UDC 575.630
https://doi.org/10.2298/GENSR2002435A
Original scientific paper

PHYLOGENY AND DIVERSITY OF LAMIACEAE BASED ON *RPS*14 GENE IN PAKISTAN

Asma AYAZ[1,2*], Wajid ZAMAN[2], Saddam SAQIB[2], Fazal ULLAH[2], Tariq MAHMOOD[2*]

[1]State Key Laboratory of Biocatalysis and Enzyme Engineering, School of Life Sciences, Hubei University, Wuhan 434200, China
[2]Department of Plant Sciences, Quaid- i- Azam University Islamabad, 45320, Pakistan

Ayaz A., W. Zaman, S. Saqib, F. Ullah, T. Mahmood (2020). *Phylogeny and diversity of Lamiaceae based on rps14 gene in Pakistan*.- Genetika, Vol 52, No.2, 435-452.
The present research work was carried out to conduct concordance between the molecular data and the phylogenetic relationship between the selected species of Lamiaceae. We studied 34 species collected from various phytogeographical areas of Pakistan. Molecular DNA was isolated from fresh leaf specimens of the selected taxa. The *rps*14 gene was amplified for the isolation of DNA sequencing and the amplified products were sequenced and analyzed for phylogenetic analysis. The sequenced products were analyzed using bioinformatics tools such as MEGA7, I-TASSER (Iterative Threading Assembly Refinement) and SAVES (The Structure Analysis and Verification Server). ExPASY translate tool was used for the translation of nucleotide sequences to amino acid sequences. *Phlomis cashmeriana* with *Origanum vulgare,* and *Lamium album* with *Lamium amplexicaul* showed a close relationship in both phylogenetic trees with well-represented bootstrap (BS) values. The pairwise distance ranged from 0.013 to 0.19 with an overall mean distance of 0.068 was also observed. I-TASSER software was utilized to predict the 3D protein structures followed by stereochemical analysis. Based on the present finding it is determined that the *rps*14 gene would be used as a DNA barcode for the identification of plant taxa. Additionally, it is needed for the development of additional reliable molecular markers to resolve the systematics issues in the family Lamiaceae.
Keywords: Lamiaceae*;* *rps*14; phylogenetic analysis; I-TASSER; molecular markers

Corresponding authors: Asma Ayaz/Tariq Mahmood, School of Life Sciences, Hubei University, Wuhan 434200, China/ Department of Plant Sciences, Quaid- i- Azam University Islamabad, 45320, Pakistan, E-mail: asmaayaz@bs.qau.edu.pk/tmahmood@qau.edu.pk

INTRODUCTION

Plants constitute a major component of foodstuffs in human life and play a key role in health, medicine and climate stability. From the beginning of life history, plants are the prestigious health facility for the preservation and treatment of different ailments and diseases. According to an estimate, approximately 4,22,000 plant species have been reported from all over the world at present, in which, more than 35,000 species are used for therapeutic purposes around the globe, 20,000 are edible and consumed by 90 % of the population (HASAN *et al.*, 2007). Among all these, spices and their derivatives are used as flavoring agents in food since ancient times and have diverse biological activities especially the members of Family Lamiaceae are known for these properties (FAKIR *et al.*, 2015).

Family Lamiaceae distributed cosmopolitan having 236 genera and 7200 species worldwide. The Center of the diversity of Lamiaceae is Central Asia and Mediterranean regions with a great diversity of the taxa of Lamiaceae (PERVEEN QAISER, 2004). Besides these phytogeographical areas, other regions of the world including Afghanistan, the Arabian Peninsula, Ethiopia, India, Iran, Nepal, Somalia, Sudan and Egypt also host major genera like *Nepeta*, *Eremostachys,* and *Salvia* species of Lamiaceae (NAGHIBI *et al.*, 2010). Pakistan has also a rich diversity of the taxa of family Lamiaceae, about 212 taxa and 60 genera have been recorded (PERVEEN QAISER, 2004). Its complex inflorescence, morphological variations, essential oils presence (Chemotaxonomy) and closely related to *Verbenaceae* serve as the basis of modern taxonomic classification (CANTINO, 1992; FEDERICI *et al.*, 2013; TALEBI, 2016).

Species identification and species boundaries defining are an important part of exploring and unfolding natural biodiversity. Traditionally, morphological characteristics have been used by expert taxonomists, systematic biologists and trained practitioners to identify species after earning substantial experience in the field (EFTEKHARIAN *et al.*, 2018); KRESS and ERICKSON, 2008). As the significance of biodiversity is increasing day by day in the field of agriculture, evolutionary biology, ecology, conservation biology and other fields, therefore correct identification of plant taxa is very necessary, while DNA barcode is an important tool for the modern-day plant identification and delimitation (BAYAT *et al.*, 2018; GALOVIĆ *et al.*, 2018; WANG *et al.*, 2018). Recent advances in the field of genomics and bioinformatics are incomplete without the use of DNA sequences (SEVĠNDĠK and YALÇIN, 2018; SAVIĆ *et al.*, 2019). DNA sequences have become a major source of collecting information. Thus, researchers have examined the idea that an easy and rapid way of species identification and defining species boundaries is possible with the use of short DNA sequence (ERIĆ *et al.*, 2019; SEVINDIK, 2019), which possess a standardized position in the genome, called as DNA barcode (HEBERT *et al.*, 2004; FRÉZAL and LEBLOIS, 2008). DNA barcoding is a new approach designed to provide rapid, accurate and automatable species identification by using short standardized gene regions as internal species tags (AKINRO *et al.*, 2019; GRITSENKO *et al.*, 2019; PATI *et al.*, 2019). Molecular identification methods have been used in forensics, food and taxonomy-based identification (TELETCHEA *et al.*, 2008). The key objectives of DNA barcoding are to allocate species to the unidentified samples and improve the identification especially in cryptic and other organisms having a complex morphological characters (HEBERT *et al.*, 2003).

Due to finding an appropriate locus for DNA barcode, the technique is complex and challenging. There are some conserved regions like *rbcL* (rubisco large subunit) that can be

used in forest dynamics and functional evolution (KRESS and ERICKSON, 2008). The issue of finding a plant barcode is complex and yet not resolved as compared to animals. According to many scientists, to distinguish a plant species, multiple molecular markers are required instead of single markers (HOLLINGSWORTH *et al.*, 2011). Here, we presented genetic characterizations of Lamiaceae, their phylogenetic relationship within the species and bring classification of the family into line with its hypothesized evolutionary history. This is the first reference documentation of Lamiaceae by I-TASSER software to predict the 3D protein structures followed by stereochemical analysis. Furthermore, genetic variation based on *rps*14 among selected species of different genera like *Mentha, Phlomis, Ajuga, Isodon, Salvia, Nepeta* and *Ocimum* collected from different regions in Pakistan were studied.

MATERIALS AND METHODS

Plant Material Collection

A total of 34 species belonging to 24 genera of family Lamiaceae were collected from various geographical areas of Pakistan including Islamabad, Quetta, Malakand, Shangla, Margalla Hills, Swat, Kalam, Matta, Madyan and Bahrain. Collections were made in the flowering season of family Lamiaceae during 2016-17 (April-October). After collection, plant species were morphologically identified by using e-flora of Pakistan and their taxonomical authentication was done (ALI QAISER, 1993). The list of plant species with their location is given in (Table 1). Fresh leaves were picked, preserved in sealed zipper bags and stored at 4°C till further process.

Table 1. List of selected species of Lamiaceae with their geographic coordinates.

S. No.	Plant Name	Location	Geographic coordinates*
1.	*Ajuga parviflora*	Kashmir	33.7782 ° N, 76.5762° E
2.	*Anisomeles indica*	Islamabad	33.7294° N, 73.0931° E
3.	*Clinopodium debile*	Kurram	33.6960° N, 70.3361° E
4.	*Clinopodium umbrosum*	Shangla	34.8015° N, 72.7570° E
5.	*Colebrookea oppositifolia*	Buner	34.3943° N, 72.6151°E
6.	*Isodon rugosus*	Shangla	34.8015° N, 72.7570° E
7.	*Lamium album*	Shangla	34.8015° N, 72.7570° E
8.	*Lamium amplexicaule*	Islamabad	33.7294° N, 73.0931° E
9.	*Leucas cephalotes*	Islamabad	33.7294° N, 73.0931° E
10.	*Marrubium vulgare*	Quetta	30.1830° N, 66.9987° E
11.	*Mentha arvensis*	Shangla	34.8015° N, 72.7570° E
12.	*Mentha longifolia*	Shangla	34.8015° N, 72.7570° E
13.	*Mentha spicata*	Shangla	34.8015° N, 72.7570° E
14.	*Micromeria biflora*	Shangla	34.8015° N, 72.7570° E
15.	*Nepeta erecta*	Parachinar	33.8837° N, 70.1099° E
16.	*Nepeta juncea*	Ziarat	30.3829° N, 67.7243°E

17.	*Nepeta laevigata*	Shangla	34.8015° N, 72.7570° E
18.	*Ocimum americanum*	Chitral	35.8523° N, 71.7871° E
19.	*Ocimum basilicum*	Bannu	32.9298° N, 70.6693° E
20.	*Origanum vulgare*	Chitral	35.8523° N, 71.7871° E
21.	*Phlomidoschema parviflorum*	Kuram	33.6960° N, 70.3361° E
22.	*Phlomis cashmeriana*	Kuram	33.6960° N, 70.3361° E
23.	*Plectranthus barbatus*	Gilgit	35.9202° N, 74.3080° E
24.	*Prunella vulgaris*	Shangla	34.8015° N, 72.7570° E
25.	*Rydingia limbata*	Swat	35.4920° N, 72.5205° E
26.	*Salvia coccinea*	Islamabad	33.7294° N, 73.0931° E
27.	*Salvia moorcroftiana*	Dir Lower	34.8453° N, 71.9046° E
28.	*Salvia plebeian*	Islamabad	33.7294° N, 73.0931° E
29.	*Teucrium stocksianum*	Kuram	33.6960° N, 70.3361° E
30.	*Thymus linearis*	Ziarat	30.3829° N, 67.7243°E
31.	*Vitex agnus-castus var. pseudo-negundo*	Quetta	30.1830° N, 66.9987° E
32.	*Vitex negundo*	Islamabad	33.7294° N, 73.0931° E
33.	*Zataria multiflora*	Ziarat	30.3829° N, 67.7243°E
34.	*Ziziphora tenuior*	Ziarat	30.3829° N, 67.7243°E

Geographic coordinates* N= North, E= East

DNA Extraction and amplification

We followed the CTAB method for DNA extraction for plant leaves. The reaction mixture was prepared with CTAB (2x) buffer, 20 mM ethylene diamine tetraacetic acid (EDTA), 100 mM tris HCl, 1.4 g Sodium chloride (NaCl), 1 % mercaptoethanol and 2 % w/v CTAB were mixed. The pH of the buffer was adjusted to 8.0 (POREBSKI *et al*., 1997). The fresh leaf part of the plant was gently excised through using autoclaved scissors, washed with distilled water and 70 % ethanol, and then dried at room temperature. CTAB method was followed for DNA extraction from plant leaves. About 0.4 g of leaves of the plant were crushed utilizing mortar and pestle through adding preheated (65°C) 2 x CTAB buffer (1-2 ml) until the material becomes homogenized. The homogenized mixture was then incubated in an autoclaved Eppendorf tube at 65°C for 45 to 60 minutes. Centrifugation of the incubated material was done for 20 minutes at 12000 rpm and then the supernatant was separated. The together supernatant was mixed with an equivalent volume of chloroform: isomyl alcohol, which was ready by a ratio of (24:1) correspondingly. The material was gently mixed by inversion 2 to 5 times (POREBSKI *et al*., 1997). Again, centrifugation at 12000 rpm for 15 minutes was performed. The supernatant was together, and the entire procedure was repeated four to five times till the supernatant became dark yellow to white. After washing, an equivalent volume of chilled isopropanol was added to the supernatant that precipitated the DNA. This was followed by gentle inversions for 2 to 5 times and the mixture was kept overnight at -20°C. Once again, the mixture was centrifuged at

12000 rpm for 20 minutes. Subsequently, aqueous phase (isopropanol) was removed and the DNA pellet was retained. To remove impurities, 70 % ethanol was used to wash the pellet. The pellet was air-dried at room temperature. To degrade RNAs, Tris EDTA (TE) (0.1 X) buffer with RNase was added to the DNA pellet. Finally, the samples were stored at -20°C for additional use. The presence of genomic DNA was confirmed utilizing gel electrophoresis containing 1 % agarose gel, followed by ethidium bromide staining. After staining the gel was observed under UV light by Dolphin Doc plus documentation system (Wealtec).

The primers were designed by using primer 3 (version 4.0) (http://bioinfo.ut.ee/primer3-0.4.0/) by executing the *rps*14 gene sequence obtained from the cpDNA sequence of *Nicotiana tabacum* available in "NCBI database (https://blast.ncbi.nlm.nih.gov/Blast.cgi)". The specified used as forward was *rps*14 - F: 5'- ATGGCAAGGAAAAGTTTGATTC-3' while reverse was *rps*14 - R: 5'- TTACCAACTTGATCTTGTTGCTCCT- 3'. The sequences were amplified through polymerase chain reaction (PCR) containing total reaction mixture (25 µL) of nano pure water, Taq buffer, forward and reverse *rps*14 primers, Taq DNA polymerase, dNTPs, MgCl2 and DNA template. The concentration and volume of PCR components used for PCR mixtures are given in Table II. The amplified regions remained purified using GeneJet PCR Purification Kit (Thermo Scientific) and sent to and Beijing Genomic Institute (BGI), Shenzhen, China for sequencing. The amplicons of the *rps*14 gene were sequenced and taxonomic relation was found for individually species separately, using the Basic Local Alignment Search Tool (BLAST) package within the "National Centre for Biotechnology Information Gene Bank (https://www.ncbi.nlm.nih.gov)". All sequences were deposited to the NCBI database.

Table 2. Concentration and volume of PCR reagents used for the amplification of the rps14 gene.

S/N	Reagents	Concentration	Volume
1	Nano pure water		16.2 µl
2	*Taq* buffer	10 X	2.5 µl
3	$MgCl_2$	25 mM	1.5 µl
4	dNTPs	2 mM	1.5 µl
5	Forward primer	25 pmoles	1 µl
6	Reverse primer	25 pmoles	1 µl
7	*Taq* DNA polymerase	1.5 U	0.3 µl
8	Genomic DNA		1 µl
9	Total volume		25 µl

Phylogenetic analysis

The obtained sequences were assembled and aligned using MEGA7 (KUMAR *et al.*, 2016). The matrices were modified by insertion and deletion manually coded as present/absent. The phylogenetic relationship of all selected species was examined using the neighbor-joining method through MEGA7. Furthermore, pairwise distance analysis was conducted using MEGA 7

to determine the evolutionary distinction among *rps*14 gene sequences. Tajima's neutrality test was calculated based on *rps*14 gene sequences to examine nucleotide diversity among 34 species of family Lamiaceae. The nucleotide average number for the *rps*14 gene was also analyzed by MEGA7 software to assess the genetic relationship among selected species of Lamiaceae. Substitution patterns and rates were projected under the TAMURA and NEI (1993) model.

Translation of nucleotide sequence and protein BLAST

ExPASY translate package was used for nucleotide sequence translation (http://web.expasy.org/translate/), whereas sequences were aligned by online web server JustBio. Amino acid sequences of all 34 members were uploaded at NCBI. Protein BLAST was carried out for each species separately and correlated with already reported amino acid sequences in Genbank.

Protein structure and function detection

3-dimensional (3-D) protein structure and function were detected using I-TASSER (Iterative Threading Assembly Refinement) server. I-TASSER is an internet service for the estimate of 3-dimensional (3-D) protein structure and function. Amino acid sequences of *rps*14 gene of selected species were uploaded one by one on I-TASSER server and five 3D atomic models for each sequence were observed following the protocol of (ZHANG, 2008).

3D atomic models of *rps*14 gene were studied and protein structure was confirmed via Ramachandran plot analysis and PROCHECK. 3D models were uploaded on SAVES (The Structure Analysis and Verification Server) in pdb files for Ramachandran plot and PROCHECK analysis. PROCHECK provides the stereochemical quality of predicted protein structures by evaluating residues by residues geometry (http://services.mbi.ucla.edu/SAVES/). Ramachandran plot assessment was analyzed by uploading pdb files of 3D models on RAMPAGE (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php).

RESULTS AND DISCUSSION

Amplification of Genomic DNA

The band size of PCR products for all species was approximately 303 bp. The average of various nucleotide numbers for *rps*14 was A=33.7, C=18.9, T (U)=24.8 and G=22.7. An overall average value of the nucleotides is 303.0, which shows that A secures the highest average number while C displays the lowest (Table 3). Due to the complexity of family Lamiaceae taxonomy, identification based on morphological features alone is difficult, hence DNA barcodes study provides an ideal case in solving the taxonomic problem in the family. Our study is in agreement with a previous study on Lamiaceae (ZAHRA *et al.*, 2016). DNA barcoding is an advanced approach for the systematic and identification of taxon using a short genomic region (HEBERT *et al.*, 2003). Information regarding DNA barcoding on Lamiaceae is very deficient in the literature (AKINRO *et al*., 2019). However, in the past few studies tested the utility of DNA barcoding in family Lamiaceae in terms of genetic variation (DE MATTIA *et al.* (2011).

Table 3. Nucleotide composition of rps14 gene sequenced from 34 different selected species of family Lamiaceae

S. No.	Plant Name	Accession number	T/U	C	A	G	Total
1.	*Ajuga parviflora*	MF678776	24.4	18.8	33.0	23.8	303.0
2.	*Anisomeles indica*	MF678774	24.4	17.5	34.3	23.8	303.0
3.	*Clinopodium debile*	MF678772	24.1	18.5	34.7	22.8	303.0
4.	*Clinopodium umbrosum*	MF678775	24.8	18.2	34.7	22.4	303.0
5.	*Colebrookea oppositifolia*	MF678778	24.1	18.2	34.0	23.8	303.0
6.	*Isodon rugosus*	MF687462	24.4	17.8	33.0	24.8	303.0
7.	*Lamium album*	MF687472	23.4	19.5	34.7	22.4	303.0
8.	*Lamium amplexicaule*	MF687475	23.1	20.3	35.3	21.4	295.0
9.	*Leucas cephalotes*	MF687471	24.1	19.5	33.7	22.8	303.0
10.	*Marrubium vulgare*	MF678782	24.1	19.5	33.0	23.4	303.0
11.	*Mentha arvensis*	MF687479	24.1	17.8	33.0	25.1	303.0
12.	*Mentha longifolia*	MF687465	24.4	17.8	33.7	24.1	303.0
13.	*Mentha spicata*	MF687463	24.4	17.8	33.7	24.1	303.0
14.	*Micromeria biflora*	MF678783	23.4	18.5	33.7	24.4	303.0
15.	*Nepeta erecta*	MF687468	24.8	18.5	34.7	22.1	303.0
16.	*Nepeta juncea*	MF678779	23.8	18.5	34.0	23.8	303.0
17.	*Nepeta laevigata*	Mf687469	24.4	18.8	33.7	23.1	303.0
18.	*Ocimum americanum*	MF678777	24.1	18.8	33.3	23.8	303.0
19.	*Ocimum basilicum*	MF687470	24.1	18.2	33.3	24.4	303.0
20.	*Origanum vulgare*	MF687467	24.4	18.5	32.7	24.4	303.0
21.	*Phlomidoschema parviflorum*	MF678781	24.1	17.8	33.3	24.8	303.0
22.	*Phlomis cashmeriana*	MF687466	24.1	17.7	34.1	24.1	311.0
23.	*Plectranthus barbatus*	MF687474	24.4	18.5	34.7	22.4	303.0
24.	*Prunella vulgaris*	MF687480	24.4	18.5	35.0	22.1	303.0
25.	*Rydingia limbata*	Mf687481	24.4	20.8	32.0	22.8	303.0
26.	*Salvia coccinea*	MF687477	23.8	18.5	33.3	24.4	303.0
27.	*Salvia moorcroftiana*	MF678784	23.4	19.1	33.3	24.1	303.0
28.	*Salvia plebeian*	MF687473	24.1	18.5	34.7	22.8	303.0
29.	*Teucrium stocksianum*	MF687478	23.8	18.8	34.0	23.4	303.0
30.	*Thymus linearis*	MF678780	24.1	18.5	34.3	23.1	303.0
31.	*Vitex agnus-castus var. pseudo-negundo*	MF678785	25.1	19.5	31.7	23.8	303.0
32.	*Vitex negundo*	MF678773	25.4	19.1	34.3	21.1	303.0
33.	*Zataria multiflora*	MF687476	23.8	18.2	32.7	25.4	303.0
34.	*Ziziphora tenuior*	MF687464	24.4	18.8	33.0	23.8	303.0
	Average		24.8	18.9	33.7	22.7	303.0

Phylogenetic Tree Based on Nucleotide Sequences of rps14

The phylogenetic tree constructed using MEGA7 software revealed two major groups, we took *Zataria multiflora* as an outgroup for the present study, the group I have two species *Vitex negundo* and *Vitex agnus-castus* var. *pseudo negundo* taxa with 93 BP values. The rest of the 31 species occur in separate group II, which is further divided into subgroups, sections, subsections, and clusters. Subgroup I have one species *Anisomeles indica*, which share 68 BP value with subgroup II, which make this species genetically different from the rest of the studied species. Subgroup II contain 30 species which is further divided into two sections, section I have 12 species which have two subsections, subsection I have *Colebrookea oppositfolia*, while subsection II has 11 species with total 76 BP values *Nepeta erecta* and *Nepeta leavigata* formed one cluster with 100 BP values, while *Lamium album* and *Lamium amplexicaule* come in another cluster. The species *Marrubium vulgare*, and *Phlomidoschema parviflorum* come in one cluster. *Rydingia limbata* and *Leucas cephalotes* are come with 91 BP and do not have a cluster with any other species. The two species of *Clinopodium umbrosum* and *Clinopodium debile* fall in the same cluster with 100 BP values. Subsection II has 18 taxa *Teucrium stocksianum*, *Prunella vulgaris* (cluster together), *Ajuga parviflora*, *Plectranthus barbatus*, *Isodon rugisus*, *Ocimum basillicum*, *Ocimum americanum*, *Salvia moorcroftiana*, *Salvia coccinea*, *Salvia plebeian*, *Zizpiphora tenuior*, *Micromeria biflora*, *Thymus linearis*, *Phlomis cashmeriana, Origanum vulgare*, *Mentha spicata*, *Mentha longifoloia* (Figure 1).

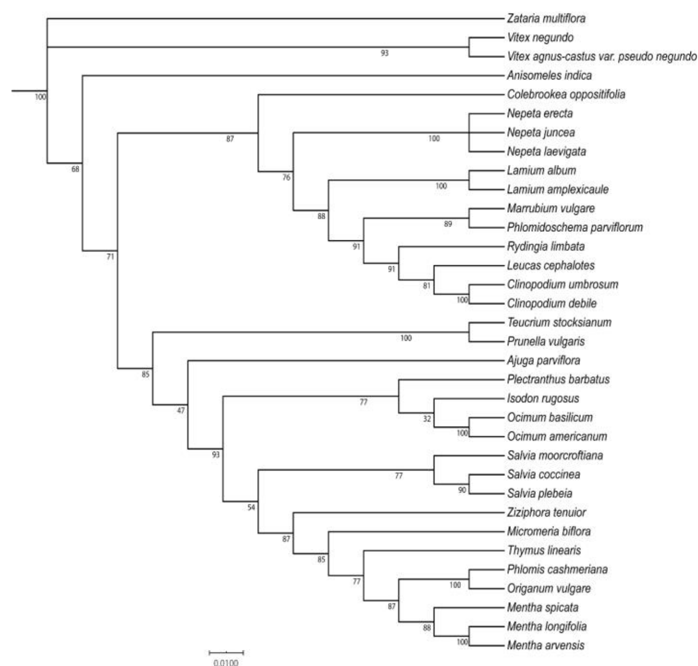


Figure 1. Phylogenetic tree of selected species of family *Lamiaceae* based on *rps*14 gene sequences

According to the molecular analysis, the species in each cluster showed genetic differences and species boundaries could be easily identified in the studied taxa. According to previous studies, several genomic regions are applied in taxonomy and phylogeny of family Lamiaceae. However, from Pakistan, a study conducted on DNA barcoding of family Lamiaceae utilized *rbcL*, *matK* and *psbA-trnH*, but our approach was different from (ZAHRA *et al.*, 2016). Their study was based on available sequences from Genbank while we made our own Genbank by collecting and extraction of DNA from natural habitat. Therefore, this study was restricted to the genus level due to the availability of insufficient sequences (ZAHRA *et al.*, 2016), whereas our study cover species level classification based on phylogenetic analysis. Our study provides some basic information based on the *rps*14 gene to delimit these selected species in the Family Lamiaceae from Pakistan. The present study is in accordance with the previously published articles on family Lamiaceae taxa (ZAHRA *et al.*, 2016; (CURTO *et al.*, 2012). Some other studied on family Lamiaceae have been done utilizing the genetic diversity through short sequencing for delimitation of the taxa in the family through using DNA barcoding (ERIĆ *et al.*, 2019; SEVINDIK, 2019). These selected species we studied in the present work showed that they are paraphyletic which have been studied in the previous work on Lamiaceae (SCHEEN *et al.*, 2009).

Tajima's Neutrality Test

We analyzed Tajima's neutrality test to detect past population growth of family Lamiaceae. Thirty-four numbers of sequences (m) have given 88 separation sites (S) revealing very little nucleotide diversity (π) of 0.064220 (Table 4). This little nucleotide diversity is a sign of the close genetic relationship among different species of Lamiaceae. The low nucleotide diversity revealed that there is an excess of rare mutations in this family. Previous studies demonstrated that direct or indirect selection in natural populations are rare (WANG H.T. *et al.*, 2013). In comparison with other genome regions, we found low nucleotide diversity for *rps*14 which shows *rps*14 the most promising gene as the land plant barcode according to the measures of universal application and high sequence divergence among the species (AKINRO *et al.*, 2019; WANG H. T. *et al.*, 2013; ZAHRA *et al.*, 2016).

Table 4. Tajima's neutrality test values based on the rps14 gene for thirty-four different species of Lamiaceae using MEGA7.

No. of sequences "m"	No. of segregating sites "S"	Ps= S/n	Θ=Ps/a1	Nucleotide diversity "π"	Tajma test statistic "D"
34	88	0.300341	0.073455	0.064220	-0.469462

Maximum Likelihood (ML) Estimate of Substitution Matrix

Substitution design and rates were valued under the TAMURA and NEI (1993) model. Relative values of instantaneous (r) would be considered when estimating them. For ease, the sum of (r) values was made equivalent to 100. The transitional substitution rates were originated higher than rates of different transversional substitutions. The nucleotide frequencies were A=

33.7 %, T/U= 24.8 %, C= 18.9 %, and G= 22.7 %. For estimating ML values, a tree topology was automatically computed. The maximum log-likelihood for this computation was -1955.610. This analysis involved 34 nucleotide sequences. There was a total of 303 positions in the final dataset. Detailed information about the rate of different transitional (bold) and transversional substitutions (italics) is shown in (Table 5). This analysis indicates that the GC content of coding regions are higher than in introns. Based on these results our study is in agreement with previous studies indicating a higher percentage of GC contents, showing minor variations (H. T. WANG *et al.*, 2013).

Table 5. Values of maximum likelihood, transitional (bold) and transversional substitution (italics) of the nucleotide sequences for rps14 gene from 34 species of Lamiaceae using MEGA7.

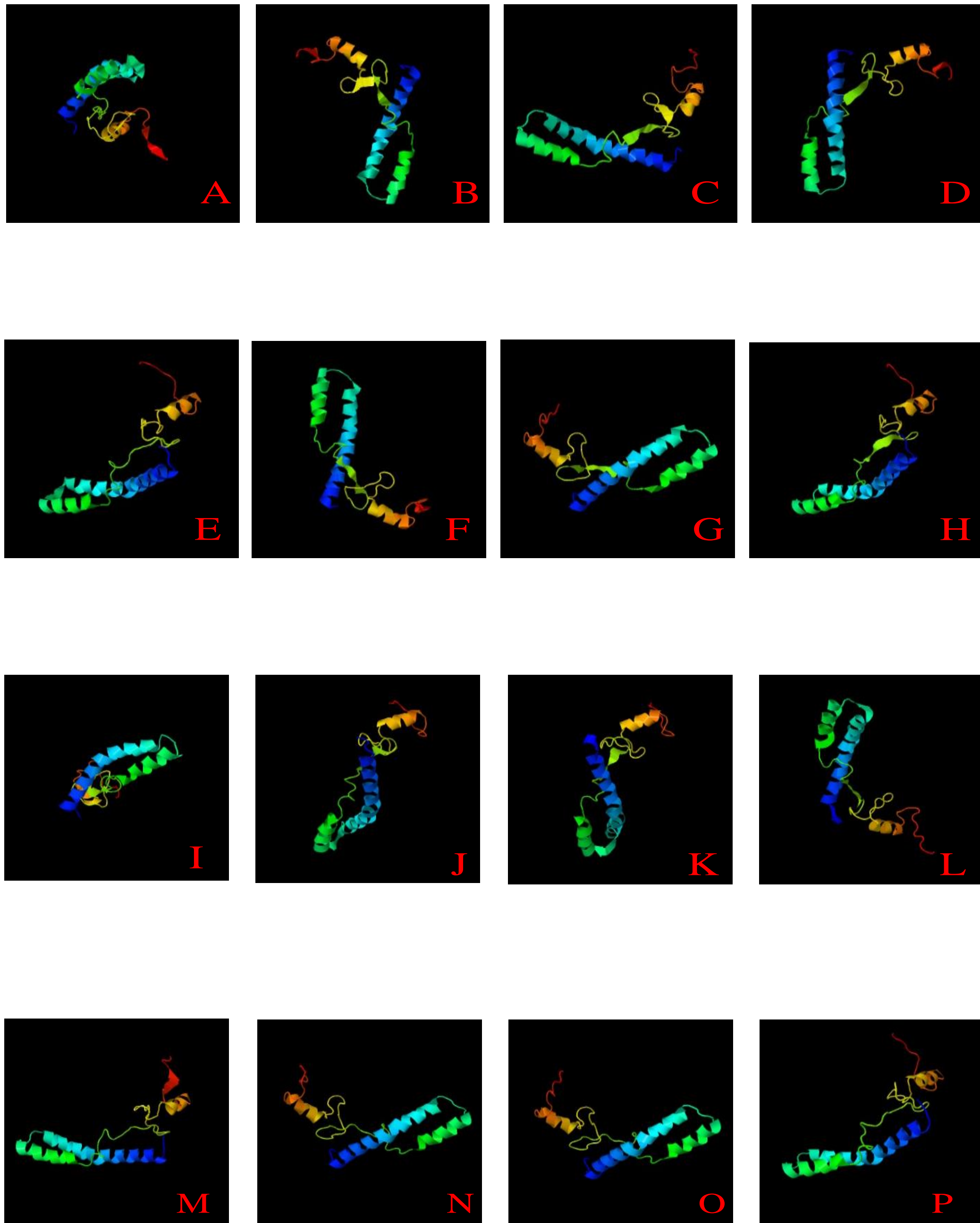
	A	T/U	C	G
A	-	*4.18*	*3.21*	20.00
T/U	*5.83*	-	7.28	*4.07*
C	*5.83*	9.47	-	*4.07*
G	28.68	*4.18*	*3.21*	-

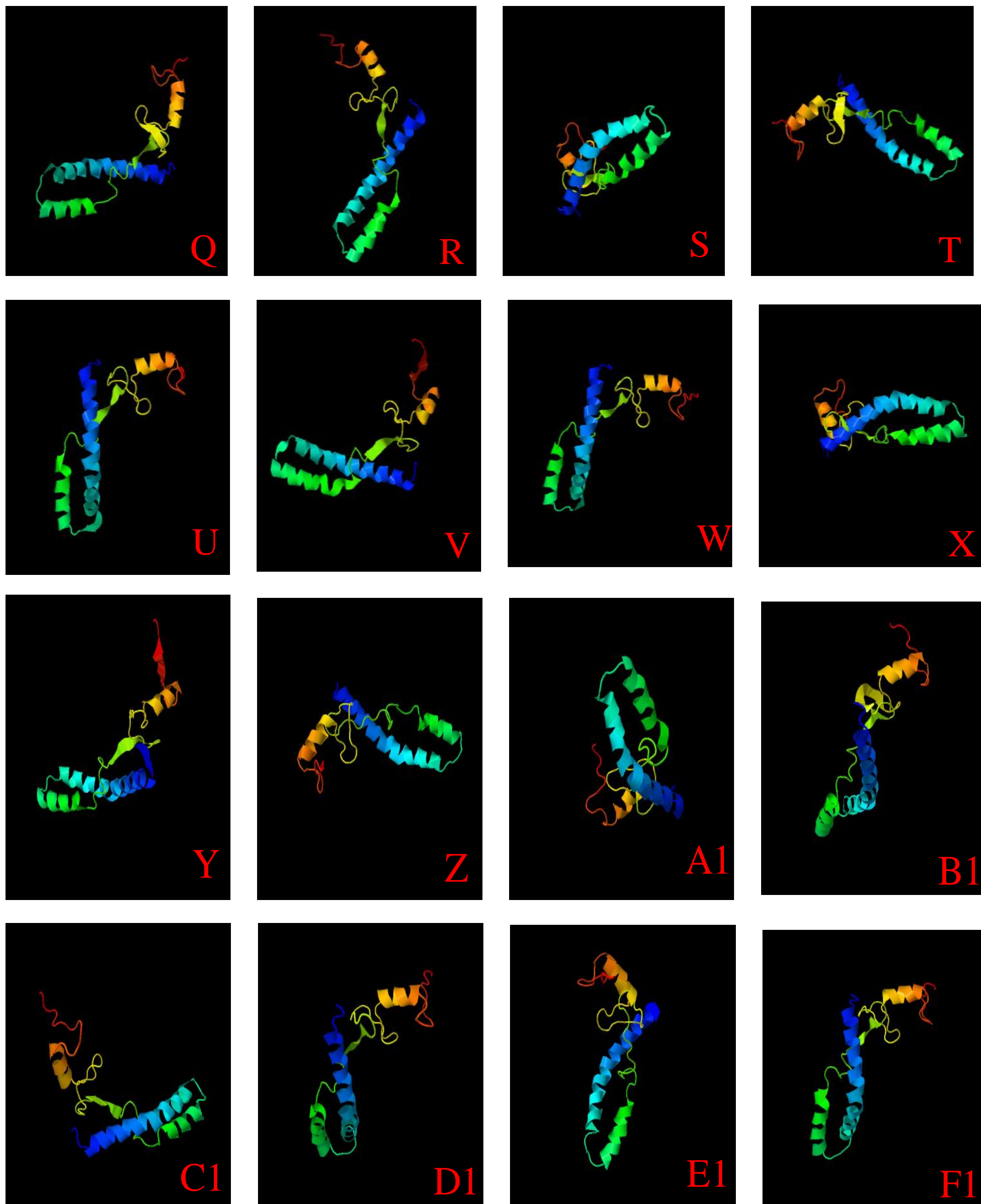
Estimation of Evolutionary Divergence by Pairwise Distance Calculation

Evolutionary divergence among *rps*14 gene sequences was estimated by conducting the pairwise distance analysis using MEGA7. The value of genetic diversity was found to be in the range of 0.013 and 0.194 with the overall mean distance of 0.068 for *rps*14 gene sequences. The very low distance values illustrate that all selected species have low genetic diversity among them based on *the rps*14 gene. The *rps*14 gene was found as a conserved gene in many species in comparison with *rbc*L, ITS, *mat*K and AFLP (WALI *et al.*, 2013). Furthermore, by comparing the genetic diversity values based on the *rps*14 gene analyzed by WALI *et al*. (2013), among different genera of *Citrus* species, the values were found to be 0.02. While this value for different selected species of *Mentha* investigated by JABEEN *et al*. (2012), was found to be 0.04. However, the present dendrogram for selected species of Lamiaceae expressed value of 0.0100, expressing high genetic similarity than the earlier reports on the *rps*14 gene. Our finding are in agreement with the previous research of SCHEEN *et al.* (2009), which were based on three plastid markers in 159 species from 50 genera, that showed *Origanum vulgare, Salvia moorcroftiana* and *Nepeta juncea* in clade I while *Clinopodium debile, Rydingia limbate, Teucrium stocksianum and Zataria multiflora* in clade II. Thus, the phylogenetic study of family Lamiaceae by using the *rps*14 gene proved increased phylogenetic resolution and stronger support for most of the clades within the family.

D Structural Analysis and Protein Validation and Structure Analysis

Nucleotide sequences were decoded into amino acid sequences via online DNA to protein conversion tool ExPasy (http://web.expasy.org/translate/). I-TASSER was used to generate five full-length atomic models. The confidence score (C-score) of the I-TASSER structural models were used to predict their functions. The 3-D protein models of *rps*14 with the highest C-score values are shown in (Figure 2).


A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P


Q
R
S
T
U
V
W
X
Y
Z
A1
B1
C1
D1
E1
F1

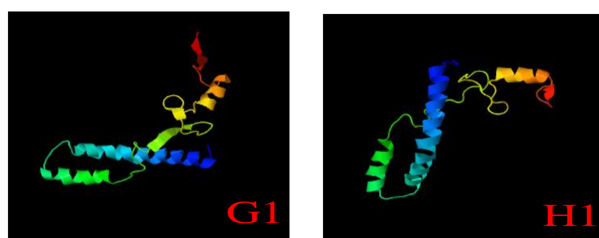


Figure 2. 3-D protein models build by I-TASSER for *rps*14 having highest C score for different species of family *Lamiaceae* A: *Ajuga parviflora* B: *Anisomeles indica* C: *Clinopodium debile* D: *Clinopodium umbrosum* E: *Colebrookea oppositifolia* F: *Isodon rugosus* G: *Lamium album* H: *Lamium amplexicaule* I: *Leucas cephalotes* J: *Marrubium vulgare* K: *Mentha arvensis* L: *Mentha longifolia* M: *Mentha spicata* N: *Micromeria biflora* O: *Nepeta erecta* P: *Nepeta juncea* Q: *Nepeta laevigata* R: *Ocimum americanum* S: *Ocimum basilicum* T: *Origanum vulgare* U: *Phlomidoschema parviflorum* V: *Phlomis cashmeriana* W: *Plectranthus barbatus* X: *Prunella vulgaris* Y: *Rydingia limbata* Z: *Salvia coccinea* A1: *Salvia moorcroftiana* B1: *Salvia plebeia* C1: *Teucrium stocksianum* D1: *Thymus linearis* E1: *Vitex agnus-castus var. pseudo-negundo* F1: *Vitex negundo* G1: *Zataria multiflora* H1: *Ziziphora tenuior*

The fidelity and stereological quality of predicted proteins from 34 species of Lamiaceae under consideration were analyzed while using the PROCHECK package (http://www.ebi.ac.uk/thorton/software.html) via residue-by-residue and overall geometry of protein structures. Analysis of the Ramachandran plots of selected species showed that on average protein model for all the species predicted by I-TASSER were reliable and can be used for further analysis because they have ≥ 81 % to ≥ 92 % residues that lie in the most allowed region and ≤ 1 % to ≤ 7 % residues that lie in the disallowed region. The protein models for *rps*14 of *Ajuga parviflora, Anisomeles indica, Clinopodium debile, Clinopodium umbrosum, Colebrookea oppositifolia, Isodon rugosus, Lamium album, Mentha longifolia, Mentha spicata, Micromeria biflora, Nepeta juncea, Nepeta laevigata, Origanum vulgare, Phlomis cashmeriana, Prunella vulgaris, Salvia moorcroftiana, Teucrium stocksianum, Vitex agnus-castus var. pseudo negundo, Zataria multiflors and Ziziphora tenuior* were found to be best quality structural models, as they have ≤ 2 % residues that lie in the disallowed region. A detailed score of Ramachandran plots for 34 sequences are given in Table 6.

The PROCHECK program provides a means of validation, the geometry and restraint violations of an entity for protein structures. In the past few decades, PROCHECK has remarkable progress in the methodology of solving protein structures. The number of spectra for the experimental techniques protocols and computer programs for the structure determinations are now available. Such techniques are helpful for the improvement and correcting diagnosed problems and validate the structure of proteins (EKINS *et al.*, 2016). Thus, the phylogenetic study of family Lamiaceae by using the *rps*14 gene proved increased phylogenetic resolution and stronger support for most of the clades within the family. Evaluation and assessment results from Ramachandran plot analysis expressed that *rps*14 structures of all the selected species have good

quality structural models. Most of the species in the present study showed the best quality protein structures predicated by I-TASSER.

Table 6. Ramachandran score of rps14 protein models for selected 34 amino acid sequences predicted by PROCHECK

S. No.	Plant Name	Most allowed region in %	Allowed region in %	Disallowed region in %
1.	*Ajuga parviflora*	85.7	13.3	1.0
2.	*Anisomeles indica*	88.8	11.2	0.0
3.	*Clinopodium debile*	90.8	9.2	0.0
4.	*Clinopodium umbrosum*	91.8	6.1	2.0
5.	*Colebrookea oppositifolia*	91.8	7.1	1.0
6.	*Isodon rugosus*	92.9	5.1	1.0
7.	*Lamium album*	92.9	6.1	1.0
8.	*Lamium amplexicaule*	89.8	6.1	4.1
9.	*Leucas cephalotes*	84.7	11.2	4.1
10.	*Marrubium vulgare*	84.7	11.2	4.1
11.	*Mentha arvensis*	87.8	9.2	3.1
12.	*Mentha longifolia*	88.8	10.2	1.0
13.	*Mentha spicata*	85.7	12.2	2.0
14.	*Micromeria biflora*	87.8	10.2	2.0
15.	*Nepeta erecta*	91.8	4.1	4.1
16.	*Nepeta juncea*	85.7	13.3	1.0
17.	*Nepeta laevigata*	86.7	11.2	2.0
18.	*Ocimum americanum*	84.7	11.2	4.1
19.	*Ocimum basilicum*	82.2	13.8	3.7
20.	*Origanum vulgare*	88.8	10.2	1.0
21.	*Phlomidoschema parviflorum*	85.7	10.2	4.1
22.	*Phlomis cashmeriana*	90.8	8.2	1.0
23.	*Plectranthus barbatus*	79.6	13.3	7.1
24.	*Prunella vulgaris*	87.8	11.2	1.0
25.	*Rydingia limbata*	88.8	8.2	3.1
26.	*Salvia coccinea*	81.6	14.3	4.1
27.	*Salvia moorcroftiana*	83.7	14.3	2.0
28.	*Salvia plebeian*	81.6	12.2	6.1
29.	*Teucrium stocksianum*	87.8	11.2	1.0
30.	*Thymus linearis*	81.6	12.2	6.1
31.	*Vitex agnus-castus var. pseudo-negundo*	82.7	15.3	2.0
32.	*Vitex negundo*	92.9	3.1	4.1
33.	*Zataria multiflora*	86.7	11.2	2.0
34.	*Ziziphora tenuior*	87.8	11.2	1.0

CONCLUSION

This study provides important insights into the genetic diversity of taxa in the family Lamiaceae and divides studied species into different clusters, providing boundaries for taxonomic classification. In order to obtain more reliable results, many species in different regions should be sequenced and identified using different markers. Bioinformatics analyses in this study can be used to understand the structure and functions of proteins using online programs. Using molecular approaches like *rps*14 marker can be used effectively to study genetic divergence among species of family Lamiaceae and may be used to distinguish between species of the same family.

ACKNOWLEDGMENT

The authors are grateful to Dr. Awais from the Beijing Genomic Institute (BGI), Shenzhen, China for the sequencing of samples. We are thankful to Herbarium of Pakistan, Quaid-i-Azam University, Islamabad (ISL) for identification of plant species used in this study.

Received, May 12th, 2019
Accepted March 18th, 2020

REFERENCES

AKINRO, L. A., A. I. .,ADESOYE, T. R., FASOLA (2019): Genetic diversity in *Cola acuminata* and *Cola nitida* using RAPD primers. Genetika*, 51*(1): 227-236.

ALI, S. ., M., QAISER (1993-2015): Flora of Pakistan:194-220.

BAYAT, M., R.,AMIRNIA, H.,ÖZKAN, A.,GEDIK, D.,ATES, B., TANYULAC, M., RAHIMI (2018): Diversity and phylogeny of saffron (*Crocus sativus* L.) accessions based on iPBS markers. Genetika*, 50*(1): 33-44.

CANTINO, P. D. (1992): Evidence for a polyphyletic origin of the Labiatae. Annals of the Missouri Botanical Garden*, 1*: 361-379.

CURTO, M. A., P., PUPPO, D., FERREIRA, M., NOGUEIRA, H., MEIMBERG (2012): Development of phylogenetic markers from single-copy nuclear genes for multi locus, species level analyses in the mint family (Lamiaceae). Molecular phylogenetics and evolution, *63*(3): 758-767.

DE MATTIA, F., I., BRUNI, A., GALIMBERTI, F., CATTANEO, M., CASIRAGHI, M., LABRA (2011): A comparative study of different DNA barcoding markers for the identification of some members of Lamiacaea. Food Research International*, 44*(3): 693-702.

EFTEKHARIAN, R., M., SHEIDAI, F., ATTAR, Z., NOORMOHAMMADI, S., GHASEMZADEH-BARAKI (2018): Molecular and morphological study of the genus *Senecio* L.(Asteraceae: Senecioneae) in Iran. Genetika*, 50*(2): 403-419.

EKINS, S., J., LIEBLER, B.J., NEVES, W.G., LEWIS, M., COFFEE, R., BIENSTOCK, C., SOUTHAN, C.H. ANDRADE (2016): Illustrating and homology modeling the proteins of the Zika virus. F1000 Research*, 5*.

ERIĆ, P., M., JELIĆ, M., SAVIĆ-VESELINOVIĆ, B., KENIG, M., ANĐELKOVIĆ, M., STAMENKOVIĆ-RADAK (2019): Nucleotide diversity of Cyt b gene in drosophila subobscura Collin. Genetika*, 51*(1): 213-226.

FAKIR, H., A.A., US, M., SAGDIC, F., TORNUK (2015): Essential Oil Composition, Antimicrobial and Bioactive Properties of *Origanuım hypericifolium*, An Endemic Plant Species grown in Turkey. Research Journal of Biotechnology*, 10*: 11.

FEDERICI, S., A., GALIMBERTI, F., BARTOLUCCI, I., BRUNI, F., DE MATTIA, P., CORTIS, M., LABRA (2013): DNA barcoding to analyse taxonomically complex groups in plants: the case of *Thymus* (Lamiaceae). Botanical Journal of the Linnean Society, *171*(4): 687-699.

GALOVIĆ, V., M., MARKOVIĆ, P., PAP, M., MULETT, M., RAKIĆ, A.,VASILJEVIĆ, S., PEKEČ (2018): Molecular taxonomy and phylogenetics of Daedaleopsis confragosa (Bolt.: Fr.) J. Schröt. From wild cherry in Serbia. Genetika, *50*(2): 519-532.

GRITSENKO, D., A., POZHARSKY, N., DERYABINA, A., KASSENOVA, N., GALIAKPAROV (2019): Genetic analysis of hemagglutinin proteins of H3 and H1 subtypes in Kazakhstan.

HASAN, A., M., KHAN, M., AHMAD (2007): Authenticity of folk medicinal plants of Pakistan. Taxonomic and Chemical Methods, *1*: 3-4.

HEBERT, P.D., A., CYWINSKA, S.L., BALL (2003): Biological identifications through DNA barcodes. Proceedings of the Royal Society of London B: Biological Sciences, *270*(1512): 313-321.

HOLLINGSWORTH, P.M., S.W.,GRAHAM, D.P., LITTLE (2011): Choosing and using a plant DNA barcode. PloS one, *6*(5): 1-13.

KRESS, W.J., D.L., ERICKSON (2008): DNA barcodes: genes, genomics, and bioinformatics. Proceedings of the National Academy of Sciences, *105*(8): 2761-2762.

KUMAR, S., G., STECHER, K., TAMURA (2016): MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Molecular biology and evolution, *33*(7): 1870-1874.

NAGHIBI, F., M., MOSADDEGH, M., MOHAMMADI MOTAMED, A., GHORBANI (2010): Labiatae family in folk medicine in Iran: from ethnobotany to pharmacology. Iranian Journal of Pharmaceutical Research, *1*: 63-79.

PATI, K., D.A., MUNSHI, M., VERMA, K.T., BEHERA, L., ARYA (2019): Genetic diversity of cucumber genotypes revealed by SSR markers and agronomic traits. Genetika, *51*(2): 405-416.

PERVEEN, A., M., QAISER (2004): Pollen morphology of the family Labiatae from Pakistan. Pakistan Journal of Botany, *35*(5; SPI): 671-694.

SAVIĆ, A., M., BRDAR-JOKANOVIĆ, M., DIMITRIJEVIĆ, S., PETROVIĆ, M., ZDRAVKOVIĆ, D., ŽIVANOV, M., VASIĆ (2019): Genetic diversity of common bean (*Phaseolus vulgaris* L.) breeding collection in Serbia. Genetika, *51*(1): 1-15.

SEVĠNDĠK, E., K., YALÇIN (2018): Phylogenetic analysis of some *Citrus* L.(Rutaceae) taxa in Turkey based on chloroplast (cPDNA) trnL intron and trnL-F DNA sequences. Genetika, *50*(3): 1035-1044.

SEVINDIK, E. (2019): Comparative and phylogenetic analysis of RuBisCO large subunit (rbcL) proteins in some *Sideritis* L.(Lamiaceae) species: A bioinformatic approach. Genetika, *51*(1): 69-80.

TALEBI, S. M. (2016): Infra-specific morphological diversity in *Phlomis olivieri* (Labiatae). Journal of Bio-Science, *22*: 59-67.

TAMURA, K., M., NEI (1993): Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular biology and evolution, *10*(3): 512-526.

TELETCHEA, F., J., BERNILLON, M., DUFFRAISSE, V., LAUDET, C., HÄNNI (2008): Molecular identification of vertebrate species by oligonucleotide microarray in food and forensic samples. Journal of Applied Ecology, *45*(3): 967-975.

WALI, S., F., MUNIR, T., MAHMOOD (2013): Phylogenetic studies of selected *Citrus* species based on chloroplast gene, rps14. International Journal of Agriculture & Biology, *15*: 357-361.

WANG, H.T., X., YU, Y., LIU, C.-Y., LIANG, W.-L., LI (2013): Analysis of genetic variability and relationships among *Mentha* L. using the limonene synthase gene, LS. Gene, *524*(2): 246-252.

WANG, W., K.,WANG, X.,CHEN, S., PRODANOVIĆ, X.,LI, X.,YE, Y., YAN (2018): Molecular characterization and phylogenetic analysis of one Omega-gliadin gene from *Aegilops speltoidesl.* Genetika, *50*(2): 503-517.

ZAHRA, N.B., Z.K., SHINWARI, M., QAISER (2016): Dna barcoding: a tool for standardization of Herbal Medicinal Products (HMPS) of Lamiaceae From Pakistan. Pakistan Journal of Botany, *48*(5): 2167-2174.

ZHANG, Y. (2008): I-TASSER server for protein 3D structure prediction. BMC bioinformatics*, 9*(1), 40

FILOGENIJA I DIVERZITET FAMILIJE LAMIACEAE U PAKISTANU NA OSNOVU GENA *RPS14*

Asma AYAZ[1,2*], Wajid ZAMAN[2], Saddam SAQIB[2], Fazal ULLAH[2], Tariq MAHMOOD[2*]

[1]Državna ključna laboratorija za biokatalizu i inženjerstvo enzima, Škola nauka o životu, Univerzitet Hubei, 434200 Wuhan, China

[2]Departman biljnih nauka, Quaid- i- Azam Univerzitet Islamabad, 45320, Pakistan

Izvod

Ovaj rad je sproveden u skladu sa molekularnim podacima i filogenetskim odnosom između odabranih vrsta *Lamiaceae*. Proučavali smo 34 vrste sakupljene iz različitih fitogeografskih područja Pakistana. Molekularna DNK je izolovana iz svežih uzoraka lista odabranih taksona. Gen *rps14* je amplifikovan radi sekvenciranja i dalje filogenetske analize. Sekvencionirani proizvodi su analizirani pomoću alata za bioinformatiku kao što su MEGA7, I-TASSER i SAVES. Alat za prevođenje ExPASY korišćen je za prevođenje nukleotidnih sekvenci u aminokiselinske sekvence. *Phlomis cashmeriana* sa *Origanum vulgare* i *Lamium album* sa *Lamium amplexicaul* pokazali su blisku vezu u oba filogenetska stable. Rastojanje u paru se kretalo od 0,013 do 0,19, sa ukupnom srednjim rastojanjem od 0,068. Softver I-TASSER korišćen je za predviđanje 3D proteinskih struktura praćenih stereohemijskim analizama. Na osnovu dobijenih rezultata utvrđeno je da bi se *rps14* gen mogao koristiti kao DNK barkod za identifikaciju biljnih taksona. Pored toga, potreban je I razvoj dodatnih pouzdanih molekularnih markera za rešavanje problema sistematike u porodici *Lamiaceae*.

Primljeno 12.V.2019.
Odobreno 18. III. 2020.