The chromosome-level genome of *Elsholtzia splendens* provides insights into genome evolution and environmental adaptation

Licao Cui
Jiangxi Agricultural University
Yiting Su
Jiangxi Agricultural University
Han Lin
Jiangxi Agricultural University
Xiaofeng Yang
Jiangxi Agricultural University
Yu Lu
Jiangxi Agricultural University
Yuzhe Song
Jiangxi Agricultural University
Yihan Li
liyihan@jxau.edu.cn
Jiangxi Agricultural University
Qinglin Ke
Jiangxi Agricultural University
Ruimin Li
Gannan Normal University

Research Article

Keywords: Elsholtzia splendens, chromosome-scale genome assembly, adaptive evolution, copper stress, HMA gene family

Posted Date: July 7th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10177130/v1

Additional Declarations: No competing interests reported.

Abstract

Novel efficient phytoremediation approaches are urgently required to mitigate global soil copper (Cu) pollution, which severely damages ecosystem stability and agricultural yields. *Elsholtzia splendens* (copper grass) ranks among the most efficient Cu-hyperaccumulating plants known, yet the genomic underpinnings of its adaptive evolution have remained largely unresolved. We constructed a chromosome-scale, high-continuity genome assembly for this species by combining PacBio HiFi long-read sequencing with Hi-C chromatin conformation capture. Phylogenomic placement positioned *E. splendens* as sister to *Salvia* and *Origanum*, with a divergence time of roughly 30.11 million years ago (Mya). Beyond marked expansion of gene families linked to stress resistance and secondary metabolism, genome-wide scans uncovered 93 positively selected genes, among them heavy metal transporters and regulators of redox balance. Recent bursts of long terminal repeat retrotransposon (LTR-RT) proliferation constituted a major driver of these genomic innovations. Transcriptome profiling across copper concentrations exposed organ-specific regulatory logic—leaves prioritized oxidative detoxification, whereas stems were specialized in metal sequestration and long-distance translocation. Within this framework, 16 Heavy Metal ATPase genes (EsHMAs), most belonging to the Cu/Ag subfamily, displayed substantial haplotype diversity and expression levels positively correlated with copper treatment concentrations; functional inference points to distinct paralogs governing xylem loading, vacuolar sequestration in roots, and metal efflux. Together, these findings reveal how transposon bursts, gene family expansion, and HMA-dominated regulatory networks jointly shape the extreme Cu hypertolerance of this species, offering concrete genetic targets for phytoremediation efforts.

1 Introduction

Rapid industrialization coupled with intensive agriculture has made soil copper (Cu) contamination a pervasive global problem [1, 2]. Cu is an essential micronutrient for plant growth; however, Cu concentrations exceeding physiological tolerance thresholds—commonly detected in mining and industrial regions—induce severe oxidative stress, disrupts redox balance, and damages nucleic acids and proteins, with photosynthesis and overall growth suffering as a result [3, 4]. Conventional physical and chemical remediation techniques suffer notable drawbacks: high cost, risk of secondary pollution, and disruption of soil microbial communities, which limit their large-scale field application [5]. Metallophyte-based phytoremediation avoids most of these drawbacks, combining low cost with minimal ecological disruption, which has made it an increasingly active frontier for restoring metal-contaminated land [6, 7].

Among metallophytes, *Elsholtzia splendens* (copper grass)—a perennial Lamiaceae herb native to temperate East Asia—stands out as a defining indicator species of copper-rich habitats [8, 9], having acquired pronounced Cu tolerance and aboveground hyperaccumulation through prolonged exposure to metalliferous soils [10–12]. A previously released chromosome-level genome of *E. splendens* was limited to basic structural assembly, lacking systematic exploration of the genetic basis underlying adaptive

evolution and Cu hypertolerance [13]. Consequently, the genomic architecture underlying this adaptation—its repeat element dynamics, lineage-specific shifts in gene families during adaptive divergence, and the regulatory networks that govern its copper stress response—remains largely undescribed.

Among the molecular players known to govern metal handling, Heavy Metal ATPases (HMAs) are P1B-type ATPase superfamily members that use ATP hydrolysis to move metal ions across membranes, thereby maintaining ion homeostasis and enabling detoxification [14]. Substrate preference splits the family into two clades: a Cu/Ag subfamily that governs copper allocation and chloroplast function, and a Zn/Co/Cd/Pb subfamily responsible for long-range metal transport through xylem loading [15]. Work in *Arabidopsis thaliana* has mapped out how individual members divide this labor. *AtHMA1* caps zinc accumulation when supply is excessive, while *AtHMA2* and *AtHMA4* drive Zn and Cd movement from root to shoot via the xylem [16, 17]; *AtHMA3*, by contrast, sequesters cadmium in root vacuoles to confer tolerance [18]. Within the Cu/Ag clade, *AtHMA5* handles copper loading into the xylem, *AtHMA6* and *AtHMA8* operate in chloroplasts to sustain photosynthesis and antioxidant capacity, and *AtHMA7* delivers copper to the endoplasmic reticulum to support ethylene receptor synthesis [19–22]. HMA gene families have since been catalogued across a range of crops and ornamentals, such as *Triticum aestivum* [23], *Hydrangea macrophylla* [24], *Medicago truncatula* [25], nevertheless, a systematic genome-wide investigation of HMA genes remains absent in *E. splendens*, a model Cu hyperaccumulator. In other metallophytes, gene duplication and transcriptional upregulation of HMA genes are tightly linked to hyperaccumulation traits. However, it remains unclear whether this evolutionary pattern applies to *E. splendens*, nor have we characterized the evolutionary origins, duplication modes, population haplotype diversity, and tissue-specific expression patterns of EsHMA genes.

Physiological characterization and candidate-gene studies have accumulated for *E. splendens*, yet they fall short of a systematic account of its molecular architecture, leaving this species behind other well-studied metallophytes in terms of genomic insight into adaptive evolution and regulation. We close this gap by assembling a high-quality *E. splendens* genome from PacBio HiFi long reads anchored with Hi-C chromatin contact data. Multi-omic analyses, paired with RNA-seq profiling across a copper-stress gradient, allowed us to dissect genome organization, characterize repeat dynamics, and trace amplification patterns of long terminal repeat retrotransposons (LTR-RTs). Comparative genomics further clarified the species' phylogenetic position, gene family evolution, and signatures of positive selection, with particular emphasis on the HMA family—its evolutionary trajectory, modes of duplication, and tissue-resolved expression. Taken together, this work moves past simple genome assembly to deliver an integrated picture of the genomic innovations and regulatory networks that underlie extreme copper tolerance in *E. splendens*, advancing our mechanistic understanding of heavy metal adaptation in Lamiaceae and providing candidate genes for breeding metal-tolerant crop varieties.

2 Materials and methods

2.1 Plant Material and Genome Survey

We collected *E. splendens* seeds from the Tongling Copper Mine Ruins Museum, Jiangxi Province, China (29.756227° N, 115.558771° E). After surface sterilization, seeds were germinated and raised in a climate chamber set to a 16/8 h light/dark cycle, 22–23°C by day and 18–20°C by night, with relative humidity held at 40–60%. Genomic DNA, extracted from roughly 100 mg of fresh young leaves via the CTAB method, was checked for quality and concentration using a Qubit fluorometer and 1% agarose gel electrophoresis. Qualified DNA was sheared to 200–400 bp on a Covaris ultrasonicator, then carried through end-repair, A-tailing, adapter ligation, and PCR amplification before sequencing as paired-end libraries on the MGISEQ-2000 platform. Raw reads were filtered using fastp v0.21.0. We then performed k-mer counting with Jellyfish v2.3.0 and estimated genome size, heterozygosity and repeat content via GenomeScope v2.0.

2.2 Nucleic Acid Extraction, Library Construction, Sequencing, and Genome Assembly

A SMRTbell Express Template Prep Kit 3.0 (PacBio, USA) was used to construct a single-molecule real-time (SMRT) sequencing library, which was sequenced on a PacBio Revio platform to generate high-fidelity (HiFi) reads. For chromosome-level scaffolding, a Hi-C library was constructed following the manufacturer's protocol, involving formaldehyde crosslinking, restriction enzyme digestion, biotin labeling, and proximity ligation. The Hi-C library was sequenced on an Illumina NovaSeq 6000 platform.

Following quality control with fastp v0.21.0, HiFi reads were assembled de novo in Hifiasm v0.21.0-r686. Hi-C reads were then mapped onto the resulting contigs with Juicer v1.2.2, and the 3D-DNA v201008 pipeline ordered and oriented these contigs into scaffolds; manual curation in Juicebox v2.17 confirmed assembly accuracy at the chromosome scale. Genome completeness was finally benchmarked against the embryophyta_odb10 dataset using BUSCO v6.0.0.

2.3 Transcriptome Sampling, RNA Extraction, and Sequencing

Genome annotation and copper-response profiling drew on tissue collected from roots, young stems, mature stems, leaves, and flowers of *E. splendens*. Separately, three-week-old seedlings were grown hydroponically for seven days under three conditions: 100 μM $CuSO_4 \cdot 5H_2O$ (Cu100), 500 μM $CuSO_4 \cdot 5H_2O$ (Cu500), or distilled water as a control (CK), after which stem and leaf tissue was harvested, snap-frozen in liquid nitrogen, and stored at − 80°C.

Total RNA, extracted with the RNeasy Plant Mini Kit (Qiagen, Germany), met quality thresholds of RIN ≥ 8.0 and $OD_{260}/_{280}$ ≈ 1.8–2.0, as verified on an Agilent 2100 Bioanalyzer and NanoDrop spectrophotometer. Strand-specific cDNA libraries, built with the TruSeq Stranded mRNA Library Prep Kit (Illumina, USA), were sequenced on a NovaSeq 6000 instrument (150 bp paired-end).

2.4 Repeat Annotation, Gene Prediction, and Functional Annotation

Repeat annotation combined homology-based and de novo strategies. Krait v1.4.0 and TRF v4.09 located simple sequence repeats (SSRs) and tandem repeats; MITE-Tracker v1.0 identified miniature inverted-repeat transposable elements (MITEs); and LTR-RT candidates, first flagged by LTR_FINDER v1.07 and LTRharvest v1.6.2, were subsequently refined with LTR_retriever v2.9.0. A consensus repeat library was generated by integrating *de novo* predictions from RepeatModeler v2.0.2a with curated entries from Dfam v3.5 and RepBase v20181026. This library was used to annotate the genome via RepeatMasker v4.1.2, and the resulting elements were classified using TEsorter v1.4.6.

LTR-RT insertion ages were calculated from the Kimura two-parameter distance (K) between aligned 5' and 3' LTRs (MAFFT v7.490), using $T = K/(2r)$ with a substitution rate r of 7×10^{-9} per site per year.

Non-coding RNAs were classified by type and tool: tRNAscan-SE v1.23 for tRNAs; RNAmmer v1.2 together with Barrnap v0.8 for rRNAs; and Infernal v1.1.4, searched against Rfam v14.7, for the remaining ncRNA classes.

Protein-coding gene prediction integrated three independent lines of evidence. Transcript-based support came from RNA-seq reads aligned with HISAT2 v2.2.1 and assembled in StringTie v2.2.1; homology-based support came from ProtHint v2.6.0 and Spaln v2.4.13f, run against protein sets from related species (*Perilla frutescens*, *Salvia hispanica*, *Sal. miltiorrhiza*, *Sal. rosmarinus*, *Sal. splendens*); and de novo predictions came from BRAKER3 v3.0.3 and Augustus v3.4.0. TSEBRA v1.1.1 merged these into a single non-redundant gene set, after which GUSHR v1.0.0 and GeMoMa v1.6.2 refined UTRs using full-length transcript evidence.

Functional annotation searched the NR, Swiss-Prot, and TrEMBL databases with DIAMOND v2.0.15. Protein domains came from Pfam v35.0; GO terms, COG/eukaryotic orthologous group (KOG) assignments, and KEGG pathways came from eggNOG-mapper v2.1.0 (eggNOG v5.0) and the KEGG database (Release 100.0). We assessed gene set completeness with BUSCO v5.2.2 and cross-checked prediction accuracy against transcriptomic support, requiring FPKM (fragments per kilobase of transcript per million mapped reads) > 1.

2.5 Gene Family Clustering, Phylogenetic Reconstruction, and Expansion/Contraction Analysis

OrthoFinder v2.5.4, using DIAMOND alignment (E-value ≤ 1×10^{-5}), grouped genes into orthologous families. From these, 201 single-copy orthologs were aligned in MAFFT v7.490 and trimmed with Gblocks v0.91b before serving as input for a maximum-likelihood (ML) tree in IQ-TREE v2.3.6, built under the ModelFinder-selected JTT + F+I+G4 model with 1,000 bootstrap replicates and rooted on *Vitis vinifera* and *A. thaliana*. Divergence times were then dated in MCMCTree (PAML v4.9i) under a correlated molecular clock, calibrated with a TimeTree-derived constraint of 142.1–163.5 Mya for the *Sesamum indicum–A. thaliana* split.

CAFE v4.2.1, applying a birth–death model, flagged gene families as significantly expanded or contracted when both family-wide and Viterbi *P*-values fell below 0.05.

2.6 Positive Selection Analysis

We screened for positive selection using CodeML in PAML v4.9i. Protein alignments (MAFFT v7.490) were converted to codon alignments with PAL2NAL v14, and branch-site models were fitted under the F3×4 codon frequency scheme. A likelihood ratio test (LRT, $P < 0.05$) distinguished selection from the null model, and Bayes empirical Bayes (BEB) analysis (posterior probability > 0.95) pinpointed the specific sites under selection.

2.7 Genome Collinearity and Whole-Genome Duplication (WGD) Analysis

MCScanX v1.0.0 mapped genome-wide collinearity. Homologous gene pairs, identified through DIAMOND v2.0.5.143 and JCVI v0.9.14, were sorted by synteny topology and physical spacing into WGD/segmental, tandem, proximal, or dispersed (transposon-mediated) duplication categories.

WGDI v0.71 then validated and dated candidate WGD events. We computed synonymous substitution rates (Ks) for syntenic paralog pairs under the Nei–Gojobori (NG86) model, excluding tandem duplicates spaced fewer than 10 genes apart to limit bias, and dated WGD timing from the dominant peak of a Gaussian-mixture fit to the Ks distribution. JCVI v0.9.14 also produced the collinearity plots.

2.8 Transcriptome Profiling and Weighted Gene Co-expression Network Analysis (WGCNA)

After aligning clean reads to the reference genome with HISAT2 v2.2.1, we quantified expression as FPKM and transcripts per million (TPM) via StringTie v2.2.1, and called differential expression in DESeq2 v1.38.3 (adjusted $P < 0.05$, $|\log_2 FC| > 1$).

A co-expression network built in WGCNA v1.74 (R) used a soft-thresholding power chosen for scale-free topology fit ($R^2 > 0.9$), with modules defined by dynamic tree cutting (minimum size = 30 genes). clusterProfiler v4.6.0 tested each module for GO and KEGG enrichment (FDR < 0.05), and we designated the ten genes with the highest intramodular connectivity in each module as hub genes.

2.9 Identification of the HMA Gene Family

To find *E. splendens* HMA genes (*EsHMA*), we used *A. thaliana* and rice (*Oryza sativa*) sequences as BLASTP v2.14.1 + queries with default parameters, then candidate proteins were further validated by the simultaneous presence of the E1-E2 ATPase domain (PF00122) and heavy-metal-associated domain (PF00403) using HMMER searches against the Pfam database. Confirmed genes were numbered *EsHMA1* through *EsHMA16* in order of chromosomal position.

2.10 Phylogenetic, Haplotype, and Gene Duplication Analysis of HMA Genes

HMA phylogenetic relationships were inferred by neighbor-joining (Poisson correction, 500 bootstraps) in ClustalW v2.1 and MEGA-X v10.1.7. Haplotype structure was characterized in DnaSP v6.12.03 and rendered as a network in PopART v1.7. Duplication events were detected via self-BLAST of *E. splendens* protein sequences, with MCScanX subsequently classifying duplication type based on the resulting syntenic blocks.

3 Results

3.1 Genome Sequencing and Chromosome-Scale Assembly

Genome profiling estimated a haploid genome size of approximately 297.74 Mb for *E. splendens*, with a heterozygosity rate of 1.24% and repetitive sequence content of 55.77% (Tables S1-S2, Fig. S1). To generate a chromosome-level reference genome, we integrated HiFi long-read sequencing with Hi-C chromatin conformation capture. A total of 34.16 Gb of high-quality long reads and 31.78 Gb of effective Hi-C data were obtained, with uniquely mapped paired reads accounting for 36.03%. This sequencing depth and quality provided a robust foundation for assembly and chromosomal scaffolding (Tables S3-S5).

De novo assembly yielded a contiguous draft genome comprising 237 contigs, with a contig N50 of 14.11 Mb and a maximum contig length of 35.08 Mb (Table S6). Subsequent integration of Hi-C contact data enabled ordering, orientation, and anchoring of contigs, resulting in 92.01% of the assembled sequences being assigned to eight chromosomes (Table S7). The final assembled genome spanned 255.39 Mb, representing 85.78% of the estimated genome size (Table S8). Hi-C heatmaps revealed strong intra-chromosomal interaction signals and minimal inter-chromosomal noise, with no detectable misassemblies, inversions, or translocations, confirming high structural integrity (Fig. S2). The LTR Assembly Index (LAI) reached an average value of 21.16, further verifying that this assembly meets reference-grade continuity and completeness criteria (Fig. S3).

Genome completeness assessed using BUSCO indicated that 98.07% of conserved land plant genes were fully recovered, including 93.55% as single copies and 4.51% as duplicates. Only 0.13% of BUSCOs were fragmented and 1.81% were missing, which demonstrates that this high-quality genome sufficiently supports subsequent functional annotation and comparative genomic analyses (Fig. S4, Table S9).

3.2 Repeat Annotation and Non-Coding RNA Identification

Repetitive elements accounted for 44.54% (113.75 Mb) of the assembled genome, which is broadly consistent with the k-mer-based estimation, and the moderate numerical discrepancy arises from

differences between short-read genome survey data and the final polished genome assembly (Table S10). Unknown repeats (19.64%) and LTR-RTs (16.61%) constituted the majority, followed by DNA transposons (5.54%), other repeats (3.77%), LINEs (0.88%), and SINEs (0.28%). This repeat composition conforms to the typical expansion pattern of plant genomes driven by retrotransposon proliferation. Compared with homology-based prediction alone, ab initio prediction greatly improved the identification of species-specific repetitive sequences.

A total of 38,685 SSR loci spanning 828,195 bp were identified (Tables S11–S12). The SSRs exhibited a strong AT bias, with TA/TA (19.82%) and AT/AT (19.81%) as the most abundant motifs, followed by AG/CT (6.14%), GA/TC (6.11%), and AAT/ATT (5.00%). GC-rich motifs such as CCG/CGG were notably scarce, consistent with the overall low genomic GC content (Table S13). These abundant and well-characterized SSRs provide valuable resources for molecular marker development and genetic diversity studies.

Non-coding RNA annotation identified 10,284 ncRNA elements totaling 11.02 Mb, occupying 4.30% of the genome (Table S14). Ribosomal RNAs predominated, with 28S (1,976), 18S (1,942), 5S (3,000), and 5.8S (883) rRNAs collectively representing 97.87% of the total ncRNA sequence. Additionally, 1,253 tRNAs, 99 miRNAs, and various snoRNAs and spliceosomal RNAs were annotated. This comprehensive ncRNA dataset lays a genomic foundation for exploring stress responses and secondary metabolic pathways in *E. splendens*.

3.3 Gene Prediction and Functional Annotation

We predicted 34,335 protein-coding genes using a combined strategy of homology-based prediction, *ab initio* modeling, and transcriptome evidence (Tables S15-S17). The average gene length was 3,406.60 bp, with 4.45 exons per gene and mean exon and intron lengths of 256.52 bp and 306.73 bp, respectively. These gene structural characteristics are comparable to those of other closely related Lamiaceae species (Table S18).

Functional annotation against multiple databases successfully assigned putative functions to 28,461 genes (82.89% annotation rate; Table S19, Fig. S5). High annotation coverage was achieved for NR (82.84%, 28,443 genes), TrEMBL (81.06%, 27,832 genes), and COG/KOG (77.72%, 26,684 genes). Pfam domain annotation covered 71.80% (24,654 genes), while GO and KEGG annotations described 36.93% (12,680 genes) and 36.87% (12,658 genes), respectively. Swiss-Prot annotation accounted for 57.04% (19,584 genes). The high annotation efficiency verifies the reliability of this gene set for subsequent functional genomics and evolutionary research.

3.4 Gene Family Dynamics

Orthologous clustering across *E. splendens* and 13 representative plant species identified 37,209 gene families, including 4,650 core families conserved across all species and 606 families specific to *E. splendens* (Fig. 2A, Tables S20-S21). Enrichment analysis revealed that species-specific genes are

primarily involved in ribosome biogenesis, nucleic acid metabolism, chromatin regulation, protein homeostasis, floral organ development, embryogenesis, seed development, and biotic defense (Table S22). KEGG analysis linked these genes to ascorbate metabolism, endoplasmic reticulum protein processing, protein export, RNA polymerase-mediated transcription, and fatty acid elongation pathways (Table S23), supporting their roles in transcriptional regulation, protein processing, metabolism, and environmental adaptation.

Gene copy number distribution analysis categorized families into single-, double-, triple-, quadruple-, and multi-copy (> 4 genes) classes. Single- and double-copy families formed the conserved genomic backbone across species (Fig. 2B). *E. splendens* exhibited a copy number distribution similar to that of *Sal*. *miltiorrhiza*, *Scu. baicalensis*, and *Ses. indicum*, reflecting Lamiaceae-wide conservation. In contrast, multi-copy families showed marked lineage-specific expansions, particularly in *V*. *vinifera*, *Sal*. *bowleyana*, and *Sal*. *rosmarinus*. Such gene family expansions reflect lineage-specific evolutionary histories and adaptive genomic plasticity.

3.5 Phylogeny and Species Divergence

A maximum likelihood phylogeny constructed from 201 single-copy orthologous genes placed *E. splendens* closest to *Salvia* and *Origanum* species, forming a distinct monophyletic clade (Fig. 2C). Divergence time estimation suggested that *E. splendens* split from its nearest relatives approximately 30.11 million years ago (Mya).

Gene family expansion and contraction analysis revealed dynamic evolutionary patterns across Lamiaceae. In *E. splendens*, 129 families expanded and 40 contracted (Table S24). Expanded families were significantly enriched in biological processes such as immune responses to biotic stress, programmed cell death, stress signaling, RNA-mediated gene silencing, and alkaloid metabolism (Table S25). KEGG pathways associated with expansion included phenylpropanoid biosynthesis, sesquiterpene and triterpene biosynthesis, brassinosteroid synthesis, RNA polymerase transcription, and ascorbate metabolism (Table S26), indicating the evolutionary enhancement of stress resistance and specialized secondary metabolism. Conversely, contracted families were associated with fungal defense, auxin response, secondary metabolism, methylation, and basal metabolic pathways such as alpha-linolenic acid, galactose, and starch/sucrose metabolism (Tables S27-S28). Collectively, these turnover events indicate that *E. splendens* improved environmental adaptability by strengthening adaptive traits while streamlining redundant metabolic pathways. Comparable expansion and contraction patterns observed in *Sal. miltiorrhiza*, *Sal. bowleyana*, *Origanum vulgare*, and *O. majorana* underscore the prevalence of gene family remodeling during Lamiaceae evolution, providing an essential evolutionary context for understanding the molecular basis of specialized traits in *E. splendens*.

3.6 Positive Selection and Genomic Collinearity

We identified 93 genes under significant positive selection (Table S29). Notably, these included heavy metal detoxification superfamily proteins (*g28509*), ABC transporters (*g9348*, *g26436*), and cadmium zinc-transporting ATPase (*g14686*), which underpin heavy metal uptake, translocation, and

sequestration. Adaptive variations were also detected in peroxidases (*g24555*), small heat shock proteins HSP20 (*g9751*), oxidoreductases (*g22562*), and mitochondrial dehydrogenases (*g12854*), enhancing reactive oxygen species scavenging and redox homeostasis. Furthermore, selective sweeps in ubiquitination enzymes (*g5995*, *g24486*) and stress-responsive transcription factors (MYB, bHLH; *g5700*, *g31464*) improved stress signal perception and transduction. Together, these findings highlight a coordinated molecular adaptation strategy integrating metal detoxification, oxidative stress tolerance, protein quality control, and developmental plasticity.

Genome-wide collinearity analysis with three representative Lamiaceae species revealed extensive conserved collinearity (Fig. 3A). Large-scale syntenic blocks confirmed close evolutionary relationships, although localized micro-rearrangements, inversions, and fragment shuffling were also detected. No whole-chromosome fusions, fissions, or translocations were observed, indicating that macro-synteny is largely conserved while microstructural variations contribute to species differentiation. These collinearity results further support our phylogenetic inference and provide structural evidence for genome evolution in *E. splendens*.

3.7 LTR-RT Insertion Dynamics and Genome Expansion

The insertion times of intact LTR-RTs were inferred to trace genome evolution, with comparisons drawn among four additional Lamiaceae species. LTR-RT insertion profiles varied markedly, suggesting independent transposition histories (Fig. 3B). *E. splendens* exhibited a pronounced recent burst of LTR-RT insertions, with a sharp unimodal peak at 0.0241 Mya (density = 2.4597), far exceeding the activity levels observed in the other species. Approximately 70% LTR-RT insertions were concentrated within the last one Mya, with ancient elements nearly absent, implying intense recent amplification coupled with efficient removal of older copies via sequence homogenization or deletion.

In contrast, *Sal. bowleyana* (0.0768 Mya, density = 1.2241) and *Sal. rosmarinus* (0.0679 Mya, density = 1.0542) displayed moderate recent peaks followed by gradual decay, indicative of sustained but milder transposition activity. *C. americana* (1.0679 Mya, density = 0.4131) and *Scu. baicalensis* (1.0567 Mya, density = 0.3617) exhibited older insertion peaks, reflecting predominantly ancient LTR-RT expansion with limited recent activity. The recent LTR-RT burst in *E. splendens* likely drove substantial genomic restructuring, influencing neighboring gene expression and contributing to stress adaptation. This lineage-specific transposition surge represents a key mechanism shaping the genomic architecture and adaptive potential of *E. splendens* in heavy metal-contaminated environments.

To elucidate the evolutionary divergence of *E. splendens*, we estimated pairwise Ks for collinear gene blocks between *E. splendens* and six representative angiosperms, visualizing distributions using kernel density estimation (Fig. 3C). Intragenomic syntenic blocks within *E. splendens* formed a sharp Ks peak near zero. Interspecific comparisons revealed progressively higher Ks peaks correlating with phylogenetic distance. Comparisons with closely related Lamiaceae species (*Sal. rosmarinus*, *Sal. bowleyana*, *Scu. baicalensis*, *C. americana*) produced narrow Ks peaks centered between 0.4 and 0.7,

reflecting relatively recent shared ancestry. By contrast, more distantly related species showed significantly elevated Ks values, as demonstrated by the two most distal pairwise comparisons. The *E. splendens – V. vinifera* ortholog peak fell near Ks = 1.5, whereas the *E. splendens – A. thaliana* comparison yielded the broadest distribution, peaking above Ks = 2.2, consistent with deep evolutionary divergence.

3.8 Transcriptomic Responses to Copper Stress in Stems and Leaves

To dissect tissue-specific copper responses, we performed RNA-seq on stems and leaves under control (CK), low copper (Cu100), and high copper (Cu500) treatments. After stringent quality filtering, 528.46 million clean reads were obtained, with an average mapping rate of 94.91% to the reference genome (Table S30). Sample correlation and principal component analysis confirmed high reproducibility among biological replicates and clear separation across tissues and copper treatments, indicating extensive transcriptional reprogramming induced by copper stress (Fig. 4A-B).

Differential expression analysis identified numerous concentration-dependent differentially expressed genes (DEGs) (Fig. 4C-H, Table S31). In leaves, Cu100 and Cu500 induced 686 and 4,576 DEGs, respectively, with 346 shared between treatments. In stems, Cu100 and Cu500 elicited 1,485 and 5,924 DEGs, respectively, with 990 shared. Overall, high-concentration copper triggered more extensive transcriptional changes. Tissue-specific analysis revealed 2,385 leaf-specific and 3,888 stem-specific DEGs, with 2,531 DEGs common to both organs, which highlights divergent yet complementary regulatory strategies in the two tissues.

GO and KEGG enrichment analyses highlighted functional specialization (Tables S32-S35). Leaf DEGs were strongly associated with copper ion binding (GO:0005507), oxidoreductase activities, transmembrane transport, and ion binding, reflecting metal chelation, oxidative detoxification, and ion sequestration. Enriched biological processes included metal ion response, abiotic stress, and phenylpropanoid metabolism, suggesting activation of antioxidant defenses. Cellular component terms pointed to plasma membrane and apoplast remodeling to limit intracellular metal influx.

In stems, DEGs were predominantly enriched in phenylpropanoid biosynthesis, lignin metabolism, glutathione metabolism, heavy metal response, xenobiotic transport, and hormone signaling (ABA, ethylene). Cellular components related to the apoplast, extracellular region, and cell wall were prominently enriched, supporting a structural barrier strategy for metal immobilization. Molecular functions emphasized oxidoreductases, metal ion binding, and transporter activities. KEGG analysis corroborated these findings, highlighting phenylpropanoid biosynthesis, cytochrome P450 metabolism, glutathione metabolism, amino acid metabolism, and transporter pathways. In summary, leaves mainly execute oxidative stress alleviation and metabolic adjustment, while stems enhance cell wall lignification and physical barriers to achieve long-distance metal sequestration and translocation.

WGCNA identified 27 modules, from which turquoise, blue, yellow, and grey60 were designated as copper-stress core modules (Fig. S6-S8). The blue module regulated copper efflux and signaling; turquoise maintained redox balance and energy metabolism; yellow preserved photosynthetic homeostasis; and grey60 facilitated DNA repair and cellular maintenance. A total of 540 hub genes were identified as central regulators orchestrating these adaptive responses(Table S36).

3.9 Characterization of the HMA Gene Family

The Heavy Metal ATPase (HMA) family comprises P-type ATPases that mediate ATP-dependent transport of Cu, Zn, Cd, and other metals. We identified 16 HMA genes in *E. splendens*, all containing canonical heavy metal-binding and ATPase domains (Fig. 5A, Table S37). Phylogenetic analysis classified them into two clades corresponding to Cu/Ag (15 members) and Zn/Co/Cd/Pb (1 member, *EsHMA2*) transporters (Fig. 5B). *E. splendens* HMAs clustered closely with homologs from *A*. *thaliana*, with several forming species-specific subclades indicative of lineage-specific diversification. Haplotype network analysis resolved 12 distinct haplotypes, with most genes possessing unique haplotypes; notably, *EsHMA2* formed a solitary haplotype (Hap2), while Hap11 encompassed five genes (*EsHMA11*-15) (Fig. 5C, Table S38). The high haplotype diversity of these copper-responsive HMA genes reflects persistent adaptive selection under heavy metal stress. Gene duplication analysis revealed that proximal duplication (10 genes) was the primary driver of family expansion, supplemented by dispersed (3 genes) and tandem duplications (2 genes), facilitating functional divergence (Table S39).

Transcriptomic profiling revealed dose-dependent and tissue-specific expression patterns (Fig. 5D). Under low copper stress (Cu100), *EsHMA3*, *EsHMA4*, and *EsHMA5* were significantly upregulated in stems, whereas *EsHMA3*, *EsHMA4*, and *EsHMA16* were downregulated in leaves, indicating stems as the principal early-response site. Under high copper (Cu500), tissues coordinately upregulated *EsHMA10*, *EsHMA11*, *EsHMA12*, and *EsHMA13*, which we define here as core high-copper-tolerance effectors. By contrast, *EsHMA3*, *EsHMA4*, and *EsHMA16* were suppressed in both tissues, and *EsHMA6* showed stem-specific downregulation, further diversifying the regulatory wiring. Collectively, these patterns outline a tissue-partitioned, HMA-mediated copper detoxification strategy in *E. splendens*.

4 Discussion

By bringing together comparative genomics, evolutionary analysis, copper-stress transcriptomics, and detailed characterization of one key gene family, this high-quality chromosome-level genome of *E. splendens* lets us trace, in a systematic way, both the evolutionary features of its genome and the molecular logic behind its heavy metal tolerance.

4.1 Genome Assembly Features and Evolutionary Patterns within the Lamiaceae

Lamiaceae span a wide range of geographies and ecological niches, and the genomic resources accumulated for this family so far point to a recurring profile: moderate genome sizes, abundant repetitive content, and heterozygosity levels shaped largely by reproductive strategy and gene flow within wild populations. *E. splendens* fits the herbaceous-wild-type pattern within this family—its genome size, heterozygosity, and repeat content from k-mer surveys align closely with *Scu. baicalensis* and *Scu. barbata*. Other Lamiaceae diverge from this baseline in different directions: *Sal. splendens* and *C. americana* carry larger genomes outright, while *Sal. rosmarinus* has expanded dramatically through repeated WGD and transposon activity [26]. Genome size across the family, in other words, varies according to each lineage's particular life history and evolutionary path rather than following a single family-wide rule.

Our chromosome-scale assembly, built from PacBio long reads anchored by Hi-C contact data, meets every standard quality benchmark expected of a high-contiguity plant genome. It also agrees closely with a previously published haplotype-resolved *E. splendens* assembly [13], a convergence that lends confidence to both the structural accuracy and gene-space completeness reported here. The eight pseudochromosomes recovered through this anchoring constitute a karyotype distinct from related Lamiaceae genera and species.

4.2 Evolutionary Functions and Applied Value of Repetitive Sequences

LTR-RTs and other repetitive elements are widely recognized as primary determinants of genome size and engines of plant adaptation [27, 28]. Repeats account for 44.54% of the *E. splendens* assembly—somewhat below the earlier k-mer-based estimate—but this figure sits in no particular relationship to overall genome size across Lamiaceae: *Sal. hispanica* (45.7%) [29], *Scu. barbata* (53.5%), and *Scu. baicalensis* (55.2%) [30] cluster near *E. splendens*, while *Sal. rosmarinus* (72.72%) [26], *O. majorana* (65.5%), and *O. vulgare* (65.4%) [31] sit considerably higher. This variation arises from both true biological divergence and technical limitations, including difficulties assembling compact heterochromatin and inconsistent repeat classification pipelines among different studies.

Within *E. splendens* itself, LTR-RTs and unclassified elements dominate the repeat landscape, followed at a distance by DNA transposons, LINEs, and SINEs—a hierarchy that matches the broader pattern by which LTR-RT proliferation drives repetitive expansion across land plants generally [32]. What sets this genome apart is timing: the recent, sharp burst of LTR-RT amplification we describe above appears to be the dominant force behind its lineage-specific genomic remodeling. This kind of late transposon reactivation is a documented route by which plants colonize extreme habitats, since insertions landing within regulatory *cis*-elements can rewire transcriptional programs and accelerate adaptive divergence [33–35]. Set against its Lamiaceae relatives, which show comparatively stable, low-level transposon activity, the magnitude of the *E. splendens* burst is unusual and points toward transposon-driven plasticity as a specific mechanism behind this species' success in colonizing copper-contaminated mine soils—a departure from the evolutionary stasis more typical of the family. Given that *E. splendens* lives

permanently in such substrates, we suspect chronic copper exposure itself sustains this LTR-RT mobilization, which in turn could reshape local genome architecture and, through insertions near regulatory regions, retune the expression of stress- and metal-transport genes. We hypothesize that this lineage-specific LTR-RT burst acts as a unique adaptive driver that differentiates this Cu-metallophyte from non-metalliferous Lamiaceae species, which requires further experimental validation.

SSR mining turned up a substantial set of polymorphic loci, again dominated by AT-rich motifs with GC-containing repeats notably underrepresented—a compositional bias that matches prior reports from *S. hispanica* [29], *Rosmarinus officinalis*, and *Ocimum basilicum* [31], and that appears to reflect a base-composition preference conserved across Lamiaceae more broadly. Beyond their immediate use as markers for population genetic studies, these repetitive loci may double as a reservoir of *cis*-regulatory elements feeding into the layered transcriptional response to copper stress described later in this discussion.

4.3 Gene Family Dynamics and Lineage-Specific Adaptive Divergence

Expansion of gene families is a recurring engine of speciation and niche adaptation, generating much of the phenotypic novelty that distinguishes related lineages [36, 37]. Among the thousands of families we annotated, 4,650 form a conserved core tied to basic cellular metabolism and characterized by single- or double-copy status across Lamiaceae—a backbone that appears deeply stable across the family. Set against this core, 606 genes unique to *E. splendens* cluster around nucleic acid metabolism, epigenetic regulation, reproductive development, and abiotic stress tolerance, a repertoire that plausibly underlies much of what separates this species phenotypically from its relatives.

Copy-number variation across Lamiaceae further illustrates how differently lineages have evolved. *Sal. bowleyana* and *Sal. rosmarinus*, for instance, owe much of their multi-copy gene family expansion to ancestral WGD compounded by tandem duplication; *E. splendens*, by contrast, shows no signature of recent WGD, and its comparatively compact genome instead achieves adaptive flexibility through targeted, localized remodeling of individual gene families. These two evolutionary strategies—whole-genome duplication versus independent gene-family expansion—represent alternative adaptive solutions under distinct ecological pressures within Lamiaceae.

Within *E. splendens* specifically, 129 families expanded against 40 contracted, and the functional split between these two sets is informative. Expansion concentrated in pathways tied to biotic stress resistance—RNA silencing, phenylpropanoid biosynthesis, terpenoid metabolism, ascorbate-based antioxidant defense—mirroring patterns reported elsewhere [38] and suggesting that the same expanded toolkit confers resistance against both pathogens and abiotic stressors simultaneously. Contraction, conversely, fell mainly on basal fungal immunity genes and conserved glycolipid metabolism. This trade-off suggests that *E. splendens* reinforces stress-adaptive pathways while pruning redundant basal metabolic functions via gene family turnover to survive metal-enriched soils.

4.4 Phylogenetic, Syntenic, and Positive Selection Evidence of Adaptive Evolution

The single-copy-ortholog phylogeny places *E. splendens* as sister to *Salvia* and *Origanum*, with a divergence time falling squarely within the period when Lamiaceae underwent its major diversification. Whole-genome synteny analysis adds a structural dimension to this placement: large collinear blocks connect *E. splendens* to its relatives, and the near-total absence of chromosome-scale fusion or fission events across these comparisons' points to a long-stable ancestral Lamiaceae karyotype. Smaller-scale rearrangements—local segmental breaks, micro-inversions—do occur, and while they are not the primary force behind family-wide differentiation, they likely still contribute something to the lineage-specific traits that set *E. splendens* apart.

Metalliferous environments are known to impose strong directional selection on functional genes, and the resulting pattern of positive selection is itself something of a hallmark among metallophytes. Consistent with this expectation, our scan identified 93 positively selected genes clustering into three functional axes—heavy metal transmembrane transport, redox homeostasis, and stress signal transduction—a distribution that closely parallels selection signatures reported in other metal-tolerant species [39, 40]. The breadth of this signal, spanning multiple pathways rather than concentrating in one or two genes, suggests that copper tolerance in *E. splendens* rests on a coordinated, multi-pathway adaptation rather than any single decisive mutation.

4.5 Copper Stress Transcriptional Responses and HMA Gene Family Adaptive Differentiation

DEG counts climbed steadily with copper concentration, the expected signature of dose-dependent transcriptional activation, and GO/KEGG enrichment sharpened this into a clear division of labor between tissues. Leaves leaned on redox-balancing and secondary-metabolic genes alongside metal chelation, limiting both oxidative damage and copper influx at the cellular boundary; stems instead combined copper sequestration, long-distance translocation, and glutathione-based detoxification with structural responses—cell wall modification, transmembrane transport, hormone signaling. This division matches established models of heavy metal handling in vascular plants generally and tracks transcriptional stress-response patterns documented elsewhere [41].

HMA transporters sit at the center of cellular metal trafficking and have already been well characterized across other hyperaccumulator species [42]. The 16 full-length paralogs we identified in *E. splendens* all retain the domains required for transport function, and phylogenetic placement split them cleanly into Cu/Ag-specific and Zn/Co/Cd clades, with 15 of 16 falling in the former—numerically consistent with this species' copper-hyperaccumulating phenotype. Most of this expansion traces to proximal duplication, with tandem and dispersed events playing smaller roles; the post-speciation timing of these duplications likely allowed paralogs to specialize into distinct transport functions, and the extensive haplotype diversity we observed across the family is itself suggestive of sustained directional selection operating

in copper-contaminated habitats. Expression data add a temporal and spatial layer to this picture: HMA activation concentrates in stems under low copper but spreads to both leaves and stems once copper concentration rises, implying a tiered defense that escalates in scope as the threat intensifies.

Several paralogs map onto known rice transporters in ways that suggest specific functions. *EsHMA3*, *EsHMA4*, and *EsHMA5*, all homologous to *OsHMA5*—which loads copper into the xylem for root-to-shoot and root-to-grain translocation in rice [20]—likely perform an analogous role here. *EsHMA10–13*, homologous to *OsHMA4* (which sequesters copper in root vacuoles and, through natural variation, modulates grain copper content in rice [43]), point toward a vacuolar sequestration function. *EsHMA16*, downregulated in both leaves and stems and homologous to *OsHMA9* (an established copper/zinc/lead efflux transporter in monocots [44]), suggests an efflux role consistent with its suppression under stress. These candidate EsHMA genes provide promising molecular targets for subsequent functional verification of copper hyperaccumulation, and lay a foundation for breeding metal-tolerant crops and developing phytoremediation germplasm.

Abbreviations

BEB
Bayes Empirical Bayes
BUSCO
Benchmarking Universal Single-Copy Orthologs
CK
control
COG/KOG
Clusters of Orthologous Groups / Eukaryotic Orthologous Groups
CTAB
Cetyltrimethylammonium bromide
Cu100
100 μM $CuSO_4 \cdot 5H_2O$ treatment group
Cu500
500 μM $CuSO_4 \cdot 5H_2O$ treatment group
DEG
Differentially Expressed Gene
FDR
False Discovery Rate
FPKM
Fragments Per Kilobase of transcript per Million mapped reads
GO
Gene Ontology
HMA
Heavy Metal ATPase

HiFi
High-Fidelity
KEGG
Kyoto Encyclopedia of Genes and Genomes
Ks
Synonymous substitution rate
LAI
LTR Assembly Index
LRT
likelihood Ratio Test
LTR-RT
Long Terminal Repeat RetroTransposon
ML
Maximum-Likelihood
MITE
Miniature Inverted-repeat Transposable Element
Mya
Million years ago
NCBI
National Center for Biotechnology Information
NGDC
National Genomics Data Center
NR
Non-Redundant
PAML
Phylogenetic Analysis by Maximum Likelihood
SMRT
Single-Molecule Real-Time
SSR
Simple Sequence Repeat
TE
Transposable Element
TPM
Transcripts Per Million
WGCNA
Weighted Gene Co-expression Network Analysis
WGD
Whole-Genome Duplication.

Declarations

Ethics approval and consent to participate

No human participants, animal subjects, or materials requiring permits or ethical approval were involved in this study.

Consent for publication

All authors have read the manuscript and approved it for publication.

Availability of data and materials

This article and its Supplementary Information files contain all original data and materials generated or analyzed during the study. Raw sequencing data (genome survey, PacBio HiFi, Hi-C, and RNA-seq) are deposited at the National Genomics Data Center (NGDC) under accession PRJCA062937, and the final genome assembly and annotation are available from NCBI under accession PRJNA1289993. Custom analysis scripts and code can be requested from the corresponding author.

Competing interests

The authors declare no commercial or financial relationships that could be construed as a potential conflict of interest in connection with this research.

Fundings

This work was supported by the Natural Science Foundation of Jiangxi Province (Grant Nos. 20232BAB205012 and 20252BAC220047), and the National Natural Science Foundation of China (Grants No. 32360157). The funding agencies had no role in the study design, data collection, analysis, interpretation, or manuscript preparation.

Authors' contributions

Licao Cui and Yiting Su contributed equally to this work. Licao Cui: Conceptualization, Formal analysis, Supervision, Writing—original draft, Writing—review & editing. Yiting Su: Formal analysis, Visualization, Writing—original draft. Han Lin: Formal analysis, Visualization. Xiaofeng Yang: Data curation, Visualization. Yu Lu: Formal analysis, Visualization. Yuzhe Song: Visualization, Validation. Yihan Li: Supervision, Writing—original draft. Qinglin Ke: Formal analysis, Visualization, Writing—review & editing. Ruimin Li: Supervision, Project administration, Writing—original draft, Writing—review & editing.

Acknowledgements

We thank the research groups whose publicly available reference genomes supported this work, and the High-Performance Computing Platform at Northwest A&F University for assistance with data processing.

References

1. Chen S, Liu W, Mei Z, Li H, Zhao W, Zhao J, Tao H. The synthesis of copper-modified biochar from Elsholtzia Harchowensis and its electrochemical activity towards the reduction of carbon dioxide. Front Chem. 2023;11:1238424.
2. Fatima S, Bashir S, Husain A, Zahra A, Shah MA, Li Z, Wang Y, Alotaibi SS. Use of organic soil conditioners to improve sunflower's phyto-stabilization in copper-contaminated soil. BMC Plant Biol. 2026;26(1):220.
3. Giannakoula A, Therios I, Chatzissavvidis C. Effect of Lead and Copper on Photosynthetic Apparatus in Citrus (Citrus aurantium L.) Plants. The Role of Antioxidants in Oxidative Damage as a Response to Heavy Metal Stress. Plants (Basel Switzerland) 2021, 10(1).
4. Tashakorizadeh M, Golkar P, Vahabi MR, Ghorbanpour M. Physiological and biochemical mechanisms of grain yield loss in fumitory (Fumaria parviflora Lam.) exposed to copper and drought stress. Sci Rep. 2023;13(1):17934.
5. Ali H, Khan E, Sajad MA. Phytoremediation of heavy metals–concepts and applications. Chemosphere. 2013;91(7):869–81.
6. Gan C, Liu Z, Pang B, Zuo D, Hou Y, Zhou L, Yu J, Chen L, Wang H, Gu L, et al. Integrative physiological and transcriptome analyses provide insights into the Cadmium (Cd) tolerance of a Cd accumulator: Erigeron canadensis. BMC Genomics. 2022;23(1):778.
7. Bai X, Bol R, Chen H, Cui Q, Qiu T, Zhao S, Fang L. A meta-analysis on crop growth and heavy metals accumulation with PGPB inoculation in contaminated soils. J Hazard Mater. 2024;471:134370.
8. Liu J, Yang D, Li X, Jin Z, Li J. In Vitro Inducted Tetraploid Elsholtzia splendens Nakai ex F. Maek. Alters Polyphenol Species and Synthesis. *Plants (Basel, Switzerland)* 2024, 13(23).
9. Ding L, Zhao X, Su L, Peng H, Sun C. The complete chloroplast genome of copper-tolerance plant Elsholtzia splendens. Mitochondrial DNA Part B Resour. 2019;4(2):2729–30.
10. Guo Z, Liu Z, Wang X, Liu W, Jiang R, Cheng R, She G. Elsholtzia: phytochemistry and biological activities. Chem Cent J. 2012;6(1):147.
11. Liu J, Yang D, Li X, Jin Z, Li J. Deciphering the effect mechanism of chromosome doubling on the biomass increase in Elsholtzia splendens. Sci Hortic. 2024;326:112751.
12. Qin C, Quan L, Wang C, Duan W, Duan K, Wang B, Yao Q, Shen Z, Xia Y, Wei Z. Differential physiological responses of copper-sensitive and copper-tolerant Elsholtzia species to copper toxicity: the character of cell walls and their subfractions. J Soil Sci. 2022;22(4):4168–78.
13. Moon SJ, Lee SH, Sim WH, Choi HS, Lee JS, Shim S. Haplotype-resolved chromosome-level genome sequence of Elsholtzia splendens (Nakai ex F.Maek). Sci data. 2025;12(1):827.
14. Møller JV, Juul B, le Maire M. Structural organization, ion transport, and energy transduction of P-type ATPases. Biochim Biophys Acta. 1996;1286(1):1–51.
15. Williams LE, Mills RF. P(1B)-ATPases–an ancient family of transition metal pumps with diverse functions in plants. Trends Plant Sci. 2005;10(10):491–502.

16. Kim YY, Choi H, Segami S, Cho HT, Martinoia E, Maeshima M, Lee Y. AtHMA1 contributes to the detoxification of excess Zn(II) in Arabidopsis. Plant journal: cell Mol biology. 2009;58(5):737–53.
17. Hussain D, Haydon MJ, Wang Y, Wong E, Sherson SM, Young J, Camakaris J, Harper JF, Cobbett CS. P-type ATPase heavy metal transporters with roles in essential zinc homeostasis in Arabidopsis. Plant Cell. 2004;16(5):1327–39.
18. Chao DY, Silva A, Baxter I, Huang YS, Nordborg M, Danku J, Lahner B, Yakubova E, Salt DE. Genome-wide association studies identify heavy metal ATPase3 as the primary determinant of natural variation in leaf cadmium in Arabidopsis thaliana. PLoS Genet. 2012;8(9):e1002923.
19. Verret F, Gravot A, Auroy P, Leonhardt N, David P, Nussaume L, Vavasseur A, Richaud P. Overexpression of AtHMA4 enhances root-to-shoot translocation of zinc and cadmium and plant metal tolerance. FEBS Lett. 2004;576(3):306–12.
20. Deng F, Yamaji N, Xia J, Ma JF. A member of the heavy metal P-type ATPase OsHMA5 is involved in xylem loading of copper in rice. Plant Physiol. 2013;163(3):1353–62.
21. Catty P, Boutigny S, Miras R, Joyard J, Rolland N, Seigneurin-Berny D. Biochemical characterization of AtHMA6/PAA1, a chloroplast envelope Cu(I)-ATPase. J Biol Chem. 2011;286(42):36188–97.
22. Li W, Lacey RF, Ye Y, Lu J, Yeh KC, Xiao Y, Li L, Wen CK, Binder BM, Zhao Y. Triplin, a small molecule, reveals copper ion transport in ethylene signaling from ATX1 to RAN1. PLoS Genet. 2017;13(4):e1006703.
23. Shao L, Guo X, Wang Y, Lei T, Wan Y, Ma L, Niu N. Genome-wide survey of the HMA gene family in wheat (Triticum aestivum L.) and its potential role in cadmium stress. BMC Genomics. 2025;26(1):619.
24. Ahmad MZ, Chen S, Qi X, Feng J, Chen H, Liu X, Sun M, Deng Y. Genome wide analysis of HMA gene family in Hydrangea macrophylla and characterization of HmHMA2 in response to aluminum stress. Plant Physiol biochemistry: PPB. 2024;216:109182.
25. Ma Y, Wei N, Wang Q, Liu Z, Liu W. Genome-wide identification and characterization of the heavy metal ATPase (HMA) gene family in Medicago truncatula under copper stress. Int J Biol Macromol. 2021;193(Pt A):893–902.
26. Han D, Li W, Hou Z, Lin C, Xie Y, Zhou X, Gao Y, Huang J, Lai J, Wang L, et al. The chromosome-scale assembly of the Salvia rosmarinus genome provides insight into carnosic acid biosynthesis. Plant journal: cell Mol biology. 2023;113(4):819–32.
27. Vitte C, Panaud O. LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. Cytogenet Genome Res. 2005;110(1–4):91–107.
28. De La Cerda GY, Landis JB, Eifler E, Hernandez AI, Li FW, Zhang J, Tribble CM, Karimi N, Chan P, Givnish T, et al. Balancing read length and sequencing depth: Optimizing Nanopore long-read sequencing for monocots with an emphasis on the Liliales. Appl plant Sci. 2023;11(3):e11524.
29. Wang L, Lee M, Sun F, Song Z, Yang Z, Yue GH. A chromosome-level genome assembly of chia provides insights into high omega-3 content and coat color variation of its seeds. Plant Commun. 2022;3(4):100326.

30. Xu Z, Gao R, Pu X, Xu R, Wang J, Zheng S, Zeng Y, Chen J, He C, Song J. Comparative Genome Analysis of Scutellaria baicalensis and Scutellaria barbata Reveals the Evolution of Active Flavonoid Biosynthesis. Genom Proteom Bioinform. 2020;18(3):230–40.
31. Bornowski N, Hamilton JP, Liao P, Wood JC, Dudareva N, Buell CR. Genome sequencing of four culinary herbs reveals terpenoid genes underlying chemodiversity in the Nepetoideae. DNA research: Int J rapid publication Rep genes genomes 2020, 27(3).
32. Park M, Sarkhosh A, Tsolova V, El-Sharkawy I. Horizontal Transfer of LTR Retrotransposons Contributes to the Genome Diversity of Vitis. *International journal of molecular sciences* 2021, 22(19).
33. Mokhtar MM, El Allali A. MegaLTR: a web server and standalone pipeline for detecting and annotating LTR-retrotransposons in plant genomes. Front Plant Sci. 2023;14:1237426.
34. Hassan AH, Mokhtar MM, El Allali A. Transposable elements: multifunctional players in the plant genome. Front Plant Sci. 2023;14:1330127.
35. Chen TH, Winefield C. Comprehensive analysis of both long and short read transcriptomes of a clonal and a seed-propagated model species reveal the prerequisites for transcriptional activation of autonomous and non-autonomous transposons in plants. Mob DNA. 2022;13(1):16.
36. Ding K, Xu Q, Zhao L, Li Y, Li Z, Shi W, Zeng Q, Wang X, Zhang X. Chromosome-level genome provides insights into environmental adaptability and innate immunity in the common dolphin (delphinus delphis). BMC Genomics. 2024;25(1):373.
37. Ajayi OO, Showalter AM. Systems identification and characterization of β-glucuronosyltransferase genes involved in arabinogalactan-protein biosynthesis in plant genomes. Sci Rep. 2020;10(1):20562.
38. Shifting the limits. in wheat research and breeding using a fully annotated reference genome. Sci (New York NY) 2018, 361(6403).
39. Ometto L, Li M, Bresadola L, Varotto C. Rates of evolution in stress-related genes are associated with habitat preference in two Cardamine lineages. BMC Evol Biol. 2012;12:7.
40. Yang Q, Shohag MJI, Feng Y, He Z, Yang X. Transcriptome Comparison Reveals the Adaptive Evolution of Two Contrasting Ecotypes of Zn/Cd Hyperaccumulator Sedum alfredii Hance. Front Plant Sci. 2017;8:425.
41. Huang Q, Huang X, Deng J, Liu H, Liu Y, Yu K, Huang B. Differential Gene Expression between Leaf and Rhizome in Atractylodes lancea: A Comparative Transcriptome Analysis. Front Plant Sci. 2016;7:348.
42. Takahashi R, Bashir K, Ishimaru Y, Nishizawa NK, Nakanishi H. The role of heavy-metal ATPases, HMAs, in zinc and cadmium transport in rice. Plant Signal Behav. 2012;7(12):1605–7.
43. Huang XY, Deng F, Yamaji N, Pinson SR, Fujii-Kashino M, Danku J, Douglas A, Guerinot ML, Salt DE, Ma JF. A heavy metal P-type ATPase OsHMA4 prevents copper accumulation in rice grain. Nat Commun. 2016;7:12138.

44. Lee S, Kim YY, Lee Y, An G. Rice P1B-type heavy-metal ATPase, OsHMA9, is a metal efflux protein. Plant Physiol. 2007;145(3):831–42.

Figures

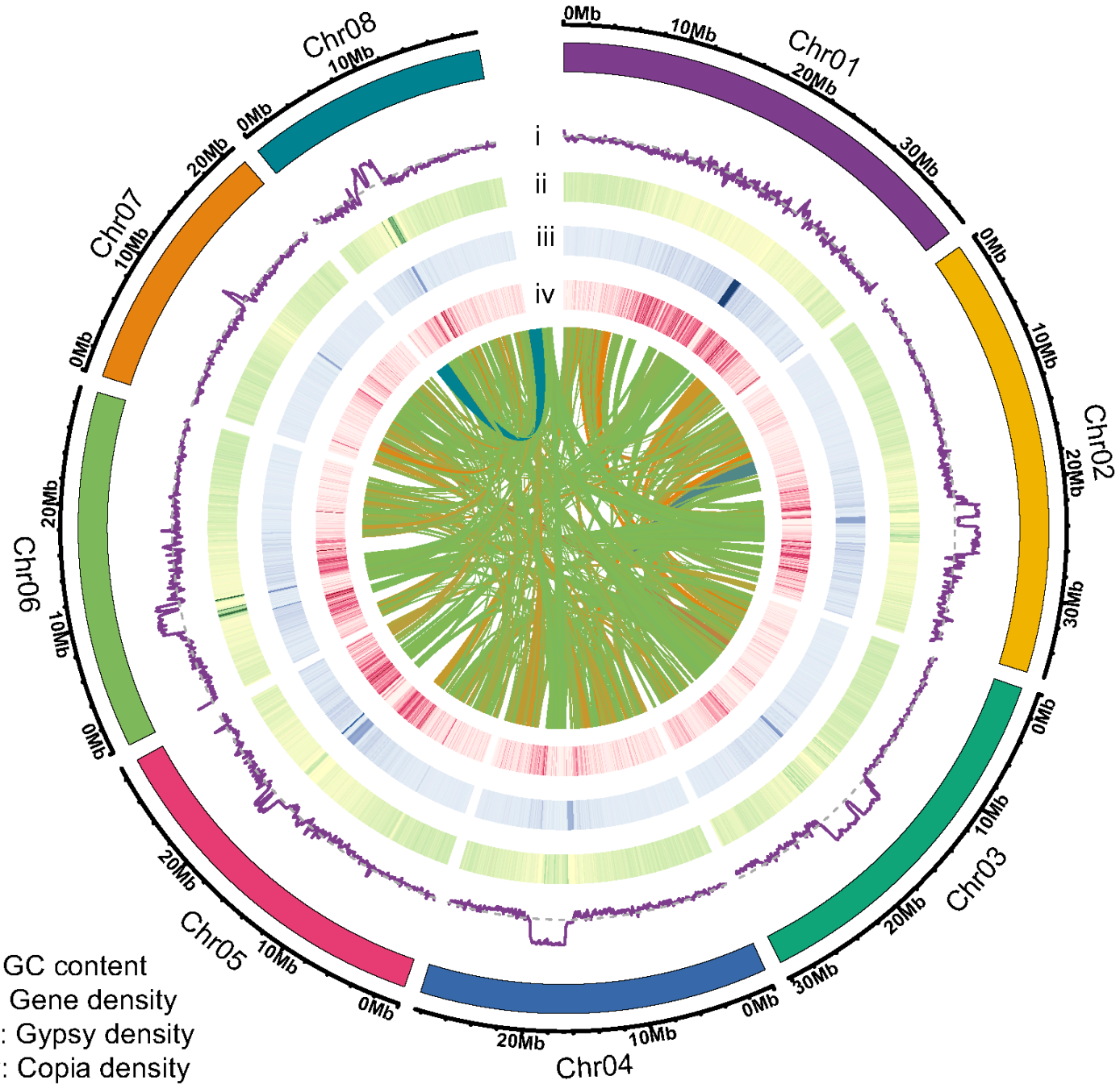

Figure 1

Circos plot showing the global genomic landscape of *E. splendens*. The outermost tracks represent eight chromosomes with physical coordinate scales (Mb). Concentric rings arranged from outer to inner

correspond to: (i) GC content, (ii) gene density, (iii) density of Gypsy-type LTR-RTs, (iv) density of Copia-type LTR-TRs. Central colored links indicate syntenic blocks across different chromosomes.

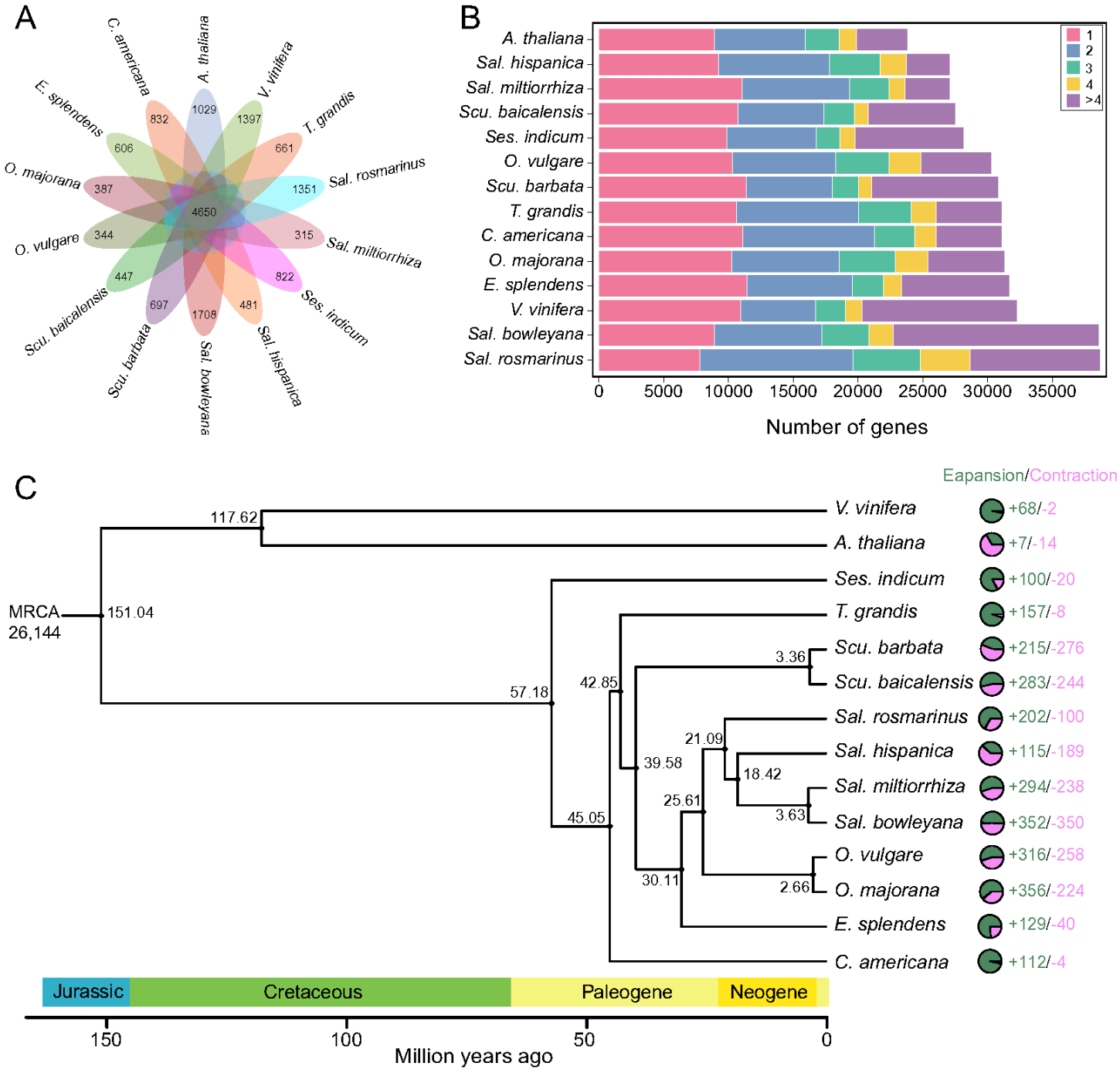

Figure 2

Comparative genomics and evolutionary dynamics of *E. splendens* and 13 representative angiosperm species. (A) Venn diagram illustrating the counts of shared orthologous gene families and species-specific gene families across the sampled taxa. (B) Cross-species distribution patterns of gene copy numbers for classified gene families. (C) Time-calibrated maximum-likelihood phylogenetic tree

depicting species divergence times and gene family expansion/contraction events along each evolutionary branch.

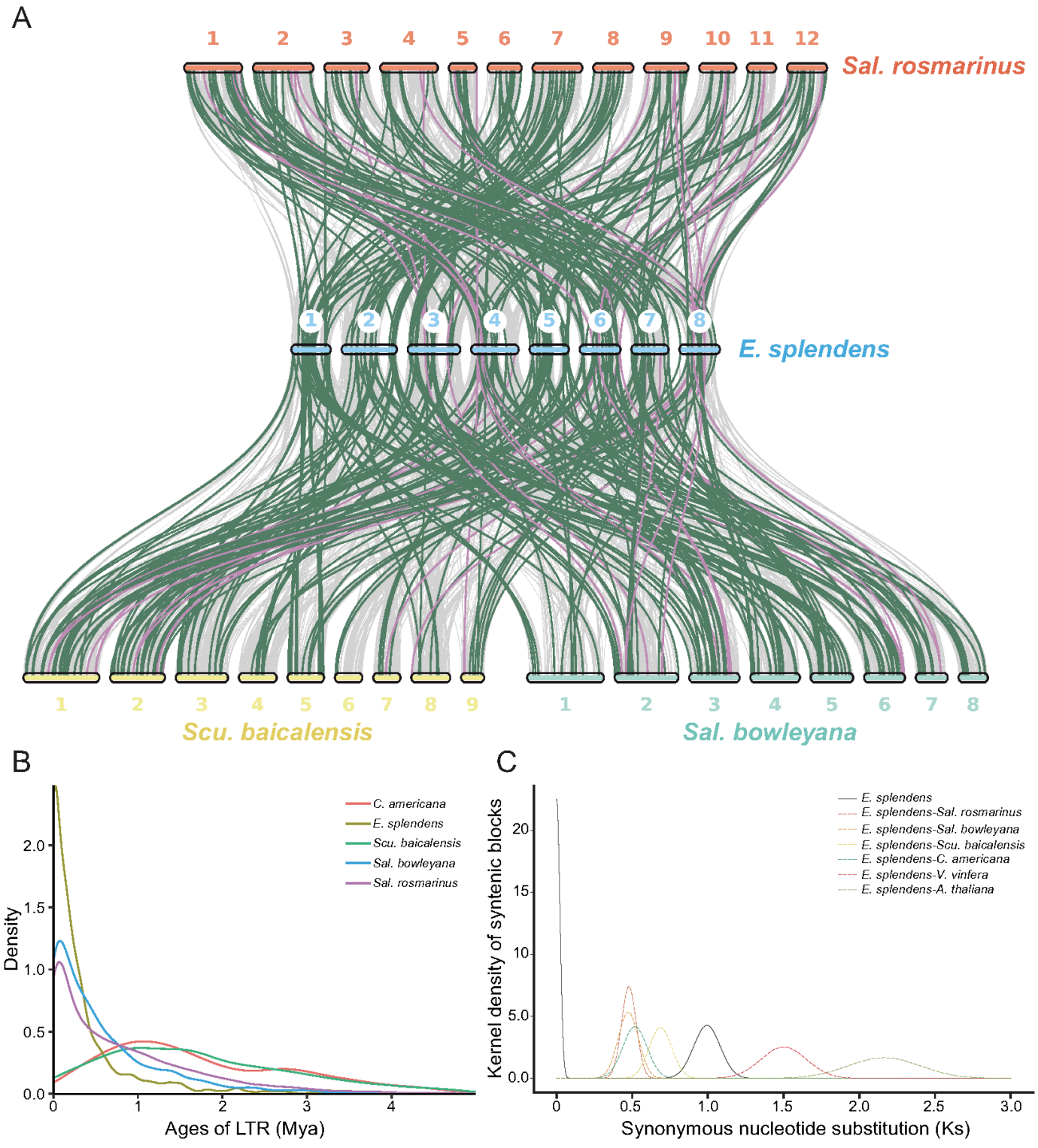

Figure 3

Genomic collinearity, LTR-RT insertion dynamics, and Ks distribution analysis in *E. splendens*. (A) Genome collinearity among *E. splendens*, *Sal. rosmarinus*, *Sal.bowleyana* and *Scu. baicalensis*. (B) Density

distribution of LTR-RT insertion ages (Mya) in *C. americana*, *E.splendens*, *Scu. baicalensis*, *Sal. bowleyana* and *Sal. rosmarinus*. (C) Pairwise synonymous Ks distributions for collinear gene pairs between *E. splendens* and selected angiosperms.

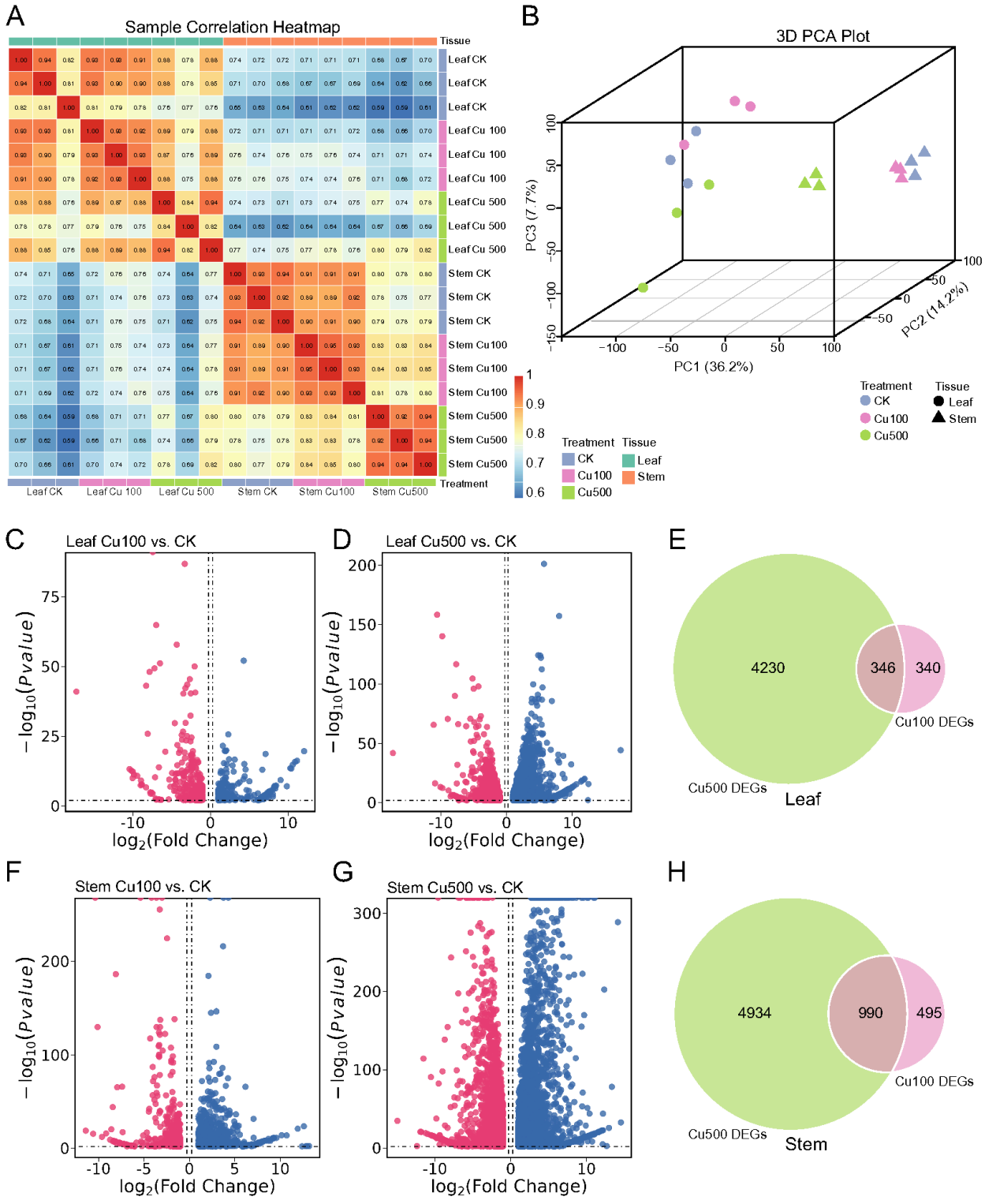

Figure 4

Overview of sample reproducibility and DEGs under copper stress. (A) Heatmap of pairwise correlation coefficients among all biological samples. (B) Three-dimensional principal component analysis of transcriptome data for all samples. (C) Volcano plot of gene expression changes in leaves between the CK and Cu100 treatment groups. (D) Volcano plot of gene expression changes in leaves between CK and Cu500 treatment groups. (E) Venn diagram showing the shared and specific DEGs in leaves under Cu100 and Cu500 treatments. (F) Volcano plot of gene expression changes in stems between CK and Cu100 treatment groups. (G) Volcano plot of gene expression changes in stems between CK and Cu500 treatment groups. (H) Venn diagram showing the shared and specific DEGs in stems under Cu100 and Cu500 treatments.

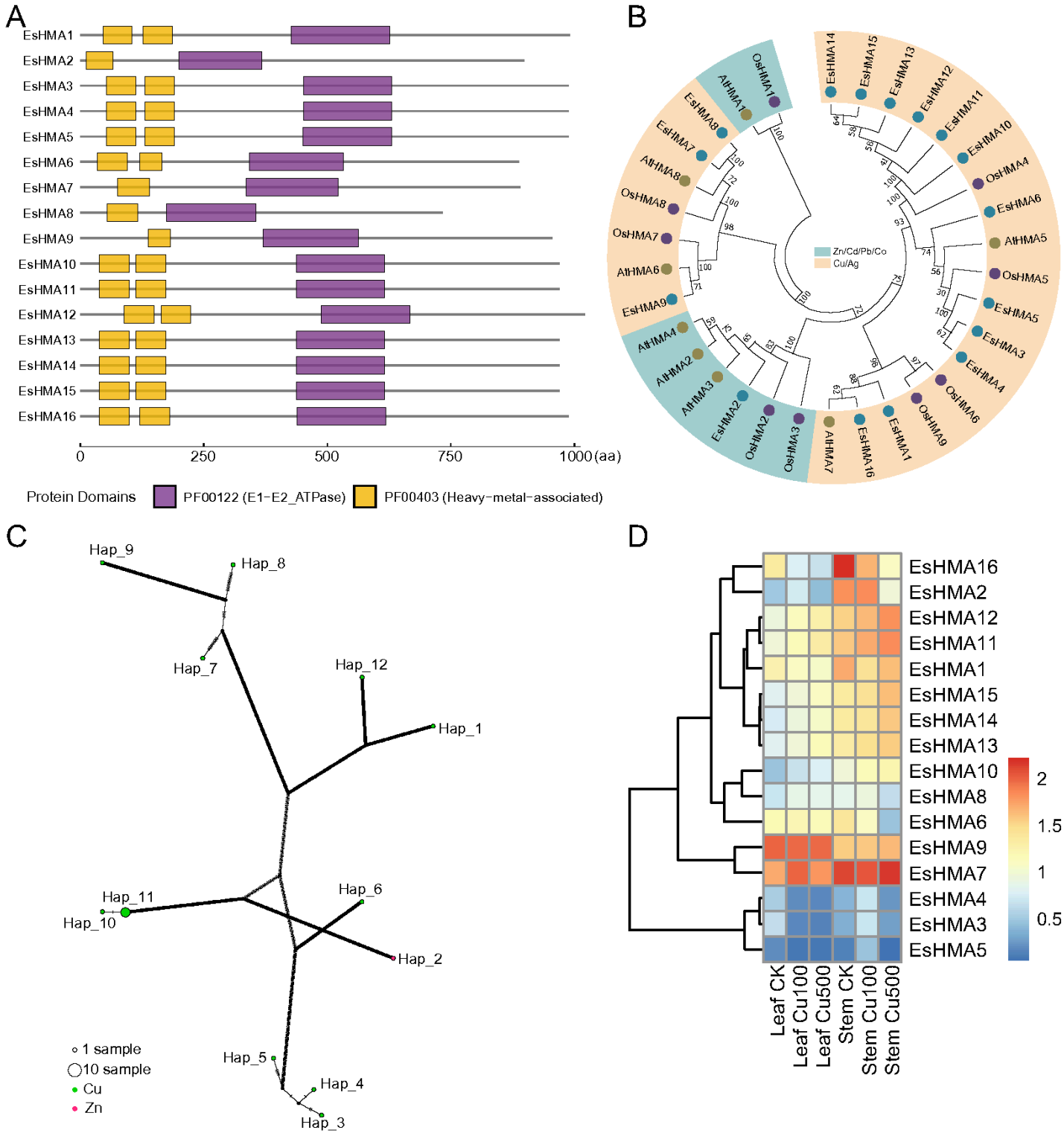

Figure 5

Characterization of the HMA gene family in *E. splendens*. (A) Conserved functional domain architecture across HMA proteins. (B) Phylogenetic clustering of HMA family members. (C) Haplotype network of the HMA gene family. (D) Heatmap showing the transcript abundance of EsHMAs in stem and leaf tissues following Cu treatment.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterials1.xlsx
- SupplementaryMaterials2.docx