**Journal of Animal & Plant Sciences, 33(2): 2023, Page: 397-408**
**ISSN (print): 1018-7081; ISSN (online): 2309-8694**
https://doi.org/10.36899/JAPS.2023.2.0641

# DIVERSITY, PHYLOGENY AND 3-D PROTEIN MODELING OF FAMILY *ACANTHACEAE* BASED ON *RPS*14 GENE SEQUENCE IN PAKISTAN

S. Shafqat* and T. Mahmood*

Biochemistry and Molecular Laboratory, Department of Plant Sciences, Quaid- i- Azam University Islamabad, 45320, Pakistan
*Corresponding author's Email: sabashafqat42@gmail.com; tmahmood@qau.edu.pk

## ABSTRACT

The *Acanthaceae* is an important family of plants and it is also known as acanthus family, comprised of 250 genera and 4000 species spread globally in subtropical and tropical areas. From various geographical areas of Pakistan twelve species of *Acanthaceae* were collected to study its phylogeny utilizing ribosomal protein subunit 14 (*rps*14) gene. CTAB method was followed for the extraction of DNA from the leaves of collected plant samples. Primer was designed by using an online software primer3 (version 4.0). The sequence analysis was done using molecular evolutionary genetic analysis (MEGA 7), Iterative Threading Assembly Refinement (I-TASSER) and The Structure Analysis and Verification Server (SAVES). Our results showed two main clades of all twelve species revealing low genetic diversity (0.01) based on well supported bootstrap (BS) values. Furthermore, the pairwise distances ranged from 0.027 to 0.130 with mean value 0.078 was observed. To determine the 3-D protein structure and function of *rps*14 amino acid sequences, stereochemical analysis was performed using Ramachandran plots and I-TASSER. 3-D protein modeling depicted that *rps*14 protein from ten species out of twelve have the best quality structure with ≤ 2.0 % residues in disallowed region. Both phylogenetic analysis (amino acid and nucleotide sequences) revealed monophyletic origin of all species. Based on these results future studies should focus on physiological, taxonomical and molecular characters of family *Acanthaceae* species to determine lineages of all taxa within this diverse family.

**Key words:** *Acanthaceae*; *rps*14 gene; Genetic diversity; 3-D protein; Phylogeny



Published first online December 18, 2022 Published final March 24, 2023

## INTRODUCTION

Plants are important source of food, health, medicine, and ecological stability. Globally about 4,22, 000 plant species have been reported and classified into different classes and families (Hasan *et al*. 2007). However, a lot of plant species are still unidentified and new approaches are required to more clarify the evolutionary positions of species in a family (Hameed *et al*. 2022, Ullah *et al*. 2022). Likewise, family *Acanthaceae* is one of the important family that exhibit 250 genera and 4,000 species (Morais *et al*. 2019). It belongs to order Lamiales and distributed throughout the tropical and subtropical regions including Australia and extending to Mediterranean regions, Indonesia, Malaysia, Brazil and Africa, and Pakistan (Perveen and Qaiser 2010). Pakistan is hosting 18 genera and 60 specific and infraspecific species; of which 44 are native mostly found in Punjab, Islamabad, Azad Kashmir, Khyber Pakhtunkhwa and Sindh (Flora of Pakistan) (Malik and Ghafoor 1988). Primarily it includes annual and perennial herbs, shrubs, climbers, and some trees. Species of family *Acanthaceae* play important role in pharmacological, phytochemical and ethnobotanical activities such as insecticidal, antioxidant, anti-cancer, antifungal, antiviral, anti-inflammatory, and antifungal (Awan *et al*. 2014). In horticultural sector it includes ornamentals plants such as, jellybean plant (Crossandra), caricature-plant (Graptophyllum), wild petunia (Ruellia), polka dot plant (Hypoestes), and blue sage (Eranthemum) (Rahman and Gulshana 2014). The species of *Acanthaceae* are characterized by exhibiting spreading lobes or strongly 2 lipped, 4-5 lobed calyx, 4 didynamous stamens zygomorphic in nature, 2 epipetals, and gamopetalous corolla (Perveen and Qaiser 2010). The leaves are opposite, simple, estipulate, and entire margined. The inflorescence is solitary or racemose (McDade *et al*. 2000). The fruit contain loculicidal capsule and explosively dehiscent (Fongod *et al*. 2013). Morphological variations also exist in the members of family *Acanthaceae*, due to which taxonomic complexities exist within this family.

During last few decades, important development has been made in determining relationships among members of *Acanthaceae*. Several attempts have been made to classify the members of *Acanthaceae* family into more identical groups. Various scientists categorized members of *Acanthaceae* family into four subgroups

namely, (1) *Thunbergioideae*, (2) *Mendoncioideae*, (3) *Nelsonioideae* and (4) *Acanthoideae* (Bremekamp 1938, Balkwill and Norris 1988, Scotland 1992). These four subfamilies were documented based on their shape types, retinacula absence or presence, number of ovules and types of fruits. Various intra familial categorizations have been proposed for the *Acanthaceae*, but no taxonomic accord has yet been obtained (Scotland *et al*. 1995). To limit these complexities and shortcomings of morphological identifications, molecular characterizations which is cost effective and sustainable can be used by molecular biologists to identify and classify such complex families (Channa *et al*. 2018, Ayaz *et al*. 2020, Khan *et al*. 2020). Recent research on the morphological variation and evolution of *Acanthaceae* provides a phylogenetic framework for understanding and identifying the taxonomic characters (including micro-morphological structures) contained by the family (McDade *et al*. 2008, Morais *et al*. 2019, Khan *et al*. 2020).

Molecular studies provide a new insight to plant taxonomy and systematics botany (Asghar *et al*. 2022, Ullah *et al*. 2022, Zaman *et al*. 2022). Researchers recommended the requirements of molecular studies to trace the phylogenetic assembling and lineages among the members of *Acanthaceae* (Scotland *et al*. 1995, McDade and Moody 1999, Khan *et al*. 2020). Assembling and analyzing of molecular markers facilitated the botanists to move toward inclusion of *Avicenniaceae* in family *Acanthaceae* (Schwarzbach and McDade 2002). Some previous studies reported higher level systematics of *Acanthaceae* by using different genetic markers such as *rbc*L*, ndh*F*, trn*L-*trn*F and *trn*L-*trn*F separately and combined with *ITS*, respectively (Hedrén *et al*. 1995, Scotland *et al*. 1995, McDade and Moody 1999, McDade *et al*. 2000, Khan *et al*. 2020). In addition, several genetic regions and combination of genetic regions have been tested by experts of *Acanthaceae* for classification, however no taxonomic agreement has yet been achieved. The consortium for the barcode of life (CBOL) plant working group (2009) formally declared *rbc*L and *mat*K as core regions for plant species identification. However, there is still need to examine additional markers for identification and classification species of family *Acanthaceae* at the species level. To cover these gaps this study, aim to utilize conserved gene sequence of *rps*14 (a chloroplast gene) to address phylogenetic issues related to family *Acanthaceae*. Furthermore, we analyzed 3D protein structure of all collected species and validated Ramachandran analysis based on *rps*14 gene sequence.

# MATERIALS AND METHODS

**Sample's collection:** Plant samples were collected in flowering season from different areas of Pakistan including Lahore, Jmashoro, Nangarparkar, Islamabad and Swat. The plants samples were examined based on morphological and taxonomic characters comparing with the *Flora of Pakistan* (http://efloras.org/flora_page.aspx?flora_id=5). The plant samples were preserved in a sterile zipper bag and stored at 4°C for further use. The plant names and geographical detail is provided in Table 1.

**DNA extraction and amplification:** CTAB method was followed for the extraction of DNA from the leaves of collected plant samples (Porebski *et al*. 1997). The mixture was prepared with 2X CTAB buffer, 100 mM tris HCl, 20 mM ethylene diamine tetraacetic acid (EDTA), 1% mercaptoethanol, NaCl (1.4 g), and CTAB 2% w/v. The buffer's pH was adjusted to 8.0. The leaves were isolated from the samples with autoclaved scissors, sterilized with 70% ethanol and dried at room temperature. About 4.0 g leaf sample was grinded with mortar and pestle by adding preheated (65°C) CTAB buffer (1-2 ml) to get homogenized mixture. The Eppendorf tube containing homogenized mixture was incubated at 65°C for 50 min. The samples were centrifuged at 12,000 rpm for 20 min and supernatant was separated. The isolated supernatant was added with an equivalent volume of chloroform including isomyl alcohol of 24:1. The mixture was gently mixed 5-8 times of inversions. The reaction mixture was centrifuged again at 12,000 rpm for 12 min to isolate the supernatant. This step was repeated 4-5 times until dark yellow to white color was appeared. Isopropanol of equal volume was mixed with DNA to get the DNA in the supernatant. Subsequently the mixture was kept overnight at -20°C. The mixture was again centrifuged at 12,000 rpm for 20 min. The aqueous phase containing isopropanol was isolated and DNA pellet was stored. The pellet was washed with 70% ethanol to remove the impurities. Tris EDTA buffer containing RNase was added to degrade the RNA. The samples were stored at -20°C for further use.

**Primer designing:** The primer designing was performed through online software primer3 (version 4.0) (http://primer3.sourseforage.net/) using *rps14* gene sequence obtained from chloroplast DNA sequence of *Nicotiana tobacum* available in NCBI Genbank (https://www.ncbi.nlm.nih.gov/). The forwards sequence was *rps*14-F: 5'- ATGGCAAGGAAAAGTTTGATTC-3' while reverse sequence was *rps*14-R: 5'- TTACCAACTTGATCTTGTTGCTCCT- 3'.

**PCR amplification:** The PCR product prepared in 25 µL reaction mixture including nano pure water (16.2 µL),

forward and reverse *rps*14 primers (both 25 pmoles) 1 µL volume of each primer, *Taq* DNA polymerase (1.5 U concentration in 0.3 µL), dNTPs (2 mM in 1.5 µL), Taq buffer (10 X in 2.5 µL), DNA template (1 µL) and $MgCl_2$ (25 mM in 1.5 µL). The DNA presence was analyzed through gel electrophoresis exhibiting 1% agarose gel, followed by staining of ethidium bromide. The gel was examined under UV light of Dolphin Doc plus documentation system (Wealtec). The amplified regions were sent to Beijing Genomic Institute (BGI), Shenzhen, China for sequencing. The amplicons of *rps*14 were obtained for every individual species and subjected to Basic Local Alignment Search Tool (BLAST) in Gnebank (https://www.ncbi.nlm.nih.gov). The sequences were deposited to database of NCBI.

**Phylogenetic pattern:** The sequences were aligned and assembled through MEGA7 (Ayaz *et al*. 2020). The matrices were adjusted manually through insertion and deletion which was coded as absent/present. Neighbor joining method with 1000 bootstrap's value using MEGA7 was utilized, to examine phylogenetic relationship of all collected species. pairwise distance evaluation was also examined through MEGA7 to find evolutionary divergence among *rps*14 gene sequences of all species. To find nucleotide diversity among all collected species, Tajima's neutrality test was performed. Average nucleotide number of *rps*14 from all collected species was examined using MEGA7 software, to examine genetic relationship among the species of *Acanthaceae* from Pakistan. Substitution rates and patterns were estimated using Tamura and Nei (1993) model.

**Protein BLAST and nucleotide translation:** Nucleotide sequences were translated by using ExPASy translate tool and aligned by online web server JustBio (http://web.expasy.org/translate/). Amino acid sequences of *rps*14 representing twelve members were uploaded to NCBI database. Protein BLAST was performed and compared with already reported amino acid sequences in Genbank. After assembling and alignment phylogenetic tree based on amino acid sequences from 12 collected species was constructed to examine the phylogenetic relationship using MEGA7 via neighbor joining method with a bootstrap's value of 1000 (Porebski *et al*. 1997).

**Determination of protein structure:** Three dimensional (3-D) atomic structure models were evaluated, using I-TASSER (Iterative Threading Assembly Refinement). The amino acid sequences of *rps*14 gene from all collected species were uploaded individually on I-TASSER server and the atomic model for each sequence was examined following Ayaz *et al*. (2020) 3D atomic model of *rps*14 gene sequence was evaluated via Ramachandran plot analysis and PROCHECK (Hooft *et al*. 1997, Laskowski *et al*. 2006). For Ramachandran plot and PROCHECK analysis Pdb files of 3D models were uploaded on SAVES. PROCHECK determine the stereo chemical quality of predicted protein structure by estimating residue by residue geometry (https://services.mbi.ucla.edu/SAVES/). Ramachandran plot assessment was done by uploading pdb files of 3-D models on RAMPAGE (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php).

## RESULTS AND DISCUSSION

**DNA amplification:** The band size of the amplicon for all collected species was about 303 bp. The sequence data was analyzed, and accession numbers were obtained from NCBI Gnebank (Table 2). The fine quality of DNA was obtained (Figure 1). The nucleotide average for all rps14 gene sequences was C=18.7, A 33.5, T (U)= 25.0, and G=22.9. Overall A showed the highest average number while C revealed the lowest average (Table 2). Molecular techniques based on sequencing give improved solution at intra-genus and higher level (Li *et al*. 2021). Various morphological and biochemical techniques have also been used in plant identification because these are economical. Molecular techniques (Hajibabaei *et al*. 2007) such as DNA barcoding, amplified fragment length polymorphism (AFLP), single nucleotide polymorphisms (SNP) (Loureiro *et al*. 2020), DNA amplification fingerprinting (DAF), random amplified polymorphic DNA (RAPD) (Hadrys *et al*. 1992), microsatellites, cleaved amplified polymorphic sequences (CAPS), simple sequence repeats (SSR), allele specific associated primers (ASAP), arbitrary primed PCR (AP-PCR), sequence characterized amplified region (SCAR), restriction fragment length polymorphism (RFLP), sequence tagged microsatellite loci (STMS) and single strand conformation polymorphism (SSCP) (Labou *et al*. 2020) can be used for plant systematics, taxonomic and phylogenetic studies (Sheth and Thaker 2017).

**Nucleotide based phylogenetic tree assessment of *rps*14:** In the present study, genetic characterization was inferred by sequences of rps14 gene. Based on nucleotide data of rps14 gene, the phylogenetic tree was constructed using neighbor Joining tree method with the help of MEGA7 (Kumar *et al*. 2016). The molecular analysis of the results revealed two main clades i.e. clade I and clade II as shown in the Figure 1. Clade I is comprised of nine species namely, *Justicia peploides, Dicliptera roxburghiana, Eranthemum alba, Peristrophe paniculata, Blepharis sindica, Strobilanthes urticifolia, Barleria acanthoides, Thunbergia laurifolia* and *Ruellia brittoniana* (Figure 2). Because of the genetic variability this clade was further divided into two clusters, cluster I and cluster II shown in Figure 1. To study this phylogram in detail, the cluster I was further divided into two groups, group I and group II. Group I comprised of two

species namely, *Justicia peploides* and *Dicliptera roxburghiana* with BS value of 51 % which showed their resemblance, while *Eranthemum alba* is forming an outgroup with group I. Group II consisted of two species namely *Peristrophe paniculata* and *Blepharis sindica*. These species showed high similarity with BS value of 72 %. Whereas Clade II is mainly comprised of three species namely, *Eranthemum pulchellum, Thunbergia alata* and *Crossandra infundibuliformis* in which *Thunbergia alata* showed close relationship with *Crossandra infundibuliformis* with BS value of 63 %. *Eranthemum pulchellum* is forming an outgroup with BS value of 56 % indicating that it might be the possible ancestor of all the species in clade II (Figure 1). Phylogeny is significant for understanding the origin of classification, evolutionary history, taxonomy, similarities, and variation of this economically and ecologically important family of angiosperms. The species identification approaches have been changed with the development of DNA-based markers (Popiela and Molnár 2020, Song *et al*. 2020). Previously (Khan *et al*. (2020)) studied phylogenetic relationship of eight species from family *Acanthaceae* based on *rbc*L, *mat*K and *trn*H-*psb*A. Our approach is different in two ways: first we utilized *rps*14 gene sequence for twelve species from Pakistan. Second, we also determined protein structure analysis using *rps*14 gene sequence. In past various researchers used molecular approaches to explore the natural relationship within taxa of *Acanthaceae* (Scotland 1992, Schwarzbach and McDade 2002, Huang *et al*. 2020). The taxonomy and systematics of *Acanthaceae* from different aspects has been significantly emphasized based on molecular markers of plastid gene region (*rbc*L, *mat*K, *atb*β, *trn*L, *ndh*F, *trn*F, *rps*16, trnG), nuclear gene region (*ITS, ITS*1, *ITS*2, *Eif3E*) and different morphological characters (McDade *et al*. 2008, Tripp *et al*. 2013, Chumchim *et al*. 2015). Our results are in consistent with previous studies that showed two major clades in phylogenetic tree based on different molecular markers (Scotland *et al*. 1995, Khan *et al*. 2020). In the present work, an overall close relationship revealed common origin of species. Species under study were monophyletic which confirmed the monophyly of family *Acanthceae* which agrees with the results of Scotland *et al*. (1995) and McDade *et al*. (2000). Monophyletic nature of these species were also confirmed by phylogenetic study of (McDade *et al*. 2008).

**Nucleotide diversity:** To find nucleotide diversity Tajima's neutrality test of selection was performed (π) with the MEGA7 software for *rps*14 gene sequences from 12 species of family *Acanthaceae*. Twelve numbers of sequences (m) have given 61 segregation sites (S) showing little diversity of nucleotides (π=0.073457) (Table 3). This low nucleotide diversity is a sign of close genetic relationship among different species of the family *Acanthaceae*. It also indicates rare mutation rate in this family. Studies suggest that indirect or direct selection in natural populations are low (Ackerly *et al*. 2000, Juenger *et al*. 2000). Comparatively *rps*14 genome region is the most stable gene as the plant barcode (Figueroa *et al*. 1999, Ayaz *et al*. 2020).

**Estimate of maximum likelihood (ML) substitution pattern:** To determine substitution rates and pattern Tamura *et al*. (2004) model was applied. While evaluating them, analogous estimation of instantaneous (r) was considered. For convenience, the sum of r values is made equivalent to 100. Rates of diverse transitional substitutions are mentioned in red font and those of transversional substitutions are shown in italics (Table 4). The nucleotide frequencies are 33.5 % (A), 25.0 % (T/U), 18.7 % (C), and 22.9 % (G). The ML for this computation was -1107.069. The calculation included 12 species of family *Acanthaceae* collected from different regions of Pakistan. These results determined that GC contents of coding regions in *rps*14 genome region are higher as compared to introns. Our results are in consistent with previous studies revealing a higher percentage of GC contents with little variations (Ayaz *et al*. 2020, Khan *et al*. 2020).

**Calculation of Evolutionary divergence:** On the basis of rps14 gene pairwise distance was computed using MEGA7. For rps14 gene the value of genetic diversity was observed to be in the range of 0.027 to 0.130 with general mean distance of 0.078. These very low distance values showed that all species are genetically related to each other and there is low genetic diversity among them based on *rps*14 gene (Table 5). In plants, molecular phylogenies are usually based on cp DNA sequence variation Després *et al*. (2003) as chloroplast genes such as *rps*14 have relatively low genetic diversity. The *rps*14 gene was reported as conserved gene in various species as compared to *rbc*L, AFLP, ITS, and *mat*K (Wali *et al*. 2013, Akhtar *et al*. 2014). Our results provided support for use of *rps*14 gene sequences and proved high phylogenetic resolution for most of the clades within family *Acanthaceae*.

**Phylogenetic pattern based on *rps*14 amino acid sequences:** For understanding the phylogenetic pattern of *rps*14 gene sequences of family *Acanthaceae*, amino acid sequences were also used. Based on the amino acid data of protein, the phylogenetic tree (Figure 3) was constructed using MEGA7 (Kumar *et al*. 2016). The phylogenetic tree showed two clades i.e. clade I and clade II. Clade I of protein based phylogenetic tree is comprised of eight species namely, *Eranthemum pulchellum, Thunbergia alata, Justicia peploides, Dicliptera roxburghiana, Eranthemum alba, Crossandra infundibuliformis, Peristrophe paniculata* and *Blepharis sindica*. Clade I is further divided into two clusters;

cluster I and cluster II (Figure 2). Cluster I is further divided into two groups; group I and group II. Group I comprised of two species *Eranthemum pulchellum* and *Thunbergia alata. Justicia peploides* and *Dicliptera roxburghiana* forming outgroups with group I species. Group II contains *Eranthemum alba* and *Crossandra infundibuliformis*. On the other hand, cluster II is comprised of only two species namely, *Peristrophe paniculata* and *Blepharis sindica* (Figure 2). Whereas Clade II is comprised of *Barleria acanthoides, Strobilanthes urticifolia, Thunbergia laurifolia* and *Ruellia brittoniana.* Clade II is further divided into group I which is comprised of two species *Thunbergia laurifolia* and *Ruellia brittoniana* depicting high similarity with BS value of 85 %. *Strobilanthes urticifol ia* and *Barleria acanthoides* formed outgroups showing that they might be the possible ancestor of group I. Protein based tree showed that *Eranthemum pulchellum* and *Eranthemum alba* are placed in same cluster depicting high similarities. The separate clustering of *Crossandara infundibuliformis* from *Dicliptera roxburghiana* and *Justicia peploides* showed genetic divergence that was also reported by McDade and Moody (1999). Based on these results species of *Acanthaceae* family showed common origin. In addition, these results showed that the species are monophyletic which confirmed the monophyly of family *Acanthceae* which is in consent with the results of Scotland *et al*. (1995) and McDade *et al*. (2000).

**Comparative study (nucleotide and amino acid sequences-based tree):** In the present study *rps*14 nucleotide and amino acid sequences were used to conclude phylogenetic relationship. The genetic distance in both phylogenetic trees is same i.e. 0.0100 showing consistency in both trees. The BS value based on nucleotides was up to 90 % (Figure 4), while for protein-based tree it was up to 85 % (Figure 5). The amino acid sequences of *rps*14 protein based phylogenetic trees was bit diversified as compared to the nucleotide sequences based phylogenetic tree. It was observed that amino acid-based tree is more supportive for explaining genetic diversity among 12 species of family *Acanthaceae* as compared to nucleotide sequence-based tree. The probable explanation behind such diversity may be since there is hardly significant evolutionary pressure to avoid a silent mutation in a protein sequence, so the mutation should not count much against the protein homology score and also the proteins are made up of twenty different amino acids while DNA sequences are made up of four nucleotides (Zhang and Yang 2015). It is simple to accomplish statistically significant alignment by comparing sizeable characters as these have much less likely to get a match by chance. The phylogenetic analysis of both nucleotide and amino acid sequences generated a dendrogram that explains the relationship among different species based on rps14. It was found that *Justicia peploides, Dicliptera roxburghiana, Eranthemum alba, Peristrophe paniculata* and *Blepharis sindica* were grouped in cluster I while cluster II was occupied by *Strobilanthes urticifolia, Barleria acanthoides, Thunbergia laurifolia* and *Ruellia brittoniana*. However, a distinct grouping was observed for *Eranthemum pulchellum, Thunbergia alata* and *Crossandra infundibuliformis* which may be due to their divergent evolutionary history. In the present work, an overall close relationship revealed common origin of species. Species under study were monophyletic in both amino acid and nucleotide based phylogenetic trees (Figure 4 and 5).

**Protein structure analysis and validation:** Five full length atomic models were created by I-TASSER (Iterative Threading Assembly Refinement). The confidence of each structural models was estimated by the confidence score (C-score). 3D protein models of rps14 that have highest C-score is shown in Figure 3. The precision and stereo chemical quality of the predicted proteins from the 12 *Acanthaceae* species under examination were observed while using the PROCHECK program (http://services.mbi.ucla.edu/PROCHECK/) through residue-by-residue and overall geometry of protein structures. Ramachandran plots were drawn for these 3D protein models, where the most allowed regions are shown by red patches, while allowed regions are indicated by yellow areas (Figure 5).

It was observed in Ramachandran plots of selected specimens that normally protein models for all the species estimated by I-TASSER were fine and acceptable that can be used for further analysis because they have ≥ 83 % to ≤ 94.9 % residues that lie in the most allowed region and ≥ 0 % to ≤ 5.1 % residues that lie in the disallowed region. All the protein models of rps14 of *Thunbergia alata, Crossandra infundibuliformis, Eranthemum alba, Thunbergia laurifolia, justicia peploides, Strobilanthes urticifolia, Dicliptera roxburghiana, Ruellia brittoniana, Barleria acanthoides* and *Blepharis sindica* were found to be the best quality structural models, as they have ≤ 2 % residues that lie in the disallowed region expect the *Peristrophe paniculata* and *Eranthemum pulchellum* that have ≤ 2.6 % and ≤ 5.1 % residues respectively that lie in the disallowed region. Detailed values of Ramachandran plots for twelve sequences are given in Table 6. After phylogenetic studies, 3D protein modeling and Ramachandran analysis were also evaluated for the first time. Formerly it was assessed by Munir *et al*. (2013) for wound-inducible proteinase inhibitor-II gene from four different varieties of tomato. The analyses and authentication of results from Ramachandran plot evaluation revealed that rps14 protein structure models of 12 species of *Acanthaceae* have good quality structural model. In this work, the molecular analysis of the selected genera of *Acanthaceae*

was studied. To the best of our knowledge, this is the first report that the *rps*14 gene was used to infer phylogeny as well as stereo chemical analysis from *Acanthaceae*. The *rps*14 gene of chloroplast has significant data with conserved evolutionary rate, so it was used by researchers for phylogenetic and comparative studies in future.


**Figure 1: Isolated genomic DNA of selected species of *Acanthaceae* 1) *Eranthemum pulchellum*, 2) *Blepharis sindica,* 3) *Ruellia brittoniana,* 4) *Peristrophe paniculata,* 5) *Thinbergia laurifolia,* 6) *Justicia peploides* and 7) *Eranthemum alba***

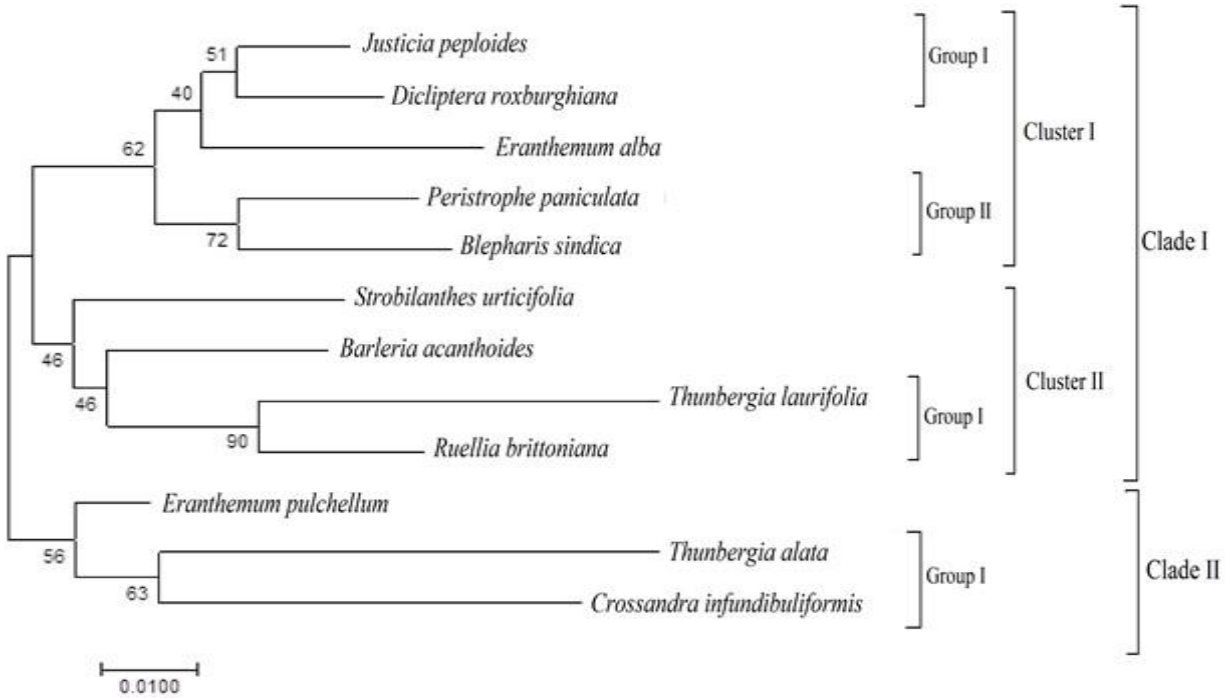


**Figure 2: Phylogenetic tree of collected species of family *Acanthaceae* based on *rps*14 nucleotide sequences constructed by MEGA7**

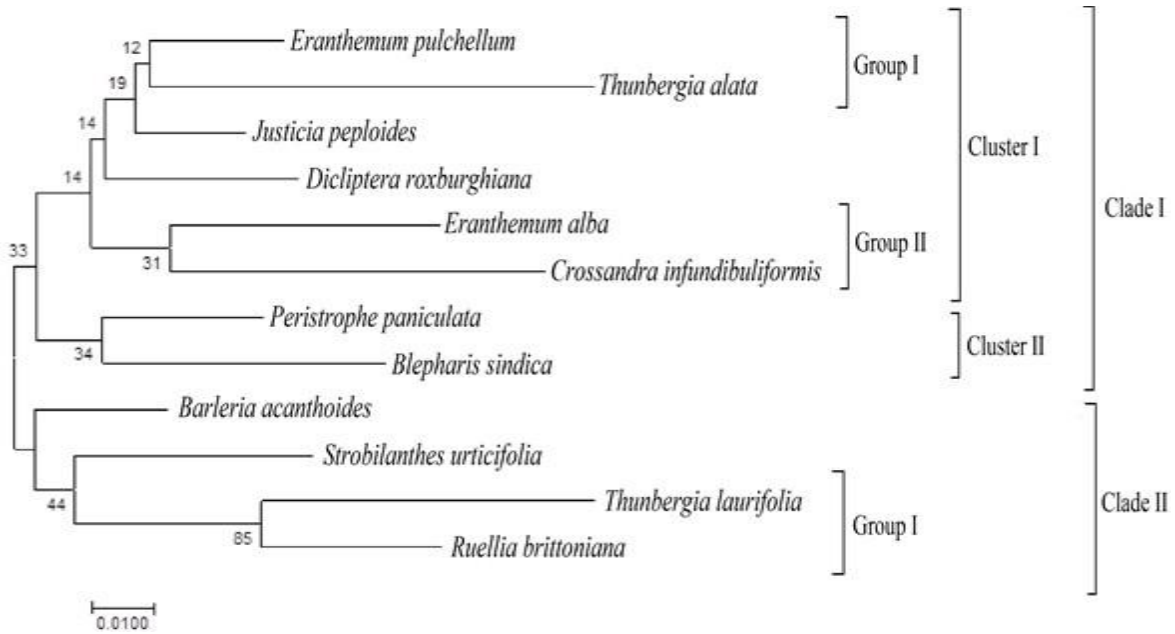


**Figure 3. Phylogenetic tree of 12 species of *Acanthaceae* based on amino acid sequences of *rps*14 constructed by MEGA7**

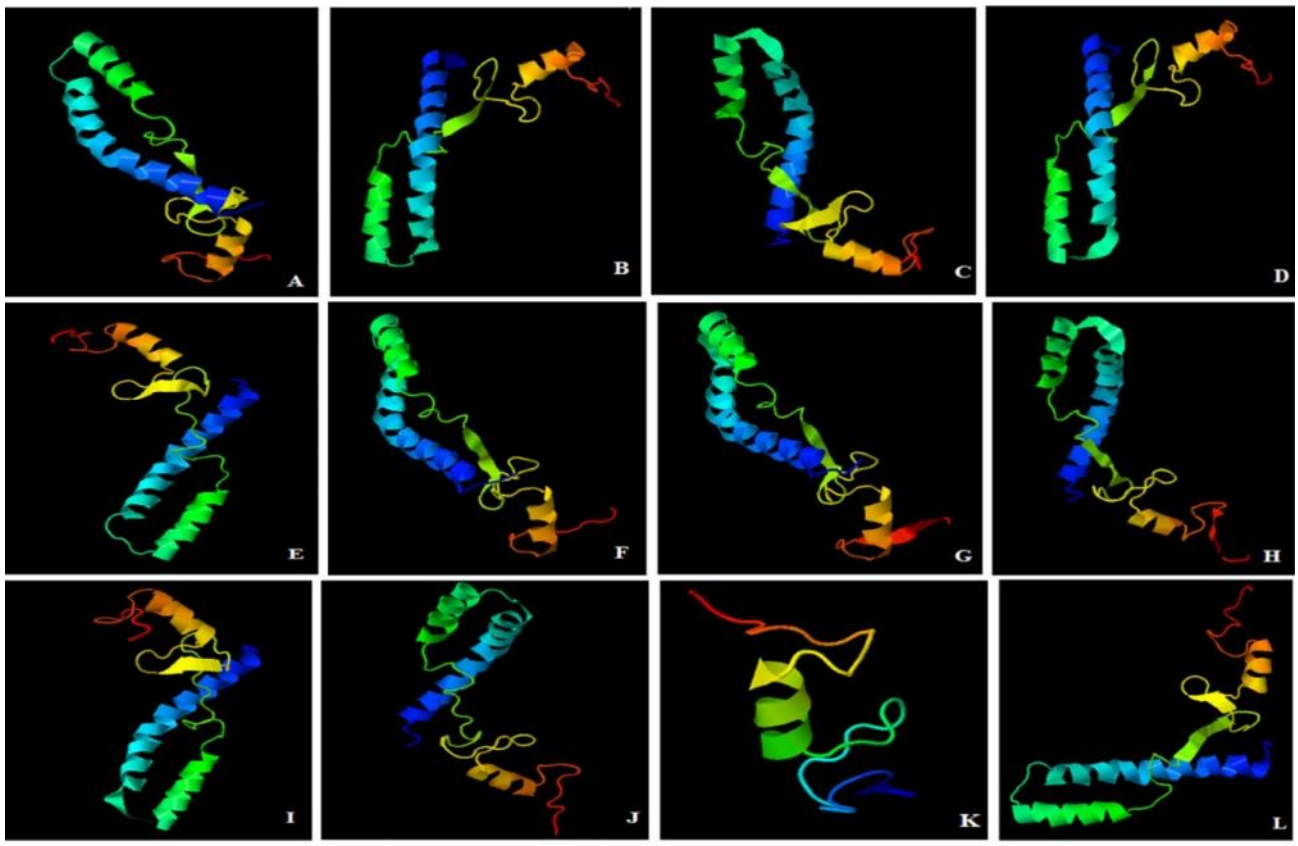


**Figure 4. The 3D protein models of *rps*14 having the highest C-score build by I-TASSER for different species of family *Acanthaceae* A*: Thunbergia laurifolia* B: *Crossandra infundibuliformis* C: *Dicliptera roxburghiana* D: t*Eranthemum alba* E: *Eranthemum pulchellum* F: *Justicia peploides* G: *Strobilanthes urticifolia* H: *Thunbergia alata* I: *Ruellia brittoniana* J: *Barleria acanthoides* K: *Peristrophe paniculata* L: *Blepharis sindica***

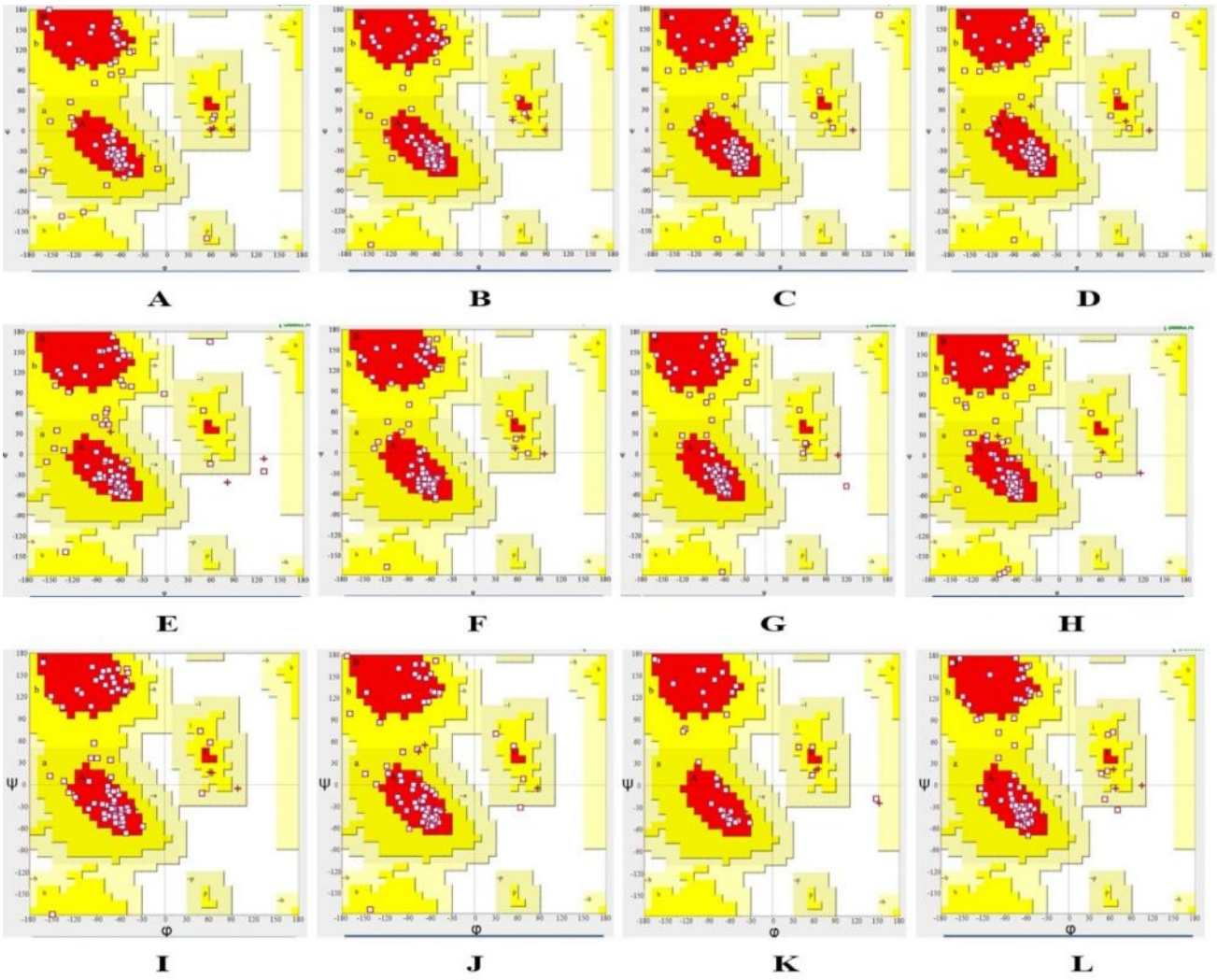


**Figure 5: The Ramachandran plot for *rps*14 protein of different species of family *Acanthaceae* A*: Thunbergia laurifolia* B: *Crossandra infundibuliformis* C: *Dicliptera roxburghiana* D: *Eranthemum alba* E: *Eranthemum pulchellum* F: *Justicia peploides* G: *Strobilanthes urticifolia* H: *Thunbergia alata* I: *Ruellia brittoniana* J: *Barleria acanthoides* K: *Peristrophe paniculata* L: *Blepharis sindica***

**Table 1. List of selected species of family *Acanthaceae* along with their longitude, latitude and area of collection.**

| S/N | Species Name | Location | Geographic coordinate |
|---|---|---|---|
| 1 | *Justicia peploides* | Islamabad | 33.7444° North, 73.0417° East |
| 2 | *Ruellia brittoniana* | Islamabad | 33.7444° North, 73.0417° East |
| 3 | *Eranthemum pulchellum* | Lahore | **31.5820**° North, **74.3293**° **East** |
| 4 | *Strobilanthes urticifolia* | Lahore | **31.5820**° North, **74.3293**° **East** |
| 5 | *Thinbergia laurfolia* | Lahore | **31.5820**° North, **74.3293**° **East** |
| 6 | *Dicliptera roxburghiana* | Lahore | **31.5820**° North, **74.3293**° **East** |
| 7 | *Eranthemum alba* | Lahore | **31.5820**° North, **74.3293**° **East** |
| 8 | *Thunbergia alata* | Lahore | **31.5820**° North, **74.3293**° **East** |
| 9 | *Crossandra infundibuliformis* | Lahore | **31.5820**° North, **74.3293**° **East** |
| 10 | *Barleria acanthoides* | Jamshoro | 25.4169° **North,** 68.2743° **East** |
| 11 | *Blepharis sindica* | Nagarparkar | 24.3605° **North,** 70.7585° **East** |
| 12 | *Peristrophe paniculata* | Swat | 35.4920° North, 72.5205° East |

**Table 2. Estimate of nucleotide composition and accession numbers of *rps*14 gene sequences representing 12 different species of *Acanthaceae.***

| S/N | Species Names | T/U | C | A | G | Total | Accession Number |
|---|---|---|---|---|---|---|---|
| 1 | *Thunbergia laurifolia* | 24.1 | 20.5 | 33.3 | 22.1 | 303.0 | KY242375 |
| 2 | *Eranthemum pulchellum* | 24.8 | 19.1 | 32.7 | 23.4 | 303.0 | KY404064 |
| 3 | *Strobilanthes urticifolia* | 25.1 | 19.1 | 33.3 | 22.4 | 303.0 | KY404065 |
| 4 | *Justicia peploides* | 25.1 | 18.2 | 33.7 | 23.1 | 303.0 | KY404066 |
| 5 | *Dicliptera roxburghiana* | 25.4 | 18.8 | 34.0 | 21.8 | 303.0 | KY404067 |
| 6 | *Eranthemum alba* | 26.1 | 17.8 | 33.7 | 22.4 | 303.0 | KY404068 |
| 7 | *Thunbergia alata* | 25.7 | 18.5 | 33.7 | 22.1 | 303.0 | KY404069 |
| 8 | *Crossandra infundibuliformis* | 25.4 | 18.5 | 31.7 | 24.4 | 303.0 | KY404070 |
| 9 | *Ruellia brittoniana* | 23.8 | 18.8 | 34.0 | 23.4 | 303.0 | MF045822 |
| 10 | *Barleria acanthoides* | 24.8 | 18.5 | 34.0 | 22.8 | 303.0 | MF045823 |
| 11 | *Peristrophe paniculata* | 24.4 | 18.2 | 34.0 | 23.4 | 303.0 | MF045824 |
| 12 | *Blepharis sindica* | 25.1 | 18.2 | 34.0 | 22.8 | 303.0 | MF045825 |
| | **Average** | **25.0** | **18.7** | **33.5** | **22.9** | **303.0** | |

**Table 3. Tajima's neutrality test values based on *rps*14 gene of 12 different species of family *Acanthaceae* using MEGA7.**

| No. of sequences "m" | No. of segregating sites "S" | Ps= S/n | Θ=Ps/a1 | Nucleotide diversity "π" | Tajima test statistic "D" |
|---|---|---|---|---|---|
| 12 | 61 | 0.201320 | 0.066665 | 0.073457 | 0.470205 |

**Table 4: Maximum likelihood values of transitional (bold) and transversional substitution (italics) of nucleotides of *rps*14 gene for 12 different species of *Acanthaceae*, calculated through MEGA7.**

| | A | T/U | C | G |
|---|---|---|---|---|
| **A** | - | *5.61* | *4.20* | **16.28** |
| **T** | *7.53* | - | **6.38** | *5.13* |
| **C** | *7.53* | **8.53** | - | *5.13* |
| **G** | **23.87** | *5.61* | *4.20* | - |

* Adenine (A), Thymine/Uracil (T/U), Cytosine (C), Guanine (G)

**Table 5: Pairwise distance of *rps*14 gene sequences from selected species of *Acanthaceae* using MEGA7. The bold value represents the highest and lowest values.**

| S/N | Species Names | Values | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | *Thunbergia laurifolia* | | | | | | | | | | | |
| 2 | *Eranthemum pulchellum* | 0.080 | | | | | | | | | | |
| 3 | *Strobilanthes urticifolia* | 0.088 | 0.051 | | | | | | | | | |
| 4 | *Justicia peploides* | 0.095 | 0.048 | 0.066 | | | | | | | | |
| 5 | *Dicliptera roxburghiana* | 0.107 | 0.058 | 0.066 | **0.027** | | | | | | | |
| 6 | *Eranthemum alba* | 0.127 | 0.069 | 0.066 | 0.048 | 0.044 | | | | | | |
| 7 | *Thunbergia alata* | **0.130** | 0.069 | 0.107 | 0.084 | 0.088 | 0.111 | | | | | |
| 8 | *Crossandra infundibuliformis* | 0.126 | 0.058 | 0.087 | 0.087 | 0.091 | 0.080 | 0.095 | | | | |
| 9 | *Ruellia brittoniana* | 0.058 | 0.055 | 0.063 | 0.077 | 0.088 | 0.092 | 0.115 | 0.115 | | | |
| 10 | *Barleria acanthoides* | 0.088 | 0.041 | 0.055 | 0.073 | 0.073 | 0.088 | 0.099 | 0.099 | 0.048 | | |
| 11 | *Peristrophe paniculata* | 0.103 | 0.066 | 0.077 | 0.048 | 0.041 | 0.066 | 0.111 | 0.122 | 0.069 | 0.055 | |
| 12 | *Blepharis sindica* | 0.099 | 0.069 | 0.080 | 0.051 | 0.059 | 0.062 | 0.115 | 0.088 | 0.080 | 0.069 | 0.041 |

**Table 6: Ramachandran score of *rps*14 protein models for all the 12 amino acid sequences predicted by PROCHECK.**

| S/N | Species Names | Most Allowed region | Allowed region | Disallowed region |
|---|---|---|---|---|
| 1 | *Thunbergia laurifolia* | 84.7 | 13.3 | 2.0 |
| 2 | *Crossandra infundibuliformis* | 94.9 | 4.1 | 1.0 |
| 3 | *Thunbergia alata* | 89.8 | 9.2 | 1.0 |
| 4 | *Eranthemum alba* | 88.8 | 10.2 | 1.0 |
| 5 | *Dicliptera roxburghiana* | 89.8 | 9.2 | 1.0 |
| 6 | *Justicia peploides* | 88.8 | 11.2 | 0.0 |
| 7 | *Strobilanthes urticifolia* | 92.9 | 5.1 | 2.0 |
| 8 | *Eranthemum pulchellum* | 83.7 | 11.2 | 5.1 |
| 9 | *Ruellia brittoniana* | 87.8 | 11.2 | 1.0 |
| 10 | *Barleria acanthoides* | 88.8 | 11.2 | 0.0 |
| 11 | *Peristrophe paniculata* | 84.2 | 13.2 | 2.6 |
| 12 | *Blepharis sindica* | 88.8 | 10.2 | 1.0 |

**Conclusion:** This study concludes that the family *Acanthaceae* has narrow genetic distance of 0.010, indicating low genetic diversity. 3D protein modeling and Ramachandran analysis of our results determined that *rps*14 protein from ten species out of twelve have the best quality structure with ≤ 2.0 % residues in disallowed region. Our study also concludes that family *Acanthaceae* is monophyletic. Moreover, the interspecific variation in the nucleotide sequence appears to be more prevalent in *rps*14 gene sequence as compared to the *mat*K and *rbc*L sequences. Therefore, it may be suggested that *rps*14 gene should be used as a barcode in the recognition of plant species. Our results offer a basis to which additional taxa can be readily added to address specific questions regarding phylogenetic relationships. For better understanding of complex taxonomic relationships of *Acanthaceae*, there is a future need of phytochemical, anatomical studies, morphological and molecular characterization of family *Acanthaceae*.

**Acknowledgment:** We are thankful to department of Laboratory of Molecular Biology, Department of Plant Science for providing basic necessities during this research work.

## REFERENCES

Ackerly, D. D., S. A. Dudley, S. E. Sultan, J. Schmitt, J. S. Coleman, C. R. Linder, D. R. Sandquist, M. A. Geber, A. S. Evans, and T. E. Dawson (2000). The evolution of plant ecophysiological traits: recent advances and future directions: new research addresses natural selection, genetic constraints, and the adaptive evolution of plant ecophysiological traits. Biosci. 50:979-995. https://doi.org/10.1641/0006-3568(2000)050[0979:TEOPET]2.0.CO;2

Akhtar, W., A. Rasheed, Z. K. Shinwari, S. M. S. Naqvi, and T. Mahmood (2014). Genetic characterization of different Pakistani date palm varieties. Pak J Bot. 46:2095-2100. https://pakbs.org/pjbot/PDFs/46(6)/23.pdf

Asghar, M., M. Younas, B. Arshad, W. Zaman, A. Ayaz, S. Rasheed, A. Shah, F. Ullah, and S. Saqib (2022). Bioactive potential of cultivated *Mentha arvensis* L. for preservation and production of health-oriented food. J. Anim. Plant Sci., 32:835-844.http://doi.org/10.36899/JAPS.2022.3.0484

Awan, A. J., C. B. Ahmed, M. Uzair, M. S. Aslam, U. Farooq, and K. Ishfaq (2014). Family Acanthaceae and genus Aphelandra: ethnopharmacological and phytochemical review. Int J Pharm Pharm Sci. 10:44-55. https://innovareacademics.in/journals/index.php/ijpps/article/view/2569.

Ayaz, A., W. Zaman, S. Saqib, F. Ullah, and T. Mahmood (2020). Phylogeny and diversity of lamiaceae based on rps14 gene in Pakistan. Genetika. 52:435-452. https://doi.org/10.2298/GENSR2002435A

Balkwill, K., and F. G. Norris (1988). Classification of the Acanthaceae: a southern African perspective. Missouri Botanical Garden. https://www.scirp.org/(S(lz5mqp453edsnp55rrgjct55.))/reference/referencespapers.aspx?referenceid=1429403

Bremekamp, C. E. B (1938). Notes on the Acanthaceae of Surinam. Recueil des travaux botaniques néerlandais. 35:130-176. https://repository.naturalis.nl/pub/534841

Channa, F. N., Z. K. Shinwari, and S. I. Ali (2018). Phylogeny of tamaricaceae using PsbA-trnH nucleotide sequences. Pak J Bot. 50:983-987. http://www.pakbs.org/pjbot/papers/1521182711.pdf

Chumchim, N., L. A. McDade, and A. E. Fisher (2015). Phylogeny of Dyschoriste (Acanthaceae). Aliso:

J Syst Evol Bot. 33:77-89. https://www.proquest.com/openview/6d9b5a3370030d589ffee31fd5e50865/1?pq-origsite=gscholar&cbl=18750

Després, L., L. Gielly, B. Redoutet, and P. Taberlet (2003). Using AFLP to resolve phylogenetic relationships in a morphologically diversified plant species complex when nuclear and chloroplast sequences fail to reveal variability. Mol Phylogenet Evol. 27:185-196. https://doi.org/10.1016/S1055-7903(02)00445-1

Figueroa, P., I. Gómez, L. Holuigue, A. Araya, and X. Jordana (1999). Transfer of rps14 from the mitochondrion to the nucleus in maize implied integration within a gene encoding the iron–sulphur subunit of succinate dehydrogenase and expression by alternative splicing. TPJ. 18:601-609. https://doi.org/10.1046/j.1365-313x.1999.00485.x

Fongod, A., N. Modjenpa, and M. Veranso (2013). Ethnobotany of Acanthaceae in the Mount Cameroon region. J. Med. Plants Res. 7:2859-2866. DOI: 10.5897/JMPR11.1267

Hadrys, H., M. Balick, and B. Schierwater (1992). Applications of random amplified polymorphic DNA (RAPD) in molecular ecology. Mol. Ecol. 1:55-63. https://doi.org/10.1111/j.1365-294X.1992.tb00155.x

Hajibabaei, M., G. A. Singer, P. D. Hebert, and D. A. Hickey (2007). DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. Trends Genet. 23:167-172. https://doi.org/10.1016/j.tig.2007.02.001

Hameed, A., M. Zafar, M. Ahmad, S. Sultana, M. Akhter, W. Zaman, S. Saqib, and F. Ullah. 2022. Micromorphology, phytochemical and pharmacological evaluation of *Isodon rugosus* (Wail. ex Benth.) Codd. JAPS. 32:736-745. http://doi.org/10.36899/JAPS.2022.3.0475

Hasan, A., M. Khan, and M. Ahmad (2007). Authenticity of folk medicinal plants of Pakistan. Tax Chem Meth. 1:1-5.

Hedrén, M., M. W. Chase, and R. G. Olmstead (1995). Relationships in the Acanthaceae and related families as suggested by cladistic analysis of rbc L nucleotide sequences. Plant Syst Evol. 194:93-109. https://doi.org/10.1007/BF00983219

Hooft, R. W., C. Sander, and G. Vriend (1997). Objectively judging the quality of a protein structure from a Ramachandran plot. Bioinform. 13:425-430. Hooft, R. W., C. Sander, and G. Vriend (1997). Objectively judging the quality of a protein structure from a Ramachandran plot. Bioinform. 13:425-430. https://doi.org/10.1093/bioinformatics/13.4.425

Huang, S., X. Ge, A. Cano, B. G. M. Salazar, and Y. Deng (2020). Comparative analysis of chloroplast genomes for five Dicliptera species (Acanthaceae): molecular structure, phylogenetic relationships, and adaptive evolution. PeerJ. 8:e8450. DOI.10.7717/peerj.8450

Juenger, T., T. Lennartsson, and J. Tuomi (2000). The evolution of tolerance to damage in Gentianella campestris: natural selection and the quantitative genetics of tolerance. Evol Ecol. 14:393-419. https://doi.org/10.1023/A:1010908800609

Khan, I., Z. K. Shinwari, N. B. Zahra, S. A. Jan, S. Shinwari, and S. Najeebullah (2020). DNA barcoding and molecular systematics of selected species of family Acanthaceae. Pak J Bot. 52:205-212. http://pakbs.org/pjbot/papers/1580200070.pdf

Kumar, S., G. Stecher, and K. Tamura (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 33:1870-1874. https://doi.org/10.1093/molbev/msw054

Labou, J. T., T. T. M. Tanekou, G. Simo, D. Kaba, S. Ravel, and F. Njiokou (2020). Single-strand conformation polymorphism (SSCP) of mitochondrial genes helps to estimate genetic differentiation, demographic parameters and phylogeny of Glossina palpalis palpalis populations from West and Central Africa. Infect Genet Evol. 82:104303. https://doi.org/10.1016/j.meegid.2020.104303

Laskowski, R., M. MacArthur, and J. Thornton (2006). PROCHECK: validation of protein-structure coordinates. https://doi.org/10.1107/97809553602060000882

Li, Z. Z., A. W. Gichira, J. K. Muchuku, W. Li, G. X. Wang, and J. M. Chen (2021). Plastid phylogenomics and biogeography of the genus *Monochoria* (Pontederiaceae). J Syst Evol. 59(5): 1027-1039. https://doi.org/10.1111/jse.12744

Loureiro, L. O., M. D. Engstrom, and B. K. Lim (2020). Single nucleotide polymorphisms (SNPs) provide unprecedented resolution of species boundaries, phylogenetic relationships, and genetic diversity in the mastiff bats (*Molossus*). Mol Phylogenet Evol. 143:106690. https://doi.org/10.1016/j.ympev.2019.106690

McDade, L. A., T. F. Daniel, and C. A. Kiel (200. Toward a comprehensive understanding of phylogenetic relationships among lineages of Acanthaceae sl (Lamiales). Am J Bot. 95:1136-1152. https://doi.org/10.3732/ajb.0800096

McDade, L. A., S. E. Masta, M. L. Moody, and E. Waters (2000). Phylogenetic relationships among Acanthaceae: evidence from two genomes. Syst

Bot.:106-121. https://www.jstor.org/stable/2666677

McDade, L. A., and M. L. Moody (1999). Phylogenetic relationships among Acanthaceae: evidence from noncoding trnL-trnF chloroplast DNA sequences. Am J Bot. 86:70-80. https://doi.org/10.2307/2656956

Morais, E. B., J. Schönenberger, E. Conti, A. Antonelli, and P. Szövényi (2019). Orthologous nuclear markers and new transcriptomes that broadly cover the phylogenetic diversity of Acanthaceae. Appl Plant Sci. 7:e11290. https://doi.org/10.1002/aps3.11290

Munir, F., S. M. S. Naqvi, and T. Mahmood (2013). In vitro and in silico characterization of *Solanum lycopersicum* wound-inducible proteinase inhibitor-II gene. Turk J Biol. 37:1-10. DOI.10.3906/biy-1111-23

Perveen, A., and M. Qaiser (2010). Pollen flora of Pakistan—LXVII: Acanthaceae. Pak J Bot. 42:175-191. https://www.pakbs.org/pjbot/PDFs/42(SI)/PJB42(SI)175.pdf

Popiela, A., and A. Molnár (2020). Recent Developments in Taxonomy and Phylogeny of Plants. Acta Soc Bot Pol. 89. https://doi.org/10.1038/s41598-022-24660-1

Porebski, S., L. G. Bailey, and B. R. Baum (1997). Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Mol Biol Rep. 15:8-15. https://doi.org/10.1007/BF02772108

Rahman, A., and M. I. A. Gulshana (2014). Taxonomy and medicinal uses on amaranthaceae family of Rajshahi, Bangladesh. Appl Ecol Environ Res. 2:54-59. DOI:10.12691/aees-2-2-3

Schwarzbach, A. E., and L. A. McDade (2002). Phylogenetic relationships of the mangrove family Avicenniaceae based on chloroplast and nuclear ribosomal DNA sequences. Syst Bot. 27:84-98. https://doi.org/10.1043/0363-6445-27.1.84

Scotland, R. W (1992). Systematics, similarity and Acanthaceae pollen morphology. Bot. J. Linn Soc. 109:529-541. https://doi.org/10.1111/j.1095-8339.1992.tb01449.x

Scotland, R. W., J. A. Sweere, P. A. Reeves, and R. G. Olmstead (1995). Higher-level systematics of Acanthaceae determined by chloroplast DNA sequences. Am J Bot. 82:266-275. https://doi.org/10.1002/j.1537-2197.1995.tb11494.x

Sheth, B. P., and V. S. Thaker (2017). DNA barcoding and traditional taxonomy: an integrated approach for biodiversity conservation. Genome. 60:618-628. DOI: 10.1139/gen-2015-0167

Song, Y.-G., Y. Li, H.-H. Meng, Y. Fragnière, B.-J. Ge, H. Sakio, H. Yousefzadeh, S. Bétrisey, and G. Kozlowski (2020). Phylogeny, Taxonomy, and Biogeography of Pterocarya (Juglandaceae). Plants. 9:1524. https://doi.org/10.3390/plants9111524

Tamura, K., and M. Nei (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 10:512-526. https://doi.org/10.1093/oxfordjournals.molbev.a040023

Tamura, K., M. Nei, and S. Kumar (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. PNAS. 101:11030-11035. https://doi.org/10.1073/pnas.0404206101

Tripp, E. A., T. F. Daniel, S. Fatimah, and L. A. McDade (2013). Phylogenetic relationships within Ruellieae (Acanthaceae) and a revised classification. Int J Plant Sci. 174:97-137. https://doi.org/10.1086/668248

Ullah, F., Y. Gao, İ. Sari, R.-F. Jiao, S. Saqib, and X.-F. Gao (2022). Macro-Morphological and Ecological Variation in Rosa sericea Complex. Agron. 12:1078. https://doi.org/10.3390/agronomy12051078

Wali, S., F. Munir, and T. Mahmood (2013). Phylogenetic studies of selected Citrus species based on chloroplast gene, rps14. Int J Agric Biol. 15:357-361. Doi.12–557/2013/15–2–357–361

Zaman, W., J. Ye, M. Ahmad, S. Saqib, Z. K. Shinwari, and Z. Chen (2022). Phylogenetic exploration of traditional Chinese medicinal plants: a case study on Lamiaceae (angiosperms). Pak J Bot. 54. http://pakbs.org/pjbot/papers/1646205657.pdf

Zhang, J., and J.-R. Yang. 2015. Determinants of the rate of protein sequence evolution. Nat Rev Genet. 16:409-420. https://doi.org/10.1038/nrg3950