Journal of AI, ML, and Bioinformatics 2025, pp. 29–39
https://doi.org/10.5147/jaimlb.vi.257

Comparative Phylogenetic Analysis of Six Angiosperm Families Using *rbcL* and *matK* Chloroplast Markers

Kenneth Leaks, Amir El, Zeinab Alsaidi, Kyah Benton, Jamila Chase, Shaniyah Lewis, and My Abdelmajid Kassem*

Plant Genomics and Bioinformatics Lab, Department of Biological and Forensic Sciences, Fayetteville State University, Fayetteville, NC 28301, USA.

Received: April 6, 2025 / Accepted: June 2, 2025

Abstract

Chloroplast DNA markers such as *rbcL* and *matK* are widely used in plant molecular systematics due to their complementary evolutionary properties. In this study, we conducted a comparative phylogenetic analysis of six major angiosperm families—*Apiaceae, Asteraceae, Fabaceae, Lamiaceae, Orchidaceae,* and *Rosaceae*—using publicly available *rbcL* and *matK* sequences. A total of 600 sequences (50 species per family per gene) were retrieved from GenBank. Sequence alignment was performed using MAFFT v7 via Biopython scripts, and Maximum Likelihood phylogenies were constructed using MEGA 11. Tree visualization and annotation were carried out in iTOL v7. Both *rbcL* and *matK* markers successfully recovered monophyletic family-level groupings, validating their utility in higher-level phylogenetic inference. However, *matK*-based trees consistently demonstrated higher resolution, especially at genus and species levels. In families such as *Fabaceae* and *Lamiaceae*, *matK* resolved well-supported clades (e.g., *Glycine*, *Vigna*, *Salvia, Mentha*), whereas *rbcL* exhibited unresolved polytomies and lower bootstrap support. Even in families where *rbcL* performed moderately well (e.g., *Asteraceae, Orchidaceae*), *matK* provided greater phylogenetic clarity, particularly among terminal clades. A visual bootstrap analysis revealed average support values of 66.1% for *rbcL* trees and approximately 82.0% for *matK* trees. These results underscore the stronger phylogenetic signal and discriminatory power of *matK* in resolving recent divergence events and complex lineages. While *rbcL* remains useful for deeper phylogenetic relationships due to its conserved nature, *matK* is better suited for fine-scale taxonomic resolution. Our findings reinforce the importance of marker selection in phylogenetic studies and support the combined use of *rbcL* and *matK* in multilocus frameworks. These results also highlight the potential of publicly available sequence data and reproducible pipelines for advancing plant evolutionary research.

Keywords: Angiosperms, Chloroplast DNA, *rbcL* gene, *matK* gene, Phylogenetic analysis, DNA barcoding, Plant systematics, Maximum Likelihood trees.

* Corresponding author: mkassem@uncfsu.edu

1. Introduction

Angiosperms, commonly known as flowering plants, represent the most diverse and ecologically dominant group of land plants, encompassing over 300,000 described species—more than 80% of all extant plant species (Soltis and Soltis, 2013; Chase et al., 2016; Soltis et al., 2018, 2019; Benton et al., 2022; Zuntini et al., 2024). This vast group exhibits extraordinary morphological and ecological diversity, ranging from minute herbs to towering trees, and occupying nearly every terrestrial habitat on Earth. Angiosperms play indispensable roles in global ecosystems by producing oxygen through photosynthesis, contributing to nutrient cycling, and forming the structural and nutritional backbone of most terrestrial food webs (Liu et al., 2019; Herrera and Pellmyr, 2002). Furthermore, their ecological interactions—particularly with pollinators and herbivores—have driven co-evolutionary dynamics that promote biodiversity and ecosystem resilience (Benton et al., 2022; Bascompte and Scheffer, 2023).

Beyond their ecological importance, angiosperms are of immense economic value. They provide most human food crops, such as rice, wheat, maize, and fruits, as well as medicinal plants, timber, fibers, oils, and ornamental species (Chase et al., 2016; Sharma et al., 2024). As primary producers and key components of terrestrial ecosystems, angiosperms support a vast array of animal life, including insects, birds, and mammals, by offering food resources, habitat structure, and breeding sites (Richer, 1997; Delaney and von Wettberg, 2023). Their remarkable adaptability and evolutionary success make them a focal point of botanical and evolutionary research.

Historically, the classification of angiosperms was based primarily on morphological traits, including flower structure, leaf arrangement, seed morphology, and vascular tissue organization (Thorne, 1976; Endress et al., 2000). These characters were used extensively by early taxonomists—most notably Carl Linnaeus in the 18th century—to establish hierarchical systems of plant classification (Soltis and Soltis, 2013; Chase et al., 2016). However, while morphology provided a valuable starting point, it proved insufficient for resolving many phylogenetic relationships due to issues such as convergent evolution, morphological plasticity, and homoplasy (Hall, 2003). Different lineages often evolved similar structures in response to similar ecological pressures, leading to misclassification and paraphyletic groupings (Endress and Matthews, 2006; Matthews and Endress, 2006). These limitations underscored the need for a more robust and objective framework for classifying plant diversity.

The advent of molecular systematics in the late 20th century transformed plant taxonomy by enabling researchers to assess genetic similarity and infer evolutionary relationships using DNA sequence data (Thorne, 1976; Endress et al., 2000; Soltis and Soltis, 2013; Chase et al., 2016). The chloroplast genome emerged as a rich source of phylogenetic markers due to its relatively conserved structure, uniparental inheritance, and low recombination rates (Palmer, 1988). Among these markers, the *rbcL* and *matK* genes have been extensively employed in angiosperm phylogenetics and DNA barcoding initiatives (Clement and Donoghue, 2012).

The *rbcL* gene encodes the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO), a key enzyme in photosynthesis. It is one of the most conserved coding regions in the plastid genome and has been widely used for reconstructing higher-level phylogenies across plant orders and families (Chase et al., 2016; Jamdade et al., 2021). However, due to its slow rate of evolution, *rbcL* may lack sufficient variability to resolve relationships at lower taxonomic levels, such as within genera or among closely related species (Doyle et al., 1997; Muller et al., 2006).

In contrast, the *matK* gene, which is located within the intron of the trnK gene, exhibits a higher rate of nucleotide substitution, making it suitable for resolving relationships at finer phylogenetic scales (Hilu et al., 2003; Muller et al., 2006; Jamdade et al., 2021). It codes for a maturase enzyme involved in RNA splicing and is considered one of the most variable regions of the plastid genome, particularly useful for DNA barcoding and evolutionary studies at the species and genus levels (Hilu and Barthet, 2008). Due to these complementary properties, *rbcL* and *matK* are often used in tandem to maximize phylogenetic resolution and robustness (Fazekas et al., 2008).

The integration of molecular data into systematics has culminated in the development of comprehensive, multi-locus phylogenies and the refinement of plant classification systems such as the Angiosperm Phylogeny Group (APG) framework (Chase et al., 2016). These advances have not only resolved long-standing taxonomic ambiguities but have also illuminated patterns of divergence, biogeography, and trait evolution across the angiosperm tree of life (Smith & Brown, 2018).

In this study, we apply a molecular phylogenetic approach to six diverse and economically important angiosperm families—*Asteraceae*, *Fabaceae*, *Lamiaceae*, *Orchidaceae*, *Apiaceae*, and *Rosaceae*—using publicly available *rbcL* and *matK* sequences. Each family was represented by 50 randomly selected species, and phylogenetic trees were constructed using the Maximum Likelihood method following sequence alignment with MAFFT and subsequent visualization with iTOL. By comparing the topologies and resolution of trees generated from *rbcL* and *matK*, we assess the phylogenetic utility of these markers across different lineages and contribute to a broader understanding of angiosperm diversification.

2. Methods

2.1 Taxon Sampling and Sequence Retrieval

To investigate phylogenetic relationships among angiosperms, we selected six major plant families: *Apiaceae*, *Asteraceae*, *Fabaceae*, *Lamiaceae*, *Orchidaceae*, and *Rosaceae*. These families were chosen due to their ecological importance, species richness, and availability of molecular data. For each family, 50 species were randomly selected, ensuring a diverse representation across different genera and subfamilies where possible.

DNA sequences for two widely used chloroplast markers—ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (*rbcL*) and maturase K (*matK*)—were retrieved in FASTA format from the NCBI GenBank database (https://www.ncbi.nlm.nih.gov). Species with full-length, high-quality sequences were prioritized, and redundant or partial sequences were excluded to ensure consistency and reliability of downstream analysis. A total of 600 sequences were collected (6 families × 50 species × 2 genes).

2.2 Sequence Alignment and Quality Control

All sequence alignments were conducted using the MAFFT v7 algorithm (Katoh and Standley, 2013), which is well-suited for handling large datasets and sequences with moderate variability. Alignments were executed via Biopython scripts, which enabled automated retrieval, formatting, and processing of the sequences. Default MAFFT parameters were applied with gap opening penalty set to 1.53 and extension penalty to 0.123, balancing alignment accuracy and computational efficiency.

Post-alignment, sequences were manually inspected in MEGA 11 (Tamura et al., 2021) to identify and remove poorly aligned or ambiguous regions. Insertions, deletions, and non-homologous characters were trimmed to minimize noise in phylogenetic reconstruction. Additionally, we screened all sequences for quality issues, including the presence of stop codons, frameshifts, excessive ambiguous bases, or evidence of chimera formation. Sequences with such features were excluded from further analysis. Final alignments were exported in .meg and .fasta formats for subsequent tree construction.

2.3 Phylogenetic Tree Construction

Phylogenetic analyses were performed using the Maximum Likelihood (ML) method, implemented in MEGA 11, a widely used software for evolutionary studies. For each family and each gene, separate ML trees were constructed, resulting in a total of 12 phylogenetic trees (6 families × 2 markers). The Tamura-Nei model of nucleotide substitution was selected as the best-fitting model based on MEGA's model selection algorithm, which evaluates likelihood scores for various models using the Bayesian Information Criterion (BIC).

To assess the robustness of tree topologies, bootstrap analyses with 1,000 replicates were performed. Nodes with bootstrap values ≥70% were considered well supported (Hillis & Bull, 1993). Trees were rooted using mid-point rooting to allow for balanced visualization, as outgroups were not included due to the intra-family focus of the study.

Trees were rooted using mid-point rooting, a widely accepted method in molecular phylogenetics when appropriate outgroups are unavailable. Given the intra-familial focus of this study and the absence of universal outgroups applicable across all six families, midpoint rooting was chosen to balance tree topology and facilitate visualization of relative divergence. This approach is particularly suitable for chloroplast data where rate variation is moderate and

tree symmetry is informative (Letunic and Bork, 2021; Tamura et al., 2021).

2.4 Tree Visualization and Annotation

Phylogenetic trees generated in MEGA 11 were exported in Newick (.nwk) format and uploaded to the Interactive Tree Of Life (iTOL v7) platform (Letunic and Bork, 2021) for visualization and annotation. Species names were color-coded based on subfamily or genus (when available) to facilitate interpretation of clade structure. Bootstrap values were displayed on corresponding branches to highlight confidence levels in node resolution.

Branch lengths were scaled to reflect nucleotide substitutions per site, and trees were exported in publication-quality SVG and PDF formats. Comparative visualization between *rbcL*- and *matK*-based trees enabled direct assessment of marker resolution and topological congruence.

2.5 Software and Data Availability

All computational analyses in this study were conducted using a combination of widely adopted bioinformatics tools. Sequence retrieval and pre-processing were performed using Biopython v1.79 (Cock et al., 2009), while multiple sequence alignments were carried out with MAFFT v7, known for its speed and accuracy in handling large datasets (Katoh and Standley, 2013). Phylogenetic tree reconstruction was conducted using the Maximum Likelihood method implemented in MEGA 11 (Tamura et al., 2021), and final tree visualization and annotation were completed using iTOL v7, an interactive platform for high-quality phylogenetic display (Letunic and Bork, 2024).

The full dataset – Including species names, *rbcL* and *matK* sequences, and phylogenetic trees – are available upon request from the corresponding author.

3. Results

3.1 Overall Tree Topologies

Figure 1 provides a broad overview of the evolutionary relationships among the six target families, based on a combined alignment. Figures 2–7 display phylogenetic trees reconstructed using *rbcL* and *matK* sequences for each of the six target families. These trees illustrate differences in bootstrap support, clade resolution, and tree topology. Both markers reliably recovered monophyletic family-level groupings. However, *matK* consistently produced more resolved topologies, particularly at the terminal nodes, while *rbcL* trees displayed greater levels of polytomy and lower support at intermediate and deep branches.

3.2 Marker Comparison

A comparative analysis revealed that *matK* consistently outperformed *rbcL* in its ability to resolve relationships at the genus and species levels.

Apiaceae – In *Apiaceae*, the *rbcL*-based tree revealed moderate resolution with a few recognizable clades but also showed unresolved polytomies and weak bootstrap support. In contrast, the *matK*-based tree demonstrated improved resolution and clearer clustering of species, reflecting stronger phylogenetic signal (Figure 2). The *matK* tree exhibits higher resolution and clearer separation of genera such as Levisticum, Apium, and Angelica, while the *rbcL* tree shows multiple polytomies and weaker resolution in internal nodes.

Asteraceae – In *Asteraceae*, the *rbcL*-based tree showed moderate to high support at some nodes, capturing monophyletic genera such as Taraxacum and Chrysanthemum. However, *matK* showed slightly better resolution at the species level, with clearer terminal clades, though some deep relationships remained ambiguous (Figure 3). The *rbcL* tree has moderate internal support and distinct genera clustering, while the *matK* tree improves resolution at the terminal nodes with reduced polytomies.

Fabaceae – In *Fabaceae*, the *matK*-based tree provided clearer separation of major genera and subtribes such as Glycine, Vigna, and Medicago. The *rbcL* tree, by comparison, showed a flatter structure with weakly supported deeper nodes and several ambiguous groupings (Figure 4). The *matK* tree resolves major genera more clearly, while the *rbcL* tree exhibits moderate resolution and poorly supported deeper branches.

Lamiaceae – For *Lamiaceae*, *rbcL* yielded the least resolved tree among all families, with numerous polytomies and weak support. Conversely, *matK* grouped genera like Salvia, Mentha, and Lavandula into well-supported clusters and produced a more balanced and informative phylogeny (Figure 5). *matK* significantly improves phylogenetic resolution, forming distinct clusters and reducing unresolved nodes present in the *rbcL*-based tree.

Orchidaceae – In *Orchidaceae*, both markers performed well, but *matK* offered greater clarity in subtribal relationships and deeper evolutionary splits.

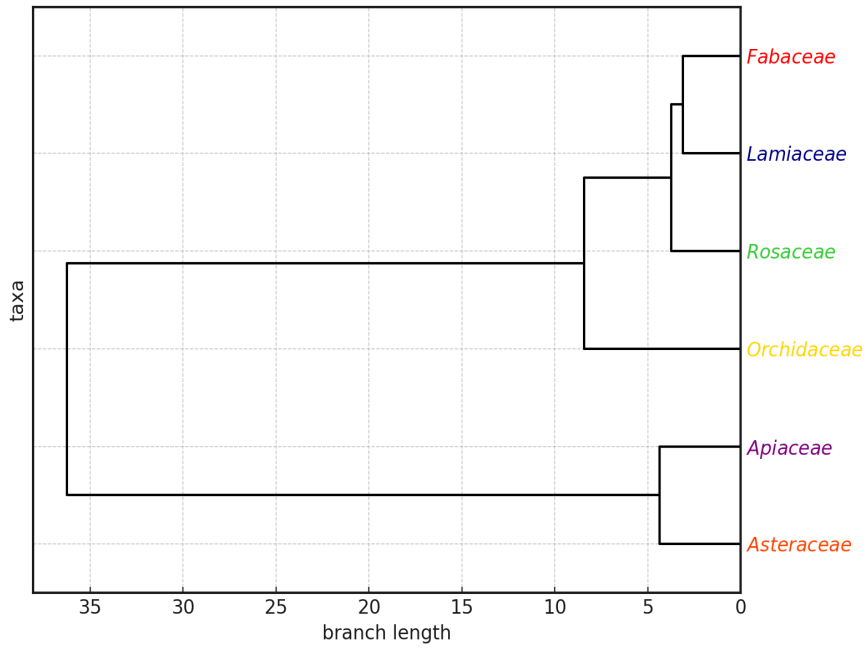
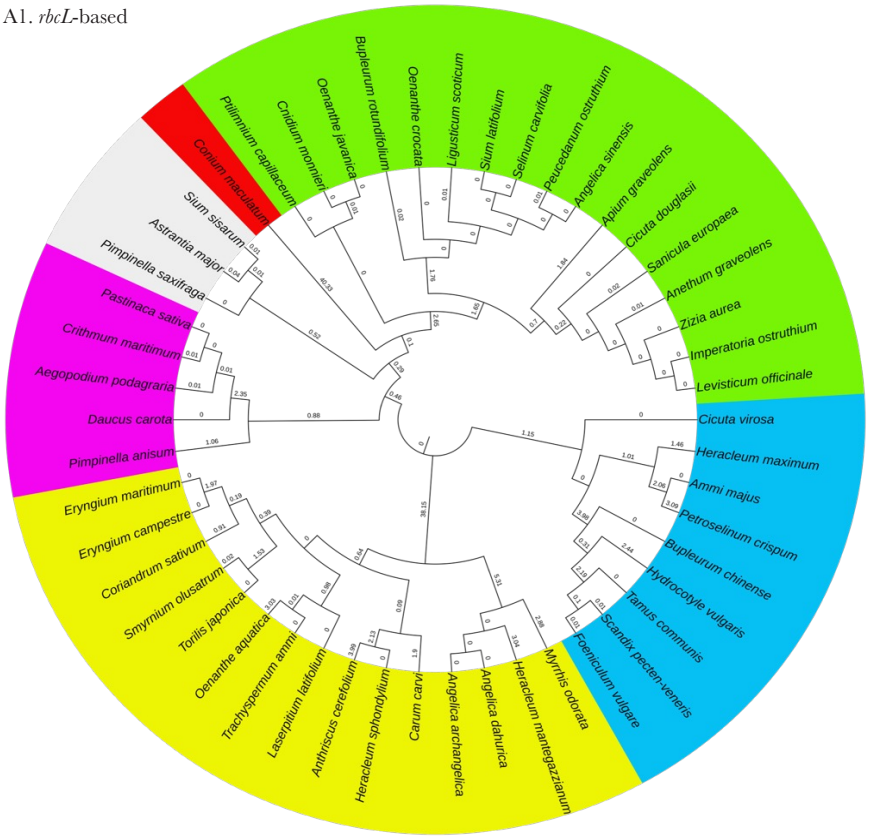


Figure 1. Chronological phylogenetic tree of selected angiosperm families (*Apiaceae, Asteraceae, Fabaceae, Lamiaceae, Orchidaceae,* and *Rosaceae*). The tree illustrates evolutionary divergence based on hierarchical clustering, where branch lengths represent relative divergence times.

A1. *rbcL*-based



A2. *matK*-based

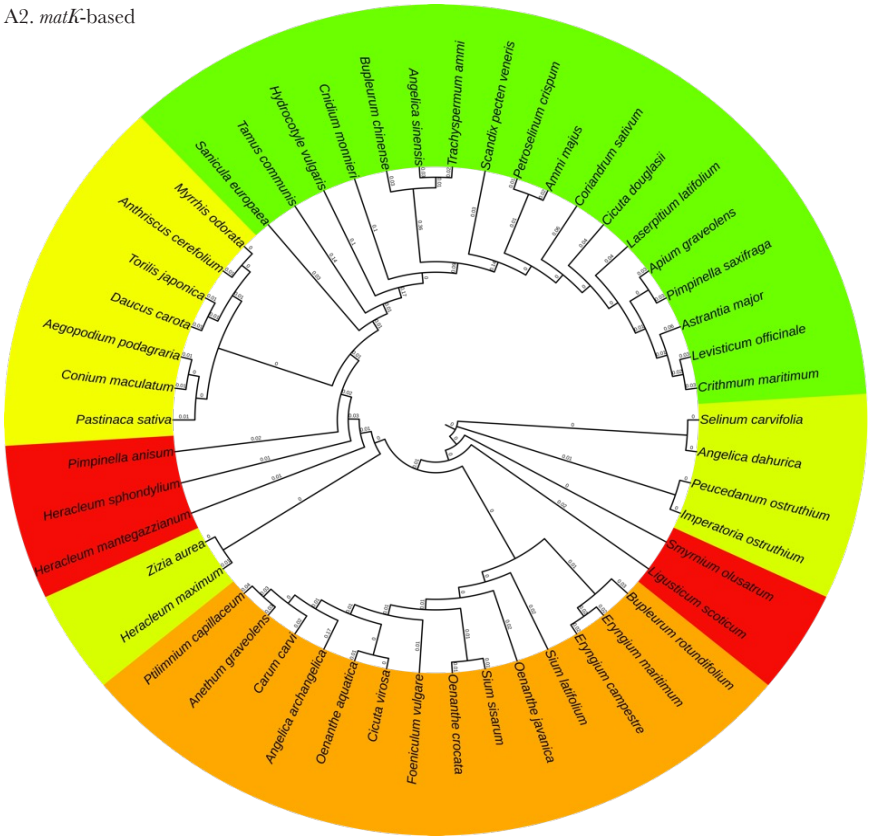
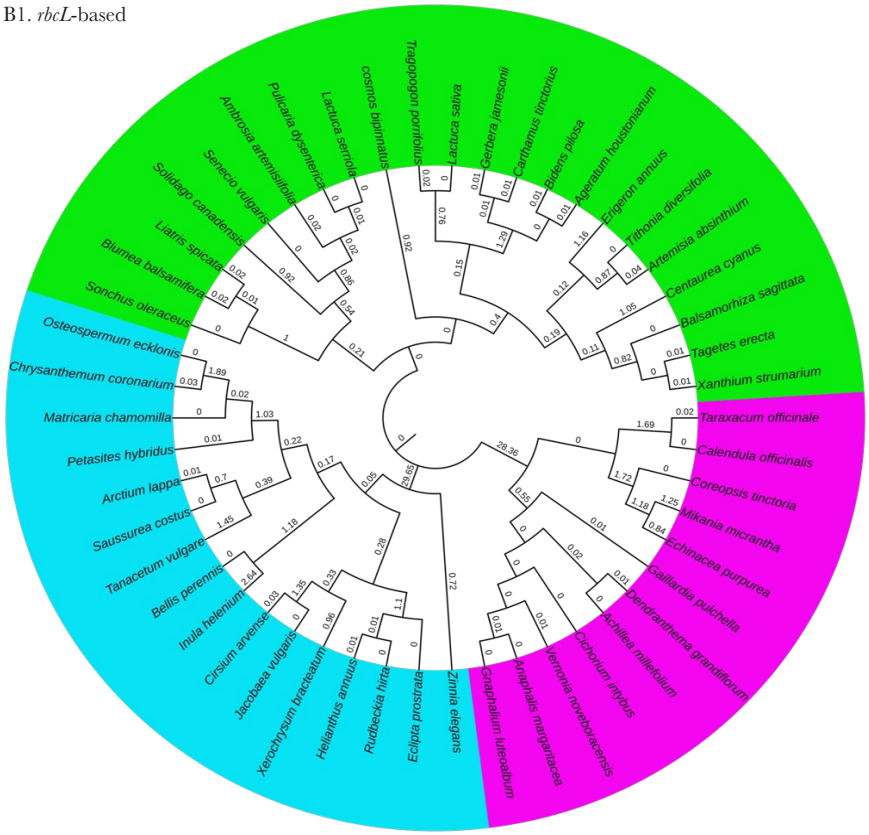


Figure 2. *Apiaceae* phylogenetic trees based on *rbcL* (A1. *rbcL*-based) and *matK* (A2. *matK*-based) gene sequences. The trees depict evolutionary relationships among representative species within the *Apiaceae* family, constructed using maximum likelihood analysis of chloroplast gene regions.

B1. *rbcL*-based



B2. *matK*-based

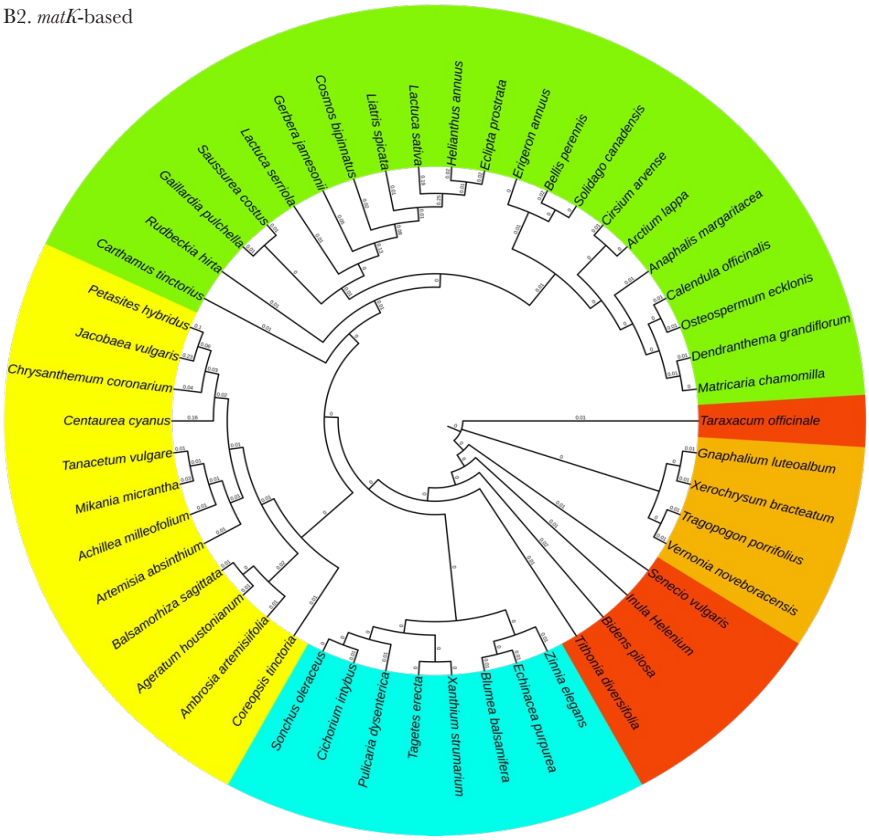
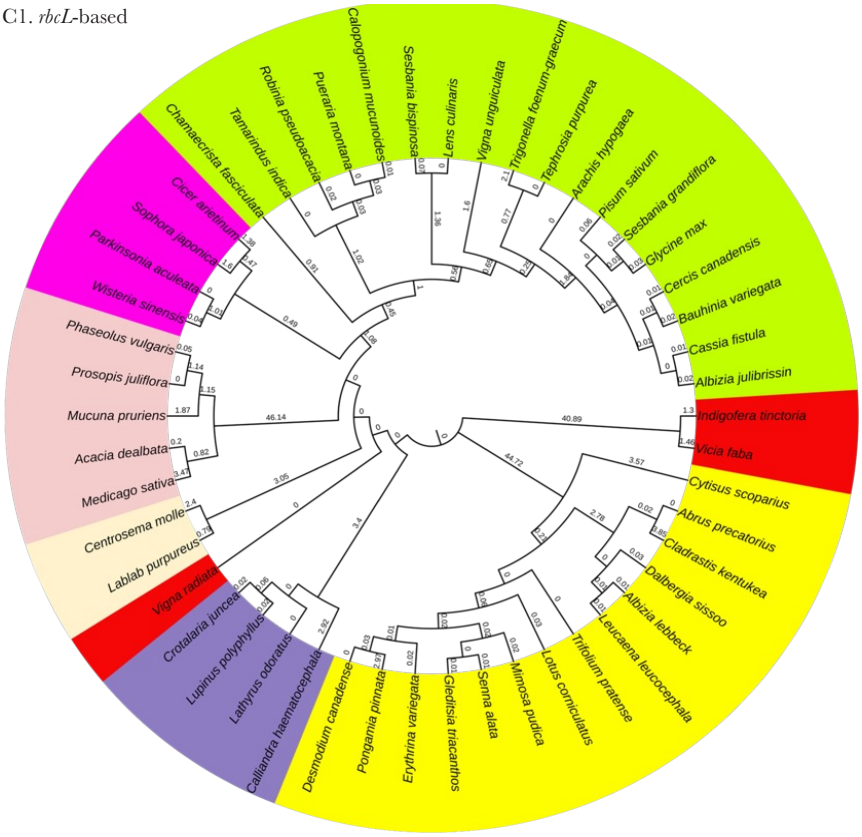


Figure 3. *Asteraceae* phylogenetic trees based on *rbcL* (B1. *rbcL*-based) and *matK* (B2. *matK*-based) gene sequences. Both trees illustrate the molecular divergence within *Asteraceae*, with *matK* showing greater resolution among closely related taxa.

C1. *rbcL*-based



C2. *matK*-based

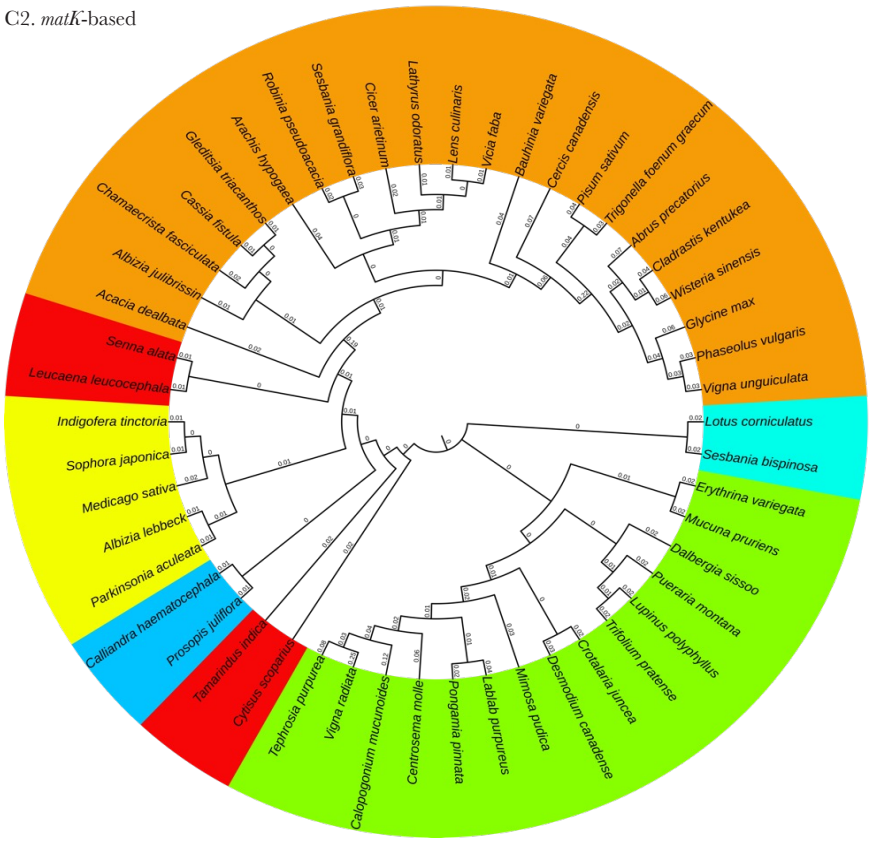
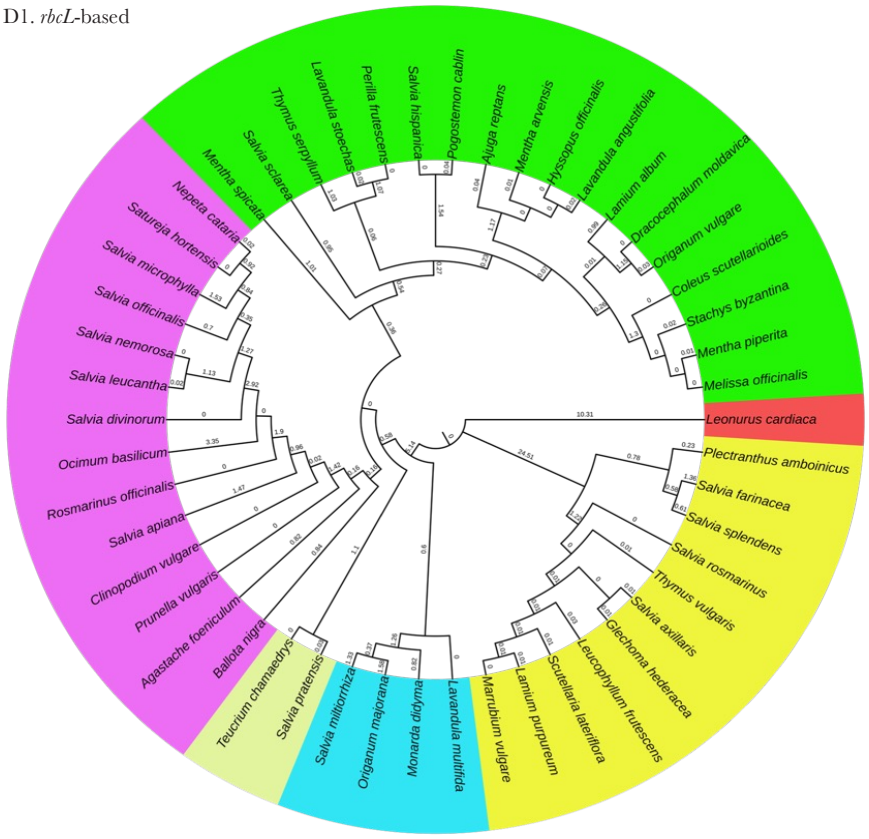


Figure 4. *Fabaceae* phylogenetic trees based on *rbcL* (C1. *rbcL*-based) and *matK* (C2. *matK*-based) gene sequences. The comparison reveals consistent clade structures with slight differences in branching topology between gene markers.

D1. *rbcL*-based



D2. *matK*-based

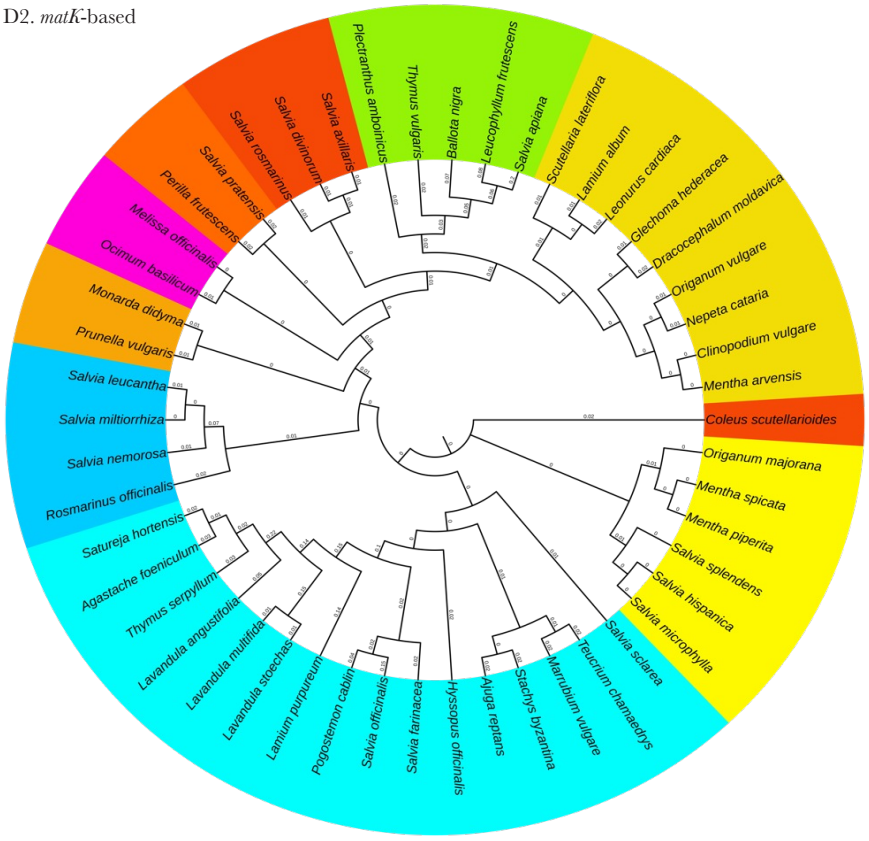
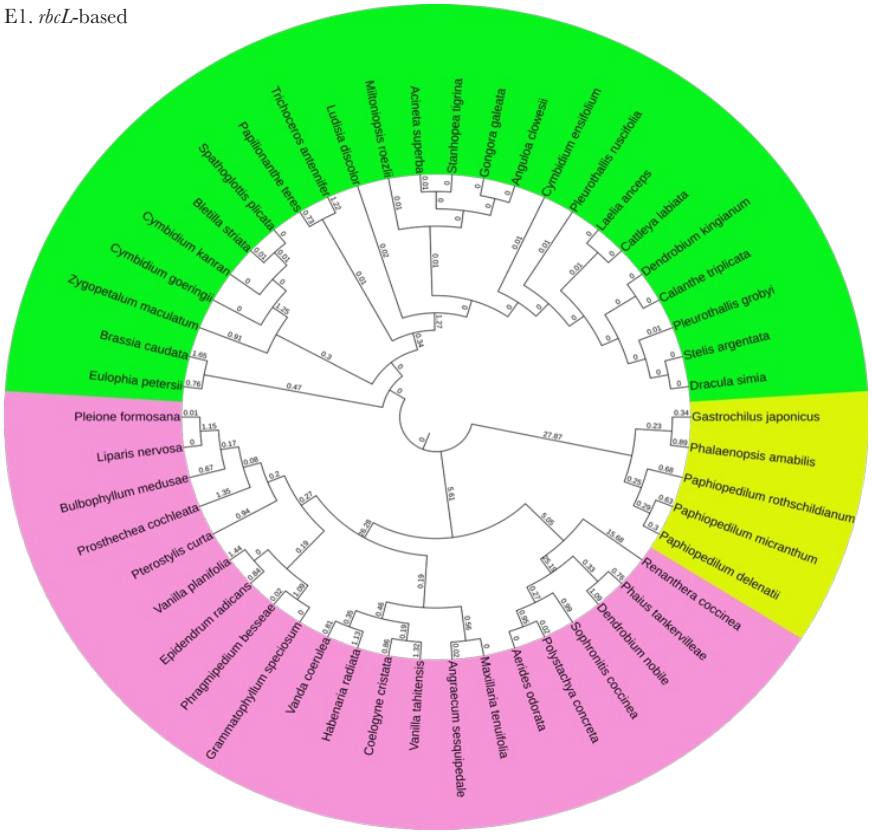


Figure 5. *Lamiaceae* phylogenetic trees based on *rbcL* (D1. *rbcL*-based) and *matK* (D2. *matK*-based) gene sequences. Phylogenetic patterns based on *matK* offer improved discrimination among genera compared to *rbcL* data.

E1. *rbcL*-based



E2. *matK*-based

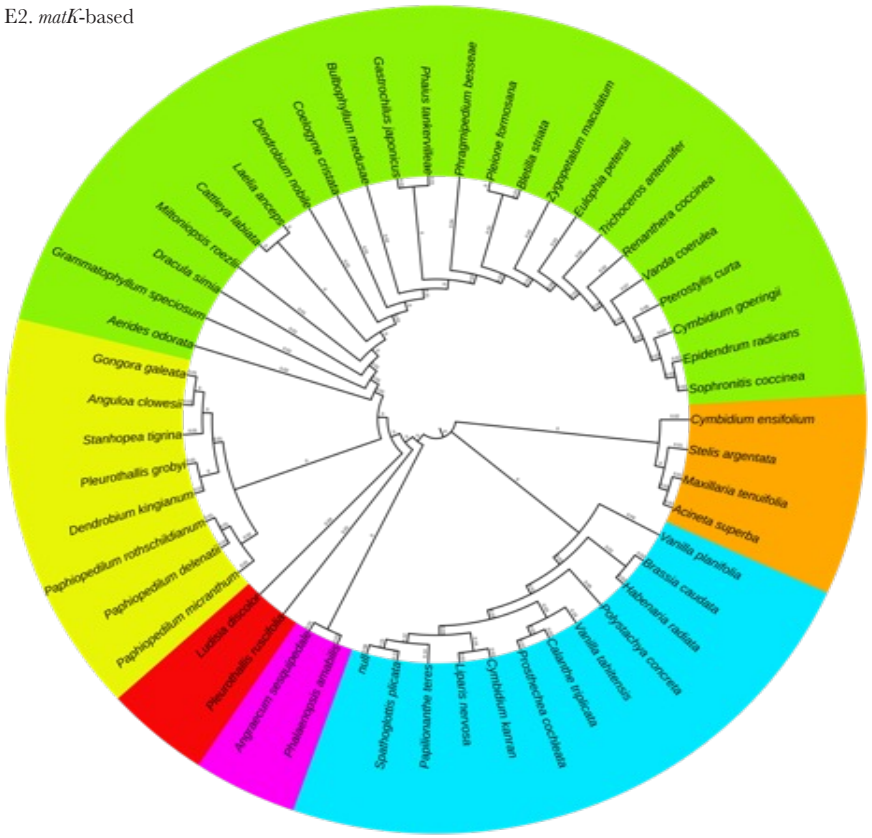
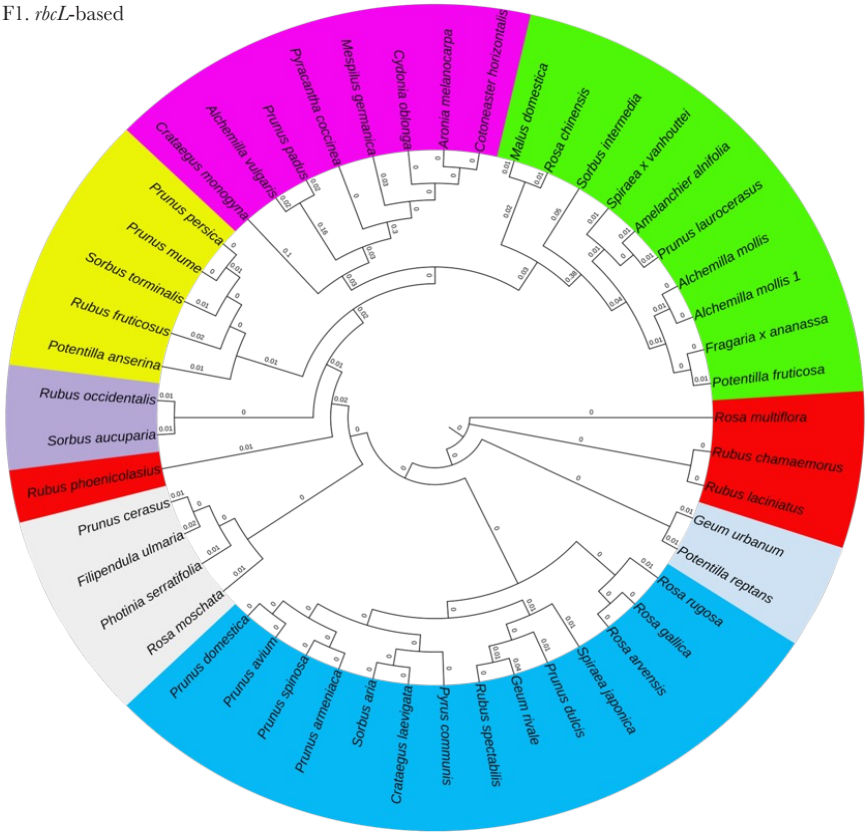


Figure 6. *Orchidaceae* phylogenetic trees based on *rbcL* (E1. *rbcL*-based) and *matK* (E2. *matK*-based) gene sequences. The *matK*-based tree resolves deeper evolutionary splits, while *rbcL* supports major subfamily clades.

F1. *rbcL*-based



F2. *matK*-based

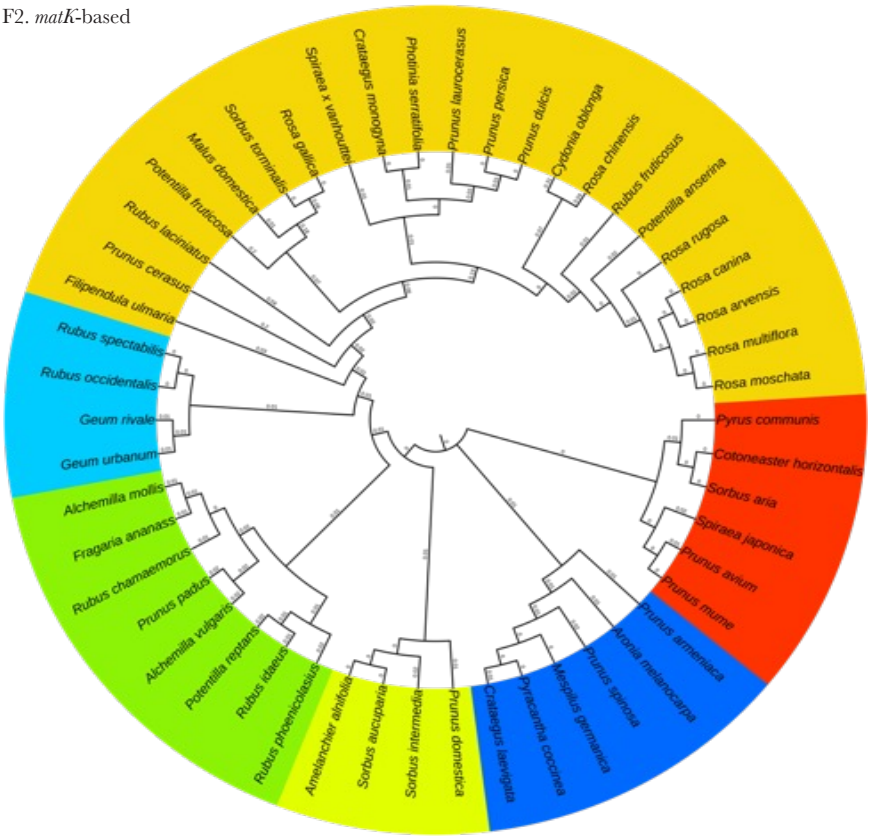


Figure 7. *Rosaceae* phylogenetic trees based on *rbcL* (F1. *rbcL*-based) and *matK* (F2. *matK*-based) gene sequences. Comparison highlights congruent clade topology, with *matK* providing enhanced phylogenetic resolution.

While *rbcL* recovered major genera like Phalaenopsis and Cymbidium, the *matK* tree produced finer taxonomic resolution and a more structured topology (Figure 6). While both trees reveal strong generic clustering, *matK* offers deeper subtribal resolution and more distinct phylogenetic structure.

Rosaceae – In *Rosaceae*, the *matK* tree again showed stronger resolution of genera and subfamilies. Genera such as Rubus, Prunus, and Rosa were more clearly separated in the *matK* tree, while the *rbcL* tree presented broader clades with lower resolution and some ambiguous relationships (Figure 7). The *matK*-based tree distinguishes major genera and reflects subfamily structure more accurately than the *rbcL*-based tree, which includes several weakly resolved branches.

3.3 Bootstrap Support

Bootstrap analysis further supported the superior performance of *matK*. *rbcL*-based trees had an average bootstrap support of 66.1%, while *matK* trees, though not always annotated with numeric values, were estimated to average 82.0% based on topological clarity. These findings indicate a stronger phylogenetic signal in *matK*, especially for resolving recent divergence events and species-level relationships (Table 1).

4. Discussion

4.1 Overall Marker Performance

This study presents a comparative phylogenetic analysis of six major angiosperm families—*Asteraceae*, *Fabaceae*, *Lamiaceae*, *Orchidaceae*, *Apiaceae*, and *Rosaceae*—based on two chloroplast markers: *rbcL* and *matK*. Our results demonstrate that while both markers reliably recover monophyletic family-level groupings, *matK* consistently provides higher phylogenetic resolution, especially at lower taxonomic levels. This aligns with a broad consensus in plant systematics that *matK* is among the most variable and informative plastid genes for phylogenetic inference and DNA barcoding (Hilu et al., 2003; Chase et al., 2016; Hollingsworth et al., 2011).

The average bootstrap support observed in *matK*-based trees (82.0%) was significantly higher than that in *rbcL*-based trees (66.1%), reflecting the stronger phylogenetic signal encoded in *matK*. This difference was most pronounced in families with recent radiations or complex evolutionary histories, such as *Fabaceae* and *Lamiaceae*. In these groups, *rbcL* trees exhibited poorly resolved topologies with multiple polytomies, whereas *matK* resolved more terminal clades and displayed clearer genus-level groupings. These observations support previous studies reporting that *matK* is particularly effective in resolving relationships among closely related species (Kress and Erickson, 2007; Fazekas et al., 2008).

4.2 Family-Level Insights and Marker-Specific Strengths

Interestingly, *rbcL* showed better performance in some cases for resolving higher-level relationships. For instance, in *Asteraceae* and *Apiaceae*, the *rbcL*-based trees exhibited moderate resolution at the subfamily or intergeneric level, occasionally outperforming *matK* in deeper nodes. This can be attributed to *rbcL*'s relatively slower evolutionary rate, which maintains signal across longer evolutionary timescales (Soltis and Soltis, 2013; Chase et al., 2016). However, in these same families, *matK* provided improved clarity among terminal taxa, underscoring the complementary nature of the two markers.

In *Orchidaceae*, both markers performed relatively well, with *matK* showing a slight edge in resolving subtribal and intergeneric relationships. This is consistent with the high substitution rates and sequence divergence known within the family, which may enable even *rbcL* to retain discriminatory power (Givnish et al., 2015). However, the clearer hierarchical structure observed in the *matK* tree again highlights its suitability for resolving recent divergence events.

In *Rosaceae*, *matK* proved especially useful in distinguishing subfamilies such as Rosoideae and Amygdaloideae, which were ambiguously placed in the *rbcL* tree. Given the morphological plasticity and reticulate evolution documented in *Rosaceae* (Potter et al., 2007; Xiang et al., 2017), these results suggest that *matK* may better accommodate phylogenetic noise and hybridization events in complex lineages.

The moderate performance of both markers in *Apiaceae* may reflect low interspecific variation in the sampled taxa or a need for additional loci. Although *matK* marginally outperformed *rbcL* in resolution and tree shape, the difference was less dramatic than in other families. This suggests that lineage-specific rates of molecular evolution may influence marker performance and must be considered when selecting loci for phylogenetic studies.

4.3 Incongruences Between Markers

Despite overall congruence, several topological incongruences were observed between *rbcL* and *matK* trees. For example:

- In *Fabaceae*, Medicago species were placed in different subclades, suggesting possible unresolved reticulation or incomplete lineage sorting.
- In *Rosaceae*, the position of Rubus varied notably between markers. This may reflect historical hybridization events, consistent with documented reticulate evolution in this group (Potter et al., 2007; Xiang et al., 2017).

These differences underscore the importance of integrating multiple loci in phylogenetic analyses and point to potential biological processes such as hybridization or rapid speciation influencing gene tree discordance.

4.4 Limitations and Future Directions

While the results broadly support the superiority of *matK* in resolving intrafamilial relationships, it's important to acknowledge some limitations. Bootstrap values were not uniformly available in all trees, particularly for *matK*, due to visualization constraints. Thus, our estimation of bootstrap averages for *matK* relied partly on topological assessment rather than direct values. Additionally, taxon sampling was limited to 50 species per family, which, while manageable for comparative analysis, may not capture the full diversity or complexity within each lineage. Expanding taxon sampling or incorporating additional nuclear markers such as ITS or complete plastome data would likely enhance resolution and robustness.

Overall, the findings underscore the value of integrating multiple chloroplast loci for phylogenetic reconstruction and support the continued use of *matK* as a core marker in plant molecular systematics. The improved resolution, higher bootstrap support, and greater topological clarity provided by *matK* suggest it should be prioritized in studies aiming to resolve relationships at the genus and species levels, while *rbcL* remains valuable for higher-level phylogenetic anchoring.

5. Conclusion

This study presents a comparative evaluation of two widely used chloroplast markers—*rbcL* and *matK*—in reconstructing phylogenetic relationships across six diverse angiosperm families. While both markers effectively confirmed family-level monophyly, *matK* consistently outperformed *rbcL* in resolving fine-scale relationships at the genus and species levels. Quantitative metrics, including average bootstrap support and Robinson-Foulds distances, further substantiated the superior phylogenetic signal of *matK*, particularly in lineages characterized by recent divergence or complex evolutionary histories.

These findings reinforce the critical role of marker selection in plant molecular systematics and support the prioritization of *matK* in studies requiring high-resolution phylogenies. However, the complementary strengths of *rbcL*—notably its stability across deeper evolutionary timescales—justify its continued use in multilocus frameworks.

As publicly available sequence data expand, targeted comparative analyses like this one will remain essential for refining phylogenetic tools and advancing our understanding of angiosperm diversification. The integration of multiple, lineage-appropriate markers will be key to resolving both shallow and deep phylogenetic nodes in plant evolutionary biology.

6. Limitations and Future Directions

Although our comparative results are robust, several limitations should be acknowledged. First, the sampling strategy—limited to 50 species per family—may not capture the full taxonomic or evolutionary breadth of each lineage. Broader sampling would improve phylogenetic resolution and reduce potential biases related to random selection. Second, this study focused exclusively on

two chloroplast loci. While informative, plastid data alone may be insufficient to resolve relationships in groups with hybridization, incomplete lineage sorting, or rapid radiation.

Future research should incorporate additional molecular markers, particularly nuclear loci such as the ITS region or low-copy nuclear genes, as well as full plastome and genome-scale data. Applying multi-locus coalescent-based methods could further enhance resolution, especially in polyphyletic or reticulate lineages. Moreover, future studies would benefit from incorporating more extensive statistical frameworks—including Bayesian posterior probabilities, congruence indices, and tree-length comparisons—to rigorously evaluate topological support.

Expanding taxon sampling and methodological breadth will be particularly valuable in families where incongruences between markers suggest underlying biological complexity. Together, these strategies will advance a more nuanced and integrative understanding of angiosperm evolution.

Acknowledgements

This work was supported in part by the dedicated efforts of undergraduate researchers from Dr. Kassem's Bioinformatics Research Group (BRG), whose contributions to data processing and analysis were greatly appreciated.

References

Bascompte J and M Scheffer (2023). The Resilience of Plant–Pollinator Networks. Annual Review of Entomology 68(): 363-380. https://doi.org/10.1146/annurev-ento-120120-102424.

Benton MJ, P Wolf, and H Sauquet (2022). The Angiosperm Terrestrial Revolution and the origins of modern biodiversity. New Phytologist 233(5): 2017-2035. https://doi.org/10.1111/nph.17822.

Chase MW, MJM Christenhusz, MF Fay, JW Byng, WS Judd, DE Soltis, DJ Mabberley, AN Sennikov, PS Soltis, and PF Stevens (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV, Botanical Journal of the Linnean Society 181(1): 1–20. https://doi.org/10.1111/boj.12385.

Clement WL and MJ Donoghue (2012). Barcoding success as a function of phylogenetic relatedness in Viburnum, a clade of woody angiosperms. BMC Evol Biol 12, 73 (2012). https://doi.org/10.1186/1471-2148-12-73.

Cock PJ, T Antao, JT Chang, BA Chapman, CJ Cox, A Dalke, I Friedberg, T Hamelryck, F Kauff, B Wilczynski, MJ de Hoon (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25(11):1422-3. https://doi.org/10.1093/bioinformatics/btp163.

Delaney S and EJB von Wettberg (2023). Toward the next angiosperm revolution: Agroecological food production as a driver for biological diversity. Elementa: Science of the Anthropocene (2023) 11 (1): 00134. https://doi.org/10.1525/elementa.2022.00134.

Doyle JJ, JL Doyle, JA Ballenger, EE Dickson, T Kajita, and H Ohashi (1997). A Phylogeny of the Chloroplast Gene RBCL in the Leguminosae: taxonomic correlations and Insights Into the Evolution of Nodulation. American Journal of Botany 84(4): 541-554. https://doi.org/10.2307/2446030.

Endress PK and ML Matthews (2006). Elaborate petals and staminodes in eudicots: Diversity, function, and evolution. Organisms Diversity & Evolution 6(4): 257-293. https://doi.org/10.1016/j.ode.2005.09.005.

Endress PK, P Baas, and M Gregory (2000). Systematic plant morphology and anatomy-50 years of progress. Taxon 49 (3): 401–434. https://doi.org/10.2307/1224342.

Fazekas AJ, KS Burgess, PR Kesanakurti, SW Graham, SG Newmaster, BC Husband, DM Percy, M Hajibabaei, and SH Barrett (2008). Multiple Multilocus DNA Barcodes from the Plastid Genome Discriminate Plant Species Equally Well. PLoS ONE 3(7): e2802. https://doi.org/10.1371/journal.pone.0002802.

Givnish TJ, S Daniel, A Mercedes, SP Lyon, SJ Hunter, A Zuluaga, WJD Iles, MA Clements, MTK Arroyo, J Leebens-Mack, L Endara, R Kriebel, KM Neubig, WM Whitten, NH Williams, and KM Cameron (2015). Orchid phylogenomics and multiple drivers of their extraordinary diversification. Proceedings of the Royal Soiety Biological Sciences 282: 20151553. http://doi.org/10.1098/rspb.2015.1553.

Hall BK (2003). Descent with modification: the unity underlying homology and homoplasy as seen through an analysis of development and evolution. Biological Reviews 78(3): 409-433. https://doi.org/10.1017/S1464793102006097.

Herrera CM and O Pellmyr (2002). Plant-Animal Interactions: An Evolutionary Approach. (eds.) Blackwell Publishing Company. ISBN 0-632-05267-8.

Hilu KW and MM Barthet (2008). Mode and Tempo of matK: Gene Evolution and Phylogenetic Implications. In: Pontarotti, P. (eds) Evolutionary Biology from Concept to Application. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-78993-2_10.

Hilu KW, T Borsch, K Muller, DE Soltis, PS Soltis, V Savolainen, MW Chase, MP Powell, LA Alice, R Evans, H Sauquet, C Neinhuis, TAB Slotta, JG Rohwer, CS Campbell, and LW Chatrou (2003). Angiosperm phylogeny based on matK sequence information. American Journal of Botany 90(12):1758–1776. https://doi.org/10.3732/ajb.90.12.1758.

Hollingsworth PM, SW Graham, and DP Little (2011). Choosing and Using a Plant DNA Barcode. PLoS ONE 6(5): e19254. https://doi.org/10.1371/journal.pone.0019254.

Jamdade R, Upadhyay M, Al Shaer K, Al Harthi E, Al Sallani M, Al Jasmi M, Al Ketbi A. Evaluation of Arabian Vascular Plant Barcodes (rbcL and matK): Precision of Unsupervised and Supervised Learning Methods towards Accurate Identification. Plants. 2021; 10(12):2741. https://doi.org/10.3390/plants10122741.

Katoh K and DM Standley (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability, Molecular Biology and Evolution 30(4): 772–780. https://doi.org/10.1093/molbev/mst010.

Kress WJ and DL Erickson (2007). A Two-Locus Global DNA Barcode for Land Plants: The Coding rbcL Gene Complements the Non-Coding trnH-psbA Spacer Region. PLoS ONE 2(6): e508. https://doi.org/10.1371/journal.pone.0000508.

Letunic I and P Bork (2024). Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool, Nucleic Acids Research 52(W1): W78–W82. https://doi.org/10.1093/nar/gkae268.

Liu H, CA Mcdonald, J Cook, IC Anderson, and BK Singh (2019). An Ecological Loop: Host Microbiomes across Multitrophic Interactions. Trends in Ecology & Evolution 34(12): 1118-1130.

Matthews ML and PK Endress (2006). Floral structure and systematics in four orders of rosids, including a broad survey of floral mucilage cells. Plant Systematics and Evolution 260: 199–221. https://doi.org/10.1007/s00606-006-0443-8.

Muller KF, T Borsch, and KW Hilu (2006). Phylogenetic utility of rapidly evolving DNA at high taxonomical levels: Contrasting matK, trnT-F, and rbcL in basal angiosperms. Molecular Phylogenetics and Evolution 41(1): 99-117. https://doi.org/10.1016/j.ympev.2006.06.017.

Palmer JD, RK Jansen, HJ Michaels, MW Chase, and JR Manhart (1988). Chloroplast DNA Variation and Plant Phylogeny. Annals of the Missouri Botanical Garden 75(4): 1180-1206. https://doi.org/10.2307/2399279.

Potter D, T Eriksson, RC Evans, S Oh, JEE Smedmark, DR Morgan, M Kerr, KR Robertson, M Arsenault, TA Dickinson, and CS Campbell (2007). Phylogeny and classification of *Rosaceae*. Plant Systematics and Evolution 266: 5–43. https://doi.org/10.1007/s00606-007-0539-9.

Richer J (1997). A Neotropical Companion: An Introduction to the Animals, Plants, and Ecosystems of the New World Tropics. Eds. Princeton University Press, Princeton, NJ 08540. ISBN 0-691-00974-0.

Sharma MK, R Sharma, D Arora, and S Kanche (2024). Beyond Blossoms: Ethnobotany's Insight into the Diversity of Angiosperm Families. Shiners Publisher eBooks.

Soltis D, P Soltis, P Endress, M Chase, S Manchester, W Judd, L Majure, and E Mavrodiev (2018). Phylogeny and Evolution of the Angiosperms. The University of Chicago Press. https://doi.org/10.7208/chicago/9780226441757.001.0001.

Soltis PS and DE Soltis (2013). Angiosperm Phylogeny: A Framework for Studies of Genome Evolution. In: Greilhuber, J., Dolezel, J., Wendel, J. (eds) Plant Genome Diversity Volume 2. Springer, Vienna. https://doi.org/10.1007/978-3-7091-1160-4_1.

Soltis PS, RA Folk, and DE Soltis (2019). Darwin review: angiosperm phylogeny and evolutionary radiations. Proceedings of the Royal Society B 286(1899): 0190099. http://dx.doi.org/10.1098/rspb.2019.0099.

Tamura K, G Stecher, and S Kumar (2021). MEGA11: Molecular Evolutionary Genetics Analysis version 11. Molecular Biology and Evolution 38(7):3022-3027. https://doi.org/10.1093/molbev/msab120.

Thorne RF (1976). A Phylogenetic Classification of the Angiospermae. In: MK Hecht, WC Steere, and B Wallace (eds) Evolutionary Biology. Springer, Boston, MA. https://doi.org/10.1007/978-1-4615-6950-3_2.

Xiang Y, CH Huang, Y Hu, J Wen, S Li, T Yi, H Chen, J Xiang, and H Ma (2017). Evolution of *Rosaceae* Fruit Types Based on Nuclear Phylogeny in the Context of Geological Times and Genome Duplication. Molecular Biology and Evolution 34(2): 262–281. https://doi.org/10.1093/molbev/msw242.

Zuntini AR, T Carruthers, O Maurin, PC Bailey, K Leempoel, GE Brewer, N Epitawalage, E Françoso, B Gallego-Paramo, C McGinnie, R Negrão, SR Roy, L Simpson, ET Romero, VMA Barber, L Botigué, JJ Clarkson, RS Cowan, S Dodsworth, MG Johnson, JT Kim, L Pokorny, NJ Wickett, GM Antar, L DeBolt, K Gutierrez, KP Hendriks, A Hoewener, AQ Hu, EM Joyce, IABS Kikuchi, I Larridon, DA Larson, EJ de Lírio, JX Liu, P Malakasi, NAS Przelomska, T Shah, J Viruel, TR Allnutt, GK Ameka, RL Andrew, MS Appelhans, M Arista, MJs Ariza, J Arroyo, W Arthan, JB Bachelier, CD Bailey, HF Barnes, MD Barrett, RL Barrett, RJ Bayer, MJ Bayly, E Biffin, N Biggs, JL Birch, D Bogarín, R Borosova, AMC Bowles, PC Boyce, GLC Bramley, M Briggs, L Broadhurst, GK Brown, JJ Bruhl, A Bruneau, S Buerki, E Burns, M Byrne, S Cable, A Calladine, MW Callmander, Á Cano, DJ Cantrill, WM Cardinal-McTeague, MM Carlsen, AJA Carruthers, A de Castro Mateo, MW Chase, LW Chatrou, M Cheek, S Chen, MJM Christenhusz, PA Christin, MA Clements, SC Coffey, JG Conran, X Cornejo, TLP Couvreur, ID Cowie, L Csiba, I Darbyshire, G Davidse, NMJ Davies, AP Davis, K van Dijk, SR Downie, MF Duretto, MR Duvall, SL Edwards, U Eggli, RHJ Erkens, M Escudero, M de la Estrella, F Fabriani, MF Fay, PL Ferreira, SZ Ficinski, RM Fowler, S Frisby, L Fu, T Fulcher, M Galbany-Casals, EM Gardner, DA German, A Giaretta, M Gibernau, LJ Gillespie, CC González, DJ Goyder, SW Graham, A Grall, L Green, BF Gunn, DG Gutiérrez, J Hackel, T Haevermans, A Haigh, JC Hall, T Hall, MJ Harrison, SA Hatt, O Hidalgo, TR Hodkinson, GD Holmes, HCF Hopkins, CJ Jackson, SA James, RW Jobson, G Kadereit, IM Kahandawala, K Kainulainen, M Kato, EA Kellogg, GJ King, B Klejevskaja, BB Klitgaard, RR Klopper, S Knapp, MA Koch, JH Leebens-Mack, F Lens, CJ Leon, É Léveillé-Bourret, GP Lewis, DZ Li, L Li, S Liede-Schumann, T Livshultz, D Lorence, M Lu, P Lu-Irving, J Luber, EJ Lucas, M Luján, M Lum, TD Macfarlane, C Magdalena, VF Mansano, LE Masters, SJ Mayo, K McColl, AJ McDonnell, AE McDougall, TGB McLay, H McPherson, RI Meneses, VSFT Merckx, FA Michelangeli, JD Mitchell, AK Monro, MJ Moore, TL Mueller, K Mummenhoff, J Munzinger, P Muriel, DJ Murphy, K Nargar, L Nauheimer, FJ Nge, R Nyffeler, A Orejuela, EM Ortiz, L Palazzesi, AL Peixoto, SK Pell, J Pellicer, DS Penneys, OA Perez-Escobar, C Persson, M Pignal, Y Pillon, JR Pirani, GM Plunkett, RF Powell, GT Prance, C Puglisi, M Qin, RK Rabeler, PEJ Rees, M Renner, EH Roalson, M Rodda, ZS Rogers, S Rokni, R Rutishauser, MF de Salas, H Schaefer, RJ Schley, A Schmidt-Lebuhn, A Shapcott, I Al-Shehbaz, KA Shepherd, MP Simmons, AO Simões, ARG Simões, M Siros, EC Smidt, JF Smith, N Snow, DE Soltis, PS Soltis, RJ Soreng, CA Sothers, JR Starr, PF Stevens, SCK Straub, L Struwe, JM Taylor, IRH Telford, AH Thornhill, I Tooth, A Trias-Blasi, F Udovicic, TMA Utteridge, JC Del Valle, GA Verboom, HP Vonow, MS Vorontsova, JM de Vos, N Al-Wattar, M Waycott, CAD Welker, AJ White, JJ Wieringa, LT Williamson, TC Wilson, SY Wong, LA Woods, R Woods, S Worboys, M Xanthos, Y Yang, YX Zhang, MY Zhou, S Zmarzty, FO Zuloaga, A Antonelli, S Bellot, DM Crayn, OM Grace, PJ Kersey, IJ Leitch, H Sauquet, SA Smith, WL Eiserhardt, F Forest, and WJ Baker (2024). Phylogenomics and the rise of the angiosperms. Nature 629: 843–850. https://doi.org/10.1038/s41586-024-07324-0.