Pak. J. Bot., 48(5): 2167-2174, 2016.

DNA BARCODING: A TOOL FOR STANDARDIZATION OF HERBAL MEDICINAL PRODUCTS (HMPS) OF LAMIACEAE FROM PAKISTAN

NADIA BATOOL ZAHRA[1], ZABTA KHAN SHINWARI[1, 2] AND MUHAMMAD QAISER[3]

[1]*Department of Biotechnology, Quaid-i-Azam University, Islamabad, Pakistan*
[2]*Pakistan Academy of Sciences, Islamabad, Pakistan*
[3]*University of Karachi, Karachi, Pakistan*
[*]*Corresponding author's email: shinwari2008@gmail.com*

Abstract

There has been a considerable interest worldwide in traditional and alternative medicine, particularly herbal products over the past few decades but the adulteration or contamination of herbal medicinal products (HMPs) is a potential threat to consumer safety. The fact highlights the importance of an effective and accurate science integrated method for taxonomic identification of the medicinal plants and their HMPs. DNA barcoding is a molecular technique which has made it possible to identify the herbs and to find the adulterants in HMPs. The current study was designed on DNA barcoding of medicinal plants of family Lamiaceae for their correct identification and to fix the problem of adulteration for protecting consumers from health risks associated with product substitution and contamination. Many Lamiaceae species are used as traditional medicines, as culinary herbs, spices and as source of essential oils. HMPs representing 32 Lamiaceae plant samples were purchased/collected from three herbal stores (Pansar stores) in Islamabad and a herbal pharmaceutical industry. We selected three plastid loci *rbcL*, *matK* and *psbA-trnH* to barcode these HMPs. MEGABLAST sequence comparison was performed to verify the taxonomic identity of the samples. We found four mislabeled samples and two product substitutions. The overall amplification success for *rbcL* and *matK* was 87% and 81% while *psbA-trnH* showed 69%. *matK* and *psbA-trnH* were able to distinguish the species relatively better with 40% success rate than *rbcL* (16%). On the whole we generated a total of 22 genus-level barcodes (78%) and 12 species-level barcodes (44%). The species-level identification was considerably low due to insufficient reference data and selection of plastid markers. Therefore, it is recommended to develop herbal barcode library for adequate availability of reference sequence data and addition of nuclear markers. DNA barcoding can help the regulatory authorities to devise a mechanism for quality control and can largely support the herbal pharmaceutical industries to restore the eroded consumer confidence.

Key words: DNA barcoding, Lamiaceae, Herbal medicinal products, *rbcL, matK* and *psbA-trnH*.

Introduction

The popularity of herbal medicine/products has increased worldwide during the past couple of decades. In many countries herbal medicine is not only regarded as a conventional treatment strategy but also acts as a health care system. This knowledge of traditional medicine based on utility of plants/herbs is playing a pivotal role in today's drug development and biological research. The active compounds of medicinal herbs are used as a lead molecule for the discovery of new drug. Globally speaking, demand for medicinal plants, herbal medicinal products, food supplements, pharmaceuticals and health products are significantly growing (Sen & Chakraborty, 2015). By 2050, the estimated rise in global market for medicinal and aromatic plants is US $5 trillion which was US $62 billion in 2002, indicating a global paradigm shift from an allopathic to a traditional healthcare system (Shinwari, 2010). Despite the increasing demand and acceptance of medicinal plants, there are pitfalls in drug standardization, efficacy, safety, quality control, information & regulatory system. The very first step in the process of quality assurance is the authentic identification of the plant species. The identification of herbs is traditionally carried out by morphological characters, but in case of cryptic species or phenotypically variable species the chances of misidentification are greater (Vijayan & Tsou, 2010). There can be serious consequences of use of a misidentified medicinal plant. For example, *Digitalis purpurea* (foxglove) could be mistaken for *Symphytum officinale* (comfrey) because of morphological similarity in their leaves. *Symphytum officinale* is used to treat fractures, tendon injury while *Digitalis purpurea* has been used to treat congestive heart failure due to the presence of cardiac glycosides. Ingesting the *Symphytum officinale* tea resulted the cardiotoxicity in nine individuals, some others even required temporary pacemakers to regulate their heart beat (Lin *et al.,* 2010). There are studies which proved the deliberate adulteration of herbal medicinal products (HMPs) with additives not found on the label (Stoeckle *et al*., 2011; Newmaster *et al*., 2013; Anon., 2015). These reports highlight the importance of an effective and accurate science integrated method for taxonomic identification of the medicinal plants and their HMPs. DNA barcoding is a molecular technique which has made it possible to identify the herbs and to find the adulterants in HMPs. A DNA barcode is a standardized, short (< 1000 bp) and highly variable segment of DNA which is compared to a reference database through a sequence alignment algorithm for species identification (Shinwari *et al*., 2014). The standard genomic regions that have been used for DNA barcoding are chloroplast ribulose 1, 5-bisphosphate carboxylase/oxygenase large subunit (*rbcL*) and maturase K (*matK*) as core barcodes (Hollingsworth *et al.,* 2009). The spacer between photosystem II protein D1 and tRNA-His (*psbA-trnH* spacer) and the nuclear ribosomal internal transcribed spacer 2 (ITS2), however, are now suggested by several studies (Chen *et al.,* 2010; Gao *et al.,* 2010; Yao *et al*., 2010; Fu *et al.,* 2011, Han *et al*., 2012; Newmaster *et al*., 2013; Michel *et al*., 2016).

In Pakistan about 70% of the population is dependent on the plant derived traditional medicine for primary health care system. In developed countries like Germany the 80% of population have used herbal medicine at least once. Therefore, to provide standardization and quality control of herbal medicines with established therapeutic value, resulting

in an increase in demand, to revive a traditional healthcare system facing extinction is an important task that can be fulfilled by DNA barcoding of medicinal plants. In Pakistan, medicinal plants are used widely often in the form of packaged medicine manufactured by herbal medicine industries and as raw herbs and their decoctions which are formulated by indigenous people by practicing their indigenous knowledge. The raw material is collected from the wild and transported to national and international markets. The long transportation chain, with many middlemen, results in increased events of misidentification and adulteration. Due to lack of proper monitoring and regulatory system there is a risk that many medicinal plants and their products sold in the market are taxonomically misidentified, mislabeled, or contaminated. The aforesaid situation emphasizes need for an effective and efficient identification system through barcoding these medicinal plants. Hence, where the barcoding will allow the pharmaceutical industry and consumers in Pakistan to authenticate the raw material, it will also provide reference sequences to the scientific community. The current study was designed on DNA barcoding of medicinal plants of family Lamiaceae for their correct identification and to fix the problem of adulteration. Many species are used as traditional medicines, as culinary herbs and spices and as source of essential oils. However, it is often difficult to differentiate closely related groups due to morphological complexity of the Lamiaceae, which leads to many taxonomic problems. Many members of Lamiaceae are available in the domestic herbal markets of Pakistan as raw herbs and packaged herbal preparations manufactured by herbal pharmaceutical industries. Barcoding of suspect raw ingredients can prove or disprove the identity of medicinal plants before they are processed. It enables the pharmaceutical industry to build its consumer confidence (Schori & Showalter, 2011). Here, we investigated the taxonomic accuracy of raw herbs (Lamiaceae) sold by three local stores within Islamabad area and the medicinal species collected from a major herbal pharmaceutical industry of Pakistan. These local markets are termed as 'Pansar stores' which are usually owned by qualified individuals of herbal medicine known as 'Hakeems' or 'Tabibs'. These small businesses sell the dried medicinal herbs and the 'Hakeems' prepare the medicinal formulations as well. We selected the three standard barcoding markers *rbcL*, *matK* and *psbA-trnH* for this study and also evaluated the ease of barcoding using these markers with different plant parts, including amplification success and ability to resolve species in comparative sequence analyses. Many previous studies employed the *rbcL* for taxonomic purposes (Shinwari *et al.,* 1994a; Shinwari *et al.,* 1994b; Shinwari, 1995; Shinwari 1998; Shinwari, 2002; Shinwari & Shinwari, 2010).

Materials and Methods

Herbal medicinal products (HMPs) representing 32 plant samples belonging to family Lamiaceae were purchased/collected from three herbal stores (Pansar stores) in Islamabad and a herbal pharmaceutical industry. These sampled HMP's came in plastic packets, pre-weighed or weighed according to customer needs. HMPs sampled included fragmented plant parts, twenty-three of which were leaf samples, two seeds, two inflorescence, four samples were a mix of different dried plant parts (shoots, flowers, leaves) and one sample was grounded plant material beyond recognition.

Standard cetyltrimethylammonium bromide (CTAB) protocol (Doyle, 1991) was used for DNA isolation. Polymerase chain reaction (PCR) amplification of the *rbcL*, *matK* and *psbA-trnH* was performed in a Veriti Thermal Cycler (Applied Biosystems, Carlsbad, California, USA) using the KAPA3G Plant PCR Kit (Kapa Biosystems, Woburn, Massachusetts, USA) (Schori *et al.,* 2013; Shinwari *et al*., 2014). Each reaction contained the KAPA3G Plant PCR Buffer (1× final concentration, includes dNTPs at 0.2 mM each), $MgCl_2$ (1.5 mM final concentration), 1 unit KAPA3G Plant DNA polymerase, primers at a final concentration of 0.3 µM each, template DNA and PCR-grade water to bring the volume to 50 µL. A higher concentration (2 mM) of $MgCl_2$ was required for successful PCR of nine of these species. For most of the species successful amplification of our selected markers was achieved with the KAPA3G enzyme from dirty pellets (not purified after CTAB extraction), while other species (*Ocimum*, *Lycopus*, *Nepeta*, *Origanum*) failed to amplify until genomic DNA had been purified or extracted with the PowerPlant Pro DNA Isolation Kit. The following cycling parameters were used for *rbcL* : 95°C 10 min; 50 cycles: 95°C 20 s, 58°C 15 s, 72°C 90 s; 72°C 90 s. The *rbcL* primers 1F (Fay *et al.,* 1997) and 1460R (Fay *et al.,* 1998; Cuénoud *et al.,* 2002) were used in this experiment. The same cycling parameters were performed using the *matK* 390F/1360R primers (Cuénoud *et al.,* 2002) with an annealing temperature of 50°C for 40 cycles (Schori *et al.,* 2013; Shinwari *et al*., 2014). A touchdown program was carried out for *psbA-trnH* using the PsbAF/PsbHR primers (Sang *et al.*, 1997; Tate & Simpson, 2003), where the annealing temperature was 58°C for initial 11 cycles followed by touchdown to 48°C for 29 cycles (Schori *et al.,* 2013; Shinwari *et al*., 2014). To ensure the successful amplification of the desired sequence, the PCR products were run on 1% agarose gel. PCR products were cleaned with the Wizard SV Gel and PCR Clean-Up System (Promega Corporation, Madison, Wisconsin, USA). The purified PCR products were sequenced at Ohio University's Genomics Facility and analyzed using an ABI 3130xl Genetic Analyzer (Applied Biosystems, Carlsbad, California, USA). Each sequencing reaction included 2 µL 5× buffer (Applied Biosystems), 0.5 µL dimethyl sulfoxide (DMSO ; Sigma), 0.5 µL BigDye (Applied Biosystems), 0.1 µL ThermoFidelase (Fidelity Systems, Gaithersburg, Maryland, USA), 10–40 ng template DNA, and PCR-grade water for a total volume of 8 µL. Cycle sequencing products were cleaned with the BigDye XTerminator Purification Kit (Applied Biosystems). In case of *rbcL*, external primers 1F and 1460R and internal primers 636F and 724R (Fay *et al.,* 1997) were used for sequencing. For *matK* and *psbA-trnH* sequencing, the similar primer pairs were used as for amplification.

Assembly of forward and reverse sequences and manual trimming were performed in Geneious R6 (Biomatters Ltd., Auckland, New Zealand). Sequences were then compared to GenBank nucleotide database using NCBI's MEGABLAST default parameters. Percent similarity was recorded for the closest matches. The highest BLAST % identity of the query sequence should be from the expected species or the species belonging to the expected genera for correct identification; ambiguous identification means that the highest BLAST % identity for a query sequence was found to match several genera of the expected family; incorrect identification means that the highest BLAST %

identity of the query sequence was not from the expected species/expected genera/expected family.

Results

In our investigation, *rbcL* proved to be the most successfully amplified and sequenced in 28 of the 32 HMPs (87%). *matK* provided 26 barcode sequences while *psbA-trnH* yielded 22 sequences. The barcode recovered for each HMP and plant part from which it was recovered can be found in Table 1. Recovered *rbcL* barcodes ranged from 576-1445 bp. Barcoding success for *matK* and *psbA-trnH* were 26/32 (81%) and 22/32 (69%), respectively (Fig. 1). *matK* sequences ranged from 506–586 bp, while *psbA-trnH* sequences ranged from 713–826 bp. All sequences (KP172036-KP172082, KP218929- KP218945) generated from this study were deposited in GenBank (excluding misidentified and substituted/contaminated samples).

Twenty-two (22) of 32 sequenced samples matched their expected genera based on MEGABLAST of the selected barcode regions. Four samples *Leucas linifolia, Lycopus europaeus, Salvia haematodes* and *Salvia moorcroftiana* II which were bought from the local stores came up as completely different species not belonging to Lamiaceae (Table 1). Sequence chromatograms of these samples showed clear peaks with no indication of admixture. With respect to their obtained barcodes, all, except *Lycopus europaeus*, showed 99% sequence identity to a different unrelated species in MEGABLAST. *Lycopus europaeus* showed *Proboscidea* and *Martynia* with 95% sequence identity as top-hits.

Excluding the four misidentified samples and three other samples for which *rbcL* could not be amplified/sequenced, 21 of 25 *rbcL* sequences (84%) matched the expected genera based on MEGABLAST. Of these 21 only four *rbcL* sequences matched the expected species (*Melissa officinalis, Nepeta cataria, Rosmarinus officinalis, Stachys byzantina*). Nine *rbcL* sequences (*Ajuga bracteosa*, *Ajuga parviflora*, *Plectranthus rugosus*, *Leucas cephalotes*, *Nepeta ruderalis*, *Otostegia limbata*, *Salvia aegyptiaca*, *Salvia lanata*, *Salvia moorcroftiana* I) could not be identified to species level due to limited reference sequence data available in GenBank. *rbcL* lacked sequence variability for resulting in multiple congeneric species hits having the similar % identity. For example, top hits for the *rbcL* sequence for *Mentha spicata* sample did not include the expected species, but included *Mentha suaveolens*. *matK* for *Mentha spicata* came up with multiple congeneric hits including *Mentha spicata*, *Mentha suaveolens* and *Mentha x piperita* with 100% identity. On the other hand *psbA-trnH* resulted in 99% identity with *Mentha spicata* but 97% identity with other congeneric species. The same was true for *Mentha pulegium* and *Mentha aquatica rbcL*, which showed *Mentha suaveolens* and *Mentha x piperita* as top hits respectively while their *matK* and *psbA-trnH* provided us with the expected closest matches. Based on our *rbcL* barcode, *Mentha suaveolens, Mentha x piperita, Mentha longifolia, Mentha arvensis, Ocimum basilicum* I, *Ocimum basilicum* II, *Ocimum x africanum, Ocimum tenuiflorum* and *Salvia plebeia* did not match their expected species. We could only prove their identification to the expected genus level.

A total of 25 *matK* barcodes excluding the four misidentified and three non-amplified/non-sequenced samples were obtained. Of these 25 *matK* sequences, 19 matched the expected genera based on MEGABLAST. 10 of these 19 *matK* sequences matched their expected species level (*Melissa officinalis, Mentha aquatica*, *Mentha pulegium*, *Mentha spicata*, *Nepeta cataria, Ocimum basilicum* I, *Ocimum x africanum, Rosmarinus officinalis, Salvia plebeia*, *Stachys byzantina*). No reference data could be retrieved for *Ajuga bracteosa, Hyssopus officinalis*, *Lallemantia royleana, Leucas cephalotes, Nepeta ruderalis. Otostegia limbata, Salvia aegyptiaca* and *Salvia moorcroftiana* I from GenBank, therefore, species level identification was not possible based on *matK* barcodes.

psbA-trnH spacer region had a lower rate of amplification/sequencing success as compared to *rbcL* and *matK*, however, it proved to be more successful for distinguishing the samples at species level. Only 20 *psbA-trnH* barcodes were found (excluding four mislabeled and eight unamplified), resulting in 14 sequences which provided identification to expected genera and eight sequences (*Mentha pulegium, Mentha spicata, Mentha longifolia, Mentha aquatica, Ocimum basilicum* I, *Ocimum basilicum* II, *Rosmarinus officinalis* and *Salvia plebeia*) came up as successful barcodes for expected species. *Ajuga bracteosa* and *Leucas cephalotes* could be identified only to the expected genera due to lack of reference *psbA-trnH* sequences for these species in GenBank.

There are three HMPs from our collection which appeared to be highly ambiguous. There was not sufficient reference data available in GenBank for these species with respect to our chosen DNA barcoding markers. Although the barcodes obtained for these species (*Lallemantia royleana*, *Origanum vulgare and Salvia aegyptiaca*) were not completely unrelated because they belonged to Lamiaceae and none of the sequences for either of the barcoding region showed up to any other plant family. Genus *Lallemantia* had no records for *rbcL*, *matK* and *psbA-trnH*. *Otostegia* had no reference sequence for *psbA-trnH*, four un-verified *matK* sequences and three *rbcL* sequences which did not matched to our query blast. *matK* and *psbA-trnH* sequences of *Origanum vulgare* did not correspond to reference sequences in top-hits, rather in case of *matK* it appeared to be 99-100% identical to several *Mentha* species but also 99% identity to *Origanum vulgare* in lower-hits. *psbA-trnH* behaved in the same fashion where top-hit was *Thymbra* while lower hits contained *Origanum vulgare*, both with 97% identity. The only reference data available for *Salvia aegyptiaca* was a *psbA-trnH* sequence which did not show up in our blast hits. In case of *Hyssopus* only two *rbcL* reference sequences were available for *Hyssopus* which did not came up as top-hits for our query sequence. No reference sequences were present in GenBank for *matK* and *psbA-trnH* for genus *Hyssopus*. Similarly, *Nepeta ruderalis* had no sequence data in GenBank. However, we have marked *Hyssopus officinalis* and *Nepeta ruderalis* as product substitutes based on the sequences we obtained. Our *rbcL*, *matK* and *psbA-trnH* sequences for *Hyssopus officinalis* and *Nepeta ruderalis* showed them as *Nepeta bracteata* and *Salvia* respectively. As all the three markers showed same identification, it indicated the product substitution.

The overall amplification success for *rbcL* and *matK* was 87% and 81% while *psbA-trnH* showed 69% (Fig. 1). On the contrary, *matK* and *psbA-trnH* were able to distinguish the species relatively better with 40% success rate than *rbcL* (16%) (Fig. 1). On the whole we generated a total of 22 genus-level barcodes (78%) and 12 species-level barcodes (44%) (Fig. 2).

Table 1. List of HMPs used in this study including scientific names based on the common names under which the products were sold, plant part used, and taxonomic ID based on MEGABLAST top hits (best max score). Empty cells suggest that barcode was not recovered due to DNA extraction and/or amplification problems. "spp." following genus name indicate multiple species possible within that genus. † represents misidentified HMPs. * indicates HMPs potentially substituted/contaminated.

Common name	Expected scientific name	Plant part	Identification based on barcodes		
			rbcL ID	*matK* ID	*psbA-trnH* ID
Kauri booti	*Ajuga bracteosa*	Leaf	*Ajuga reptans*	*Ajuga orientalis*	*Ajuga ciliata*
Tarkha booti	*Ajuga parviflora*	Leaf	*Ajuga decumbens*	-	-
Zufa Khushk	*Hyssopus officinalis*	Inflorescence	**Nepeta bracteata*	**Nepeta bracteata*	**Nepeta bracteata*
Tukhm malanga	*Lallemantia royleana*	Leaf	-	*Glechoma hederaceae*	-
Chatra, Gul dode	*Leucas cephalotes*	Leaf	*Marrubium peregrinum*	*Leucas* sp.	*Leucas* sp.
Gomi	*Leucas linifolia*	Shoot, leaf, flower	†*Portulaca oleraceae*	-	†*Portulaca oleraceae*
Gypsywort	*Lycopus europaeus*	Shoot, leaf, flower	-	-	†*Proboscidea triloba*
Lemon balm	*Melissa officinalis*	Leaf	*Melissa officinalis*	*Melissa officinalis*	-
Watermint	*Mentha aquatica*	Leaf	*Mentha x piperita*	*Mentha aquatica*	*Mentha aquatic*
Wild mint	*Mentha arvensis*	Leaf	*Mentha suaveolens*	*Mentha spicata*	*Mentha canadensis*
Peppermint	*Mentha x piperita*	Leaf	*Mentha suaveolens*	*Mentha aquatica*	*Mentha canadensis*
Pennyroyal	*Mentha pulegium*	Leaf	*Mentha suaveolens*	*Mentha pulegium*	*Mentha pulegium*
Velanay, Horse mint	*Mentha longifolia*	Leaf	*Mentha suaveolens*	*Mentha aquatica*	*Mentha longifolia*
Spearmint	*Mentha spicata*	Leaf	*Mentha suaveolens*	*Mentha spicata*	*Mentha spicata*
Apple mint	*Mentha suaveolens*	Leaf	*Mentha spicata*	*Mentha spicata*	*Mentha Canadensis*
Catnip	*Nepeta cataria*	Leaf	*Nepeta cataria*	*Nepeta cataria*	-
Badranj Boya, Indian catnip	*Nepeta ruderalis*	Inflorescence	**Salvia* spp.	**Salvia plebeia*	**Salvia* spp.
Nyazbo, Basil	*Ocimum basilicum* I	Leaf	*Ocimum tenuiflorum*	*Ocimum basilicum*	*Ocimum basilicum*
Nyazbo, Basil	*Ocimum basilicum* II	Seed	*Ocimum tenuiflorum*	*Ocimum americanum*	*Ocimum basilicum*
Lemon basil	*Ocimum x africanum*	Leaf	*Ocimum tenuiflorum*	*Ocimum x africanum*	*Ocimum basilicum*
Tulsi	*Ocimum tenuiflorum*	Leaf	*Ocimum basilicum*	*Ocimum basilicum*	-
Chiti Booti	*Otostegia limbata*	Leaf	*Stachys, Phlomis, Lamium, Ballota*	*Leucas cephalotes*	*Lamium galeobdolon* subsp. *flavidum*
Banjawain, Wild marjoram	*Origanum vulgare*	Shoot, leaf	-	*Mentha* spp.	*Thymbra, Mentha*
Khwangere	*Plectranthus rugosus*	Leaf	*Plectranthus barbatus*	-	-
Rosemary	*Rosmarinus officinalis*	Leaf	*Rosmarinus officinalis*	*Rosmarinus officinalis*	*Rosmarinus officinalis*
Tak malanga	*Salvia aegyptiaca*	Seed	*Dracocephalum grandiflorum*	*Lallementia royleana*	*Rosmarinus officinalis*
Behman Surkh, Red sage	*Salvia haematodes*	Roots	†*Biebersteinia multifida*	-	-
Kianr	*Salvia lanata*	Leaf	*Savia moorcroftiana*	-	-
Kalli-jarri	*Salvia moorcroftiana* I	Leaf	*Salvia nemorosa*	*Salvia nemorosa*	-
Kalli-jarri	*Salvia moorcroftiana* II	Powder	-	†*Baccharoides adoensis*	-
Samundar sokh	*Salvia plebeia*	Leaf	*Salvia miltiorrhiza*	*Salvia plebeia*	*Salvia plebeia*
Lamb's ears	*Stachys byzantina*	Leaf	*Stachys byzantina*	*Stachys byzantina*	*Prasium majus*

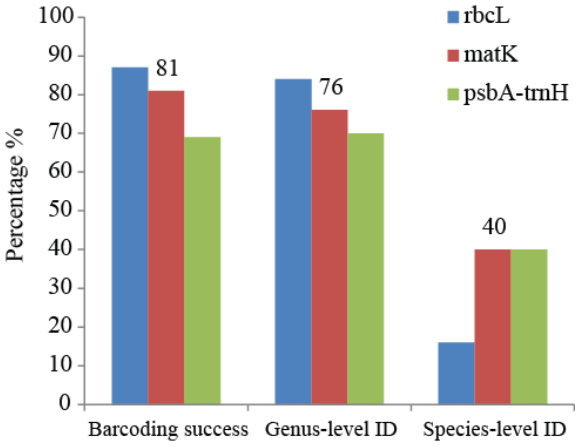

Fig. 1. *rbcL* demonstrated the highest barcoding and genus-level identification success, but with the lowest species-level identification. *matK* and *psbA-trnH* showed the same percentage of species-level identification. Our results proposed *matK* as the potential barcode for Lamiaceae HMPs.

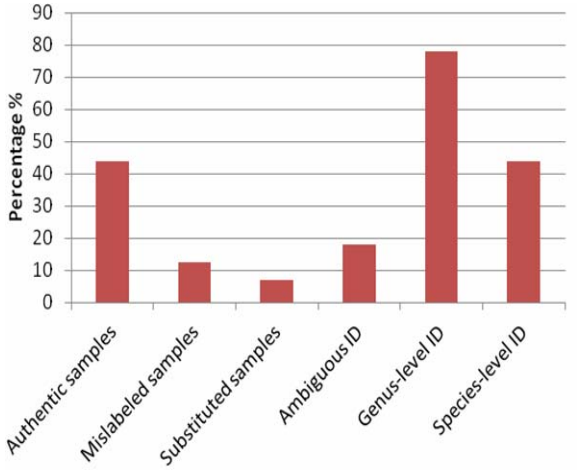

Fig. 2. DNA barcodes of 32 HMPs belonging to Lamiaceae, collected from local herbal stores and herbal pharmaceutical industry.

Discussion

1. Difficulty in DNA extraction and amplification: Wallace *et al*. (2012) reported a relatively low DNA barcode success rate of 48% in their study. There may be different reasons of this low success rate including different manufacturing protocols of HMPs and the type of plant material (for example, seeds, leaf, stem, roots, barks) used in the herbal formulations. The natural presence of several secondary metabolites, including polysaccharides, glycoproteins and phenolics in plants can result in interference with DNA isolation, amplification and sequencing (Schori *et al.,* 2013). Therefore, inefficient laboratory protocols used for DNA isolation, amplification and sequencing of haplotypes from herbal products may impede the success. In addition, degradation at primer binding sites may also contribute to differential amplification success of selected genes in samples with potentially degraded DNA (Newmaster *et al.,* 2013). We faced some failures in DNA extraction by using the standard CTAB method for *Ocimum*, *Lycopus*, *Nepeta* and *Origanum* which we overcame by replacing the CTAB protocol with the PowerPlant Pro DNA Isolation Kit. It suggests that DNA extraction from HMPs is not unachievable from common forms of herbal materials (leaf, seed, root, flowers, and powder). There are previous studies (Cimino, 2010; Stoeckle *et al*., 2011; Baker *et al*., 2012; Wallace *et al*., 2012; Newmaster *et al.,* 2013; Michel *et al*., 2016) in support of DNA based methods for quality control of herbal products wherein DNA was successfully extracted. Majority of our samples amplified at 1.5 mM $MgCl_2$, a higher concentration (2 mM) of $MgCl_2$ was required for successful PCR of nine samples only. We tried plant enhancer which comes with the KAPA3G Plant PCR Kit according to manufacturer's protocol, yet failed to get the desired results for few samples. Hence, it suggests that there are instances when PCR from herbal pant material become challenging in the presence of inhibitors and absence of primers which do not perfectly match the target sequences.

2. Insufficient reference sequence data: Several Lamiaceae species used as medicinal plants do not have reference sequence data available in GenBank and it became a major problem for our species-level identification of *Ajuga bracteosa, Ajuga parviflora, Lallemantia royleana, Plectranthus rugosus, Leucas cephalotes, Nepeta ruderalis, Otostegia limbata, Salvia lanata and Salvia moorcroftiana* I. There was not a single reference sequence available for these species for either of our selected markers. Hence, MEGABLAST hits showed congeneric species but not the said species, resulting in the genus-level identification only and decreasing the success rate of species-level discrimination. Our approach was in accordance with Hollingsworth *et al.* (2009) who stated that species with a single sample were considered potentially distinguishable if the sequence was unique (i.e. there was a potential that successful species-level discrimination may be achieved, but further sampling was needed to verify this). For *Hyssopus officinalis* only two *rbcL* reference sequences were available while *Salvia aegyptiaca* had only one *psbA-trnH* sequence in GenBank. *Lallemantia* was another exception for which no reference data was found for any member of this genus. Our voucher specimen are the first ever sequences for these aforementioned species deposited in GenBank and publicly available. With such scant data the possibility of misidentification by the suppliers, or the chances of mislabeled sequences in GenBank always remain. It is not always possible to identify an unknown or suspect plant where there is no reference sequence generated. The family or genus level identification may take place but it is unlikely that species identification will be confirmed. Newmaster *et al*. (2013) established a reference database of barcodes termed as Sequence Reference Material (SRM) Library to which they compared sequences to prevent these problems. However GenBank is adequately informative when we are working with limited resources, or performing preliminary research. Another potential problem which we

encountered was the deposited reference data in GenBank do not always include the complete gene region. A partial sequence from a less variable portion of a gene may lead to a high match percentage that does not reflect an accurate identification of the query sequence. *rbcL* and *matK* regions are relatively long (approximately 1430 and 1550 bp respectively) but the rate of evolution might not be same for complete gene regions. The *psbA-trnH* spacer may be much shorter (200-650 bp, Kress *et al.,* 2005) than *rbcL* and *matK*, so a 99% match using it as a barcode may be more accurate than a 99% match using either of these gene regions (Schori and Showalter, 2011).

3. Pros and cons of *rbcL, matK* and *psbA-trnH* regions: CBOL (The Consortium for the Barcode of Life) recommended the two-locus *rbcL–matK* combination as the universal plant DNA barcode in 2009. The *psbA-trnH* spacer and the nuclear internal transcribed spacer 2 (ITS2), are also widely used (Chen *et al.,* 2010; Gao *et al.,* 2010; Yao *et al*., 2010; Fu *et al.,* 2011, Han *et al*., 2012; Newmaster *et al*., 2013; Michel *et al.,* 2016). Our findings showed that *rbcL* and *matK* reasonably amplified among Lamiaceae with 89% success for both but *psbA-trnH* demonstrated a relatively lower success (71%). However, we were able to get the good quality bidirectional reads for *rbcL* and *matK* but *rbcL* was not variable enough to discriminate species. *rbcL* was able to distinguish only four samples at species-level while *matK* provided species identification of ten samples. *psbA-trnH* performed better than *rbcL* by pinpointing the taxonomic identity of eight species, although its amplification/ sequencing success remained lower. The concatenated phylogenetic analyses of *rbcL* and *matK* was unable to resolve species because of insufficient reference sequence data in GenBank which resulted in extensive missing data of either *rbcL* or *matK* sequences in GenBank for many accessions per species (results not shown here). Both *matK* and *psbA-trnH* carried out 40% success rate for species-level identification wherein *matK* had better amplification/sequencing success. Therefore, based on this investigation, *matK* proved to be the potential barcode for Lamiaceae HMPs used in Pakistan. The overall success rate (excluding the four misidentified samples) of barcoding the Lamiaceae HMPs in this study is 82% at genus-level, however, it reduces to 44% at species-level identification attributed to the inadequate availability of reference sequence data and related issues discussed earlier.

Challenges, improvements and regional recommendations for DNA barcoding of HMPs: To date, as far as we know this is the first study published from Pakistan on the quality assurance of HMPs generally and particularly on Lamiaceae where we targeted a big herbal pharmaceutical industry and local herbal stores of the country's capital city Islamabad. Contamination and adulteration in herbal products offers considerable health risks for consumers. Previously, contamination and substitution in several products with plants that have known toxicity, side effects and/or negatively interact with other herbs, supplements, or medications has been reported by other studies (Lin *et al*., 2010; Stoeckle *et al*., 2011; Newmaster *et al*., 2013). Our results found two product substitutions but it was difficult to affirm whether it was an accidental misidentification of a bulk product or a fraudulent market substitution for a cheaper product. Unlabeled plant fillers may also be found in herbal products, which may pose a potential health risk for consumers. With the recent advances in DNA barcoding methods, it is potentially possible to meet the challenge of authentication of herbal products or their raw material for routine market analysis (Sucher & Carles, 2008).

The first major challenge which created the limitations in our study was the lack of herbal barcode library for reference sequence data. The similar problem has been reported in market studies from North America, where scientists tested the authenticity of herbal products without herbal SRM barcode library (Stoeckle *et al.,* 2011; Baker *et al*., 2012; Wallace *et al*., 2012). Therefore, it is important to develop the regional herbal barcode library where we can find the reference sequence data of all the medicinal plants and their haplotypes being used in the region. Once such a regional herbal barcode library will be created, it would be possible to overcome the insufficiency of authentic reference sequence data for quality control by regulatory authorities as well as the herbal pharmaceutical industries. A similar approach was taken by Newmaster *et al*. (2013) where they developed a SRM herbal barcode library for their study. Moreover, it is important to Pakistan that a broader regional herbal barcode library is established which not only includes the reference sequence data from Pakistani voucher specimen's but also from its neighboring countries i.e. China, Iran, India and Afghanistan. It is because we found in our study that the raw material for HMPs do not necessarily comes from within the country but has been imported from the border countries. The 'GenBank' and 'Medicinal Materials DNA Barcode Database' (Lou *et al*., 2010) provides barcodes for many species of medicinal plants but without the information about voucher specimens which is an essential component of any DNA Barcode (Hebert *et al*., 2004). Therefore, it provides another reason for construction of herbal barcode library.

The second challenge occurred due to the selection of plastid barcode regions only, which provided low species resolution for herbal products. Chen *et al*. (2010) selected ITS2 for use in identifying medicinal plants provided the reasons (i) it comes from the nuclear genome, which has a different rate of evolution than the plastid genome (ii) it provides high species resolution (iii) it is a much shorter sequence allowing higher recovery from processed plant materials found within herbal products. Han *et al*. (2012) suggested the use of ITS2 for identification of medicinal plants of Lamiaceae. Hence, ITS2 may present a more successful barcode for Pakistani Lamiaceae HMPs and should be tested in future investigations.

There has been a considerable interest worldwide in traditional and alternative medicine, particularly herbal products over the past few decades. The World Health Organization also emphasizes the crucial role of alternative and traditional medicines in preventive and curative health, especially in developing countries and encourages member states to support traditional medicines and plan for formulation of policies with appropriate regulations (Anon., 2002).The herbal medicine as an alternative system of medicine is largely

prevailing in South Asian countries such as Pakistan, India, Srilanka and Bangladesh. About half of the industrialized world's population and 75-90% of the developing countries population still depend on the alternative systems of medicine. In Pakistan, traditional medicines have been a strong part of our cultural heritage and playing a significant role in providing health care to a large part of the population. The family Lamiaceae is one of the largest family of angiosperms found in Pakistan. Hedge, (1990) reported 60 genera and about 212 species of Lamiaceae present in Pakistan. Many species in the family Lamiaceae have been widely used for the treatment of cardiovascular diseases, stroke, and other conditions, and are largely used by indigenous people, herbal pharmaceutical industries and the local herb stores. Hence authenticating these species has become an important area of research. We need to have scientific evidence on quality, efficacy, purity and safety of our HMPs in addition to good manufacturing practice (GMP), compliant facilities and manufacturing processes. The quality assurance or quality control starts right away from the authentication of raw material used in the preparation of HMPs, therefore the suppliers must be placed in quality chain, more importantly bringing our herbal collectors and sellers in spot light to ensure quality.

There has been lack of concerted efforts for proper utilization of traditional medicines in the health care system. Globally there is a serious focus on regional and traditional ways of treatment to cure modern diseases. The Ministry of Health Pakistan along with National Council of Tibb (NCT), a regulatory authority for traditional and herbal system of medicine is determined to preserve and promote primitive practices by old practitioners and have recently initiated efforts to take measurements for quality assurance of herbal medicines, as it has bountiful capacity to enhance Pakistan's economy in traditional medicine. The demand for a product authentication service that utilizes molecular biotechnology can be met by DNA barcoding which will restore the consumer confidence. The regulatory authorities should not only adopt DNA barcoding as an improved scientific method for standardization of herbs but also for conservation of these resources.

Conclusion

Our study which was primarily designed to authenticate the HMPs of Lamiaceae can be taken as a case study for all the herbal products of Pakistan. Presently there is no standardization of herbal products based on an efficient and cost-effective scientific method. The herbal industry is suffering in terms of losing the customer confidence including the serious health risks. It is crucial to authenticate the socioeconomically vital medicinal plants including Lamiaceae species. Our findings showed that the herbal pharmaceutical industry and the local herb markets are probably not completely guilty on their part only because they lack the proper species identification system and mostly the contamination or adulteration occurs during the long supply chain of raw materials. DNA barcoding can help the regulatory authorities to devise a mechanism for quality control and may develop check-points during the long supply chain starting from local collectors to the market shelf. The herbal pharmaceutical industries are recommended that they should voluntarily establish the DNA barcoding for the testing of raw materials used in manufacturing their medicinal products. A herbal pharmaceutical industry may build his own herbal barcode library for reference sequence data depending on which plant species do they use in their products. It will be cost effective and will facilitate the rapid authentication to increase sovereign business interests and provide considerable health safety to consumers.

Acknowledgements

This research was supported by a grant from Higher Education Commission, Pakistan to Zabta Shinwari and from the US National Academy of Sciences (PGA-P210852) to Allan Showalter at Ohio University. The authors extend their gratitude to Allan Showalter and Melanie Schori for assistance with the research and providing the laboratory facilities at Department of Environmental and Plant Biology, Molecular and Cellular Biology Program, 315 Porter Hall, Ohio University, Athens, OH 45701 USA where the whole experimental work was performed.

References

Anonymous. 2002-2005. WHO Traditional Medicine Strategy 2002–2005. http://whqlibdoc.who.int/hq/2002/who_edm_trm_2002.1.pdf.

Anonymous. 2015. New York City Press Office. A.G. Schneiderman asks major retailers to halt sales of certain herbal supplements as DNA tests fail to detect plant materials listed on majority of products tested. 3 Feb, available from http://www.ag.ny.gov/press-release/ag-schneiderman-asks-major-retailers-halt-sales-certain-herbal-supplements-dna-tests (accessed 16.01.2016).

Baker, D.A., D.W. Stevenson and D.P. Little. 2012. DNA barcode identification of black cohosh herbal dietary supplements. *J. AOAC. Int*., 95: 1023-1034.

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

Chen, S., H. Yao, J. Han, C. Liu, J. Song, L. Shi, Y. Zhu, X. Ma, T. Gao , X. Pang, K. Luo, Y. Li , X. Li, X. Jia, Y. Lin and C. Leon, 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS One*, 5(1): e8613,

Cimino, M.T. 2010. Successful isolation and PCR amplification of DNA from National Institute of Standards and Technology herbal dietary supplement standard reference material powders and extracts. *Planta. Med.*, 76: 495-497.

Cuénoud, P., V. Savolainen, L.W. Chatrou, M. Powell, R.J. Grayer and M.W. Chase. 2002. Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbcL* , *atpB* , and *matK* DNA sequences. *Am. J. Bot*., 89: 132-14 .

Doyle, J.J. 1991. DNA protocols for plants, pp. 283-293. In: *Molecular techniques in taxonomy*. (Eds.): G. Hewitt, A.W.B. Johnson, and J.P.W. Young. NATO ASI Series H, Cell Biology 57. Springer-Verlag, Berlin, Germany.

Fay, M.F., C. Bayer, W.S. Alverson, A.Y. DeBruijn and M.W. Chase. 1998. Plastid *rbcL* sequence data indicate a close affinity between *Digodendron* and *Bixa*. *Taxon*., 47: 43-50.

Fay, M.F., S.M.S. Wensen and M.W. Chase. 1997. Taxonomic affinities of *Medusagyne oppositifolia* (Medusagynaceae). *Kew Bulletin*, 52: 111-120.

Fu, Y.M., W.M. Jiang and C.X. Fu. 2011. Identification of species within *Tetrastigma* (Miq.) Planch. (Vitaceae) based on DNA barcoding techniques. *J. Syst. Evol.,* 49(3): 237-245.

Gao, T., H. Yao, J. Song, Y. Zhu, C. Liu and S. Chen. 2010. Evaluating the feasibility of using candidate DNA barcodes in discriminating species of the large Asteraceae family. *BMC Evol. Biol.,* 10: 324-330.

Han, J.P, L.C. Shi, X.C. Chen and Y.L. Lin. 2012. Comparison of four DNA barcodes in identifying certain medicinal plants of Lamiaceae. *J. Syst. Evol*., 50: 227-234.

Hebert, P.D.N., M.Y. Stoeckle, T.S. Zemlak and C.M. Francis. 2004. Identification of birds through DNA barcodes. *PLoS. Bio.l*, 2: e312.

Hedge, I.C. 1992. A global survey of the biogeography of the Labiatae. In: *Advances in Labiate science*. (Eds.): Harley, R.M. & T. Reynolds. Kew: Royal Botanic Gardens, pp. 7-17.

Hollingsworth, P.M., L.L. Forrest, J.L. Spouge, M. Hajibabaei, S. Ratnasingham, M. van der Bank, M.W. Chase, R.S. Cowan, D.L. Erickson, A.J. Fazekas, S.W. Graham, K.E. James, K.-J. Kim, W.J. Kress, H. Schneider, J. van AlphenStahl, S.C.H. Barrett, C. van den Berg, D. Bogarin, K.S. Burgess, K.M. Cameron, M. Carine, J. Chacón. A. Clark, J.J. Clarkson, F. Conrad, D.S. Devey, C.S. Ford, T.A.J. Hedderson, M.L. Hollingsworth, B.C. Husband, L.J. Kelly, P.R. Kesanakurti, J.S. Kim, Y.-D. Kim, R. Layahe, H.-L. Lee, D.G. Long, S. Madriñán. O. Maurin, I. Meusnier, S.G. Newmaster, C.-W. Park, D.M. Percy, G. Petersen, J.E. Richardson, G.A. Salazar, V. Savolainen, O. Seberg, M.J. Wilkinson, D.-K. Yi and D. P. Little. 2009. A DNA barcode for land plants. *Proc. Nat. Acad. Sci.,* 106(31): 12794-12797.

Kress, W.J., K.J. Wurdack, E.A. Zimmer, L.A. Weigt and D.H. Janzen. 2005. Use of DNA barcodes to identify flowering plants. *Proc. Nat. Acad. Sci.,* 102(23): 8369-8374.

Lin, C.C, C.C. Yang, D.H. Phua, J.F. Deng and L.H. Lu. 2010. An outbreak of foxglove leaf poisoning. *J. Chin. Med. Assoc*., 73: 97-100.

Lou, S.K., K.L. Wong, M. Li, P.P. But, S.K. Tsui and P.C. Shaw. 2010. An integrated web medicinal materials DNA database: MMDBD (Medicinal Materials DNA Barcode Database). BMC Genomics, 11: 402.

Michel, C.I., S.M. Rachel, T. Yanille and M. Jeanmaire. 2016. The nuclear internal transcribed spacer (ITS2) as a practical plant DNA barcode for herbal medicines. *J. Appl. Res. Med. Aromat. Plants*. http://dx.doi.org/10.1016/ j.jarmap.2016.02.002

Newmaster, S.G., M. Grguric, D. Shanmughanandhan, S. Ramalingam and S. Ragupathy. 2013. DNA barcoding detects contamination and substitution in North American herbal products. *BMC Med*., 11: 222.

Sang, T., D.J. Crawford and T.F. Stuessy. 1997. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *Am. J. Bot*., 84: 1120-1136.

Schori, M. and A.M. Showalter. 2011. DNA barcoding as a means for identifying medicinal plants of Pakistan. *Pak. J. Bot*., 43: 1-4. Special Issue, December, 2011 (Medicinal Plants: Conservation & Sustainable Use).

Schori, M., M. Appel, A. Kitko and A.M. Showalter. 2013. Engineered DNA polymerase improves PCR results for plastid DNA. *Applications in Plant Sciences*, 1: 1-7.

Sen, S. and R. Chakraborty. 2015."Toward the Integration and Advancement of Herbal Medicine: A Focus on Traditional Indian Medicine." *BTAT Botanics: Targets and Therapy*, 33. doi:10.2147/btat.s66308.

Shinwari, Z.K.; R. Terauchi; F.H. Utech and S. Kawano 1994a Recognition of the New World Disporum Section Prosartes as Prosartes (Liliaceae) based on the sequence data of the rbcL gene. Taxon 43(3): 353-366.

Shinwari, Z.K.; R. Terauchi and S. Kawano 1994b Phylogenetic relationships among genera in the Liliaceae-Asparagoideae-Polygonatae sensu lato inferred from rbcL gene sequence data. PI. Systematic & Evolution 192: 263-277.

Shinwari, Z.K. 1995 Congruence between morphology and molecular phylogeneties in Prosartes (Liliaceae). Pak. J. Bot. 27(2): 361-369.

Shinwari, Z.K 1998 Molecular Systematics of the genus *Uvularia* and related taxa based upon *rbc*L gene sequence data. Pak. J. Bot. 30(2): 161-172

Shinwari, Z. K. 2002 Sequence divergence of *rbc*L gene and Phylogenetic relationships in Liliales. Pak.J.Bot. 34(2): 191-204

Shinwari, Z.K. 2010. Medicinal plants research in Pakistan. *J. Med. Plant. Res*., 4: 161-176.

Shinwari, Z.K. and Shehla Shinwari 2010 Molecular Data and Phylogeny of Family Smilacaceae. *Pak. J. Bot.,* Special Issue (S.I. Ali Festschrift) 42: 111-116

Shinwari, Z.K., K. Jamil and N.B. Zahra. 2014 Molecular systematics of selected genera of subfamily Mimosoideae-fabaceae. *Pak. J. Bot.,* 46(2): 591-598, .

Stoeckle, M.Y., C.C. Gamble, R. Kirpekar, G. Young, S. Ahmed and D.P. Little. 2011. Commercial teas highlight plant DNA barcode identification successes and obstacles. *Sci. Rep*., 1: 42.

Sucher, N.J. and M.C. Carles. 2008. Genome-based approaches to the authentication of medicinal plants. *Planta. Med*., 74: 603-623.

Tate, J.A. and B.B. Simpson. 2003. Paraphyly of *Tarasa* (Malvaceae) and diverse origins of the polyploid species. *Syst. Bot*., 28: 723-737.

Vijayan, K. and C.H. Tsou. 2010. DNA barcoding in plants: taxonomy in a newperspective. *Curr. Sci*., 99: 1530-1541.

Wallace, L.J., S.M. AL. Boilard, S.H.C. Eagle, J.L. Spall, S. Shokralla and M. Hajibabaei. 2012. DNA barcodes for everyday life: routine authentication of natural health products. *Food. Res. Int.*, 49: 446-452.

Yao, H., J. Song, C. Liu, K. Luo, J. Han, Y. Li , X. Pang, H. Xu, Y. Zhu, P. Xiao and S. Chen. 2010. Use of ITS2 region as the universal DNA barcode for plants and animals. *PLoS One*, 5: e13102.

(Received for publication 26 August 2015)