Buitenzorg
Journal of Tropical Science

Research Article

Comparative Evaluation of ITS1, ITS2, and *LEAFY* (*LFY*) CDS Markers for Species Discrimination in Asteraceae

Aulia Fitriana Ardhyatul Jannah[1], Ira Erdiandini[2], Mufidah[1], Thoriq Ahmad Syauqy[3], & Aqwin Polosoro[4]*

[1] Innovation Centre for Tropical Sciences (ICTS), Bogor 16113, West Java, Indonesia
[2] Department of Agrotechnology, Universitas Tanjungpura, Jalan Prof. Dr. H. Hadari Nawawi, Pontianak 78124, West Kalimantan, Indonesia
[3] Department of Agronomy and Horticulture, Faculty of Agriculture, IPB University, Jalan Meranti, IPB Dramaga Campus, Bogor 16680, West Java, Indonesia
[4] Research Center for Genetic Engineering, National Research and Innovation Agency, Jalan Raya Jakarta-Bogor Km. 46, Cibinong, Bogor 16911, West Java, Indonesia

*Corresponding author: aqwi001@brin.go.id

ABSTRACT

Asteraceae is one of the largest and most diverse angiosperm families, making species identification challenging, especially among closely related taxa with overlapping morphological characters. This study evaluated the performance of internal transcribed spacer (ITS), namely ITS1 and ITS2, and the coding sequence (CDS) of the *LEAFY* (*LFY*) gene as molecular markers for discriminating selected Asteraceae species and assessing relationships at the genus, tribe, and subfamily levels. Publicly available sequences were retrieved from NCBI databases and curated based on taxonomic identity, marker annotation, sequence quality, and accession consistency. The ITS1, ITS2, and *LFY* CDS datasets were aligned separately using MUSCLE in MEGA, followed by sequence variation analysis, Maximum Likelihood phylogenetic reconstruction, and pairwise genetic distance analysis. All three markers recovered broad taxonomic patterns, including the clustering of *Chrysanthemum* species and the separation of more distant genera, such as *Helianthus*, *Tagetes*, *Erigeron*, *Cynara*, and *Lactuca*. *LFY* CDS showed the highest absolute number of variable sites and parsimony-informative sites, followed by ITS2 and ITS1. ITS2 provided slightly greater phylogenetic information than ITS1, while *LFY* CDS provided an independent nuclear coding signal that complemented the ribosomal ITS markers. Overall, each marker contributed useful phylogenetic information for assessing taxonomic relationships within Asteraceae.

Keywords: Asteraceae, ITS1, ITS2, *LFY*, DNA barcoding

Introduction

Asteraceae, also known as Compositae, is one of the largest and most diverse families of flowering plants. Gao et al. (2010) reported more than 1,600 genera and approximately 23,000 species, while the Global Compositae Database currently records more than 33,000 accepted species. The family contains many economically, medicinally, ornamentally, and ecologically important taxa, including *Helianthus annuus* L., *Lactuca sativa* L., *Artemisia annua* L., *Chrysanthemum morifolium* Ramat., *Tagetes erecta* L.,

Ageratum conyzoides L., *Chromolaena odorata* (L.) R.M. King & H. Rob., and *Tithonia diversifolia* (Hemsl.) A. Gray. However, species identification in Asteraceae remains difficult because of morphological similarity, phenotypic plasticity, hybridization, polyploidy, and rapid diversification. In addition, the capitulum, a highly condensed inflorescence composed of multiple florets that resembles a single flower, contributes to the morphological complexity of the family (Elomaa et al., 2018; Zhao et al., 2016).

DNA barcoding is widely used for species identification through short and standardized DNA regions. In land plants, *rbcL* and *matK* were recommended as the core plastid barcode because of their broad universality and sequence recoverability (CBOL Plant Working Group, 2009). However, these plastid loci often have limited resolution among closely related species due to their conserved evolutionary rate and possible sharing of chloroplast haplotypes. Consequently, nuclear ribosomal regions, especially internal transcribed spacer (ITS), namely ITS1 and ITS2, have been widely used as complementary markers because they generally provide higher sequence variation and improved discrimination at lower taxonomic levels (Awaliah & Polosoro, 2024; Chen et al., 2010; China Plant BOL Group, 2011; Hollingsworth et al., 2011).

The ITS region of nuclear ribosomal DNA, comprising ITS1, 5.8S rDNA, and ITS2, is one of the most widely used nuclear markers in plant systematics and DNA barcoding. ITS has been extensively applied in phylogenetic inference at generic and infrageneric levels because it usually provides higher sequence variation than many plastid loci (Álvarez & Wendel, 2003). In Asteraceae, ITS2 has shown strong discriminatory potential. Gao et al. (2010) evaluated five candidate barcode regions, namely *rbcL*, *matK*, ITS, ITS2, and *psbA-trnH*, and reported that ITS2 was the most useful marker based on universality, sequence variation, and identification capability. Their analysis of 3,490 ITS2 sequences representing 2,315 species and 494 genera showed correct identification rates of 76.4% at the species level and 97.4% at the genus level. However, ITS belongs to a multicopy nuclear ribosomal DNA array, and its use may be affected by intragenomic variation, incomplete concerted evolution, pseudogenes, paralogous copies, and possible contamination, especially when using public database sequences or analyzing taxa with recent radiation, hybridization, or polyploidy (Bailey et al., 2003; Gao et al., 2010).

Single-copy or low-copy nuclear genes may provide complementary information to multicopy ribosomal DNA markers because they can represent orthologous nuclear variation more directly and may improve phylogenetic inference in plants (Sang, 2002). One potential candidate is *LEAFY* (*LFY*) gene, a plant-specific transcription factor gene involved in floral meristem identity and floral development. *LFY* is generally maintained as a single-copy gene in most angiosperms, although duplicated paralogs may occur in recent polyploids (Baum et al., 2005). Previous studies have also suggested that *LFY*-associated regions, particularly the second intron of *FLORICAULA*/*LEAFY*, can provide useful phylogenetic information at lower taxonomic levels (Grob et al., 2004). However, because the present study focuses on the *LFY* CDS, which is expected to be more functionally constrained than intronic regions, *LFY* CDS should be evaluated as a potentially cleaner but possibly less variable nuclear marker rather than assumed to outperform ITS1 or ITS2.

Asteraceae is one of the largest and most taxonomically complex families of flowering plants, containing numerous species with overlapping morphological characteristics. Accurate species identification is essential for taxonomy, conservation, evolutionary studies, and

breeding programs. Although ITS1, ITS2, and *LFY* CDS have been widely used as molecular markers in plant molecular systematics, most studies have evaluated these markers independently. Direct comparisons of their discriminatory power and phylogenetic utility within Asteraceae remain scarce. This study aimed to compare the performance of ITS1, ITS2, and *LFY* CDS markers in species discrimination and phylogenetic inference across selected Asteraceae species using sequence similarity, genetic distance, and phylogenetic analysis.

Materials and Methods

Study Design

The study was designed to compare the effectiveness of ITS1, ITS2, and *LFY* CDS markers for species discrimination in Asteraceae. The main steps of the study were sequence retrieval, data curation, marker validation, sequence alignment, phylogenetic reconstruction, and species discrimination analyses.

In the present study, ITS1, ITS2, and *LFY* CDS were evaluated using publicly available and quality-filtered sequence data from the National Center for Biotechnology Information (NCBI) databases. ITS1 and ITS2 sequences were retrieved primarily from the NCBI GenBank/Nucleotide database, whereas *LFY* CDS sequences were obtained from available nucleotide, mRNA, transcript, or RefSeq RNA records. GenBank is a comprehensive public database of nucleotide sequences and associated biological annotations maintained by NCBI (Sayers et al., 2024), while NCBI Taxonomy was used to verify the taxonomic placement of retrieved Asteraceae accessions (Schoch et al., 2020). This *in silico* sequence-mining approach enabled broad comparative sampling across selected Asteraceae taxa without initial wet-laboratory sequencing. All retrieved sequences were subjected to quality filtering, taxonomic validation, marker verification, and removal of incomplete, redundant, ambiguously annotated, or potentially misidentified records. The curated datasets were then subjected to taxonomic verification, sequence quality filtering, marker validation, multiple sequence alignment, phylogenetic reconstruction, and pairwise genetic distance analysis.

Sequence Retrieval from NCBI

Sequence data were mined from the NCBI GenBank database using a marker-specific strategy. *LFY* CDS sequences were retrieved using BLASTn searches against NCBI RNA-related databases, including mRNA, transcript, and RefSeq RNA records. Representative *LFY* sequences from Asteraceae were used as query sequences, and searches were restricted to the family Asteraceae using the organism or taxonomy filter.

In contrast, ITS sequences were retrieved from the NCBI Nucleotide/GenBank database (https://www.ncbi.nlm.nih.gov/genbank/) in May 2026 using a combination of keyword-based searches and BLASTn searches with representative ITS sequences. Search terms included “Asteraceae ITS”, “Asteraceae internal transcribed spacer”, “ITS1”, and “ITS2”. This approach was used because ITS is a nuclear ribosomal spacer region and is more commonly deposited as nucleotide sequence data rather than RNA or transcript data. The complete list of species names, accession numbers, and sequence lengths is provided in Table 1.

Sequence Selection and Quality Filtering

Retrieved sequences were filtered using marker-specific criteria. ITS1 and ITS2 records were retained only when assigned to Asteraceae and clearly annotated as ITS, ITS1, ITS2, or internal transcribed spacer sequences provided in GenBank records. *LFY* records were retained when annotated as *LFY* CDS, mRNA, transcript, or RefSeq RNA records. All retained sequences were required to have sufficient length, clear marker

identity, reliable species-level taxonomic assignment, and minimal ambiguous nucleotide characters.

Sequences with uncertain taxonomy, incomplete or inconsistent annotation, excessive ambiguous bases, abnormal length, very short fragments, or suspected misannotation were excluded. Duplicate, identical, or highly redundant records were removed where appropriate. Sequences were retained when they showed ≥ 90% sequence identity, ≥ 90% query coverage. Scientific names and taxonomic placement were verified using NCBI Taxonomy before downstream analysis.

Validation of LFY CDS, ITS1, and ITS2 Sequences

Candidate *LFY* CDS sequences were validated using BLASTn similarity searches, open reading frame inspection, and amino acid translation. Because *LFY* is a protein-coding gene, sequences were retained only when they contained an intact coding region and translated into amino acid sequences consistent with known *LFY* proteins. Sequences with frameshifts, premature stop codons, truncated coding regions, inconsistent annotations, or similarity to unrelated genes were excluded. Sequences were retained when they showed ≥ 90% sequence identity, ≥ 90% query coverage, and an e-value ≤ 1e-20 relative to the *LFY* reference sequence.

Candidate ITS1 and ITS2 sequences were validated separately based on marker annotation, sequence length, BLASTn similarity to known ITS regions, and sequence quality. Because ITS1 and ITS2 are non-coding ribosomal spacer regions, they were not translated. Sequences with abnormal length, unclear identity, excessive ambiguous bases, poor quality, or suspected contamination were removed before alignment.

Multiple Sequence Alignment, Sequence Variation, Phylogenetic, and Pairwise Distance Analyses

Curated ITS1, ITS2, and *LFY* CDS datasets were analyzed separately using MEGA software. Multiple sequence alignment was performed using the MUSCLE algorithm, which is widely used for accurate and high-throughput sequence alignment (Edgar, 2004). ITS1 and ITS2 sequences were aligned as nucleotide sequences, whereas *LFY* CDS sequences were aligned while maintaining the coding frame. Poorly aligned regions, ambiguous sites, and low-quality sequence ends were manually trimmed before further analysis.

Sequence variation was assessed based on alignment length, conserved sites, variable sites, parsimony-informative sites, nucleotide composition, and taxonomic coverage. Phylogenetic trees were reconstructed separately for ITS1, ITS2, and *LFY* CDS using the Maximum Likelihood method in MEGA, with the best-fit nucleotide substitution model selected before tree construction and branch support evaluated using 1,000 bootstrap replicates (Kumar et al., 2024). Pairwise genetic distances were calculated to compare interspecific divergence. A marker was considered effective when it showed clear interspecific divergence and clustered taxa according to species.

Pairwise distance matrices generated from ITS1, ITS2, and *LFY* CDS sequence datasets were imported into Heatmapper2 (https://heatmapper2.ca/site/pairwise/index.html) with pairwise distance values displayed as a continuous color scale for visualization. Heatmaps were generated using pairwise distance values as input. A hierarchical clustering dendrogram and heatmap arrangement were used to facilitate the visualization of genetic relationships among species. Hierarchical clustering was then performed using the average linkage method based on the pairwise distance matrix, and the resulting leaf order was used to arrange taxa with similar sequence profiles closer together in the heatmap. Darker colors indicate lower genetic distances, whereas brighter colors indicate higher genetic distances. Final figures were exported as high-resolution PNG and PDF files.

Results and Discussion

Dataset Composition and Marker Availability

Based on Table 1, the dataset consisted of 16 Asteraceae species representing three subfamilies, six tribes, and eight genera. The largest proportion of the dataset belonged to subfamily Asteroideae, which was represented by 13 species, followed by Cichorioideae with two species and Carduoideae with one species. At the tribal level, the dataset included Anthemideae, Astereae, Heliantheae, Tageteae, Cichorieae, and Cardueae. This taxonomic structure was intentionally designed to evaluate marker performance at different discriminatory levels, from closely related species within the same genus to more divergent taxa across genera, tribes, and subfamilies. Broad sampling across major Asteraceae lineages is important because molecular phylogenetic studies have shown that the family contains several deeply structured subfamilial and tribal lineages (Panero & Funk, 2002).

The genus *Chrysanthemum* was deliberately overrepresented, with eight species: *C. indicum*, *C. chanetii*, *C. zawadzkii*, *C. potentilloides*, *C. dichrum*, *C. lavandulifolium*, *C. oreastrum*, and *C. hypargyrum*. This sampling design provides a focused test of whether ITS1, ITS2, and *LFY* CDS can distinguish species within a closely related genus. Such within-genus resolution is critical for DNA-based species identification because closely related Asteraceae taxa often show high morphological similarity and may be difficult to distinguish using morphology alone. The inclusion of additional Anthemideae genera, namely *Phaeostigma* and *Argyranthemum*, further allows the evaluation of whether the markers can separate species within the same tribe but belong to different genera.

To test broader taxonomic discrimination, the dataset also included representatives from other Asteroideae tribes, namely *Erigeron canadensis* from Astereae, *Helianthus annuus* from Heliantheae, and *Tagetes erecta* from Tageteae. In addition, *Lactuca saligna* and *Lactuca sativa* were included as representatives of Cichorioideae–Cichorieae, while *Cynara cardunculus* represented Carduoideae–Cardueae. This broader sampling enables assessment of whether each marker can recover expected clustering not only at the species and genus

Table 1. Taxonomic composition and accession numbers of ITS1, ITS2, and *LFY* CDS sequences from 16 Asteraceae species*.

No.	Species	Subfamily	Tribe	Genus	*LEAFY*	ITS1	ITS2
1	*Chrysanthemum indicum*	Asteroideae	Anthemideae	*Chrysanthemum*	PQ134099.1	KC694293.1	MG731087.1
2	*Chrysanthemum chanetii*	Asteroideae	Anthemideae	*Chrysanthemum*	KF151324.1	KX421715.1	KX421714.1
3	*Chrysanthemum zawadzkii*	Asteroideae	Anthemideae	*Chrysanthemum*	KF151355.1	EF577307.1	KX394604.1
4	*Chrysanthemum potentilloides*	Asteroideae	Anthemideae	*Chrysanthemum*	MT648297.1	KC694534.1	KC694532.1
5	*Chrysanthemum dichrum*	Asteroideae	Anthemideae	*Chrysanthemum*	MT648309.1	KX352148.1	KX352148.1
6	*Chrysanthemum lavandulifolium*	Asteroideae	Anthemideae	*Chrysanthemum*	AY672542.1	AF314600.1	MH712594.1
7	*Chrysanthemum oreastrum*	Asteroideae	Anthemideae	*Chrysanthemum*	MT648306.1	EF577304.1	EF577304.1
8	*Chrysanthemum hypargyrum*	Asteroideae	Anthemideae	*Chrysanthemum*	AB808646.1	KC694194.1	KC694194.1
9	*Phaeostigma variifolium*	Asteroideae	Anthemideae	*Phaeostigma*	MT648353.1	EF577283.1	EF577283.1
10	*Argyranthemum frutescens*	Asteroideae	Anthemideae	*Argyranthemum*	MK990597.1	KX446802.1	KX446802.1
11	*Erigeron canadensis*	Asteroideae	Astereae	*Erigeron*	XM_043761487.1	LC387645.1	OK376295.1
12	*Helianthus annuus*	Asteroideae	Heliantheae	*Helianthus*	XM_022128524.1	KX671853.1	MG219856.1
13	*Tagetes erecta*	Asteroideae	Tageteae	*Tagetes*	JF825877.2	KJ525046.1	MK087966.1
14	*Cynara cardunculus*	Carduoideae	Cardueae	*Cynara*	XM_025114808.1	AJ831528.1	JX867643.1
15	*Lactuca saligna*	Cichorioideae	Cichorieae	*Lactuca*	XM_078991094.1	MT002897.1	MT002897.1
16	*Lactuca sativa*	Cichorioideae	Cichorieae	*Lactuca*	XM_023888266.3	AJ633337.1	AJ633337.1

*ITS1, ITS2, and *LFY* CDS sequences were retrieved from NCBI GenBank (May, 2026).

Table 2. Summary statistics of the sequence datasets of 16 Asteraceae species.

Marker	Number of sequences (N)	Length range (bp)	Mean length (bp)
ITS1	16	401–718	492.8
ITS2	16	350–1,019	678.5
LFY CDS	16	1,236–1,311	1,261.3

levels, but also at higher taxonomic levels. Therefore, the dataset was structured to compare marker resolution across a gradient of relatedness: intra-genus, inter-genus within the same tribe, inter-tribe within the same subfamily, and inter-subfamily.

All 16 species had available accession records for the three target markers, resulting in an initial comparative matrix of 48 sequence records: 16 *LFY* accessions (1,236–1,311 bp), 16 ITS1 accessions (401–718 bp), and 16 ITS2 accessions (350–1019 bp) (Table 2). This balanced marker availability is advantageous because it permits direct comparison of ITS1, ITS2, and *LFY* CDS across the same taxon set. ITS2 is particularly relevant as a benchmark because a previous Asteraceae-wide barcode evaluation reported that ITS2 showed strong performance, with 76.4% correct identification at the species level and 97.4% at the genus level using 3,490 Asteraceae ITS2 sequences from GenBank (Gao et al., 2010).

Nevertheless, because the dataset was mined from public NCBI records, sequence curation remained essential before alignment and phylogenetic analysis. GenBank is a comprehensive public database of nucleotide sequences and associated biological annotations, but sequence length, annotation quality, and taxonomic accuracy vary among submitted records (Sayers et al., 2024). Therefore, all accessions should be verified using NCBI Taxonomy and marker-specific validation before analysis; NCBI Taxonomy provides a curated framework linked to sequence records, but NCBI also cautions that it is not the final authority for nomenclature or classification. Accessions used for both ITS1 and ITS2 should be checked to confirm whether they represent a complete ITS region containing ITS1–5.8S–ITS2, after which ITS1 and ITS2 must be extracted and analyzed separately.

The alignment statistics showed that the *LFY* CDS marker had the longest alignment length (1,386 bp), number of variable sites (529 bp), and parsimony-informative sites (353 bp) compared to ITS1 and ITS2 (Table 3). The result indicates that *LFY* CDS showed the highest absolute number of variable and parsimony-informative sites, thus potentially providing better phylogenetic resolution in revealing phylogenetic relationships between species than ITS1 and ITS2. The greater number of informative characters did not show into complete separation between related *Chrysanthemum* species. This indicates that marker performance is not only based on the quantity of variable sites, but also on how mutations are distributed across taxa. Similar patterns have been reported by Zhang et al. (2026), in phylogenomic studies based on low-copy nuclear genes, still show limited resolution in diverged lineages because of incomplete lineage sorting, hybridization, and reticulate evolution. Conversely, ITS1 had the lowest

Table 3. Alignment statistics of 16 Asteraceae species.

Marker	Alignment length (bp)	Conserved sites (bp)	Variable sites (bp)	Parsimony informative sites (bp)
ITS1	754	420	318	174
ITS2	1,038	521	338	186
LFY CDS	1,386	797	529	353

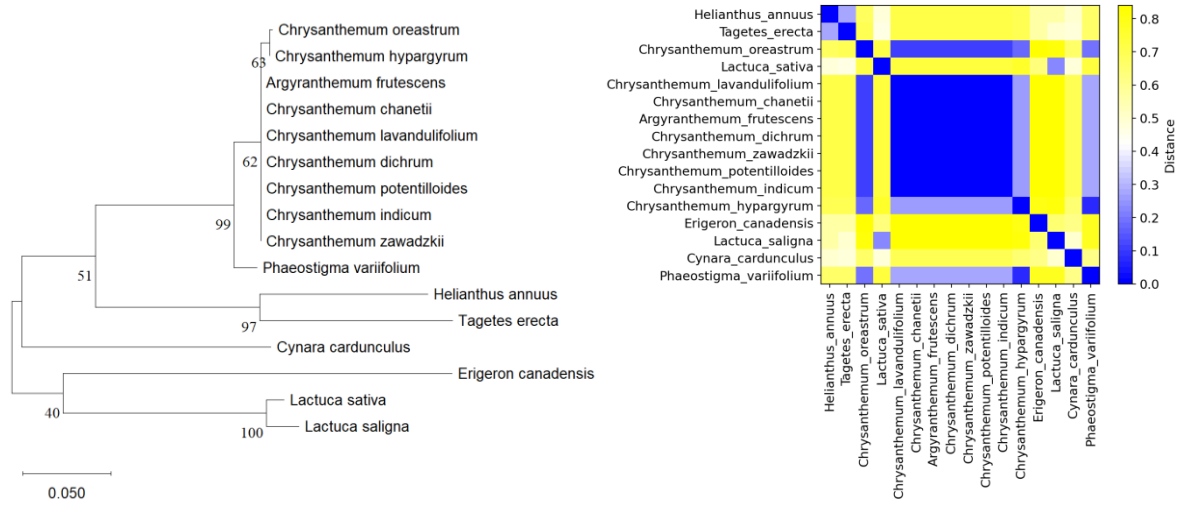

Fig. 1. ITS1-based phylogenetic tree (left) and clustered pairwise genetic distance heatmap (right) of selected Asteraceae species.

number of informative sites (174 bp), so its discriminatory power was relatively lower compared to *LFY* CDS and ITS2.

Phylogenetic Analysis Based on ITS1

The ITS1-based phylogenetic tree showed that most *Chrysanthemum* species formed a compact cluster, indicating close genetic relationships among species (Figure 1). *A. frutescens* and *P. variifolium* were positioned near the *Chrysanthemum* cluster, suggesting that ITS1 was able to recover genetic similarity among closely related Anthemideae taxa. In contrast, *H. annuus* and *T. erecta* formed a separate group from the *Chrysanthemum*-dominated clade, reflecting their placement in different tribes within Asteroideae. The two *Lactuca* species, *L. sativa* and *L. saligna*, were grouped closely together, consistent with their shared generic affiliation, while *C. cardunculus* was separated from but positioned near the *Lactuca* clade.

The ITS1 pairwise-distance heatmap supported the phylogenetic pattern. A large low-distance block was observed among *Chrysanthemum* species, indicating high ITS1 similarity within the genus. Relatively lower pairwise distance patterns were also observed between *Chrysanthemum*, *A. frutescens*, and *P. variifolium*, suggesting close relationships among Anthemideae-related taxa. In contrast, higher pairwise distances were observed between the *Chrysanthemum* cluster and more distant taxa, such as *E. canadensis*, *H. annuus*, *T. erecta*, *C. cardunculus*, and *Lactuca* sp. The result indicates that ITS1 is useful for separating broader taxonomic groups in Asteraceae, but low divergence among closely related species suggests that ITS1 alone may have limited power for resolving species boundaries within closely related genera (Álvarez & Wendel, 2003; Bailey et al., 2003). The limited divergence reflects recent diversification and extensive reticulate evolution within the genus (Liu et al. 2012). Therefore, comparison with ITS2 and *LFY* CDS is necessary to determine whether additional markers can improve species discrimination.

Phylogenetic Analysis Based on ITS2

The ITS2-based phylogenetic tree showed that most *Chrysanthemum* species formed a compact clade together with closely related Anthemideae taxa (Figure 2). This pattern indicates that ITS2 was able to recover genetic similarity among closely related taxa within Asteroideae. However, the short branches among *Chrysanthemum* species suggest low ITS2 divergence within the genus, indicating that ITS2 may have limited resolution for discriminating closely related species. The ITS2 results

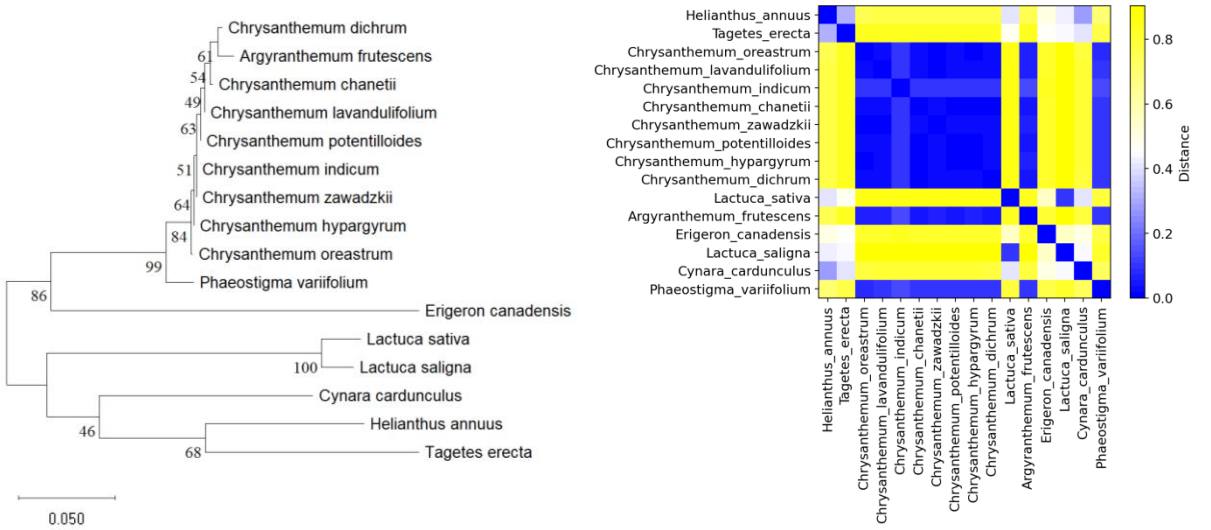

Fig. 2. ITS2-based phylogenetic tree (left) and clustered pairwise genetic distance heatmap (right) of selected Asteraceae species.

were consistent with Gao et al. (2010), showed high identification success at broader taxonomic levels, but reduced discriminatory among closely related taxa. In contrast, *Erigeron canadensis*, *H. annuus* and *T. erecta* were separated from the *Chrysanthemum*-dominated clade, reflecting broader divergence among different Asteroideae tribes. The two *Lactuca* species formed a close pair, while *C. cardunculus* was separated from the Lactuca clade, consistent with broader subfamilial differentiation in Asteraceae.

The ITS2 pairwise-distance heatmap supported the phylogenetic results. A large low-distance block was observed among *Chrysanthemum* species and related Anthemideae taxa, indicating high ITS2 similarity within this group. Higher pairwise distances were observed between this group and more distant taxa, such as *E. canadensis*, *H. annuus*, *T. erecta*, *C. cardunculus*, and *Lactuca* sp. These results suggest that ITS2 is effective in distinguishing major taxonomic groups in the dataset. Nevertheless, the low genetic distances among closely related *Chrysanthemum* species indicate that ITS2 alone may not be sufficient for resolving all species boundaries within closely related genera. These findings suggest that future combined-marker analysis should be tested to determine whether integrating ITS1, ITS2, and *LFY* CDS can improve species discrimination in Asteraceae.

Phylogenetic Analysis Based on LFY

The *LFY* CDS-based phylogenetic tree showed that most *Chrysanthemum* species formed a compact clade, indicating close nuclear sequence similarity within the genus (Figure 3). Several internal subgroups were observed, including the grouping of *C. potentilloides*, *C. dichrum*, and *C. lavandulifolium*, as well as the close placement of *C. indicum*, *C. chanetii*, and *C. zawadzkii*. *P. variifolium* and *A. frutescens* were positioned near the *Chrysanthemum* clade, suggesting that *LFY* CDS retained phylogenetic signal among closely related Anthemideae taxa. In contrast, *E. canadensis*, *H. annuus*, *T. erecta*, *C. cardunculus*, and *Lactuca* spp. were separated from the *Chrysanthemum*-dominated clade, indicating that *LFY* CDS was informative for distinguishing broader taxonomic groups within Asteraceae.

The *LFY* CDS pairwise-distance heatmap supported the phylogenetic topology. A large low-distance block was observed among *Chrysanthemum* species, indicating high sequence similarity within the genus. Lower distances were also observed between *Chrysanthemum* and nearby Anthemideae-related taxa, whereas

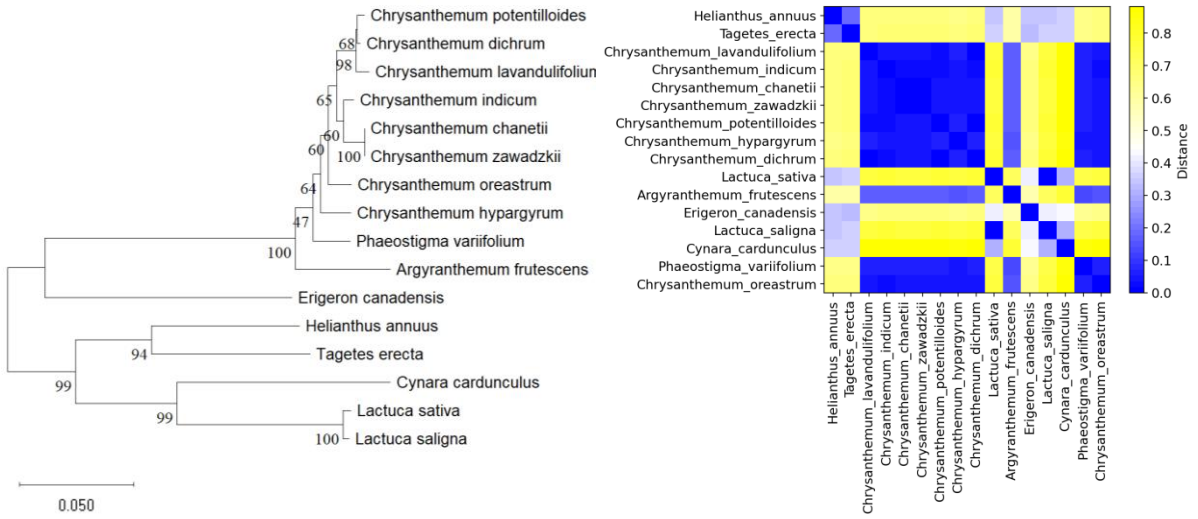

Fig. 3. Comparative *LFY* CDS-based phylogenetic tree (left) and clustered pairwise genetic distance heatmap (right) of selected Asteraceae species.

higher distances were detected between this group and more distant genera, such as *Erigeron*, *Helianthus*, *Tagetes*, *Cynara*, and *Lactuca*. The close relationship between *L. sativa* and *L. saligna* was also recovered. These results suggest that *LFY* CDS can provide useful complementary nuclear information for Asteraceae species identification, especially for separating genera, tribes, and broader taxonomic groups. However, its relatively low divergence among closely related species indicates that *LFY* CDS alone may not provide sufficient resolution for species delimitation within recently diverged lineages, despite the substantial phylogenetic information associated with low-copy nuclear genes (Baum et al., 2005; Gao et al., 2019; Sang, 2002). Comparison with ITS1 and ITS2 remains necessary to determine the most effective marker combination for species discrimination in Asteraceae.

The minimum pairwise distance (0.0000) indicates that at least one pair of taxa had identical sequences for a given marker. The maximum pairwise distance reflects the greatest divergence among the sampled taxa. ITS2 showed the highest mean interspecific distance among the three markers, followed by ITS1 and *LFY* CDS (Table 4). This suggests that ITS2 provided slightly greater overall sequence divergence in the present dataset. However, the differences among markers were relatively small, indicating that marker performance cannot be evaluated solely from average genetic distances. Effective species discrimination also depends on the distribution of genetic variation among taxa and the ability of a marker to discriminate closely related species (China Plant BOL Group, 2011; Gao et al., 2010; Hollingsworth et al., 2011).

The comparison of ITS1, ITS2, and *LFY* CDS showed that the three markers recovered broadly similar taxonomic patterns but differed in their discriminatory strength. ITS1 and ITS2 consistently grouped the *Chrysanthemum*-related taxa together and separated them from more distant genera, such as *Helianthus*, *Tagetes*, *Erigeron*, *Cynara*, and *Lactuca*. This pattern supports previous evidence that nuclear ribosomal ITS regions are useful for plant systematics because they often provide higher variation than many plastid loci, especially at generic and infrageneric levels (Álvarez & Wendel, 2003; Bailey et al., 2003). However, the low pairwise distances among closely related *Chrysanthemum* species indicate that ITS1 and ITS2 may still have limited resolution

Table 4. Pairwise distance among the three markers ITS1, ITS2, and *LFY* CDS.

Marker	Minimum pairwise distance	Maximum pairwise distance	Mean interspecific distance
ITS1	0.0000	0.3013	0.1534
ITS2	0.0000	0.3364	0.1621
LFY CDS	0.0000	0.4153	0.1408

for distinguishing very closely related species.

Among the two ribosomal spacers, ITS2 appears to be the stronger benchmark marker for Asteraceae. Gao et al. (2010) evaluated candidate DNA barcodes in Asteraceae and reported that ITS2 correctly identified 76.4% of samples at the species level and 97.4% at the genus level, although its ability varied among genera. This supports the present result that ITS2 can clearly separate broader taxonomic groups but may not fully resolve all species within closely related genera such as *Chrysanthemum*. Therefore, ITS2 is appropriate as the main comparative barcode marker, whereas ITS1 provides additional ribosomal variation but should be interpreted carefully because ITS regions can be affected by intragenomic variation, pseudogenes, paralogy, and incomplete concerted evolution (Álvarez & Wendel, 2003; Bailey et al., 2003; Gao et al., 2010).

In contrast, *LFY* CDS produced phylogenetic relationships that differed from those inferred from the ribosomal ITS regions. The *LFY*-based tree recovered the main *Chrysanthemum* cluster and separated broader groups, such as *Erigeron*, *Helianthus*, *Tagetes*, *Cynara*, and *Lactuca*. This is consistent with the expected behavior of a protein-coding nuclear gene: *LFY* CDS is more functionally constrained than ITS1 or ITS2, so it may show lower sequence divergence among closely related species. Nevertheless, its value lies in providing an independent nuclear signal. Low-copy nuclear genes are considered a rich source of phylogenetic information and can improve inference when plastid or multicopy nuclear ribosomal markers are insufficient (Sang, 2002).

The biological relevance of *LFY* further strengthens its use as a complementary marker. *LFY* encodes a plant-specific transcription factor involved in floral meristem identity and flowering development. Earlier molecular evolutionary work reported that *LFY* is generally single-copy in many angiosperms, although duplicated paralogs may occur in recent polyploids (Baum et al., 2005). Recent broad-scale analyses of the *LFY* gene family indicate a complex evolutionary history involving ancient duplication events and lineage-specific changes, including grass-specific transposition; therefore, orthology validation is essential before using *LFY* CDS as a species-identification marker (Gao et al., 2019). Therefore, *LFY* CDS is promising as a low-copy nuclear marker, but orthology validation remains essential before using it for species identification.

Although the markers overall could differentiate major taxonomic groups, none of the three markers fully resolved relationships among closely related *Chrysanthemum* species. Several factors may contribute to this limited resolution, including recent diversification, hybridization, introgression, polyploidization, and incomplete lineage sorting. These processes can reduce phylogenetic signal and generate conflicting genetic patterns among closely related taxa. Hybridization and polyploidy have played important roles in the evolutionary history of *Chrysanthemum*, contributing to reticulate evolutionary patterns and limited phylogenetic resolution (Hao et al., 2022; Liu et al., 2012). Consequently, additional low-copy nuclear genes, genomic-scale datasets, or multilocus approaches may be required to

achieve stronger resolution within *Chrysanthemum*.

This study provides useful insights into the comparative performance of ITS1, ITS2, and *LFY* CDS markers in selected Asteraceae species. However, the results should be interpreted considering the limited taxon sampling and the reliance on publicly available sequence data. In addition, the markers were evaluated separately, and no combined-marker analysis was performed. Therefore, the potential benefits of integrating ITS1, ITS2, and *LFY* CDS remain to be tested in future studies. Additional analyses involving broader species representation, multiple accessions per species, and concatenated datasets would further strengthen the assessment of marker performance and taxonomic resolution within Asteraceae.

Conclusion

ITS1, ITS2, and *LFY* CDS are useful molecular markers for evaluating genetic relationships and species discrimination in Asteraceae. All three markers recovered the major taxonomic structure of the family, including the clustering of *Chrysanthemum* species and the separation of more distantly related genera, such as *Helianthus*, *Tagetes*, *Erigeron*, *Cynara*, and *Lactuca*. Among the ribosomal markers, ITS2 provided slightly greater phylogenetic information than ITS1, as indicated by its higher number of variable sites and parsimony-informative sites, as well as its higher mean interspecific distance. Nevertheless, ITS2 still showed limited resolution among closely related *Chrysanthemum* species.

LFY CDS showed the highest absolute number of variable sites and parsimony-informative sites and provided an independent nuclear coding signal that complemented the ribosomal ITS regions. However, LFY CDS did not fully resolve relationships among closely related *Chrysanthemum* species, indicating that the number of informative characters alone is insufficient for complete species discrimination in recently diverged or reticulate lineages. Therefore, *LFY* CDS should not be considered a replacement for ITS1 or ITS2, but rather a complementary marker. Future studies should include broader taxon sampling, multiple accessions per species, orthology validation, and combined multilocus or genomic-scale analyses to improve phylogenetic resolution within Asteraceae.

Acknowledgements

This research received no specific funding from public, commercial, or not-for-profit agencies.

Conflict of Interest

The authors declare that they have no conflict of interest.

Data Availability

The sequence data used in this study are publicly available in the NCBI GenBank database. The accession numbers of ITS1, ITS2, and *LFY* CDS sequences analyzed in this study are provided in Table 1. The aligned sequence datasets, phylogenetic trees, and pairwise genetic distance matrices generated during the current study are available from the corresponding author upon reasonable request.

References

Álvarez, I., & Wendel, J. F. (2003). Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution*, *29*(3), 417–434. https://doi.org/10.1016/S1055-7903(03)00208-2

Awaliah, S. R., & Polosoro, A. (2024). A comparative study of DNA barcoding markers for bamboo. *Buitenzorg: Journal of Tropical Science*, *1*(2), 30–41. https://doi.org/10.70158/buitenzorg.v1i2.9

Bailey, C. D., Carr, T. G., Harris, S. A., & Hughes, C. E. (2003). Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. *Molecular Phylogenetics and Evolution*, *29*(3),

435–455. https://doi.org/10.1016/j.ympev.2003.08.021

Baum, D. A., Yoon, H. S., & Oldham, R. L. (2005). Molecular evolution of the transcription factor *LEAFY* in Brassicaceae. *Molecular Phylogenetics and Evolution*, *37*(1), 1–14. https://doi.org/10.1016/j.ympev.2005.07.002

CBOL Plant Working Group (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(31), 12794–12797. https://doi.org/10.1073/pnas.0905845106

Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., et al. (2010). Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS ONE*, *5*(1), e8613. https://doi.org/10.1371/journal.pone.0008613

China Plant BOL Group (2011). Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(49), 19641–19646. https://doi.org/10.1073/pnas.1104551108

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Elomaa, P., Zhao, Y., & Zhang, T. (2018). Flower heads in Asteraceae—recruitment of conserved developmental regulators to control the flower-like inflorescence architecture. *Horticulture Research*, *5*, 36. https://doi.org/10.1038/s41438-018-0056-8

Gao, T., Yao, H., Song, J., Zhu, Y., Liu, C., & Chen, S. (2010). Evaluating the feasibility of using candidate DNA barcodes in discriminating species of the large Asteraceae family. *BMC Evolutionary Biology*, *10*, 324. https://doi.org/10.1186/1471-2148-10-324

Gao, B., Chen, M. X., Li, X. S., & Zhang, J. H. (2019). Ancient duplications and grass-specific transposition influenced the evolution of *LEAFY* transcription factor genes. *Communications Biology*, *2*, 237. https://doi.org/10.1038/s42003-019-0469-4

Grob, G. B. J., Gravendeel, B., & Eurlings, M. C. M. (2004). Potential phylogenetic utility of the nuclear *FLORICAULA/LEAFY* second intron: Comparison with three chloroplast DNA regions in *Amorphophallus* (Araceae). *Molecular Phylogenetics and Evolution*, *30*(1), 13–23. https://doi.org/10.1016/S1055-7903(03)00183-0

Hao, D., Song, Y., Xiao, P., Zhong, Y., Wu, P., & Xu, L. (2022). The genus *Chrysanthemum*: phylogeny, biodiversity, phytometabolites, and chemodiversity. *Frontiers in Plant Science*, *11*(13), 973197. https://doi.org/10.3389/fpls.2022.973197

Hollingsworth, P. M., Graham, S. W., & Little, D. P. (2011). Choosing and using a plant DNA barcode. *PLoS ONE*, *6*(5), e19254. https://doi.org/10.1371/journal.pone.0019254

Kumar, S., Stecher, G., Suleski, M., Sanderford, M., Sharma, S., & Tamura, K. (2024). MEGA12: molecular evolutionary genetic analysis version 12 for adaptive and green computing. *Molecular Biology and Evolution*, *41*(12), msae263. https://doi.org/10.1093/molbev/msae263

Liu, P., Wan, Q., Guo, Y., Yang, J., & Rao, G. (2012). Phylogeny of the genus *Chrysanthemum* L.: evidence from

single-copy nuclear gene and chloroplast DNA sequences. *PLoS ONE*, *7*(11), e48970. https://doi.org/10.1371/journal.pone.0048970

Panero, J. L., & Funk, V. A. (2002). Toward a phylogenetic subfamilial classification for the Compositae (Asteraceae). *Proceedings of the Biological Society of Washington*, *115*(4), 909–922.

Sang, T. (2002). Utility of low-copy nuclear gene sequences in plant phylogenetics. *Critical Reviews in Biochemistry and Molecular Biology*, *37*(3), 121–147. https://doi.org/10.1080/10409230290771474

Sayers, E. W., Cavanaugh, M., Clark, K., Pruitt, K. D., Sherry, S. T., Yankie, L., & Karsch-Mizrachi, I. (2024). GenBank 2024 update. *Nucleic Acids Research*, *52*(D1), D134–D137. https://doi.org/10.1093/nar/gkad903

Schoch, C. L., Ciufo, S., Domrachev, M., Hotton, C. L., Kannan, S., Khovanskaya, R., et al. (2020). NCBI taxonomy: a comprehensive update on curation, resources and tools. *Database*, *2020*, baaa062. https://doi.org/10.1093/database/baaa062

Zhang, R., Stull, G. W., Jin, J., Wang, Y., Guo, Y., Yang, Z., et al. (2026). Phylogenetic resolution and conflict in the species-rich flowering plant family Leguminosae. *Systematic Biology*, *75*(2), 236–258. https://doi.org/10.1093/sysbio/syaf057

Zhao, Y., Zhang, T., Broholm, S. K., Tähtiharju, S., Mouhu, K., Albert, V. A., Teeri, T. H., & Elomaa, P. (2016). Evolutionary co-option of floral meristem identity genes for patterning of the flower-like Asteraceae inflorescence. *Plant Physiology*, *172*(1), 284–296. https://doi.org/10.1104/pp.16.00779