Check for updates

Analysis of gene expansion and defense-related genes in Anacardiaceae family from an evolutionary aspect

Bing-Liang Fan, Lin-Hua Chen and Ling-Ling Chen*

State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources, College of
Life Science and Technology, Guangxi University, Nanning, China

Introduction: The Anacardiaceae family, encompassing economically and
ecologically significant genera such as *Rhus*, *Mangifera*, and *Pistacia*, exhibits
substantial genomic diversity and adaptive complexity. However, comparative
genomic analysis and evolutionary insights into defense-related mechanisms
within this family remain underexplored.

Methods and results: This study employed a comprehensive phylogenomic,
synteny, and gene family analysis across six *Rhus* species and three additional
Anacardiaceae plants (*Mangifera indica*, *Pistacia vera*, and *Anacardium
occidentale*). Our findings revealed distinct evolutionary trajectories:
Mangifera/Anacardium underwent lineage-specific whole-genome
duplications (WGDs) with chromosomal rearrangements, while *Rhus/Pistacia*
retained only the ancestral gamma duplication. *Pistacia*'s genome expanded
via transposable elements (TEs), whereas *Rhus* conserved chromosomal synteny
despite accumulating TE-mediated structural variations. Some defense-related
gene families, including WRKY transcription factors and nucleotide-binding
leucine-rich repeat (NLR) genes, displayed substantial expansions and stress-
responsive expression patterns, with 31 WRKY genes significantly upregulated
during aphid infestation. NLRs clustered on chromosomes 4/12 showed positive
selection signatures. Long terminal repeat retrotransposons exhibited
Pleistocene-era activation bursts, potentially linked to climatic adaptation.

Discussion: This study unveils TE-mediated diversification as a key driver of
Anacardiaceae evolution, complementing WGD-dependent strategies in tropical
lineages. The identification of lineage-specific structural variations, dynamic TE
activities, and clustered defense-related genes highlights adaptive trade-offs
shaped by biotic stresses and the biogeographic history of Anacardiaceae
species. This study lays the groundwork for leveraging these genomic
resources to enhance stress resilience and adaptive potential in economically
important Anacardiaceae crops.

1 Introduction

The family Anacardiaceae, first proposed and systematically classified in the late 18th to early 19th century, belongs to the order Sapindales within the eudicot clade Rosids (Müller-Wille and Reeds, 2007; Lindley, 2015). This family comprises over 80 genera and 900 species, predominantly distributed across tropical regions of Africa, Asia, and the Americas, with some species extending into subtropical and temperate zones (Dahlgren, 1982; Wannan, 2006). It is classified into five tribes: Anacardieae, Dobineeae, Rhoeae, Semecarpeae, and Spondiadeae (Pell, 2004). Members of the Anacardiaceae family are mainly woody plants, including trees, shrubs, and lianas, while herbaceous forms are rare (Baskin and Baskin, 2022). The family is distinguished by its remarkable ecological adaptability, diverse array of secondary metabolites, and substantial economic importance, making it a vital component of both natural ecosystems and human industries. Notable examples include tropical fruits such as the mango (*Mangifera indica*), cashew (*Anacardium occidentale*), and pistachio (*Pistacia vera*), which are globally cultivated for their edible fruits and seeds (Zeng et al., 2019; Wang et al., 2020c; Mango Genome Consortium et al., 2021). Additionally, several species yield high-quality timber that is resistant to decay, contributing to niche wood markets. The family also holds horticultural importance; for instance, *Cotinus coggygria* (smoke tree) is widely planted for its ornamental foliage (Matić et al., 2016). Medicinally, extracts from plants such as *Pistacia chinensis* (Chinese pistache) have been utilized in traditional remedies for their anti-inflammatory and antimicrobial properties (Paterniti et al., 2017). Beyond its horticultural and medicinal applications, *Toxicodendron vernicifluum* (Chinese lacquer tree) serves as a vital source of raw lacquer (Wang et al., 2020b). By tapping its trunk, this species yields a natural adhesive and coating material renowned for its remarkable properties, such as anticorrosion, antirust, nonoxidation, and resistance to acids, alcohol, and high temperatures (Suzuki et al., 2007; Zhao et al., 2013; Suzuki et al., 2014). These qualities make it indispensable for both traditional craftsmanship and modern industrial applications. Furthermore, a unique feature of *Rhus* species is their intricate interaction with aphids, which induces the formation of Chinese gallnuts (Wu-bei-zi) (Chen et al., 2020; Wang et al., 2020a; Wei et al., 2022). These galls are exceptionally rich in tannins, predominantly gallotannins, and are extensively utilized in traditional Chinese medicine, the chemical industry, and as animal feed additives due to their potent bioactive properties (Zhang et al., 2014; Ren et al., 2021). The family produces a wide array of secondary metabolites, many of which underpin its economic and ecological roles. Key compounds include urushiols (allergenic phenols in lacquer trees) (Ma et al., 2012), anacardic acids (cashew-derived antimicrobial agents) (Dendena and Corsi, 2014), mangiferolic acids (bioactive triterpenoids in mango) (Ediriweera et al., 2017), and tannins (e.g., gallic acid and catechols in gallnuts) (Beretta et al., 2009). These metabolites contribute to defense mechanisms against herbivores and pathogens while offering industrial applications, such as tannins for leather processing and urushiols for producing durable lacquers (Kumanotani, 1995; Chen et al., 2018).

Anacardiaceae exhibits specialized adaptations for pollination and seed dispersal. Floral nectaries attract insect pollinators, while fruits or seeds facilitate dispersal by insects, birds, and mammals (Herrera et al., 2018). Fossil evidence suggests an ancient origin, with genera such as *Pistacia* and *Cotinus* identified in Mediterranean Tertiary deposits (Palamarev, 1989; Marino et al., 2018), and *Toxicodendron* (lacquer tree) and *Rhus* fossils documented across Europe, North America, and Northeast Asia (Andreánszky, 1959; Wolfe, 1966; Yi et al., 2004; Jiang et al., 2019). This paleobotanical record highlights the family's long-term evolutionary success and adaptability to diverse climates. The evolutionary history of Anacardiaceae is deeply intertwined with the Boreotropical flora that dominated the Northern Hemisphere during much of the Tertiary period, particularly from the Eocene to the middle Miocene (Wolfe, 1975). This tropical flora extended across North America, Europe, and into Asia. During the early Tertiary, the western part of North America and East Asia emerged as hotspots for the evolution and diversification of Tertiary relict plants, which later migrated to other parts of the Northern Hemisphere (Xiang, 2001). Botanists have long noted the significant floristic similarities across the Northern Hemisphere, dating back to observations made in the eighteenth century (Gray, 1857; Boufford and Spongberg, 1983; Wen, 1999). Among these, the well-known disjunction between East Asia and eastern North America is particularly striking, often referred to as the East Asia–eastern North America floristic disjunction. Climate deterioration following the late Eocene disrupted the Boreotropical belt, forcing thermophilic plants toward equatorial refugia. Southeastern Asia became a major sanctuary due to its continuous land connection to tropical zones (Morley, 2007), while Central/South America hosted fewer survivors (Burnham and Graham, 1999; Montes et al., 2015). Europe retained minimal Boreotropical elements due to Mediterranean barriers, Alpine uplift, and Sahara expansion (Tiffney, 1985; Morley, 2001). Within Anacardiaceae, the genus *Rhus* exemplifies this biogeographic narrative (Figure 1A). The oldest *Rhus* fossils from western North America (Early Eocene) and continuous records through Pliocene confirm North America as its origin center (Wilf, 2000; Yi et al., 2004). Dual eastward migrations occurred via Beringian and North Atlantic routes during the Eocene thermal maximum, when Boreotropical flora blanketed northern latitudes. Late Eocene *Rhus* fossils in Alaska corroborate Bering Land Bridge (BLB) dispersal, while late Miocene specimens resembling North American *R. glabra* and *R. typhina* in Hungary suggest North Atlantic Land Bridge (NALB) migrations (Andreánszky, 1959; Wolfe, 1966). *Toxicodendron* presents a complementary case of Boreotropical dynamics. Originating in the New World during late Eocene, it diverged into subtropical-temperate and tropical lineages before migrating to East Asia via NALB routes during Oligocene-early Miocene (Jiang et al., 2019). Most species spread from North America to East Asia, where they underwent adaptive radiation and increased species diversity, which is associated with East Asia's climate and continuous land connection.

WRKY transcription factors constitute a major family of plant-specific regulators that play pivotal roles in modulating defense-related gene expression during biotic and abiotic stress responses
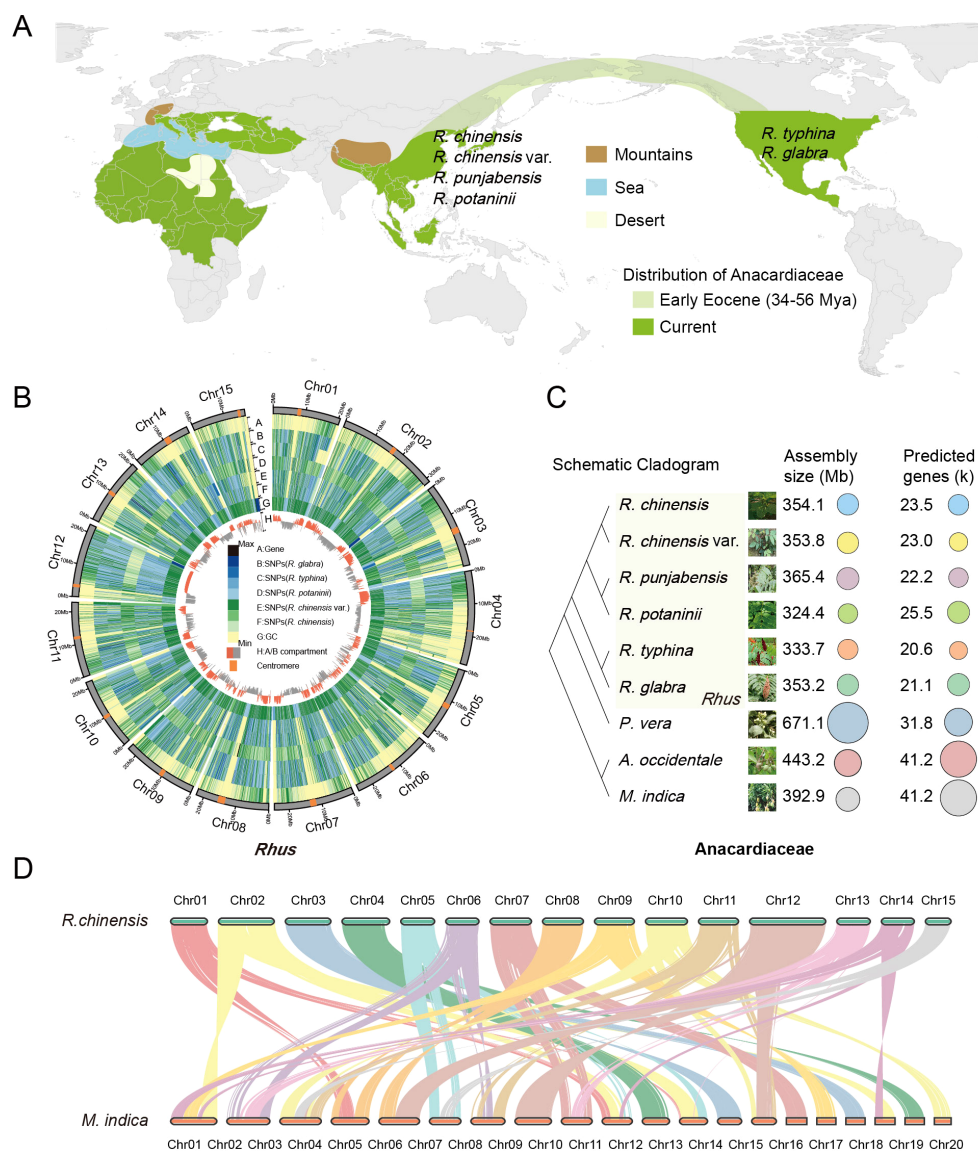
FIGURE 1

Origins and genomic features of Anacardiaceae. **(A)** Current and Eocene distribution of Anacardiaceae is shown in green and light green, respectively. **(B)** Circos plot of distribution of the genomic elements in five *Rhus* species, with *R. punjabensis* as the reference genome. (A–H) Concentric circles from outermost to innermost display protein-coding genes, SNP density in *R. glabra*, SNP density in *R. typhina*, SNP density in *R. potaninii*, SNP density in *R. chinensis* var. *roxburghii*, SNP density in *R. chinensis*, GC content and A/B compartment respectively. Chromosomes with centromeres are highlighted in orange. **(C)** The assembled evolutionary relationships, genome sizes, and annotated gene counts of nine Anacardiaceae genomes are presented in this study. **(D)** Synteny and micro-synteny among *R. chinensis* and *M. indica*.

(Wani et al., 2021; Javed and Gao, 2023). Acting as both positive and negative regulators of plant immunity, WRKYs participate in complex transcriptional networks that contribute to rapid transcriptional reprogramming and enhanced stress resilience (Eulgem and Somssich, 2007). Meanwhile, nucleotide-binding leucine-rich repeat (NLR) proteins are critical components of plant immune systems, mediating defense responses against a broad spectrum of pathogens through the recognition of specific pathogen-derived signals (Jacob et al., 2013; Yang et al., 2017). Despite their independent evolutionary origins in plants and animals, plant NLRs are often organized in genomic clusters, facilitating coordinated immune activation and contributing to

both race-specific and broad-spectrum disease resistance (Wersch and Li, 2019). Furthermore, plant hormones such as salicylic acid (SA), jasmonic acid (JA), abscisic acid (ABA), and ethylene orchestrate intricate signaling networks and crosstalk mechanisms, fine-tuning immune responses, balancing defense and growth, and optimizing resource allocation under stress conditions (Vos et al., 2013; Verma et al., 2016; Berens et al., 2017). Collectively, these regulatory components underpin the sophisticated defense strategies plants employ to mitigate environmental challenges. With its rich phytochemical diversity, ecological versatility, and multifaceted human uses, Anacardiaceae represents a vital plant family bridging natural biodiversity and

socioeconomic development. Despite its importance, evolutionary and adaptive studies within the family have historically been constrained by limited genomic resources, with prior phylogenetic reconstructions relying predominantly on chloroplast genomes or transcriptome-level data. To address these gaps, this study leverages chromosome-level genomes of six *Rhus* species assembled in our laboratory, combined with publicly available genomic data from globally cultivated species (*M. indica*, *P. vera*, and *A. occidentale*), to establish a robust phylogenomic framework. This approach not only resolves long-standing ambiguities in the family's evolutionary relationships but also provides the first comparative genomic perspective on lineage-specific gene family expansions, particularly those linked to environmental adaptation and chemical defense. Furthermore, through bioinformatic analyses, we systematically unravel the structural diversity, copy-number variation, and selection pressures acting on defense-related genes (e.g., WRKY transcription factors, NLR genes, and plant hormone-related genes). By integrating evolutionary genomics with functional annotation, this work advances our understanding of how genomic innovation has driven the diversification and ecological dominance of Anacardiaceae across tropical to temperate ecosystems. Ultimately, these findings offer a foundation for targeted breeding of high-value species, sustainable utilization of bioactive compounds, and conservation strategies tailored to this economically and ecologically pivotal plant family.

2 Materials and methods

2.1 Data sources

The genomic and annotation data utilized in this study were obtained from two sources. First, six *Rhus* genomes assembled by our laboratory were included, achieving gold-standard reference genome quality to ensure the reliability and completeness of this research. Second, publicly available genomes of three additional Anacardiaceae species *Mangifera indica*, NCBI accession GCA_011075055.1 (Wang et al., 2020c), *Pistacia vera*, GCA_008641045.1 (Zeng et al., 2019), and *Anacardium occidentale* (Phytozome13, https://phytozome-next.jgi.doe.gov) were incorporated to broaden the phylogenetic scope (Goodstein et al., 2011). To provide an evolutionary framework for comparative analyses, the *Arabidopsis thaliana* genome (TAIR version 10) was included as an outgroup (Lamesch et al., 2011). This integrated dataset enabled a comprehensive investigation of genomic diversity and evolutionary dynamics within Anacardiaceae (Supplementary Table S1).

2.2 Synteny and whole-genome duplication analysis

Genome collinearity was investigated through JCVI (v1.2.7) using protein-based reciprocal best-hit alignments to detect conserved syntenic blocks (Tang et al., 2008). Putative orthologous relationships were established through pairwise genome comparisons, with aligned regions representing ancestral genomic linkages. Whole-genome duplication (WGD) chronology and speciation divergence times were resolved via WGDI (v0.6.1) analysis (Sun et al., 2022), incorporating calculation of nonsynonymous (Ka) and synonymous (Ks) substitution rates through the NG86 method (Nei and Gojobori, 1986). This integrative approach enabled temporal reconstruction of lineage-specific WGD events and interspecies divergence patterns.

2.3 Long terminal repeat retrotransposon characterization

Full-length long terminal repeat retrotransposons (LTR-RTs) were identified through complementary *de novo* prediction using LTRharvest (v1.5.10) (Ellinghaus et al., 2008) and LTR_FINDER_parallel (v1.1) (Ou and Jiang, 2019). Consensus predictions were integrated and refined through LTR_retriever (v2.9.0) (Ou and Jiang, 2018), with LTR insertion timing calculated using a lineage-specific substitution rate of 7×10^{-9} (Ossowski et al., 2010).

2.4 Variant detection and annotation

Genome-wide SNP identification was performed through MUMmer (v4.0.0beta2) alignment using stringent filtering parameters (-r -q -l 1000) to ensure reciprocal best matches (Marçais et al., 2018). The show-snps utility was implemented with -ClrTH thresholds to detect single nucleotide variants and small indels (< 100 bp). Functional impacts of identified polymorphisms were annotated using SnpEff (v5.0e) through comprehensive variant effect prediction (Cingolani et al., 2012). Visualization of the interchromosomal distribution of structural variations in *Rhus* species using SyRI (Goel et al., 2019).

2.5 Phylogenomic analysis

To investigate the evolutionary relationships within Anacardiaceae, we performed phylogenetic reconstruction incorporating six *Rhus* species from this study along with three additional Anacardiaceae taxa (*P. vera*, *M. indica*, *A. occidentale*) and the outgroup species *A. thaliana*. Single-copy ortholog identification was achieved through OrthoFinder (v2.5.4) analysis (Emms and Kelly, 2019), employing the longest transcript isoforms from ten plant species as input data. Multiple sequence alignment of conserved orthologous sequences was executed using MUSCLE (v5.1.linux64) (Edgar, 2022), followed by identification of phylogenetically informative regions through Gblocks (v0.91b) for site selection (Castresana, 2000).

Optimal substitution model selection via Prottest (v3.4.2) identified Blosum62+I+G+F as the most appropriate amino acid substitution matrix (Darriba et al., 2011). Maximum likelihood phylogenetic inference was subsequently conducted using RAxML

(v8.2.12) with 200 bootstrap replicates to assess nodal support (Stamatakis, 2014). Temporal calibration of divergence events was implemented in MCMCTREE (PAML v4.9j) (Yang, 2007), incorporating a key fossil-derived calibration point (*Rhus-P. vera* split at ~48 mya) from TimeTree and prior paleobotanical studies (Moran, 1989; Kumar et al., 2017).

Gene family dynamics were analyzed through CAFE (v5.0) (Mendes et al., 2021), leveraging OrthoFinder-derived gene clusters to detect significant expansions/contractions (conditional *P* < 0.05 threshold). Functional enrichment analysis of dynamically evolving gene families was conducted using the R package clusterProfiler by performing Gene Ontology (GO) term and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses (Ogata et al., 1999; Ashburner et al., 2000; Yu, 2024). KEGG pathway annotations were retrieved from the official database (http://www.kegg.jp/kegg/pathway.html).

2.6 Defense-related gene annotation and functional classification

To systematically identify and characterize defense-associated genes, we implemented an integrated annotation framework combining structural domain verification and functional ontology analysis. The WRKY transcription factors were rigorously identified by detecting the presence of the PF03106 WRKY domain through Pfam profiles. Specifically, the Pfam database integrated within the InterPro platform (https://www.ebi.ac.uk/interpro/) was queried using Hidden Markov Model (HMM) profiles with a stringent E-value cutoff of 1e-5 to ensure high-confidence annotation of these conserved protein domains (Blum et al., 2024). NLR-type resistance genes were classified based on the presence of characteristic nucleotide-binding domains (PF00931/NB-ARC, PF01582/TIR, PF05659/RPW8) coupled with leucine-rich repeat motifs (PF00560/LRR_1, PF07725/LRR_8, PF13306/LRR_9, PF13855/LRR_6). Functional annotation was extended through GO enrichment analysis targeting seven phytohormone signaling pathways: auxin (GO:0009734), cytokinin (GO:0009736), abscisic acid (GO:0009738), brassinosteroid (GO:0009742), jasmonic acid (GO:0009867), ethylene-activated signaling pathway (GO:0009873), and strigolactone (GO:0010476) responses. Additionally, selective pressure analysis was performed using PAML to calculate the dN/dS ratios of defense-related genes relative to background genes.

2.7 Phylogenetic relationship, gene structure, and conserved motifs analysis

To investigate the evolutionary relationships of defense-associated genes in Anacardiaceae, multiple sequence alignment of WRKY proteins identified in *Rhus* and *Arabidopsis* was performed using MUSCLE (v5.1.linux64). A neighbor-joining phylogenetic tree was subsequently constructed with 1,000 bootstrap replicates in MEGA11 to assess node reliability

(Koichiro et al., 2021). Conserved motifs were predicted using the MEME Suite (v5.5.7) with default parameters (Bailey et al., 2015), while functional domains were annotated via the NCBI Conserved Domain Database (CDD) (Marchler-Bauer et al., 2014). Gene structure features (coding sequences and untranslated regions) and chromosomal localization patterns were visualized using TBtools (v2.225) to integrate genomic architecture and positional context (Chen et al., 2023).

2.8 Cis-regulatory element identification and protein network prediction

We retrieved 2-kb upstream promoter regions of *R. chinensis* WRKY family genes using genome annotation data. Putative cis-regulatory elements were systematically characterized through computational analysis using PlantCARE (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/), with particular emphasis on stress-responsive motifs. Core promoter elements, such as the TATA-box and CAAT-box, were excluded from subsequent analyses to specifically focus on defense-related regulatory sequences. Visualization of spatial element distribution was performed using TBtools with customized annotation parameters (Chen et al., 2023). For protein interaction network inference, a homology-based approach was implemented. All annotated *R. chinensis* protein sequences were compiled into a local BLAST database. WRKY family proteins served as query sequences for iterative BLAST searches with E-value < 1e-5 and sequence identity >= 30%. The network topology was constructed using Gephi (Bastian et al., 2009), in which nodes represent individual WRKY proteins and edges denote significant sequence homology, indicating potential functional associations.

2.9 Transcriptomic profiling and differential expression analysis

Following specimen acquisition, tissues were rapidly preserved through cryogenic freezing to maintain biomolecular stability. Nucleic acid isolation was performed employing commercial plant RNA extraction kits (Tiangen Biotech) with subsequent enzymatic purification using recombinant DNase I (Takara Bio) to eliminate genomic DNA interference. RNA integrity verification was conducted through nucleic acid electrophoresis utilizing 1% agarose matrices impregnated with fluorescent nucleic acid stain, supplemented by dual-platform validation of nucleic acid purity and concentration through spectrophotometric (NanoDrop 2000C) and microfluidic (Agilent Bioanalyzer 2100) quantification systems. For Illumina sequencing, poly-A enriched transcripts were converted into amplifiable cDNA libraries through reverse transcription employing ultra-high-fidelity cDNA synthesis reagents (NEBNext series, NEB). The resulting PCR products were purified using the AMPure XP system, and the library quality was assessed on the Agilent Bioanalyzer 2100 system.

To ensure data reliability, three biological replicates were included for each sample. Raw sequencing data underwent comprehensive quality evaluation through FastQC (v0.12.1) prior to pre-processing (Andrews, 2010). Adapter trimming and read quality enhancement were implemented via Trim Galore (v0.6.7) (Krueger et al., 2021), followed by alignment of processed reads to the reference genome using HISAT2 (v2.2.1) with standard parameters (Kim et al., 2015). Transcript assembly was performed through StringTie (v2.1.0) (Pertea et al., 2015), with subsequent quantification of gene expression levels expressed as fragments per kilobase of transcript per million fragments mapped (FPKM). Differential expression analysis was executed through DESeq2 (v1.38.3) (Love et al., 2014), employing a stringent significance threshold ($P < 0.01$) coupled with a |log2 fold change| ≥ 1 for candidate gene identification.

2.10 Statistical analysis

The statistical details of analysis applied in this paper are provided alongside in the results and methods section. Statistical analyses were performed in R 4.0.0 (R Core Team, 2014).

3 Results

3.1 Genomic features of Anacardiaceae species

The Anacardiaceae family, which includes economically and ecologically important genera such as *Rhus*, *Mangifera*, and *Pistacia*, exhibits substantial genomic diversity and evolutionary complexity. To explore these variations, we used *R. punjabensis* as the reference genome for SNP calling across five other *Rhus* species (*R. chinensis*, *R. chinensis* var. *roxburghii*, *R. typhina*, *R. potaninii*, and *R. glabra*). Analysis of the chromosomal distribution revealed distinct compartmentalization, with the B compartment concentrated in centromeric regions, characterized by higher transposon density, while the A compartment was primarily distributed along chromosome arms, enriched with gene-dense regions (Figure 1B). This spatial organization reflects a high-quality genome assembly, reinforcing the reliability of SNP identification and interspecies comparisons. To further expand our comparative framework, we included three economically important species from the Anacardiaceae family: *P. vera*, *M. indica*, and *A. occidentale*, providing a broader perspective on genomic evolution within the family. As shown in Figure 1C, genome sizes varied significantly, ranging from 324.4 Mb in *R. potaninii* to 671.1 Mb in *P. vera*, with the number of genes spanning 20.6 k to 41.3 k (Supplementary Table S2). Notably, when *P. vera* was excluded, genome sizes of the remaining species showed relatively minor differences, indicating that large-scale expansions may be lineage-specific rather than characteristic of the entire family. Synteny analysis between *R. chinensis* and *M. indica* (Figure 1D) revealed a clear 1:2 syntenic relationship, suggesting

that *M. indica* underwent a WGD event followed by chromosomal rearrangements, resulting in its 20-chromosome configuration. In contrast, *Rhus* species displayed a conserved chromosomal structure without evidence of recent WGD, implying that their diversification was primarily driven by smaller-scale duplications, SNP accumulation, and transposable element activities. These findings highlight not only the high assembly quality of the *Rhus* genomes but also the distinct evolutionary trajectories among major Anacardiaceae lineages.

3.2 Divergent genomic evolution in Anacardiaceae

Phylogenetic analysis based on whole-genome sequencing and orthologous gene families from nine representative Anacardiaceae species, with *A. thaliana* as an outgroup, enabled the reconstruction of evolutionary trajectories and divergence times. The resulting phylogenetic tree revealed that *Rhus* species form a distinct clade closely related to *Pistacia*. Molecular dating suggests that Anacardiaceae originated approximately 97 million years ago (Mya), with *Rhus* diverging from its common ancestor around 48 Mya (Figure 2A). This split likely reflects geographical isolation and subsequent ecological adaptation of *Rhus* across temperate and subtropical regions. Notably, significant gene family expansions were observed in *P. vera*, *M. indica*, and *A. occidentale*, potentially driven by WGD events and transposable element proliferation. Prior studies indicate that *M. indica* and *A. occidentale* share a common WGD event (Mango Genome Consortium et al., 2021), whereas synteny analysis between *R. chinensis* and *M. indica* demonstrated that *Rhus* did not undergo an independent WGD (Figure 2B). Comparative genomic analysis revealed that *Rhus* shares 13,369 gene families with three other Anacardiaceae species while possessing 1,652 unique gene families (Figure 2C). Further syntenic analysis with the ancestral eudicot karyotype (AEK) confirmed that *Rhus* experienced only the gamma duplication event shared among eudicots, without any lineage-specific WGD (Figure 2D). In total, 23,530 genes from *R. chinensis* were clustered into 16,411 gene families, of which 81.5% were conserved across *Rhus* and three other Anacardiaceae species, while 10.0% represented *Rhus*-specific families (Figures 2C, E). These unique gene families may contribute to the species-specific traits and ecological adaptability of *Rhus*.

3.3 Structural variations and retrotransposon dynamics in Anacardiaceae genomes

We conducted a comprehensive comparative analysis of structural variations (SVs) across *Rhus* species using *R. chinensis* as the reference genome. The analysis identified substantial differences in the number and types of genetic variations, including single nucleotide polymorphisms (SNPs), insertions/deletions (InDels), and SVs such as presence-absence variations
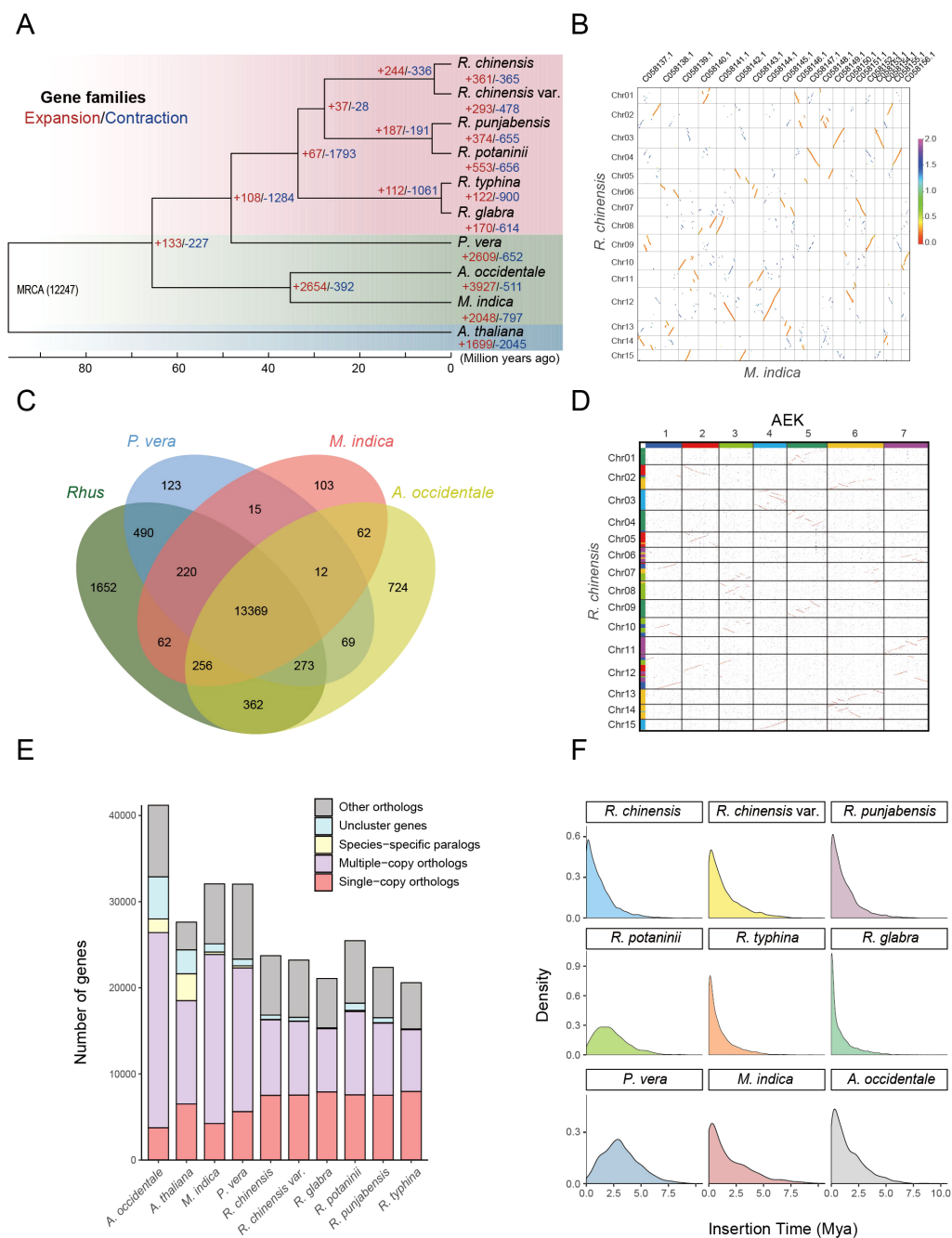
FIGURE 2
Comparative genomic and evolutionary analysis of Anacardiaceae species. **(A)** Phylogenetic tree of 9 Anacardiaceae species and *A. thaliana*, including their estimated divergence times based on orthologs from a single-gene family. Node supports were computed using 200 bootstrap replicates, with all nodes exhibiting 100% support. **(B)** Synteny blocks between *R. chinensis* and *M. indica*, with *M. indica* demonstrating a whole-genome duplication. **(C)** Clusters of orthologous and paralogous gene families in *Rhus* and three other Anacardiaceae species. **(D)** Synteny blocks between *R. chinensis* and the ancestral eudicot karyotype (AEK) indicate that *R. chinensis* has undergone a triplication event. **(E)** Clustering of gene families across different species. **(F)** Insertion times of long-terminal repeat retrotransposons (LTR-RTs) in 9 Anacardiaceae species.

(PAVs), inversions, and translocations. Among these species, *R. chinensis* var. *roxburghii* exhibited the highest number of SVs, with a total of 6,099 events, followed closely by *R. glabra* (5,355) and *R. typhina* (5,218). In terms of specific variation types, *R. potaninii* and *R. glabra* displayed the greatest number of deletions, with 143,237 and 143,254 events, respectively. Notably, the number of insertions was also prominent in *R. potaninii* (132,100), while *R. punjabensis*

demonstrated the highest count of translocations (1,131). Comparative analysis illustrated clear lineage-specific patterns of SV distribution, as reflected in the contrasting SV profiles between *R. chinensis* var. *roxburghii* and *R. chinensis* (Supplementary Figure S1), *R. punjabensis* and *R. potaninii* (Supplementary Figure S2), and *R. typhina* and *R. glabra* (Supplementary Figure S3). These variations may underlie phenotypic divergence, ecological adaptation, and

evolutionary differentiation among Anacardiaceae species. The identification of such structural variations provides valuable insights into the genomic mechanisms contributing to species-specific traits and adaptive capacities within Anacardiaceae.

Furthermore, transposable elements (TEs) have played a pivotal role in shaping genome architecture within Anacardiaceae. In contrast to *M. indica*, *P. vera* does not exhibit lineage-specific genome duplications, although it shares the ancient gamma duplication event common to eudicots. Instead, its larger genome size is primarily driven by substantial TE expansion, with TEs occupying a remarkable 70.7% of its genome and LTR-RTs accounting for 46.75%, rather than typical WGD observed in other species (Zeng et al., 2019). The insertion time distribution analysis of LTR-RTs in various Anacardiaceae species reveals multiple waves of transposition, potentially triggered by environmental shifts that facilitated adaptive evolution (Figure 2F). In *Rhus*, the insertion time analysis of intact LTR-RTs demonstrated that *R. chinensis* and *R. chinensis* var. *roxburghii* exhibit similar insertion patterns (Supplementary Figure S4), with recent transposition bursts observed in both lineages. Likewise, *R. typhina* and *R. glabra* showed comparable insertion dynamics (Supplementary Figure S5). In contrast, *R. punjabensis* and *R. potaninii* displayed distinct LTR insertion timelines, with *R. potaninii* experiencing a transposition burst approximately two million years ago, while *R. punjabensis* showed more recent LTR insertions around one million years ago (Supplementary Figure S6). The chromosomal distribution of retrotransposons suggests a potential role in enhancing genomic plasticity and adaptive capacity.

3.4 Gene family expansion and functional enrichment analysis in Anacardiaceae

Gene family expansion, a critical driver of adaptive evolution (Guo, 2013), was analyzed using the CAFE program across nine Anacardiaceae species. Notably, *P. vera*, *M. indica*, and *A. occidentale* exhibited significant expansions of 2,609, 2,048, and 3,927 gene families, respectively, with expansions notably outpacing contractions, suggesting that these species undergo adaptive evolution primarily through the expansion of specific gene families. Functional enrichment analyses revealed defense-related pathways as central to these expansions. For *P. vera* (Figure 3; Supplementary Tables S3, S4), GO terms such as "defense response to bacterium," "xenobiotic detoxification by transmembrane export across the plasma membrane," and "protein phosphorylation" were prominently enriched (Figure 3A). Correspondingly, KEGG pathways included "plant–pathogen interaction" and "biosynthesis of various plant secondary metabolites" (Figure 3B), highlighting their involvement in redox regulation, phytohormone-mediated defense signaling, and defense-associated metabolic processes. In Anacardiaceae, enriched GO terms included "defense response to bacterium," "defense response," and "signal transduction," reflecting defense-related biological processes (Supplementary Figure S7; Supplementary Tables S5, S6). Similarly, *M. indica* showed

significant enrichment in "regulation of defense response to fungus" and "response to auxin," suggesting that the expanded gene families are predominantly associated with defense-related functions (Supplementary Figure S8; Supplementary Tables S7, S8). For *A. occidentale*, KEGG pathways such as "Plant hormone signal transduction" were highlighted, indicating its involvement in hormone-mediated defense responses (Supplementary Figure S9; Supplementary Tables S9, S10).

3.5 Phylogenetic and structural analysis of WRKY transcription factors in *Rhus* species

WRKY transcription factors (TFs) are pivotal regulators of plant defense mechanisms, orchestrating responses to biotic and abiotic stresses (Jiang et al., 2017). In *R. chinensis*, a comprehensive analysis was performed to investigate the phylogenetic relationships, conserved domains, and structural organization of WRKY TFs. Based on the conserved WRKY domain PF03106, a total of 55 WRKY genes were identified in *R. chinensis*. To elucidate their evolutionary relationships, 15 representative WRKY sequences from *A. thaliana* were retrieved from the TAIR database and aligned with the 55 *R. chinensis* WRKY sequences for phylogenetic analysis (Supplementary Figure S10; Supplementary Table 11). Structural analysis of the WRKY proteins revealed the presence of 10 conserved motifs distributed across the 55 members. These motifs exhibited distinct patterns among different WRKY subgroups, suggesting divergent evolutionary trajectories and specialized physiological roles (Figure 4; Supplementary Table S12). All *R. chinensis* WRKY proteins contained at least one WRKY superfamily domain. Additionally, several members harbored other conserved domains, such as plant-specific zinc finger motifs and bZIP superfamily elements, which may enhance transcriptional regulation under pathogen invasion and environmental challenges (Figure 4). Gene structure analysis further highlighted substantial variation in intron-exon arrangements among WRKY members. Some genes exhibited multiple introns, indicative of potential alternative splicing events that could contribute to functional diversity and regulatory complexity.

3.6 Cis-regulatory element profiling of WRKY gene promoters

The functional diversity of genes is closely associated with the composition and abundance of cis-regulatory elements in their promoter regions. To investigate the potential regulatory mechanisms of WRKY transcription factors, we systematically analysed 2 kb upstream sequences of WRKY genes, with particular emphasis on cis-elements associated with abiotic stress adaptation, biotic stress responses, and defense metabolism regulation (Figure 5A). Our analysis revealed three predominant regulatory elements: MeJA-responsiveness (involved in jasmonate signaling), abscisic acid (ABA) responsiveness (central to drought and osmotic stress adaptation), and salicylic acid (SA) responsiveness (critical for systemic acquired resistance against pathogens), which emerged as the most prominently represented stress-related motifs (Figure 5B).
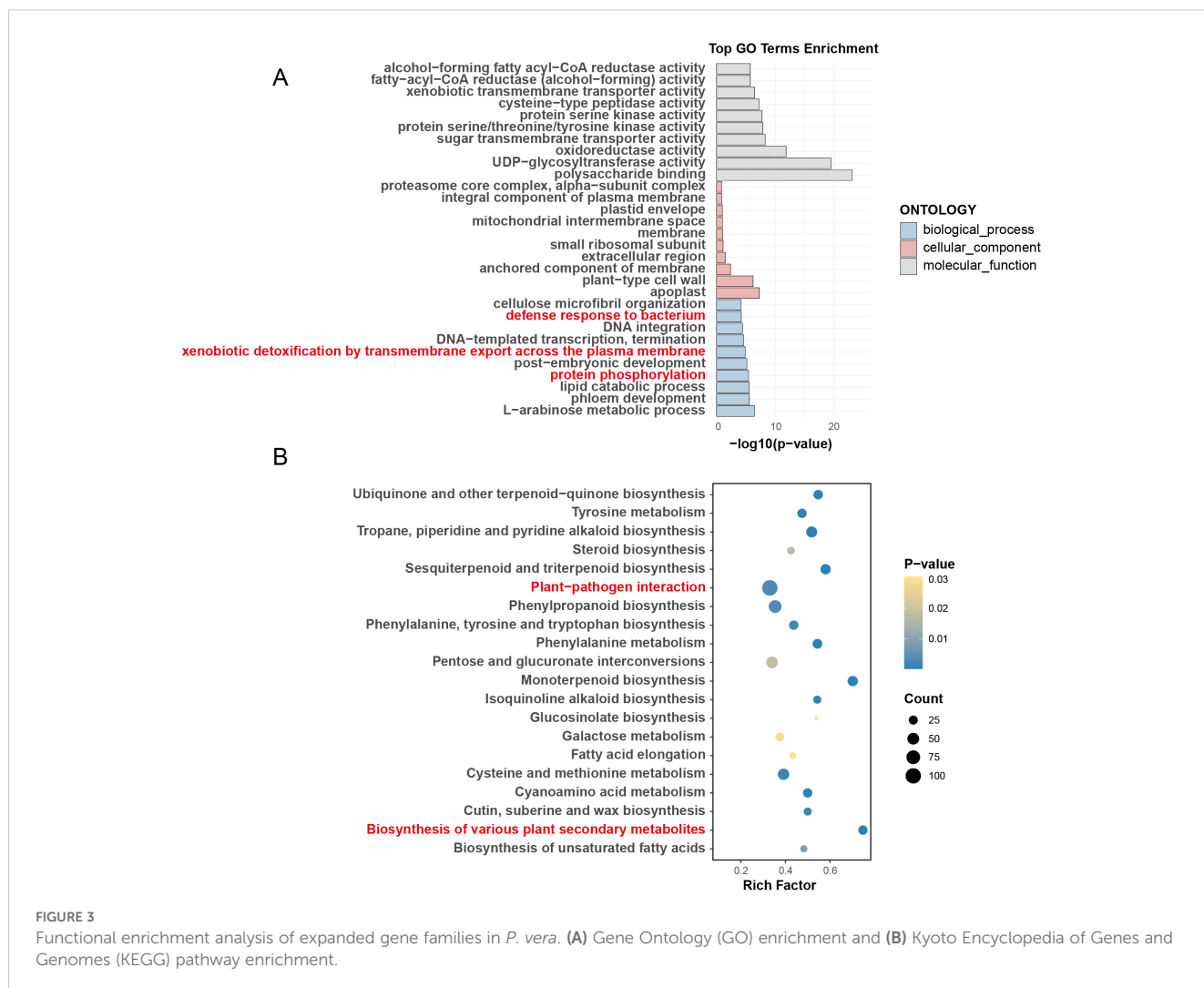
FIGURE 3
Functional enrichment analysis of expanded gene families in *P. vera*. **(A)** Gene Ontology (GO) enrichment and **(B)** Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment.

These elements coordinate plant responses to environmental challenges through phytohormone-mediated signalling pathways. Furthermore, the promoter regions contained multiple specialized cis-regulatory modules, including elements governing low-temperature adaptation (e.g., dehydration, low-temperature, and salt stress-responsive motifs), hypoxia-inducible enhancer-like sequences, and MYB-binding sites regulating both drought-inducible gene expression and flavonoid biosynthesis. Notably, wound-responsive elements and pathogen defense motifs (e.g., elicitor-mediated activation elements) were selectively retained in specific WRKY promoters, implying specialized adaptation for localized stress recognition. The homology-based protein interaction network, constructed using sequence similarity thresholds, revealed a striking pattern of intra-family connectivity among WRKY proteins (Figure 5C). Nodes representing WRKY members exhibited dense clustering, with extensive edge formation signifying robust sequence homology-driven associations. Notably, no significant interactions were detected between WRKY proteins and non-WRKY proteins, suggesting functional modularity or selective evolutionary constraints that prioritize intramolecular cooperation within the WRKY family.

3.7 Distribution and evolutionary selection pressure of defense-associated genes

Chromosomal mapping revealed that WRKY, NLR, and phytohormone-related genes are unevenly distributed across the chromosomes of *R. chinensis*. Notably, these genes exhibit clustered arrangements in specific chromosomal regions (Figure 6A; Supplementary Figure S11). Among them, NLR genes are particularly enriched on chromosomes 4 and 12 (Supplementary Figure S12), suggesting that these loci may play crucial roles in adaptive defense mechanisms. In *R. chinensis*, genome-wide identification and characterization of these defense-associated genes indicate their extensive presence across multiple chromosomes, with pronounced clustering at certain loci. This non-random distribution pattern implies potential hotspots for evolutionary adaptation, possibly driven by selective pressures to enhance pathogen defense. To further understand the evolutionary dynamics of these genes, we performed evolutionary analysis using the dN/dS ratio (nonsynonymous to synonymous substitution ratio). The analysis identified evidence of positive selection in these defense-related gene families (Student's t-test, $P < 0.05$;

FIGURE 4
Structure of *R. chinensis* WRKY family members. **(A)** Phylogenetic relationships among WRKY family members (left) and distribution of conserved motifs in *R. chinensis* WRKY proteins (right). **(B)** Architecture of conserved domains in *R. chinensis* WRKY proteins. **(C)** Schematic structures of *R. chinensis* WRKY genes.
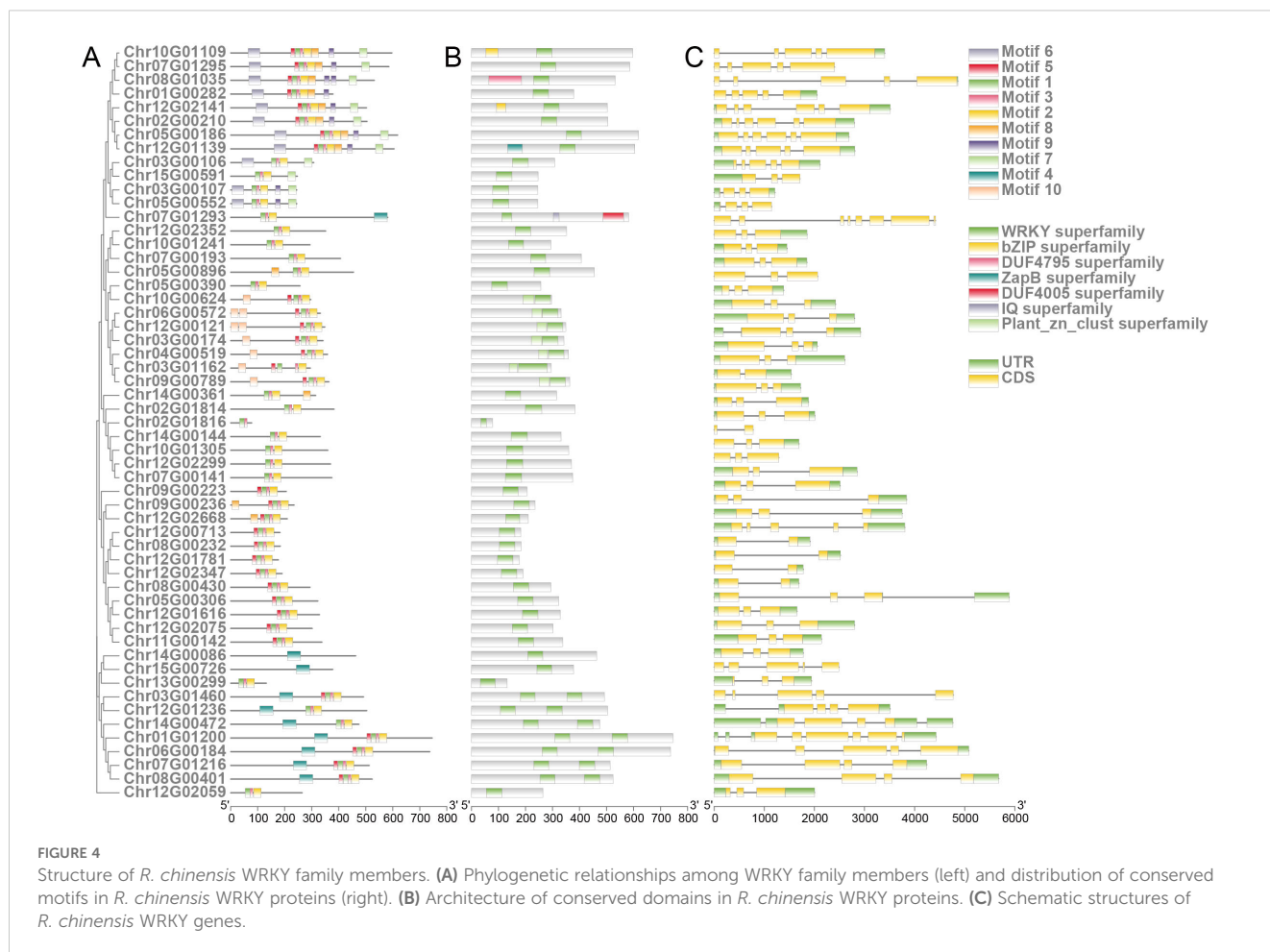
Figure 6B), indicating that mutations in these loci are likely beneficial for improving defense responses against biotic stresses. These findings underscore the evolutionary importance of defense-associated genes in shaping the adaptive landscape of *Rhus* species.

3.8 Synteny and comparative analysis of defense-related genes across Anacardiaceae species

In this study, the synteny relationships of defense-related genes were analyzed among six *Rhus* species and three additional members of the Anacardiaceae family, namely *M. indica*, *P. vera*, and *A. occidentale*. High-resolution synteny maps revealed both the conservation and specificity of defense-related gene loci, with particular emphasis on regions associated with the WRKY gene family. The analysis primarily focused on syntenic clusters of WRKY-related genes, as illustrated in Figure 6C. Notably, conserved WRKY gene fragments were identified at the terminal or proximal regions of chromosomes 2, 3, and 7 across the six *Rhus* species, indicating strong syntenic relationships. Typically, these segments corresponded to a single copy in *P. vera*, whereas *M. indica* and *A. occidentale* maintained two copies, suggesting lineage-specific duplication events. Interestingly, certain deviations were observed.

For example, on chromosome 7, the two North American *Rhus* species (*R. typhina* and *R. glabra*) exhibited specific loss of syntenic segments compared to the four East Asian *Rhus* species. Conversely, these syntenic regions were well-preserved in the more distantly related *P. vera*, indicating potential species-specific adaptive evolution.

3.9 Expression dynamics of defense-related genes under biotic stress

To validate the functional relevance of defense-associated genes identified in *Rhus* species, we investigated their transcriptional responses to infestation by gall-forming aphids, obligate parasites that induce tannin-rich galls (Wei et al., 2022). Samples were collected from four tissue-time point combinations: leaf-wing before infection (WBI), leaf-wing after 7 days of infection (WAI-7d), leaf before infection (LBI), and leaf after 7 days of infection (LAI-7d). RNA-seq analysis revealed that biotic stress significantly upregulated 31 WRKY transcription factor genes in leaves (LAI-7d vs. LBI; fold change >2; Figure 6D; Supplementary Tables S13, S14). Moreover, WRKY genes in leaf-wings (WAI-7d vs. WBI) exhibited a broad upregulation trend across the gene family, with expression levels increasing in the majority of members, though these changes did not reach statistical significance. Similarly, defense-related genes,
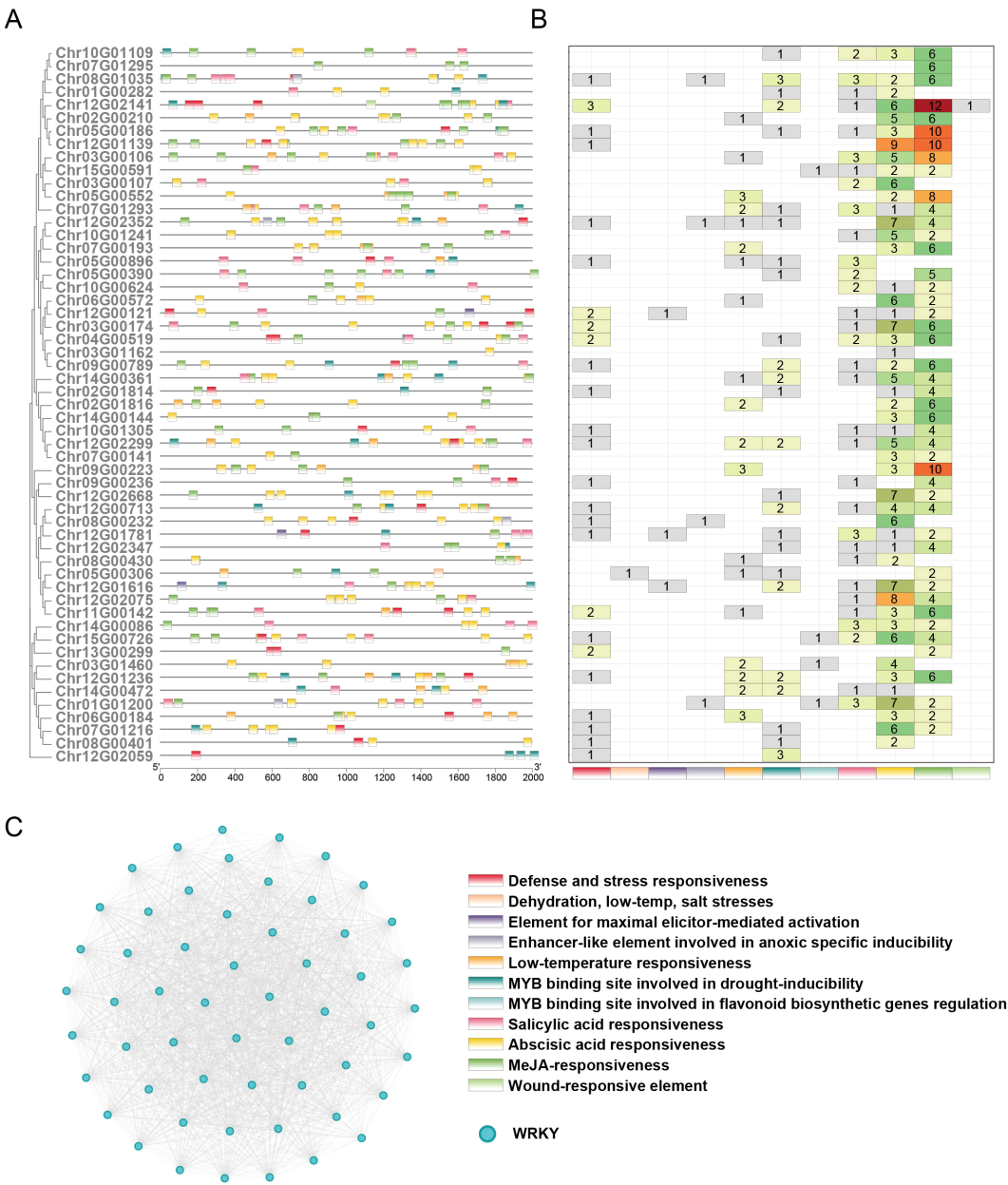
FIGURE 5
Defense-related cis-regulatory elements and protein interaction networks of WRKY gene family members. **(A)** Localization of 11 defense-related cis-acting elements within the 2-kb upstream promoter regions of WRKY genes. Distinct hues represent individual element types, and the horizontal scale denotes nucleotide positions relative to the transcription initiation site. **(B)** Frequency distribution of defense-related cis-element variants across WRKY gene family members. **(C)** The protein-protein interaction (PPI) network represents WRKY proteins as nodes, with edges (gray lines) denoting pairwise connections between proteins exhibiting >=30% sequence identity.

including NLRs and phytohormone signaling components, exhibited pronounced expression shifts post-stress (Supplementary Figure S13). These results demonstrate that defense-related gene families under positive selection are functionally responsive to biotic stress.

4 Discussion

The Anacardiaceae family exemplifies contrasting evolutionary strategies shaped by chromosomal stability and TE-driven plasticity. While *M. indica* (2n = 40) and *A. occidentale* (2n = 42) underwent lineage-specific WGDs post-eudicot gamma duplication (Mango Genome Consortium et al., 2021; Savadi et al., 2022), *P. vera* (2n = 30) and *Rhus* species (2n = 30) retained ancestral karyotypes (Zeng et al., 2019; Ni et al., 2024), diverging via smaller-scale genomic changes. This aligns with prior hypotheses that TE-mediated expansion, rather than WGD, drives adaptation in taxa occupying stable niches (Bennetzen, 2005; Fedoroff, 2012; Lisch, 2013). Our phylogenomic analysis of nine Anacardiaceae species further resolves their Boreotropical origins, linking
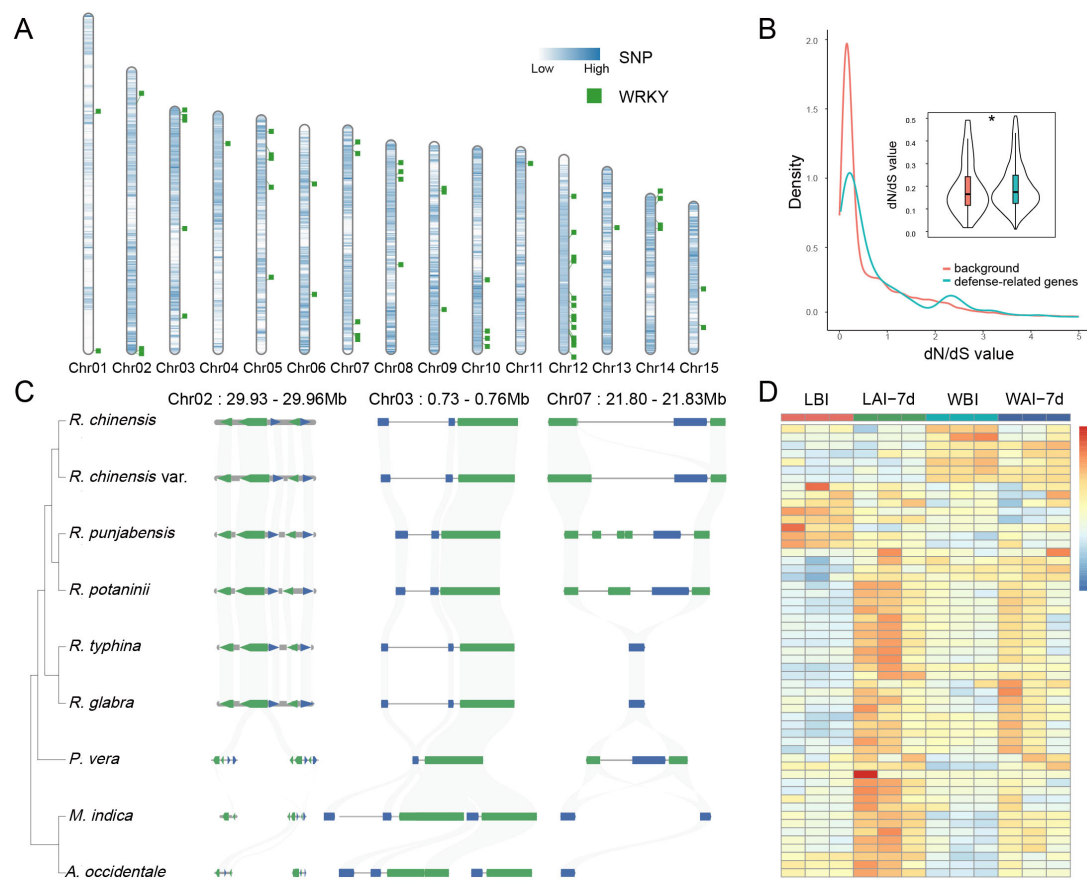
FIGURE 6
Genomic and expression features of WRKY-related defense genes in *R. chinensis*. **(A)** Chromosomal distribution of WRKY-related genes in the *R. chinensis* genome. Blue represents the density distribution of SNPs unique to Southeast Asian *Rhus* species. **(B)** Comparative assessment of evolutionary selection pressure (dN/dS ratios) between plant defense-associated genes and genome-wide background genes (*P < 0.05).
(C) Synteny of defense-related WRKY genes across nine Anacardiaceae species. **(D)** Expression levels of WRKY-related genes in four conditions: leaf-wing before infection (WBI), leaf-wing after 7 days of infection (WAI-7d), leaf before infection (LBI), and leaf after 7 days of infection (LAI-7d).

Miocene-Pliocene climatic shifts (~48 Mya) to the intercontinental disjunction of *Rhus* lineages. Notably, North American *R. typhina* and *R. glabra* exhibit morphological and genomic affinities to East Asian *R. potaninii* and *R. punjabensis* (Manchester and Judd, 2022), suggesting adaptive divergence during post-Eocene migrations. However, limited subspecies-level genomic data for *P. vera*, *M. indica*, and *A. occidentale* hinder precise reconstruction of their Tertiary biogeography, underscoring a need for expanded sampling.

Lineage-specific expansions of defense-related genes, including WRKY transcription factors and NLR receptors, highlight adaptive trade-offs within Anacardiaceae. In *Rhus*, chromosomal clustering of these genes (e.g., on chromosomes 4 and 12) mirrors defense "hotspots" observed in *Arabidopsis* and rice (Meyers et al., 2003; Singh et al., 2015), suggesting conserved mechanisms for co-regulation under stress. Synteny loss in North American *Rhus* WRKY loci contrasts with conservation in *P. vera*, implying geographic isolation reshaped regulatory architectures. Similarly, the TE-enriched genome of *P. vera*, which consists of 70.7% TEs,

shows an expansion of NB-ARC and cytochrome P450 families (McHale et al., 2006; Zeng et al., 2019), potentially contributing to enhanced abiotic stress tolerance. This pattern is not observed in the WGD-dependent genomes of *M. indica* and *A. occidentale*. These findings challenge the primacy of WGD in plant diversification, instead implicating TE-mediated plasticity as a driver of ecological resilience. However, reliance on computational annotations limits functional validation, as syntenic deviations in *R. typhina* or *R. glabra* could reflect pseudogenization or regulatory divergence, necessitating transcriptomic profiling.

Our study repositions TEs as central players in Anacardiaceae evolution, particularly during Pleistocene climatic upheavals. Recent LTR retrotransposon bursts in *R. punjabensis* (~1 Mya) and *R. potaninii* (~2 Mya) coincide with major environmental shifts (Caves et al., 2015; Ding et al., 2017), suggesting that fluctuating climates may have stimulated TE mobilization. These findings align with the 'genomic shock' hypothesis, where environmental upheaval triggers TE activation to drive adaptive innovation

(McClintock, 1984). Such TE dynamics mirror patterns observed in *Quercus* and *Pinus* (Plomion et al., 2018; Niu et al., 2022), where Pleistocene climatic oscillations contributed to genome expansion. In *P. vera*, rampant TE activity, which is absent in WGD-prone relatives, may have helped buffer against extinction during postglacial range fragmentation. Future studies should prioritize functional interrogation of TE impacts: CRISPR-edited *Rhus* lines could test whether LTR insertions near WRKY/NLR loci modulate stress responses. Additionally, comparative epigenomic analyses across Anacardiaceae may reveal how heterochromatin stability constrains or potentiates TE-driven adaptation (Slotkin and Martienssen, 2007), offering insights for breeding climate-resilient cultivars.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

BLF: Data curation, Software, Visualization, Writing – original draft. LHC: Writing – original draft, Data curation. LLC: Writing – review & editing.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2025.1638044/full#supplementary-material

References

Andreánszky, G. (1959). *Die flora der sarmatischen stufe in ungarn: die paläoökologischen und zönologischen beziehungen ihrer entwicklungsgeschichte* (Budapest: Akadémiai Kiadó).

Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data.* Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (Accessed March 1, 2025).

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556

Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME suite. *Nucleic Acids Res.* 43, W39–W49. doi: 10.1093/nar/gkv416

Baskin, J. M., and Baskin, C. C. (2022). Seed (true seed plus endocarp) dormancy in Anacardiaceae in relation to infrafamilial taxonomy and endocarp anatomy. *Seed Sci. Res.* 32, 187–199. doi: 10.1017/S096025852200023X

Bastian, M., Heymann, S., and Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. *Proc. Int. AAAI Conf. Web Soc. Media* 3, 361–362. doi: 10.1609/icwsm.v3i1.13937

Bennetzen, J. L. (2005). Transposable elements, gene creation and genome rearrangement in flowering plants. *Curr. Opin. Genet. Dev.* 15, 621–627. doi: 10.1016/j.gde.2005.09.010

Berens, M. L., Berry, H. M., Mine, A., Argueso, C. T., and Tsuda, K. (2017). Evolution of hormone signaling networks in plant defense. *Annu. Rev. Phytopathol.* 55, 401–425. doi: 10.1146/annurev-phyto-080516-035544

Beretta, G., Rossoni, G., Santagati, N. A., and Facino, R. M. (2009). Anti-ischemic activity and endothelium-dependent vasorelaxant effect of hydrolysable tannins from the leaves of *Rhus coriaria* (Sumac) in isolated rabbit heart and thoracic aorta. *Planta Med.* 75, 1482–1488. doi: 10.1055/s-0029-1185797

Blum, M., Andreeva, A., Florentino, L. C., Chuguransky, S. R., Grego, T., Hobbs, E., et al. (2024). InterPro: the protein sequence classification resource in 2025. *Nucleic Acids Res.* 53, D444–D456. doi: 10.1093/nar/gkae1082

Boufford, D. E., and Spongberg, S. A. (1983). Eastern Asian-Eastern North American phytogeographical relationships-a history from the time of linnaeus to the twentieth century. *Ann. Missouri Botanical Garden* 70, 423–439. doi: 10.2307/2992081

Burnham, R. J., and Graham, A. (1999). The history of neotropical vegetation: new developments and status. *Ann. Missouri Botanical Garden* 86, 546–589. doi: 10.2307/2666185

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552. doi: 10.1093/oxfordjournals.molbev.a026334

Caves, J. K., Winnick, M. J., Graham, S. A., Sjostrom, D. J., Mulch, A., and Chamberlain, C. P. (2015). Role of the westerlies in Central Asia climate over the Cenozoic. *Earth Planet. Sci. Lett.* 428, 33–43. doi: 10.1016/j.epsl.2015.07.023

Chen, H., Liu, J., Cui, K., Lu, Q., Wang, C., Wu, H., et al. (2018). Molecular mechanisms of tannin accumulation in *Rhus* galls and genes involved in plant-insect interactions. *Sci. Rep.* 8, 9841. doi: 10.1038/s41598-018-28153-y

Chen, C., Wu, Y., Li, J., Wang, X., Zeng, Z., Xu, J., et al. (2023). TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining. *Mol. Plant* 16, 1733–1742. doi: 10.1016/j.molp.2023.09.010

Chen, X., Yang, Z., Chen, H., Qi, Q., Liu, J., Wang, C., et al. (2020). A complex nutrient exchange between a gall-forming aphid and its plant host. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00811

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 6, 80–92. doi: 10.4161/fly.19695

Dahlgren, R. (1982). Angiosperm orders and families: an integrated system of classification of flowering plants. Arthur Cronquist. Columbia University Press, New York 1981. xviii 2162 pp., illus. $100. *Science* 216, 1217–1218. doi: 10.1126/science.216.4551.1217.b

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27, 1164–1165. doi: 10.1093/bioinformatics/btr088

Dendena, B., and Corsi, S. (2014). Cashew, from seed to market: a review. *Agron. Sustain. Dev.* 34, 753–772. doi: 10.1007/s13593-014-0240-7

Ding, L., Spicer, R. A., Yang, J., Xu, Q., Cai, F., Li, S., et al. (2017). Quantifying the rise of the Himalaya orogen and implications for the South Asian monsoon. *Geology* 45, 215–218. doi: 10.1130/G38583.1

Edgar, R. C. (2021). Muscle5: High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nat. Commun.* 13, 6968. doi: 10.1038/s41467-022-34630-w

Ediriweera, M. K., Tennekoon, K. H., Samarakoon, S.R.J.E.B.C., and Medicine, A. (2017). A review on ethnopharmacological applications, pharmacological activities, and bioactive compounds of Mangifera indica (Mango). *Evidence-Based Complement. Altern. Med.* 2017, 6949835. doi: 10.1155/2017/6949835

Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for *de novo* detection of LTR retrotransposons. *BMC Bioinf.* 9, 18. doi: 10.1186/1471-2105-9-18

Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20, 238. doi: 10.1186/s13059-019-1832-y

Eulgem, T., and Somssich, I. E. (2007). Networks of WRKY transcription factors in defense signaling. *Curr. Opin. Plant Biol.* 10, 366–371. doi: 10.1016/j.pbi.2007.04.020

Fedoroff, N. V. (2012). Transposable elements, epigenetics, and genome evolution. *Science* 338, 758–767. doi: 10.1126/science.338.6108.758

Goel, M., Sun, H., Jiao, W.-B., and Schneeberger, K. (2019). SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol.* 20, 277. doi: 10.1186/s13059-019-1911-0

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2011). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186. doi: 10.1093/nar/gkr944

Gray, A. (1857). Diagnostic characters of new species of phaenogamous plants collected in Japan by Charles Wright. With observations upon the relations of the Japanese flora to that of North America and of other parts of the northern temperate zone. *Memoirs Am. Acad. Arts Sci.* 6, 377–452.

Guo, Y.-L. (2013). Gene family evolution in green plants with emphasis on the origination and evolution of *Arabidopsis thaliana* genes. *Plant J.* 73, 941–951. doi: 10.1111/tpj.12089

Herrera, F., Mitchell, J. D., Pell, S. K., Collinson, M. E., Daly, D. C., and Manchester, S. R. (2018). Fruit morphology and anatomy of the spondioid Anacardiaceae. *Bot. Rev.* 84, 315–393. doi: 10.1007/s12229-018-9201-1

Jacob, F., Vernaldi, S., and Maekawa, T. (2013). Evolution and conservation of plant NLR functions. *Front. Immunol.* 4. doi: 10.3389/fimmu.2013.00297

Javed, T., and Gao, S.-J. (2023). WRKY transcription factors in plant defense. *Trends Genet.* 39, 787–801. doi: 10.1016/j.tig.2023.07.001

Jiang, Y., Gao, M., Meng, Y., Wen, J., Ge, X.-J., and Nie, Z.-L. (2019). The importance of the North Atlantic land bridges and eastern Asia in the post-Boreotropical biogeography of the Northern Hemisphere as revealed from the poison ivy genus (*Toxicodendron*, Anacardiaceae). *Mol. Phylogenet. Evol.* 139, 106561. doi: 10.1016/j.ympev.2019.106561

Jiang, J., Ma, S., Ye, N., Jiang, M., Cao, J., and Zhang, J. (2017). WRKY transcription factors in plant responses to stresses. *J. Integr. Plant Biol.* 59, 86–101. doi: 10.1111/jipb.12513

Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi: 10.1038/nmeth.3317

Koichiro, T., Glen, S., and Sudhir, K. (2021). MEGA11 molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/molbev/msab120

Krueger, F., James, F., Ewels, P., Afyounian, E., and Schuster-Boeckler, B. (2021). *FelixKrueger/TrimGalore* (Geneva, Switzerland: Zenodo). doi: 10.5281/zenodo.5127898

Kumanotani, J. (1995). Urushi (oriental lacquer)—a natural aesthetic durable and future-promising coating. *Prog. Organic Coatings* 26, 163–195. doi: 10.1016/0300-9440(95)00559-5

Kumar, S., Stecher, G., Suleski, M., and Hedges, S. B. (2017). TimeTree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* 34, 1812–1819. doi: 10.1093/molbev/msx116

Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., et al. (2011). The *Arabidopsis* information resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 40, D1202–D1210. doi: 10.1093/nar/gkr1090

Lindley, J. (2015). *An introduction to the natural system of botany: or, a systematic view of the organisation, natural affinities, and geographical distribution, of the whole vegetable kingdom* (Cambridge: Cambridge University Press).

Lisch, D. (2013). How important are transposons for plant evolution? *Nat. Rev. Genet.* 14, 49–61. doi: 10.1038/nrg3374

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. doi: 10.1186/s13059-014-0550-8

Ma, X.-M., Lu, R., and Miyakoshi, T. J. A. I. (2012). Recent advances in research on lacquer allergy. *Allergol. International: Off. J. Japanese Soc. Allergol.* 61, 45–50. doi: 10.2332/allergolint.11-RA-0324

Manchester, S. R., and Judd, W. S. (2022). Extinct anacardiaceous samaras and sumac-like leaves from the Eocene of Western North America. *Int. J. Plant Sci.* 183, 357–366. doi: 10.1086/719948

Mango Genome Consortium, C., Bally, I. S. E., Bombarely, A., Chambers, A. H., Cohen, Y., Dillon, N. L., et al. (2021). The 'Tommy Atkins' mango genome reveals candidate genes for fruit quality. *BMC Plant Biol.* 21, 108. doi: 10.1186/s12870-021-02858-1

Marçais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L., and Zimin, A. (2018). MUMmer4: A fast and versatile genome alignment system. *PloS Comput. Biol.* 14, e1005944. doi: 10.1371/journal.pcbi.1005944

Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., et al. (2014). CDD: NCBI's conserved domain database. *Nucleic Acids Res.* 43, D222–D226. doi: 10.1093/nar/gku1221

Marino, G., Di Martino, S., Roxas, A. A., Caruso, T., Ferguson, L., Barone, E., et al. (2018). Sustainability of pistachio production (*Pistacia vera* L.) under supplemental irrigation in a Mediterranean climate. *Scientia Hortic.* 241, 260–266. doi: 10.1016/j.scienta.2018.06.032

Matić, S., Stanić, S., Mihailović, M., and Bogojević, D. (2016). *Cotinus coggygria* Scop.: An overview of its chemical constituents, pharmacological and toxicological potential. *Saudi J. Biol. Sci.* 23, 452–461. doi: 10.1016/j.sjbs.2015.05.012

McClintock, B. (1984). The significance of responses of the genome to challenge. *Science* 226, 792–801. doi: 10.1126/science.15739260

McHale, L., Tan, X., Koehl, P., and Michelmore, R. W. (2006). Plant NBS-LRR proteins: adaptable guards. *Genome Biol.* 7, 212. doi: 10.1186/gb-2006-7-4-212

Mendes, F. K., Vanderpool, D., Fulton, B., and Hahn, M. W. (2021). CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* 36, 5516–5518. doi: 10.1093/bioinformatics/btaa1022

Meyers, B. C., Kozik, A., Griego, A., Kuang, H., and Michelmore, R. W. (2003). Genome-wide analysis of NBS-LRR–encoding genes in *Arabidopsis*. *Plant Cell* 15, 809–834. doi: 10.1105/tpc.009308

Montes, C., Cardona, A., Jaramillo, C., Pardo, A., Silva, J. C., Valencia, V., et al. (2015). Middle miocene closure of the central American seaway. *Science* 348, 226–229. doi: 10.1126/science.aaa2815

Moran, N. A. (1989). A 48-million-year-old aphid—host plant association and complex life cycle: biogeographic evidence. *Science* 245, 173–175. doi: 10.1126/science.245.4914.173

Morley, R. J. (2001). *Why are there so many primitive angiosperms in the rain forests of Asia-Australasia*. Eds. I. Metclfe, J. M. B. Smith, M. Morwood and I. Davidson. Lisse, Netherlands: Swets & Zeitlinger b.v. 185–199.

Morley, R. J. (2007). "Cretaceous and Tertiary climate change and the past distribution of megathermal rainforests," in *Tropical Rainforest Responses to Climatic Change*. Eds. M. B. Bush and J. R. Flenley (Springer Berlin Heidelberg, Berlin, Heidelberg), 1–31.

Müller-Wille, S., and Reeds, K. (2007). A translation of Carl Linnaeus's introduction to Genera plantarum, (1737). *Stud. History Philosophy Sci. Part C: Stud. History Philosophy Biol. Biomed. Sci.* 38, 563–572. doi: 10.1016/j.shpsc.2007.06.003

Nei, M., and Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* 3, 418–426. doi: 10.1093/oxfordjournals.molbev.a040410

Ni, B.-b., Liu, H., Wang, Z.-S., Zhang, G.-Y., Sang, Z.-Y., Liu, J.-J., et al. (2024). A chromosome-scale genome of *Rhus chinensis* Mill. provides new insights into plant–insect interaction and gallotannins biosynthesis. *Plant J.* 118, 766–786. doi: 10.1111/tpj.16631

Niu, S., Li, J., Bo, W., Yang, W., Zuccolo, A., Giacomello, S., et al. (2022). The Chinese pine genome and methylome unveil key features of conifer evolution. *Cell* 185, 204–217.e214. doi: 10.1016/j.cell.2021.12.006

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 27, 29–34. doi: 10.1093/nar/27.1.29

Ossowski, S., Schneeberger, K., Lucas-Lledó, J. I., Warthmann, N., Clark, R. M., Shaw, R. G., et al. (2010). The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science* 327, 92–94. doi: 10.1126/science.1180677

Ou, S., and Jiang, N. (2018). LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* 176, 1410–1422. doi: 10.1104/pp.17.01310

Ou, S., and Jiang, N. (2019). LTR_FINDER_parallel: parallelization of LTR_FINDER enabling rapid identification of long terminal repeat retrotransposons. *Mob DNA* 10, 48. doi: 10.1186/s13100-019-0193-0

Palamarev, E. (1989). Paleobotanical evidences of the Tertiary history and origin of the Mediterranean sclerophyll dendroflora. *Pl Syst. Evol.* 162, 93–107. doi: 10.1007/BF00936912

Paterniti, I., Impellizzeri, D., Cordaro, M., Siracusa, R., Bisignano, C., Gugliandolo, E., et al. (2017). The anti-inflammatory and antioxidant potential of pistachios (*Pistacia vera* L.) *in vitro* and in vivo. *Nutrients* 9, 915. doi: 10.3390/nu9080915

Pell, S. K. (2004). *Molecular systematics of the cashew family (Anacardiaceae)* Vol. 1472 (Baton Rouge, LA: LSU Doctoral Dissertations). Available online at: https://repository.lsu.edu/gradschool_dissertations/1472 (Accessed March 9, 2025).

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. doi: 10.1038/nbt.3122

Plomion, C., Aury, J.-M., Amselem, J., Leroy, T., Murat, F., Duplessis, S., et al. (2018). Oak genome reveals facets of long lifespan. *Nat. Plants* 4, 440–452. doi: 10.1038/s41477-018-0172-3

R Core Team (2014). *R: A language and environment for statistical computing*. Available online at: https://www.R-project.org/ (Accessed September 1, 2021).

Ren, Y. Y., Zhang, X. R., Li, T. N., Zeng, Y. J., Wang, J., and Huang, Q. W. (2021). *Galla Chinensis*, a traditional Chinese medicine: comprehensive review of botany, traditional uses, chemical composition, pharmacology and toxicology. *J. Ethnopharmacol.* 278, 114247. doi: 10.1016/j.jep.2021.114247

Savadi, S., Muralidhara, B. M., Godwin, J., Adiga, J. D., Mohana, G. S., Eradasappa, E., et al. (2022). *De novo* assembly and characterization of the draft genome of the cashew (*Anacardium occidentale* L.). *Sci. Rep.* 12, 18187. doi: 10.1038/s41598-022-22600-7

Singh, S., Chand, S., Singh, N. K., and Sharma, T. R. (2015). Genome-wide distribution, organisation and functional characterization of disease resistance and defence response genes across rice species. *PloS One* 10, e0125964. doi: 10.1371/journal.pone.0125964

Slotkin, R. K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* 8, 272–285. doi: 10.1038/nrg2072

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Sun, P., Jiao, B., Yang, Y., Shan, L., Li, T., Li, X., et al. (2022). WGDI: A user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. *Mol. Plant* 15, 1841–1851. doi: 10.1016/j.molp.2022.10.018

Suzuki, M., Noshiro, S., Tanaka, T., Kobayashi, K., Wang, Y., Liu, J., et al. (2014). Origin of urushi (*Toxicodendron vernicifluum*) in the neolithic jomon period of Japan. *Natl. Museum Japanese History Repository* 187, 49–70.

Suzuki, M., Yonekura, K., and Noshiro, S. (2007). Distribution and habitat of *Toxicodendron vernicifluum* (stokes) FA Barkl.(Anacardiaceae) in China. *Japanese J. Historical Bot.* 15, 58–62. doi: 10.34596/hisbot.15.1_58

Tang, H., Bowers, J. E., Wang, X., Ming, R., Alam, M., and Paterson, A. H. (2008). Synteny and collinearity in plant genomes. *Science* 320, 486–488. doi: 10.1126/science.1153917

Tiffney, B. H. (1985). Perspectives on the origin of the floristic similarity between Eastern Asia and Eastern North America. *J. Arnold Arboretum* 66, 73–94. doi: 10.5962/bhl.part.13179

Verma, V., Ravindran, P., and Kumar, P. P. (2016). Plant hormone-mediated regulation of stress responses. *BMC Plant Biol.* 16, 86. doi: 10.1186/s12870-016-0771-y

Vos, I. A., Pieterse, C. M. J., and van Wees, S. C. M. (2013). Costs and benefits of hormone-regulated plant defences. *Plant Pathol.* 62, 43–55. doi: 10.1111/ppa.12105

Wang, L., He, N., Li, Y., Fang, Y., and Zhang, F. (2020b). Complete chloroplast genome sequence of Chinese lacquer tree (*Toxicodendron vernicifluum*, Anacardiaceae) and its phylogenetic significance. *BioMed. Res. Int.* 2020, 9014873. doi: 10.1155/2020/9014873

Wang, C., Liu, P., Chen, X., Liu, J., Lu, Q., Shao, S., et al. (2020a). Microenvironmental analysis of two alternating hosts and their impact on the ecological adaptation of the horned sumac gall aphid *Schlechtendalia chinensis* (Hemiptera, Pemphiginae). *Sci. Rep.* 10, 435. doi: 10.1038/s41598-019-57138-8

Wang, P., Luo, Y., Huang, J., Gao, S., Zhu, G., Dang, Z., et al. (2020c). The genome evolution and domestication of tropical fruit mango. *Genome Biol.* 21, 60. doi: 10.1186/s13059-020-01959-8

Wani, S. H., Anand, S., Singh, B., Bohra, A., and Joshi, R. (2021). WRKY transcription factors and plant defense responses: latest discoveries and future prospects. *Plant Cell Rep.* 40, 1071–1085. doi: 10.1007/s00299-021-02691-8

Wannan, B. S. (2006). Analysis of generic relationships in Anacardiaceae. *Blumea: Biodivers. Evol. Biogeogr. Plants* 51, 165–195. doi: 10.3767/000651906X622427

Wei, H. Y., Ye, Y. X., Huang, H. J., Chen, M. S., Yang, Z. X., Chen, X. M., et al. (2022). Chromosome-level genome assembly for the horned-gall aphid provides insights into interactions between gall-making insect and its host plant. *Ecol. Evol.* 12, e8815. doi: 10.1002/ece3.8815

Wen, J. (1999). Evolution of Eastern Asian and Eastern North American disjunct distributions in flowering plants. *Annu. Rev. Ecol. Syst.* 30, 421–455. doi: 10.1146/annurev.ecolsys.30.1.421

Wersch, S., and Li, X. (2019). Stronger when together: clustering of plant NLR disease resistance genes. *Trends Plant Sci.* 24, 688–699. doi: 10.1016/j.tplants.2019.05.005

Wilf, P. (2000). Late Paleocene–early Eocene climate changes in southwestern wyoming: paleobotanical analysis. *GSA Bull.* 112, 292–307. doi: 10.1130/0016-7606(2000)112<0292:LPEECC>2.3.CO;2

Wolfe, J. A. (1966). *Tertiary plants from the cook inlet region* (Washington, D.C.: Alaska.U.S. Government Printing Office). doi: 10.3133/pp398B

Wolfe, J. A. (1975). Some aspects of plant geography of the northern hemisphere during the late cretaceous and tertiary. *Ann. Missouri Botanical Garden* 62, 264–279. doi: 10.2307/2395198

Xiang, J. (2001). Dispersal-vicariance analyses of intercontinental disjuncts: historical biogeographical implications for angiosperms in the northern hemisphere. *S29 Int. J. Plant Sci.* 162, 29–39. doi: 10.1086/323332

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088

Yang, R.-C., Peng, F. Y., and Hu, Z. (2017). Inferring defense-related gene families in *Arabidopsis* and wheat. *BMC Genomics* 18, 980. doi: 10.1186/s12864-017-4381-3

Yi, T., Miller, A. J., and Wen, J. (2004). Phylogenetic and biogeographic diversification of *Rhus* (Anacardiaceae) in the northern hemisphere. *Mol. Phylogenet. Evol.* 33, 861–879. doi: 10.1016/j.ympev.2004.07.006

Yu, G. (2024). Thirteen years of clusterProfiler. *Innovation* 5, 100722. doi: 10.1016/j.xinn.2024.100722

Zeng, L., Tu, X. L., Dai, H., Han, F. M., Lu, B. S., Wang, M. S., et al. (2019). Whole genomes and transcriptomes reveal adaptation and domestication of pistachio. *Genome Biol.* 20, 79. doi: 10.1186/s13059-019-1686-3

Zhang, B., Wang, L., Luo, L., and King, M. W. (2014). Natural dye extracted from Chinese gall – the application of color and antibacterial activity to wool fabric. *J. Clean. Prod.* 80, 204–210. doi: 10.1016/j.jclepro.2014.05.100

Zhao, M., Liu, C., Zheng, G., Wei, S., and Hu, Z. (2013). Comparative studies of bark structure, lacquer yield and urushiol content of cultivated *Toxicodendron vernicifluum* varieties. *New Z. J. Bot.* 51, 13–21. doi: 10.1080/0028825X.2012.731005