BMC Genomics

RESEARCH

Open Access

Polyploidy-driven expansion and regulatory diversification of the Kelch repeat F-box gene family in sweetpotato

Yanxingzi Ma[1], Guangli Zhou[2], Jinchai Chen[2], Fanqing Meng[3], Runze Wei[4], Shijian Zhou[2], Jiayuan Shi[1], Jiahui Chen[1], Liheng He[2,5], Ling Yuan[6] and Xiaoyun Jia[1,5*]

Abstract

Background Kelch repeat F-box (KFB) proteins are substrate-recognition components of SCF (SKP1-CULLIN1-F-box) E3 ubiquitin ligases and play important roles in protein turnover, specialized metabolism, and stress responses in plants. However, the evolutionary history of the KFB gene family across different ploidy levels remains unclear in sweetpotato (*Ipomoea batatas*), an important polyploid crop. Here, we tested the hypothesis that KFB genes in sweetpotato and its relatives underwent expansion and transcriptional diversification during polyploid evolution while retaining the conserved structural framework required for SCF-mediated proteolysis.

Results We performed a cross-ploidy comparative analysis of the KFB gene family in four *Ipomoea* species representing three ploidy levels: diploid *I. trifida* and *I. triloba*, tetraploid *I. tabascana*, and hexaploid cultivated sweetpotato (*I. batatas*). We identified 24, 27, 16, and 102 KFB genes in these species, respectively. The marked expansion in hexaploid sweetpotato, together with the reduced KFB repertoire in tetraploid *I. tabascana*, indicates that KFB family size does not scale linearly with ploidy and instead reflects lineage-specific retention and loss in gene family evolution. Segmental duplication was the predominant mode associated with KFB expansion, whereas tandem duplication was not detected. Most duplicated gene pairs had Ka/Ks ratios below 1, indicating strong purifying selection on retained copies. Despite substantial variation in copy number, KFB proteins retained conserved F-box/Kelch domain organization, with moderate variation in motif composition and gene structure. Transcriptome profiling, co-expression analysis, and qRT-PCR further revealed substantial tissue-, stress-, and hormone-responsive expression divergence among KFB genes.

Conclusions These findings establish an evolutionary framework for KFB family diversification in *Ipomoea* and show that expansion of the KFB family in sweetpotato is shaped not only by genome multiplication, but also by lineage-specific gene retention, loss, and regulatory divergence. The identified KFB candidates provide a useful foundation for future functional studies of stress adaptation and metabolic regulation in sweetpotato.

Keywords Sweetpotato, KFB gene family, Polyploidy, WGCNA, GO and KEGG enrichment

*Correspondence:
Xiaoyun Jia
jiaxiaoyun@sxau.edu.cn

Full list of author information is available at the end of the article

Background

Sweetpotato (*Ipomoea batatas* L.) is a globally important food, feed, and industrial crop valued for its exceptional adaptability, rich nutritional profile, and resilience in diverse agroecosystems. Originating from tropical America [1], sweetpotato contributes approximately 89 million tons of global production annually, providing a critical source of calories, vitamins, and bioactive compounds essential for addressing global food and nutrition security challenges [2]. Its closely related wild diploid species, *I. trifida* and *I. triloba*, exhibit strong disease resistance, stress tolerance, and rich genetic diversity, making them important reservoirs for sweetpotato improvement [3]. The tetraploid *I. tabascana*, proposed to have arisen from a natural hybridization between *I. trifida* and *I. batatas* [4], together with cultivated hexaploid cultivars such as 'Beauregard', offers a comparative system for investigating polyploid evolution, gene family expansion, and functional diversification within the genus *Ipomoea* [5]. The *Ipomoea* lineage provides an opportunity to examine how ploidy variation may influence the retention, expansion, and diversification of regulatory gene families relevant to stress adaptation and crop improvement. The ubiquitin–26 S proteasome system (UPS) is a central regulatory machinery in plants that modulates protein abundance and turnover, thereby influencing virtually all aspects of plant growth, metabolism, and environmental adaptation [6]. Within the UPS, the Skp1-Cullin-F-box (SCF) complex is a major class of E3 ubiquitin ligases, in which the F-box protein functions as the substrate-recognition component [7–9]. F-box proteins typically contain a conserved N-terminal F-box domain that mediates interaction with Skp1, along with variable C-terminal protein–protein interaction domains, such as Kelch repeats, WD40 motifs, and leucine-rich repeats, that confer substrate specificity [7, 10]. Kelch repeat F-box (KFB) proteins form a distinct subfamily characterized by multiple Kelch motifs that assemble into β-propeller structures capable of recognizing diverse target proteins. Growing evidence indicates that KFB proteins play pivotal roles in developmental regulation, specialized metabolism, and responses to biotic and abiotic stresses in multiple plant species. A particularly well-studied example involves the regulation of the phenylpropanoid pathway, in which KFB proteins in *Arabidopsis*, rice, and other species mediate the ubiquitin-dependent degradation of phenylalanine ammonia-lyase (PAL), a rate-limiting enzyme for lignin, flavonoid, and other phenylpropanoid-derived metabolites [11–13]. In sweetpotato, our previous studies have revealed that *IbKFB* is a direct target of the microRNA Ib-miR2111 and negatively regulates anthocyanin biosynthesis by promoting degradation of IbPAL1 and IbGAPCp1, highlighting the KFB family as a key post-translational regulatory node for specialized metabolism [14].

In the post-genomic era, genome-wide characterization of gene families has become an effective strategy for uncovering functional diversification, evolutionary dynamics, and regulatory complexity across species. Although the F-box superfamily has been cataloged in sweetpotato [15], the KFB subfamily has not been systematically examined in a comparative framework spanning different ploidy levels within *Ipomoea*. This represents an important gap, because it is unclear whether expansion and diversification of KFB genes are associated with polyploidization and subsequent evolutionary retention. Previous studies in other species support the biological importance of this family; for example, 31 KFB genes were identified in *Salvia miltiorrhiza*, including *SmKFB1/2/5* as potential regulators of phenolic acid biosynthesis [7], 19 in Moso bamboo with *PeKFB9* implicated in lignification and stress responses [16], and 44 in potato with some StKFB members associated with anthocyanin regulation [17]. Together, these reports indicate that KFB genes are both functionally important and evolutionarily flexible, but they do not clarify how this family has diversified across diploid and polyploid *Ipomoea* species.

Given the established roles of KFB proteins in metabolism and stress-related regulation, and the availability of genomic resources from diploid, tetraploid, and hexaploid *Ipomoea* species, a comparative analysis of the KFB family across ploidy levels is both timely and necessary. At the same time, the evolutionary relationship between ploidy level and KFB family size is unlikely to be explained by ploidy alone, because lineage-specific duplication, retention, and loss may also shape the final gene repertoire. Most recently, the post-translational regulatory mechanism of IbKFB in sweetpotato has been further clarified, with its role in mediating ubiquitination of key enzymes involved in anthocyanin biosynthesis demonstrated [14]. These advances provide a mechanistic foundation for asking whether KFB family members in *Ipomoea* have undergone not only numerical expansion, but also structural and transcriptional diversification that may be associated with stress responses and metabolic regulation.

In this study, we performed a comprehensive, multilevel investigation of the KFB gene family across four *Ipomoea* genomes representing three ploidy levels: diploids (*I. trifida*, *I. triloba*), tetraploid (*I. tabascana*), and hexaploid cultivated sweetpotato (*I. batatas* 'Beauregard'). We combined genome-wide identification, chromosomal mapping, phylogenetic classification, motif and gene structure analysis, synteny and evolutionary selection assessment, tissue- and stress-responsive transcriptome profiling, and weighted gene co-expression network

analysis (WGCNA) coupled with GO enrichment. We further integrated hormone-response expression patterns and qRT-PCR validation to identify candidate KFB genes associated with hormonal signaling and phenylpropanoid metabolism.

Collectively, this study provides a systematic cross-ploidy characterization of the KFB family in *Ipomoea* and establishes a comparative framework for evaluating how gene family expansion, conservation, and expression divergence are related to polyploid evolution. Rather than inferring direct regulatory functions from co-expression alone, our results identify candidate KFB genes and conserved or lineage-specific patterns that may underlie stress adaptation and metabolic regulation. Harnessing the untapped biodiversity of *Ipomoea* germplasm through genomics approaches is critical for unlocking novel genetic variation, which can further enhance the resilience and adaptive potential of sweetpotato via targeted molecular breeding [18]. Therefore, this study provides a comparative framework and candidate gene set for future functional analyses of KFB genes in sweetpotato and its relatives.

Materials and methods

Identification of KFB family members

Protein sequences, genomic sequences, coding sequences (CDS), and genome annotation files for diploid *I. trifida* V3, *I. triloba* V3, and hexaploid *I. batatas* 'Beauregard' V1 were obtained from the Sweetpotato Genomics Resource database (http://sweetpotato.uga.edu/). Genomic resources for *I. tabascana* used in this study were obtained from the Sweetpotato Research Institute, Chinese Academy of Agricultural Sciences, Xuzhou, China. A conference abstract describing the whole-genome assembly of *I. tabascana* has been reported previously [19]. Reference KFB protein sequences from *Arabidopsis thaliana* (TAIR10) were downloaded from TAIR (https://www.arabidopsis.org/Blast/index.jsp) and used as queries for homology-based identification.

Candidate KFB proteins were identified using a combined homology- and domain-based strategy. First, BLASTP searches were performed against the proteomes of the four *Ipomoea* species using known Arabidopsis KFB proteins as queries, with an E-value threshold of < 1e-5. Second, Hidden Markov Model (HMM) profiles corresponding to the F-box domain (PF00646) and Kelch domain (PF01344) were retrieved from the Pfam (http://pfam-legacy.xfam.org/) database and were subsequently used as queries to search against the protein sequences of the *Ipomoea* proteome using HMMER 3.0. All candidates recovered by either approach were pooled and further subjected to domain confirmation using the NCBI Conserved Domain Database (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi).

Only proteins containing both an F-box domain and at least one Kelch motif were retained as bona fide KFB family members. Candidates with partial, fragmented, or ambiguous domain matches were excluded. To avoid redundant counting caused by highly similar models representing the same locus or haplotype-resolved assemblies, redundant sequences were collapsed using CD-HIT (https://github.com/weizhongli/cdhit/releases) on a Linux platform with a sequence identity threshold of 0.9, and representative non-redundant sequences were retained for downstream analyses. The final KFB datasets for each species were used for chromosomal localization, phylogenetic analysis, motif and gene structure analysis, duplication analysis, and expression profiling.

Prediction of physicochemical properties and subcellular localization

Physicochemical properties of KFB proteins, including amino acid length, molecular weight, theoretical isoelectric point (pI), and Grand Average of Hydropathy (GRAVY), were predicted using ExPASy ProtParam (https://web.expasy.org/protparam/). Putative subcellular localization was predicted using WOLF PSORT (https://wolfpsort.hgc.jp/).

Chromosomal localization and gene duplication analysis

Chromosomal coordinates of KFB genes were extracted from the corresponding genome annotation files for each species and visualized using the Gene Location Visualize (Advanced) function in TBtools-II v2.1. Physical mapping was used to examine the genomic distribution pattern of KFB genes across chromosomes or subgenomes.

Gene duplication analysis was performed using MCScanX with default parameters. Both tandem duplication and segmental/collinear duplication were examined based on the MCScanX [20] output. Tandem duplication events were defined as two or more homologous genes located within a 200 kb chromosomal region, consistent with previous studies [21]. Segmental duplication events were defined according to collinear blocks identified either between chromosomes or among subgenomes. The duplication relationships of KFB genes were visualized using the Advanced Circos function in TBtools-II v2.1.

For each duplicated gene pair, non-synonymous substitution rate (Ka), synonymous substitution rate (Ks), and their ratio (Ka/Ks) were calculated using the Simple Ka/Ks Calculator function in TBtools-II v2.1 to evaluate evolutionary selection pressure. Ka/Ks > 1, Ka/Ks < 1, and Ka/Ks = 1 were interpreted as indicative of positive, purifying, and neutral selection, respectively [22].

Analysis of conserved motifs and gene structures

Multiple sequence alignments of KFB proteins were generated using MUSCLE [23] with default parameters. Based on the aligned protein sequences, a Maximum Likelihood (ML) phylogenetic tree was constructed using IQ-TREE v2.2.2.6 with the best-fit substitution model selected by ModelFinder and 1,000 bootstrap replicates to define KFB subgroups. Conserved motifs were identified using the MEME [24] Suite (https://meme-suite.org/meme/tools/meme) with the maximum number of motifs set to 20, while other parameters were kept at default settings. Motif annotations were further examined using InterProScan [25] (https://www.ebi.ac.uk/interpro/). Exon–intron structures of KFB genes were extracted from genome annotation files. Motif distribution and gene structure were integrated and visualized using the Gene Structure View (Advanced) function in TBtools-II v2.1.

Cross-species evolutionary analysis and classification

Protein sequence files for diploid and hexaploid sweetpotato, tetraploid *I. tabascana*, *Arabidopsis thaliana*, *Oryza sativa*, *Gossypium hirsutum*, and *Ipomoea aquatica* were obtained from the Sweetpotato Genomics Resource database (http://sweetpotato.uga.edu/), the TAIR database (https://www.arabidopsis.org/download/list? dir=Proteins ? dir=Proteins), the Rice Genome Annotation Project database (http://rice.plantbiology.msu.edu/), the Cotton Research Institute database (https://mascotton.njau.edu.cn/info/1054/1118.htm), and the National Genomics Data Center (https://ngdc.cncb.ac.cn/gwh/), respectively. KFB family members in *A. thaliana*, *O. sativa*, *G. hirsutum*, and *I. aquatica* were identified using the same workflow described above for sweetpotato KFB identification.

The identified KFB protein sequences from the four *Ipomoea* species and the four reference species were aligned using MUSCLE [23] with default parameters. An ML phylogenetic tree was then constructed using IQ-TREE v2.2.2.6 with 1,000 bootstrap replicates. The resulting phylogenetic tree was visualized using TVBOT (https://chiplot.online/tvbot.html) [26], and KFB proteins were classified into evolutionary subfamilies based on phylogenetic topology and previously reported functional annotations of representative KFB members in *Arabidopsis thaliana* and *Oryza sativa* [27–29].

Synteny analysis

Collinearity relationships of KFB genes among *Ipomoea* genomes with different ploidy levels were analyzed using MCScanX [20] (https://github.com/wyp1125/MCScanx). The following pairwise genome comparisons were performed: diploid *I. trifida* versus tetraploid *I. tabascana*; diploid *I. triloba* versus tetraploid *I. tabascana*; and tetraploid *I. tabascana* versus hexaploid *I. batatas* 'Beauregard'. Homologous KFB gene pairs and collinear blocks were visualized using the Multiple Synteny Plot function in TBtools-II v2.1. These analyses were used to assess the conservation of ancestral KFB loci and the extent of lineage-specific retention or expansion across ploidy levels.

Tissue-specific and stress-induced expression analysis

Public transcriptome datasets for *I. trifida* and *I. triloba* were obtained from the GT4SP Data Download page of the Sweetpotato Genomics Resource. Specifically, the v3 high-confidence gene model tissue-expression matrices and stress-expression matrices were used for both species. For *I. trifida*, the tissue-expression matrix (I_trifida_NSP306_FPKM_expression_matrix_v3_anno.xlsx) included eight RNA-seq libraries (callus_flower, callus_stem, flower, flowerbud, leaf, root1, root2, and stem), and the stress-expression matrix (I_trifida_NSP306_stress_FPKM_expression_matrix_v3_anno.xlsx) included 15 abiotic and biotic stress RNA-seq libraries. For *I. triloba*, the tissue-expression matrix (I_triloba_NSP323_FPKM_expression_matrix_v3_anno.xlsx) included six RNA-seq libraries (flower, flowerbud, leaf, root1, root2, and stem), and the stress-expression matrix (I_triloba_NSP323_stress_FPKM_expression_matrix_v3_anno.xlsx) included 15 abiotic and biotic stress RNA-seq libraries. For cultivated sweetpotato (*I. batatas* Xu Shu 18), transcriptome data for stem, leaf, and fibrous root tissues under control conditions, three abiotic stress treatments (cold, salt, and drought), and three hormone treatments (ABA, SA, and MeJA), each with two biological replicates, were downloaded from the NCBI SRA database (accession number: PRJNA511028). For the *I. batatas* RNA-seq data, raw reads were filtered using Trimmomatic [30] to remove adapters and low-quality sequences. Clean reads were aligned to the *I. batatas* 'Beauregard' V1 genome using HISAT2 [31], and mapped reads were quantified using FeatureCounts [32] to obtain FPKM values. Differential expression analysis was conducted using the DESeq2 [33] R package, with biological replicates modeled within the DESeq2 framework. Genes with $|\log2(\text{fold change})| \geq 1$ and adjusted p-value < 0.05 were defined as significantly differentially expressed genes (DEGs). Raw counts generated by FeatureCounts [32] were used for DESeq2, and FPKM values were used for expression visualization and WGCNA-based co-expression analysis. Expression heatmaps were generated using the MicroBioinformatics platform (https://bioinformatics.com.cn) based on hierarchical clustering with Euclidean distance. Given the differences in experimental design and data sources among the available datasets, transcript abundance and differential expression results were interpreted primarily as comparative expression profiles and candidate associations rather than direct evidence of regulatory function.

Weighted gene co-expression network analysis

Weighted gene co-expression network analysis (WGCNA) was implemented using the WGCNA package in R v4.3.1 [34] to identify gene co-expression modules in *I. trifida*, *I. triloba*, and *I. batatas*. The analysis was based on the same transcriptome datasets described in Sect. 2.7, and species-specific FPKM expression matrices were used as input for network construction. The FPKM expression matrix was imported into R and transposed so that rows represented samples and columns represented genes. Data quality was assessed using the goodSamplesGenes function, and genes or samples failing the quality check were removed before network construction. To reduce background noise, genes with mean FPKM values ≤ 0.2 across all samples were excluded.

Sample clustering was conducted using hierarchical clustering with average linkage to assess potential outliers. No obvious outlier samples were removed after sample clustering. A series of candidate soft-thresholding powers (1–30) was evaluated using the pickSoftThreshold function, and the optimal soft-thresholding power was selected for each species based on the scale-free topology criterion. The final soft-thresholding powers used for network construction were $\beta = 19$ for *I. batatas*, $\beta = 8$ for *I. triloba*, and $\beta = 7$ for *I. trifida* (Fig. S1). Although the scale-free topology fit for *I. trifida* did not reach 0.8 at $\beta = 7$, this power was selected as a compromise between an acceptable topology fit and sufficient mean connectivity, ensuring the biological relevance of the co-expression network. Adjacency matrices were converted into topological overlap matrices (TOM), and genes were clustered using average-linkage hierarchical clustering based on TOM dissimilarity.

Co-expression modules were identified using the dynamic tree cut algorithm with minModuleSize = 30 and deepSplit = 2. Closely related modules were merged using mergeCloseModules with a cut height of 0.25. Modules containing KFB genes were retained for downstream analyses. Because WGCNA captures correlated expression patterns rather than direct regulatory relationships, the resulting modules were interpreted as candidate co-expression associations for gene prioritization rather than evidence of causal control (Fig. S2).

Gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) enrichment analysis

GO and KEGG enrichment analyses were performed for KFB-associated co-expression modules using the clusterProfiler 4.0 package [35, 36] in R. Species-specific eggNOG-mapper annotation output files were used to generate functional annotation backgrounds for *I. trifida*, *I. triloba*, and *I. batatas*. The gene lists subjected to enrichment analysis were derived from significant KFB-associated WGCNA modules. For GO analysis, the background gene set consisted of all genes with valid GO annotations in the corresponding species-specific eggNOG annotation file and custom OrgDb. For KEGG analysis, the background gene set consisted of all genes included in the species-specific gene-to-pathway mapping table generated from the KEGG_Pathway field of the eggNOG annotation file. For GO analysis, GO annotations were extracted from the GOs field of the eggNOG annotation files, and only valid GO terms were retained after filtering and deduplication. Species-specific custom annotation databases were then constructed using AnnotationForge and GO.db, and GO enrichment was performed with the enrichGO function. For KEGG analysis, gene-to-pathway mapping tables were generated from the KEGG_Pathway field of the eggNOG annotation files. Multiple pathway assignments were separated, and KEGG pathway identifiers were standardized by converting ko entries to map format. Pathway names were retrieved using KEGGREST, and KEGG enrichment was performed using the enricher function. Terms with adjusted p-values < 0.05 after Benjamini–Hochberg correction were considered significantly enriched. Enrichment results were visualized as bar plots and dot plots using clusterProfiler [35, 36] and ggplot2. GO and KEGG analyses were used to infer the predominant biological processes and pathways associated with KFB-enriched co-expression modules, but not to assign direct biochemical functions to individual KFB genes.

qRT-PCR validation of gene expression

Uniform plantlets of *Ipomoea triloba*, *I. trifida*, and *I. batatas* cv. Beauregard with 4-5 nodes and similar growth vigor were selected from field-grown stocks and acclimated indoors for 14 d under controlled environmental conditions. Subsequently, plantlets of each species were sprayed separately with 289 μmol/L gibberellic acid (GA), 200 μmol/L indole-3-acetic acid (IAA), or 50 μmol/L methyl jasmonate (MeJA), while distilled water was used as the mock treatment. Each treatment consisted of three biological replicates, and each replicate contained six independent plantlets. Leaf samples were harvested at 0, 12, 24, and 48 h after treatment for downstream gene expression analysis.

Total RNA was extracted using the FastPure Plant Total RNA Isolation Kit (Polysaccharides & Polyphenolics-rich) (Vazyme, China) according to the manufacturer's instructions. First-strand cDNA was synthesized using the Thermo Scientific Maxima First Strand cDNA Synthesis Kit with dsDNase (Thermo Fisher Scientific, USA). A 169-bp fragment of sweetpotato *actin2* (GenBank AY905538) was used as the internal reference gene. Gene-specific primers were designed in untranslated regions (UTRs) whenever possible to distinguish highly

similar homologs sharing more than 90% sequence identity (Table S1).

RT-qPCR was performed on a QuantStudio™ 7 Flex Real-Time PCR System (Applied Biosystems, USA) using Applied Biosystems™ Power SYBR™ Green PCR Master Mix. ROX was used as the passive reference dye. Each reaction was performed in a 20 μL volume containing 10 μL 2 × SYBR Green Master Mix, 0.8 μL forward primer, 0.8 μL reverse primer, 2 μL cDNA template, and nuclease-free water to the final volume. The amplification program consisted of 95 °C for 30 s, followed by 40 cycles of 95 °C for 10 s, 60 °C for 30 s, and 72 °C for 30 s. Melt curve analysis (60–95 °C, 0.5 °C increments) was performed after amplification to confirm product specificity.

For each biological replicate, qRT-PCR was performed with three technical replicates. Relative gene expression levels were calculated using the 2^-ΔΔCt method [37]. Statistical significance between treatment and control groups at each time point was evaluated using one-way ANOVA followed by Duncan's multiple range test, and $p < 0.05$ was considered statistically significant. qRT-PCR results were used to validate transcriptome-based expression trends of selected candidate genes.

Results

The KFB gene family is markedly expanded in cultivated sweetpotato, but its size does not scale linearly with ploidy

A genome-wide search identified 24 KFB family genes in *I. trifida* (*ItfKFB01–ItfKFB24*), 27 in *I. triloba* (*ItbKFB01–ItbKFB27*), 16 in *I. tabascana* (*ItaKFB01–ItaKFB16*), and 102 in cultivated sweetpotato *I. batatas* (*IbKFB01–IbKFB102*). All retained candidate proteins contained both F-box (PF00646) and Kelch (PF01344) domains, and proteins lacking either domain were excluded from subsequent analyses. Notably, the tetraploid *I. tabascana* harbored fewer KFB genes than the two diploid relatives, indicating that KFB family size in *Ipomoea* does not increase in a simple linear manner with ploidy level. Because the *I. tabascana* assembly was represented in a haplotype-resolved format, 16 non-redundant family members were retained for this species. KFB genes were unevenly distributed across chromosomes within each species (Fig. 1). In *I. trifida*, KFB genes were distributed across 12 chromosomes, with Chr01, Chr10, and Chr11 containing the highest numbers. In *I. triloba*, KFB genes were also mapped to all 12 chromosomes, with Chr04 and Chr11 each containing four genes. In *I. tabascana*, the 16 ItaKFB genes were located on 11 chromosomes, with Chr04, Chr08, Chr09, Chr10, and Chr11 each containing two genes. In hexaploid *I. batatas*, 102 IbKFB genes were distributed across 59 chromosomes, generally with 1–2 genes per chromosome, although several chromosomes harbored higher local densities. Overall, these results indicate substantial expansion of the KFB family in cultivated sweetpotato together with broad but uneven chromosomal distribution across *Ipomoea* genomes.

KFB proteins retain highly similar physicochemical features across species, and several homologs show nearly identical predicted properties

Across the four *Ipomoea* species, 169 KFB proteins were identified, including 102 from *I. batatas*, 24 from *I. trifida*, 27 from *I. triloba*, and 16 from *I. tabascana*. These proteins ranged from 230 to 687 amino acids in length, with MW of 25.8–74.6 kDa and pI of 5.04–9.62 (Table S2).

Despite major differences in family size, the overall physicochemical characteristics of KFB proteins were broadly conserved across ploidy levels. The average amino acid length of KFB proteins in *I. triloba*, *I. trifida*, *I. tabascana*, and *I. batatas* was 427, 452, 471, and 449, respectively. The corresponding average MW were 47.67, 50.53, 52.50 and 50.18 kDa, and the mean pI values were 7.29, 7.32, 7.20 and 7.38, respectively. Most KFB proteins had negative GRAVY values, with species-specific means of -0.15, -0.18, -0.17, and – 0.20, indicating their predominant hydrophilicity. The proportions of hydrophilic KFB proteins reached 85.19% in *I. triloba*, 95.83% in *I. trifida*, 93.75% in *I. tabascana*, and 91.18% in *I. batatas*.

Predicted subcellular localization also showed broad conservation. Most KFB proteins were predicted to localize to the cytoplasm (42.6%, 72 proteins) or nucleus (36.1%, 61 proteins), whereas a small subset was predicted to localize to chloroplasts (20.7%, 35 proteins).

Notably, Euclidean distance analysis (similarity > 0.9) identified 181 pairs of cross-species KFB proteins with highly similar physicochemical properties, with average differences of only 0.87%, 0.92%, 1.23% and 1.56% in amino acid number, MW, pI and GRAVY, respectively. Among these similar pairs, 98.9% shared the same predicted subcellular localization. Several homologs displayed essentially identical predicted characteristics, including nuclear-localized *ItfKFB20* and *IbKFB81*, cytoplasm-localized *ItfKFB21* and *ItaKFB15*, and chloroplast-localized *ItfKFB22*, *IbKFB83*, and *ItbKFB25*. These results indicate that core physicochemical properties are highly conserved across *Ipomoea* KFB proteins, even where family size differs dramatically among species.

Segmental duplication is the dominant duplication mode, whereas most duplicated KFB genes remain under purifying selection

Duplication analysis revealed clear differences among species in the extent of KFB family expansion (Fig. 2; Table S3). No tandem duplication events were detected in any of the four species, whereas segmental duplication was evident in all of them (Table S3). In both diploid species, three segmentally duplicated gene pairs were

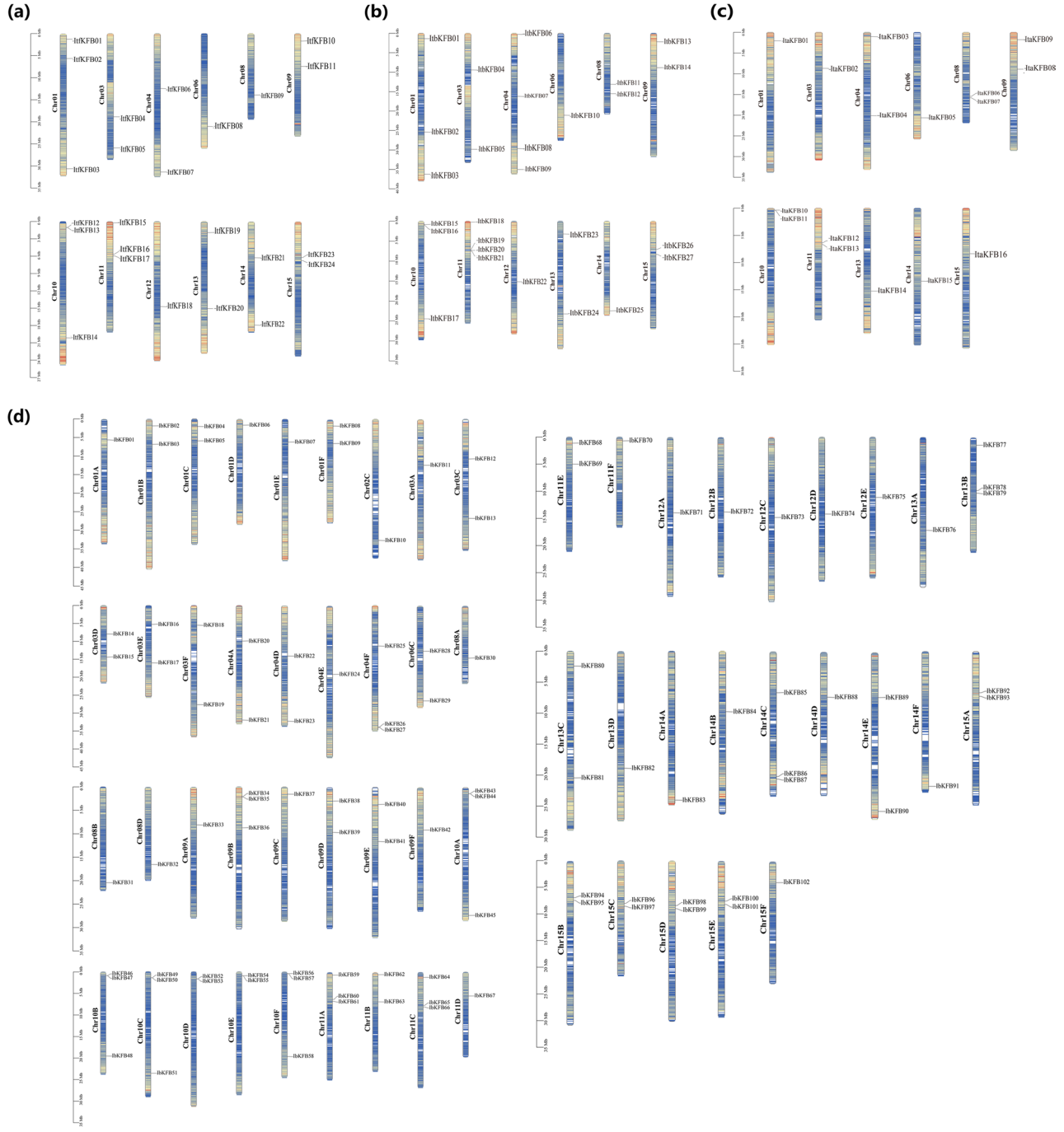

Fig. 1 Physical mapping of KFB genes in *Ipomoea* genomes. Chromosomal distribution of KFB gene family members in (**a**) *Ipomoea trifida* (ItfKFB), (**b**) *I. triloba* (ItbKFB), (**c**) *I. tabascana* (ItaKFB), and (**d**) *I. batatas (*IbKFB*)*. Blue rectangular bars represent sweetpotato chromosomes, with chromosome number and length (Mb) indicated. KFB genes are numbered consecutively based on their physical position on the chromosomes (*ItfKFB01–ItfKFB24*, *ItbKFB01–ItbKFB27*, *ItaKFB01–ItaKFB16*, *and IbKFB01–IbKFB102)*. Physical mapping of KFB genes was visualized using TBtools-II v2.1

identified, and all exhibited Ka/Ks < 1, indicating predominant purifying selection. In *I. tabascana*, only one segmental duplication pair *(ItaKFB10/16*) was detected, and this pair also showed Ka/Ks < 1. In contrast, hexaploid *I. batatas* displayed extensive segmental duplication, with 110 duplicated gene pairs identified. Among these, 102 pairs had Ka/Ks < 1, consistent with widespread purifying selection, whereas one pair (*IbKFB72/74*) had Ka/Ks > 1, implying possible positive selection and functional divergence. Seven additional pairs showed indeterminate Ks values (NaN), which may reflect either recent duplication or synonymous substitution saturation. Collectively,

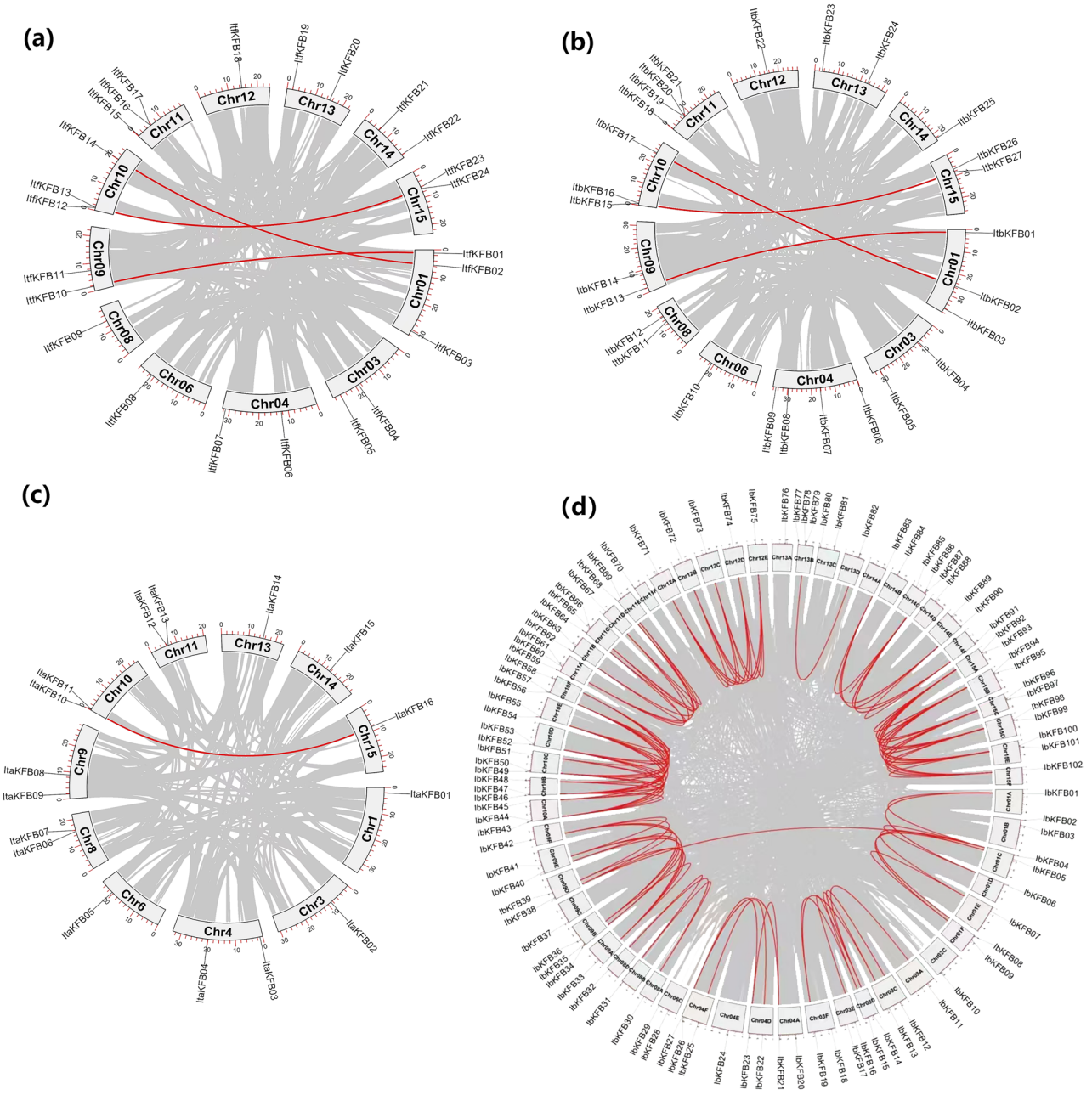

Fig. 2 Gene duplication analysis of KFB genes. Duplicated KFB gene pairs in (**a**) *I. trifida* (ItfKFB), (**b**) *I. triloba* (ItbKFB), (**c**) *I. tabascana* (ItaKFB), and (**d**) *I. batatas* (IbKFB), identified using MCScanX. Red lines indicate duplicated KFB gene pairs, while black lines denote duplicated genes from other families in sweetpotato. Rectangular boxes represent chromosomes, and chromosome numbers are labeled on each chromosome. Diploids (*I. trifida*, *I. triloba*) each had 3 duplicated gene pairs; tetraploid (*I. tabascana*) had only 1 duplicated gene pair; hexaploid (*I. batatas*) had 110 duplicated gene pairs

these results support segmental duplication as the main duplication mechanism associated with KFB family expansion in *Ipomoea*, particularly in hexaploid sweetpotato. At the same time, the reduced KFB complement and limited duplication observed in tetraploid *I. tabascana* indicate that ploidy increase alone is insufficient to explain family size, and that lineage-specific retention or loss also contributed to present-day KFB repertoires.

KFB proteins retain a conserved F-box/Kelch core architecture but show lineage-specific variation in motif composition and exon–intron structure

I. trifida KFB proteins show a conserved motif backbone with limited exon-intron complexity

The ItfKFB proteins contain 5 to 12 conserved motifs, with motifs 1, 2, 3, 4, and 7 being the most widely conserved among members (Fig. 3a). Several motifs showed subgroup-specific distributions, indicating structural

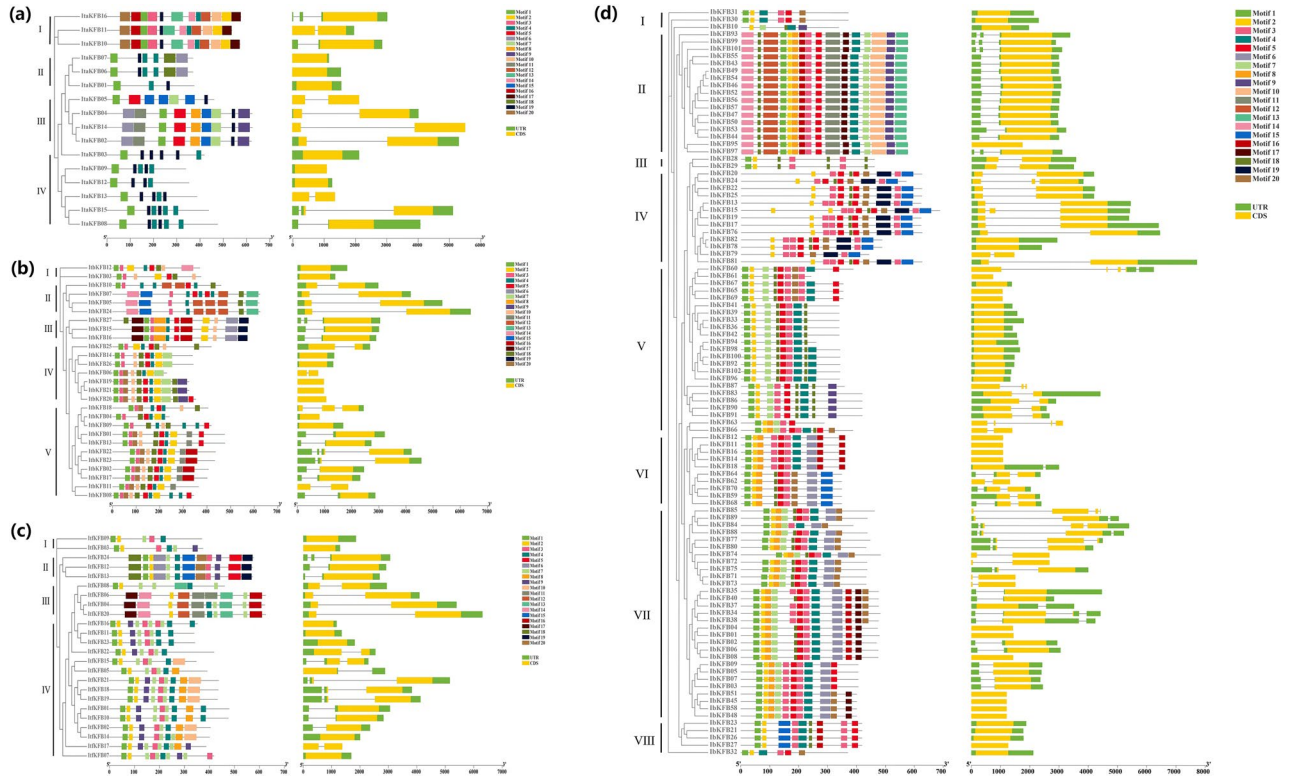

Fig. 3 Gene structure and conserved protein domain organization of KFB genes. Comparative analysis of KFB genes in (**a**) *I. trifida*, (**b**) *I. triloba*, (**c**) *I. tabascana*, and (**d**) *I. batatas*, arranged according to their phylogenetic relationships. Branches indicate distinct KFB subgroups. The phylogenetic tree was constructed using IQ-TREE software using the maximum likelihood method with 1000 bootstrap replicates. Conserved motifs were identified using MEME v5.3.0. The motifs 1–20 are displayed as colored boxes. Protein lengths are indicated by the scale bar at the bottom. Gene structures were visualized using TBtools-II v2.1, with yellow boxes representing exons, green boxes representing untranslated regions (UTRs), and black lines indicating introns. Gene lengths are shown to scale

divergence among subfamilies. Gene structure analysis showed that most ItfKFB genes contained one to two exons, and seven genes were intronless. Motif annotation indicated that motifs 1 and 2 corresponded to the F-box domain, whereas multiple additional motifs encoded Kelch repeats (Table S4).

I. triloba KFB proteins display similar structural conservation with subgroup-specific motif patterns

The ItbKFB proteins contained 7 to 12 conserved motifs, with motifs 1–5, 10, and 18 being highly conserved (Fig. 3b). Most ItbKFB genes contained zero to three introns. As in *I. trifida*, motif composition showed subgroup-specific organization, whereas motif annotation confirmed the conservation of both F-box- and Kelch-associated regions (Table S4).

I. tabascana KFB proteins preserve the KFB core despite reduced family size

The ItaKFB proteins contained 4 to 12 conserved motifs, with motif 1 corresponding to the F-box domain and motif 19 associated with Kelch repeats; both were present in all members (Fig. 3c). Exon–intron structure was relatively simple, with most genes containing zero to two introns (Table S4).

I. batatas KFB proteins show expanded structural diversity consistent with family expansion

The IbKFB proteins contain 5 to 16 conserved motifs. Motifs 1 and 2 were highly conserved and corresponded to the F-box region, whereas motifs 3, 4, 5, and 18 were associated with Kelch repeats (Fig. 3d). A set of 17 genes in subgroup II shared nearly identical motif arrangements, consistent with expansion from closely related ancestral duplicates. Gene structures ranged from zero to four introns, indicating greater structural variation in *I. batatas* than in the other species. Overall, motif and gene-structure analyses across the four species show that KFB proteins retain a conserved F-box/Kelch core architecture while exhibiting subgroup- and lineage-specific structural diversification (Table S4).

Phylogenetic analysis resolves KFB proteins into five major evolutionary clades across multiple plant species

A phylogenetic tree constructed from 432 KFB proteins across eight plant species, including *Ipomoea* spp.,

Arabidopsis, rice, cotton, and water spinach, grouped the proteins into five major clades (I–V) (Fig. 4). Group I and II, containing 16 and 65 members, respectively, were composed mainly of classical Kelch + F-box proteins with relatively simple domain structures. Group III contained 55 members and was distinguished by the presence of additional PAS and GAF domains in some proteins, suggesting greater structural complexity. Group IV, with 226 members, was the largest clade and included many proteins related to previously characterized KFBs involved in metabolic regulation. Group V contained 70 members and appeared comparatively divergent, with a greater proportion of species-specific proteins. This classification indicates that KFB proteins are evolutionarily conserved across land plants while also showing substantial lineage-specific diversification. The broad representation of *Ipomoea* proteins across multiple clades suggests

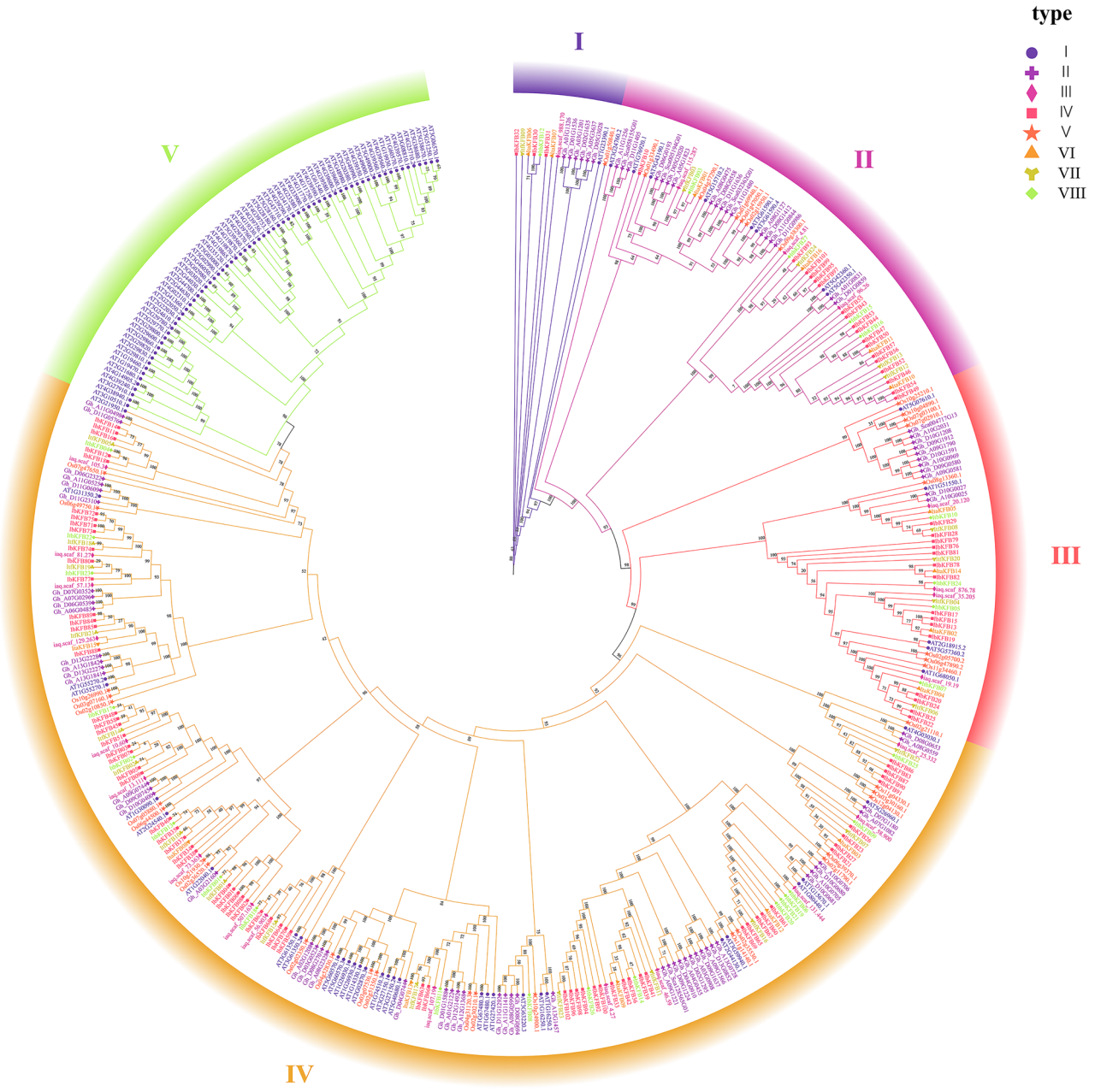

Fig. 4 Phylogenetic analysis of KFB domain-containing proteins from multiple plant species. Phylogenetic relationships of KFB proteins from *Ipomoea batatas*, *I. trifida*, *I. triloba*, *I. tabascana*, *Arabidopsis thaliana*, *Oryza sativa*, *Gossypium hirsutum*, and *Ipomoea aquatica* were analyzed. The phylogenetic tree was constructed using the maximum-likelihood method implemented in IQ-TREE with 1,000 bootstrap replicates. Different KFB subfamilies are highlighted by colored arcs

both retention of ancestral KFB lineages and expansion of particular subclades during *Ipomoea* evolution.

Synteny relationships support conservation of ancestral KFB loci together with expansion in polyploid sweetpotato

Comparative synteny revealed 15 homologous KFB gene pairs between *I. trifida* and *I. tabascana*, 43 pairs between *I. tabascana* and *I. batatas* (Fig. 5a), and 12 pairs between *I. triloba* and *I. tabascana* (Fig. 5b). The larger number of homologous pairs between tetraploid *I. tabascana* and hexaploid *I. batatas* is consistent with greater retention or expansion of KFB loci in the cultivated polyploid genome. These syntenic relationships support the view that a substantial fraction of KFB genes in *I. batatas* are evolutionarily related to loci already present in lower-ploidy *Ipomoea* genomes.

KFB genes show strong tissue-, stress-, and hormone-responsive expression divergence across *Ipomoea*

To examine transcriptional divergence of KFB genes across *Ipomoea*, we analyzed publicly available RNA-seq datasets and generated clustered heatmaps of KFB expression (Fig. 6). In the diploid species *I. trifida* and *I. triloba*, KFB genes displayed clear tissue-preferential expression patterns. A subset of KFB genes, such as *ItfKFB03*, *ItfKFB12*, *ItbKFB02*, *ItbKFB15*, was highly expressed in leaves, whereas others, including *ItfKFB11/23* and *ItbKFB14*, showed root-enriched expression. Only a few genes, such as *ItbKFB06* and *ItfKFB16*, displayed relatively broad expression across tissues. In hexaploid cultivated sweetpotato, multiple *IbKFB* genes responded strongly to abiotic stress treatments. Under cold stress, genes such as *IbKFB33*, *IbKFB41*, *IbKFB96* were strongly upregulated, whereas others were transiently repressed. Under salt stress (Fig. 6e), *IbKFB89*, *IbKFB95*, *and IbKFB100* were strongly upregulated, while some homologs, including *IbKFB36* and *IbKFB39* showed reduced expression. Under drought stress, multiple genes, including *IbKFB42*, *IbKFB92*, and *IbKFB98*, displayed significant induction. These patterns indicate that stress responsiveness is highly diversified across the IbKFB family.

Phytohormone treatments likewise elicited distinct transcriptional responses. Homologous groups containing *ItfKFB11/23* and *ItbKFB14* showed marked induction under IAA and GA treatments. MeJA produced the most

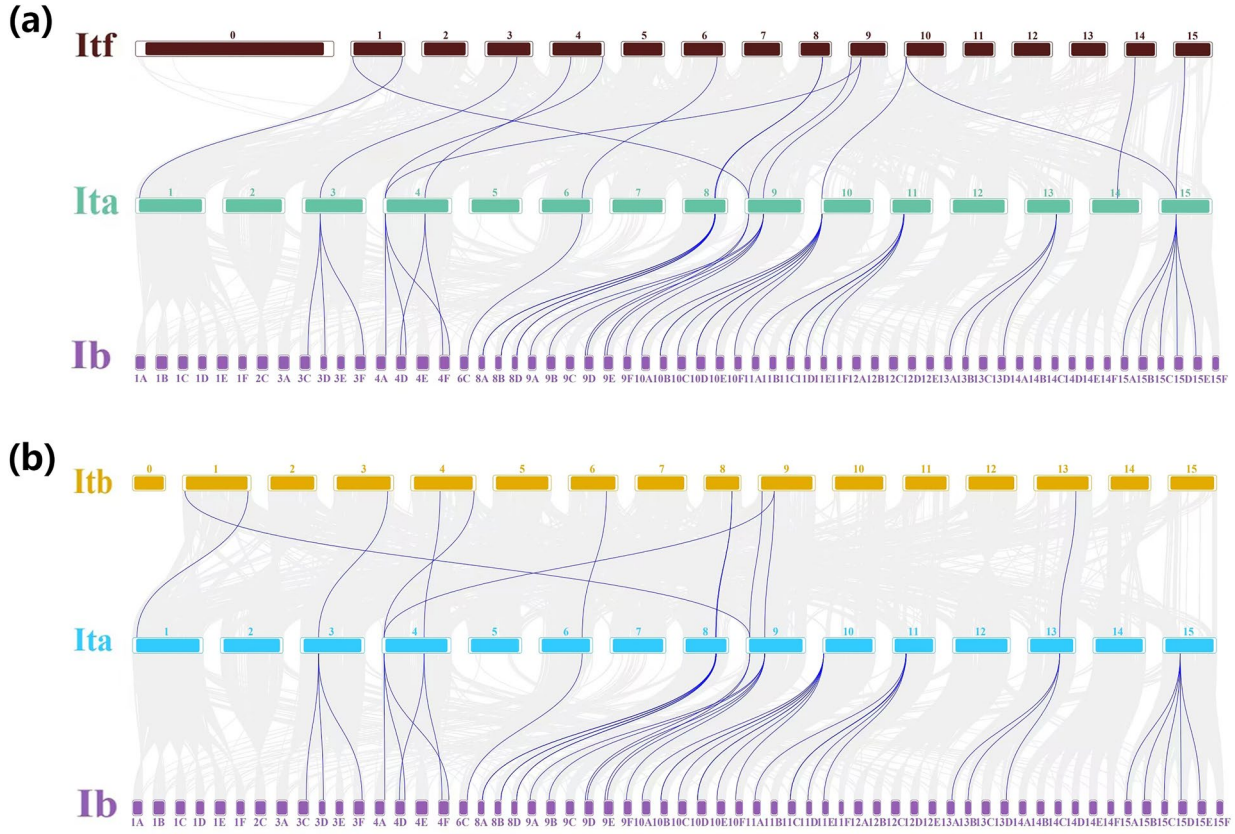

Fig. 5 Collinearity analysis of KFB genes in sweetpotato and related species. Syntenic relationships of KFB genes among (**a**) *I. trifida*, *I. tabascana* and *I. batatas* cv. Beauregard, and (**b**) *I. triloba*, *I. tabascana* and *I. batatas* cv. Beauregard were analyzed. The syntenic relationship was analyzed using MCScanX. Gray lines indicate collinear blocks, while the blue lines highlight the syntenic KFB gene pairs

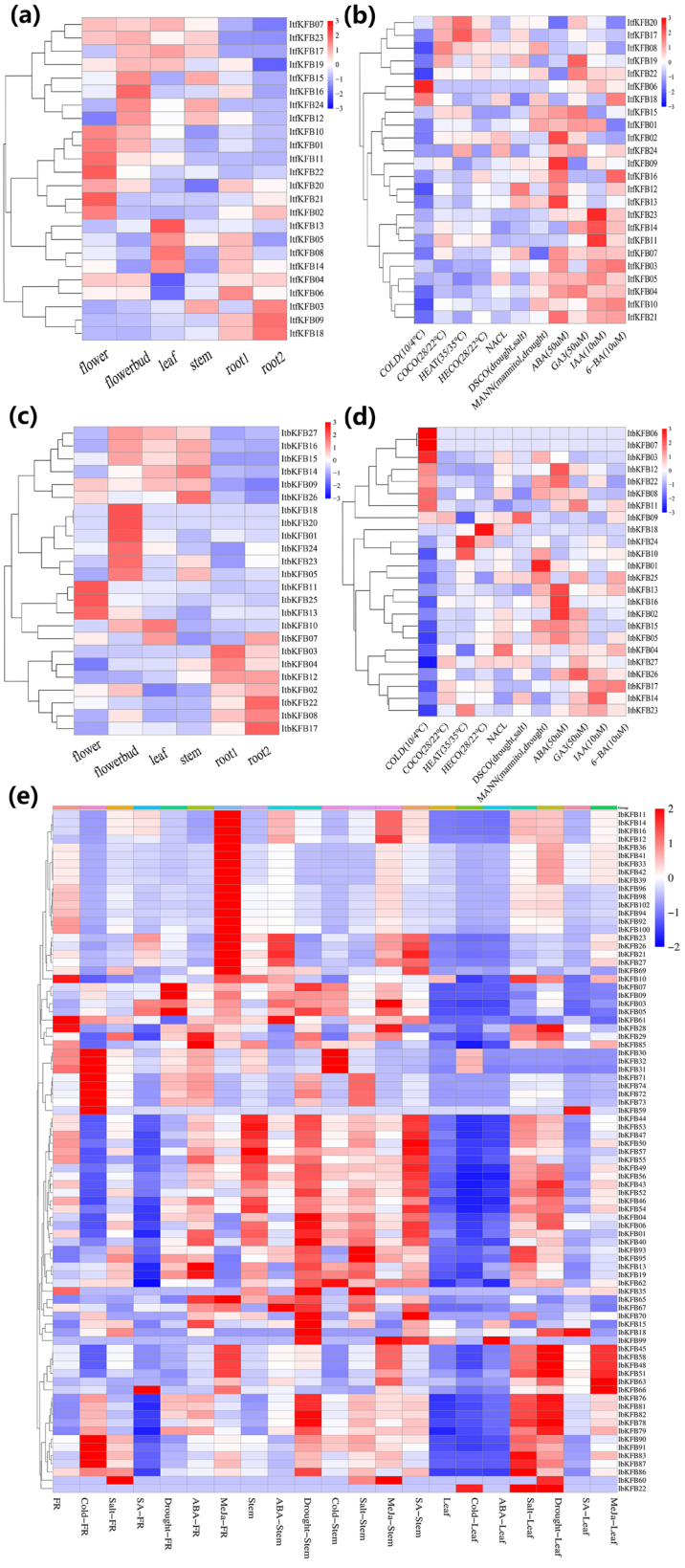

Fig. 6 (See legend on next page.)

(See figure on previous page.)

Fig. 6 Global expression patterns of KFB genes across tissues and treatments. Heatmap visualization of KFB gene expression based on FPKM values: (**a**) ItfKFB genes across different tissues, (**b**) ItfKFB genes under different stress treatments, (**c**) ItbKFB genes across different tissues, (d) ItbKFB genes under different stress treatments, and (**e**) IbKFB genes across tissues and treatments. Public transcriptome datasets for *I. trifida* and *I. triloba* were obtained from the GT4SP Data Download page of the Sweetpotato Genomics Resource database. For *I. trifida*, the tissue-expression matrix included eight RNA-seq libraries (callus_flower, callus_stem, flower, flowerbud, leaf, root1, root2, and stem), and the stress-expression matrix included 15 abiotic and biotic stress RNA-seq libraries. For *I. triloba*, the tissue-expression matrix included six RNA-seq libraries (flower, flowerbud, leaf, root1, root2, and stem), and the stress-expression matrix included 15 abiotic and biotic stress RNA-seq libraries. For cultivated sweetpotato (*I. batatas* cv. Xu Shu 18), RNA-seq data from stem, leaf, and fibrous root tissues under control conditions, three abiotic stress treatments (cold, salt, and drought), and three hormone treatments (ABA, SA, and MeJA), each with two biological replicates, were retrieved from the NCBI SRA database under BioProject PRJNA511028

pronounced transcriptomic response, with twelve *IbKFB* paralogs showing strong induction. In contrast, *ItbKFB26* showed no obvious response in the transcriptome data, suggesting regulatory divergence from otherwise closely related homologs.

qRT-PCR validation supports the hormone responsiveness of most selected KFB candidates

To validate RNA-seq-based expression patterns, 15 candidate genes from homologous clusters were analyzed by qRT-PCR (Fig. 7; Table S5). Fourteen genes exhibited expression trends consistent with RNA-seq results and were significantly induced under at least one hormone treatment. In particular, several IbKFB genes in the JA-responsive cluster exhibited 2- to 10-fold increases in expression after MeJA treatment. By contrast, *ItbKFB26*, which showed little or no responsiveness in the transcriptome analysis, was likewise non-responsive in qRT-PCR assays. This strong concordance between RNA-seq and qRT-PCR supports the reliability of the hormone-response patterns identified for most KFB candidates.

Co-expression analysis identifies conserved and divergent KFB-associated modules across ploidy levels

WGCNA was performed for *I. batatas*, *I. trifida*, and *I. triloba* using whole-genome transcriptome matrices [38]. In *I. batatas*, IbKFB genes were distributed across four co-expression modules, among which the lightcyan1 module contained the largest number of KFB members (15 IbKFB genes) together with 2,482 co-expressed genes. This concentration of IbKFB genes in a single module suggests coordinated transcriptional behavior among a substantial subset of the family.

In the diploid wild progenitor *I. trifida*, ItfKFB genes were distributed across 10 modules. Notably, *ItfKFB11* and *ItfKFB23*, which are homologous to a large IbKFB subfamily (*IbKFB33/36/39/41/42/92/94/96/98/100/102*), were both assigned to the tan module, which contains 534 co-expressed genes. In *I. triloba*, KFB genes were distributed across 12 modules, indicating the greatest module diversification among the three species. Of particular note, *ItbKFB14* was located in the grey60 module containing 4,638 co-expressed genes, whereas *ItbKFB26*, despite its close homology to the same IbKFB group, was placed in the distinct lightyellow module containing 397 co-expressed genes. The separation of closely related orthologs into different modules in *I. triloba* is consistent with transcriptional divergence among KFB homologs across *Ipomoea* species.

GO enrichment links KFB-associated modules to phenylpropanoid regulation, cytokinin signaling, and SCF-related functions

GO enrichment analysis was conducted for co-expression modules containing 11 IbKFB orthologs (*IbKFB33/36/39/41/42/92/94/96/98/100/102*), 2 ItfKFB orthologs *(ItfKFB11/23*), and 2 ItbKFB orthologs (*ItbKFB14/26*), all of which shared > 95% sequence similarity.

In *I. batatas*, the lightcyan1 module containing 2,482 co-expressed genes yielded 373 enriched GO terms (Fig. 8a). Among these, *IbKFB33/36/39/41/42/92/94/96/98/100/102* were associated with four notable terms: negative regulation of cytokinin-activated signaling pathway, regulation of cytokinin-activated signaling pathway, regulation of phenylpropanoid metabolic process, and regulation of secondary metabolic process. These enrichments indicate that the lightcyan1 module is strongly associated with hormone- and metabolism-related processes. In *I. trifida*, the tan module containing *ItfKFB11* and *ItfKFB23* yielded 36 enriched GO terms (Fig. 8b). Two GO terms were shared with the corresponding *I. batatas* ortholog module: regulation of phenylpropanoid metabolic process and SCF ubiquitin ligase complexes. This overlap suggests conservation of a core regulatory context across ploidy levels.

In *I. triloba*, GO enrichment analysis was performed using the combined gene set from the grey60 module containing *ItbKFB14* and the lightyellow module containing *ItbKFB26*. Together, these two modules yielded 82 enriched GO terms, and their enrichment profiles are shown in Fig. 8c. *ItbKFB14* and *ItbKFB26* shared four GO terms, including regulation of phenylpropanoid metabolic process, response to cytokinin, cullin-RING ubiquitin ligase complex, and SCF ubiquitin ligase complex. These results support conservation of a KFB-associated regulatory context related to phenylpropanoid metabolism and ubiquitin ligase functions, while also suggesting species-specific divergence in module composition.

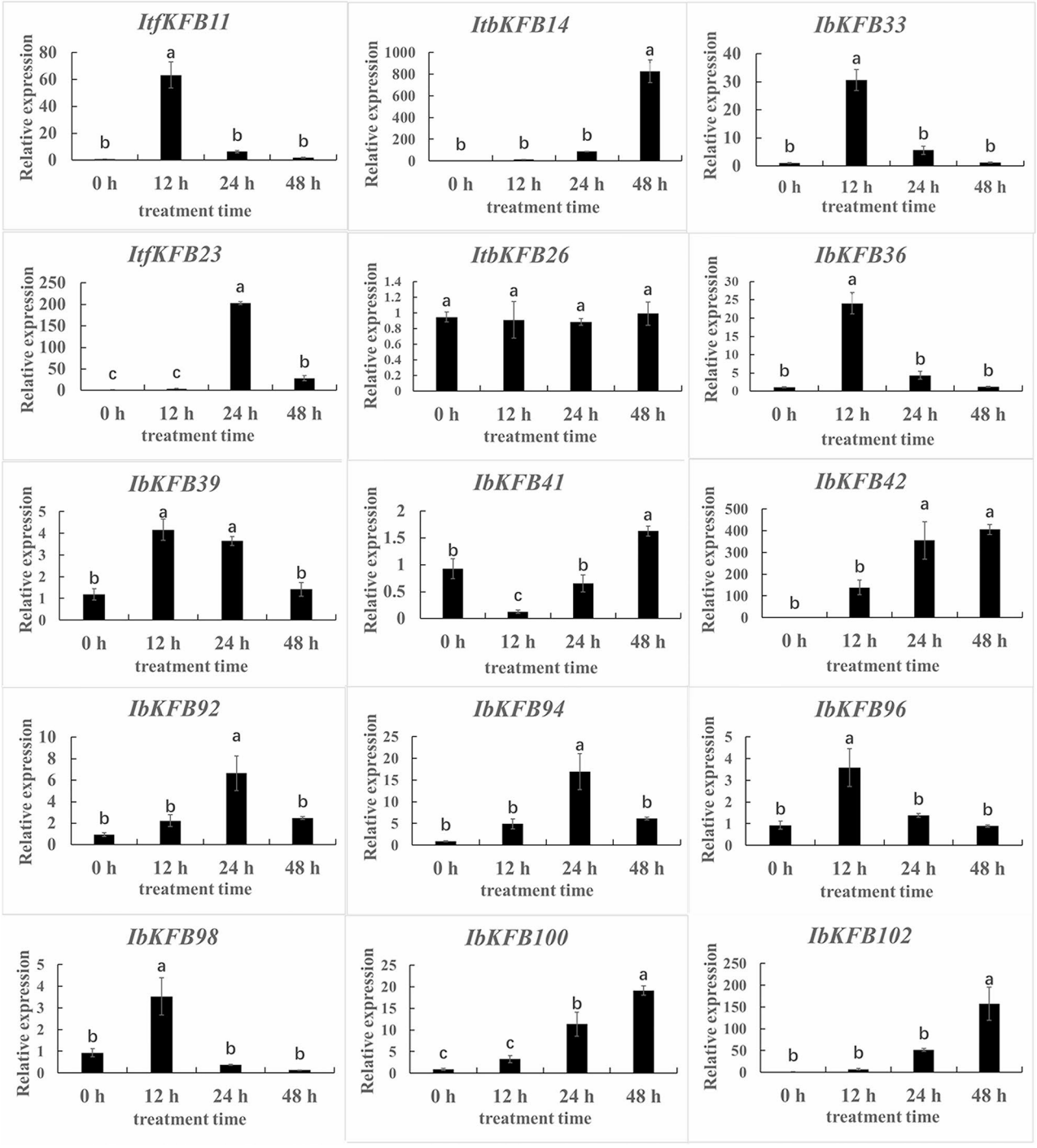

Fig. 7 Expression profiling of KFB genes in response to hormone treatments. Relative expression levels of fifteen KFB genes analyzed by qRT-PCR in *I. trifida*, *I. triloba*, and *I. batatas* cv. Beauregard. Samples were collected at 0, 12, 24, and 48 h after treatment for subsequent analysis. Specifically, *ItfKFB11/23* and *ItbKFB14* were analyzed under IAA treatment, *ItbKFB26* under GA treatment, and twelve *IbKFB* genes under MeJA treatment. Each treatment included three biological replicates. Different lowercase letters represent significant differences between groups, as determined by one-way ANOVA followed by Duncan's multiple range test ($P < 0.05$)

KEGG enrichment highlights module-level functional contexts of KFB-associated co-expression networks

KEGG enrichment analysis was performed for KFB-associated co-expression modules containing the same 15 highly similar orthologs. In *I. batatas*, the lightcyan1 module showed 14 significantly enriched KEGG pathways (all adjusted p-value < 0.05) (Fig. 9a). The most significant pathway was Systemic lupus erythematosus

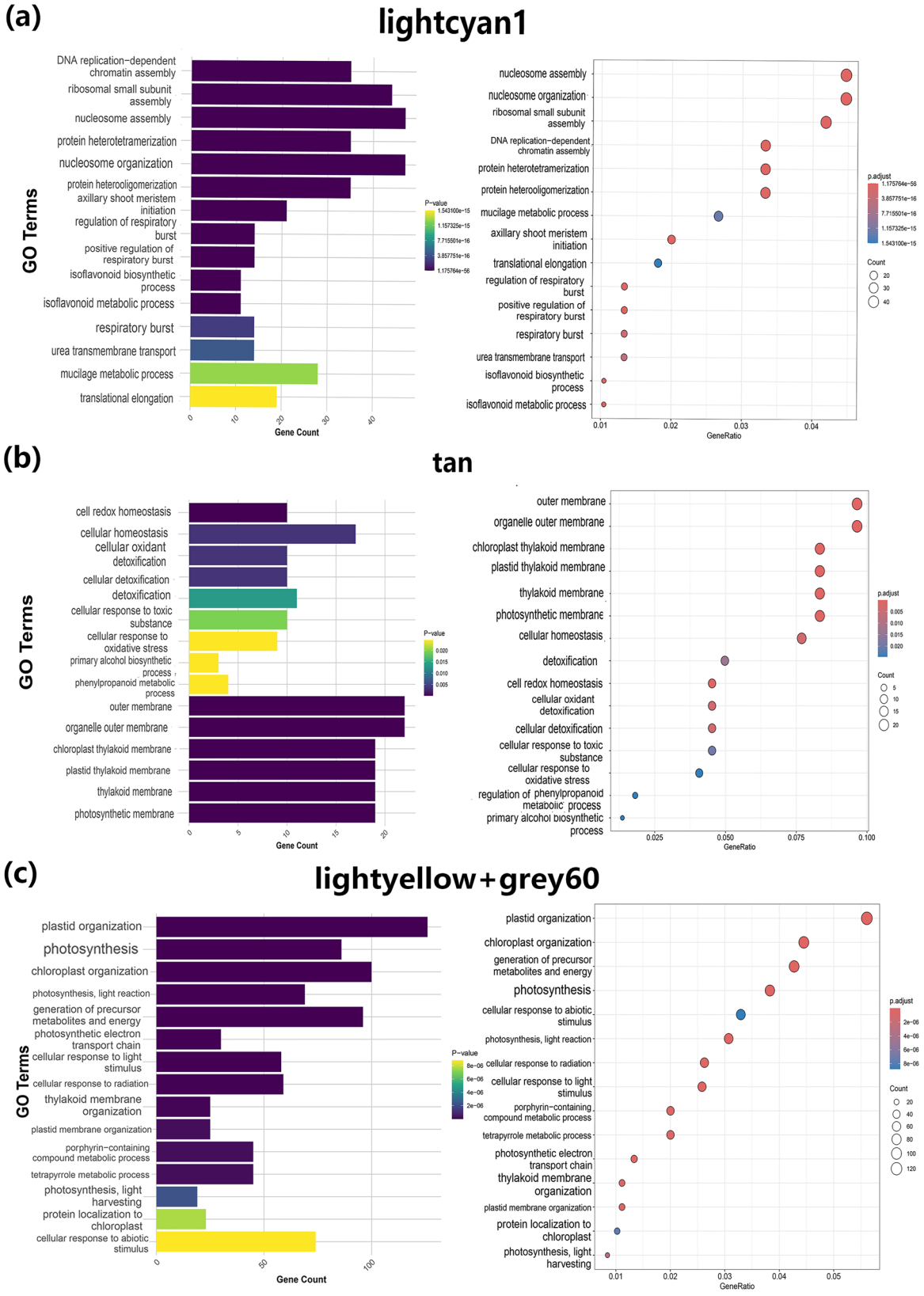

Fig. 8 (See legend on next page.)

(See figure on previous page.)

Fig. 8 Gene Ontology (GO) enrichment analysis of co-expression modules. Bar plots and bubble plots showing GO enrichment results for (**a**) the lightcyan1 module of *I. batatas*, (**b**) the tan module of *I. trifida*, (**c**) the combined lightyellow+grey60 modules of *I. triloba*. The bar plots display GO terms on the y-axis and the number of enriched genes (Gene Count) on the x-axis, with color intensity representing P-value significance. The bubble plots show GO terms on the y-axis and gene ratio (GeneRatio) on the x-axis; bubble size corresponds to the number of enriched genes, and color indicates the adjusted *p*-value, with darker colors denoting higher significance

(map05322), which contained 123 co-expressed genes, with a Fold Enrichment of 36.87 and an adjusted *p*-value of 1.99×10^{-150}. Other significant pathways included Pathogenic *Escherichia coli* infection (map05130), Two-component system (map02020), and Mineral absorption (map04978). However, none of the 11 IbKFB orthologs were directly annotated within these enriched pathways.

In *I. trifida*, KEGG analysis of the tan module identified 183 pathways, but none reached statistical significance after multiple-testing correction (Fig. 9b). Among the non-significant pathways, Thiamine metabolism (map00730) showed the highest Fold Enrichment (8.54; p.adjust = 0.099), followed by Ovarian steroidogenesis (map04913) and Steroid hormone biosynthesis (map00140). The two ItfKFB orthologs, *ItfKFB11* and *ItfKFB23*, were not annotated in any of the 183 pathways.

In *I. triloba*, KEGG enrichment analysis was performed using the combined gene set from the grey60 and lightyellow modules. This combined analysis identified 14 significantly enriched pathways (all adjusted p-value < 0.05) (Fig. 9c). Photosynthesis (map00195) and Photosynthesis—antenna proteins (map00196) were the two most significant pathways, containing 38 and 15 co-expressed genes, respectively; the latter had the highest Fold Enrichment (5.41; adjusted p-value = 1.06×10^{-6}). Biosynthesis of amino acids (map01230) contained the largest number of co-expressed genes (86 genes), with a Fold Enrichment of 1.80 and adjusted p-value = 6.47×10^{-6}. As in the other species, *ItbKFB14* and *ItbKFB26* were not directly mapped to any of the enriched KEGG pathways.

Taken together, these KEGG results indicate that KFB-associated co-expression modules are embedded in transcriptional networks enriched for multiple biological pathways, but the focal KFB orthologs themselves are not directly assigned to the enriched KEGG terms. This pattern is consistent with the interpretation that these KFB genes may function primarily as regulatory components rather than as enzymes directly represented in pathway maps.

Discussion

Our results support the hypothesis that KFB family evolution in *Ipomoea* reflects both polyploidy-associated expansion and post-duplication regulatory divergence, while preserving the core F-box/Kelch framework required for SCF function. Rather than showing a simple dosage effect of ploidy, the KFB repertoire appears to have been shaped by a combination of genome multiplication, lineage-specific retention/loss, and subsequent transcriptional divergence.

A central finding of this study is that the KFB family is dramatically expanded in hexaploid cultivated sweetpotato, with 102 members in *I. batatas*, compared with 24–27 in the two diploid relatives; however, this expansion is not simply proportional to ploidy level because tetraploid *I. tabascana* contains only 16 KFB genes (Fig. 1). This non-linear pattern indicates that polyploidization alone does not determine gene family size in a simple manner. Instead, our data support a more balanced model in which whole-genome multiplication creates the opportunity for expansion, whereas lineage-specific retention and loss shape the final gene repertoire. Similar patterns of lineage-specific expansion and retention have been reported in other plant gene families, including MADS-box genes in potato [39], MADS-box genes in *Brassica* [40, 41], and *NBS-encoding* genes in *Ipomoea* [3]. Therefore, our study implies that polyploidy interacts with lineage-specific retention and divergence to shape KFB family evolution.

Our duplication analyses further support this interpretation. Segmental duplication, rather than tandem duplication, was the major duplication mode detected across the four *Ipomoea* genomes, and this pattern was especially pronounced in *I. batatas*, which contained 110 duplicated KFB pairs (Fig. 2; Table S3). By contrast, the diploid species each contained only three duplicated pairs, and *I. tabascana* retained just one duplicated pair. These results indicate that large-scale genomic duplication and chromosomal rearrangement are associated with KFB family expansion in sweetpotato, consistent with previous studies showing that polyploid-associated segmental duplication is a major driver of gene family expansion in plants [42]. However, the low number of duplicated pairs in tetraploid *I. tabascana* suggests that segmental duplication alone is not sufficient to explain family expansion across all polyploid lineages. At the same time, most duplicated KFB pairs exhibited Ka/Ks < 1 (Table S3), indicating predominant purifying selection and suggesting that the majority of retained duplicates preserve biologically important functions rather than evolving under relaxed constraint [22]. Only one duplicated pair in *I. batatas* showed Ka/Ks > 1, implying that possible neofunctionalization occurred only in a limited subset of paralogs (Fig. 2; Table S3).

Despite the pronounced difference in family size, KFB proteins across *Ipomoea* species retained a conserved

architecture consisting of an N-terminal F-box domain and C-terminal Kelch repeats (Fig. 3; Table S4), consistent with their canonical role as substrate-recognition subunits of SCF ubiquitin ligase complexes [43]. This structural conservation was also reflected in their broadly similar physicochemical properties and predicted subcellular localization patterns across species (Table S2). In particular, many highly similar cross-species protein pairs showed nearly identical amino acid length, molecular weight, pI, GRAVY, and predicted localization, supporting strong evolutionary constraint on core KFB properties. At the same time, variation in motif composition, exon–intron structure, and subgroup organization, especially in *I. batatas*, suggests that structural conservation coexists with lineage-specific diversification (Fig. 3; Table S4). Such a pattern suggests that duplicated KFB genes retain the core biochemical framework required for SCF assembly while potentially diverging in substrate recognition, regulatory responsiveness, or expression pattern [27, 44]. However, whether this structural divergence leads to functional differences requires further investigation.

Phylogenetic analysis further supports the dual pattern of conservation and diversification. Across eight plant species, KFB proteins were resolved into five major clades, with *Ipomoea* proteins represented broadly across these groups (Fig. 4). This indicates that the KFB family contains deeply conserved evolutionary lineages that predate *Ipomoea* diversification, while also showing evidence of lineage-specific expansion in certain clades [8]. The large representation of *Ipomoea* proteins in several clades, particularly in *I. batatas*, is consistent with the retention and expansion of specific KFB *sublineages* during sweetpotato evolution.

Comparative synteny analysis adds an additional evolutionary dimension to this interpretation. The detection of 15 homologous KFB pairs between *I. trifida* and *I. tabascana*, 12 between *I. triloba* and *I. tabascana*, and 43 between *I. tabascana* and *I. batatas* indicates that many KFB loci in cultivated sweetpotato are evolutionarily traceable to loci already present in lower-ploidy *Ipomoea* genomes (Fig. 5). Notably, the highest number of homologous pairs was found between tetraploid *I. tabascana* and hexaploid *I. batatas*, suggesting closer conservation of KFB loci between these two genomes and supporting the possibility that many sweetpotato KFB loci are evolutionarily traceable to lower-ploidy Ipomoea lineages. These collinearity relationships support the view that sweetpotato KFB family expansion involved both retention of ancestral loci and subsequent expansion or diversification after polyploidization, rather than *de novo* appearance of a wholly novel gene set [4]. However, distinguishing between simple retention of ancestral loci and active post-polyploidization expansion would require additional analysis of copy number dynamics and sequence divergence within the hexaploid genome.

A major biological insight from this study is that KFB genes show extensive transcriptional divergence across tissues, abiotic stresses, and hormone treatments (Fig. 6). In the diploid relatives, several KFB genes showed clear tissue-preferential expression, indicating that some degree of functional specialization predates polyploidization. In cultivated sweetpotato, numerous IbKFB genes exhibited strong responsiveness to cold, salt, drought, and multiple phytohormones, with especially notable induction under MeJA treatment (Fig. 6b, d, e). These patterns suggest that expansion of the KFB family in sweetpotato may have increased the regulatory flexibility of this family in environmental and hormonal responses. However, expression responsiveness alone does not establish direct molecular function. Therefore, these data should be interpreted as identifying candidate KFB genes with possible roles in stress and hormone signaling, rather than proving that they directly control these processes.

The hormone-response patterns are particularly noteworthy in the context of previous work showing that IbKFB negatively regulates anthocyanin biosynthesis by promoting the degradation of IbPAL1 and IbGAPCp1 [14]. That study provides a mechanistic precedent linking sweetpotato KFB proteins to phenylpropanoid-related metabolism. In the present work, the strong induction of multiple IbKFB genes under hormone treatments, especially MeJA, together with the co-expression of several orthologous KFB genes with genes enriched for phenylpropanoid-related GO terms, suggests that KFB-mediated post-translational regulation may involve a broader subset of the family than previously appreciated (Figs. 6 and 8). Nevertheless, this inference remains indirect. Our current data support candidate prioritization, but do not yet identify the specific enzyme targets, substrate interactions, or biochemical pathways directly controlled by these KFB proteins.

The qRT-PCR analysis strengthens the reliability of the transcriptome-based observations. Fourteen of the 15 selected candidate genes showed expression trends consistent with the RNA-seq results, and most were significantly induced under at least one hormone treatment (Fig. 7; Table S5). The consistency between transcriptome profiling and qRT-PCR supports the robustness of the hormone-responsive expression patterns identified here. At the same time, qRT-PCR validation confirms transcriptional responsiveness, not downstream protein activity or physiological function. Therefore, these validated genes represent a strong set of candidates for future functional studies rather than definitive regulators of hormone-mediated phenotypes.

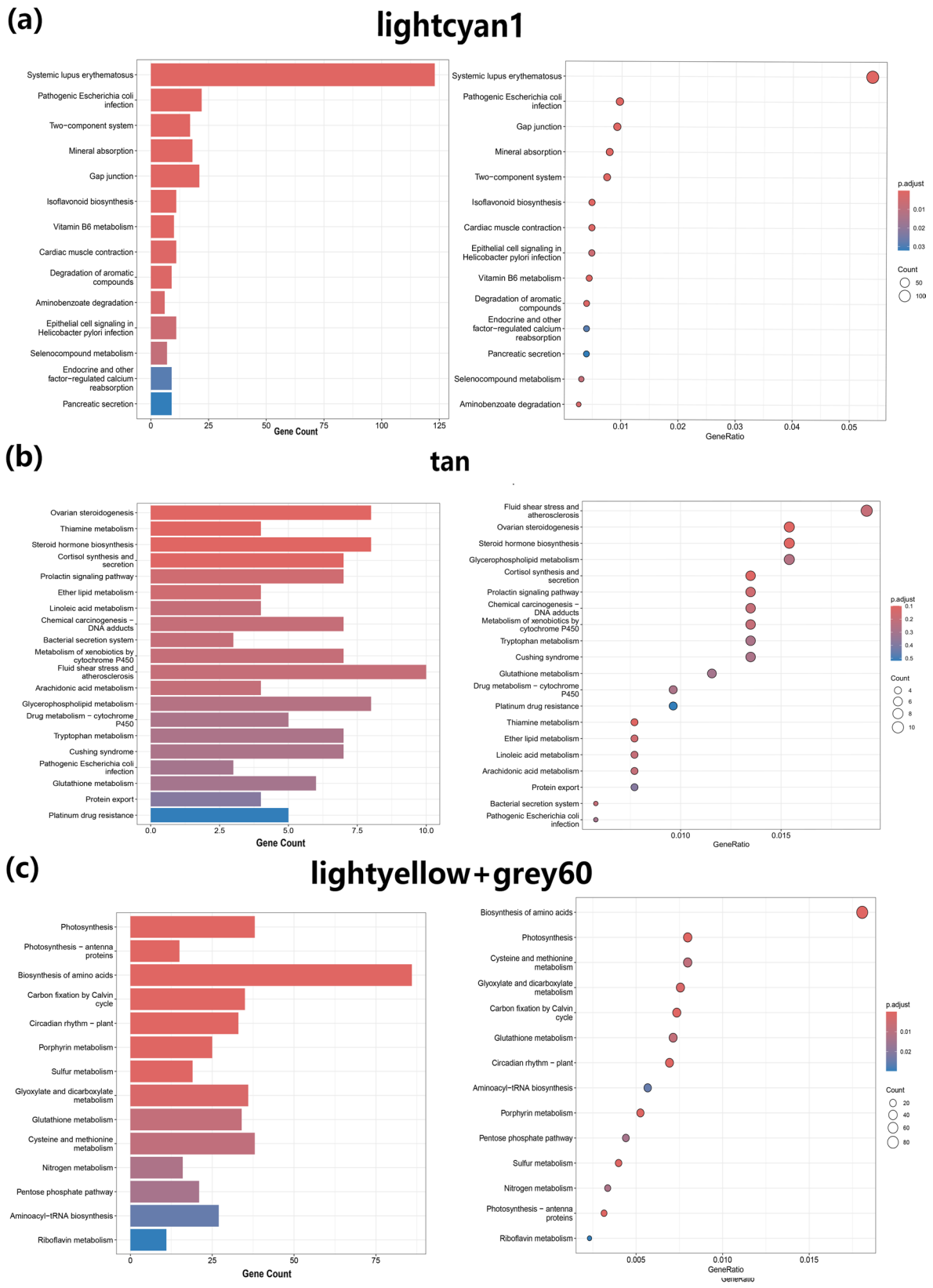

Fig. 9 (See legend on next page.)

(See figure on previous page.)

Fig. 9 KEGG pathway enrichment analysis of KFB-containing co-expression modules. Bar plots and bubble plots showing KEGG enrichment results for (**a**) the lightcyan1 module of *I. batatas*, (**b**) the tan module of *I. trifida*, and (**c**) the combined lightyellow+grey60 modules of *I. triloba*. In the bar plots, KEGG pathway terms are displayed on the y-axis and the number of enriched genes (Gene Count) on the x-axis, with color intensity representing statistical significance (adjusted *p*-value). The bubble plots show KEGG pathway terms on the y-axis and gene ratio (GeneRatio) on the x-axis; bubble size corresponds to the number of enriched genes, and bubble color represents the adjusted *p*-value, with darker colors indicating higher significance

The WGCNA and GO enrichment analyses provide additional support for regulatory diversification while also requiring careful interpretation. Homologous KFB genes from diploid and polyploid *Ipomoea* species were recovered in co-expression modules associated with phenylpropanoid metabolism, cytokinin signaling, secondary metabolism, and SCF ubiquitin ligase-related functions (Fig. S2; Fig. 8). This suggests that the regulatory context of KFB genes has been at least partially conserved across ploidy levels. At the same time, the placement of closely related orthologs into different modules, particularly in *I. triloba*, indicates substantial transcriptional divergence, consistent with subfunctionalization or rewiring of regulatory context after duplication (Fig. S2; Fig. 8) [44]. However, co-expression modules and GO enrichment reflect association rather than direct regulatory control, and therefore these findings should be interpreted as evidence of candidate functional context rather than proof of causal mechanism [34].

The KEGG enrichment results reinforce this interpretation. Several KFB-associated co-expression modules showed significant pathway enrichment, particularly in *I. batatas* and *I. triloba* (Fig. 9). However, the focal KFB orthologs themselves were not directly assigned to the enriched KEGG pathways. One possible explanation is that KFB proteins function primarily as substrate adaptors within ubiquitin ligase complexes and therefore are more likely to act as regulatory components than as enzymes directly represented in canonical pathway maps. Therefore, the KEGG analysis is most informative here as a description of the broader transcriptional environment in which KFB genes operate, rather than as direct evidence of biochemical pathway association (Fig. 9). In this sense, the KEGG data complement the GO results by supporting the idea that KFB genes are embedded within hormone-, stress-, and metabolism-related regulatory networks.

The novelty of this work lies primarily in its cross-ploidy comparative framework. To our knowledge, this is the first systematic analysis of the KFB gene family across diploid, tetraploid, and hexaploid *Ipomoea* species. Previous studies in other plants, including *Salvia miltiorrhiza*, moso bamboo, and potato, have documented species-specific KFB family composition and candidate biological roles [7, 16, 17], but they did not address how KFB family repertoires evolve across related genomes with contrasting ploidy levels. By integrating comparative genomics, expression profiling, qRT-PCR validation, and co-expression analyses, our study provides an evolutionary perspective on sweetpotato KFB genes that complements previous species-specific studies. A second notable aspect is the identification of conserved yet transcriptionally divergent homologous KFB clusters associated with hormone responsiveness and phenylpropanoid-related GO terms (Figs. 6, 7 and 8), which provides a focused set of targets for future mechanistic work.

The significance of this study is both conceptual and practical. Conceptually, the results highlight that variation in KFB family size across ploidy levels is not simply linear with genome duplication events. Factors such as structural conservation, selective constraint, lineage-specific retention, and regulatory divergence appear to be associated with this variation. Practically, the KFB candidates identified here provide a genomic resource for future functional dissection in sweetpotato. Because KFB proteins are implicated in protein turnover, specialized metabolism, and stress-related regulation, these candidates may ultimately prove useful for manipulating traits such as abiotic stress resilience or metabolic composition. Harnessing the untapped biodiversity of *Ipomoea* germplasm through genomics approaches is critical for unlocking novel genetic variation, which can further enhance the resilience and adaptive potential of sweetpotato via targeted molecular breeding.

This study also has several limitations. First, the comparative transcriptome datasets were derived from different experimental sources and designs, which limits strict cross-species quantitative comparability. Second, the WGCNA, GO, and KEGG analyses identify association patterns rather than direct regulatory relationships. Third, although qRT-PCR supported the expression trends of selected candidates, the biochemical substrates and in vivo regulatory functions of individual KFB proteins remain to be experimentally validated. In addition, differences in genome assembly quality and annotation completeness among the four *Ipomoea* species, particularly the haplotype-resolved representation of *I. tabascana*, may have influenced the estimated family sizes and synteny relationships. Future work combining substrate identification, protein interaction assays, and functional genetic analysis will be necessary to establish causal mechanisms.

Conclusions

In conclusion, this study provides the first cross-ploidy comparative characterization of the Kelch repeat F-box gene family in *Ipomoea*. Our results show that KFB family size varies markedly among diploid, tetraploid, and hexaploid genomes and does not scale linearly with ploidy level. Expansion of the KFB family in cultivated sweetpotato was associated primarily with segmental duplication, whereas most retained duplicated pairs remained under purifying selection and preserved the conserved F-box/Kelch structural framework. In parallel, transcriptome profiling, co-expression analysis, and qRT-PCR supported substantial tissue-, stress-, and hormone-responsive expression divergence among KFB genes. Together, these findings establish a comparative evolutionary framework for understanding KFB family diversification in *Ipomoea* and identify candidate genes for future studies of stress adaptation and metabolic regulation in sweetpotato. However, the precise biological functions and regulatory mechanisms of individual KFB genes remain to be validated experimentally.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-026-12867-9.

Supplementary Material 1: Fig. S1. WGCNA soft-threshold analysis of KFB genes in sweetpotato across different ploidies. Determination of the optimal soft-thresholding power for constructing a scale-free co-expression network for (a) hexaploid *I. batatas* (IbKFB genes), (b) diploid *I. triloba* (ItbKFB genes), and (c) diploid *I. trifida* (ItfKFB genes). Left panels show the scale-free topology fit index (R^2) as a function of the soft-threshold power, with the red line indicating the threshold for approximate scale-free topology (= 0.8). Right panels display the mean connectivity (average degree) across different soft-threshold powers. These analyses guided the selection of appropriate power values for WGCNA to ensure biologically meaningful network construction.

Supplementary Material 2: Fig. S2. WGCNA module maps of KFB genes in sweetpotato across different ploidy levels. Module map of (a) IbKFB Genes, (b) ItfKFB Genes, (c) ItbKFB Genes. According to WGCNA analysis, different colors represent different modules, and genes with similar expression patterns are clustered within the same module.

Supplementary Material 3: Table S1. Primer sequences for qRT-PCR validation of transcriptome data.

Supplementary Material 4: Table S2. Physicochemical characteristics and subcellular localization of KFB proteins.

Supplementary Material 5: Table S3. Segmentally duplicated KFB gene pairs.

Supplementary Material 6: Table S4. Motif sequences and corresponding domains of KFB proteins.

Supplementary Material 7: Table S5. Expression levels of candidate genes in response to hormone treatments as determined by RT-qPCR.

Acknowledgements
Not applicable.

Authors' contributions
XJ and LY: conceptualization; YM, GZ, JCC, FM, RW: performed experiments and data analysis; YM wrote the first draft; SZ, JS, JCC prepared the figures; GZ, JHC, LH, XJ and LY: reviewed and edited the manuscript. All authors have read and approved the final version of the manuscript.

Funding
This work was supported by Special Project for Scientific and Technological Cooperation and Exchange in Shanxi Province (202304041101009), The earmarked fund for Modern Agro-industry Technology Research System (2025CYJSTX06-04/-22), Shanxi Province Foreign Experts Introduction Project (202504051011006), Central Government Guides Local Science and Technology Development Fund (YDZJSX2024B007).

Data availability
The RNA-seq data analyzed in this study are available in the NCBI Sequence Read Archive under BioProject accession number PRJNA511028. Genome assemblies and annotation files for *I. batatas* ('Beauregard'), *I. trifida* , and *I. triloba* were obtained from the Sweetpotato Genomics Resource database (http://sweetpotato.uga.edu/), as described in the Materials and Methods. Genomic resources for *I. tabascana* used in this study were obtained from the Sweetpotato Research Institute, Chinese Academy of Agricultural Sciences, Xuzhou, China. A conference abstract describing the whole-genome assembly of *I. tabascana* has been reported previously [19] . All other data generated or analyzed during this study are included in this published article and its supplementary information files.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]College of Life Sciences, Shanxi Agricultural University, Taigu, Shanxi 030801, China
[2]College of Agriculture, Shanxi Agricultural University, Taigu, Shanxi 030801, China
[3]School of Life Sciences, Jilin University, Changchun, Jilin 130012, China
[4]College of Life Sciences, North West Agriculture and Forestry University, Yangling, Shaanxi 712100, China
[5]Shanxi Engineering Research Center for Genetics and Metabolism of Special Crops, Taigu, Shanxi 030801, China
[6]Department of Plant and Soil Sciences and Kentucky Tobacco Research and Development Center, University of Kentucky, Lexington, Kentucky 40546, USA

Received: 28 December 2025 / Accepted: 14 April 2026
Published online: 28 April 2026

References

1. Roullier C, Benoit L, McKey DB, Lebot V. Historical collections reveal patterns of diffusion of sweet potato in Oceania obscured by modern plant movements and recombination. Proc Natl Acad Sci U S A. 2013;110(6):2205–10.
2. Padmaja G. Uses and nutritional data of sweetpotato. In: Loebenstein G, Thottappilly G, editors. The sweetpotato. Dordrecht: Springer; 2009. pp. 189–234.
3. Si Z, Wang L, Qiao Y, Roychowdhury R, Ji Z, Zhang K, Han J. Genome-wide comparative analysis of the nucleotide-binding site-encoding genes in four *Ipomoea species*. Front Plant Sci. 2022;13:960723.
4. Muñoz-Rodríguez P, Wells T, Wood JRI, Carruthers T, Anglin NL, Jarret RL, Scotland RW. Discovery and characterization of sweetpotato's closest tetraploid relative. New Phytol. 2022;234(4):1185–94.
5. Wu S, Lau KH, Cao Q, Hamilton JP, Sun H, Zhou C, Eserman L, Gemenet DC, Olukolu BA, Wang H, et al. Genome sequences of two diploid wild relatives of cultivated sweetpotato reveal targets for genetic improvement. Nat Commun. 2018;9(1):4580.

6. Sadanandom A, Bailey M, Ewan R, Lee J, Nelis S. The ubiquitin-proteasome system: central modifier of plant signalling. New Phytol. 2012;196(1):13–28.
7. Yu H, Jiang M, Xing B, Liang L, Zhang B, Liang Z. Systematic Analysis of Kelch Repeat F-box (KFB) Protein Gene Family and Identification of Phenolic Acid Regulation Members in *Salvia miltiorrhiza* Bunge. Genes (Basel). 2020;11(5):557.
8. Xu G, Ma H, Nei M, Kong H. Evolution of F-box genes in plants: different modes of sequence divergence and their relationships with functional diversification. Proc Natl Acad Sci U S A. 2009;106(3):835–40.
9. Yang Q, Zhao J, Chen D, Wang Y. E3 ubiquitin ligases: styles, structures and functions. Mol Biomed. 2021;2(1):23.
10. Pengyun C, Jingxia Z, Yu C, Guodong L, Chuanyun Z, Furong W, Zhi L, Jun Z. Genome-Wide Identification and Expression Analysis of the KFB Gene Family in *Gossypium hirsutum*. Cotton Sci. 2017;29(03):222–32.
11. Wang B, Zhao X, Zhao Y, Shanklin J, Zhao Q, Liu CJ. *Arabidopsis* SnRK1 negatively regulates phenylpropanoid metabolism via Kelch domain-containing F-box proteins. New Phytol. 2021;229(6):3345–59.
12. Yu SI, Kim H, Yun DJ, Suh MC, Lee BH. Post-translational and transcriptional regulation of phenylpropanoid biosynthesis pathway by Kelch repeat F-box protein SAGL1. Plant Mol Biol. 2019;99(1–2):135–48.
13. Feder A, Burger J, Gao S, Lewinsohn E, Katzir N, Schaffer AA, Meir A, Davidovich-Rikanati R, Portnoy V, Gal-On A, et al. A Kelch Domain-Containing F-Box Coding Gene Negatively Regulates Flavonoid Accumulation in Muskmelon. Plant Physiol. 2015;169(3):1714–26.
14. Tang R, Zhao C, Dong J, Liu X, Chang L, Li J, Dong H, Lv Y, Luo Z, Wu M, et al. Post-transcriptional and post-translational regulation of anthocyanin biosynthesis in sweetpotato by Ib-miR2111 and IbKFB: Implications for health promotion. J Adv Res. 2025;77:57–73.
15. Amoanimaa-Dede H, Shao Z, Su C, Yeboah A, Zhu H. Genome-wide identification and characterization of F-box family proteins in sweet potato and its expression analysis under abiotic stress. Gene. 2022;817:146191.
16. Yang K, Li Z, Zhu C, Liu Y, Sun H, Li X, Gao Z. Identification of KFB Family in Moso Bamboo Reveals the Potential Function of PeKFB9 Involved in Stress Response and Lignin Polymerization. Int J Mol Sci. 2022;23(20):12568.
17. Tang R, Dong H, He L, Li P, Shi Y, Yang Q, Jia X, Li XQ. Genome-wide identification, evolutionary and functional analyses of KFB family members in potato. BMC Plant Biol. 2022;22(1):226.
18. Tripodi P, Cocozza A. Harnessing Crop Diversity through Genetics, Genomics and Phenomics Approaches. Plants (Basel). 2023;12(8):1685.
19. Zhang A, Cao Q. The whole genome of the sweetpotato tetraploid relatives Ipomoea tabascana. In: PAG ASIA2019 Final Program, Abstract & Exhibit Guide; Shenzhen, China. Abstract P0051. 2019. p. 54.
20. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40(7):e49.
21. Guo C, Guo R, Xu X, Gao M, Li X, Song J, Zheng Y, Wang X. Evolution and expression analysis of the grape (Vitis vinifera L.) WRKY gene family. J Exp Bot. 2014;65(6):1513–28.
22. Li J, Zhang Z, Vang S, Yu J, Wong GK, Wang J. Correlation between Ka/Ks and Ks is related to substitution model and evolutionary lineage. J Mol Evol. 2009;68(4):414–23.
23. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–7.
24. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. Nucleic Acids Res. 2015;43(W1):W39–49.
25. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30(9):1236–40.
26. Xie J, Chen Y, Cai G, Cai R, Hu Z, Wang H. Tree Visualization By One Table (tvBOT): a web application for visualizing, modifying and annotating phylogenetic trees. Nucleic Acids Res. 2023;51(W1):W587–92.
27. Zhang X, Gou M, Liu CJ. Arabidopsis Kelch repeat F-box proteins regulate phenylpropanoid biosynthesis via controlling the turnover of phenylalanine ammonia-lyase. Plant Cell. 2013;25(12):4994–5010.
28. Chen Y, Xu Y, Luo W, Li W, Chen N, Zhang D, Chong K. The F-box protein OsFBK12 targets OsSAMS1 for degradation and affects pleiotropic phenotypes, including leaf senescence, in rice. Plant Physiol. 2013;163(4):1673–85.
29. Zhang X, Abrahan C, Colquhoun TA, Liu CJ. A Proteolytic Regulator Controlling Chalcone Synthase Stability and Flavonoid Biosynthesis in Arabidopsis. Plant Cell. 2017;29(5):1157–74.
30. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
31. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37(8):907–15.
32. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923–30.
33. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
34. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
35. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innov (Camb). 2021;2(3):100141.
36. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics. 2012;16(5):284–7.
37. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001;25(4):402–8.
38. Horvath S, Dong J. Geometric interpretation of gene coexpression network analysis. PLoS Comput Biol. 2008;4(8):e1000117.
39. Gao H, Wang Z, Li S, Hou M, Zhou Y, Zhao Y, Li G, Zhao H, Ma H. Genome-wide survey of potato MADS-box genes reveals that StMADS1 and StMADS13 are putative downstream targets of tuberigen StSP6A. BMC Genomics. 2018;19(1):726.
40. Sheng XG, Zhao ZQ, Wang JS, Yu HF, Shen YS, Zeng XY, Gu HH. Genome wide analysis of MADS-box gene family in Brassica oleracea reveals conservation and variation in flower development. BMC Plant Biol. 2019;19(1):106.
41. Song M, Zhang Y, Jia Q, Huang S, An R, Chen N, Zhu Y, Mu J, Hu S. Systematic analysis of MADS-box gene family in the U's triangle species and targeted mutagenesis of BnaAG homologs to explore its role in floral organ identity in Brassica napus. Front Plant Sci. 2022;13:1115513.
42. Flagel LE, Wendel JF. Gene duplication and evolutionary novelty in plants. New Phytol. 2009;183(3):557–64.
43. Lechner E, Achard P, Vansiri A, Potuschak T, Genschik P. F-box proteins everywhere. Curr Opin Plant Biol. 2006;9(6):631–8.
44. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. Genetics. 1999;151(4):1531–45.

Publisher's Note