The Journal of Animal & Plant Sciences, 31(4): 2021, Page: 1060-1069
ISSN (print): 1018-7081; ISSN (online): 2309-8694

PHYLOGENETIC RELATIONSHIP AMONG SELECTED SPECIES OF *LAMIACEAE* INFERRED FROM CHLOROPLAST *RPS*14 GENE

A. Malik[1*], T. Shahzad[1], S. Arif[1], W. Akhtar[2] and T. Mahmood[1]

[1]Department of Plant Sciences, Quaid-i-Azam University, Islamabad-45320, Pakistan.
[2]Department of Botany, University of Azad Jammu and Kashmir, Muzaffarabad, Pakistan.
[*]Corresponding author's email: ayesha.malik7390@gmail.com.

ABSTRACT

The *Lamiaceae* is commonly known as Mint family of order Lamiales. The present study was conducted for assessment of genetic diversity, phylogeny of *Lamiaceae* based on *rps*14 gene and validation of *rps*14 protein by Ramachandran plots. The *rps*14 gene was amplified and sequence analyzed through MEGA7, I-TASSER and RAMPAGE. The phylogram revealed genetic distance (number of base substitutions per site) of 0.0100 among *Lamiaceae* species. In Pairwise distance, the range of genetic diversity was 0.013 to 0.089 with a mean of 0.060. *Vitex agnus castus* and *Vitex trifolia* were grouped in same cluster with Bootstrap value of 99 %. *Caryopteris odorata* revealed close relationship with *Volkameria inermis, Clerodendrum calamitosum* and *Clerodendrum indicum* in same cluster having BS of 58 %. *C. indicum, C. calamitosum* and *V. inermis* showed close relationship in same cluster with BS of 51%. The 3D protein model designed by I-TASSER and its validation by RAMPAGE revealed ≤ 2 % amino acid in outlier regions of *Gmelina philippensis*, *Clerodendrum. umbellatum* and *C. calamitosum*. This study concluded close genetic relationship and low genetic diversity that indicated potential of *rps*14 gene for phylogeny and diversity analysis among *Lamiaceae* species. The validation of 3D protein models by RAMPAGE showed good protein models. This study highlights the evolutionary relationship, earliest, recent evolution of *Lamiaceae* species. A very low genetic diversity indicates ex-situ conservation and good protein models in *Lamiaceae* species.

Key words: *Lamiaceae*, Iterative Threading Assembly Refinement, *rps*14 gene, Pairwise distance, Genetic diversity.

https://doi.org/10.36899/JAPS.2021.4.0303 *Published online December 18, 2020*

INTRODUCTION

The *Lamiaceae* is family of flowering plant in the order Lamiales that comprised of annual or perennial herbs and shrubs with opposite leaves. The stem is mostly square and leaves when grinded releases pleasant odors (Metcalfe and Chalk, 1950). The *Lamiaceae* included 240 genera, more than 7,000 taxa in the World and seven subfamilies namely Viticoideae, Ajugoideae, Nepetoideae, Scutellarioideae, Prostantheroideae, Symphorematoideae and Lamioideae (Harley *et al*., 2004), now new subfamilies are comprised of Peronematoideae, Tectonoideae, Cymarioideae, Callicarpoideae and Premnoideae (Li *et al*., 2016; Li and Olmstead, 2017). The *Lamiaceae* is found in various altitudes and habitats varying from North Eastern Asia to Hawaii, North Pole areas to the Himalayas, America, Africa and Australia (Erdem *et al*., 2017). The *Lamiaceae* is economically important for various timber trees like *Tectona*, many species are medicinal and utilized as culinary herbs. The *Lamiaceae* species are also used for horticultural purposes, in perfumery and nectar producing species that produce high-quality honey (Harley, 2012).

The molecular phylogenetic studies are conducted to determine the relationship among organisms or genes by comparing the DNA or protein sequences (Patwardhan *et al.*, 2014). The chloroplast genome (*cp*) comprised of 120-130 genes with conserved gene content and gene order (Lei *et al.,* 2016). The *cp* genome experienced selection pressure during course of evolution and recent phylogenetic study indicate numerous positive selection genes such as *psaA*, *atpA, atpB, rbcL*, *rpI*20 and *ndhI* genes in *Urophysa* (Xie *et al*., 2018).The *cp* genome has uniparental inheritance, moderate evolutionary rate and conserved structure that makes *cp* genome applicable for DNA barcoding and phylogenetic studies (Luo *et al*., 2016: Li *et al.,* 2015: Dong *et al.,* 2018; Yang *et al*., 2018).

A wide range of molecular markers have been utilized to study phylogeny of *Lamiaceae* such as *matK*, *ndhF*, *rbcL*, *rps*16 and *trnL-F* (Li *et al*., 2016) ITS and *ndhF* (Steane *et al*., 2004), *trnL* intron, *trnL- trnF* and *rps*16 (Paton *et al*., 2004), on the basis of four cpDNA regions *ycf1*, *ycf1- rps*15 spacer, *trnL-F* and *rpI*32- *trnL* (Drew and Sytsma, 2012). The plastid genes *psaA* and *psaB* in rice encoded the two apoproteins of P700 chlorophyll *a* protein complex of photosystem I reaction center and *rps*14 gene that encodes the ribosomal protein S14 are organized into a transcription unit (Chen *et al.,* 1992). The *rps*14 gene has been reported from the *cp* genome of *N. tabacum* (Wakasugi *et al*., 1998). In the past, *rps*14 gene has been used for the assessment of phylogenetic relationship among *Plantago* species,

Mentha, Citrus species and date palm varieties (Saeed *et al*., 2011; Jabeen *et al*., 2012; Wali *et al*., 2013; Akhtar *et al*., 2014). In the past, phylogenetic relationship among various genera of *Lamiaceae* such as *Clerodendrum, Vitex, Tectona*, *Gmelina, Callicarpa* and *Caryopteris* had not been focused based on *rps*14 gene. Therefore, the hypothesis of present study is an assessment of genetic diversity and phylogenetic relationship among selected *Lamiaceae* species based on *rps*14 gene and structural validation of *rps*14 protein via Ramachandran plots.

MATERIALS AND METHODS

Plant Collection: The leaves of plant species were collected from Islamabad, Rawalpindi, Lahore and Faisalabad regions of Pakistan and stored at 4 ºC for future purpose (Table 1). The correct taxonomic placement of plant species was done by following the International Plant name Index the Royal Botanic Gardens Kew, UK (http://plantsoftheworldonline.org/).

Table 1: List of selected species of family *Lamiaceae*.

S/N	Plant Species	Cities	Nature of Plant Species
1	*Callicarpa macrophylla*	Faisalabad	Shrub
2	*Tectona grandis*	Lahore	Tree
3	*Caryopteris odorata*	Islamabad	Shrub erect or suberect
4	*Gmelina philippensis*	Islamabad	Straggling or scandent spinose shrub
5	*Clerodendrum umbellatum*	Islamabad	Shrub, climbing to suberect
6	*Vitex trifolia*	Islamabad	Shrub or small tree
7	*Clerodendrum calamitosum*	Islamabad	Small to medium size woody shrub
8	*Clerodendrum indicum*	Islamabad	Shrub
9	*Volkameria inermis*	Islamabad	Erect to scandent or trailing ever green shrub
10	*Vitex agnus castus*	Islamabad	Erect shrub

DNA isolation and Primer designing: The genomic DNA was isolated from stored leaves of *Lamiaceae* by CTAB method (Richards *et al,.* 1997) with little modifications. The extraction buffer 2 % CTAB (2.4 g of Tris HCl, 1.16 g of EDTA, 16.3 g of Nacl and 4 g of CTAB) was prepared in 200 ml distilled water. The plant material (0.4 g) was crushed in 1500 µl pre-heated CTAB at 65 °C along with 15 µl of mercaptoethanol. The homogenized mixture was transferred to eppendorf tubes, incubated at 65 °C for 1 hour and centrifuged at 12,000 rpm for 15 minutes. The supernatant was transferred to a new eppendorf and washed 3-4 times with equal volume of Chloroform isoamyl alcohol (24:1) and centrifuged at 12,000 rpm for 15 minutes. The pre-chilled isopropanol was added to clear supernatant and tubes were left overnight at -20 °C, then sample was centrifuged and supernatant was discarded. The pellet was washed with cold 70 % ethanol and air dried at room temperature for 1 hour. The air-dried pellet was dissolved in TE (Tris EDTA) buffer containing RNase (7 µl of RNase added in 1mL buffer) and incubated at 37 ºC for 10 minutes. The obtained DNA was stored at -20 °C. The DNA isolation and PCR amplification were performed in the Plant Molecular biology and Biochemistry Lab situated in Quaid-i-Azam University Islamabad. The primers were designed for *rps*14 gene through online primer 3 (version 4.10) (http://bioinfo.ut.ee/primer3/) based on *cp* genome of *N. tabacum* available in NCBI Genbank (https://www.ncbi.nlm.nih.gov). The primer sequence is mentioned below;

*rps*14 F: 5′-ATGGCAAGGAAAAGTTTGATTC-3'
*rps*14 R: 5′- TTACCAACTTGATCTTGTTGCTCCT- 3'

PCR amplification conditions and sequencing: The following conditions were used for the amplification of *rps*14 gene in PCR Multi Gene thermal cycler (Labnet). The initial denaturation at 94 °C for 5 minutes was followed by 35 cycles of denaturation at 94 °C for 30 seconds, annealing in the range of 45-51 °C for 1 minute, extension at 72 °C for 1 minute and final extension at 72 °C for 20 minutes. The amplified PCR product was purified by Gene JET PCR Purification Kit and sequenced from the Beijing Genomic Institute, Shenzhen, China.

Sequence analysis: The *rps*14 gene sequences of *Lamiaceae* species were uploaded and BLASTn (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was performed one by one in query form for comparison with already reported sequences in Genbank. The nucleotide sequences were cleaned and aligned with the help of online server JustBio (https://www.justbio.com/hosted-tools.html). The nucleotide sequences were converted into amino acid sequences through EXPASY- Translate tool (https://Web.expasy.org/ translate/). The sequences data of *rps*14 gene of all ten *Lamiaceae* species were submitted to the Genbank for accession numbers (https://www.ncbi.nlm.nih.gov/home/submit/) (Table 2).

Table 2: Accession numbers and Nucleotide composition of *rps*14 gene from selected species of family *Lamiaceae.*

S/N	Plant species	Accession Number	T(U)	C	A	G	Total
1	*Vitex agnus castus*	MH382179	23.8	18.2	34.0	24.1	303
2	*Callicarpa macrophylla*	MH382180	24.8	18.8	34.3	22.1	303
3	*Tectona grandis*	MH382183	24.4	18.5	34.7	22.4	303
4	*Caryopteris odorata*	MH382184	24.1	18.2	34.0	23.8	303
5	*Gmelina philippensis*	MH382186	24.4	18.2	36.0	21.5	303
6	*Clerodendrum umbellatum*	MH382188	24.1	19.5	34.0	22.4	303
7	*Vitex trifolia*	MH382190	23.8	18.2	34.0	24.1	303
8	*Clerodendrum calamitosum*	MH382191	24.4	19.1	34.3	22.1	303
9	*Clerodendrum indicum*	MH382192	24.4	19.1	34.7	21.8	303
10	*Volkameria inermis*	MH382194	24.4	19.1	34.0	22.4	303
	Average		**24.3**	**18.7**	**34.4**	**22.7**	**303**

Phylogenetic analysis: The aligned sequences were subjected to MEGA7 with Maximum Likelihood (ML) method and 1000 BS value was used for construction of phylogenetic tree (Kumar *et al.,* 2016). The DNA sequences were aligned into FASTA format and uploaded in MEGA7 with nucleotide sequences, protein coding nucleotide sequence data and standard genetic code. The gaps/missing data treatment was completely deleted for calculation of Pairwise distance, Tajima's Test of Neutrality and Substitution patterns. In MEGA7 the selection tool was used for calculation of Tajima's Test of Neutrality. The pairwise distance was calculated by Tamura-Nei model with substitution of transitions and transversions. The substitution Matrix (ML) values were determined by Tamura-Nei model and Maximum Likelihood method.

Statistics: The Tajima's D statistics was used for detection of selection using *rps*14 gene sequences in MEGA7 (Tajima, 1989). The Tajima's D can be described as if D = 0 observed data =expected = no selection, D = negative = D < 0 observed < expected = fewer haplotypes and lower average heterozygosity as compared to segregating sites showed size expansion after a bottle neck effect or selective sweep or positive selection and then expected under a standard neutral model, D=positive = D > 0 observed > expected= More average heterozygosity as compared to segregating sites showed balancing selection. The best DNA model (ML) was determined by MEGA7 (Kumar *et al.,* 2016) in which the lowest BIC (Bayesian information criterion) scores indicated the substitution pattern best. The model with best substitution pattern was then used to build the substitution pattern estimation (ML), Pairwise distance and phylogenetic tree. The statistical support of phylogenetic tree was determined by Bootstrapping method performed with 1000 bootstrap replications in MEGA7 (Kumar *et al.,* 2016).

The statistics used in the I TASSER was C-score, Z-score and TM-scores for 3D protein models. The threading alignment of 3D protein models was denoted by Z. The Z is measured as the difference between the raw and average scores in the unit of standard deviation. The Normalized Z-score > 1 indicates good alignment. The C-score (Confidence score) is measured as significance of threading template alignments and the convergence parameters of the structure assembly simulation. The value of C-score was in the range of -5 to 2 that indicates higher confidence and more reliable prediction. The TM-score (Template based score) measuring the structural similarity between two structures. A TM-score has value ranges from (0-1) and TM-score > 0.5 indicates a model of similarity of protein to the template (https:// Zhanglab.ccmb.med. umich.edu/I-TASSER)**.** The 3D models was validated by RAMPAGE that shows the scores of allowed region, favored region and outlier region in which the < 2 % amino acid in the outlier region of different *Lamiaceae* species showed good protein models (http://mordred.bioc.cam.ac.uk/ ~rapper/rampage.php).

Protein structure prediction and validation: I-TASSER is online software used for prediction of 3D protein structures (https:// Zhanglab.ccmb.med.umich. edu/I-TASSER)**.** The validation of 3D Protein models was done by RAMPAGE (http://mordred.bioc.cam.ac.uk /~rapper/ rampage.php). The Ramachandran plots formed by uploading Pdb files of 3D protein models on RAMPAGE. The Ramachandran plots were used for structural validation of 3D protein structures and to visualize the conformation of dihedral angle φ and ψ of amino acid residues present in proteins.

RESULTS

Amplification of *rps*14 gene: The target gene *rps*14 was amplified by specific forward and reverse primers. The band was observed under UV light in Dolphin-Doc Plus gel documentation system. The amplification process gave clear bands of 303 bp for *Lamiaceae* species (Figure 1).

Sequence analysis: The BLASTn comparisons of chloroplast *rps*14 gene sequence with already reported sequence revealed 99 % sequence identity with chloroplast

genome of *Vitex negundo* (Accession number # MF678773).

Estimation of nucleotide composition: The average of nucleotide number was T (U) = 24.3, C= 18.7, A = 34.4 and G = 22.7. The A has highest average nucleotide number of 34.4 and C has lowest average nucleotide number of 18.7. The total average number of all nucleotides was 303 (Table 2).

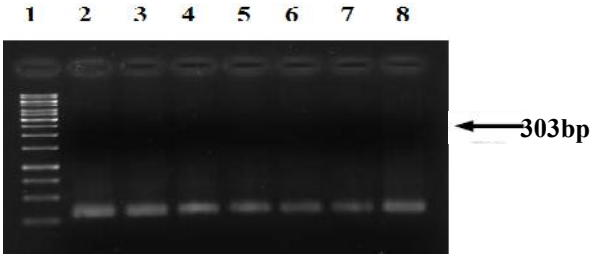

Figure 1: Amplified PCR product of some selected species of family *Lamiaceae* based on *rps*14 gene. 1: Marker 1000 bp Ladder 2: *Clerodendrum indicum* 3: *Gmelina philippensis* 4: *Caryopteris odorata* 5: *Tectona grandis* 6: *Vitex agnus castus* 7: *Vitex trifolia* 8: *Callicarpa macrophylla.*

Tajima's neutrality test: In *Lamiaceae* ten numbers of sequences gave 41 segregation sites *(S)* and nucleotide diversity *(π)* of 0.056912. The low nucleotide diversity of 0.056912 revealed close genetic relationship among studied *Lamiaceae* species. The positive value of Tajima's D statistics was 0.922432 which showed more haplotypes that increase the chance of average heterozygosity than segregating sites and balancing selection would operate (Tajima, 1989) (Table 3).

Estimation of substitution matrix (ML): The lowest BIC score of 1790.792 indicated TN93 model that was Tamura-Nei (1993) model so the substitutions pattern and rates were determined through Tamura-Nei (1993) model. The relative values of instantaneous (r) should be included when calculating them, for easiness the sum of (r) values is made equal to 100. The values of transitional substitution were higher than transversionsal substitution. The nucleotide frequencies for *Lamiaceae* were A = 34.39 %, T/U = 24.26 %, C = 18.68 % and G = 22.67 %. The calculation of ML values for *Lamiaceae* were done by computing the tree topology. The ML Log for *Lamiaceae* was -863.363 (Table 4). The statistical result of lowest BIC score showed that in Tamura- Nei model there were different nucleotide frequencies. So, substitution matrix (ML) values estimated by Tamura-Nei model also have different nucleotide frequencies (Table 5).

Table 3: Tajima's neutrality test values based on *rps*14 gene of *Lamiaceae.*

No. of sequences ''m''	Number of segregating sites ''S''	*Ps*	Θ	Nucleotide diversity ''*π*''	Tajima's test Statistics "D"
10	41	0.135314	0.047831	0.056912	0.922432

Table 4: Maximum likelihood values of transitional (bold) and transversionsal substitutions (italics) of nucleotides of *rps*14 gene for *Lamiaceae.*

	A	T/U	C	G
A	-	*4.02*	*3.09*	**18.28**
T/U	*5.69*	-	**9.09**	*3.75*
C	*5.69*	**11.80**	-	*3.75*
G	**27.72**	*4.02*	*3.09*	-

The nucleotide frequencies for *Lamiaceae* were **A** = 34.39 %, **T/U** = 24.26 %, **C** = 18.68 % and **G** = 22.67 %. The calculation of ML values for *Lamiaceae* were done by computing the tree topology. The ML Log for *Lamiaceae* was -863.363.

Calculation of Pairwise distance: In *Lamiaceae* the value of genetic diversity lies in the range of 0.013 to 0.089 with overall mean distance of 0.060 for *rps*14 gene. The low mean distance value of 0.060 show close genetic relationship among *Lamiaceae* species and low genetic diversity among them on the basis of *rps*14 gene (Table 6).

Phylogenetic tree: The phylogenetic tree was built through MEGA7 based on *rps*14 gene sequences (Kumar *et al.,* 2016). The dendrogram revealed little genetic distance of 0.0100 that showed close genetic relationship among *Lamiaceae* species. The phylogram was divided in two clusters and three groups denoted by cluster I, cluster II, group I, group II and group III. The cluster I consisted of *C. umbellatum*, *C. odorata, C. indicum*, *C. calamitosum* and *Volkameria inermis*. In cluster I, *C. odorata* showed close phylogenetic relationship with *C. indicum, C. calamitosum* and *Volkameria inermis* with BS value of 58%. The *C. indicum*, *C. calamitosum* and *Volkameria inermis* also showed close phylogenetic association among them with BS value of 51 %. The *C. umbellatum* showed recent evolution with small branch length of 0.021. While *C. odorata* were earliest evolved member with more branch length of 0.039. The *C. odorata* also depicted close relationship with *C. umbellatum* in the same cluster. In group I, *C. calamitosum* and *Volkameria inermis* also revealed close phylogenetic relationship between them. The Cluster II included *V. agnus castus*. *V. trifolia, C. macrophylla, T. grandis* and *G. philippensis* and divided into group II and group III. In group II, *C. macrophylla* and *G.philippensis* revealed close relationship and

grouped together in the cluster II. The *G. philippensis* was recently evolved member with branch length of 0.017. While *C. macrophylla* were earliest evolved with branch length of 0.021. In group III, *V. agnus castus* and *V. trifolia* depicted close phylogenetic relationship with BS value of 99 %. The *C. macrophylla* and *T. grandis* also showed close relationship among them in the cluster II. The *C. macrophylla* showed earliest evolution with branch length of 0.021.While *T. grandis* was recently evolved member with branch length of 0.010. The *G. philippensis* and *T. grandis* revealed close relationship and were grouped together in the cluster II. The *V. trifolia* and *C. macrophylla* also showed close phylogenetic association among them. The *V. agnus castus* and *T. grandis* showed close relationship and were grouped together in the cluster II. The *V. trifolia* showed phylogenetic association with *G. philippensis* in the same cluster based on *rps*14 gene. The *V. agnus castus* and *G. philippensis* revealed close relationship among them. The *V. agnus castus* showed recent evolution with branch length of 0.003 while *G. philippensis* showed the earliest evolution with branch length of 0.017. In Cluster II, *V. trifolia* and *T. grandis* revealed close relationship among them (Figure 2).

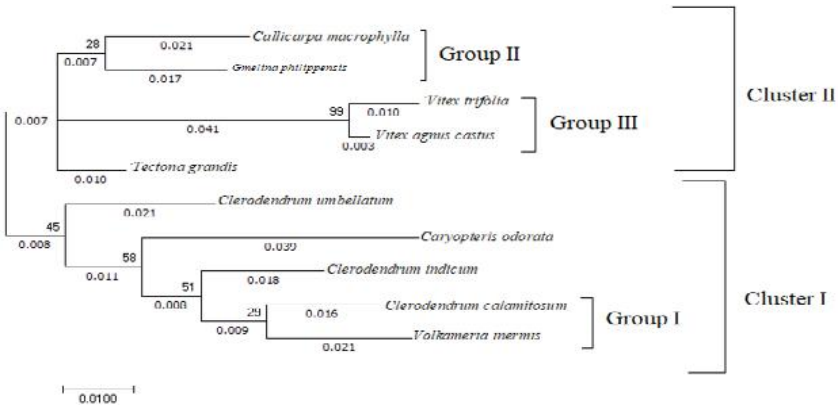

Figure 2: Phylogenetic tree among selected species of *Lamiaceae* based on *rps*14 gene.

Analysis of 3D proteins structures: The 3D protein models were constructed through I-TASSER and top five models of 3D protein structures for each species were predicted. The confidence of each model was calculated on the basis of confidence score (C-score). The value of C-score was in the range of -5 to 2. The 3D protein model having higher C-score was indication of higher confidence and more reliable prediction. The *rps*14 protein contains 100 amino acid residues and 3D protein models consists of the alpha helices, beta strands and coils the number of alpha helices are more than beta strands and coils. The predicted solvent accessibility of 3D protein model is from hydrophilic to hydrophobic. The C-score, Z-score and TM-score is good in all 3D protein model of *Lamiaceae* (Table 7).

Protein structure validation: The Ramachandran plots consisted of quadrant of different colors. The dark blue and dark orange areas indicated the favored region. Whereas the light blue and light orange were allowed region. The unshaded areas showed outlier region (Figure 3). In *Lamiaceae* ≥ 81 to ≤ 88 residues in favored region, ≥ 8 to ≤ 13 in allowed region and ≤ 2 to ≥ 3 in outlier region. The protein models of *Gmelina philippensis, Clerodendrum umbellatum* and *Clerodendrum calamitosum* were considered to be good protein models as they had ≤ 2 % amino acid residues in the outlier region (Table 8).

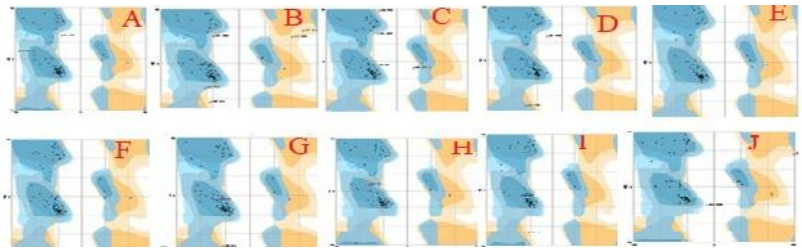

Figure 3: Ramachandran plots of ten selected species of family *Lamiaceae*. A: *Vitex agnus castus* B: *Callicarpa macrophylla* C: *Tectona grandis* D: *Caryopteris odorata* E: *Gmelina philippensis* F: *Clerodendrum umbellatum* G: *Vitex trifolia* H: *Clerodendrum calamitosum* I: *Clerodendrum indicum* J: *Volkameria inermis.*

Table 5: Maximum Likelihood fits of 24 different nucleotide substitution models.

Models	Parameter	BIC	AICc	*inL*	(+I)	(+G)	R	*f*(A)	*f*(T)	*f*(C)	*f*(G)	*r*(AT)	*r*(AC)	*r*(AG)	*r*(TA)	*r*(TC)	*r*(TG)	*r*(CA)	*r*(CT)	*r*(CG)	*r*(GA)	*r*(GT)	*r*(GC)
TN93	22	**1790.792**	1664.756	**-863.363**	n/a	n/a	2.09	**0.344**	**0.243**	**0.187**	**0.277**	0.040	0.031	0.183	0.057	0.091	0.038	0.057	0.118	0.038	0.277	0.040	0.031
K2+G+I	20	1791.518	1671.471	-815.596	0.063	0.29	2.92	0.250	0.250	0.250	0.250	0.032	0.032	0.186	0.032	0.186	0.032	0.032	0.186	0.032	0.186	0.032	0.032
T92+G	20	1792.723	1672.676	-816.198	n/a	0.05	2.47	0.293	0.293	0.207	0.207	0.041	0.029	0.148	0.041	0.148	0.029	0.041	0.211	0.029	0.211	0.041	0.029
K2+G	19	1792.750	1678.62	-820.220	n/a	0.05	2.51	0.250	0.250	0.250	0.250	0.036	0.036	0.179	0.036	0.179	0.036	0.036	0.179	0.036	0.179	0.036	0.036
HKY+G+I	23	1797.672	1658.664	-806.648	0.61	0.27	2.73	0.344	0.243	0.187	0.227	0.033	0.025	0.166	0.046	0.136	0.046	0.031	0.177	0.031	0.251	0.033	0.025
HKY+G	22	1799.067	1667.045	-811.354	n/a	0.05	2.40	0.344	0.243	0.187	0.227	0.036	0.028	0.160	0.051	0.132	0.033	0.051	0.171	0.033	0.242	0.036	0.028
TN93+G+I	24	1804.138	1660.145	-805.873	0.058	0.25	2.83	0.344	0.243	0.187	0.227	0.033	0.025	0.205	0.046	0.094	0.030	0.046	0.122	0.030	0.311	0.033	0.025
TN93+G	23	1804.236	1666.228	-809.931	n/a	0.05	2.53	0.344	0.243	0.187	0.227	0.035	0.027	0.202	0.050	0.087	0.050	0.033	0.112	0.033	0.307	0.305	0.027
JC+G	18	1826.785	1718.719	-841.246	n/a	0.05	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
GTR+G+I	27	1828.103	1666.166	-805.831	0.58	0.25	2.82	0.344	0.243	0.187	0.227	0.033	0.028	0.204	0.047	0.095	0.027	0.051	0.123	0.029	0.310	0.029	0.024
GTR+G	26	1828.151	1672.194	-809.863	n/a	0.05	2.53	0.344	0.243	0.187	0.277	0.037	0.025	0.202	0.052	0.087	0.029	0.047	0.113	0.039	0.306	0.031	0.032
JC+G+I	19	1828.463	1714.405	-853.014	0.060	0.29	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
T92+I	20	1855.601	1735.554	-847.637	0.43	n/a	2.15	0.293	0.293	0.227	0.227	0.046	0.032	0.143	0.046	0.143	0.032	0.046	0.202	0.032	0.202	0.046	0.032
K2+I	20	1858.367	1744.309	-853.028	0.43	n/a	2.15	0.250	0.250	0.250	0.250	0.040	0.040	0.171	0.040	0.171	0.040	0.040	0.171	0.040	0.171	0.040	0.040
HKY+I	19	1861.651	1729.628	-842.646	0.43	n/a	2.14	0.344	0.293	0.187	0.227	0.039	0.030	0.154	0.055	0.127	0.036	0.055	0.165	0.036	0.234	0.039	0.030
TN93+I	22	1865.632	1727.629	-839.625	0.43	n/a	2.14	0.344	0.243	0.187	0.227	0.039	0.041	0.170	0.059	0.105	0.041	0.059	0.137	0.041	0.244	0.041	0.032
JC+I	23	1866.390	1728.382	-841.007	0.43	n/a	2.21	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
GTR+I	26	1890.274	1734.365	-840.925	0.43	n/a	2.21	0.344	0.243	0.187	0.227	0.035	0.030	0.190	0.050	0.086	0.055	0.037	0.112	0.042	0.289	0.040	0.035
T92	19	1894.234	1734.34	-849.732	0.43	n/a	2.21	0.29	0.29	0.210	0.210	0.048	0.035	0.140	0.048	0.140	0.035	0.048	0.193	0.035	0.193	0.048	0.035

			7					0	0													
K2	18	1891.812	1777.755	-869.751	n/a	n/a	2.09	0.250	0.250	0.250	0.250	0.040	0.040	0.169	0.040	0.169	0.040	0.040	0.169	0.040	0.169	0.040	0.040
HKY	21	1893.928	1785.862	-874.817	n/a	n/a	2.09	0.341	0.239	0.183	0.237	0.039	0.030	0.153	0.056	0.126	0.037	0.056	0.164	0.037	0.232	0.039	0.030
T92+G+I	21	1897.839	1771.804	-864.748	n/a	n/a	2.08	0.344	0.344	0.210	0.210	0.037	0.026	0.155	0.037	0.155	0.026	0.037	0.220	0.026	0.220	0.037	0.026
JC	17	1903.084	1771.062	-811.225	0.61	0.27	2.89	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
GTR	25	1923.993	1821.919	-893.858	n/a	n/a	0.50	0.341	0.239	0.183	0.237	0.036	0.031	0.182	0.051	0.091	0.039	0.058	0.118	0.042	0.276	0.041	0.035

Models with lowest **BIC scores** (Bayesian information criterion) **are considered to describe the substitution pattern best** AICc (Akaike information criterion values, *InL* (Maximum Likelihood values) +G (discrete Gamma distribution) +I (invariant sites) *f (*nucleotide frequency*) r* (rate of base substitution) Assumed or estimated values of Transition/ Transversion bias (R) for each model. The sum of r values is equal to 1 for each model.

Abbreviation: GTR=General Time Reversible, HKY=Hasegawa-Kishino-Yano, TN93=Tamura-Nei Model, T92= Tamura 3-parameter, K2=Kimura 2-parameter, JC=Jukes-Cantor.

Table 6: Pairwise distance of *rps*14 gene of family *Lamiaceae* using MEGA7.

Plant species	1	2	3	4	5	6	7	8	9	10
Callicarpa macrophylla	0.00									
Tectona grandis	0.037									
Caryopteris odorata	0.063	0.062								
Gmelina philippensis	0.038	0.034	0.077							
Clerodendrum umbellatum	0.059	0.037	0.067	0.055						
Vitex trifolia	0.056	0.059	0.070	0.059	0.081					
Clerodendrum calamitosum	0.041	0.055	0.056	0.070	0.056	0.081				
Clerodendrum indicum	0.059	0.059	0.067	0.066	0.052	0.081	0.037			
Volkameria inermis	0.052	0.059	0.056	0.073	0.066	0.081	0.038	0.048		
Vitex agnus castus	0.063	0.052	0.084	0.059	0.067	**0.013**	0.074	0.088	**0.089**	

Table 7: Statistical analysis of *rps*14 protein models of *Lamiaceae* by I-TASSER.

S/N	Plant species	Z Score	C-Score	TM-Score	Description of Statistical values Good C score, Z Score and TM Score for all species
1	*Tectona grandis*	3.53	0.91	0.892	Good alignment, higher confidence and best match with template
2	*Vitex agnus castus*	3.51	0.90	0.896	Good alignment, higher confidence and best match with template
3	*Vitex trifolia*	3.50	0.93	0.905	Good alignment, higher confidence and best match with template
4	*Gmelina philippensis*	2.79	0.96	0.900	Good alignment, higher confidence and best match with template
5	*Callicarpa macrophylla*	3.56	0.91	0.896	Good alignment, higher confidence and best match with template
6	*Caryopteris odorata*	3.16	0.93	0.888	Good alignment, higher confidence and best match with template
7	*Volkameria inermis*	3.54	0.94	0.899	Good alignment, higher confidence and best match with template
8	*Clerodendrum indicum*	3.06	0.89	0.882	Good alignment, higher confidence and best match with template
9	*Clerodendrum umbellatum*	2.79	0.94	0.906	Good alignment, higher confidence and best match with template
10	*Clerodendrum calamitosum*	3.25	0.92	0.872	Good alignment, higher confidence and best match with template

Table 8: Ramachandran scores of *rps*14 protein model of *Lamiaceae* by RAMPAGE.

S/N	Plant species	Favored region	Allowed region	Outlier region
1	*Vitex agnus castus*	88	8	2
2	*Callicarpa macrophylla*	81	10	7
3	*Tectona grandis*	85	9	4
4	*Caryopteris odorata*	87	9	2
5	***Gmelina philippensis***	**85**	**13**	**0**

6	***Clerodendrum umbellatum***	**85**	**13**	**0**
7	*Viex trifolia*	88	8	2
8	***Clerodendrum calamitosum***	**84**	**13**	**1**
9	*Clerodendrum indicum*	83	13	2
10	*Volkameria inermis*	84	11	3

DISCUSSION

The present study reveals a close relationship of *Caryopteris* and *Clerodendrum* based on *rps*14 gene. This results are in congruent with the findings of Li *et al*. (2016) in which *Caryopteris* and *Clerodendrum* showed close relationship in the same clade of subfamily Ajugoideae based on *matK, ndhF, rbcL, rps*16 and *trnL-F*. The close relationship of *Caryopteris* with *Clerodendrum* was also studied on the basis of *ndhF* marker (Steane *et al*., 1997). The close relationship of *C. indicum, C. calamitosum and V. inermis* based on *rps*14 gene was also supported by (Steane *et al*., 1997, 1999 and 2004) based on cpDNA restriction site, ITS marker and cpDNA restriction site data, ITS and *ndhF* markers. Xiang *et al*. (2018) also reported close phylogenetic relationship of *C. indicum, C. calamitosum* and *V. inermis* based on *matK*, *rbcL*, *trnL* intron, *trnL-F* and *rps*16 markers. The *V. agnus castus* and *V. trifolia* depicted close phylogenetic relationship based on *rps*14 gene which was in accordance with the findings of Li *et al*. (2016) in which *Vitex agnus castus* and *Vitex trifolia* showed close phylogenetic association based on *matK, ndhF*, *rbcL, rps*16 *and trnL-F* markers. Bramley *et al*. (2009) reported close relationship of *V. agnus castus* and *V. trifolia* based on combined analysis of ITS and *ndhF* sequences.

The present study indicated that *C. macrophylla* and *T. grandis* showed close relationship in the same cluster based on *rps*14 gene that was also studied by Paton *et al*. (2004) on the basis of *trnL* intron, *trnL-F* and *rps*16 markers. The close relationship was also seen by Steane *et al*. (2004) on the basis of *ndhF* sequences, Wagstaff *et al*. (1998) on the basis of combined analysis of *rbcL* and *ndhF* sequences. Scheen *et al*. (2010) studied the close relationship of *C. macrophylla* with *T. grandis* based on *trnL-F* regions and *rps*16 intron. The close phylogenetic relationship of *G. philippensis* and *T. grandis* based on *rps*14 gene were also consistent with the findings of Paton *et al*. (2004) in which *Gmelina* showed close relationship with *T. grandis* on the basis of *trnL* intron, *trnL-F* and *rps*16, Xiang *et al.* (2018) based on *matK*, *rbcL*, *trnL-F* and *rps*16 molecular markers, Bendiksby *et al.* (2011) based on the *cp* genes *matK*, *rps*16, *trnL* intron and *trnL-F* spacer. Scheen *et al.* (2010) also studied the close relationship of *Gmelina* with *T. grandis* based on *trnL-F* regions and *rps*16 intron.

The *T. grandis* and *V. agnus castus* revealed close relationship based on *rps*14 gene was also supported by Li *et al.* (2016) based on *matK, ndhF, rbcL, rps*16 and *trnL-F* markers. Wagstaff *et al.* (1998) reported close phylogenetic relationship of *T. grandis* with *V. agnus castus* on the basis of *rbcL* and *ndhF* data sets. The *V. trifolia* showed phylogenetic association with *G. philippensis* in the same cluster in the present study that are similar to Bendiksby *et al*. (2011) on the basis of *trnL*-intron, *trnL-F* spacer, *rps*16 intron and *matK,* Paton *et al*. (2004) on the basis of *trnL* intron, *trnL–trnF* intergene spacer and *rps*16 intron. Scheen *et al*. (2010) also reported close phylogenetic association of *V. trifolia* and *Gmelina* based on the *trnL-F* regions and *rps*16 intron. The close association of *V. trifolia* and *C. macrophylla* on the basis of *rps*14 gene was supported by Paton *et al*. (2004) in which *Vitex trifolia* showed close relationship with *Callicarpa* on the basis of *trnL* intron, *trnL-F* and *rps*16.

The *V. agnus castus* and *G. philippensis* depicted close phylogenetic association and were grouped together in the same cluster on the basis of *rps*14 gene and comparable with the findings of Chen *et al*. (2014) in which *V. agnus castus* and G. *philippensis* showed close relationship in the same cluster based on *ndhF* and *rbcL*. The close association of *Callicarpa* and *Gmelina* in the same cluster based on *rps*14 gene was also reported by Scheen *et al*. (2010) on the basis of the *trnL-F* regions and *rps*16 intron. The *V. trifolia* revealed close relationship with *T. grandis* in the same cluster based on the *rps*14 gene was also studied by Scheen *et al*. (2010) based on the *trnL-F* regions and *rps*16 intron.

The present study showed lowest genetic diversity of (0.060) among *Lamiaceae* species based on *rps*14 gene. These results were comparable with diversity accessed by Talebi *et al*. (2019) in which low genetic diversity (0.063) was observed among *Salvia nemorosa* of family *Lamiaceae* collected from different parts of Iran and consists of several local populations based on ISSR marker. The method of conservation among *Lamiaceae* species are ex-situ conservation (Sun *et al*., 2019), as low genetic diversity in present study revealed ex-situ conservation that leads to diversification between *Lamiaceae* species. The result of the present study also supports the earlier work based on different DNA markers indicating that *rps*14 gene can be used for evaluation of the phylogenetic relationship and genetic diversity. The present study shows that *rps*14 gene is a potential marker as it revealed the close relationship among *Lamiaceae* species and low genetic diversity which is indicative of the conservative trend among *Lamiaceae* species. The validation of *rps*14 protein by RAMPAGE also indicated good protein models in different *Lamiaceae* species.

Conclusion: The present study revealed the close genetic relationship and low genetic diversity among different *Lamiaceae* species based on *rps*14 gene showing their usefulness for the phylogenetic relationship and genetic diversity analysis among *Lamiaceae* species. The 3D protein structures formed by I-TASSER and its validation through RAMPAGE indicated good protein models.

Author's contribution: All authors have equal contribution data analysis, drafted and review the manuscript.

Acknowledgements: We would like to thank Dr. Amir Sultan, Dr. Muhammad Zafar and Dr. Mushtaq Ahmad for identification of plant species.

Conflict of interest: The authors declare that there is not any conflict of interest in the manuscript.

REFERENCES

Akhtar, W., A. Rasheed, Z.K. Shinwari, S.M.S. Naqvi and T. Mahmood (2014). Genetic characterization of different Pakistani date palm varieties. Pakistan J. Bot. 46(6): 2095-2100.

Bramley, G.L., F. Forest and R.P. dekok (2009). Troublesome tropical mints: re-examining generic limits of Vitex and relations (Lamiaceae) in South East Asia. Taxon. 58(2): 500-510.

Bendiksby, M., L. Thorbek, A.C. Scheen, C. Lindqvist and O. Ryding (2011). An updated phylogeny and classification of Lamiaceae subfamily Lamioideae. Taxon. 60(2): 471-484.

Chen, S.C.G., M.C. Cheng, K.R. Chung, N.J. Yu and M.C. Chen (1992). Expression of the rice chloroplast psaA-psaB-rps14 gene cluster. Plant Sci. 81(1): 93-102.

Chen, Y.P., B. Li, R.G. Olmstead, P.D. Cantino, E.D. Liu and C.L. Xiang (2014). Phylogenetic placement of the enigmatic genus Holocheila (Lamiaceae) inferred from plastid DNA sequences. Taxon. 63(2): 355-366.

Dong, W., Xu. Chao, Wu. Ping, T. Cheng, Yu. Jing, S. Zhou and D.Y. Hong (2018). Resolving the systematic positions of enigmatic taxa: Manipulating the chloroplast genome data of Saxifragales. Mol. Phylogenet. Evol. 126: 321-330.

Drew, B.T., and K.J. Sytsma (2012). Phylogenetics, biogeography, and staminal evolution in the tribe Mentheae (Lamiaceae). Am. J. Bot. 99(5): 933-953.

Erdem, F., G. Doğan, Y. Kıran and H. Evren (2017). Morphological, anatomical, palynological and karyological characters of endemic Sideritis vulcanica Hub.-Mor.(Lamiaceae) from Turkey. IJNLS. 1(1): 1-11.

Harley, R.M. (2012). Checklist and key of genera and species of the Lamiaceae of the Brazilian Amazon. Rodriguésia. 63(1): 129-144.

Harley, R.M., S. Atkins, A.L. Budantsev, P.D. Cantino, B.J. Conn, R. Grayer, M.M. Harley, R. dekok, T. Krestovskaja, R. Morales, A.J. Paton, O. Ryding and T. Upson. (2004). Labiatae. In Flowering Plants Dicotyledons. Springer; Berlin, Heidelberg. 167-275p.

Jabeen, A., B. Guo, B.H. Abbasi, Z.K. Shinwari and T. Mahmood (2012). Phylogenetics of selected Mentha species on the basis of rps8, rps11 and rps14 chloroplast genes. J. Med. Plants Res. 6(1): 30-36.

Kumar, S., G. Stecher and K. Tamura (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33(7): 1870-1874.

Li, X., Y. Yang, R.J. Henry, M. Rossetto, Y. Wang and S. Chen (2015). Plant DNA barcoding: from gene to genome. Biol. Rev. 90(1): 157-166.

Li, B.o., P.D. Catino, R.G. Olmstead, G.L. Bramley, C.L. Xiang, Ma. Zhong-Hui. Y.H. Tan and D.X. Zhang (2016). A large-scale chloroplast phylogeny of the Lamiaceae sheds new light on its subfamilial classification. Scientific reports. 6(1): 1-18.

Li, B.o., and R.G. Olmstead (2017). Two new subfamilies in Lamiaceae. Phytotaxa. 313(2): 222-226.

Lei, W., D. Ni, Y. Wang, J. Shao, X. Wang, Y. Dan, J. Wang, H. Chen and Liu. Chang (2016). Intraspecific and heteroplasmic variations, gene losses and inversions in the chloroplast genome of Astragalus membranaceus. Scientific reports. 6(1): 1-13.

Luo, Y., P.F. Ma, H.T. Li, J.B. Yang, H. Wang and D.Z. Li (2016). Plastid phylogenomic analyses resolve Tofieldiaceae as the root of the early diverging monocot order Alismatales. Genome Biol. Evol. 8(3): 932-945.

Metcalfe, C.R., and L. Chalk. (1950). Anatomy of the Dicotyledons. 2 vols. Clarendon Press; Oxford (UK). 1030-1045p.

Paton, A.J., D. Springate, S. Suddee, D. Otieno, R. Grayer, M.M. Harley, F. Willis, M.S.J. Simmmonds, M.P. Powell and V. Savolainen (2004). Phylogeny and evolution of basils and allies (Ocimeae, Labiatae) based on three plastid DNA regions. Mol. Phylogenet. Evol. 31(1): 277-299.

Patwardhan, A., S. Ray and A. Roy (2014). Molecular markers in phylogenetic studies-a review. J Phylogen Evolution Biol. 2(2): 131.

Richards, E.J., M. Reichardt and S. Rogers (1997). Preparation of plant DNA using CTAB. Short protocol in Molecular biology. 3: 2- 10.

Saeed, S., F. Munir, I. Naveed, G.K. Raja and T. Mahmood

(2011). Phylogenetics of selected Plantago species on the basis of rps14 chloroplast gene. J. Med. Plants Res. 5(19): 4888-4891.

Scheen, A.C., M. Bendiksby, O. Ryding, C. Mathiesen, V.A. Albert and C. Lindqvist (2010). Molecular phylogenetics, character evolution, and suprageneric classification of Lamioideae (Lamiaceae) ANN. MISSOURI BOT. GARD. 97(2): 191-217.

Steane, D.A., R.W. Scotland, D.J. Mabberley, S.J. Wagstaff, P.A. Reeves and R.G. Olmstead (1997). Phylogenetic relationships of Clerodendrum sl (Lamiaceae) inferred from chloroplast DNA. Systematic Botany. 22(2): 229-243.

Steane, D.A., R.W. Scotland, D.J. Mabberley and R.G. Olmstead (1999). Molecular systematics of Clerodendrum (Lamiaceae): ITS sequences and total evidence. Am. J. Bot. 86(1): 98-107.

Steane, D.A., R.P. Dekok and R.G. Olmstead (2004). Phylogenetic relationships between Clerodendrum (Lamiaceae) and other Ajugoid genera inferred from nuclear and chloroplast DNA sequence data. Mol. Phylogenet. Evol. 32(1): 39-45.

Sun, Y., H. Yang, Q. Zhang, Qin. Luping, Li. Pan, J. Lee, S. Chen, K. Rahman, T. Kang and M. Jia (2019). Genetic diversity and its conservation implications of Vitex rotundifolia (Lamiaceae) populations in East Asia. Peer J. 7: e6194.

Talebi, S.M., R. Rezakhanlou and V.A. Matsyura (2019). Infraspecific genetic variation and population structure of Salvia nemorosa L.(Lamiaceae) in Iran. Ecologica Montenegrina. 26: 127-136.

Tajima, F. (1989). Statistical methods to test for neutral mutation hypothesis by DNA polymorphism. Genetics. 123(3): 585-595.

Tamura, K., and M. Nei (1993). Estimation of the number of nucleotides substitutions in the control region of mitochondial DNA in humas and chimpanzees. Mol. Biol. Evol. 10(3): 512-526.

Wakasugi, T., M. Sugita, T. Tsudzuki and M. Sugiura (1998). Updated gene map of tobacco chloroplast DNA. Plant Molecular Biology Reporter. 16(3): 231-241.

Wali, S., F. Munir and T. Mahmood (2013). Phylogenetic studies of selected Citrus species based on chloroplast gene, rps14. Int. J. Agric. Biol. 15(2): 357-361.

Wagstaff, S.J., L. Hickerson, R. Spangler, P.A. Reeves and R.G. Olmstead (1998). Phylogeny in Labiatae sl, inferred from cpDNA sequences. Pl. Syst. Evol. 209(3-4): 265-274.

Xiang, C.L., F. Zhao, P.D. Cantino, B.T. Drew, B. Li, E. Liu, DE. Soltis, PS. Soltis and H. Peng (2018). Molecular systematics of Caryopteris (Lamiaceae) and its allies with reference to the molecular phylogeny of subfamily Ajugoideae. Taxon. 67(2): 376-394.

Xie, D.F.,Y. Yan, Y. Deng, Li. Juan, Liu. Hai-Yang, S.D. Zhou and X.J. He (2018). Comparative analysis of the chloroplast genomes of the Chinese endemic genus Urophysa and their contribution to chloroplast phylogeny and adaptive evolution. Int. J. Mol. Sci. 19(7): 1847.

Yang, Z., T. Zhao, Q. Ma, L. Liang and G. Wang (2018). Comparative genomics and phylogenetic analysis revealed the chloroplast genome variation and interspecific relationships of Corylus (Betulaceae) species. Front. Plant Sci. 9: 927.