

Systematic Review

Open Access

Comparative Genomics: Insights into the Evolutionary History of *Eucommia* ulmoides

Xin Xie^{1,3}, Atul Kumar Srivastava³, Mehran Khan³, Yichen Zhao² 🗷, Degang Zhao^{1,2} 🗷

1 Plant Conservation and Breeding Technology Center, Guizhou Key Laboratory of Agricultural Biotechnology, Biotechnology Institute of Guizhou Province, Guizhou Academy of Agricultural Sciences, Guiyang, 550006, China

2 The Key Laboratory of Plant Resource Conservation and Germplasm Innovation in Mountainous Region (Ministry of Education), College of Tea Sciences, Guizhou University, Guiyang, 550025, China

3 College of Agriculture, Guizhou University, Guiyang, 550025, China

Co-corresponding authors: <u>yczhao@gzu.edu.cn</u>; dgzhao@gzu.edu.cn

Genomics and Applied Biology, 2024, Vol.15, No.2 doi: <u>10.5376/gab.2024.15.0010</u>

Received: 05 Feb., 2024

Accepted: 08 Mar., 2024

Published: 19 Mar., 2024

Copyright © 2024 Xie et al., This is an open access article published under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preferred citation for this article:

Xie X., Srivastava A.K., Khan M., Zhao Y.C., and Zhao D.G., 2024, Comparative genomics: insights into the evolutionary history of *Eucommia ulmoides*, Genomics and Applied Biology, 15(2): 75-88 (doi: 10.5376/gab.2024.15.0010)

Abstract *Eucommia ulmoides*, commonly known as the hardy rubber tree, is a unique and economically significant tree species with applications in rubber production as well as traditional medicine. This review delves into the evolutionary history of *E. ulmoides* through comparative genomics, highlighting key findings from recent genomic studies. High-quality genome assemblies have revealed significant insights into the genetic architecture and evolutionary mechanisms of this tree. Notably, the genome of *E. ulmoides* has undergone a whole-genome duplication event, contributing to its complex genomic structure and the expansion of gene families involved in rubber biosynthesis and stress responses. Comparative analyses of chloroplast genomes have identified heterogeneous sequence divergence and mutation hotspots, providing valuable information for conservation genetics. Transcriptome studies have uncovered sex-biased gene expression and potential sex-determination genes, shedding light on the genetic basis of sexual dimorphism in this dioecious species. Additionally, high-density genetic maps and QTL analysis have facilitated the identification of growth-related traits, paving the way for genetic improvement and breeding programs. This review presents a comprehensive understanding of the evolutionary history and genomic innovations of *E. ulmoides*, offering new perspectives for its conservation and utilization.

Keywords *Eucommia ulmoides*; Comparative genomics; Whole-genome duplication; Rubber biosynthesis; Chloroplast genome; Sex-biased gene expression; QTL analysis

1 Introduction

Eucommia ulmoides, commonly known as the Hardy Rubber Tree, is a unique and economically significant tree endemic to China. It belongs to the monotypic family *Eucommiaceae* and is renowned for its medicinal and industrial applications. The tree is a source of gutta-percha, a natural rubber, and various secondary metabolites with pharmacological properties, such as aucubin and chlorogenic acid (Li et al., 2020; Liu et al., 2021b). This species has garnered attention not only for its economic value but also for its ecological importance, including applications in landscaping, wind sheltering, and sand fixation (Jin et al., 2020).

The genetic and genomic studies of *E. ulmoides* have revealed its complex evolutionary history. It has undergone multiple whole-genome duplication events, including a recent duplication approximately 27.3 million years ago and an earlier γ paleohexaploidization event (Li et al., 2020; Du et al., 2023). These events have significantly contributed to the expansion of its genome and the diversification of its genetic traits. The high-quality chromosome-level genome assemblies for both male and female *E. ulmoides* have provided valuable resources for understanding the genetic basis of its unique traits, such as sex differentiation and α -linolenic acid biosynthesis (Du et al., 2023).

The primary objective of this study is to conduct a comprehensive comparative genomics analysis of *E. ulmoides* to gain deeper insights into its evolutionary history and genetic diversity. By leveraging high-quality genome assemblies and advanced genomic technologies, this study aims to elucidate the evolutionary events by



investigating the whole-genome duplication events and other significant evolutionary processes that have shaped the genome of E. *ulmoides*. Furthermore, it seeks to understand the genetic basis of traits by identifying and characterizing the genetic loci and pathways involved in key traits such as rubber biosynthesis, growth traits, and secondary metabolite production. Additionally, the study aims to enhance breeding programs by providing genomic resources and markers that can be utilized in marker-assisted selection and breeding programs to develop superior varieties of E. *ulmoides* with improved yield and quality. Lastly, this study contributed to the conservation of this endangered species and explore its potential for industrial and medicinal applications. By achieving these objectives, this study will not only advance our understanding of the genetic and evolutionary mechanisms underlying the unique traits of E. *ulmoides* but also pave the way for its sustainable utilization and conservation.

2 Overview of E. ulmoides

E. ulmoides, or the hardy rubber tree, is a deciduous tree native to China, growing up to 15-20 meters with a broad canopy. Its grayish-brown bark becomes deeply furrowed with age, and its dark green, glossy leaves exude latex when broken. Small, greenish flowers appear in early spring before the leaves, and its fruit is a single-seed samara ripening in autumn. Economically valuable for its natural rubber, it also provides medicinal products. The bark, known as Du Zhong, is used in traditional Chinese medicine to treat hypertension, inflammation, and oxidative stress, and to support bone, muscle, and cardiovascular health (Kim et al., 2009; Jin et al., 2020). Recent research suggests potential anti-diabetic, anti-cancer, and neuroprotective effects. Genetic and molecular studies have identified key genes in the biosynthesis of rubber and bioactive compounds, using techniques like CRISPR/Cas9 for improved production. Study also focuses on genetic diversity and breeding for superior industrial and medicinal traits (Figure 1).



Figure 1 Illustrating the botanical description, economic, medicinal significance, key genetic and molecular research of E. ulmoides

2.1 Botanical description and distribution

E. ulmoides, commonly known as the Hardy Rubber Tree, is a dioecious woody perennial species native to China. It is characterized by its ability to produce rubber and its significant medicinal properties. This specie is typically found in temperate regions and is well-adapted to various environmental conditions, making it a valuable resource for both ecological and economic purposes (Li et al., 2020; Du et al., 2023).



2.2 Economic and medicinal importance

E. ulmoides holds substantial economic and medicinal value. It is widely utilized in traditional Chinese medicine for its pharmacological properties, including anti-inflammatory, anti-hypertensive, and anti-oxidative effects. The tree is also a source of gutta-percha, a form of rubber used in various industrial applications. The biosynthesis of secondary metabolites such as aucubin, chlorogenic acid, and polyphenols in its leaves further enhances its medicinal importance (Li et al., 2020; Liu et al., 2021b; Du et al., 2023). Additionally, the tree is employed in landscaping, wind sheltering, and sand fixation, contributing to its ecological significance (Jin et al., 2020).

2.3 Previous studies on *E. ulmoides*

Several studies have been conducted to understand the genetic and molecular basis of *E. ulmoides*. A genetic linkage map was constructed to identify quantitative trait loci (QTL), affecting growth-related traits, providing a tool for marker-assisted selection and genomic studies (Li et al., 2014; Jin et al., 2020). Genome-wide association studies have revealed genetic loci associated with the biosynthesis of key leaf metabolites, offering insights into the genetic mechanisms underlying these traits (Liu et al., 2021b). High-quality genome assemblies have been developed, shedding light on the evolutionary history, sex differentiation, and rubber biosynthesis pathways of these species (Li et al., 2020; Du et al., 2023). Transcriptome analyses have identified genes related to floral development, further contributing to the understanding of the species' reproductive biology (Liu et al., 2016). These studies collectively provide a comprehensive understanding of the genetic and molecular framework of *E. ulmoides*, facilitating its conservation and improvement for industrial and medicinal applications.

3 Comparative Genomics: Concepts and Methods

3.1 Definition and scope of comparative genomics

Comparative genomics is a field of biological research in which the genomic features of different organisms are compared. This comparison can involve the entire genome or specific parts of the genome, such as genes, regulatory sequences, and non-coding regions. The primary goal is to understand the structure, function, and evolutionary relationships of genomes. Comparative genomics provides insights into the genetic basis of phenotypic differences and similarities among species, and it can reveal the evolutionary processes that shape genomes over time (Filipski and Kumar, 2005; Cordone et al., 2021).

3.2 Techniques and tools for comparative genomic analysis

3.2.1 Sequencing technologies

Advancements in sequencing technologies have been pivotal in the field of comparative genomics. High-throughput sequencing methods, such as PacBio and Illumina sequencing, have enabled the generation of high-quality genome assemblies. For instance, the high-quality haploid chromosome-scale genome assembly of E. *ulmoides* was achieved using PacBio and Hi-C technologies, significantly improving the assembly quality and providing new insights into the species' evolution and rubber biosynthesis (Li et al., 2020). These technologies allow for the detailed comparison of genomic sequences across different species, facilitating the identification of conserved and divergent genomic regions.

3.2.2 Bioinformatics tools and databases

Bioinformatics tools and databases are essential for analyzing and interpreting the vast amounts of data generated by sequencing technologies. Tools such as Genomicus provide a platform for comparative genomic visualization, allowing researchers to explore gene organization and evolutionary relationships across more than 150 eukaryote genomes (Louis et al., 2012). These tools can reveal differential gene loss and gain, segmental or genome duplications, and the evolution of loci through homology relationships. Additionally, databases and software for differential expression analysis, such as those used in the study of chlorogenic acid biosynthesis in *E. ulmoides*, help identify key genes and regulatory elements involved in specific biological pathways (Ye et al., 2019).

3.3 Data sources and selection criteria

The selection of appropriate data sources is crucial for effective comparative genomic analysis. High-quality genome assemblies, such as the one generated for *E. ulmoides*, provide a robust foundation for comparative



studies (Li et al., 2020). The choice of species for comparison should be guided by the research question, phylogenetic relationships, and the availability of genomic data. For example, comparative genomics can be used to study evolutionarily conserved protein functions in marine microbes, providing insights into their adaptation and evolutionary history (Cordone et al., 2021). The inclusion of a diverse range of species, including non-model organisms, enhances the power of comparative analyses to uncover functional genomic elements and evolutionary patterns (Stone et al., 2005). By leveraging advanced sequencing technologies, bioinformatics tools, carefully selected data sources, and comparative genomics continues to expand our understanding of genome structure, function, and evolution across the tree of life.

4 Genome Structure and Organization of E. ulmoides

4.1 Overview of *E. ulmoides* genome

E. ulmoides, a tree species known for its medicinal and industrial applications, has recently been the subject of extensive genomic research. A high-quality haploid chromosome-scale genome assembly was achieved using PacBio and Hi-C technologies, resulting in a significantly improved assembly quality compared to previous versions. The scaffold N50 increased 28-fold to 53.15 MB, and the repetitive sequence content increased by 158.24 MB, while the number of gaps decreased dramatically (Li et. al., 2020). The genome assembly anchored 92.87% of the 26,001 predicted protein-coding genes to 17 chromosomes, providing a robust foundation for further genomic studies (Table 1) (Li et. al., 2020). Additionally, a high-quality chromosome-level female genome was assembled, revealing 31,665 protein-coding genes and confirming two whole-genome duplication events (Du et al., 2023).

Table 1 Statistics for the Eucommia genome and gene annotation (Adopted From Li et. al., 2020)

Assembly	
Estimated genome size	1.02 Gb
Total assembly size	947.84 Mb
Number of contigs	564
N50 of contigs	13.16 Mb
Longest contigs	34.99 Mb
Sequence anchored to the Hi-C map	947.86 Mb
Number of scaffolds after Hi-C assembly	501
N50 of scaffolds after Hi-C assembly	53.15 Mb
Longest scaffold after Hi-C assembly	79.92 Mb
Estimated genome size	1.02 Gb
Total assembly size	947.84 Mb
Annotation	
GC content	0.3517
Number of genes	26001
Percentage of gene length in genome	16.84%
Mean gene length	6138.21
Mean coding sequence length	1108.8
Mean exon number per gene	4.85
Mean exon length	228.56
Mean intron length	1305.93
rRNAs	2099
tRNAs	825
miRNAs	1032
snRNAs	875
Repeat content	62.50%

4.2 Comparative analysis with related species

Comparative genomic analyses have provided insights into the evolutionary history of *E. ulmoides*. The chloroplast genome of *E. ulmoides* was compared with that of *Aucuba japonica*, confirming their sister



relationship within the Garryales order (Wang et al., 2018b). The genome of *E. ulmoides* has undergone two significant whole-genome duplication events: the core eudicot γ whole-genome triplication event and a more recent whole-genome duplication, approximately 27.3 million years ago (Du et al., 2023). These events have contributed to the complexity and expansion of the *E. ulmoides* genome, distinguishing it from related species.

4.3 Genomic features and annotations

The *E. ulmoides* genome exhibits several unique features. The expansion of long terminal repeats has played a significant role in its evolution (Li et al., 2020). The genome contains 21 ribonuclease T2 (EURNS) genes, which are involved in RNA cleavage and have been conserved across eukaryotic organisms (Qing et al., 2021). Additionally, the WRKY gene family, consisting of 45 genes, has been identified and is implicated in various biological processes, including stress responses and leaf development (Wang et al., 2018a). The genome also includes genes related to sex differentiation, such as EuAP3 and EuAG, which are crucial for understanding the dioecious nature of the species (Du et al., 2023). Furthermore, the high expression of ω -3 fatty acid desaturase coding gene EU0103017 is linked to the high α -linolenic acid content in *E. ulmoides* (Du et al., 2023). Overall, the comprehensive genomic data of *E. ulmoides* not only enhances our understanding of its genome structure and organization but also provides valuable resources for future research and breeding programs aimed at improving its industrial and medicinal applications.

Recent advances in genetic research on E. ulmoides have significantly enhanced our understanding of its genome and medicinal properties. High-quality genome sequencing using PacBio and Hi-C technologies has produced a haploid chromosome-scale genome assembly, offering detailed genetic insights (Li et al., 2020). Genome-wide association studies (GWAS) have identified loci linked to the biosynthesis of key metabolites in the leaves, shedding light on the genetic basis of its medicinal properties (Liu et al., 2021). Genetic linkage mapping has pinpointed quantitative trait loci (QTL) for growth traits, aiding marker-assisted selection (Li et al., 2014; Jin et al., 2020). Transcriptome analyses have revealed genes involved in floral development and rubber biosynthesis, enhancing knowledge of its reproductive biology and rubber production (Liu et al., 2016). Comparative genomics, facilitated by tools like Genomicus, has illustrated gene organization, evolutionary relationships, and the mechanisms of gene loss/gain and segmental duplications (Louis et al., 2012). This comparative approach has also provided evolutionary insights into sex differentiation and adaptation mechanisms (Qing et al., 2022). Research on secondary metabolite biosynthesis has focused on compounds like aucubin and chlorogenic acid, crucial for the tree's medicinal uses (Ye et al., 2019; Du et al., 2023). The selection of comparative species based on phylogenetic relationships and available genomic data has enhanced the understanding of functional genomic elements and evolutionary patterns (Cordone et al., 2021). These genetic and molecular studies offer a comprehensive framework for the conservation and improvement of E. ulmoides for industrial and medicinal applications (Table 2) (Filipski and Kumar, 2005; Li et al., 2020).

5 Evolutionary Insights from Comparative Genomics

5.1 Phylogenetic relationships

5.1.1 Methods for phylogenetic analysis

Phylogenetic analysis of *E. ulmoides* has been conducted using various genomic data, including chloroplast genomes and nuclear genomes. Techniques such as genome skimming, which involves sequencing a subset of the genome to obtain high-coverage data for specific regions, have been employed to generate complete chloroplast genomes (Wang et al., 2018b). Additionally, high-quality *de novo* assembly using PacBio and Hi-C technologies has provided comprehensive nuclear genome data, facilitating detailed phylogenetic studies (Li et al., 2020).

5.1.2 Phylogenetic position of E. ulmoides

E. ulmoides, a member of the order Garryales, has been confirmed to have a sister relationship with *Aucuba japonica*, based on chloroplast phylogenomic analyses (Wang et al., 2018b). This phylogenetic positioning is further supported by nuclear genomic data, which places *E. ulmoides* as a sister taxon to lamiids and campanulids, indicating its unique evolutionary lineage within the core eudicots (Wuyun et al., 2017).



Aspect	Description References	
Genome Sequencing	High-quality genome sequencing using PacBio and Hi-C technologies has enabled theLi et al., 2020	
	construction of a haploid chromosome-scale genome assembly for E. ulmoides, providing	
	detailed insights into its genetic makeup.	
Genome-Wide	GWAS has identified genetic loci associated with the biosynthesis of key metabolites inLiu et al., 2021	
Association Studie	esE. ulmoides leaves, enhancing the understanding of the genetic basis of its medicinal	
(GWAS)	properties.	
Genetic Linkag	eGenetic linkage maps have been constructed to identify quantitative trait loci (QTL)Li et al., 2014; Jin	
Mapping	affecting growth-related traits, aiding in marker-assisted selection and genome studies. et al., 2020	
Transcriptome Analys	isTranscriptome analyses have identified genes related to floral development and rubberLiu et al., 2016	
	biosynthesis pathways, contributing to the understanding of the species' reproductive	
	biology and rubber production.	
Comparative Genom	icTools like Genomicus facilitate the visualization of gene organization and evolutionaryLouis et al., 2012	
Visualization	relationships across eukaryote genomes, revealing gene loss/gain, segmental	
	duplications, and homology relationships.	
Evolutionary Insights	Comparative genomics provides insights into the evolutionary history, sex differentiation, Qing et al., 2022	
	and adaptation mechanisms of E. ulmoides, highlighting the evolutionary processes	
	shaping its genome.	
Secondary Metaboli	teStudies have focused on the biosynthesis of secondary metabolites such as aucubin,Ye et al., 2019; Du	
Biosynthesis	chlorogenic acid, and polyphenols, which are crucial for the medicinal properties of E.et al., 2023	
	ulmoides.	
Selection of	ofThe choice of species for comparative analysis is guided by research questions,Cordone et al.,	
Comparative Species	phylogenetic relationships, and the availability of genomic data, enhancing the2021	
	understanding of functional genomic elements and evolutionary patterns.	
Applications of	ofComparative genomics aids in identifying conserved protein functions, adaptationFilipski and Kumar,	
Comparative Genomic	esmechanisms, and evolutionary history, providing a comprehensive framework for the2005; Li et al., 2020	
	conservation and improvement of Eu. ulmoides for industrial and medicinal uses.	

Table 2 Comprehensive comparative genomics analysis of E. ulmoides with relevant references

5.1.3 Evolution and synteny of the E. ulmoides genome

The study of Li et. al. (2020) on *E. ulmoides* genome evolution reveals significant insights into its genomic structure and history (Figure 2). The phylogenetic analysis confirms *E. ulmoides* as part of the lamiid lineage, supporting its classification within the Angiosperm Phylogeny Group system. Synteny analysis with grape and coffee genomes highlights a stable 2:1 synteny ratio, suggesting an additional whole-genome duplication event specific to *E. ulmoides* lineage. This event, superimposed on an earlier paleohexaploidization, significantly shaped its genome. Moreover, the study identifies unique gene families and expanded ones, which likely contribute to the species' environmental adaptability and defense mechanisms. The detailed insertion times for LTR retrotransposons indicate periods of genomic expansion correlating with environmental changes during the Pliocene epoch. Overall, these findings enhance our understanding of the evolutionary processes influencing *E. ulmoides* and provide a robust framework for future genomic and functional studies.

5.2 Gene family evolution

5.2.1 Expansion and contraction of gene families

Comparative genomic studies have revealed significant expansion and contraction of gene families in *E. ulmoides*. For instance, the WRKY transcription factor family, which plays crucial roles in plant development and stress responses, has been extensively studied. A total of 45 WRKY genes were identified and classified into three groups, with further subdivisions based on phylogenetic analysis (Liu et al., 2021a). Similarly, the RNase T2 gene family, involved in RNA cleavage, has shown segmental duplication as the dominant mode of duplication, indicating evolutionary adaptation (Qing et al., 2021).



5.2.2 Functional implications

The expansion of specific gene families in *E. ulmoides* has functional implications for its adaptability and secondary metabolite biosynthesis. For example, the expansion of genes involved in stress responses and secondary metabolite biosynthesis, such as those related to polyisoprene biosynthesis, highlights the species' ability to adapt to environmental stresses and its unique rubber biosynthesis pathway (Wuyun et al., 2017). Additionally, the identification of sex-biased gene expression and putative sex-associated genes provides insights into the genetic mechanisms, underlying sexual dimorphism in this dioecious species (Wang and Zhang, 2017).



Figure 2 Evolution and synteny of the E. ulmoides genome (Adopted From Li et. al., 2020)

Imagine Caption (Adopted From Li et. al., 2020): a. The insertion times for intact LTR-RTs in the *E. ulmoides* genome. The insertion times for LTR-RTs were calculated by the formula T = K/2r. T insertion time; r synonymous mutations/site/Mya; K the divergence between the two LTRs. A substitution rate of $8.25 \times 10-9$ per site per year was used to calculate the insertion times. b. Venn diagram of shared orthologs among the five species. Each number represents a gene family number. c. Phylogenetic tree of 12 species based on orthologs of single-gene families. The number at the root (28,004) represents the number of gene families related to the common ancestor. The value above each branch denotes the number of gene families gained/lost during each round of genome duplication after diversification from the common ancestor. The red number below each branch denotes the speculated divergence time of each node. Bootstrap values for all nodes are above 50%. d. Density distributions of Ks for paralogous genes. The peak values are shown in insets for *E. ulmoides* and *C. canephora*. e. Density distributions of Ks for paralogous genes. The peak values are shown in insets for *E. ulmoides* and *S. lycopersicum*. f. Schematic representation of syntenic genes among *E. ulmoides*, *V. vinifera* and *C. canephora*. Gray lines in the background indicate the collinear blocks of at least twenty genes within the *E. ulmoides* genome and other plant genomes, while the red lines highlight the syntenic gene pairs



5.3 Genome duplication events

5.3.1 Whole-genome duplications

E. ulmoides has undergone significant genome duplication events, which have played a crucial role in its evolutionary history. A notable whole-genome duplication event was identified, superimposed on an earlier γ paleohexaploidization event. This duplication has contributed to the complexity and expansion of the genome, facilitating the evolution of new functions and traits (Li et al., 2020).

5.3.2 Segmental duplications and their evolutionary significance

Segmental duplications have also been a significant factor in the evolution of E. *ulmoides*. These duplications have led to the expansion of specific gene families, such as the RNase T2 family, which has undergone segmental duplication as the dominant mode of duplication (Qing et al., 2021). The evolutionary significance of these duplications lies in their contribution to genetic diversity and the potential for novel gene functions, enhancing the species' adaptability and resilience.

In summary, comparative genomics has provided profound insights into the evolutionary history of *E. ulmoides*, revealing its phylogenetic relationships, gene family evolution, and genome duplication events. These findings not only enhance our understanding of the species' evolutionary trajectory but also have practical implications for its conservation and utilization in various industries.

6 Adaptive Evolution and Natural Selection

6.1 Identification of adaptive genes

The identification of adaptive genes is a crucial step in understanding the evolutionary history of *E. ulmoides*. Adaptive genes are those that have undergone changes due to natural selection, allowing organisms to better survive and reproduce in their environments. Recent advancements in genomic techniques have significantly enhanced our ability to identify these genes. For instance, the integration of natural history collections with comparative genomics has provided a more comprehensive understanding of phenotypic diversity and its genetic underpinnings (Lamichhaney et al., 2019). By combining detailed phenotypic data with genomic information, researchers can better identify genes that have adapted to specific environmental pressures.

6.2 Signatures of positive selection

Positive selection refers to the process where advantageous genetic variants increase in frequency within a population. Identifying signatures of positive selection in the genome of *E. ulmoides* can provide insights into the adaptive changes that have occurred over time. Genomic signatures, which are species-specific patterns in DNA sequences, play a key role in this identification process. These signatures can reveal the statistical properties of DNA sequences that are associated with adaptive traits (De la Fuente et al., 2023). By analyzing these genomic signatures, researchers can pinpoint specific regions of the genome that have been subject to positive selection, thereby uncovering the genetic basis of adaptation.

6.3 Comparative genomic studies on stress response and adaptation

Comparative genomic studies are essential for understanding how different species, including *E. ulmoides*, respond to environmental stress and adapt to their surroundings. These studies involve comparing the genomes of different species to identify common and unique genetic features associated with stress response and adaptation. The integration of natural history collections with comparative genomics has been particularly powerful in elucidating the genomic basis of convergent evolution, where different species independently evolve similar traits (Lamichhaney et al., 2019). This approach allows researchers to explore the genetic mechanisms underlying stress response and adaptation, providing a deeper understanding of how *E. ulmoides* has evolved to cope with various environmental challenges.

By leveraging the synergistic and complementary roles of phenomic data derived from natural history collections and methods of comparative genomics, scientists can gain valuable insights into the evolutionary history and adaptive strategies of *E. ulmoides*.. This integrated approach not only enhances our understanding of the genetic architecture of adaptation but also contributes to the broader field of evolutionary biology.



7 Functional Genomics and Metabolic Pathways

7.1 Key metabolic pathways in E. ulmoides

E. ulmoides, a tree species known for its medicinal and industrial applications, exhibits several key metabolic pathways that are crucial for its unique properties. One of the primary pathways is the methylerythritol-phosphate (MEP) pathway, which is responsible for rubber biosynthesis. Unlike *Hevea brasiliensis*, which uses the mevalonate pathway, *E. ulmoides* predominantly utilizes the MEP pathway to synthesize isoprenyl diphosphate, a precursor for trans-polyisoprene found in its leaves and central peels (Figure 3) (Li et al., 2020). Additionally, the glycolytic pathway in *E. ulmoides* kernels provides cellular energy and intermediates for other biosynthetic processes, playing a significant role in the accumulation of α -linolenic acid (ALA) (Feng et al., 2016). The chlorogenic acid (CGA) biosynthesis pathway is another critical metabolic route, with enzymes preferentially expressed in the leaves (Ye et al., 2019; Li et al., 2020).



Figure 3 The *E. ulmoides* rubber biosynthesis pathway and expression profiles of genes involved in the pathway (Adopted From Li et. al., 2020)

Imagine Caption (Adopted From Li et. al., 2020): The expression level is presented by log2-transformed fragments mapped per kilobase of transcript length per million total mapped reads (log2-FPKM). ACAT acetyl-coenzyme A (CoA) C-acetyltransferase; HMGS hydroxymethylglutaryl-CoA synthase; HMGR hydroxymethylglutaryl-CoA reductase; MVK mevalonate kinase; PMK 5-phosphomevalonate kinase; MPD mevalonate pyrophosphate decarboxylase; DXS 1-deoxy-d-xylulose 5-phosphate synthase; DXR 1-deoxy-d-xylulose 5-phosphate reductoisomerase; MCT 2-C-methyl-d-erythritol 4-phosphate cytidylyltransferase; CMK 4-(cytidine 5'-diphospho)-2-C-methyl-d-erythritol kinase; MDS 2-C-methyl-d-erythritol 2,4-cyclodiphosphate HDS synthase: 4-hydroxy-3-methylbut-2-enyl diphosphate synthase; HDR 4-hydroxy-3-methylbut-2-enyl diphosphate reductase; IDI isopentenyl diphosphate isomerase; GPS geranyl diphosphate synthase; FPS farnesyl diphosphate synthase; GGPS, geranyl geranyl diphosphate synthase; SRPP small rubber particle protein; Acetyl-CoA acetyl coenzyme-A; Acetoacetyl-CoA, 3-acetoacetyl-CoA; HMG-CoA, 3-hydroxy-3-methylglutaryl-CoA; MVA mevalonate; MVA-5P mevalonate-5-phosphate; MVA-5PP mevalonate-5-diphosphate; GA-3-P, glyceraldehyde 3-phosphate; DXP, 1-deoxy-d-xylulose 5-phosphate; MEP 2-C-methyl-d-erythritol 4-phosphate; CME 4-(cytidine 5'-diphospho)-2-C-methyl-d-erythritol; PCME 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-d-erythritol; CMEC 2-C-methyl-d-erythritol 2,4-cyclodiphosphate; HMED 4-hydroxy-3-methylbut-2-enyl diphosphate; IPP isopentenyl diphosphate; DMAPP, dimethylallyl diphosphate; GPP geranyl diphosphate; FPP farnesyl diphosphate; GGPP, geranylgeranyl diphosphate. LF leaf; CP central peel; PE peel edge; XM xylem; SD seed



7.2 Comparative analysis of secondary metabolite biosynthesis

Secondary metabolites in *E. ulmoides*, such as chlorogenic acid, aucubin, and various glycosides, are of significant interest due to their pharmacological activities. Comparative transcriptome analyses have identified key genes involved in the biosynthesis of CGA, including PALs, 4CL, and several transcription factors like ERF2 and WRKY (Ye et al., 2019). The uridine diphosphate glycosyltransferases (UGTs) family also plays a crucial role in the biosynthesis of glycosides, catalyzing glycosyl-transferring reactions essential for the formation of pharmacologically active compounds (Ouyang et al., 2021). Furthermore, the WRKY transcription factors, which are involved in multiple biological processes, have been shown to regulate the expression of genes related to secondary metabolite biosynthesis (Liu et al., 2021a).

7.3 Insights into unique functional traits

E. ulmoides exhibits several unique functional traits that distinguish it from other species. The high-quality *de novo* assembly of its haploid genome has revealed a new whole-genome duplication event and the expansion of long terminal repeats, contributing to its evolutionary history and functional diversity (Li et al., 2020). The identification of 21 ribonuclease T2 (RNase T2) genes, which are involved in RNA cleavage and stress responses, highlights the plant's adaptation mechanisms (Qing et al., 2021). Additionally, the construction of a high-density genetic map and the identification of quantitative trait loci (QTLs) associated with growth traits provide insights into the genetic basis of its growth and development (Jin et al., 2020; Liu et al., 2022). These findings underscore the potential of *E. ulmoides* for genetic improvement and its application in various industries.

8 Applications and Implications of Comparative Genomics

8.1 Conservation genetics and biodiversity

Comparative genomics has significant implications for conservation genetics and biodiversity. By comparing the genomic data of *E. ulmoides* with other species, researchers can identify genetic variations and evolutionary patterns that are crucial for the conservation of this endangered species. For instance, the study of the chloroplast genome of *E. ulmoides* revealed heterogeneous sequence divergence patterns and identified potential molecular markers for population genetics studies, which are essential for conservation efforts (Wang et al., 2018b). Additionally, the high level of genetic diversity observed within *E. ulmoides* populations, despite a genetic bottleneck in one population, underscores the importance of maintaining genetic variation for the species' long-term survival (Zhang et al., 2013). Comparative genomic approaches can also help in understanding the processes that influence the diversity of *E. ulmoides*, thereby aiding in the development of effective conservation strategies (Grueber, 2015).

8.2 Genetic improvement and breeding

The insights gained from comparative genomics are invaluable for the genetic improvement and breeding of *E. ulmoides*. High-density genetic maps and QTL analyses have identified numerous markers and candidate genes associated with growth traits, which can be utilized in marker-assisted selection to enhance breeding programs (Li et al., 2014; Jin et al., 2020; Liu et al., 2022). For example, the construction of a high-density genetic map using SNP markers has provided a solid foundation for chromosome assembly and the localization of growth-trait QTLs, advancing forestry breeding efforts in *E. ulmoides* (Liu et al., 2022). Moreover, the identification of key genes involved in chlorogenic acid biosynthesis through comparative transcriptome analysis offers potential targets for genetic engineering to improve the medicinal value of *E. ulmoides* (Ye et al., 2019). The integration of these genomic resources can significantly accelerate the breeding of superior varieties with desired traits.

8.3 Medicinal and industrial applications

Comparative genomics also plays a crucial role in enhancing the medicinal and industrial applications of *E. ulmoides*. The high-quality genome assembly of *E. ulmoides* has provided new insights into rubber biosynthesis and the biosynthesis pathways of other valuable compounds such as chlorogenic acid and α -linolenic acid (Li et al., 2020; Du et al., 2023). Understanding these biosynthetic pathways at the genomic level allows for the optimization of metabolic engineering strategies to increase the yield of these compounds. For instance, the discovery that *E. ulmoides* relies on the methylerythritol-phosphate pathway for rubber biosynthesis, as opposed



to the mevalonate pathway used by *Hevea brasiliensis*, opens up new avenues for enhancing rubber production through genetic modifications (Li et al., 2020). Additionally, the identification of genes regulating sex differentiation and α -linolenic acid biosynthesis provides opportunities to improve the quality and consistency of *E. ulmoides* products for medicinal and industrial use (Du et al., 2023). These advancements underscore the potential of comparative genomics to drive innovation in the utilization of *E. ulmoides* for various applications.

9 Challenges and Future Directions

9.1 Technical and methodological challenges

The field of comparative genomics, particularly in non-model organisms like *E. ulmoides*, faces several technical and methodological challenges. One significant challenge is the generation of high-quality genome assemblies. Although recent advancements have been made, such as the high-quality haploid chromosome-scale genome assembly using PacBio and Hi-C technologies (Li et al., 2020), the complexity of plant genomes, including high levels of repetitive sequences and polyploidy, continues to pose difficulties. Additionally, the identification and annotation of sex-biased genes and sex-associated genes through comparative transcriptome analyses have revealed the complexity of genetic mechanisms underlying sexual dimorphism in *E. ulmoides* (Wang and Zhang, 2017). Another challenge is the accurate detection and interpretation of sequence divergence and mutation hotspots, as seen in the chloroplast genome of *E. ulmoides*, which requires comprehensive genome-scale comparisons (Wang et al., 2018b).

9.2 Future prospects in comparative genomics research

Future research in comparative genomics of *E. ulmoides* holds promising prospects. The development of high-density genetic maps using genotyping-by-sequencing (GBS) and SNP markers has laid a solid foundation for chromosome assembly and the localization of growth-trait QTLs (Liu et al., 2022). This advancement will facilitate the investigation of genetic mechanisms underlying important traits and improve breeding efforts. Moreover, the identification of candidate genes related to chlorogenic acid biosynthesis through comparative transcriptome analysis provides valuable genomic resources for genetic improvements (Ye et al., 2019). The integration of these genomic resources with phenotypic data, such as the phenotypic variation in leaf, fruit, and seed traits, will enhance our understanding of the relationship between genetic diversity and phenotypic traits (Wang et al., 2023).

9.3 Integrative approaches combining multi-omics data

To overcome the challenges and fully realize the future prospects in comparative genomics research, integrative approaches that combine multi-omics data are essential. The integration of genomics, transcriptomics, and epigenomics data can provide a comprehensive understanding of the genetic and epigenetic regulation of important traits. For instance, the study of the RNase T2 gene family in *E. ulmoides*, which involved the analysis of gene structure, chromosomal location, and expression patterns, highlights the importance of combining different types of omics data (Qing et al., 2021). Additionally, the use of molecular markers and QTL analysis for growth traits, as demonstrated in the updated genetic linkage map of *E. ulmoides*, underscores the need for integrative approaches to elucidate the genetic mechanisms underlying complex traits (Jin et al., 2020). By leveraging multi-omics data, researchers can gain deeper insights into the evolutionary history and genetic improvement of *E. ulmoides*.

10 Concluding Remarks

The comparative genomics studies of *E. ulmoides* have provided significant insights into its evolutionary history and genetic makeup. The high-quality haploid genome assembly revealed a new whole-genome duplication event and highlighted the primitive rubber biosynthesis pathway unique to *E. ulmoides*, which relies on the methylerythritol-phosphate pathway rather than the mevalonate pathway. Additionally, the construction of genetic linkage maps and QTL analyses have identified numerous quantitative trait loci associated with growth traits, providing valuable tools for marker-assisted selection and genetic improvement. The chromosome-level genome assembly has also shed light on sex differentiation mechanisms and the biosynthesis of α -linolenic acid, with key genes identified for these processes. Furthermore, chloroplast genome studies have revealed heterogeneous divergence and mutation hotspots, contributing to our understanding of the genetic diversity and evolutionary relationships within the species.



The findings from these genomic studies have profound implications for evolutionary biology. The identification of whole-genome duplication events and the expansion of long terminal repeats suggest significant evolutionary pressures and adaptations in *E. ulmoides*. The unique rubber biosynthesis pathway and the high α -linolenic acid content highlight the species' specialized metabolic adaptations, which could be a result of its long evolutionary history and ecological niche. The detailed genetic linkage maps and QTL analyses provide a framework for understanding the genetic basis of important traits, facilitating evolutionary studies on trait selection and adaptation. The chloroplast genome analyses further enhance our understanding of the species' genetic diversity and evolutionary relationships, which are crucial for conservation and breeding programs.

The comprehensive genomic data obtained from these studies mark a significant advancement in our understanding of *E. ulmoides*. These findings not only elucidate the evolutionary history and genetic diversity of the species but also provide valuable resources for future research and breeding programs. Future studies should focus on functional genomics to explore the roles of identified genes in trait development and adaptation. Additionally, integrating genomic data with ecological and phenotypic data will further enhance our understanding of the evolutionary dynamics of *E. ulmoides*. The development of molecular markers and genetic tools will also facilitate the conservation and sustainable utilization of this economically and medicinally important species. Overall, the insights gained from these genomic studies pave the way for innovative approaches in evolutionary biology, conservation, and genetic improvement of *E. ulmoides*.

Acknowledgments

The authors thank the two anonymous peer reviewers for their comments and suggestions on the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China [31870285, 31660076] and the Guizhou Academy of Agricultural Sciences Talent Special Project [2022-02 & 2023-02].

Conflict of Interest Disclosure

The authors affirm that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Cordone A., Coppola A., Severino A., Correggia M., Selci M., Cascone A., Vetriani C., and Giovannelli D., eds., 2021, From sequences to enzymes: comparative genomics to study evolutionarily conserved protein functions in marine microbes, In Marine Genomics: Methods and Protocols, Springer, New York, NY, US, pp.77-88.

https://doi.org/10.31219/osf.io/yacvr

- De la Fuente R., Díaz-Villanueva W., Arnau V., and Moya A., 2023, Genomic signature in evolutionary biology: a review. Biology, 12(2): 322. https://doi.org/10.3390/biology12020322
- Du Q.X., Wu Z.X., Liu P.F., Qing J., He F., Du L.Y., Sun Z.Q., Zhu L.L., Zheng H.C., Sun Z.Y., Yang L., Wang L., and Du H.Y., 2023, The chromosome-level genome of Eucommia ulmoides provides insights into sex differentiation and α-linolenic acid biosynthesis, Frontiers in Plant Science, 14: 1118363. https://doi.org/10.3389/fpls.2023.1118363
- Feng Y.Z., Zhang L., Fu J.M., Li F.D., Wang L., Tan X.F., Mo W.J., and Cao H.P., 2016, Characterization of glycolytic pathway genes using rna-seq in developing kernels of Eucommia ulmoides, Journal of Agricultural and Food Chemistry, 64(18): 3712-3731. <u>https://doi.org/10.1021/acs.jafc.5b05918</u>
- Filipski A., and Kumar S., 2005, Comparative genomics in eukaryotes, in: the evolution of the genome, Academic Press, 521-583. https://doi.org/10.1016/B978-012301463-4/50011-5
- Filipski A., and Kumar S., 2005, Comparative genomics in eukaryotes, In The evolution of the genome, Academic Press, pp.521-583.
- Grueber C.E., 2015, Comparative genomics for biodiversity conservation, Computational and Structural Biotechnology Journal, 13: 370-375. https://doi.org/10.1016/j.csbj.2015.05.003
- Jin C.F., Li Z.Q., Li Y., Wang S.H., Li L., and Liu M.H., 2020, Update of genetic linkage map and QTL analysis for growth traits in Eucommia ulmoides oliver, Forests,11(3): 311.

https://doi.org/10.3390/f11030311



- Jin C., Li Z., Li Y., Wang S., Li L. and Liu M., 2020, Update of genetic linkage map and QTL analysis for growth traits in *Eucommia ulmoides* Oliver, Forests, 11(3): 311.
- Jin C., Li Z., Li Y., Wang S., Li L., Liu M. and Ye J., 2020, Transcriptome analysis of terpenoid biosynthetic genes and simple sequence repeat marker screening in Eucommia ulmoides, Molecular Biology Reports, 47: 1979-1990.
- Kim B.H., Park K.S. and Chang I.M., 2009, Elucidation of anti-inflammatory potencies of Eucommia ulmoides bark and Plantago asiatica seeds, Journal of medicinal food, 12(4): 764-769.
- Lamichhaney S., Card D.C., Grayson P., Tonini J.F.R., Bravo G.A., Näpflin K., Termignoni-García F., Torres C., Burbrink F., Clarke J.A., Sackton T.B., and Edwards S.V., 2019, Integrating natural history collections and comparative genomics to study the genetic architecture of convergent evolution, Philosophical Transactions of the Royal Society B, 374(1777): 20180248. https://doi.org/10.1098/rstb.2018.0248
- Li Y., Wang D.W., Li Z.Q., Wei J.K., Jin C.F., and Liu M.H., 2014, A molecular genetic linkage map of Eucommia ulmoides and quantitative trait loci (QTL) analysis for growth traits, International Journal of Molecular Sciences, 15(2): 2053-2074. https://doi.org/10.3390/ijms15022053
- Li Y., Wei H.R., Yang J., Du K., Li J., Zhang Y., Qiu T., Liu Z., Ren Y.Y., Song L.J., and Kang X.Y., 2020, High-quality de novo assembly of the Eucommia ulmoides haploid genome provides new insights into evolution and rubber biosynthesis, Horticulture Research, 7: 183. https://doi.org/10.1038/s41438-020-00406-w
- Li Y., Wang D., Li Z., Wei J., Jin C. and Liu M., 2014, A molecular genetic linkage map of Eucommia ulmoides and quantitative trait loci (QTL) analysis for growth traits, International Journal of Molecular Sciences, 15(2): 2053-2074.
- Li Y., Wei H., Yang J., Du K., Li J., Zhang Y., Qiu T., Liu Z., Ren Y., Song L. and Kang X., 2020, High-quality de novo assembly of the Eucommia ulmoides haploid genome provides new insights into evolution and rubber biosynthesis, Horticulture Research, 7.
- Liu C.L., Wang L., Lu W.J., Zhong J., Du H.Y., Liu P.F., Du Q.X., Du L.Y., and Qing J., 2022, Construction of SNP-based high-density genetic map using genotyping by sequencing (GBS) and QTL analysis of growth traits in *Eucommia ulmoides* Oliv, Forests, 13(9): 1479. <u>https://doi.org/10.3390/f13091479</u>
- Liu H.M., Fu J.M., Du H.Y., Hu J.J., and Wuyun T.N., 2016, De novo sequencing of Eucommia ulmoides flower bud transcriptomes for identification of genes related to floral development, Genomics Data, 9: 105-110. <u>https://doi.org/10.1016/j.gdata.2016.07.001</u>
- Liu J., Wang X.Y., Chen Y.L., Liu Y., Wu Y.S., Ren S.S., and Li L., 2021a, Identification, evolution and expression analysis of WRKY gene family in Eucommia ulmoides, Genomics, 113(5): 3294-3309. https://doi.org/10.1016/j.ygeno.2021.05.011
- Liu Y.C., Peng X.X., Lu Y.B., Wu X.X., Chen L.W., and Feng H., 2021b. Genome-wide association study reveals the genes associated with the leaf inclusion contents in Chinese medical tree Eucommia ulmoides, Bioscience, Biotechnology, and Biochemistry, 85(2): 233-241. https://doi.org/10.1093/bbb/zbaa005
- Liu H., Fu J., Du H., Hu J. and Wuyun T., 2016, De novo sequencing of Eucommia ulmoides flower bud transcriptomes for identification of genes related to floral development, Genomics Data, 9: 105-110.
- Liu Y.C., Peng X.X., Lu Y.B., Wu X.X., Chen L.W. and Feng H., 2021, Genome-wide association study reveals the genes associated with the leaf inclusion contents in Chinese medical tree *Eucommia ulmoides*, Bioscience, Biotechnology, and Biochemistry, 85(2): 233-241.
- Louis A., Muffato M., and Crollius H.R., 2012, Genomicus: five genome browsers for comparative genomics in eukaryota, Nucleic Acids Research, 41(D1): D700-D705.

https://doi.org/10.1093/nar/gks1156

- Louis A., Muffato M. and Roest Crollius H., 2012, Genomicus: five genome browsers for comparative genomics in eukaryota, Nucleic acids research, 41(D1): D700-D705.
- Ouyang D., Wang L.C., Tang T., and Feng H., 2021, Genomic-wide identification and characterization of the uridine diphosphate glycosyltransferase family in *Eucommia ulmoides* oliver, Plants, 10(9): 1934. https://doi.org/10.3390/plants10091934
- Qing J., Du Q.X., Meng Y.D., Liu P.F., Du H.Y., and Wang L., 2021, Genome-wide identification and expression pattern analysis of the ribonuclease T2 family in *Eucommia ulmoides*, Scientific Reports, 11: 6900.

https://doi.org/10.1038/s41598-021-86337-5

- Qing J., Meng Y.D., He F., Du Q.X., Zhong J., Du H.Y., Liu P.F., Du L.Y. and Wang L., 2022, Whole genome re-sequencing reveals the genetic diversity and evolutionary patterns of Eucommia ulmoides, Molecular Genetics and Genomics, 297(2): 485-494.
- Stone E.A., Cooper G.M., and Sidow A., 2005, Trade-offs in detecting evolutionarily constrained sequence by comparative genomics, Annual review of genomics and human genetics, 6: 143-164.

https://doi.org/10.1146/ANNUREV.GENOM.6.080604.162146

Wang C.C., Gong H.M., Feng M., and Tian C.L., 2023, Phenotypic variation in leaf, fruit and seed traits in natural populations of eucommia ulmoides, a relict chinese endemic tree, Forests, 14(3): 462.

https://doi.org/10.3390/f14030462

Wang L., Du H.Y., Li T.Z., and Wuyun T.N., 2018a, De novo transcriptome sequencing and identification of genes related to salt stress in Eucommia ulmoides Oliver, Trees, 32: 151-163.

https://doi.org/10.1007/s00468-017-1620-9



- Wang W.C., and Zhang X.Z., 2017, Identification of the sex-biased gene expression and putative sex-associated genes in *Eucommia ulmoides* oliver using comparative transcriptome analyses, Molecules: A Journal of Synthetic Chemistry and Natural Product Chemistry, 22(12): 2255. <u>https://doi.org/10.3390/molecules22122255</u>
- Wang W.C., Chen S.Y., and Zhang X.Z., 2018b, Whole-genome comparison reveals heterogeneous divergence and mutation hotspots in chloroplast genome of Eucommia ulmoides oliver, International Journal of Molecular Sciences, 19(4): 1037. <u>https://doi.org/10.3390/ijms19041037</u>
- Wuyun T.N., Wang L., Liu H.M., Wang X.W., Zhang L.S., Bennetzen J.L., Li T.Z., Yang L.R., Liu P.F., Du L.Y., Wang L., Huang M.Z., Qing J., Zhu L.L., Bao W.Q., Li H.G., Du Q.X., Zhu J.L., Yang H., Yang S.G., Liu H., Yue H., Hu J., Yu G.L., Tian Y., Liang F., Hu J.J., Wang D.P., Gao R.W., Li D.J., and Du H.Y., 2017, The hardy rubber tree genome provides insights into the evolution of polyisoprene biosynthesis, Molecular Plant, 11(3): 429-442. https://doi.org/10.1016/j.molp.2017.11.014
- Ye J., Han W.J., Deng P., Jiang Y.Y., Liu M.H., Li L., and Li Z.Q., 2019, Comparative transcriptome analysis to identify candidate genes related to chlorogenic acid biosynthesis in *Eucommia ulmoides* Oliv, Trees, 33: 1373-1384. <u>https://doi.org/10.1007/s00468-019-01865-y</u>
- Ye J., Han W., Deng P., Jiang Y., Liu M., Li L. and Li Z., 2019, Comparative transcriptome analysis to identify candidate genes related to chlorogenic acid biosynthesis in *Eucommia ulmoides* Oliv, Trees, 33: 1373-1384.
- Zhang J.J., Xing C., Tian H., and Yao X.H., 2013, Microsatellite genetic variation in the Chinese endemic Eucommia ulmoides (*Eucommiaceae*): implications for conservation, Botanical Journal of the Linnean Society, 173: 775-785. <u>https://doi.org/10.1111/BOJ.12116</u>



Disclaimer/Publisher's Note

The statements, opinions, and data contained in all publications are solely those of the individual authors and contributors and do not represent the views of the publishing house and/or its editors. The publisher and/or its editors disclaim all responsibility for any harm or damage to persons or property that may result from the application of ideas, methods, instructions, or products discussed in the content. Publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.