Future application potential of *Ormosia* plant shown by chromosome-level genome of the rare *Ormosia henryi*

Xiao-ming Tian
Hunan botanical garden
Bin Yuan
Zhejiang University
Cun Mou
Hunan botanical garden
Guang-feng Xiang
Hunan botanical garden
Lu Zhu
Hunan botanical garden
Gao-fei Li
Hunan botanical garden
Chao Liu
Hunan botanical garden
Xiang-peng Li
Hunan botanical garden
Fu-liang Hu
flhu@zju.edu.cn

Zhejiang University
Hao Lv
lvhao0203@163.com

Hunan botanical garden

Research Article

Keywords: Ormosia henryi, chromosome-level genome, comparative genomics, gene family expansion, conservation genetics, sustainable utilization

Posted Date: October 31st, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7984394/v1

Additional Declarations: The authors declare no competing interests.

Abstract

Background.

Ormosia henryi, an endemic tree species in China. Its wood is characterized by a distinctive fragrance, dense texture, and high hardness, has substantial economic worth. However, wild *O. henryi* are continuously affected from illegal logging and over-exploitation. Compounded by its inherently slow growth rate, low natural seed set, difficulties in seed germination and habitat fragmentation, its wild populations experienced drastic decline.

Methods.

We present the chromosome-level genome assembly of the endangered timber tree *Ormosia henryi* (2.64 Gb, Contig N50 = 39.17 Mb, Scaffold N50 = 338.40 Mb), integrating comparative genomics to elucidate its adaptive evolution and utilization potential.

Results.

Genome annotation revealed 83.89% repetitive sequences and 39,017 protein-coding genes, of which 99.32% are functionally annotated. Phylogenetic analysis identified *Lupinus albus* as the closest relative (divergence time ~ 53.82 million years ago). Two whole-genome duplication events (Ks peaks at 0.47 and 2.26) and significant gene family expansion (276 families) were detected, alongside 122 positively selected genes enriched in photosynthesis, phenylpropanoid biosynthesis, and DNA repair pathways, demonstrating enhanced intrinsic adaptability. This evidence indicates that the endangerment of *O. henryi* may primarily stem from overexploitation rather than that by genetic vulnerability. This genomic resource enables molecular breeding (e.g., growth-regulating tree height using regulating ATP-dependent DNA helicase PIF1), conservation strategies, and sustainable industrial applications (high-value wood processing, medicinal resource development, ecological restoration).

Introduction

Ormosia henryi (2n = 16, Fig. 1), an endemic tree species in China, belonging to *Fabaceae* family (Huang and Xu, 1986). It possesses significant ecological, medicinal (e.g., its leaves, roots, stems, and seeds are used for treating ailments such as bruises and arthritis), cultural, and ornamental value. Crucially, its wood is characterized by a distinctive fragrance, dense texture, and high hardness, making it an ideal raw material for furniture and handicrafts with substantial economic worth (Wang et al., 2023, Wu et al., 2023a). However, due to its exceptional timber value, wild *O. henryi* are continuously affected from illegal logging and over-exploitation (Fan et al., 2007). Compounded by its inherently slow growth rate, low natural seed set, difficulties in seed germination (attributed to hard seed coats), and habitat fragmentation, its wild populations experienced drastic decline (Ge and Wei, 2024a, Zhou et al., 2023). It is currently listed as a National Class II Protected Wild Plant in China.

To enable the sustainable utilization and effective conservation of this precious timber resource, a thorough understanding of its genetic foundation is urgently needed (Luo et al., 2024, Yin et al., 2024). Nevertheless, genomic research on *Ormosia* species remains relatively limited. Although draft genomes of three closely related species (*O. purpureiflora*, *O. emarginata*, *O. semicastrata*) were published, revealing common genomic features within the genus including large size, high repetitive sequence content, and high heterozygosity, their assemblies based on Oxford Nanopore technology were constrained by high error rates, limiting genetic assembly quality (Wang et al., 2025, Liu et al., 2023). Zhou et al. (2025) generated a chromosome-scale genome assembly for *O. henryi* utilizing PacBio HiFi sequencing among other technologies. However, in-depth comparative genomic analyses to elucidate its adaptive evolutionary mechanisms are still lacking, which continues to hinder the development of targeted utilization and conservation strategies for *O. henryi*.

Herein, we aimed to construct a higher-quality chromosome-level reference genome for *O. henryi* and conduct cross-species comparative genomic analyses encompassing 10 Papilionoideae species. We focus on elucidating the evolutionary position, environmental adaptation mechanisms through phylogenetic reconstruction, divergence time estimation, investigation of whole-genome duplication (WGD) events, genome collinearity comparisons, and identification of species-specific gene families. Our findings will provide core genomic resources and a theoretical foundation for the conservation, genetic improvement, and trait-directed breeding of *O. henryi*.

Materials & Methods

Plant material

Healthy *O. henryi* seedlings (n = 3, three-year-old) were sampled from the nursery of the Hunan Botanical Garden (28°20'N, 113°E; altitude 70–85 m). Root, stem, and leaf tissues were separately collected from each individual plant. Thereafter, the samples were flash-frozen in liquid nitrogen and subsequently stored at -80°C.

Genomic sequencing

A multi-platform sequencing strategy was employed for genome sequencing. Initially, high-quality genomic DNA was extracted from young leaves using a plant DNA extraction kit (TIANGEN, China). This DNA was used to construct an Illumina library with an insert size of ~ 350 bp, which was sequenced on the DNBSEQ-G400 platform. Raw data underwent quality control and filtering using SOAPnuke v1.5.6.

Subsequently, HiFi sequencing was performed. DNA was sheared using a Megaruptor, and fragments of 13–16 kb were size-selected using a Sage ELF system. A 20 kb SMRTbell library was constructed and sequenced in CCS (Circular Consensus Sequencing) mode on the PacBio Sequel II platform. CCS v5.0.0 processing yielded 96.52 Gb of HiFi clean data.

For chromosomal conformation capture (Hi-C) analysis, fresh leaf tissue was cross-linked with formaldehyde. Cells were lysed, and genomic DNA was digested with the restriction enzyme MboI. Following end repair, biotin labeling was introduced, and proximity ligation was performed using DNA ligase. After protease digestion to reverse cross-links, DNA was purified, sheared to 300–700 bp, and interacting fragments were specifically captured using streptavidin beads to construct the Hi-C library. Library quality was assessed using Qubit 2.0 and Agilent 2100, with effective concentration validated by qPCR prior to sequencing on the DNBSEQ platform (insert size 350 bp), generating ~ 434.2 Gb of valid data. Contigs were subsequently clustered, ordered, and anchored into chromosomal scaffolds using 3D-DNA v180114, followed by manual correction of scaffold order and orientation using Juicebox v2.1.0.

Additionally, total RNA was extracted from *Ormosia henryi* root, stem, and leaf tissues. Following library construction and quality control (Qubit 2.0, Agilent 2100), RNA sequencing was conducted on the DNBSEQ platform. Data quality was assessed using FastQC v0.11.8.

Genome assembly and assessment

Assembly commenced with k-mer frequency analysis of Illumina sequencing data using Jellyfish v2.1.4, followed by genome size, heterozygosity, and repeat content estimation via the Genomescope model. The PacBio HiFi reads were then assembled *de novo* using Hifiasm v0.16, with subsequent removal of redundant sequences performed by purge_dups. Finally, Hi-C sequencing data were employed to anchor the assembled contigs onto chromosomes, resulting in a high-quality chromosome-level genome assembly for *O. henryi*.

Assembly quality was assessed through three complementary approaches: (1) Genome completeness was evaluated using BUSCO v5.2.2 with the embryophyta_odb10 dataset; (2) Base-level accuracy was assessed by mapping Illumina short reads to the assembly using BWA v0.7.17; (3) Spatial organization validation was achieved by generating a normalized heatmap from Hi-C interaction data, where the characteristic diagonal pattern of intra-chromosomal interactions confirmed the structural integrity of the pseudochromosomes.

Gene prediction and annotation

The assembled genome underwent comprehensive annotation, encompassing repetitive sequences, protein-coding genes, functional annotation against multiple databases (NR, eggNOG, GO, KEGG, TrEMBL, SwissProt, KOG, Pfam, InterPro), pseudogene identification, and non-coding RNA (ncRNA) annotation.

Repetitive sequence annotation employed a hierarchical strategy: *De novo* prediction was conducted using RepeatModeler v2.0.1, integrating RECON v1.0.8, and RepeatScout v1.0.6, LTR retrotransposon structures were specifically identified with LTR_retriever v2.9.0, and the combined custom repeat library

was used to mask repetitive regions via RepeatMasker v4.1.2. Tandem repeats were detected jointly by MISA v2.1 and TRF v4.0.9.

Protein-coding gene prediction integrated evidence from three sources: (1) *Ab initio* prediction using Augustus v3.1.0 and SNAP; (2) Homology-based prediction with GeMoMa v1.7 against protein sequences from nine legume species (*Arachis hypogaea*, *Cajanus cajan*, *Glycine max*, *Lupinus albus*, *L. angustifolius*, *Lotus japonicus*, *Oxytropis ochrocephala*, *Phaseolus vulgaris*, *Trifolium pratense*); (3) Transcriptome evidence derived by mapping RNA-seq data to the genome using Hisat2 v2.1.0 and StringTie v2.1.4, with *de novo* transcript assembly provided by Trinity v2.8.5. Predictions from these methods were integrated and optimized using EVM v1.1.1, and UTR annotation, along with alternative splicing isoforms, were further refined by PASA v2.3.3.

Non-coding RNA annotation involved: tRNA identification using tRNAscan-SE v1.3.1; rRNA prediction with Barrnap v0.9; and identification of other RNAs (miRNA, snRNA, snoRNA) through alignment against the Rfam database.

Pseudogenes were predicted by initially screening for homologous sequences using GenBlastA v1.0.4, followed by detection of frameshift mutations and premature stop codons via GeneWise v2.4.1.

Functional annotation was performed by conducting BLAST searches against eight major databases (NR, SwissProt, TrEMBL, KOG, Pfam, InterPro, GO, KEGG). The completeness of the final annotated gene set was assessed using BUSCO.

Gene family clustering

Orthologous clustering analysis of the complete protein sequences from *Ormosia henryi* and nine other legume species (Table S1) was performed using OrthoFinder v2.4 (DIAMOND alignment, e-value threshold 1e-5). Following gene family classification, functional annotation was conducted using the PANTHER v15 database. Gene Ontology (GO) functional enrichment and KEGG pathway enrichment analyses (significance threshold q-value < 0.05) were specifically performed on *O. henryi*-specific gene families using clusterProfiler v3.4.4.

Phylogeny

A species phylogenetic tree was reconstructed based on 444 single-copy orthologous genes. Multiple sequence alignment was first executed using MAFFT v7.205 (parameters: --localpair --maxiterate 1000). Low-confidence alignment regions were filtered using Gblocks v0.91b (parameters: -b5 = h) to enhance data quality. Subsequently, the maximum likelihood tree was constructed with IQ-TREE v1.6.11, where the ModelFinder module identified the JTT + F + I + G4 model as the optimal substitution model. Node support was assessed using 1000 bootstrap replicates. Divergence times were estimated using the MCMCTree program within the PAML package, calibrating the evolutionary timescale using *Oryza sativa* as the outgroup.

Gene family expansion/contraction

Gene family dynamics were analyzed using CAFE v4.2. Inputting the phylogenetic tree topology and gene family clustering results, the software inferred gene family counts at ancestral nodes via birth-death models and detected significant expansion/contraction events along modern species branches (likelihood ratio test $p < 0.01$). KEGG pathway and GO functional enrichment analyses were separately performed on expanded and contracted gene families to elucidate functional evolutionary trends.

Positive selection

Positive selection signals were detected using the Branch-site model in CodeML (PAML v4.9). Single-copy gene families shared between *O. henryi* and closely related species (*C. cajan*, *G. max*, etc.) were screened. Protein alignments were converted to those of codon-based using PAL2NAL. Likelihood ratio tests (LRT) were conducted by comparing the fit of Model A (allowing $\omega > 1$ on foreground branches, *O. henryi* was defined as foreground branches) against that of the null model ($\omega \leq 1$ globally). Genes showing significant LRT differences ($p < 0.05$) were further analyzed using the Bayes Empirical Bayes (BEB) method to identify positively selected sites (posterior probability > 0.95 considered significant).

WGD analysis

Paleopolyploidy events were investigated through synteny analysis and molecular clock modeling. Syntenic blocks between *O. henryi* and *C. cajan*, *Medicago truncatula*, and *G. max* were identified using MCScanX (parameter: -m 25). Homologous gene pairs within these blocks were extracted, and their synonymous substitution rates (Ks) were calculated using PAML. Ks distribution histograms were plotted to detect peaks indicative of WGD event timing. Additionally, the rate of transversions at four-fold degenerate sites (4DTv) was calculated; its bimodal distribution was used to validate the number and timing of paleopolyploidization events.

LTR insertion time

Candidate Long Terminal Repeat retrotransposon (LTR-RT) sequences were identified using LTR_FINDER v1.07 and LTRharvest v1.5.9 in parallel, with results integrated and filtered by LTR_retriever v2.8. To estimate insertion time, the 5' and 3' LTR sequences of each element were extracted. These sequences were aligned using MAFFT (parameters: --localpair --maxiterate 1000). The Kimura 2-parameter genetic distance (K) was calculated using the EMBOSS v6.6.0 distmat program (Rice; Longden & Bleasby, 2000). Insertion time (T) was calculated using the formula $T = K/(2\times r)$, where the molecular clock rate (r) was set to 7×10^{-9} substitutions per site per year.

Results

High-quality chromosome-scale genome assembly

Through the integration of PacBio HiFi long-read sequencing (96.52 Gb, depth 36.7×), Illumina short-read error correction, and Hi-C chromosome conformation capture technology (434.2 Gb clean data, depth 165.10×), we successfully assembled a chromosome-level genome for *O. henryi* (Table S2). The assembled genome size is 2.64 Gb, with a contig N50 of 39.17 Mb and a scaffold N50 enhanced to 338.40 Mb, representing the highest continuity assembly among currently published *Ormosia* species (Table 1, Table S3, and Figure S1). A total of 97.8% of the sequences (2,639,023,864 bp) were successfully anchored onto 8 pseudochromosomes (Table S2). The Hi-C interaction map revealed distinct chromosome boundaries and clear intra-chromosomal interaction signals (Fig. 2). BUSCO assessment demonstrated a high assembly completeness of 98.3% (90.7% and 7.6% single-copy and duplicated genes, respectively), confirming the high completeness and reliability of the assembly (Table 2).

Table 1
Summary of Ormosia henryi genome assembly.

Statistic	Scaffold	Contig
Total number	361	537
Total length of (bp)	2,639,023,864	2,638,935,864
Gap number (bp)	586,000	0
N50 length (bp)	338,397,820	39,169,180
N90 length (bp)	240,162,212	9,385,987
Maximum length (bp)	380,859,801	148,418,790
Minimum length (bp)	240,162,212	9,385,987

Table 2
Genome assembly evaluation using BUSCO

Type	Contig level		Chromosome level	
	Number	Percentage(%)	Number	Percentage(%)
Complete BUSCOs (C)	1586	98.3	1587	98.3
Complete and single-copy BUSCOs (S)	1464	90.7	1464	90.7
Complete and duplicated BUSCOs (D)	122	7.6	123	7.6
Fragmented BUSCOs (F)	11	0.7	10	0.6
Missing BUSCOs (M)	17	1.0	17	1.1
Total BUSCO groups searched	1614	-	1614	-

Genomic features and gene annotation

Repetitive sequences constitute 83.89% of the *O. henryi* genome, with transposable element (TE), tandem repeats (TRF), long terminal repeat (LTR) retrotransposons, and DNA transposons accounting for 78.45%, 5.44%, 74.14%, and 3.90%, respectively (Fig. 3 and Table S4). A total of 182,994 intact Copia and 631,036 intact Gypsy elements were identified among the LTR retrotransposons (Table S4). Non-coding RNAs (ncRNAs) represent 0.0005% of the genome, comprising ribosomal RNAs (rRNAs, 9,174), transfer RNAs (tRNAs, 2,068), and small RNAs (including miRNAs, snRNAs, and snoRNAs, totaling 2,954, Table S5). Pseudogenes account for 0.21% of the genome, totaling 776 in number with cumulative and average lengths of 5,649,925 bp and 7,280.83 bp (Table S6), respectively.

Based on this genomic assembly, we predicted 39,017 protein-coding genes using integrated *Ab initio*, homology-based, and transcriptome-supported approaches (Table S7). Functional annotation was achieved for 99.32% (38,753) of these genes in at least one public database (Table 3). Annotation revealed that 72.57%, 79.92%, and 77.79% of the genes matched entries in the SwissProt database, assigned Gene Ontology (GO) functional classifications (Fig. S2), and mapped to KEGG metabolic pathways (Fig. S3), respectively. Finally, the BUSCO analysis showed that 98.57% of the genes were identified during the annotation, indicating that they were complete and reliable (Table S8).

Table 3
Statistical results of gene functional annotations.

Database	Number	Percentage (%)
GO_Annotation	31181	79.92
KEGG_Annotation	30352	77.79
KOG_Annotation	22984	58.91
Pfam_Annotation	32964	84.49
Swissprot_Annotation	28316	72.57
TrEMBL_Annotation	38735	99.28
eggNOG_Annotation	33323	85.41
nr_Annotation	38532	98.76
All_Annotated	38753	99.32

Comparative genomics and evolutionary insights

Comparative genomic analysis across 10 Leguminosae species (Fig. 4A) identified 37,523 gene families. Among these, 427 gene families (comprising 2,630 genes, Fig. 4 and Table S9) were unique to *O. henryi* and significantly enriched in pathways related to alanine, aspartate, and glutamate metabolism, diterpenoid biosynthesis, and photosynthesis (Fig. 4B and Table S10). Phylogenetic reconstruction indicated the closest relationship between *O. henryi* and *Lupinus albus*, with an estimated divergence time of ~ 53.82 million years ago (Fig. 4C). Gene family expansion-contraction analysis revealed that, compared with that of the ancestral state, *O. henryi* exhibited significant expansion of 276 gene families, which were significantly enriched ($q < 0.05$) in pathways associated with protein export, RNA polymerase activity, and photosynthesis (Fig. 4D and Table S11). Contrastingly, only 9 gene families (containing 12 genes) underwent contraction, including genes regulating plant height such as ATP-dependent DNA helicase PIF1 (Gene ID: Omi05G027420.1, Table S12). Positive selection analysis identified 122 genes ($p < 0.05$) enriched in base excision repair, spliceosome, and endoplasmic reticulum protein processing pathways (Fig. 4E and Table S13), highlighting adaptive evolution in core cellular maintenance pathways potentially critical for stress resilience.

Synteny

Genomic synteny analysis between *O. henryi* and closely related species (Fig. S4) identified 22, 20, and 20 syntenic blocks with *Cajanus cajan* (pigeon pea), *Medicago truncatula* (barrel medic), and *Glycine max* (soybean), respectively. These blocks correspond to those of 33,380 (49.67%), 38,875 (46.58%), and 56,407 (59.36%) syntenic gene pairs (Table S14). The highest degree of synteny was observed between *O. henryi* and *G. max*, consistent with that of their phylogenetic positions within the Fabaceae.

WWGD analysis further revealed that *O. henryi* underwent two independent WGD events, evidenced by Ks peaks at 0.47 and 2.26. The more recent event postdates the WGD in *G. max* but predates that in *M. truncatula* (Fig. 5A). The result of 4DTv also supported this (Fig. S5). Analysis of long terminal repeat (LTR) insertion times indicated that the burst of LTR retrotransposon (LTR-RT) activity in *O. henryi* predominantly occurred within the last million years (< 1 Mya; Fig. 5B and Fig S6). Copia element insertions (peak ~ 0.06 Mya) occurred later than those of Gypsy, potentially reflecting an adaptive response to the aridification associated with Late Pleistocene climatic shifts.

Discussion

Genome resource significance

Compared with that of earlier studies (Table S15-17), we successfully constructed a higher-quality chromosome-level reference genome (2.64 Gb; Contig N50 = 39.17 Mb; Scaffold N50 = 338.40 Mb; Hi-C anchoring rate = 97.8%) for the rare and endangered *O. henryi* (Liu et al., 2023, Wang et al., 2025, Zhou et al., 2025). In plant genomics research, generating high-quality chromosome-level reference genomes is necessary for in-depth exploration of plant evolutionary history, environmental adaptive mechanisms, and the genetic basis of key traits (Chen et al., 2025, Xu et al., 2025). To effectively overcome the assembly challenges posed by the high heterozygosity (1.37%) and high repetitive sequence content (83.89%) of *O. henryi* genome, we employed PacBio Sequel II CCS mode sequencing combined with Hi-C technology, balanced read length, and high accuracy. The resulting assembly exhibits superior contiguity and completeness compared with those of the reported *O. henryi* genomic, as evidenced by BUSCO assessment (Table S17) (Zhou et al., 2025). A total of 38,753 protein-coding genes were annotated, with functional annotation achieved for 99.32% of them. Compared with those of the previously reported genome assemblies of other *Ormosia* species (*O. purpureiflora*, *O. emarginata*, *O. semicastrata*) based on Oxford Nanopore (Table S15–16), this assembly demonstrates significant advantages in assembly accuracy and gene annotation rate, thereby providing a more reliable and validated foundation for subsequent in-depth analyses (Liu et al., 2023, Wang et al., 2025).

Demystifying the endangered status

While genome assembly provides fundamental genetic information for a species, its utility remains limited as it cannot dynamically reveal the driving forces of species evolution, changes in gene function over time, or the shared and unique features of genome structure among closely related species (Cheng et al., 2025). This limitation hinders a deeper understanding of the unique biological characteristics of *Ormosia henryi* and its endangerment mechanisms, and fails to fully leverage the significant potential of the genome in both fundamental theoretical and applied research. Therefore, building conducted systematic comparative genomic and evolutionary analyses of *O. henryi*. Our findings suggest that the endangerment of *O. henryi* may not primarily stem from intrinsic adaptive deficiencies, although it is attributable to persistent and intense external anthropogenic pressures. These include overexploitation

(including illegal logging) driven by its exceptionally high economic value (premium timber and medical value), as well as habitat loss and fragmentation (Xiao et al., 2024).

The traditional view largely attributes the decline in *O. henryi* population size to its inherent physiological disadvantages, such as slow growth, low natural seed set, and difficulties in seed germination (Zhang; et al., 2003, Ge and Wei, 2024b). Our genomic findings suggest that these traits do not necessarily reflect an inherent lack of genetic fitness or adaptive potential. Rather, in the absence of intense anthropogenic pressure, the species demonstrates the capacity for successful regeneration and persistence, as evidenced by some geographic reports (*O. henryi* was found to be a co-dominant or dominant canopy tree in specific areas) (Peng et al., 2018). Furthermore, some populations exhibit successful seedling recruitment (Yang, 2023, Li, 2024). Critically, *O. henryi* is one of the most adaptable plants distributed in southern China (Song et al., 2021). The slow life history strategy may be evolutionarily stable under natural disturbance regimes (when growth in a stable environment), although it becomes a critical vulnerability under current levels of exploitation and habitat degradation (Santini et al., 2019).

Collectively, our genomic analyses provide compelling evidence that *O. henryi* possesses a strong genomic foundation and evolutionary history indicative of significant adaptive potential. This strongly suggests that its current endangered status is likely driven predominantly by persistent and intense anthropogenic pressures, rather than that stemming fundamentally from intrinsic genetic deficiencies or a lack of evolutionary resilience. Molecular evolutionary evidence indicates that *O. henryi* experienced two ancient WGD events, occurring ~ 58 Mya and at an earlier time. The latter corresponds to the core-eudicot shared WGD, while the former aligns with the WGD event shared by the subfamily Papilionoideae (Chen et al., 2022). Crucially, gene family evolution analysis revealed significant gene family expansion (276 families) throughout the evolutionary history of *O. henryi*, with minimal contraction (only 9 families). These significantly expanded gene families are functionally highly enriched in pathways critically associated with environmental adaptation, including protein export, RNA polymerase activity, and photosynthesis, phenylpropanoid biosynthesis, and alanine, aspartate and glutamate metabolism (de Vries et al., 2021, Lai et al., 2024). This pattern indicates the potential for enhanced primary metabolism (e.g., efficient photosynthetic carbon fixation and nitrogen assimilation), potentially providing ample precursors for secondary metabolism (such as the synthesis of medicinally valuable phenylpropanoids). Furthermore, by optimizing intracellular protein synthesis, processing, folding (i.e., proteostasis maintenance), and RNA transport efficiency, it ensures the smooth operation of complex metabolic networks (Hao et al., 2012, Tao et al., 2016). Positive selection analysis further identified 122 genes under significant positive selection, which are significantly enriched in pathways including base excision repair, spliceosome, homologous recombination, non-homologous end joining, and endoplasmic reticulum protein processing. These findings indicate that throughout its prolonged evolutionary history, *O. henryi* has continuously reinforced its capacity to maintain genome stability and achieve efficient, high-fidelity protein synthesis, which is critical for effectively countering cellular damage caused by diverse historical environmental stresses (Fry et al., 2005).

Additionally, insertion time analysis of Long Terminal Repeat Retrotransposons (LTR-RTs) revealed that their primary period of activity occurred within the last million years (< 1 Mya), with Copia elements exhibiting a later burst of insertion activity than that by Gypsy elements. The burst of LTR-RT activity, particularly Copia elements, coinciding with Late Pleistocene aridification, represents a period of significant genomic flux (Qiu et al., 2017). Such transposon activity can serve as a source of potential adaptive variation and genomic plasticity, although it also carries risks of deleterious mutations (Gu et al., 2022).

While these genomic signatures indicate an inherent capacity for metabolic robustness and stress resilience, future studies correlating gene expression patterns under various stresses with physiological and phenotypic measurements are crucial to validate the functional significance of these expansions for the specific adaptations by *O. henryi*.

Implications for conservation and sustainable utilization

This high-quality genome resource reveals the substantial potential of *O. henryi* for molecular breeding and paves the way for transitioning towards sustainable utilization. This genomic resource not only facilitates precision conservation strategies in the genomics era but also presents opportunities for the industrial application of this rare woody species in high-value-added wood processing, development of unique medicinal resources, ecological restoration, and landscape horticulture (Wang et al., 2024, Varshney et al., 2020). Although realizing this potential will require concerted efforts in conservation management, overcoming propagation challenges, and developing responsible harvesting and cultivation practices.

Building upon this theoretical breakthrough and the valuable genomic resource, future research should focus on the following interconnected directions: (1) Highly polymorphic molecular markers should be immediately developed to systematically assess the genetic diversity and structure of extant populations, guiding the formulation of scientific conservation strategies and the establishment of a core germplasm repository (Kress et al., 2015, Salgotra and Chauhan, 2023). Concurrently, population genomics analyses integrated with environmental variables can reveal differential genetic responses to environmental stresses (e.g., climate change, pests, and diseases), identify key adaptive genetic loci, and enable the selection or development of stress-resistant lines (e.g., drought, cold, and disease tolerance) to enhance cultivation stability (Chi et al., 2024, Tsumura et al., 2012). (2) The genetic basis of key economic traits (such as wood formation, bioactive compound synthesis, and growth rate) requires in-depth elucidation. This necessitates integrating multi-omics technologies like transcriptomics and metabolomics to identify core regulatory genes and key allelic variants, e.g., the annotated ATP-dependent DNA helicase PIF1, confirmed to influence growth in other plants (Borrill et al., 2022), coupled with efficient somatic embryogenesis-based rapid propagation for molecular marker-assisted breeding (Wu et al., 2023b). This integrated approach aims to develop new varieties with shortened growth cycles and enhanced biomass accumulation efficiency. (3) Large-scale breeding populations should be established, utilizing genome-wide markers to build genomic selection models for accelerating genetic improvement of complex traits such as growth, wood quality, and stress resistance. (4) The

methodologies and resources from our study should be extended to other economically or ecologically significant species within the genus *Ormosia* to promote systematic research and sustainable resource development across the genus.

Conclusions

Summarily, this study provides critical foundations for the scientific conservation of *O. henryi* and charts a genome-enabled pathway for its future sustainable resource utilization.

Declarations

Acknowledgements

We thank for the support from the 2023 Central Finance Second Batch of Subsidized Projects for Conservation of National Key Wildlife and Animals.

Data availability

The whole genome sequence data of O. henryi (No. GWHFICR00000000.1) reported in this paper have been deposited in the Genome Warehouse in National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation, under accession number that is publicly accessible at https://ngdc.cncb.ac.cn/gwh.

References

1. Borrill P, Mago R, Xu T, Ford B, Williams SJ, Derkx A, Bovill WD, Hyles J, Bhatt D, Xia X, Macmillan C, White R, Buss W, Molnár I, Walkowiak S, Olsen OA, Doležel J, Pozniak CJ, Spielmeyer W. 2022. An autoactive NB-LRR gene causes Rht13 dwarfism in wheat. *Proceedings of the National Academy of Sciences of the United States of America* 119**:**e2209875119. DOI: 10.1073/pnas.2209875119.
2. Chen C, Hu H, Guo H, Xia X, Zhang Z, Nong B, Feng R, Liang S, Liu B, Liu J, Li D, Zhao J, Yang X. 2025. Revealing genomic traits and evolutionary insights of Oryza officinalis from Southern China through genome assembly and transcriptome analysis. *Rice* 18**:**15. DOI: 10.1186/s12284-025-00769-5.
3. Chen H, Yao X, Cao B, Zhang B, Lu L, Mao P. 2022. A chromosome-level genome assembly of *Styphnolobium japonicum* combined with comparative genomic analyses offers insights on the evolution of flavonoid and lignin biosynthesis. *Industrial Crops and Products* 187**:**115336. DOI: 10.1016/j.indcrop.2022.115336.
4. Cheng L-T, Wang Z-L, Zhu Q-H, Ye M, Ye C-Y. 2025. A long road ahead to reliable and complete medicinal plant genomes. *Nature Communications* 16**:**2150. DOI: 10.1038/s41467-025-57448-8.
5. Chi Y, Liu C, Liu W, Tian X, Hu J, Wang B, Liu D, Liu Y. 2024. Population genetic variation and geographic distribution of suitable areas of *Coptis* species in China. *Frontiers in Plant Science* 15.

DOI: 10.3389/fpls.2024.1341996.

6. De Vries S, Fürst-Jansen JMR, Irisarri I., Dhabalia A, Ischebeck T, Feussner K, Abreu IN, Petersen M, Feussner I, de Vries J. 2021. The evolution of the phenylpropanoid pathway entailed pronounced radiations and divergences of enzyme families. *The Plant Journal,* 107**,** 975-1002.
7. Fan SUN, Yangwen DU, Xia LI, Boyun LIU, Song Y. 2007. Quantitative assessment of priority for conservation of the National protected plants in Xuebaoshan Nature Reserve*. Natural Science Edition. Journal of Southwest University* 29**:**101–107.
8. Fry RC, Begley TJ, Samson LD. 2005. Genome-wide responses to DNA-damaging agents. *Annual Review of Microbiology* 59**:**357–377. DOI: 10.1146/annurev.micro.59.031805.133658.
9. Ge M, Wei X. 2024a. Spermosphere bacterial community at different germination stages of *Ormosia henryi* and its relationship with seed germination. *Scientia Horticulturae* 324**:**112608. DOI: 10.1016/j.scienta.2023.112608.
10. Gu X, Su Y, Wang T. 2022. Roles of transposable elements on plant genome evolution, epigenetics and adaptation*. Chinese Science Bulletin* 67**:**3024–3035. DOI: 10.1360/TB-2022-0296.
11. Hao JH, Dong CJ, Zhang ZG, Wang XL, Shang QM. 2012. Insights into salicylic acid responses in cucumber (*Cucumis sativus* L.) cotyledons based on a comparative proteomic analysis. *Plant Science: an International Journal of Experimental Plant Biology* 187**:**69–82. DOI: 10.1016/j.plantsci.2012.01.001.
12. Huang S, Xu B. 1986. Chromosome karyotype of *Ormosia henryi*. *Tropical Forestry Science and Technology***,** 14–17.
13. Kress WJ, García-Robledo C, Uriarte M, Erickson DL. 2015. DNA barcodes for ecology, evolution, and conservation. *Trends in Ecology and Evolution* 30**:**25–35. DOI: 10.1016/j.tree.2014.10.008.
14. Lai Y, MA, Ma J, Zhang X, Xuan X, Zhu F, Ding S, Shang F, Chen Y, Zhao B, Lan C, Unver T, Huo G, Li X, Wang Y, Liu Y, Lu M, Pan X, Yang D, Li M, Zhang B, Zhang D. 2024. High-quality chromosome-level genome assembly and multi-omics analysis of rosemary (*Salvia Rosmarinus*) reveals new insights into the environmental and genome adaptation. *Plant Biotechnology Journal* 22**:**1833–1847. DOI: 10.1111/pbi.14305.
15. Li T. 2024. The research on seed germination and seedling drought Resistance of *Ormosia henryi* Prain from different provenances*.* Master.
16. Liu P-P, Yu E-P, Tan Z-J, Sun H-M, Zhu W-G, Wang Z-F, Cao H-L. 2023. Genome assemblies of two *Ormosia* species: gene duplication related to their evolutionary adaptation. *Agronomy* 13**:**1757. DOI: 10.3390/agronomy13071757.
17. Luo L, Ye P, Lin Q, Liu M, Hao G, Wei T, Sahu SK. 2024. From sequences to sustainability: exploring dipterocarp genomes for oleoresin production, timber quality, and conservation. *Plant Science: an International Journal of Experimental Plant Biology* 346**:**112139. DOI: 10.1016/j.plantsci.2024.112139.
18. Peng Y, Zhou S, Zhan X, Gui Z, Liang T, Zhang L, Wan M, Du J, Nie J. 2018. Community characteristics and spatial patterns of *Ormosia henryi* in Youling village of Jiangxi Province. *Journal*

of Central South University of Forestry and Technology 38**:**81–88.

19. Qiu Y, Lu Q, Zhang Y, Cao Y. 2017. Phylogeography of East Asia & apos;s tertiary relict plants: current progress and future prospects. *Biodiversity Science* 25**:**136–146.
20. Salgotra RK, Chauhan BS. 2023. Genetic diversity, conservation, and utilization of plant genetic resources. *Genes* 14:174. DOI: 10.3390/genes14010174.
21. Santini F, Climent JM, Voltas J. 2020. Phenotypic integration and life history strategies among populations of *Pinus halepensis*: an insight through structural equation modelling. *Annals of Botany* 124**:**1161–1172. DOI: 10.1093/aob/mcz088.
22. Song Y, Zhang G, Jia Q, Wen Q, Mo X, Liu L. 2021. Prediction of potential distribution of *Ormosia henryi* in China under climate change. *Journal of Northwest Forestry University* 36**:**108–115, 273.
23. Tao Y, Geng L, Shuting D, Peng L, Jiwang Z, Bin Z. 2016. Proteomic analysis of maize grain development using iTRAQ reveals temporal programs of diverse metabolic processes. *BMC Plant Biology,* 16**,** 241 241:[14pp].
24. Tsumura Y, Uchiyama K, Moriguchi Y, Ueno S, Ihara-Ujino T. 2012. Genome scanning for detecting adaptive genes along environmental gradients in the Japanese conifer, *Cryptomeria japonica*. *Heredity* 109**:**349–360. DOI: 10.1038/hdy.2012.50.
25. Varshney RK, Sinha P, Singh VK, Kumar A, Zhang Q, Bennetzen JL. 2020. 5gs for crop genetic improvement. *Current Opinion in Plant Biology* 56**:**190–196. DOI: 10.1016/j.pbi.2019.12.004.
26. Wang JQ, Li L, Wang ZH, Feng AR, Li HL, Qaseem MF, Liu LT, Deng XM, Wu AM. 2023. Integrative analysis of the metabolome and transcriptome reveals the molecular regulatory mechanism of isoflavonoid biosynthesis in *Ormosia henryi* Prain. *International Journal of Biological Macromolecules* 246. DOI: 10.1016/j.ijbiomac.2023.125601.
27. Wang Y, Sun C, Ye Z, Li C, Huang S, Lin T. 2024. The genomic route to tomato breeding: past, present, and future. *Plant Physiology* 195**:**2500–2514. DOI: 10.1093/plphys/kiae248.
28. Wang Z-F, Yu E-P, Fu L, Deng H-G, Zhu W-G, Xu F-X, Cao H-L. 2025. Chromosome-scale assemblies of three *Ormosia* species: repetitive sequences distribution and structural rearrangement. *GigaScience* 14:giaf047. DOI: 10.1093/gigascience/giaf047.
29. Wu G, Wei X, Wang X, Wei Y. 2023a. Changes and transcriptome regulation of endogenous hormones during somatic embryogenesis in *Ormosia henryi* Prain. *Frontiers in Plant Science,* Volume 14 - 2023 14:1121259. DOI: 10.3389/fpls.2023.1121259.
30. Xiao Y, Tian Y, Deng N. 2024. Geometric morphologic variation *Ormosia henryi* prain leaves from different hunan provenances. *Bangladesh Journal of Botany* 53**:**773–781. DOI: 10.3329/bjb.v53i30.76677.
31. Xu Q, Liu C, Li B, Tian K, You L, Xie L, Wang H, Zhang M, Zhou W, Zhang Y, Zhou C. 2025. A chromosome-scale genome assembly and epigenomic profiling reveal temperature-dependent histone methylation in iridoid biosynthesis regulation in *Scrophularia ningpoensis*. *Horticulture Research* 12:uhae328. DOI: 10.1093/hr/uhae328.

32. Yang K. 2023. Ecological characteristics of the populations of the precious tree species *Ormosia henryi* in Shaowu Jiangshi provincial nature reserve. *Chinese Wild Plant Resources* 42:65–69.
33. Yin K, Chung MY, Lan B, Du FK, Chung MG. 2024. Plant conservation in the age of genome editing: opportunities and challenges. *Genome Biology* 25:279. DOI: 10.1186/s13059-024-03399-0.
34. Zhang D, Yuan W, Chen C, Zhu J, Jiang B. 2003. Preliminary study on growth regularity of man-made *Ormosia henryi* Forest. *Journal of Zhejiang Forestry Science and Technology* 23:9–11, 27.
35. Zhou C, Wen Q, Zeng D, Guo C, Guo Z, Liu L, Ouyang T. 2025. Chromosome-level genome assembly of the endangered tree species *Ormosia henryi* Prain. *Scientific Data* 12:1065. DOI: 10.1038/s41597-025-05402-0.
36. Zhou C, Xia S, Wen Q, Song Y, Jia Q, Wang T, Liu L, Ouyang T. 2023. Genetic structure of an endangered species *Ormosia henryi* in southern China, and implications for conservation. *BMC Plant Biology* 23:220. DOI: 10.1186/s12870-023-04231-w.

Figures

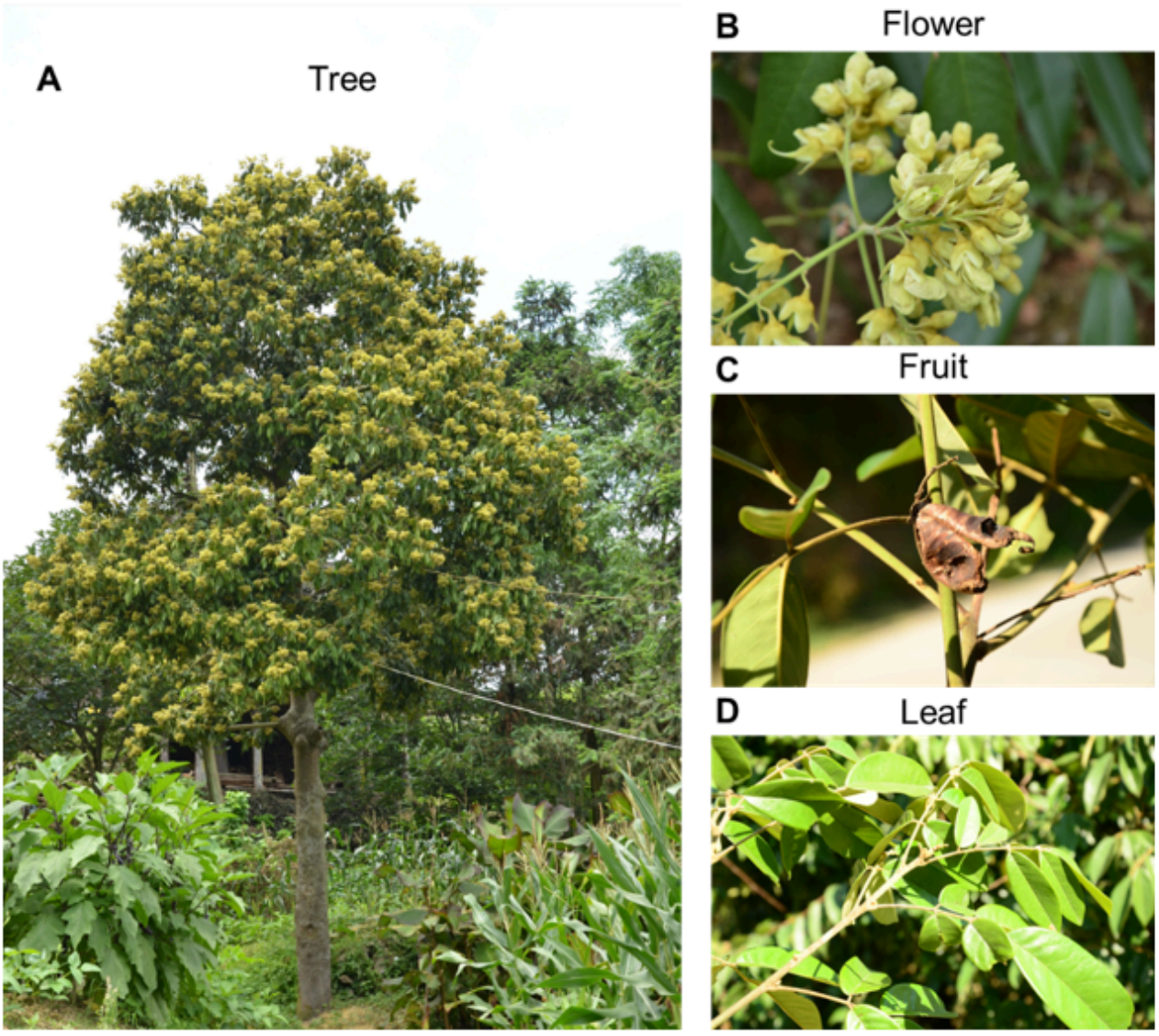

Figure 1

Tree and organs of *Ormosia henryi.* (A) Tree; (B) Flower; (C) Fruit; (D) Leaf.

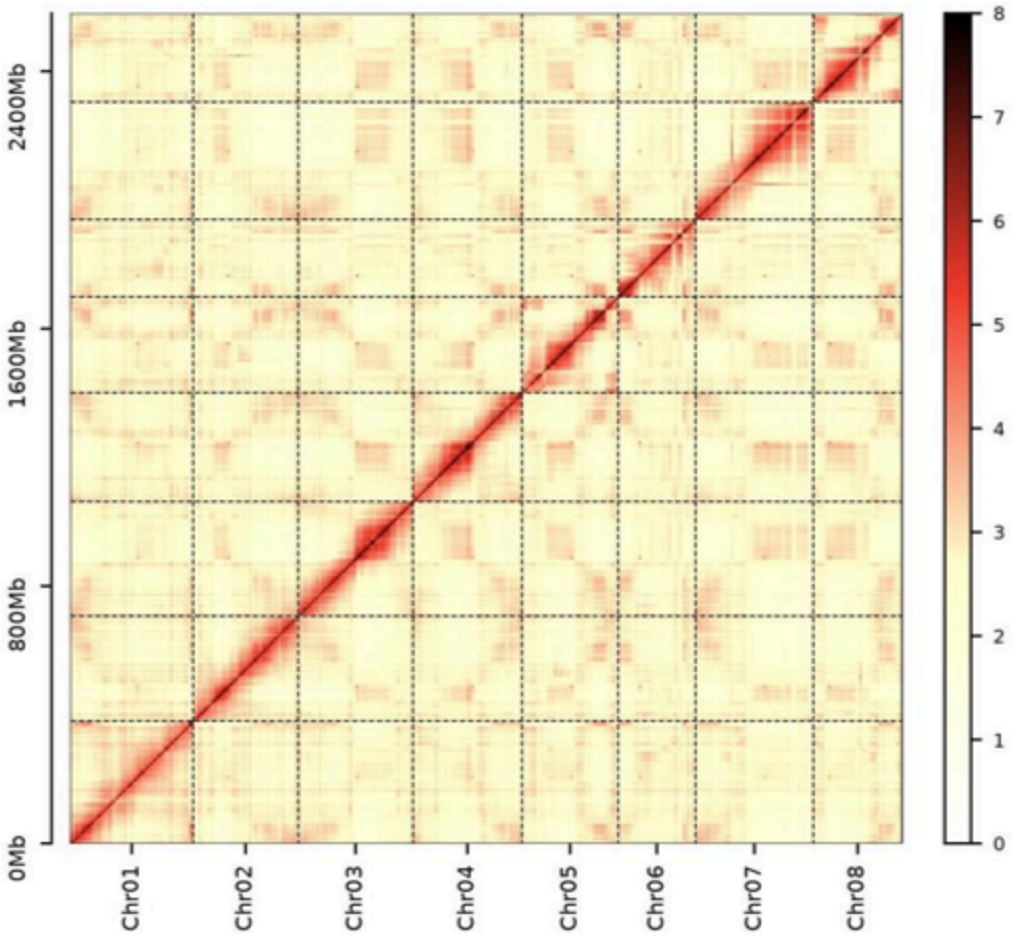

Figure 2

Hi-C heatmap for the genome assembly of *Ormosia henryi* (bin size = 900,000b).

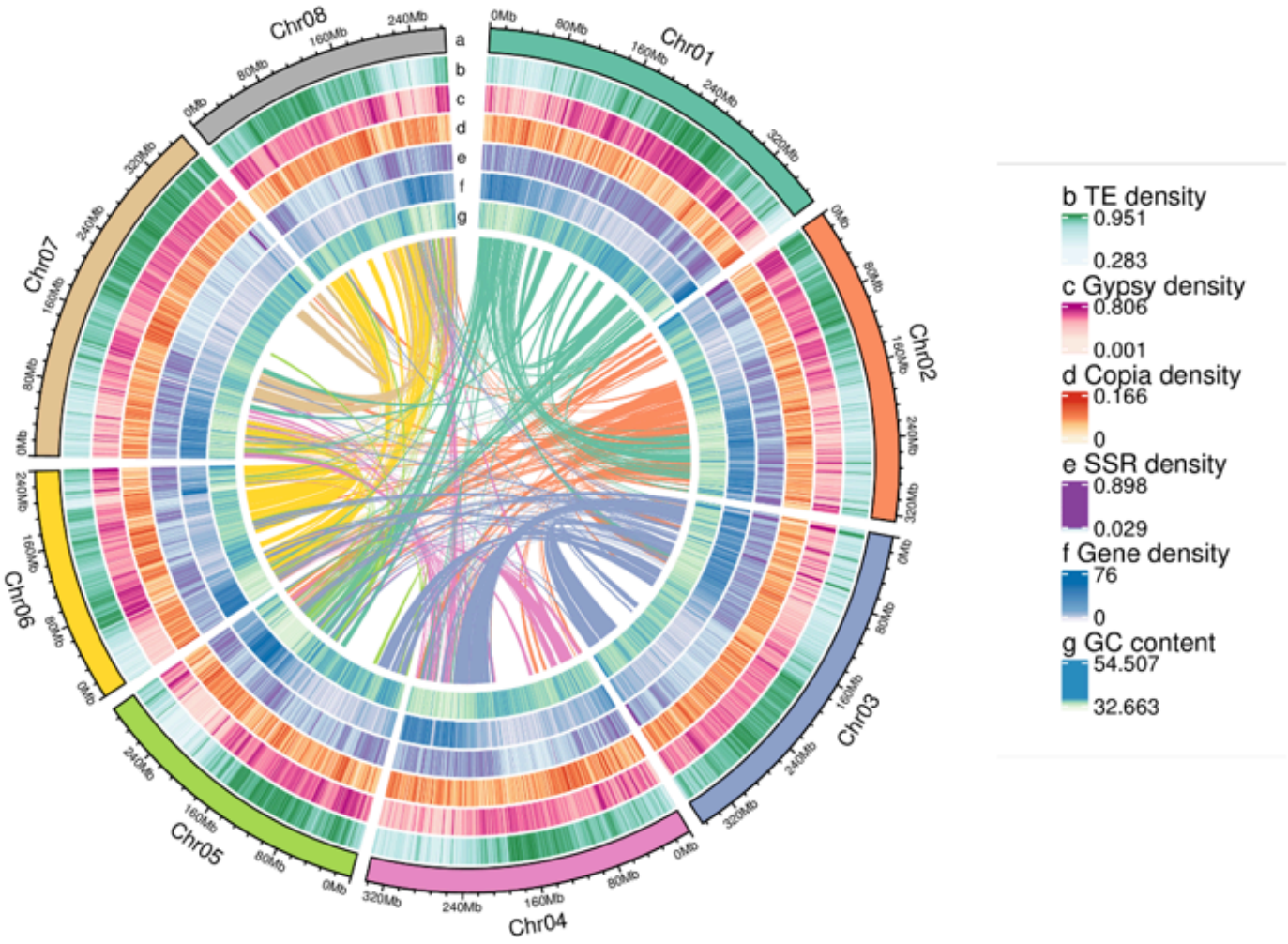

Figure 3

Genomic features of *Ormosia henryi*. The features are arranged in the order of chromosomes, GC content, gene density, SSR density, Copia density, Gypsy density, TE density, and syntenic blocks from outside to inside across the 8 pseudochromosomes.

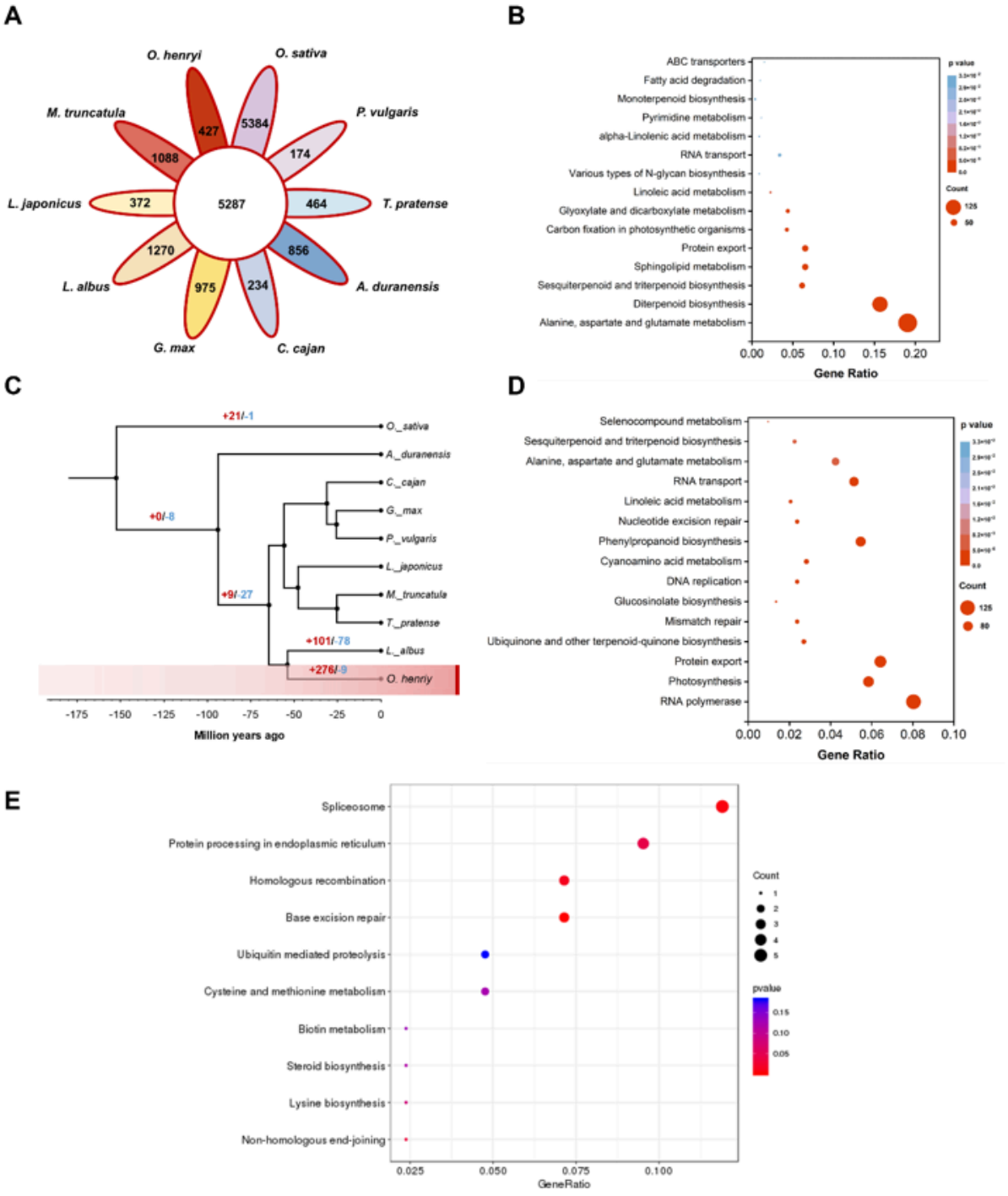

Figure 4

Comparative genomics and evolution. (A) Comparative analysis of gene families of *O. henryi* and nine widely distributed Fabaceae plants (*Trifolium pratense*, *Lupinus albus*, *Glycine max*, *Arachis duranensis*, *Medicago truncatula*, *Lotus japonicus*, *Cajanus cajan*, *Phaseolus vulgaris*); (B) KEGG enrichment of unique gene families to *O. henryi*compared with that of other plants in (A); (C) Evolutionary tree of 10 plant species in (A), *Oryza sativa* as the outgroup; (D) KEGG enrichment of expanded gene families in *O. henryi*; (E) KEGG enrichment of positively selected genes.

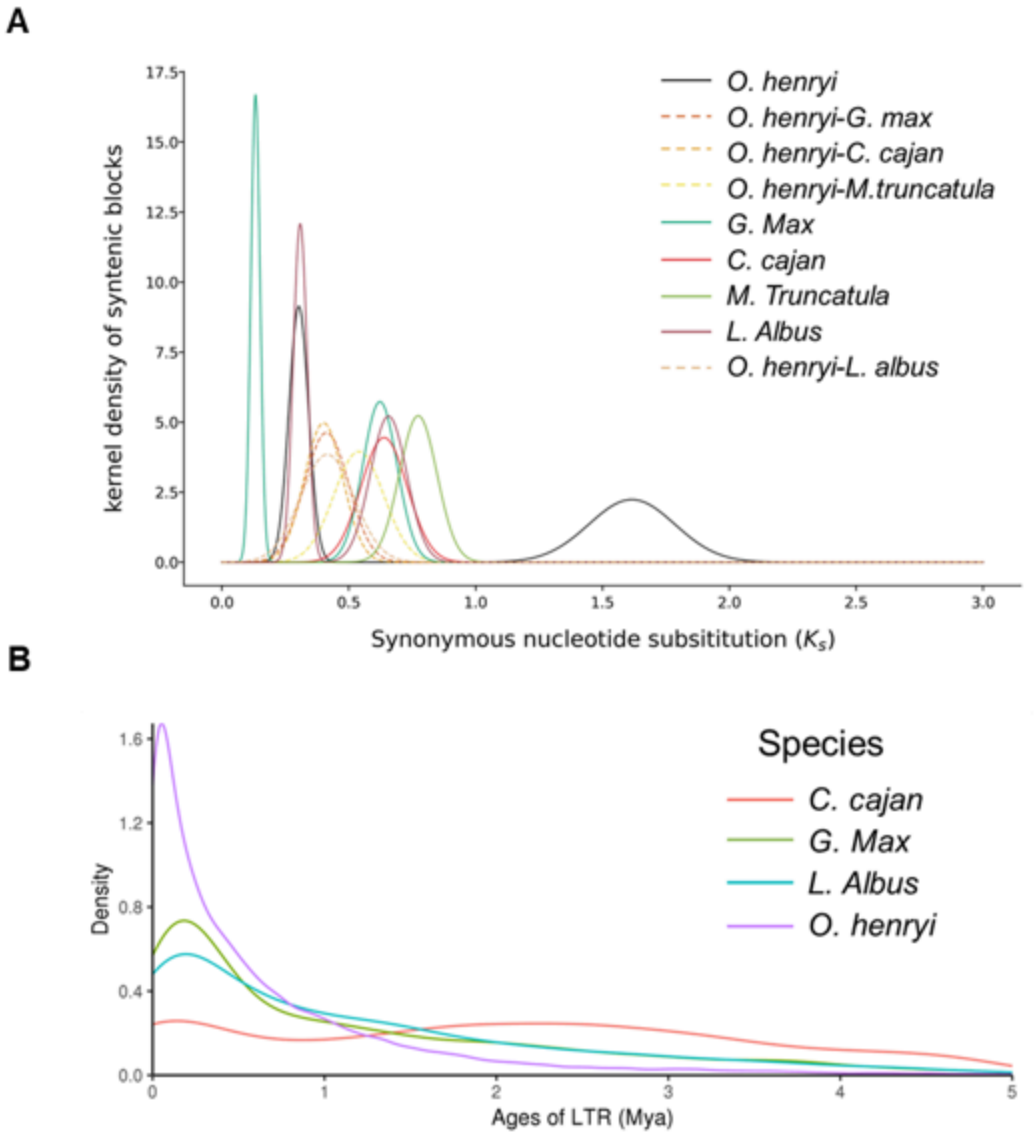

Figure 5

WGD (A) and LTR insertion time analysis (B).