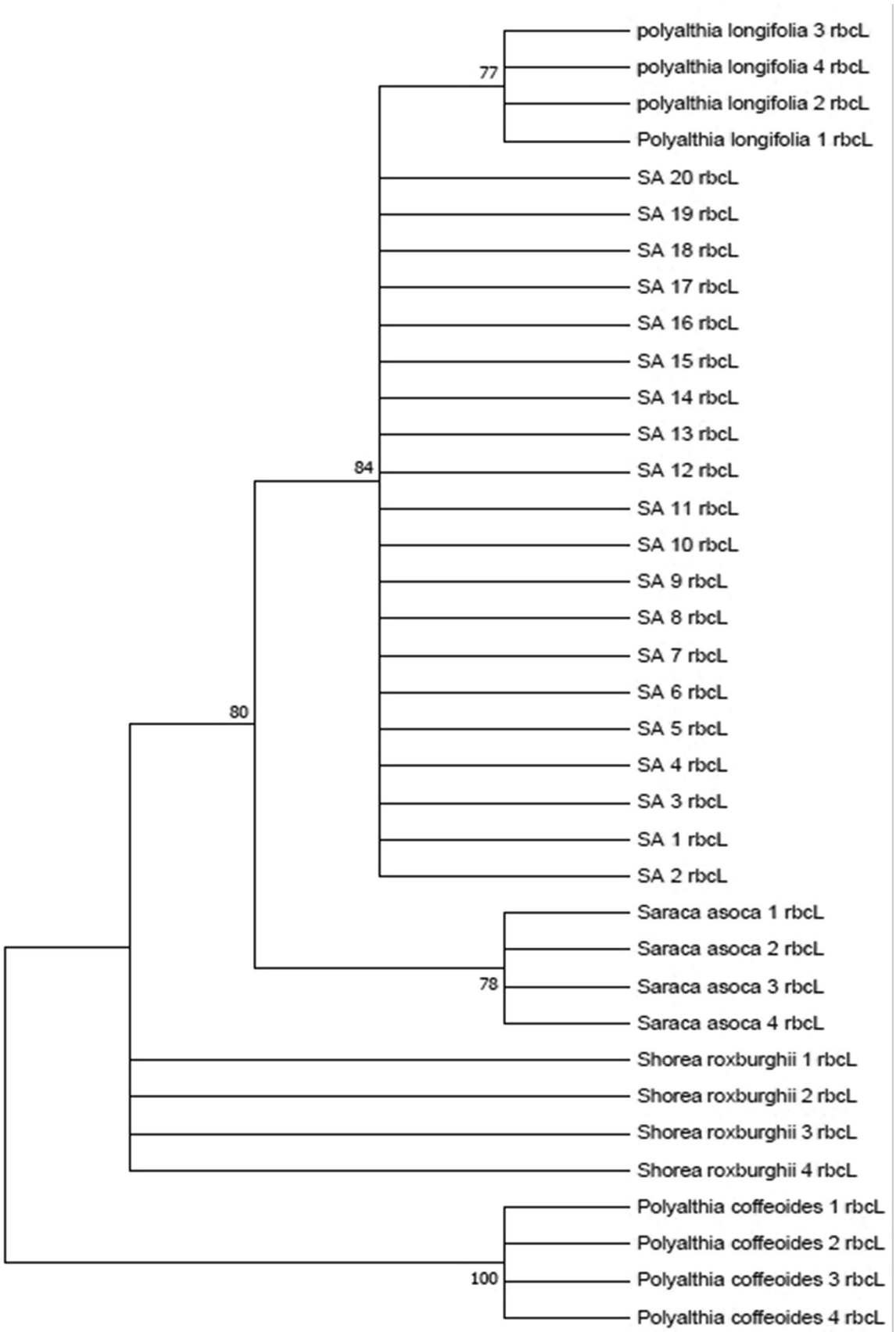

Fig. 3 Maximum Likelihood tree (ML) of market samples along with BRM (*S. asoca*)

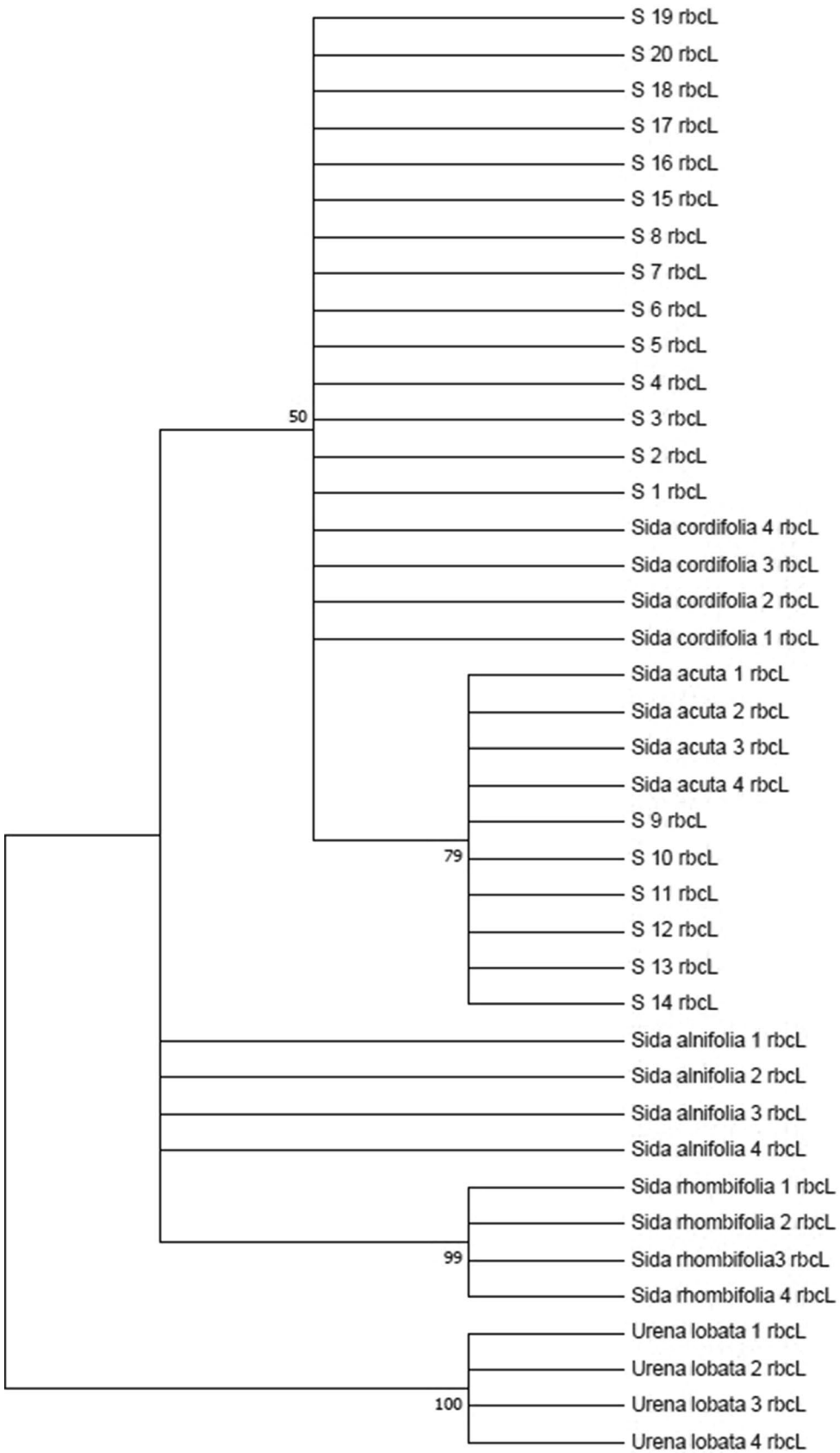

Fig. 4 Maximum Likelihood tree (ML) of market samples along with BRM (*S. alnifolia*)

Table 2 Species identification rates as percentage (correctly identified/misidentified/not identified) for the BRM using Machine Learning Algorithm (WEKA and BLOG)

Barcoding primers	WEKA				BLOG
	Naive Bayes	SMO	J48	Jrip	
Desmodium gangeticum + adulterants					
*rbc*L	30/70/0	100/0/0	70/30/0	65/35/0	100/0/0
*mat*K	100/0/0	100/0/0	100/0/0	95/5/0	100/0/0
*psb*A-*trn*H	100/0/0	100/0/0	100/0/0	100/0/0	100/0/0
ITS	95/5/0	100/0/0	95/5/0	90/10/0	55/45/0
Saraca asoca + adulterants					
*rbc*L	100/0/0	100/0/0	100/0/0	100/0/0	100/0/0
*mat*K	60/40/0	100/0/0	100/0/0	63/37/0	100/0/0
*psb*A-*trn*H	100/0/0	100/0/0	100/0/0	100/0/0	100/0/0
ITS	100/0/0	100/0/0	100/0/0	100/0/0	100/0/0
Sida alnifolia+adulterants					
*rbc*L	40/60/0	100/0/0	75/25/0	100/0/0	100/0/0
*mat*K	0/100/0	100/0/0	75/25/0	100/0/0	100/0/0
*psb*A-*trn*H	20/0/0	20/0/0	20/0/0	100/0/0	100/0/0
ITS	90/10/0	100/0/0	95/5/0	95/5/0	100/0/0

(Moustafa et al. 2016), *Garcinia* species, *Terminalia* species and *Saraca asoca* (Nithaniyal and Parani. 2016; Mishra et al. 2017; Santhosh Kumar et al. 2018; Seethapathy et al. 2018). However, studies have also reported the discrepancy of chemical markers in delineating medicinal plants, owing to variation with age of plant, dioecious nature and environmental heterogeneity (Liu et al. 2011; Cao et al. 2012; Joshi et al. 2016; Kaur et al. 2016; Upton et al. 2019). Analytical methods also failed to differentiate closely related species containing similar chemical constituents in some instances (Bajpai et al. 2017).

An integrated approach involving CBOL recommended barcode regions such as *rbc*L, *mat*K, *psb*A-*trn*H, ITS as well as HPTLC profiling was investigated in the present study to create a reference database for majorly traded raw drugs and its market adulterants in India. However, accessions of species collected from different geographic locations of unknown age, showed variations in the HPTLC fingerprints which restricted their further use in raw drug authentication. Similarly, accessions of species such as *Artemisia japonica*, *Cinnamomum glaucescens* and *Cymbopogon distans* collected from different geographical locations also showed variation in their chemical pattern (Joshi et al. 2017). Moreover, sex specific discrepancies in the phytochemical constituents were also reported in *Tinosporia cordifolia* and *Dioscorea* (Chase et al. 2007; Paul et al. 2020).

Even though, DNA barcode regions like ITS showed promising results in the case of *D. gangeticum*, *T. arjuna*, *S. asoca* and *psb*A-*trn*H in *S. alnifolia*, high number of indels along with huge interspecific variation failed to provide consistent bidirectional unambiguous sequencing reads which limited their utility for authentication (CBOL 2009; Hollingsworth et al. 2009; Roy et al. 2010; Kool et al. 2012). Further, degraded DNA obtained from the market samples with impurities like fungal contamination and presence of extraneous matters hindered the primer annealing and subsequent PCR amplification of ITS/*psb*A-*trn*H barcode gene regions. Similar hurdles were also reported in the traded samples of medicinal plants in Morocco, India and Brazil (Srirama et al. 2010; Santhosh Kumar et al. 2018; Parveen et al. 2020). Consequently, *rbc*L and *mat*K barcode sequence database which was discriminant enough to identify adulterants in all the cases were used to validate the traded market raw drugs.

Ability to analyse large data is one of the major advantages of Artificial Intelligence. It also provides quite a large platform for rapid authentication of raw drug samples, which subsequently reduces human labour (He et al. 2018). MLA's created specifically for DNA barcoding can be effectively used for species authentication in herbal industries, timber industries and for biodiversity conservation (He et al. 2018; Heinrich et al. 2018). Recent studies authenticated the efficiency of Machine Learning Approaches (WEKA and BLOG) for species delineation of even recently diverged species, and morphologically similar timber species such as *Dalbergia* and *Pterocarpus* as compared to conventional barcode sequence analysis methods [Maximum Likelihood (ML), BLAST and nearest-neighbour] (He et al. 2018, 2019). The present study demonstrated the efficiency of MLA (WEKA and BLOG) for DNA barcode sequence data analysis to confirm the presence of adulterants in commercially traded samples of *S. asoca*, *S. alnifolia* and *D. gangeticum*. Traded samples of *S. asoca* showed similarity with *Polyalthia longifolia*, *S. alnifolia* with *S. cordifolia*/*S. acuta* and *D. gangeticum* traded samples showed no similarity

Table 3 Species identification rate of market samples along with BRM

BRM + Traded samples	Barcode region	WEKA				BLOG
		Naive Bayes	SMO	J48	Jrip	
Desmodium gangeticum+traded sample+adulterants	*mat*K	34/66/0	100/0/0	95/5/0	–	100/0/0
Saraca asoca+adulterant+traded samples	*rbc*L	100/0/0	100/0/0	100/0/0	100/0/0	100/0/0
Sida alnifolia+adulterants+traded samples	*rbc*L	–	100/0/0	100/0/0	–	100/0/0

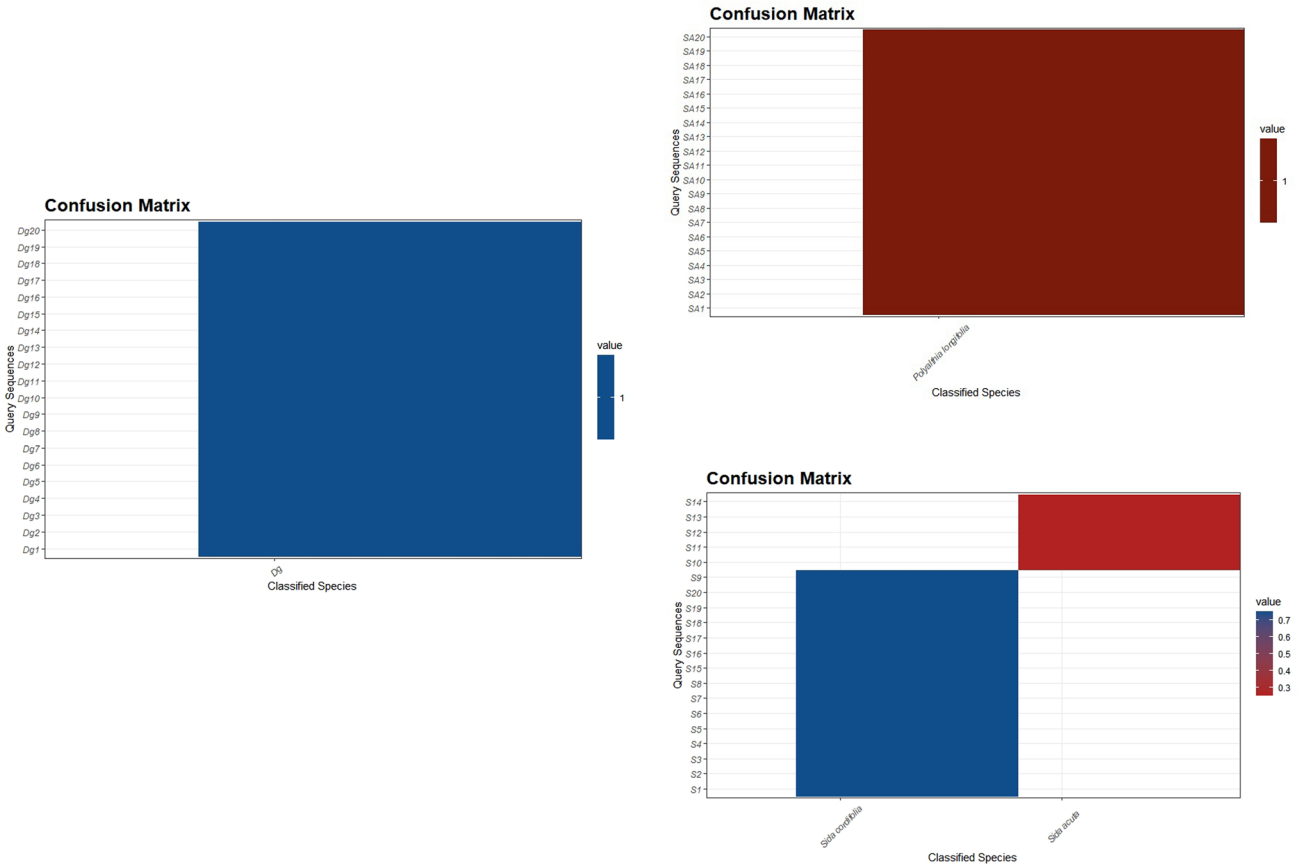

Fig. 5 The confusion matrix (output of WEKA) showing identification rate of market samples along with BRM (**a**. *Desmodium gangeticum*, **b**. *Saraca asoca*, **c**. *Sida alnifolia*)

with the developed reference database. MLA barcode data analysis thus can be recommended as a powerful, accurate and rapid means over genetic distance-based methods to discriminate original raw drugs from its adulterants based on the case studies.

DNA barcode authentication analysis revealed the presence of adulteration in the traded market samples of the studied species. Market samples of *S. asoca* were adulterated with *P. longifolia*, *S. alnifolia* with *S. cordifolia/S. acuta* and *D. gangeticum* with other similar species of *Desmodium*. Similar vernacular names, presence of potent chemicals, morphological similarities as well as overlapping species distributions are considered as primary reasons for adulteration (Srirama et al. 2010). In the studied species, *Sida alnifolia*, *S. cordifolia* and *S. acuta* are locally known as "bala" and contain the potent chemicals like ephedrine, vasicinone and rutin (Dhalwal and Shinde 2010; Chaves et al 2013). Similarly, *S. asoca* and *P. longifolia* known as 'asoka', contain caffeic acid and ellagic acid in common (Sharma et al. 2018). Similar issues of adulteration were reported wherein *Myristica fragrans* was adulterated with *M. malabarica*, *Cinnamomum verum* with *C. cassia*/*C. malabatrum* (Swetha et al. 2014, 2017). Adverse consequences of herbal drug adulteration were reported from the countries like Australia, Japan, China and regions like Taiwan, where chronic use of adulterated *Artistolochia fangchi*, led to the death of patients owing to renal failure (Michl et al. 2013; Jadot et al. 2017). Similarly, US Food and Drug Administration banned *Piper methysticum* containing products in Germany, Switzerland, France, Canada and UK due to health issues related to hepatitis, cirrhosis and liver failure (U.S. Food and Drug Administration 2002).

In India, there are no proper guidelines to coordinate and maintain the information related to collection, supply, trade and consumption of botanicals (Goraya and Ved 2017; Gauthum et al. 2020). Herbal sector in India is a large network involving various government agencies, stake holders, herbal healers and traders. Though all of them are linked in the trade network, authenticity of the raw drugs extracted from wild through formal and informal sectors is mostly ambiguous. Recently, to bring in more transparency, National Medicinal Plant Board (NMPB), Govt. of India, has launched an online virtual platform 'www.e-charak.in' and a mobile application 'e-charak'

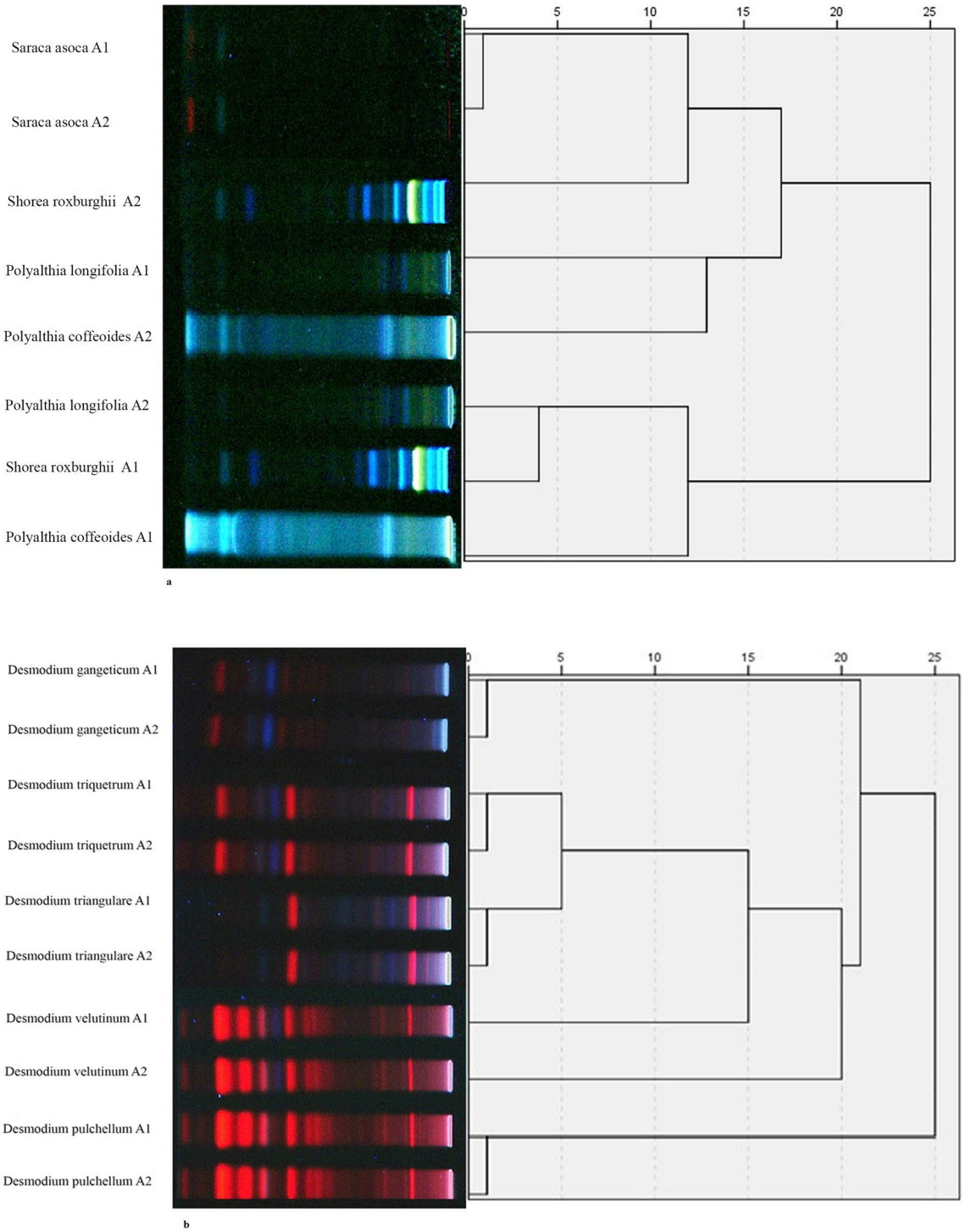

Fig. 6 Dendrogram constructed using HPTLC banding pattern (RF value) (**a**. *Saraca asoca* and its adulterants, **b**. *Desmodium gangeticum* and its adulterants)

with a view to provide online market portal and to ensure the transparent trade of medicinal plants among producers and purchasers in the country (https://www.nmpb.nic.in/content/e-charak-online-market-place-medicinal-plants).

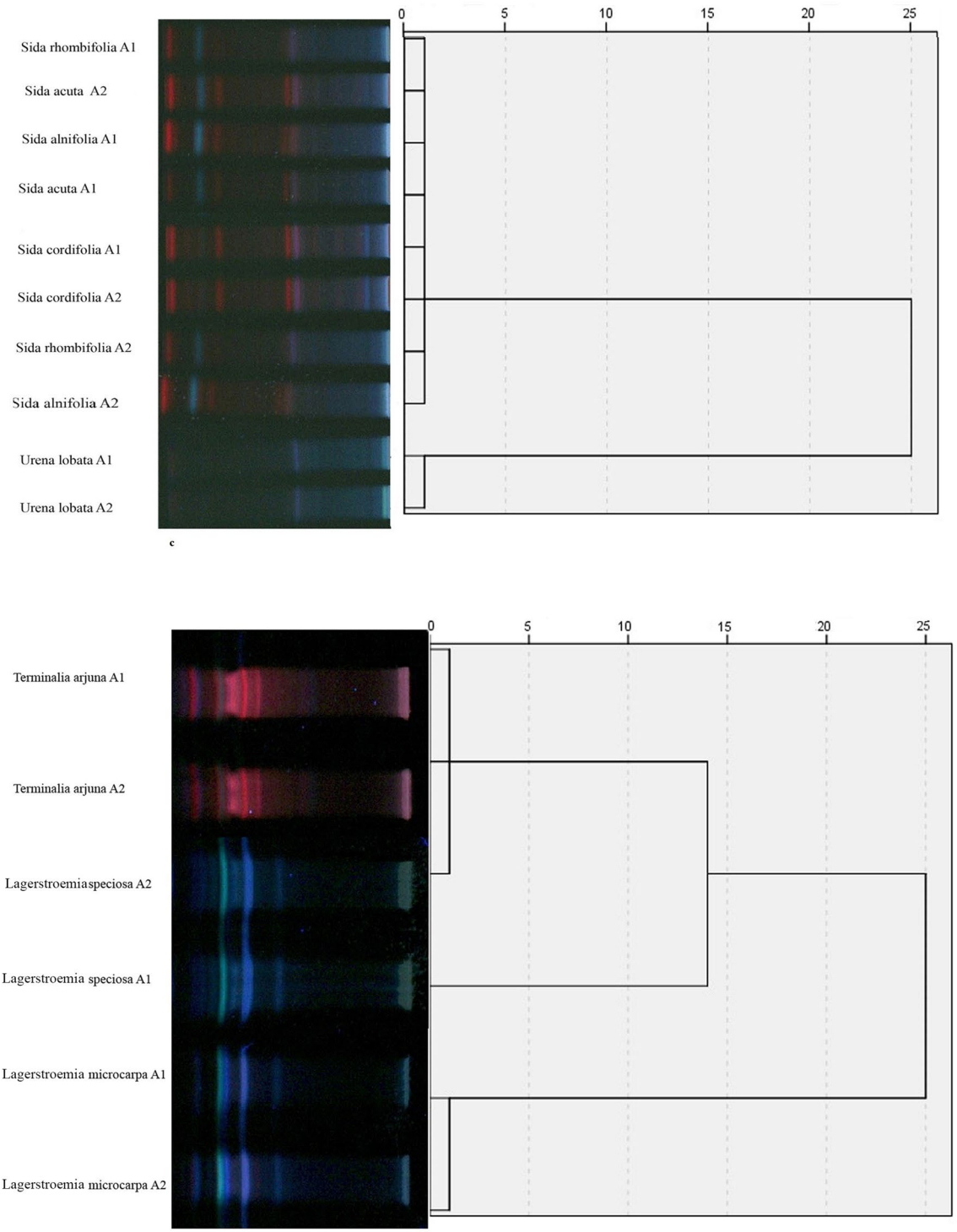

Fig. 7 Dendrogram constructed using HPTLC banding pattern (RF value) (**c**. *Sida alnifolia* and its adulterants, **d**. *Terminalia arjuna* and its adulterants)

Conclusions

Our study could demonstrate the effectiveness of DNA barcode analysis through Artificial Intelligence (AI) platform to authenticate ayurvedic raw drug species in addition to the recommended standard organoleptic and analytical methods. DNA barcode reference library once created and deposited in the public database can be further accessed

for species identification of unknown samples, whenever required for certification purposes. British pharmacopeia is the first agency to globally implement DNA barcoding tool for authentication, given its ability to identify the source of herbal products accurately (Heinrich et al. 2018; Sgamma et al. 2018). Though herbal medicine has gained much popularity in India, proper certification procedures and agencies are yet to be established. It is therefore imperative to bring forth a statutory body to monitor proper collection, processing, certification and sale of raw drugs in the country. Substitution of authentic species with species of similar therapeutic effects based on the ancient ayurvedic scripts can alleviate the extraction pressure and conservation measures can augment the depletion of wild resources.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13205-021-03001-5.

Acknowledgements We are thankful to the State Forest Department of Kerala, for their permission to collect wild samples. The financial support received from National Medicinal Plant Board (NMPB), Ministry of AYUSH, Government of India (No. F.No. Z.18017/187/CSS/R&D/KE-01/2016-17-NMPB-IV A) is gratefully acknowledged.

Author contributions SAD, JR: Fund acquisition, conceptualization, methodology and designing the experiments, writing, editing and finalization of manuscript. RU: Conducted wet lab experiments as part of doctoral research, data analysis and manuscript writing. SP, AV: Sample collection and manuscript writing.

Declarations

Conflict of interests The authors declare no conflict of interests.

References

Bajpai V, Kumar S, Singh A, Singh J, Negi MPS, Bag SK, Kumar B (2017) Chemometric based identification and validation of specific chemical markers for geographical, seasonal and gender variations in *Tinospora cordifolia* stem using HPLC-ESI-QTOF-MS. Phytochem Anal 28:277–288. https://doi.org/10.1155/2012/18010310.1002/pca.2673

Baker DA, Stevenson DW, Little DP (2012) DNA barcode identification of black cohosh herbal dietary supplements. J AOAC Int 95:1023–1034. https://doi.org/10.1155/2012/18010310.5740/jaoacint.11-261

Cao Y, Fang S, Yin FuZ, Shang X, Yang W, Yang H (2012) Chemical fingerprint and multicomponent quantitative analysis for the quality evaluation of *Cyclocarya paliurus* leaves by HPLC–Q–TOF–MS. Molecules 22:1927

CBOL (2009) Plant working group a DNA barcode for land plants. Proc Natl Acad Sci USA 1061:2794–12797

Chase MW, Cowan RS, Hollingsworth PM, van den Berg C, Madrinan S, Petersen G, Seberg O, Jorgsensen T, Cameron KM, Carine M, Pedersen N, Hedderson TAJ, Conrad F, Salazar GA, Richardson JE, Hollingsworth ML, Barraclough TG, Kelly L, Wilkinson M (2007) A proposal for a standardised protocol to barcode all land plants. Taxon 56:295–299

Chaves OS, Gomes RA, Tomaz ACA, Fernandes MG, Mendes Junior LDG, Agra MF, Braga VA, Souza MFV (2013) Secondary metabolites from *Sida rhombifolia* L. (Malvaceae) and the vasorelaxant activity of cryptolepinone. Molecules 18:2769–2777

CISION PR News Wire (2020) India Ayurveda industry outlook 2019–2024—market size and growth forecasts, product launches, competitive landscape Report p. 161, Retrieved from https://www.prnewswire.com/newsreleases/india-ayurveda-industry-outlook-2019-2024—market-size–growthforecasts-product-launches-competitive-landscape-301007651.html. Accessed 24 Feb 2021

Coghlan ML, Haile J, Houston J, Murray DC, White NE, Moolhuijzen P (2012) Deep sequencing of plant and animal DNA contained within traditional Chinese medicines reveals legality issues and health safety concerns. PLoS Genet. https://doi.org/10.1371/journal.pgen.1002657

Cortes N, Mora C, Munoz K, Díaz J, Serna R, Castro D, Osorio E (2014) Microscopical descriptions and chemical analysis by HPTLC of *Taraxacum officinale* in comparison to *Hypochaeris radicata* a solution for mis-identification. Rev Bras Farmacogn 24:381–388. https://doi.org/10.1016/j.bjp.2014.07.018

De Boer HJ, Ichim MC, Newmaster SG (2015) DNA barcoding and pharmacovigilance of herbal medicines. Drug Saf 38:611–620. https://doi.org/10.1007/s40264-015-0306-8

Dechbumroong P, Aumnouypol S, Denduangboripant J, Sukrong S (2018) DNA barcoding of *Aristolochia* plants and development of species-specific multiplex PCR to aid HPTLC in ascertainment of *Aristolochia* herbal materials. PLoS ONE 13:e0202625. https://doi.org/10.1371/journal.pone.0202625

Dev SA, Muralidharan EM, Sujanapal P, Balasundaran M (2013) Identification of market adulterants in East Indian sandalwood using DNA barcoding. Ann Sci 71:517–522. https://doi.org/10.1007/s13595-013-0354-0

Dhalwal K, Shinde V (2010) Optimization and validation of reverse phase HPLC and HPTLC method for simultaneous quantification of vasicine and vasicinone in *Sida* species. J Med Plant Res 4:1289–1296

Doebley J, Durbin M, Golenberg EM, Clegg MDP (1990) Evolutionary analysis of the large subunit of carboxylase *rbc*L nucleotide sequence among the grasses *Gramineae*. Evolution 44:1097–1108. https://doi.org/10.2307/2409569

FAO and UNEP (2020) The state of the world's forests 2020. Forests, biodiversity and people. Rome. https://doi.org/10.4060/ca8642en

Ganesan A, Marichamy K, Thagamariappan J (2016) Development of medicinal plants sector in India—an empirical study. IJETMAS 4:82–89

Gautam RS, Mishra M, Prasad R (2020) Development and trade of medicinal and aromatic plants (MAPs): learnings from comparative analysis of MAPs export of India and China. Working Paper, Rajiv Gandhi Institute for Contemporary Studies (RGICS), New Delhi, p 59

Goraya GS, Ved DK (2017) Medicinal plants in India an assessment of their demand and supply. Ministry of AYUSH, Dehradun

Hall TA (1999) BioEdit a user-friendly biological sequence alignment editor and analysis Program for Windows 95/98/NT. Nucleic Acids Sympo Ser 41:95–98

Hartvig I, Czako M, Kjær ED, Nielsen LR, Theilade I (2015) The use of DNA barcoding in identification and conservation of rosewood (*Dalbergia* spp.). PLoS ONE. https://doi.org/10.1371/journal.pone.0138231

He T, Jiao L, Yu GJ, Jiang X, Yin Y (2018) DNA barcoding authentication for the wood of eight endangered *Dalbergia* timber species using machine learning approaches. Holzforschung. https://doi.org/10.1515/hf-2018-0076

He T, Jiao L, Wiedenhoeft AC, Yin Y (2019) Machine learning approaches outperform distance- and tree-based methods for DNA

barcoding of *Pterocarpus* wood. Planta. https://doi.org/10.1007/s00425-019-03116-3
Heinrich M, Lardos A, Leonti M, Weckerle C, Willcox M, Applequist W, Stafford G (2018) Best practice in research consensus statement on ethnopharmacological field studies–Con SEFS. J Ethnopharmacol 211:329–339
Hollingsworth ML, Andra AC, Forrest LL, Richardson J, Pennington RT, Long DG, Hollingsworth PM (2009) Selecting barcoding loci for plants evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. Mol Ecol Resour 9:439–457. https://doi.org/10.1111/j.1755-0998.2008.02439.x
Ichim MC (2019) The DNA-based authentication of commercial herbal products reveals their globally widespread adulteration. Front Pharmacol. https://doi.org/10.3389/fphar.2019.0122
SPSS Inc (2007) Released. SPSS for Windows, Version 160. SPSS Inc, Chicago
Jadot I, Decleves A, Nortier J, Caron N (2017) An integrated view of aristolochic acid nephropathy update of the literature. Int J Mol Sci. https://doi.org/10.3390/ijms18020297
Joshi R, Satyal P, Setzer W (2016) Himalayan aromatic medicinal plants a review of their ethnopharmacology, volatile phytochemistry and biological activities. Medicines. https://doi.org/10.3390/medicines3010006
Joshi VK, Joshi A, Dhiman K (2017) The ayurvedic pharmacopoeia of India, development and perspectives. J Ethnopharmacol. https://doi.org/10.1016/j.jep.2016.07.030
Kaur T, Bhat R, Vyas D (2016) Effect of contrasting climates on antioxidant and bioactive constituents in five medicinal herbs in Western Himalayas. J Mt Sci 13:484–492. https://doi.org/10.1007/s11629-014-3380-y
Khatoon S, Srivastava M, Rawat A, Mehrotra S (2005) HPTLC method for chemical standardization of *Sida* species and estimation of the alkaloid ephedrine. J Planar Chroma 18:364–367. https://doi.org/10.1556/JPC.18.2005.5.5
Kool A, de Boer HJ, Krüger A, Rydberg A, Abbad A, Björk L, Martin G (2012) Molecular identification of commercialized medicinal plants in southern Morocco. PLoS ONE. https://doi.org/10.1371/journal.pone.0039459
Kumar S, Stecher G, Tamura K (2016a) MEGA7 molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874. https://doi.org/10.1093/molbev/msw054
Kumar SJ, Gogna N, Newmaster SG, Venkatarangaiah K, Subramanyam R, Saroja SG, Ramanan US (2016b) DNA barcoding and NMR spectroscopy-based assessment of species adulteration in the raw herbal trade of *Saraca asoca* Roxb. Willd, an important medicinal plant. Int J Legal Med 130:1457–1470. https://doi.org/10.1007/s00414-016-1436-y
Lewis DD (1998) Naïve Bayes at forty the independence assumption in information retrieval. Mach Learn 98:4–15
Little DP, Jeanson ML (2013) DNA barcode authentication of saw palmetto herbal dietary supplements. Sci Rep. https://doi.org/10.1038/srep03518
Liu Y, Reich PB, Li G, Sun S (2011) Shifting phenology and abundance under experimental warming alters trophic relationships and plant reproductive capacity. Ecology 92:1201–1207. https://doi.org/10.1890/10-2060.1
Meier R, Shiyang K, Vaidya G (2006) DNA barcoding and taxonomy in *Diptera* a tale of high intraspecific variability and low identification success. Syst Biol 55:715–728 (**PMID 17060194**)
Meyer CP, Paulay G (2005) DNA barcoding error rates based on comprehensive sampling. PLoS ONE. https://doi.org/10.1080/10635150600969864
Michl J, Jennings HM, Kite GC, Ingrouille MJ, Simmonds MSJ, Heinrich M (2013) Is aristolochic acid nephropathy a widespread problem in developing countries? A case study of *Aristolochia indica* L. in Bangladesh using an ethnobotanical-phytochemical approach. J Ethnopharmacol 149:235–244. https://doi.org/10.1016/j.jep.2013.06.028
Mishra P, Kumar A, Nagireddy A, Mani DN, Shukla AK, Tiwari R, Sundaresan V (2016) DNA barcoding an efficient tool to overcome authentication challenges in the herbal market. J Plant Biotechnol. https://doi.org/10.1111/pbi.12419
Mishra P, Kumar A, Nagireddy A, Shukla AK, Sundaresan V (2017) Evaluation of single and multilocus DNA barcodes towards species delineation in complex tree genus *Terminalia*. PLoS ONE. https://doi.org/10.1371/journal.pone.0182836
More RP, Mane RC, Purohit HC (2016) MatK-QR classifier: a patterns based approach for plant species identification. BioData Min 9:39
Moustafa MF, Hesham AEL, Quraishi MS, Alrumman SA (2016) Variations in genetic and chemical constituents of *Ziziphus spina-christi* L. populations grown at various altitudinal zonation up to 2227m height. J Genet Eng Biotechnol 14:349–362
Newmaster SG, Grguric M, Shanmughanandhan M, Ramalingam S, Ragupathy S (2013) DNA barcoding detects contamination and substitution in North American herbal products. BMC Med 11:222–235. https://doi.org/10.1186/1741-7015-11-222
Nithaniyal S, Parani M (2016) Evaluation of chloroplast and nuclear DNA barcodes for species identification in *Terminalia* L. Biochem Syst Ecol 68:223–229. https://doi.org/10.1016/j.bse.2016.08.001
Ouarghidi A, Powell B, Martin GJ, HugoAbbad A (2013) Species substitution in medicinal roots and possible implications for toxicity in Morocco. Econ Bot 66:370–382. https://doi.org/10.1007/s12231-012-9215-2
Palhares RM, Drummond MG, Brasil BS, Krettli AU, Oliveira GC, Brandao MG (2014) The use of an integrated molecular, chemical, and biological based approach for promoting the better use and conservation of medicinal species a case study of Brazilian quinas. J Ethnopharmacol 155:815–822. https://doi.org/10.1016/j.jep.2014.06.040
Palhares RM, Gonçalves DM, Alves BDS, Cosenza GP, Lins MDG, Brandao OG (2015) Medicinal plants recommended by the World Health Organization DNA barcode identification associated with chemical analyses guarantees their quality. PLoS ONE 10:e0127866. https://doi.org/10.1371/journal.pone.0127866
Parveen A, Adams JS, Raman V, Budel JM, Zhao BGNM, Ali Z, Khan IA (2020) Comparative morpho-anatomical and HPTLC profiling of *Tinospora* species and dietary supplements. Planta Med. https://doi.org/10.1055/a-1120-3711
Paul C, Debnath A, Chakraborty K (2020) Sex-specific variations in phytochemicals and antimicrobial potentiality of *Dioscorea*. Future J Pharm Sci. https://doi.org/10.1186/s43094-020-00076-4
Quinlan JR (1996) Improved use of continuous attributes in C4.5. J Artif Intell Res 4:77–90
Raclariu AC, Tebrencu CE, Ichim MC, Ciuperca OT, Brysting AK, de Boer H (2018) What's in the box? Authentication of *Echinacea* herbal products using DNA metabarcoding and HPTLC. Phytomedicine. https://doi.org/10.1016/j.phymed.2018.03.058
Roy S, Tyagi A, Shukla V, Kumar A, Singh UM, Chaudhary LB, Tuli R (2010) Universal plant DNA barcode loci may not work in complex groups a case study with Indian *Berberis* Species. PLoS ONE. https://doi.org/10.1371/journal.pone.0013674
Santhosh Kumar JU, Krishna V, Seethapathy GS, Ganesan R, Ravikanth G, Shanker RU (2018) Assessment of adulteration in raw herbal trade of important medicinal plants of India using DNA barcoding. 3 Biotech. https://doi.org/10.1007/s13205-018-1169-3
Seethapathy GS, Tadesse M, Urumarudappa SKJV, Gunaga S, Vasudeva R, Malterud KE, Wangensteen H (2018) Authentication of *Garcinia* fruits and food supplements using DNA barcoding and NMR spectroscopy. Sci Rep. https://doi.org/10.1038/s41598-018-28635

Sgamma T, Masiero E, Mali P, Mahat M, Slater A (2018) Sequence-specific detection of *Aristolochia* DNA – a simple test for contamination of herbal products. Front Plant Sci 91:828. https://doi.org/10.3389/fpls.2018.01828

Shahzad W, Asad S, Khan MA (2013) Feature subset selection using association rule mining and JRip classifier. Int J Phys Sci 88:85–96

Sharma P, Sharma S, Yadav S, Srivastava A, Varma A, Luhana K, Shrivastava N (2018) Categorization of *Saraca asoca* and *Polyalthia longifolia*, using chemical and genetic fingerprinting associated with multivariate statistical analysis. J Appl Res Med Aromat Plants 9:91–96. https://doi.org/10.1016/j.jarmap.2018.03.004

Shawky E, Abou EKM (2018) Rapid discrimination of different Apiaceae species based on HPTLC fingerprints and targeted flavonoids determination using multivariate image analysis. Phytochem Anal. https://doi.org/10.1002/pca.2749

Simmler C, Graham JG, Chen SN, Pauli GF (2018) Integrated analytical assets aid botanical authenticity and adulteration management. Fitoterapia 129:401–414. https://doi.org/10.1016/j.fitote.2017.11.017

Srirama R, Senthilkumar U, Sreejayan N, Ravikanth G, Gurumurthy BR, Shivanna MB, Uma Shaanker R (2010) Assessing species admixtures in raw drug trade of *Phyllanthus*, a hepato-protective plant using molecular tools. J Ethnopharmacol 130:208–215. https://doi.org/10.1016/j.jep.2010.04

Srirama R, Santhosh Kumar JU, Seethapathy GS, Newmaster SG, Ragupathy S, Ganeshaiah KN, Ravikanth G (2017) Species adulteration in the herbal trade: causes, consequences and mitigation. Drug Saf 40(8):651–661. https://doi.org/10.1007/s40264-017-0527-0

Stoeckle MY, Gamble CC, Kirpekar R, Young G, Ahmed S, Little DP (2011) Commercial teas highlight plant DNA barcode identification successes and obstacles. Sci Rep. https://doi.org/10.1038/srep00042

Suykens JAK, Vandewalle J (1999) Least squares support vector machine classifiers. Neural Process Lett 9:293–300

Swetha VP, Parvathy VA, Sheeja TE, Sasikumar B (2014) DNA barcoding for discriminating the economically important *Cinnamomum verum* from its adulterants. Food Biotechnol 28:183–194. https://doi.org/10.1080/08905436.2014.931239

Swetha VP, Parvathy VA, Sheeja TE, Sasikumar B (2017) Authentication of *Myristica fragrans* Houtt. using DNA barcoding. Food Control 73:1010–1015. https://doi.org/10.1016/j.foodcont.2016.10.004

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6 molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729

Techen N, Parveen I, Pan Z, Khan IA (2014) DNA barcoding of medicinal plant material for identification. Curr Opin Biotechnol 25:103–110. https://doi.org/10.1016/j.copbio.2013.09.010

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL X windows interface flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Unnikrishnan R, Sumod M, Jayaraj R, Sujanapal P, Dev SA (2021) The efficacy of machine learning algorithm for raw drug authentication in *Coscinium fenestratum* (Gaertn.) Colebr. employing a DNA barcode database. Physiol Mol Biol Plants 27(3):605–617. https://doi.org/10.1007/s12298-021-00965-9

Upton RT (2010) Use of High-performance thin layer chromatography by the American herbal pharmacopoeia. J AOAC Int 93:1349–1354. https://doi.org/10.1093/jaoac/93.5.1349

Upton R, David B, Gafner S, Glasl S (2019) Botanical ingredient identification and quality assessment strengths and limitations of analytical techniques. Phytochem Res. https://doi.org/10.1007/s11101-019-09625-z

Walker KM, AppleQuist WL (2012) Adulteration of selected unprocessed botanicals in the US retail herbal trade. Econ Bot 66:321–327. https://doi.org/10.1007/s12231-012-9211-6

Wallace LJ, Boilard SMAL, Eagle SHC, Spall JL, Shokralla S, Hajibabaei M (2012) DNA barcodes for everyday life routine authentication of natural health products. Food Res Int 49:446–452. https://doi.org/10.1016/j.foodres.2012.07.048

Weitschek E, Fiscon G, Felici G (2014) Supervised DNA barcodes species classification analysis, comparisons and results. Biodata Minining. https://doi.org/10.1186/1756-0381-7-4

WHO (2011) The World medicines situation—traditional medicines global situation, issues and challenges. WHO Press, Geneva