



NexBioMed: Journal of Biological, Biomedical, Public Health & Nutrition Sciences

To Cite:

Ishaq, F., Gul, S., Ashfaq, S., & Zaib, U. (2025). Whole Chloroplast Genome and Sequences Variation in the Family Lamiaceae. *NexBioMed: Journal of Biological, Biomedical, Public Health & Nutrition Sciences*, Vol. 107, Issue 1, pp. 14110-14133.
<https://doi.org/10.5281/zenodo.16418754>

Author Affiliation:

¹Department of Biology, Faculty of Biological and Biomedical Sciences, The University of Haripur, Haripur Khyber Pakhtunkhwa, Pakistan.

Email: ishaqfatima21111@gmail.com

*Corresponding author

¹Department of Biology, Faculty of Biological and Biomedical Sciences, The University of Haripur, Haripur Khyber Pakhtunkhwa, Pakistan.

Email: ishaqfatima21111@gmail.com (F.I)

Peer-Review History

Received: 12th June, 2025

Reviewed & Revised: 13th June/2025 to 22nd/July/2025

Accepted: 23rd July 2025; Published: 25th July 2025

Peer-Review Model

External peer-review was done through double-blind method.

URL:

<https://sites.google.com/view/sherwanjournals/nexbiomed>

Ethics

No animal studies are presented in this manuscript.

No human studies are presented in this manuscript.

No potentially identifiable human images or data is presented in this study.



License: Creative Commons Attribution 4.0 International



NexBioMed

Journal of Biological, Biomedical
Public Health & Nutrition Sciences

Sherwan Institute of Online Education

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of Sherwan Journals and/or the editor(s). Sherwan Journals and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

NexBioMed: Journal of Biological, Biomedical, Public Health & Nutrition Sciences, 107(1), 14110-14133 (2025) 14110 of 14133

WHOLE CHLOROPLAST GENOME AND SEQUENCE VARIATION IN THE FAMILY LAMIACEAE

Fatima Ishaq^{1*}, Shumaila Gul¹, Saleha Ashfaq¹, Umer Zaib¹

¹Department of Biology, Faculty of Biological and Biomedical Sciences,
The University of Haripur, Haripur
Khyber Pakhtunkhwa, Pakistan.

E-mail: ishaqfatima21111@gmail.com

DOI: <https://doi.org/10.5281/zenodo.16418754>

ABSTRACT

The Lamiaceae family, a large and cosmopolitan group comprising 236 genera and over 7,000 species, is primarily found in the Mediterranean basin and Central Asia. This family is characterized by bilabiate corollas, epipetalous or didynamous stamens, verticillaster flowers, and schizocarp fruits. In this study, we analyzed the complete chloroplast (cp) genome sequences of 51 species to explore the evolutionary trends and phylogenetic relationships within the Lamiaceae family. The complete chloroplast genome is valuable for studying sequence divergence, phylogenomics, and molecular evolution. We characterized the chloroplast genome structure, including gene number and order, simple sequence repeats (SSRs), and the variety and distribution of large repeat sequences. The cp genome sizes of Lamiaceae species ranged from 149,095 bp (*Elsholtzia densa*) to 154,495 bp (*Vitex agnus-castus*). The large single copy (LSC) region varied from 81,610 bp (*Alyssum gmelinii*) to 84,907 bp (*Lamium album*), the small single copy (SSC) region from 17,621 bp (*Ajuga nipponensis*) to 26,234 bp (*Origanum vulgare*), and the inverted repeats (IRs) ranged from 26,051 bp (*Origanum vulgare*) to 26,890 bp (*Salvia officinalis*). These genomes consist of circular DNA molecules, similar to those found in most angiosperms. The complete cp genomes of Lamiaceae species contain between 80 and 134 functional genes, including 84–89 protein-coding genes, 8 rRNA genes, and 35–39 tRNA genes. Among the 50 plastomes, the highest GC content (38.7%) was observed in *Stachys byzantina* and *Leonotis leonurus*, while the lowest GC content (37.7%) was found in *Coleus hadiensis*, *Coleus xanthanthus*, and *Coleus scutellarioides*. We identified three main types of repeat sequences: dispersed, palindromic, and tandem. A total of 5,349 repeat elements were identified, with palindromic repeats being the most abundant (2,895; 55%), followed by tandem repeats (1,528; 28%) and dispersed repeats (926; 17%). Ultimately, the phylogenetic analysis revealed that the 50 Lamiaceae species are closely related, providing critical insights into the sequence variation, phylogenetic connections, and molecular evolution of Lamiaceae plastid genomes.

Keywords: Lamiaceae family; Sequence variation; Repeat sequence; Phylogenetic relationship.

1. INTRODUCTION

The Lamiaceae family is one of the most important families containing volatile oils and was previously called Labiatae, or the mint family, a cosmopolitan distribution and includes over 236 genera and 7000 species (Hamed et al., 2021). It is a large natural distribution, a family

which comprises aromatic herbs, but also includes shrubs and herbs, trees are extremely rare. Often recognized by serrate opposite leaves, square stems, and monosymmetric flowers (do Rocio Duarte & Carvalho de Souza, 2014). It has known for their aromatic oils, which played an undeniably significant role within culinary, medicinal, and horticultural aspects of human history. Species of *Lamiaceae* are economic importance as sources of wood, landscape ornamentals, cosmetics, culinary herbs, and medicinal herbs. (Zhao et al., 2021; Van Wyk, 2014). Mint herbs consist of perennial aromatic members of the *Lamiaceae* family, the *Mentha* L. genus, classified into 42 species, 15 hybrids, and hundreds of subspecies, varieties and cultivars. Among the best-known mint species, we find watermint (*Mentha aquatic* L.), spearmint (*Mentha spicata* L.), wildmint (*Mentha arvensis* L.), horsemint (*Mentha longifolia* (L.) L.), pennyroyal (*Mentha pulegium* L.), and peppermint (*Mentha x piperita* L.), the last of which is a natural sterile hybrid of watermint and spearmint (Silva, 2020). The mint family (*Lamiaceae*) is one of the largest families on earth. When analyzing the use of species of the *Lamiaceae* family in the national economy, the following was revealed: 143 species out of 238 species are with honey-juce (60%), 69 medicinal species (29%), 100 species with essential oil (42%), 27 decorative (11.3%), 19 food made (8%). It is known that 74 types of plants belonging to the family are not used yet (Yusupova & Baratjon, 2022).

1.1. Family *Lamiaceae*

During the past decade, the number of molecular phylogenetic studies on plants has increased at a rapid pace. Advances in comparative DNA sequencing and methods for phylogeny reconstruction have made it possible to investigate a broad range of systematic and evolutionary questions for large numbers of taxa. One important result of this has been that molecular phylogenetic studies have sometimes overturned relationships suggested from traditional classifications (Harley et al., 2004). A close yet discrete relationship with *Verbenaceae* was also suggested by the early system builders. However, modern phylogenetic analyses of morphological data have suggested that *Lamiaceae* was polyphyletically derived within *Verbenaceae* (Abu-Asab & Cantino, 1987). Phylogenetic conclusions were later corroborated with comparative DNA sequence analysis. The complete chloroplast DNA sequences of closely related species confides several evolutionary hotspots region for mutations in the whole chloroplast genomes. Phylo-genomics study provides an excessive ability determine historically severe issues in phylogeny by decreasing sampling mistake (Langmead et al., 2009). Plastid genome is identified in the plant phylogeny, evolution, and divergence of a species. Some works supported that phylogenetic analyses not only determine the previously discussed phylogeny but also increase accurate phylogenetic tree. Nowadays, such type of studies is essential to point out the difference between various tree-building methods used for phylogenetic evaluations based on systematic errors. However, the systematic mistake will be eliminated by improving the dataset, which leads to improving the size of data (Ahmad et al., 2010).

1.2. Whole Chloroplast Genome

The complete chloroplast genomes typically have a quadripartite structure that is 120-160 kb in size and 110-130 unique genes. These genes are divided into four regions, the large single copy (LSC) region and the small single copy (SSC), and two inverted repeats (IRa and IRb). The majority of genes are involved in photosynthesis, translation, and transcription (Ahmad et al., 2021). Most angiosperms have quadripartite, double-stranded chloroplast genomes that are circular in shape. These characteristics, along with slow progression and uniparental inheritance make the chloroplast genomes an excellent runner for phylogenetic inference that ranges from profound divergence to

population genetics [12]. Chloroplasts are a type of cellular organelle in the plastid genome that are also found in some protists, algae, higher and lower plant species. Plastids also contain several metabolic reactions, their own genomes, and a huge number of nuclear genome-encoded proteins (Sim et al., 2015).

Chloroplasts are vital organelles of photosynthetic algae and plants. They are mostly inherited from the maternal parent (Bi et al., 2018). Green plants' chloroplasts, which participate in the production of starch, pigments, amino acids, and fatty acids because they have their own genetic material called the chloroplast genome (cpDNA) which play a vital role in photosynthesis, development, and physiology of plants (Biscotti et al., 2015). Because of their conserved structure and comparatively high replacement rate, plastomes are a significant source of genetic data for studies on species identification, phylogeny, biology, and photosynthetic gene degradation. They are also conserved in size, gene content, gene order, gene structure, and genome (Yu et al., 2022). Furthermore, the chloroplasts contain valuable information and have been used as ideal research models, particularly for molecular markers, barcoding identification, plant phylogenetics, evolution, and comparative genomic studies. The chloroplasts support plant growth and development by converting energy into carbohydrates through photosynthesis (Yuan et al., 2014)]. Additionally, the plastome products of metabolic processes are important for the relationship of trees with their environment. Chloroplasts, therefore, offer metabolism hubs in cellular operations to deliver cues and return via a decreasing signal (Yun & Kim, 2022). Plastids are well known for their crucial role in photosynthesis, but they also play crucial roles in other aspects of plant physiology and development (Zhang et al., 2019). A potential source of data for the evolutionary reconstruction of plant species relationships is the chloroplast genome's extremely conserved structure (Zhou et al., 2018). Due to its uniparental inheritance, haploid origin, highly preserved features, and slower evolutionary rate of change compared to nuclear genomes, the cp genome is regarded as the "workhorse" in plant systematics study (Zhu et al., 2021).

For comparative genome evolution, phylogenetic analysis, and population studies, high-copy organelle sequencing is a crucial tool and one of the most technically accessible areas of the genome due to its sequence conservation. The plastid is the perfect target for universal primers due to recombination and low amounts of nucleotide substitution. Plastid investigation offers insights into important metabolic pathways and cellular processes more generally because the plastid comprises a core set of genes for photosynthesis, protein synthesis, and ribosome development (Wang et al., 2015). Many of the plastome regions have been established as DNA barcode marker organizations to recognize taxa. These whole genome sequences are broadly used for genome evolution, DNA barcoding, conservation of species, and molecular identification of higher plants. The amount of full plastome sequences as a result of advancements in genome sequencing based on next generation sequencing (NGS) machineries and bioinformatics implementations public databases has increased. Moreover, the result based on high-resolution of phylogenetic investigation and species identification can use the whole genome as a super marker (Wang et al., 2023).

It has been demonstrated that the chloroplast genome sequences deliver a useful and perceptive source of cp DNA barcode for assessing genetic variation. Because of their relatively stable genome structure, gene contents, and gene order, which has many features due to its preserved structure, maternal characteristics. regarded as an informative and valuable data source for understanding the evolutionary studies of plastid genome sequences (Wheeler et al., 2014).

1.3. Repeat Sequence Analysis

In the plastome, repeated DNA sequences are crucial in cp genome reorganization, and also beneficial for evolutionary studies (Wicke et al., 2011). These repetitive sequence are considered a homologous DNA fragments which are present in several copies of whole genome. In addition, in plants about ninety percent of the repeated sequence of the total genome mass were present in chloroplast. Therefore, repetitive arrangements are being referred to be selfish elements or junk DNA. Repetitive elements motifs repeating units or repeat sequence in the whole genome are repeated hundreds or thousands of time which are abundant in all eukaryotic genomes, representing more than half number of the total content of the DNA in the cell (Williams et al., 2015). The key mechanism and the evolutionary apparatus of repetitive arrangements are concerned many developments such as chromosome movement, recombination, gene association with mitotic spindle, and pairing of chromosome, centromeric condensation, sister chromatid pairing, interaction of chromatin proteins, histone binding, determination of genome structure, regulation of gene expression and physiological changes reaction by genome to environmental stimuli (Wu et al., 2018). The distribution of genomic association perception into the organization, behavior, evolution and important potential of repetitive sequences in genomes of eukaryotic genomes, to understand more information on primary structure, evolution, molecular organization and purpose of repetitive sequences. Moreover, also with recent genome sequence technologies produce extraordinary rich information about origin, diversity and impact on genomic repetitive sequences (Xu et al., 2021)]. Repeat sequence analysis is also probably essential for ecological investigation of plants and regularly contributed in the plastomes variations in various cellular functions with evolution of genes, gene mobility and RNA editing (Yan et al., 2019).

1.4. Phylogenetic Analysis

Phylogenetics is the study of the evolutionary history and relationships among individuals and groups of organisms. The phylogenetic tree displayed evolutionary relationships among different biological species based on similarities and differences for their maternal characteristics [30]. The evolutionary links between individuals or groups of organisms are reflected in phylogenetic trees which are used in various taxonomical level of plants species based on resemblances and variances in maternal characteristics among biological entities. of species, individual or genes to show the evolutionary pathway and connection among organisms [31]. Generally, the chloroplast genome sequence undertaking the most important component of knowledge for plant phylogeny, widely due to the affluence with which plastome genes can be accumulated, sequenced, and examined and finally resolved by the true evolutionary relationships and phylogenetic lineages [32]. In addition, Chloroplast genes have been utilized in the majority of large-scale phylogenetic analyses of plants, and the derived phylogenies have been constructed in numerous additional comparative analyses looking at ancestral states, historical biogeography, and other evolutionary patterns [33]. It has been demonstrated that plastid genome are effective tools for examining multifarious relations in the plants and its life cycle. Additionally, they have provided on the molecular evolutionary patterns linked to gene duplication, deletion, and rearrangement; in some circumstances, these structural alterations are phylogenetically useful. Further, in clarifying the links of extremely obstinate lineages, such as those that have seen significant evolutionary radiations [34]. Plastid regions have substantial utility in phylogenetic inference, but whole chloroplast genomes have recently been used for improved inference of relationships among species. The identification of effective plastid genetic markers for plant species phylogeny and classification needs extensive knowledge of the whole

plastome genome. The phylogenetic tree, which has been a crucial component of evolutionary studies in genome sequencing, has a substantial performance inconsiderate molecular divergence [35]. Because of the chloroplast's small size, conserved nature, predetermined gene organization, and favorable capacity for transgenic expression, it is crucial to understand its molecular genetics evolution. As a result, the sequences of the chloroplast have contributed to the molecular systematics and classification of plants [36].

The results of many studies in comparison because of phylogenetic trees and data are stored in consistent and compatible forms that contain trees with analysis outcomes that are frequently matched with one another which shows close relation [37]. A broad variety of DNA markers have been found in the three plant genomes (nuclear, mitochondrial, and chloroplast), which are commonly employed for molecular identification at various taxonomic levels. DNA-based markers have proven to be particularly useful for identifying species, however, it is frequently required to incorporate additional sequences for better identification resolution in various research situations [38]. On this family, numerous studies take attentive, particularly on phylogenetic relationships and the timing of diversification events. However, comparative analysis has fully utilized all available data to produce a fully resolved phylogenetic tree that includes all currently existent species. Similar to this, extensively detailed and broadly dated phylogenetic relationships will allow us to close significant gaps in our knowledge of the evolutionary history of Lamiaceae species [39]. Numerous biological disciplines, such as comparative genomics, epidemiology, and microbiome, frequently employ for reconstructing of phylogenetic tree models. For hundreds of samples to be used in the reconstruction of a phylogenetic tree representative the evolutionary relationships among communities made up of thousands of species, assimilating and imagining phylogenetic trees with multidimensional associated data groups helps to recognize patterns and produce new postulates, the attributes on phylogenetic tree's exterior panels and aids users in exploring and contrasting various heterogeneous data sets in an evolutionary setting [40].

1.5. Objectives of the Study

The study has the following research objectives, i.e.,

- To Investigate the sequence variation of 50 species of Lamiaceae family
- To find out the gene order, gene content and repetitive sequence in whole plastid genomes of family lamiaceae.
- To reconstruct phylogenetic tree model of whole chloroplast genome of family

Most angiosperm chloroplast DNAs are maternally inherited and do not recombine, which makes them excellent for analyzing evolutionary interactions among species, particularly related species. The whole chloroplast genome and its protein-coding genes are crucial to resolving the evolutionary relationships and phylogenetic lineages among species to reconstruct the phylogeny of different entities.

2. REVIEW OF LITERATURE

Tamboli et al., (2015) focused on two Lamiaceae family members: *Arabis stelleri* and *A. takesimana*. *Arabis stelleri* is widely distributed in East Asia, and *A. takesimana* is a Korean endemic plant. Following the sequencing of their entire chloroplast genomes, which revealed a similar gene content and organization, they were assigned to the *Arabis* clade, show a close relationship to *A. flagellosa*, *A. nipponica*, *A. paniculata*, and *A. hirsuta*. For future research on the genetic diversity and

evolutionary history of these species, this work offers invaluable information. These species' evolutionary place within the Arabis clade has been established, and research has been done on how they relate to other Lamiaceae members. Okan et al., (2024) used nuclear ribosomal DNA and chloroplast DNA(trnL intron, trnL-F, rbcL and trnQ-rps16 DNA)sequences for phylogenetic analysis on Turkish endemic Bornmuellera species, and the results showed that the genus is monophyletic. This is in line with the results of German (2009), who also established the monophyly of specific tribes and genera within the family using nuclear ribosomal DNA sequences. These investigations give the evolutionary study of Bornmuellera species a more comprehensive framework. Zhang et al., (2024) implanted research on *Eutrema deltoideum* (Hook. f. et Thoms.), is a potentially significant vegetable and medicinal resource. After being thoroughly examined, the 154,051 bp chloroplast genome of *Eutrema deltoideum* was found to have 132 whole genes, including 87 genes that code for proteins, 8 ribosomal RNA genes, and 37 tRNA genes. Phylogenetic study confirms that the *Eutrema* genus is monophyletic, placing *E. deltoideum* and *E. heterophyllum* in close relationship. This data improves our comprehension of *E. deltoideum*'s taxonomy classification.

Zhao et al., (2024) studied Kohlrabi Speice, is a significant swollen-stem cabbage variety that is a member of Lamiaceae family. Few whole chloroplast genome sequences for this genus have been reported in this study. According to the base composition study, 36.36% of the whole chloroplast genome sequence was made up of GC. Most codons with values larger than one finished with A or U, but most codons with values less than one concluded with C or G, according to a relative synonymous codon use frequency (RSCU) study. The majority of the 35 scattered repetitions that were found were dispersed throughout the large single-copy (LSC) area. 188 of the 290 simple sequence repeats (SSRs) that were discovered were dispersed across the LSC region. Analysis of phylogenetic relationships showed that five *Brassica oleracea* subspecies were clustered into one group and the kohlrabi chloroplast genome was closely related to that of *B. oleracea* var. *botrytis*. Results provide a basis for understanding chloroplast-dependent metabolic studies and provide new insight for understanding the polyploidization of Lamiaceae species. Li et al., (2024) carry out the study on *Capsella bursa-pastoris*, *Descurainia sophia*, and *Lepidium apetalum*. The genus *Capsella* Medik. belongs to the tribe Camelinae and comprises seven species and one hybrid . *Capsella bursa-pastoris* (L.) Medik. is a weed in winter fields that has developed a high level of resistance to the herbicide acetolactate synthase, which has led to a significant loss of wheat yield in China. The genus *Descurainia* Webb & Berthel. belongs to the tribe Descurainieae and includes 47 accepted species. *Descurainia sophia* (L.)Webb ex Prantl is also a notorious weed in China in the winter field, as it has developed less resistance to tribenuron-methyl . *Lepidium* L. belongs to the tribe Lepidieae and is among the largest genera of Lamiaceae comprising 263 accepted species. *Lepidium apetalum* Willd. is an annual or biennial weed widely distributed in Asia and Europe. Among these species, the seeds of *D. sophia* and *L. apetalum* are used in traditional Chinese medicine for relieving asthma, clearing the lungs, reducing swelling, and promoting diuresis, whereas *C. bursa-pastoris* is considered an adulterant affecting medicinal efficacy. A series of studies have contributed to our understanding of chloroplast genomes in the Lamiaceae family. provided a comprehensive analysis of the chloroplast genomes of *Capsella bursa-pastoris*, *Descurainia sophia*, and *Lepidium apetalum*, revealing their quadripartite structure and unique gene composition. We used chloroplast DNA variation to explore the evolutionary relationships within the *Brassica* and allied genera. further expanded this knowledge by analyzing the chloroplast genomes of *Brassica nigra* and *B. oleracea*, providing insights into their genetic diversity and phylogenetic relationships. This studies collectively contribute to our understanding of chloroplast genome structure and evolution in the Lamiaceae

family.

Al-Shehbaz & Marhold (2024) led research on *Rorippa* Scop belongs to the tribe Cardamineae, and approximately 75 species within this genus are distributed worldwide. In South Korea, six species have been recorded, including *Rorippa apetala* Y.Y. Kim & B.U. Oh, a species endemic to the Korean peninsula. Compared to its relative species, *R. apetala* is characterized by a pinnatisect basal leaf, globose to subglobose ovary, and having no or rarely 1e4 petals, and empty or rarely 1e2 seeds per locule. Unfortunately, the number of individuals has decreased, and this is rare in their natural habitat. It cannot circumvent regions affected by environmental factors or anthropogenic activities. Here, we have effectively assembled the whole chloroplast genome of the Korean Peninsula-endemic species *Rorippa apetala*. The 154,818 bp long chloroplast genome has a normal quadripartite structure with two inverted repeats (26,474 bp each) and large single-copy and tiny single-copy sections. A total of 129 genes—37 tRNA genes, 8 rRNA genes, and 84 protein-coding genes were encoded. A tight connection between *R. apetala* and *Rorippa indica* (L.) Hiern was found by phylogenetic analysis based on the concatenated sequences of 78 protein-coding genes from 25 species. Further research with more illuminating markers and a wider variety of species will yield important insights into the biogeographical distribution of the genus *Rorippa* and offer a deeper understanding of the speciation process in *R. apetala*. Feng et al., (2024) conducted a research on the subfamily Polygonoideae, which encompasses a wide range of horticultural and medicinal plants, with the goal of elucidating its taxonomy and evolutionary links. The chloroplast genomes of certain plants in this subfamily were distinguish between different clades and possible molecular markers. By showing a multi-gene phylogeny of the pantropical subfamily Chrysophylloideae and demonstrating the polyphyletic character of several taxa, we made a significant contribution to this subject, and built a strong phylogeny of the subfamily Polygonoideae utilizing nuclear genes and plastome sequence variation, building on earlier work by defining several clades and providing insight into the subfamily's evolutionary diversity. When taken as a whole, these studies offer insightful information on the evolutionary history and classification of the subfamily Polygonoideae. Ren et al., (2024) studied the genus *Rorippa*, member of the family Lamnaceae typically have significant therapeutic significance. With the exception of Antarctica, the genus, which has about 75 species, is mostly found in the Northern Hemisphere. Here, we used Illumina paired-end sequencing to sequence all four of the *Rorippa* species' entire plastid genomes. The overall size of the four newly discovered *Rorippa* plastid genomes varied. The four plastomes contain 130 genes total, including 85 protein-coding genes, 37 tRNA, and 8 rRNA. Combining with six plastid genomes that have been released, this study conducted comparative as well as phylogenetic studies. In terms of gene number and order, overall size, genomic organization, codon use, long repeat sequence, and SSR, we discovered that the ten *Rorippa* plastid genomes were conservative. In order to differentiate *Rorippa* plants, fourteen mutational hotspot sites might be chosen as potential DNA barcoding candidates. Based on plastomes and nrDNA ITS sequences, the phylogenetic trees unmistakably showed that 10 *Rorippa* species showed monophyletic connections within the tribe Cardamineae. On the other hand, both the interspecific and intergeneric connections within *Rorippa* and its allied genera exhibit notable cytonuclear discordances. We deduced that interspecific hybridization within *Rorippa* and intergeneric hybridization with similar genera are most likely the causes of the cytonuclear discordance. These plastid genomes can provide essential data for research.

Wang et al., (2023) worked on Chinese kale is a commonly grown plant belonging to the genus *Brassica*. While the origins of *Brassica* have been thoroughly investigated, the origins of Chinese kale are still unknown. Chinese kale originates in southern China, as opposed to *Brassica oleracea*, which

originated in the Mediterranean area. The chloroplast genome's high conservatism makes it a popular choice for phylogenetic study. The chloroplast genomes of the white-flower Chinese kale (*Brassica oleracea* var. *alboglabra* cv. *Sijicutiao* (SJCT)) and yellow-flower Chinese kale (*Brassica oleracea* var. *alboglabra* cv. *Fuzhouhuanghua* (FZHH)) were amplified using fifteen pairs of universal primers using PCR. Eight additional Lamiaceae members and the two Chinese kale types' chloroplast genomes were examined. DNA barcodes were examined for simple sequence repetitions, lengthy repeats, and variable areas. Despite the little variations that were found, an examination of inverted repeat boundaries, relative synonymous codon use, and synteny showed that the ten species shared a great deal of commonality. Based on phylogenetic study and Ka/Ks ratios, Chinese kale is likely a *B. oleracea* variation. The phylogenetic tree demonstrates that *B. oleracea* var. *oleracea* and Chinese kale cultivars were grouped together in a same group. According to the study's findings, Chinese kale with white and yellow flowers belong to a same phyletic group, and their variations in bloom color developed throughout time. Yang & Zhou (2023) carry out study on Subtribe Swertiinae, a medicinally significant and highly speciose Subtribe of family Gentianaceae, includes approximately 539–565 species, which are distributed in alpine, temperate and alpine regions around the world but rarely in tropical and subtropical regions at low latitudes. East Asia and North America are the centers of diversification of this subtribe, with 137 species of 11 genera in China. Studies have utilized chloroplast genomes to explore the group's genetic characteristics, revealing a conserved structure and gene content, as well as the potential for mutation hotspot regions to serve as molecular markers. However, the relationships within the subtribe remain controversial, with some genera being non-monophyletic. The use of chloroplast genomes has also provided insights into the evolutionary history of *Swertia* L. species, highlighting the need for further exploration of their systematic positions. The phylogenetic analysis has shown that the subtribe Swertiinae forms a monophyletic clade, with some genera being non-monophyletic. These findings underscore the importance of continued research to clarify the relationships within this subtribe.

Shang et al., (2023) performed work on the *Ajuga* genus is a member of the Lamiaceae family and encompasses numerous economically and medically significant plants. Many *Ajuga* species have been employed in Traditional Chinese Medicine for relieving cough, reducing sputum, and arresting bleeding. Among the *Ajuga* species, *A. bracteosa* is particularly prevalent and extensively utilized in folk medicine. The potential contamination of herbal medicinal products, such as those from the *Ajuga* species, is a significant concern for consumer health. To address this, a range of molecular markers, including chloroplast (cp) genomes and protein-coding genes (PCGs), have been identified as reliable tools for species identification and phylogenetic analysis. These markers have been particularly useful in distinguishing *A. bracteosa* from its contaminant. However, the use of DNA barcoding, such as the *matK* and *rbcL* regions, has also been proposed as a valuable tool for identifying medicinal plant species. The incorporation of DNA sequence data with other taxonomical data has been suggested to enhance the accuracy of species identification. The repetitive sequences, codon uses, and cp genomes of seven species were highly conserved, and PCGs were the reliable molecular markers for investigating the phylogenetic relationship within the *Ajuga* genus. Moreover, four mutation hotspot regions (*accD-psaI*, *atpH-atpI*, *ndhC-trnV(UAC)*, and *ndhF-rpl23*) were identified within cp genomes of *Ajuga*, which could help distinguish *A. bracteosa* and its contaminants. Based on cp genomes and PCGs, the phylogenetic tree preliminary confirmed the position of *Ajuga* within the Lamiaceae family. Molecular clock analysis estimated the divergence time of *Ajuga* to be around 7.78 million years ago. Zhu et al., (2021) revealed that antibacterial properties of seed oil, *Eruca sativa* Mill. is a significant food vegetable that may also be used

medicinally. Here, a mix of long PacBio reads and short Illumina reads was used to reconstruct the whole chloroplast (cp) genome of *E. sativa* from scratch. The 153,522 bp *E. sativa* cp genome was organized into a quadripartite structure. It was made up of two big single-copy areas, measuring 83,320 bp and 17,786 bp, respectively, and two tiny single-copy regions, measuring 17,786 bp. These regions were divided by two inverted repeat (IRa and IRb) regions. The *E. sativa* included 69 basic sequence repeats and 49 lengthy repeating sequences Genome of *E. sativa* cp. The *E. sativa* cp genome's codon use study revealed a bias in favor of codons that end in A/T. Comparing the *E. sativa* cp genome to other Lamiaceae cp genomes, similar traits were observed in terms of size, gene content, and structural region linearity. Furthermore, with the exception of *ycf1*, *ycf2*, and *rps12*, the examination of the synonymous (Ks) and non-synonymous (Ka) substitution rates showed that purifying selection pressure was typically applied to protein-coding genes. According to a phylogenetic study, *E. sativa* shares evolutionary relationships with several significant Brassica species, suggesting that advantageous *E. sativa* genes might be passed on to other Brassica species. Our findings will contribute to the advancement of *E. sativa* breeding and genetic enhancement, as well as offer useful information for *E. sativa* utilization as an important resource to improve other Brassica species.

Du et al. (2020) led the research on yellow mustard (*Sinapis alba* L.), a member of the broad Lamiaceae family, has been exploited as a valuable gene pool for the genetic enhancement of cash crops in the Lamiaceae family. Knowledge of the evolutionary relationships among other Lamiaceae crops and *Sinapis alba* (*S. alba*) might help guide the introgression of its advantageous genes into related species. For evaluating genome evolution and the evolutionary connections of intricate angiosperm families, the chloroplast (cp) genome is a perfect example. A 153,760 bp gapless quadripartite cycle including two inverted repeats (IRa and IRb) separated by a large single copy (LSC) a small single copy (SSC) region. In this cp genome, 78 protein-coding genes, 30 tRNA genes, 4 rRNA genes, and 89 simple sequence repeat (SSR) loci of 18 different kinds were found. The Leu codon with the A/U ending is preferred, according to the codon use study. In addition to showing a strong link with significant Brassica and Raphanus species, the phylogenetic analysis utilizing 82 Lamiaceae species revealed that *S. alba* most likely emerged from a different evolutionary pathway than the congeneric *Sinapis arvensis*. According to the synonymous (Ks) and non-synonymous (Ks) substitution rate study, the purifying selection pressure was lowest for genes encoding "Subunits of cytochrome b/f complex," whereas those linked to Subunits of NADH-dehydrogenase, acetyl-CoA subunit, and maturase were subjected to comparatively stronger purifying selection pressures. Our findings offer important insights on how to make the most of the *S. alba* cp genome as a prospective genetic resource for the enhancement of Brassica and Raphanus species.

3. MATERIAL AND METHODS

3.1. Plastid Genome Dataset

The whole plastid genome dataset of Fifty species of family Lamiaceae including out group species, *Berberis bealei*, downloaded from NCBI (<https://www.ncbi.nlm.nih.gov/>). The whole chloroplast genome of every species were added to the Geneious R Software, Auckland, New Zealand) and then find the length, total genes and percentage of GC contents.

3.2. Whole Chloroplast Genome Sequencing: Annotations Analysis

The insertion and deletion of annotation Geneious software R software version., Auckland, was used

which detected chloroplast genomic data produces a consensus sequence. The stop and start codons are manually adjusted in Geneious R v 2023.2.1. All whole chloroplast genome of family Lamiaceae we constructed the four regions in chloroplast the large single copy (LSC), small single copy (SSC) region and inverted repeats (IRa) region with the help of mauve alignment tool in Geneious software R V 2023.2.1.

3.3. Repeat Sequence Analysis

The REPuter online server (<https://bibiserver.cebitec.unibielefeld.de/reputer>) to investigate the repeat sequences analysis in 50 chloroplast genome, we found three types of repeats: dispersed, tandem and palindromic. The following parameters were used to position of dispersed and palindromic repeating sequences: (1) sequence identity = 90%; (2) Hamming distance = 3; (3) minimal repeat size = 30bp; (4) maximum computed repeats = 50. The online Tandem Repeat Finder program (<https://tandem.bu.edu/trf/trf.html>) was used to examine the tandem repeats motifs, with parameters for alignment match, mismatch and minimum alignment score repeats were 80, and maximum period size was (500).

3.4. Simple Sequence Repeats Analysis SSRs

The Perl script MISA program (<http://pgrc.ipk-gatersleben.de/misa/>) was used which detected the simple repeat sequence in the plastid genome of Lamiaceae dataset of 50 species with motifs between 1-6 nucleotide and number of repetition units mono-, di-, tri-, tetra-, penta-, and hexanucleotides respectively.

3.5. Phylogenetic Analysis

Whole genome sequence data set of 50 species of the Lamiaceae were used to reconstruct the phylogenetic. These plastid chloroplast DNA sequence separately aligned using MAFFT v through in Geneious software. We also added *Berberis bealei* (NC_022457) cp genome as an outgroup. After, these nucleotide sequence was further aligned in cluster W method using the MEGA v 7.0.18 software. However, we evaluated inferred evolutionary tree were performed using neighbor joining setup. Further, one hundred (100) bootstrap was used to estimate the relationship among all species. Lastly, the output was assessed through construction of phylogenetic tree.

4. RESULTS AND DISCUSSIONS

4.1. The Characteristics of Complete Chloroplast Genomes in Family Lamiaceae Species

The complete chloroplast genomes of Fifty Lamiaceae species ranged in size from 154,495 (*Elsholtzia densa*) to 149,095 bp () (see, Table.1). These plastid DNA contains distinctive quadripartite circular structure, comparable to those in angiosperm species. The complete genomes of Lamiaceae species comprised of a large single copy (LSC) region ranged from 81,610 (*Alyssum gmelinii*) to 84,478 bp (*Arabidopsis pedemontana*), and a small single copy (SSC) region ranged from 17,631 (*Cakile edentula* subsp. *lacustris*) to 18,152 bp (*Cochlearia pyrenaica*), and inverted repeats (IRs) ranged from 26,051 (*Eutrema yunnanense*) to 26,490 bp in two species (*Camelina hispida*; *Camelina sativa*) in size (Table 1). Furthermore, the whole GC content of the plastomes ranged 38.7%-37.7% in which three species (*Coleus xanthanthus*, *Coleus scutellarioides*) have the lowest GC content and the highest GC content in two species (*Leonotis leonurus*, *Stachys byzantina*). The family Lamiaceae complete cp genomes encoded of 80-134 functional genes, in which contain 84-89 protein coding genes, with 6-8

rRNA,35-39 tRNA respectively. Lamiaceae species complete cp genome consisted of 123 functional genes, with 37 tRNA, 8 rRNA and 88 protein-coding. Among 134 genes, 11 genes for small ribosome subunits, 9 genes for large ribosome subunits, 4 genes for DNA-dependent RNA polymerase subunits and 50 genes fragments were related to self-replication. The translational initiation factor (infA) gene, 38 genes for photosynthesis, 6 genes for ATP synthesis, and 11 genes encoding subunits of photosystem I.

4.2. Repeat Sequence Variations and Genome Structure Comparison

In this study, we calculated three types of repetitions, i.e. dispersed, palindromic and tandem repeats. Among these repeat variations, a number of divisions and repeats were analyzed. There were 5349 total repeats found. Palindromic repeats were the most frequent 2895 (54%), Tandem were 1528 (28%), and Dispersed repeats were 926 (17%). *Lavandula angustifolia*, which has 43 dispersed repeats and 36 tandem repeats, was the species with the most number of repetitive sequences. The lowest number of repeat sequences was found in *Leonotis leonurus* with only 16 tandem repeats. In the palindromic sequences repeats *Ocimum basilicum* were the most frequent with 34 repeat sequences. The lowest palindromic repeats were found in *Lavandula angustifolia* which has only 2 repeat sequences. In addition, the highest tandem repeats were found in *Teucrium fruticans* with 57 repeat, and the lowest number of tandem repeats found in two *leonuru*, *Ajuga forrestii* with 16 repeats (see, Table 2). Previous studies suggested that the repetitive sequence variations played a significant role in the reorganization and maintenance of the cp genomes (Zeb *et al.*, 2021).

Figure 1 Pie chart showing the percentage distribution of three types of repeat sequences—dispersed, palindromic, and tandem—in the chloroplast genomes of Lamiaceae species. Palindromic repeats accounted for the highest proportion, followed by tandem and dispersed repeats.

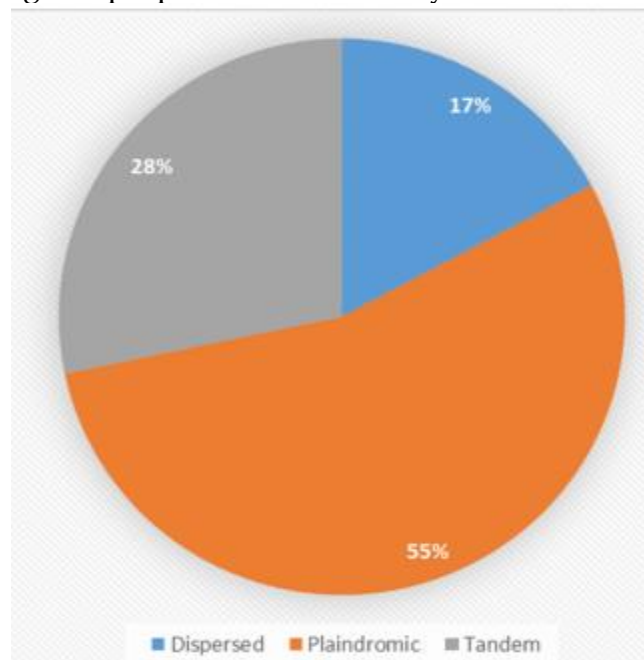


Figure 1: Distribution of Repeat Sequence Types

Source: Author's work.

Table 1: The characteristics of whole chloroplast genomes in Fifty-three species of Family Lamiaceae
[Large single copy(LSC), Small single copy (SSC), Inverted Repeats (IRs)]

S.No	Species	Size (bp)	LSC (bp)	SSC (bp)	IR (bp)	Coding Genes (CDS)	rRNA	tRNA	GC Content	GenBank Number
1	Thymus vulgaris	151649	83137	17835	26197	87	8	37	37.90%	NC_086487
2	Prunella vulgaris	151342	84097	18031	26437	87	8	37	37.90%	NC_039654
3	Satureja montana	151858	83670	17788	26253	87	8	37	37.90%	ON641342
4	Salvia prattii	151690	83070	17698	26230	87	8	37	38.00%	MK944407
5	Salvia roborowskii	151654	83225	17757	26229	87	8	37	38.00%	MK944406
6	Salvia yunnanensis	151396	83216	17691	26229	87	8	37	38.00%	MK944405
7	Salvia hispanica	150980	82808	18027	26424	87	8	37	38.00%	MN520017
8	Pogostemon cablin	152461	83946	17883	26200	87	8	37	38.20%	NC_042796
9	Origanum majorana	151841	83621	17980	26437	87	8	37	37.80%	MT385088
10	Origanum vulgare	151935	83580	26234	17807	86	8	37	37.80%	JX880022
11	Ocimum tenuiflorum	152419	84432	17784	26261	87	8	37	37.80%	MN687904
12	Ocimum basilicum	152407	83818	17853	26453	84	8	37	37.80%	KY623639
13	Ocimum gratissimum	152463	84130	17878	26261	87	8	36	37.90%	OR058741
14	Ocimum americanum	152460	84449	17785	26260	88	8	37	37.80%	OQ689804
15	Ocimum x africanum	152365	84458	17849	26141	86	8	37	37.80%	NC_082140
16	Mentha canadensis	152154	83499	17821	26220	87	8	37	37.80%	MN047448
17	Mentha spicata	152132	84006	18007	26492	87	8	37	37.80%	NC_037247
18	Mentha x piperita	152048	83313	17779	26251	87	8	37	37.80%	NC_047475
19	Melissa officinalis	152112	83849	17635	26124	87	8	37	38.10%	NC_066033
20	Lavandula angustifolia	153448	83770	17793	26183	88	8	37	38.00%	NC_029370
21	Lavandula stoechas	152100	82436	17891	26490	87	8	37	37.90%	ON641264
22	Stachys byzantina	149749	82267	17844	26484	88	8	37	38.70%	NC_029825
23	Hyssopus officinalis	149817	83630	17825	26486	85	8	37	37.80%	NC_066037
24	Lavandula dentata	151696	82255	17844	26490	87	8	38	37.90%	NC_046835
25	Teucrium fruticans	150808	82678	17835	26484	85	8	36	38.00%	NC_061173
26	Teucrium mascatense	150499	84009	17756	26051	88	8	39	38.40%	NC_044073

S.No	Species	Size (bp)	LSC (bp)	SSC (bp)	IR (bp)	Coding Genes (CDS)	rRNA	tRNA	GC Content	GenBank Number
27	Physostegia virginiana	152880	83954	17766	26079	86	8	37	38.40%	NC_085368
28	Dracocephalum moldavica	149868	83954	17766	26079	87	8	37	37.80%	NC_057509
29	Dracocephalum heterophyllum	150860	83660	18140	26398	88	8	37	37.80%	MW970109
30	rupestre	151230	83623	18152	26397	88	8	37	37.80%	NC_068682
31	Coleus hadiensis	152484	84160	17830	26257	87	8	36	37.70%	NC_069583
32	Vitex agnus-castus	154495	84478	17873	26257	86	8	36	38.30%	NC_088534
33	Coleus xanthanthus	152190	84478	17878	26273	88	8	37	37.70%	NC_058328
34	Coleus scutellarioides	152603	83050	17666	26253	85	8	37	37.70%	MW538954
35	Nepeta cataria	152339	83345	17763	26262	87	8	37	37.90%	NC_051544
36	Ajuga forrestii	150472	83532	17696	26257	86	8	35	38.30%	OR038698
37	Ajuga nipponensis	150442	83511	17621	26124	88	8	37	38.30%	NC_084167
38	Ajuga reptans	149825	83581	17631	26124	85	8	37	38.20%	PP073955
39	Ajuga campylantha	150453	83815	17854	26488	86	8	35	38.30%	NC_084165
40	Ballota nigra	151909	84287	17702	26541	87	8	37	38.50%	NC_066022
41	Salvia officinalis	151163	84312	17943	26890	87	8	37	38.00%	MZ636550
42	Glechoma longituba	151768	84356	17457	26755	86	8	37	37.90%	MF967576
43	Leonurus sibiricus	151689	84618	17890	26090	88	8	37	38.40%	NC_Unknown
44	Pogostemon cablin	152461	84569	17688	26433	87	8	37	38.20%	NC_042796
45	Pogostemon stellatus	151824	84680	17655	26709	87	8	37	38.20%	NC_031434
46	Phyllostegia velutina	150131	84123	17654	26132	88	8	37	38.40%	NC_029820
47	Galeopsis bifida	151890	84259	17323	26545	88	8	37	38.50%	MT473759
48	Stenogyne haliakalae	149736	84320	17899	26545	88	8	37	38.50%	NC_029817
49	Elsholtzia densa	149095	84313	17324	26098	84	8	37	37.90%	NC_056994
50	Ballota nigra	151909	84303	17800	26788	87	8	37	38.50%	NC_066022
51	Lamium album	150505	84907	17866	26098	89	8	37	38.60%	NC_036971

**Table 2: Repeat sequences in the 50 species of family
Lamiaceae chloroplast genomes**

Species	Dispersed				Palindromic	Tandem
	F	R	C	total		
Origanum vulgare	11	0	0	11	39	26
Ocimum basilicum	17	2	1	20	19	34
Glechoma longituba	23	1	0	24	20	31
Salvia yunnanensis	21	0	0	21	24	22
Salvia roborowskii	24	0	0	24	26	18
Salvia prattii	24	0	0	24	26	18
Mentha Canadensis	25	0	1	26	24	37
Salvia hispanica	25	0	0	25	25	30
Ocimum tenuiflorum	19	2	1	22	19	35
Origanum majorana	16	2	0	18	17	31
Galeopsis bifida	23	0	0	23	25	23
Coleus scutellarioides	21	0	0	21	27	42
Dracocephalum heterophyllum	15	0	0	15	19	32
Salvia officinalis	20	0	0	20	21	27
Lavandula angustifolia	16	27	0	43	2	36
Stenogyne haliakalae	12	2	0	14	16	21
Phyllostegia velutina	0	0	0	0	0	22
Stachys byzantine	0	0	0	0	0	24
Pogostemon stellatus	13	1	0	14	21	29
Lamium album	13	0	0	13	19	26
Mentha spicata	25	1	0	26	24	32
Prunella vulgaris	23	0	0	23	27	29
Pogostemon cablin	0	0	0	0	0	38
Teucrium mascatense	0	0	0	0	0	33
Lavandula dentate	25	0	0	25	25	20
Mentha x piperita	24	0	0	24	24	33
Nepeta cataria	19	2	1	22	20	31
Elsholtzia densa	13	1	1	15	10	25
Dracocephalum moldavica	17	0	0	17	19	24
Coleus xanthanthus	20	0	0	20	29	31
Teucrium fruticans	20	1	0	21	23	57
Ballota nigra	20	1	0	21	22	26
Melissa officinalis	22	0	0	22	20	30
Hyssopus officinalis	18	0	0	18	19	17
Leonurus sibiricus	18	1	0	19	19	16
Dracocephalum rupestre	17	0	0	17	18	21
Coleus hadiensis	21	0	1	22	20	35
Ocimum x africanum	20	1	1	22	20	35
Ajuga campylantha	19	1	0	20	21	16

Species	Dispersed				Palindromic	Tandem
	F	R	C	total		
Ajuga nipponensis	19	0	0	19	23	17
Physostegia virginiana	13	0	0	13	21	20
Thymus vulgaris	21	0	0	21	21	23
Leonotis leonurus	16	1	0	17	19	16
Vitex agnus-castus	22	0	1	23	22	29
Lavandula stoechas	24	0	0	24	25	34
Satureja montana	15	0	0	15	13	28
Ocimum americanum	20	1	1	22	20	34
Ajuga forrestii	18	0	0	18	21	16
Ocimum gratissimum	21	1	0	22	19	38
Percentage				926	2895	1528
				17%	54%	Total
						5349
						28%

Figure 2 Histogram illustrating the repeat sequence analysis of plastid genomes in selected Lamiaceae species. The chart compares the frequency and distribution of the three repeat types across species, highlighting interspecific variation.

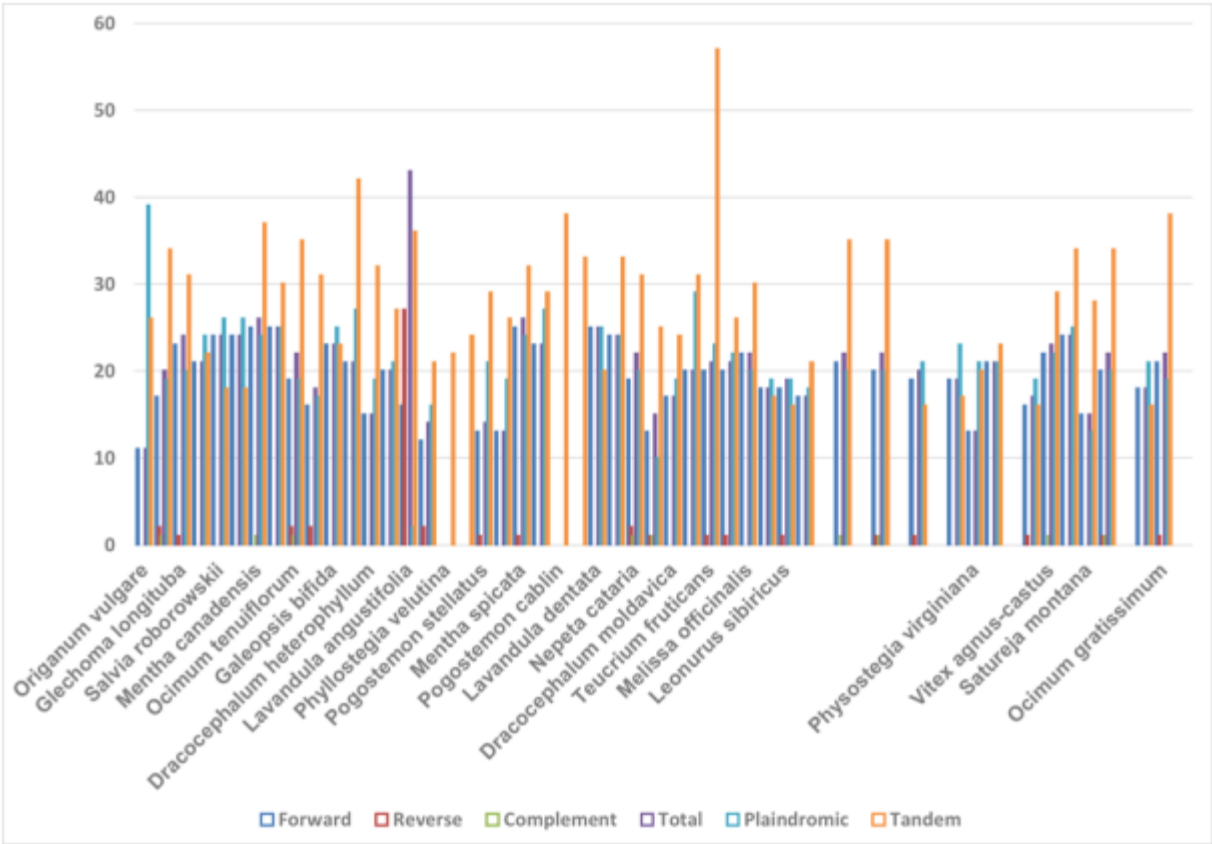


Figure 2: Repeat Sequence Analysis of Plastid Genomes in Family Lamiaceae Species
Histogram

Source: Author’s work.

Figure 3 shows line graph representing the distribution of simple sequence repeats (SSRs) in the chloroplast genomes of 50 Lamiaceae species. The analysis includes six SSR types—mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide—depicting their relative proportions and frequency trends across the species studied.

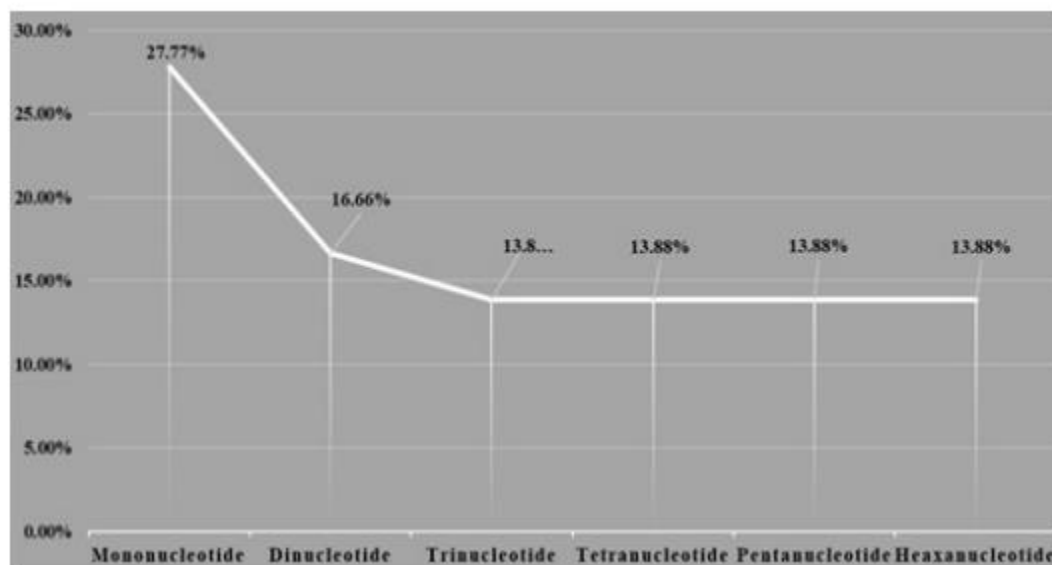


Figure 3: SSR Types and Frequencies in Lamiaceae cp Genomes

Source: Author's work.

Table 3 shows the Summary of simple sequence repeats (SSRs) identified in the chloroplast genomes of fifty Lamiaceae species. The table lists the number, types, and distribution of SSR markers, providing insights into their potential use in genetic diversity, phylogenetic, and evolutionary studies.

Table 3: Simple Sequence Repeats (SSRS) Marker In Chloroplast Genomes of Fifty Species

Species	Mono-nucleotide	Di-nucleotide	Tri-nucleotide	Tetra-nucleotide	Penta-nucleotide	Heza-nucleotide
Thymus vulgaris	10	6	5	5	5	5
Prunella vulgaris	10	6	5	5	5	5
Satureja montana	10	6	5	5	5	5
Salvia prattii	10	6	5	5	5	5
Salvia roborowskii	10	6	5	5	5	5
Salvia yunnanensis	10	6	5	5	5	5
Salvia hispanica	10	6	5	5	5	5
Pogostemon cablin	10	6	5	5	5	5
Origanum majorana	10	6	5	5	5	5
Origanum vulgare	10	6	5	5	5	5
Ocimum tenuiflorum	10	6	5	5	5	5
Ocimum basilicum	10	6	5	5	5	5
Ocimum gratissimum	10	6	5	5	5	5
Ocimum americanum	10	6	5	5	5	5

Species	Mono-nucleotide	Di-nucleotide	Tri-nucleotide	Tetra-nucleotide	Penta-nucleotide	Heza-nucleotide
Ocimum x africanum	10	6	5	5	5	5
Mentha canadensis	10	6	5	5	5	5
Mentha spicata	10	6	5	5	5	5
Mentha x piperita	10	6	5	5	5	5
Melissa officinalis	10	6	5	5	5	5
Lavandula angustifolia	10	6	5	5	5	5
Lavandula stoechas	10	6	5	5	5	5
Stachys byzantina	10	6	5	5	5	5
Hyssopus officinalis	10	6	5	5	5	5
Lavandula dentata	10	6	5	5	5	5
Teucrium fruticans	10	6	5	5	5	5
Teucrium mascatense	10	6	5	5	5	5
Physostegia virginiana	10	6	5	5	5	5
Dracocephalum moldavica	10	6	5	5	5	5
Dracocephalum heterophyllum	10	6	5	5	5	5
Dracocephalum rupestre	10	6	5	5	5	5
Coleus hadiensis	10	6	5	5	5	5
Vitex agnus-castus	10	6	5	5	5	5
Coleus xanthanthus	10	6	5	5	5	5
Coleus scutellarioides	10	6	5	5	5	5
Nepeta cataria	10	6	5	5	5	5
Ajuga forrestii	10	6	5	5	5	5
Ajuga nipponensis	10	6	5	5	5	5
Ajuga reptans	10	6	5	5	5	5
Ajuga campylantha	10	6	5	5	5	5
Ballota nigra	10	6	5	5	5	5
Salvia officinalis	10	6	5	5	5	5
Glechoma longituba	10	6	5	5	5	5
Leonurus sibiricus	10	6	5	5	5	5
Pogostemon cablin	10	6	5	5	5	5
Pogostemon stellatus	10	6	5	5	5	5
Phyllostegia velutina	10	6	5	5	5	5
Galeopsis bifida	10	6	5	5	5	5
Stenogyne haliakalae	10	6	5	5	5	5
Elsholtzia densa	10	6	5	5	5	5
Ballota nigra	10	6	5	5	5	5

Species	Mono-nucleotide	Di-nucleotide	Tri-nucleotide	Tetra-nucleotide	Penta-nucleotide	Heza-nucleotide
Leonotis leonurus	10	6	5	5	5	5
Lamium album	10	6	5	5	5	5

4.3. Phylogenetic Relationship of Family Lamiaceae in 50 Species

Numerous studies have extensively investigated and analyzed the phylogenetic relationships within the Lamiaceae family using plastid genomes, focusing on both coding and non-coding regions to elucidate evolutionary patterns. In this study, we performed a comprehensive phylogenetic analysis based on eighty-eight conserved protein-coding genes derived from the complete chloroplast genomes of fifty representative species within the Lamiaceae family, along with one outgroup species, *Berberis bealei* (NC_022457), which belongs to the Berberidaceae family. The inclusion of *Berberis bealei* as an outgroup allowed for better rooting and resolution of the evolutionary tree. To construct the phylogenetic tree, we employed the neighbor-joining (NJ) method, which was supported by a bootstrap analysis with 100 replicates to evaluate the reliability and robustness of the inferred topologies. The results showed no significant divergence in the coding regions among the examined Lamiaceae species, indicating a high level of conservation within the family. In contrast, *Berberis bealei* was clearly separated from the Lamiaceae clade, confirming its distinct phylogenetic position. Furthermore, the phylogenetic topology demonstrated that all fifty Lamiaceae species exhibited a relatively close evolutionary relationship with one another, forming a well-supported monophyletic group. This suggests a shared evolutionary lineage and possible recent diversification within the family. The high bootstrap support values further reinforced the credibility of the inferred evolutionary relationships among the taxa. These findings provide critical insights into the molecular systematics and evolutionary history of the Lamiaceae and offer a solid foundation for future comparative genomic and evolutionary studies. The detailed evolutionary relationships and clustering patterns of these species are illustrated in the phylogenetic tree presented in **Figure 4**.

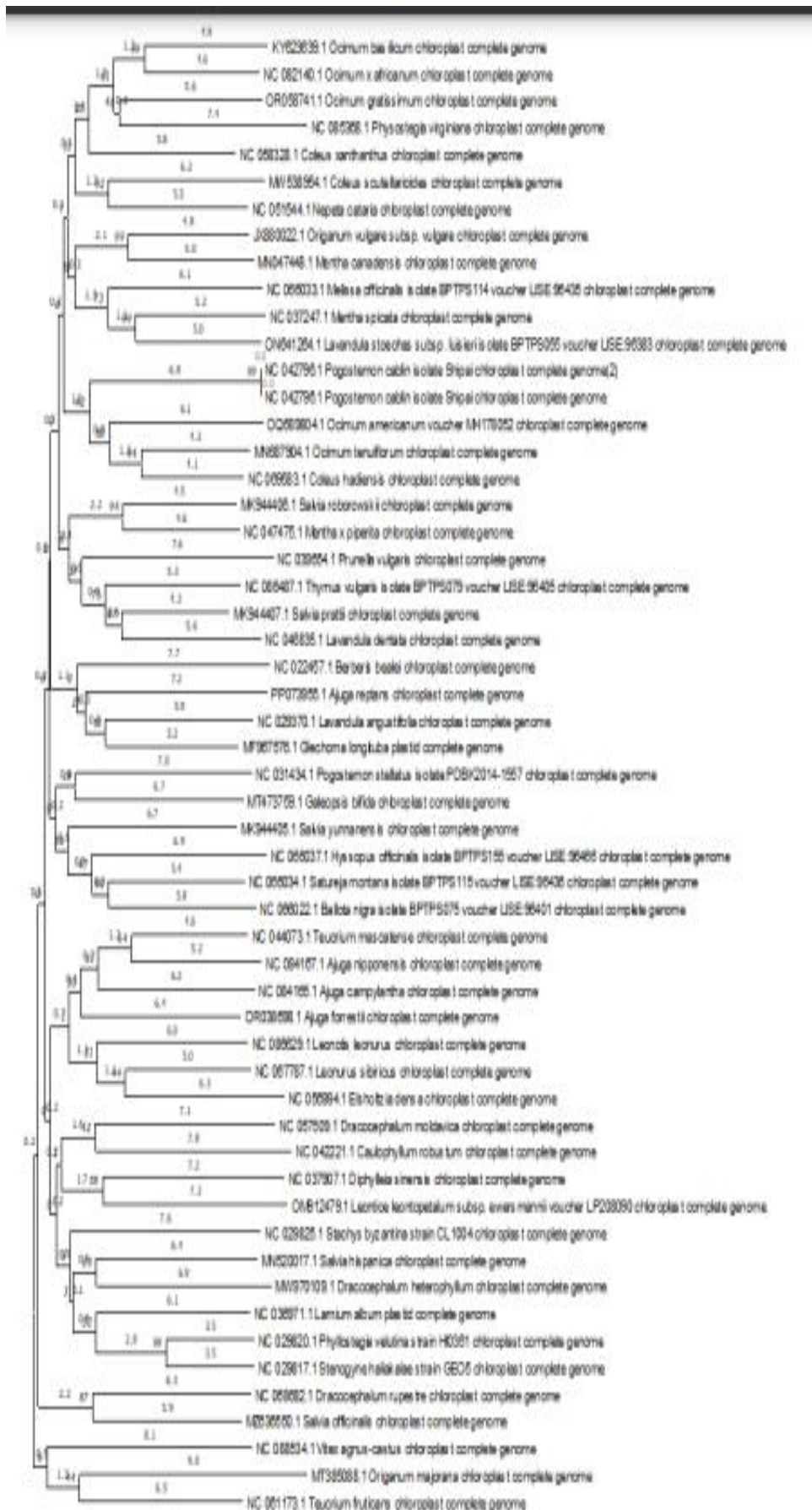


Figure 4: Phylogenetic Tree Obtained for 50 Species of Lamiaceae Based on Whole Chloroplast Genome

Source: Author's work.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1. Conclusions

The study investigated the chloroplast genome as a valuable resource for understanding evolutionary history. Specifically, we analyzed the chloroplast (cp) genomes of various species of Lamiaceae family and observed similarities in genome structure and order across these species. Additionally, the study identified the location and distribution of repeat sequences within the cp genomes. By comparing pairwise sequence divergences among interrelated species, we gained insights into their evolutionary relationships. Whole-genome sequencing of the chloroplasts provided essential information for taxonomic classification. Furthermore, the study shed light on the phylogeny of Lamiaceae by examining the cp genome. Comparative analyses of plastid genome sequences yielded DNA markers that facilitate species identification and classifications.

5.2. Recommendations

The following are the study's recommendation, i.e.,

- I. This comparative sequence analysis and divergence hotspot knowledge of family Lamiaceae existing here recommend for future studies on the population genetics, molecular phylogeny, and divergence time for other angiosperm species.
- II. The Composite Lamiaceae species and analysis on the basis of whole chloroplast genome which will further utilized for the investigation of complete phylogeny and cosmopolitan distributions.
- III. The current information of plastid genome and other previous studies of whole chloroplast genome showed multifunctional organelle which suggested for plants adaptation suffering from stresses.
- IV. We recognized significant gene-tree conflict within the chloroplast genome of angiosperm phylogeny, highlighting the requirement of future research into the plastome and some genes share different evolutionary histories.
- V. Simple sequence repeats and large repetitive sequences and exceeding the variable sites are widely used, for phylogenetic and evolutionary studies and useful indicators for the population genetics.

Ethical approval

All international standards have been adopted and compliance.

Informed consent

The study was conducted with equal participation by all authors.

Conflicts of interests

The authors declare that there are no conflicts of interests.

Funding

The study has not received any external funding.

Data and materials availability

All data associated with this study are present in the paper.

REFERENCES AND NOTES

1. Abu-Asab, M. S., & Cantino, P. D. (1987). Phylogenetic implications of leaf anatomy in subtribe Melittidinae (Labiatae) and related taxa. *Journal of the Arnold Arboretum*, 68, 1–34.
2. Ahmad, M., Bano, A., Zafar, M., Khan, M. A., Chaudhry, S. J. L., & Sultana, S. (2013). Pollen morphology of some species of the family Asteraceae from the alpine zone, Deosai Plateau, northern Pakistan. *Palynology*, 17(2), 189–195.
3. Al-Shehbaz, I. A., & Marhold, K. (2024). Notes on the Caribbean, Mexican, and Central and South American Cardamine (Cardamineae, Brassicaceae). *Harvard Papers in Botany*, 29(2), 195–204.
4. Asaf, S., Khan, A. L., Lubna, Khan, A., Khan, A., Khan, G., & Al-Harasi, A. (2010). Expanded inverted repeat region with large-scale inversion in the first complete plastid genome sequence of *Plantago ovata*. *Scientific Reports*, 10(1), 3881.
5. Azim, M. K. (2021). Date palm (*Phoenix dactylifera* L.) chloroplast genome. In *The Date Palm Genome, Vol. 1: Phylogeny, Biodiversity and Mapping* (pp. 201–209). Cham: Springer International Publishing.
6. Bac, K. M., Sim, S. C., Hong, J. H., Choi, K. J., Kim, D. H., & Kwon, Y. S. (2015). Development of genomic SSR markers and genetic diversity analysis in cultivated radish (*Raphanus sativus* L.). *Horticulture, Environment, and Biotechnology*, 56, 216–224.
7. Bi, Y., Zhang, M. F., Xue, J., Dong, R., Du, Y. P., & Zhang, X. H. (2018). Chloroplast genomic resources for phylogeny and DNA barcoding: A case study on *Fritillaria*. *Scientific Reports*, 8(1), 1184.
8. Biscotti, M. A., Olmo, E., & Heslop-Harrison, J. S. (2015). Repetitive DNA in eukaryotic genomes. *Chromosome Research*, 23, 415–420.
9. Bobik, K., & Burch-Smith, T. M. (2015). Chloroplast signaling within, between and beyond cells. *Frontiers in Plant Science*, 6, 781.
10. Du, X., Zeng, T., Feng, Q., Hu, L., Luo, X., Weng, Q., ... & Zhu, B. (2020). The complete chloroplast genome sequence of yellow mustard (*Sinapis alba* L.) and its phylogenetic relationship to other Brassicaceae species. *Gene*, 731, 144340.
11. Duarte, M. D. R., & de Souza, D. C. (2014). Microscopic characters of the leaf and stem of *Lavandula dentata* L. (Lamiaceae). *Microscopy Research and Technique*, 77(8), 647–652.
12. Feng, Z., Zheng, Y., Jiang, Y., Pei, J., & Huang, L. (2024). Phylogenetic relationships, selective pressure and molecular markers development of six species in subfamily Polygonoideae based on complete chloroplast genomes. *Scientific Reports*, 14(1), 9783.
13. Hamed, A. N., Attia, E., & Desoukey, S. Y. (2021). A review on various classes of secondary metabolites and biological activities of Lamiaceae (Labiatae) (2002–2018). *Journal of Advanced Biomedical and Pharmaceutical Sciences*, 4(1), 16–31.
14. Harley, R. M., Atkins, S., Budantsev, A. L., Cantino, P. D., Conn, B. J., Grayer, R., ... & Upson, T. (2004). Labiatae. In J. W. Kadereit (Ed.), *The families and genera of vascular plants* (Vol. 7, pp. 167–275). Springer-Verlag.
15. Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3), R25. <https://doi.org/10.1186/gb-2009-10-3-r25>
16. Li, H., Zheng, Q., Wang, Y., & Tian, X. (2024). Unveiling the chloroplast genomes of *Capsella bursa-pastoris*, *Descurainia sophia*, and *Lepidium apetalum* (Brassicaceae). *Journal of Asia-Pacific Biodiversity*, 17(2), 245–249.
17. Okan, K., Sevindik, E., & Paksoy, M. Y. (2024). Phylogenetic analysis of the endemic *Bornmuellera Hausskn. spp.* (Brassicaceae) in Türkiye based on nuclear ITS and chloroplast trn L intron, trn LF, rbc L and trn Q-rps 16 DNA sequences. *Genetic Resources and Crop Evolution*, 71(4), 1529–1539.

18. Ren, T., Xun, L., Jia, Y., & Li, B. (2024). Complete plastomes of ten *Rorippa* species (Brassicaceae): Comparative analysis and phylogenetic relationships. *Agronomy*, 14(5), 913.
19. Shang, M., Wang, J., Dai, G., Zheng, J., Liao, B., Wang, J., & Duan, B. (2023). Comparative analysis of chloroplast genome and new insights into phylogenetic relationships of *Ajuga* and common adulterants. *Frontiers in Plant Science*, 14, 1251829.
20. Silva, H. (2020). A descriptive overview of the medical uses given to *Mentha* aromatic herbs throughout history. *Biology*, 9(12), 484. <https://doi.org/10.3390/biology9120484>
21. Tamboli, E. T., Chester, K., & Ahmad, S. (2015). Quality control aspects of herbs and botanicals in developing countries: *Coleus forskohlii*: Briq a case study. *Journal of Pharmacy and Bioallied Sciences*, 7(4), 254-259.
22. Van Wyk, B.-E. (2014). Culinary herbs and spices of the world. University of Chicago Press.
23. Wang, M., Cui, L., Fang, K., Deng, P., Wan, F., Weining, S., & Ning, X. (2015). Comparative analysis of Asteracean chloroplast genomes: Structural organization, RNA editing and evolution. *Plant Molecular Biology Reporter*, 33(5), 1526–1538.
24. Wang, Y., Liang, Q., Zhang, C., Huang, H., He, H., Wang, M., ... & Sun, B. (2023a). Sequencing and analysis of complete chloroplast genomes provide insight into the evolution and phylogeny of Chinese kale (*Brassica oleracea* var. *alboglabra*). *International Journal of Molecular Sciences*, 24(12), 10287.
25. Wang, Y., Zhan, H., Zhang, Y., Long, Z., & Yang, X. (2023b). Mitochondrial Genome Analysis and Phylogeny and Divergence Time Evaluation of the *Strix aluco*. *bioRxiv*, 2023-01. Online available at: <https://www.biorxiv.org/content/10.1101/2023.01.20.524943v1.abstract> (accessed on 24th May, 2025).
26. Wheeler, G., Dorman, H. E., Buchanan, A., Challagundla, L., & Wallace, L. E. (2014). A review of the prevalence, utility, and caveats of using chloroplast simple sequence repeats for studies of plant biology. *Applications in Plant Sciences*, 2(12), 1400059.
27. Wicke, S., Schneeweiss, G. M., Depamphilis, C. W., Müller, K. F., & Quandt, D. (2011). The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Molecular Biology*, 76(3), 273–297.
28. Williams, A. V., Boykin, L. M., Howell, K. A., Nevill, P. G., & Small, I. (2015). The complete sequence of the *Acacia ligulata* chloroplast genome reveals a highly divergent *clpP1* gene. *PLOS ONE*, 10(3), e0125768.
29. Wu, Y., Liu, F., Yang, D. G., Li, W., Zhou, X. J., Pei, X. Y., Liu, Y. G., He, K. L., Zhang, W. S., Ren, Z. Y., & Zhou, K. H. (2018). Comparative chloroplast genomics of *Gossypium* species: Insights into repeat sequence variations and phylogeny. *Frontiers in Plant Science*, 9, 376.
30. Xu, S., Dai, Z., Guo, P., Fu, X., Liu, S., Zhou, L., Tang, W., Feng, T., Chen, M., Zhan, L., & Wu, T. (2021). *ggTreeExtra*: Compact visualization of richly annotated phylogenetic data. *Molecular Biology and Evolution*, 38(9), 4039–4042.
31. Yan, C., Du, J., Gao, L., Li, Y., & Hou, X. (2019). The complete chloroplast genome sequence of watercress (*Nasturtium officinale* R. Br.): Genome organization, adaptive evolution and phylogenetic relationships in *Cardamineae*. *Gene*, 699, 24–36.
32. Yang, L., & Zhou, G. (2023). Phylogeography and ecological niche modeling implicate multiple microrefugia of *Swertia tetraptera* during quaternary glaciations. *BMC Plant Biology*, 23(1), 450.
33. Yao, J., Zhao, F., Xu, Y., Zhao, K., Quan, H., Su, Y., Hao, P., Liu, J., Yu, B., Yao, M., & Ma, X. (2020). Complete chloroplast genome sequencing and phylogenetic analysis of two *Dracocephalum* plants. *BioMed Research International*, 2020.

- <https://doi.org/10.1155/2020/3561893>
34. Yu, S., Yang, X., Tian, X., Liu, X., Lu, X., Huang, C., & Fu, Z. (2022). The complete chloroplast genome sequence of the monotypic and enigmatic genus *Cavea* (tribe *Gymnarrheneae*) and a comparison with other species in *Asteraceae*. *Journal of Genetics*, 101(1), 20.
 35. Yuan, N., Sun, Y., Comes, H. P., Fu, C. X., & Qiu, Y. X. (2014). Understanding population structure and historical demography in a conservation context: Population genetics of the endangered *Kirengeshoma palmata* (*Hydrangeaceae*). *American Journal of Botany*, 101(3), 521–529.
 36. Yun, S., & Kim, S. C. (2022). Comparative plastomes and phylogenetic analysis of seven Korean endemic *Saussurea* (*Asteraceae*). *BMC Plant Biology*, 22(1), 550.
 37. Yusupova, Z. A., & Baratjon o'g'li, S. F. (2022). Bioecological properties of medicinal species of the mint family (*Lamiaceae*). *Galaxy International Interdisciplinary Research Journal*, 10(10), 285–291.
 38. Zhang, D. Y., Tu, X. D., Jiang, Y. T., Liu, Z. J., & Ma, L. (2020). The complete chloroplast genome sequence of *Acorus calamus* (*Acoraceae*) from Fujian, China. *Mitochondrial DNA Part B*, 5(2), 1334–1335.
 39. Zhang, X., Deng, T., Moore, M. J., Ji, Y., Lin, N., Zhang, H., Meng, A., Wang, H., Sun, Y., & Sun, H. (2019). Plastome phylogenomics of *Saussurea* (*Asteraceae*, *Cardueae*). *BMC Plant Biology*, 19(1), 1–10.
 40. Zhao, F., Chen, Y. P., Salmaki, Y., Drew, B. T., Wilson, T. C., Scheen, A. C., ... & Li, B. (2021). An updated tribal classification of *Lamiaceae* based on plastome phylogenomics. *BMC Biology*, 19(2), 331. <https://doi.org/10.1186/s12915-021-01165-0>
 41. Zhao, M., Wu, Y., & Ren, Y. (2024). Complete Chloroplast Genome Sequence Structure and Phylogenetic Analysis of *Kohlrabi* (*Brassica oleracea* var. *gongylodes* L.). *Genes*, 15(5), 550.
 42. Zhou, J., Cui, Y., Chen, X., Li, Y., Xu, Z., Duan, B., Li, Y., Song, J., & Yao, H. (2018). Complete chloroplast genomes of *Papaver rhoeas* and *Papaver orientale*: Molecular structures, comparative analysis, and phylogenetic analysis. *Molecules*, 23(2), 437.
 43. Zhu, B., Qian, F., Hou, Y., Yang, W., Cai, M., & Wu, X. (2021). Complete chloroplast genome features and phylogenetic analysis of *Eruca sativa* (*Brassicaceae*). *PLOS ONE*, 16(3), e0248556.