# Integrating QTL Mapping and Genomic Selection in *Eucommia ulmoides* Breeding

Ruoruo Wang [1], Luo Wang [1], Degang Zhao [1,2] ✉

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

2 The Key Laboratory of Plant Resources Conservation and Germplasm Innovation in Mountainous Region (Ministry of Education), College of Life Sciences / Institute of Agro- Bioengineering, Guizhou University, Guiyang, 550025, Guizhou, China

✉ Corresponding email: dgzhao@gzu.edu.cn

**Abstract** *Eucommia ulmoides* is a tree species with significant medicinal and industrial value. In recent years, some progress has been made in the genetic improvement of *E. ulmoides* through the construction of high-density genetic maps and the identification of massive quantitative trait loci (QTLs). This study focuses on the integrated application of QTL mapping and genomic selection (GS) in the breeding of *E. ulmoides*. By integrating current researches, the study analyzes the role of QTL mapping in identifying loci associated with growth traits, secondary metabolites, and yield, and discusses the potential of applying this information in GS models to enhance breeding efficiency. The results indicate that combining QTL mapping with GS methods can significantly improve the accuracy of selecting complex traits and help accelerate the breeding of superior *Eucommia* varieties. This study provides theoretical support for future genetic research and breeding strategies in *E. ulmoides*, promoting the development of varieties with higher economic and ecological value.

**Keywords** *Eucommia ulmoides*; QTL mapping; Genomic selection; Genetic linkage map; Breeding efficiency

## 1 Introduction

*Eucommia ulmoides*, commonly known as the hardy rubber tree, is a significant tree species with both economic and ecological value. It is prized for its applications in pharmacology, landscaping, wind protection, and sand stabilization (Jin et al., 2020). The tree is notable for producing Eucommia rubber (Eu-rubber), a trans-polyisoprene that serves as a unique strategic resource in China and can be used as a substitute for natural rubber (Wuyun et al., 2017; Zhang et al., 2023). Additionally, *E. ulmoides* is rich in chlorogenic acid and flavonoids, both of which possess high medicinal value (Xie et al., 2023). Endemic to China, this species has been classified as a national level II key protected tree species due to its economic and ecological significance (Wang et al., 2023).

Despite its significance, the breeding and genetic improvement of *E. ulmoides* face several challenges. One major issue is the inadequate comprehension of the genetic basis of key traits, which limits efforts to breed varieties with enhanced yield and quality (Li et al., 2014). The dioecious nature of the species complicates early-stage sex identification, hindering efficient breeding practices (Wang et al., 2020). Additionally, the lack of a high-quality genome sequence has restricted both fundamental biological research and applied studies (Wuyun et al., 2017; Li et al., 2020; Du et al., 2023). Furthermore, the phenotypic variation within and among natural populations poses a challenge for genetic improvement and resource management (Wang et al., 2023).

Quantitative trait loci (QTLs) mapping is a method used to identify regions of the genome associated with specific phenotypic traits. This technique is crucial for understanding the genetic architecture of complex traits and for identifying candidate genes that influence these traits. In *E. ulmoides*, QTL mapping has been employed to identify loci linked to growth traits, which is essential for improving breeding efficiency and genetic improvement at the molecular level (Li et al., 2014; Jin et al., 2020; Liu et al., 2022). GS is a modern breeding approach that utilizes genome-wide markers to predict the genetic value of individuals in a breeding population. Unlike

traditional marker-assisted selection, GS considers the effects of all markers across the genome, making it more efficient for selecting complex traits. Integration GS into *E. ulmoides* breeding can significantly accelerate the breeding cycle and enhance the accuracy of selection for desirable traits (Liu et al., 2022). This approach is particularly advantageous for perennial species like *E. ulmoides*, where long generation times can impede breeding progress.

This study explores the application of combining QTL mapping with GS in the breeding and genetic improvement of *E. ulmoides*. By integrating and analyzing existing research, we sort out candidate genes and loci linked to key growth traits and assesses the potential of genomic selection to accelerate the breeding process and enhance the accuracy of selecting complex traits. This study aims to provide theoretical support for the genetic research and breeding strategies of *E. ulmoides*, contributing to the development of superior varieties with higher economic and ecological value.

## 2 Genetic Linkage Maps in *Eucommia ulmoides*

### 2.1 Development of genetic linkage maps

The construction of genetic linkage maps in *E. ulmoides* has employed various molecular marker technologies, including genotyping-by-sequencing (GBS) and traditional marker systems such as sequence-related amplified polymorphism (SRAP), amplified fragment length polymorphism (AFLP), inter-simple sequence repeat (ISSR), and simple sequence repeat (SSR) markers. For instance, one study utilized GBS to construct a high-density genetic map by sequencing an $F_1$ population and identifying single-nucleotide polymorphism (SNP) markers (Liu et al., 2022). Another study combined SRAP, AFLP, ISSR, and SSR markers to develop a genetic linkage map from a full-sib family, which encompasses approximately 89% of the estimated *E. ulmoides* genome (Li et al., 2014).

There are four types of markers employed in the construction of genetic linkage maps for *E. ulmoides*. SSR markers, AFLP markers, SNP (single nucleotide polymorphism) markers, and GBS. SSR markers are highly polymorphic and have been widely used due to their co-dominant inheritance and reproducibility (Jin et al., 2020; Li et al., 2014; Wang et al., 2014). AFLP markers are effective for detecting a large number of polymorphisms and have been utilized in several studies (Aabidine et al., 2010; Li et al., 2014). The first genetic linkage map of *Eucommia ulmoides* was developed using a pseudo-testcross strategy with AFLP markers from an F1 population comprising 122 plants, laying the groundwork for mapping quantitative trait loci and facilitating breeding efforts (Wang et al., 2014).

Furthermore, through the application of SRAP, AFLP, ISSR, and SSR analyses, a comprehensive genetic linkage map was established, spanning 2 133 cM and encompassing 25 linkage groups. This map accounts for nearly 89% of the genome, serving as a valuable resource for identifying quantitative trait loci (QTL) associated with growth-related characteristics and supporting marker-assisted selection as well as genomic research. SNP markers have gained popularity due to their abundance and the high resolution they provide in genetic mapping. The use of GBS has facilitated the identification of a substantial number of SNPs, which are crucial for developing high-density genetic maps (Jin et al., 2020; Liu et al., 2022).

The coverage and resolution of genetic maps in *E. ulmoides* have significantly improved through the utilization of high-density SNP markers. Previous study reported a genetic linkage map covering approximately 89% of the estimated genome, with an average marker interval of 3.1 cM (Li et al., 2014). Recently, using SNP markers and GBS technology, Liu et al. (2022) constructed the first high-density genetic map of *E. ulmoides*, revealing 10 103 SNP markers distributed across 17 linkage groups. The genetic map constructed using GBS covered 90% of the genome, with a total genetic distance of 4 051.11 cM and an average distance between markers of 0.45 cM (Liu et al., 2022). These high-density maps offer enhanced resolution, facilitating the precise localization of QTLs.

### 2.2 Applications in breeding

Genetic linkage maps are invaluable tools for marker-assisted selection (MAS) in the breeding of *E. ulmoides*. These maps enable the identification of markers linked to desirable traits, enabling the selection of individuals

with favorable genotypes. The genetic linkage map developed using SSR and SNP markers has been employed to identify QTLs associated with growth traits, aiding in the selection of superior genotypes (Li et al., 2014; Jin et al., 2020).

The identification of QTLs for growth traits in *Eucommia ulmoides* has been a major focus of genetic mapping studies. Several QTLs associated with growth-related traits such as tree height, ground diameter, and crown diameter have been identified. One study detected 44 QTLs for growth traits on various linkage groups, with phenotypic variance ranging from 10.0% to 14.2% (Liu et al., 2022). Another study identified 89 hypothetical QTLs for growth traits measured over ten years, highlighting the genetic basis of these traits and providing targets for MAS (Jin et al., 2020).

# 3 QTL Mapping in *Eucommia ulmoides*

## 3.1 Methodologies for QTL mapping

Traditional QTL mapping techniques in *E. ulmoides* have primarily relied on the construction of genetic linkage maps using various molecular markers. For instance, a genetic linkage map was constructed using SRAP, AFLP, ISSR, and SSR markers. The previous map constructed by Li et al. (2014) using aforementioned molecular markers covered approximately 89% of the *E. ulmoides* genome, with an average distance of 3.1 cM between adjacent markers, facilitating the identification of 18 QTLs associated with growth traits. Additionally, another study refined the genetic linkage map using 452 polymorphic markers from 365 SSR primers, covering 94.10% of the estimated genome and identifying 89 hypothetical QTLs for growth traits (Jin et al., 2020).

With the advancement of sequencing technologies and decreasing costs, bulked segregant analysis (BSA) and QTL-seq have emerged as powerful techniques for rapid QTL mapping. BSA involves pooling DNA from individuals exhibiting extreme phenotypes to identifying significant SNPs associated with the trait of interest. For example, PyBSASeq, a Python-based algorithm, has been developed to enhance the sensitivity of BSA-Seq, enabling the detection of SNP-trait associations at lower sequencing coverage and reduced costs (Zhang and Panthee, 2020). QTL-seq, another analytical method, combines whole-genome resequencing of bulked populations to identify QTLs, which has been successfully applied to various plant species, including rice, for the rapidly mapping of QTLs associated with traits such as disease resistance and seedling vigor (Takagi et al., 2013).

## 3.2 Identified QTLs for key traits

Several studies have identified QTLs associated with growth traits in *E. ulmoides*. A high-density genetic map constructed using GBS identified 44 QTLs linked to growth traits on linkage groups LG02, LG06, LG07, LG08, and LG10, accounting for 10.0% to 14.2% of the phenotypic variance (Liu et al., 2022). Additionally, another study revealed 25 QTLs for tree height, 32 QTLs for ground diameter, and 15 QTLs for crown diameter, highlighting the significant correlation of growth traits measured over a span of ten years (Jin et al., 2020).

While the primary focus has been on growth traits, there is also interest in identifying QTLs for secondary metabolites and yield-related traits. Although specific QTLs associated with these traits in *E. ulmoides* have not been extensively reported, methodologies such as QTL-seq and BSA-Seq, which have been successfully applied in other species, hold promise for future research. For instance, QTG-seq has been utilized to fine-map QTLs in maize to identify candidate genes for plant height (Figure 1) (Zhang et al., 2019). Similar approaches could be adapted for *E. ulmoides*.

## 3.3 Challenges and limitations

Traditional QTL mapping techniques encounter several limitations, such as the requirement for large populations and extensive genotyping efforts. The resolution of QTL mapping is frequently constrained by markers density and the size of the mapping population, which can lead to broad confidence intervals for QTL locations (Li et al., 2014; Jin et al., 2020). Furthermore, the time-consuming and labor-intensive nature of developing and selecting DNA markers for linkage analysis presents significant challenges (Takagi et al., 2013).
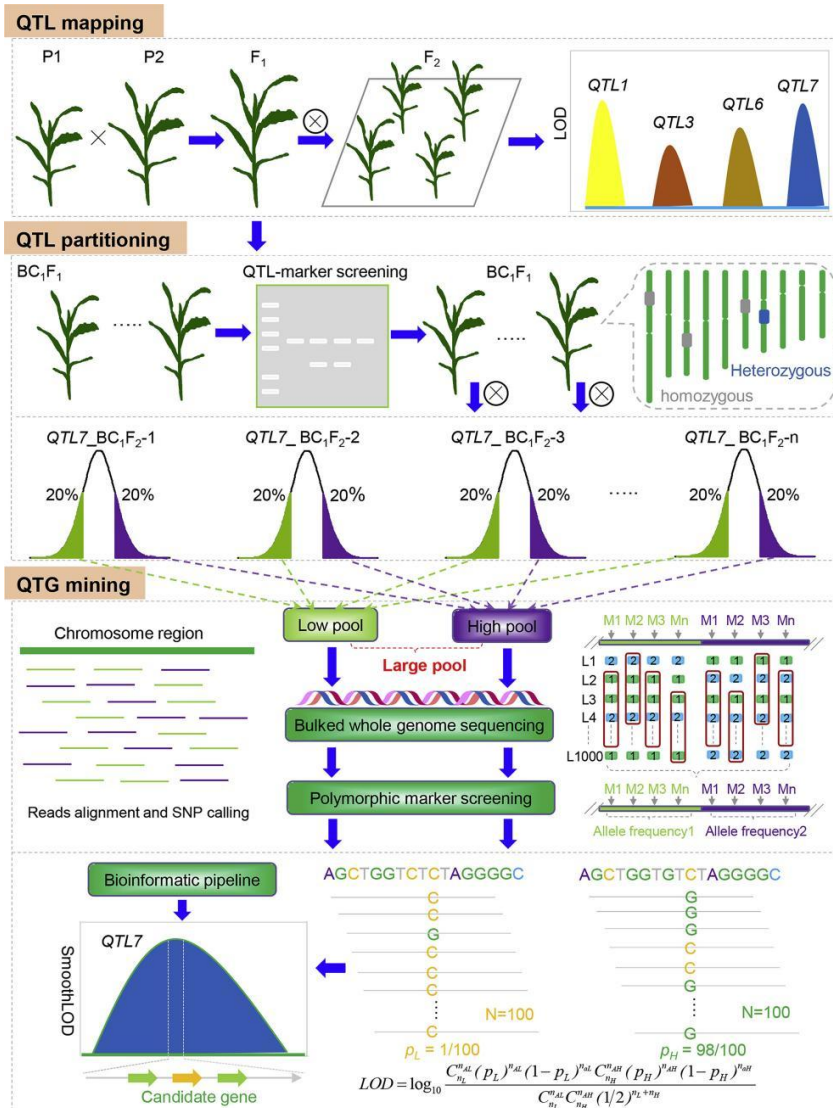
Figure 1 The underlying principles of QTG-Seq (Adopted from Zhang et al., 2019)

Image caption: QTG-seq consists of three major phases: first, QTL mapping; second, QTL partitioning, in which a numerous number of segregating $BC_1F_2$ populations (totaling over 1 000 individuals) are generated though self-crosses of various $BC_1F_1$ individuals that are heterozygous for the target QTL while being homozygous for the other QTLs; and third, QTG mining, which is characterized by high-throughput sequencing of bulked samples from the "high" and "low" pools derived from several segregating $BC_1F_2$ populations, along with a bioinformatic pipeline for efficient QTG fine-mapping. Notably, the relatively extensive high-throughput sequencing coverage in relation to the large number of individuals in the bulked samples, combined with the present of recombination blocks, facilitates precise quantification of *A* and *a* alleles frequencies based on random sampling throughout the high-throughput sequencing process. SNP, single-nucleotide polymorphism (Adopted from Zhang et al., 2019)

To overcome these limitations, there is a need for high-resolution mapping techniques and larger populations. Methods such as QTL-seq and BSA-Seq offer higher resolution and faster identification of QTLs by leveraging next-generation sequencing technologies (Takagi et al., 2013; Zhang and Panthee, 2020). These approaches can significantly reduce the time and cost associated with traditional QTL mapping, making them more accessible for species with large genomes. Furthermore, integrating multiple progenies and using advanced statistical methods can enhance the precision and accuracy of QTL detection, as demonstrated in peach breeding (Mora et al., 2017).

## 4 Genomic Selection in *Eucommia ulmoides*
### 4.1 Principles of genomic selection
GS is a form of marker-assisted selection that employs genetic markers covering the entire genome to ensure that all QTL are in linkage disequilibrium with at least one marker. This method leverages the large number of SNPs

identified through genome sequencing and advanced genotyping techniques to predict breeding values using genetic markers alone (Goddard and Hayes, 2007). The process involves estimating the breeding value from genomic data by calculating the conditional mean of the breeding value given the genotype at each QTL, which is approximated using marker genotypes.

Traditional MAS has proven effective for traits controlled by major QTLs; however, it encounters challenges when addressing quantitative traits influenced by multiple minor effect alleles. In contrast, GS estimates all marker effects simultaneously, thereby capturing the effects of both major and minor QTLs and providing a more comprehensive prediction model (Heffner et al., 2009; Merrick et al., 2022). This approach mitigates the biases associated with individual marker effect estimation and accounts for greater variation stemming from small-effect QTLs, making it more suitable for polygenic traits (Heffner et al., 2009).

### 4.2 Implementation in breeding programs
Implementing GS in breeding programs involves several key steps. Firstly, collect high-density marker data and phenotypic data from a training population. Secondly, develop prediction models using the collected data to estimate marker effects. Then, validate the prediction models with independent test populations to ensure their accuracy. In the end, use the validated models to predict the breeding values of selection candidates and make selection decisions based on these predictions (Heffner et al., 2009; Merrick et al., 2022).

GS has been successfully implemented in various plant species, significantly improving breeding efficiency. For instance, in crop improvement, GS has shown high accuracy in predicting breeding values for polygenic traits, thereby accelerating the breeding cycle and increasing genetic gains per unit time (Heffner et al., 2009). Moreover, GS has been applied to perennial ryegrass, where it has demonstrated the potential to predict breeding values with high accuracy, even for traits with low heritability (Rabier et al., 2016).

### 4.3 Benefits and challenges
GS offers several advantages over traditional breeding methods. By incorporating all marker information, GS provides more accurate predictions of breeding values, particularly for polygenic traits (Habier et al., 2007). Additionally, GS reduces the need for extensive phenotyping in each generation, facilitating faster selection and breeding cycles. GS also captures the effects of both major and minor QTLs, providing a more holistic approach to trait improvement (Heffner et al., 2009).

Despite its advantages, GS faces several challenges. Collecting high-quality genotypic and phenotypic data, essential for developing accurate prediction models, can be resource-intensive (Merrick et al., 2022). Ensuring the accuracy of these models across different populations and environments requires continuous validation and updates (Goddard and Hayes, 2007; Heffner et al., 2009). Additionally, maintaining strong linkage disequilibrium between markers and QTLs over multiple generations can be challenging, affecting the long-term accuracy of GS (Habier et al., 2007).

Integrating GS into *Eucommia ulmoides* breeding programs holds significant potential for improving breeding efficiency and achieving greater genetic gains. However, it is crucial to address the associated challenges and continuously refine the prediction models to ensure successful implementation.

## 5 Integrating QTL Mapping and Genomic Selection
### 5.1 Synergies between QTL mapping and genomic selection
Quantitative trait loci (QTL) mapping identifies specific genome regions associated with phenotypic traits, providing valuable insights into the genetic architecture of these traits. This information is crucial for GS models, which rely on high-density markers to predict breeding values. Incorporating QTL data allows for the refinement of GS models by including markers in strong linkage disequilibrium with the QTL, thus improving prediction accuracy (Goddard and Hayes, 2007; Rabier et al., 2016). For instance, integrating QTL mapping with genomic best linear unbiased prediction (GBLUP) has been shown to enhance the precision of genomic predictions by reducing false positives and increasing mapping accuracy (Li et al., 2017).

Incorporating QTL information into GS models can significantly enhance prediction accuracy by ensuring that the markers are closely linked to the traits of interest. This approach captures both major and minor genetic effects, resulting in more reliable breeding value predictions (Goddard and Hayes, 2007; Desta and Ortiz, 2014; Meuwissen et al., 2021). For example, the StepLMM model, which integrates QTL mapping with GS, has demonstrated higher prediction accuracy compared to traditional GBLUP models by effectively utilizing significant SNPs identified through QTL mapping (Li et al., 2017). Additionally, including whole-genome sequence data in genomic prediction models further improves accuracy by capturing fine-scale genetic variations (Meuwissen et al., 2021).

## 5.2 Case studies and examples

Several studies have successfully integrated QTL mapping and GS across various species, demonstrating the potential of this approach in breeding programs. For instance, in dairy cattle, the combination of whole-genome sequence data from different breeds and Bayesian genomic prediction has improved prediction accuracies and enabled the fine-mapping of QTLs associated with milk production traits (Meuwissen et al., 2021). Similarly, in plant breeding, GS models incorporating QTL information have shown higher prediction accuracies for traits such as yield and disease resistance (Desta and Ortiz, 2014).

In *E. ulmoides*, the integration of QTL mapping and GS holds significant promise for improving growth traits and other economically important characteristics. Recent studies have constructed high-density genetic maps and identified numerous QTLs associated with growth traits in *E. ulmoides* (Li and Sillanpää, 2012; Jin et al., 2020; Liu et al., 2022). Incorporating these QTL data into GS models can improve the accuracy of breeding value predictions and accelerate the development of superior *E. ulmoides* varieties. For example, the identification of 44 QTLs related to growth traits provides a solid foundation for implementing GS in *E. ulmoides* breeding programs (Liu et al., 2022).

## 5.3 Future directions

The future of integrating QTL mapping and GS lies in the adoption of emerging technologies and methodologies. Advances in high-throughput sequencing and genotyping technologies will enable the generation of more comprehensive genomic data, facilitating the identification of additional QTLs and improving the resolution of genetic maps (Meuwissen et al., 2021; Liu et al., 2022). Moreover, the development of more sophisticated statistical models, including those incorporating machine learning algorithms, will enhance the predictive power of GS models (Li and Sillanpää, 2012; Li et al., 2017).

Integrating multi-omics data, such as transcriptomics, proteomics, and metabolomics, with QTL mapping and GS represents a promising direction for future research. By combining different layers of biological information, researchers can achieve a more holistic understanding of the genetic and molecular mechanisms underlying complex traits (Goddard and Hayes, 2007; Meuwissen et al., 2021). This integrative approach has the potential to further improve the accuracy of genomic predictions and facilitate the development of more effective breeding strategies for *E. ulmoides* and other species (Desta and Ortiz, 2014).

# 6 Statistical Methods and Tools

## 6.1 Variable selection techniques

The least absolute shrinkage and selection operator (LASSO) is a widely used method for variable selection in QTL mapping and GS. LASSO imposes a penalty on the absolute values of the regression coefficients, effectively shrinking some of them to zero. This results in a sparse model that selects only the most relevant variables. Generalizations of LASSO, such as elastic net and adaptive LASSO, have been developed to address some of its limitations. The elastic net combines the penalties of LASSO and ridge regression, which is particularly useful when dealing with correlated predictors. Adaptive LASSO, on the other hand, assigns different penalties to different coefficients, offering greater flexibility in variable selection (Li and Sillanpää, 2012; Wimmer et al., 2013).

Bayesian methods provide a probabilistic framework for QTL mapping and GS, integrating prior knowledge about the distribution of QTL effects. Bayesian LASSO, for instance, extends the LASSO method by employing a Bayesian approach to estimate shrinkage parameters, thereby enhancing the accuracy of QTL detection and breeding value prediction. Other Bayesian methods, such as BayesA, BayesB, and BayesCπ, have been developed to address different genetic architectures and marker densities. These methods have demonstrated strong performance under various conditions, with BayesCπ being particularly recommended when a small number of loci have large effects on the trait (Clark et al., 2011; Wang et al., 2015).

### 6.2 Model validation and accuracy

Validating GS models is crucial to ensure their predictive accuracy and reliability. Common validation methods include cross-validation, which involves dividing the data into multiple training and testing sets to evaluate the model's performance, and independent testing, where the model is validated using a completely separate dataset. These methods help evaluate the model's ability to generalize to new data and avoid overfitting (Goddard and Hayes, 2007; Li et al., 2017).

Cross-validation and independent testing are essential for evaluating the robustness and accuracy of GS models. Cross-validation estimates the model's performance by repeatedly splitting the data into training and testing sets, which helps identify potential overfitting. Independent testing, on the other hand, validates the model on a separate dataset not used during training, providing a more stringent assessment of the model's predictive power. Both methods are critical for ensuring that GS models are reliable and applicable to diverse populations (Goddard and Hayes, 2007; Li et al., 2017).

### 6.3 Software and computational tools

Several software tools are available for QTL mapping and GS, each with its own strengths and limitations. Some of the widely used software includes the following: R/qtl is a comprehensive tool for QTL mapping that supports various types of genetic crosses and provides a range of statistical methods for QTL analysis. GEMMA is a software for genome-wide association studies (GWAS) and GS that implements linear mixed models to account for population structure and relatedness. BGLR is a Bayesian generalized linear regression package in R that supports various Bayesian methods for genomic prediction, including Bayesian LASSO, BayesA, BayesB, and BayesCπ (Li and Sillanpää, 2012; Wimmer et al., 2013; Wang et al., 2015).

The increasing availability of high-density marker data presents significant computational challenges for QTL mapping and GS. These challenges include managing large datasets, ensuring computational efficiency, and developing robust statistical methods. Addressing these issues requires the development of more efficient algorithms, the use of parallel computing, and leveraging high-performance computing resources. For instance, the Stepwise Linear Mixed Model (StepLMM) has been proposed to integrate GWAS and GS into a single statistical model, improving computational efficiency and accuracy (Eeuwijk et al., 2009; Li et al., 2017). Additionally, Bayesian methods and penalized regression techniques like LASSO can help manage the complexity of the data and improve the robustness of the models (Li and Sillanpää, 2012; Wimmer et al., 2013; Wang et al., 2015).

## 7 Case Studies in *Eucommia ulmoides*

### 7.1 Growth traits

QTL mapping has been extensively used to identify genetic regions associated with growth traits in *E. ulmoides*. Jin et al. (2020) constructed a comprehensive genetic linkage map using 452 polymorphic markers, covering 94.10% of the estimated genome (Figure 2). This study identified 25 QTLs for tree height, 32 QTLs for ground diameter, and 15 QTLs for crown diameter, distributed across various linkage groups. Li et al. (2014) mapped 706 markers across 25 linkage groups, covering approximately 89% of the genome, and identified 18 QTLs explaining 12.4% to 33.3% of the phenotypic variance in growth traits. Additionally, Liu et al. (2022) used GBS to construct a high-density genetic map and identified 44 QTLs associated with growth traits, with phenotypic variance ranging from 10.0% to 14.2%.
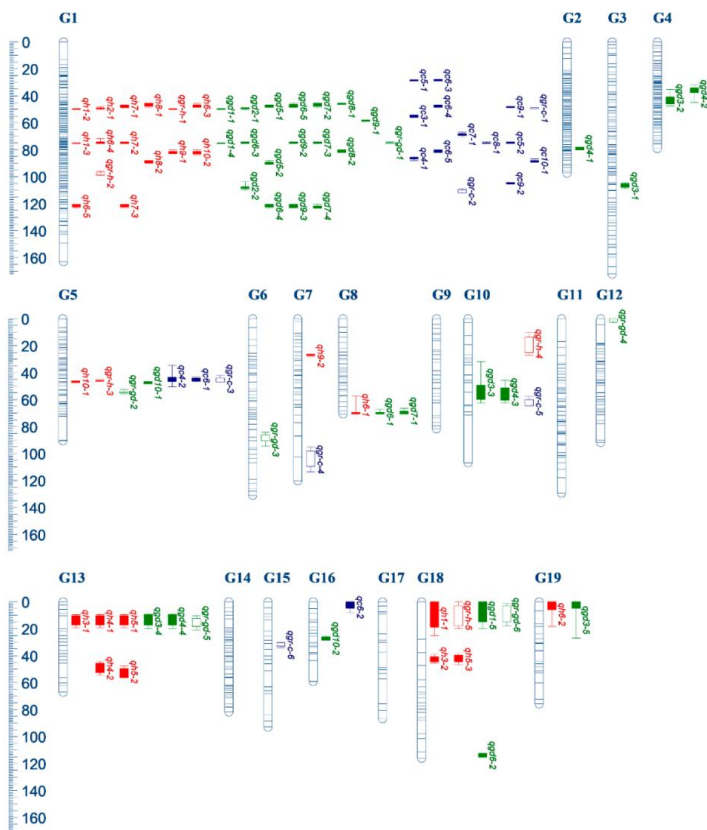
Figure 2 QTL mapping on genetic linkage groups for growth traits in *E. ulmoides* (Adopted from Jin et al., 2020)

Image caption: The red, green and blue bars denote QTLs associated with height, ground diameter and crown diameter, respectively. The hollow bars represent QTLs for growth rates. The thick bars and thin lines represent 1-LOD and 2-LOD confidence intervals for each QTL, respectively (Adopted from Jin et al., 2020)

GS models have shown promise in improving growth traits in forest trees, including *E. ulmoides*. The effectiveness of GS was demonstrated in Eucalyptus, where prediction models for tree circumference and height growth achieved accuracies between 0.55 and 0.88, capturing 74%~97% of trait heritability (Resende et al., 2012). Similarly, Grattapaglia and Resende (2011) highlighted the potential of GS to revolutionize tree breeding by improving selection efficiency and shortening breeding cycles. These models can be adapted for *E. ulmoides* to enhance the accuracy and efficiency of selecting for growth traits.

### 7.2 Secondary metabolites

Secondary metabolites in *E. ulmoides*, which are significant in pharmacology, contribute substantially to its economic value. Although specific QTLs for secondary metabolite content in *E. ulmoides* were not detailed in the provided studies, the methodologies used for growth traits can be adapted to identify QTLs for secondary metabolites. For instance, Jin et al. (2020) and Liu et al. (2022) utilized high-density genetic maps and QTL analysis to identified QTLs for growth traits, which can be similarly employed to map QTLs for secondary metabolite content.

GS can significantly enhance the yield of secondary metabolites by capturing the genetic variance associated with these traits. The principles demonstrated in *Eucalyptus* for growth and wood quality traits (Resende et al., 2012) can be extended to secondary metabolites in *E. ulmoides*. By developing prediction models that incorporate genome-wide markers, it is possible to improve the selection accuracy for high-yielding genotypes, thereby accelerating the breeding process for enhanced metabolite production.

### 7.3 Yield-related traits

Yield-related traits in *E. ulmoides* are critical for its commercial cultivation. Although the studies reviewed did not explicitly focus on yield traits, the methodologies used for growth traits can be adapted. For example, Jin et al.

(2020) and Liu et al. (2022) identified multiple QTLs associated with growth traits, which are frequently correlated with yield. By extending these QTL mapping techniques, it becomes feasible to identify QTLs specifically associated with yield and related traits.

Integrating QTL mapping with GS offers a powerful strategy to improve yield in *E. ulmoides*. The effectiveness of GS in capturing the heritability of complex traits in *Eucalyptus* (Resende et al., 2012) and its potential to enhance tree breeding efficiency (Grattapaglia and Resende, 2011) suggest that similar strategies could be applied for yield traits in *E. ulmoides*. By combining QTL information with GS models, breeders can make more informed selection decisions, resulting in significant improvements in yield and overall breeding efficiency.

## 8 Practical Implications for Breeding Programs

### 8.1 Breeding strategies

Integrating QTL mapping and GS in breeding programs for *E. ulmoides* can significantly enhance breeding strategies. QTL mapping identifies specific genomic regions associated with desirable traits, such as growth, yield, and secondary metabolite production, which can be targeted for selection (Li et al., 2014; Li et al., 2015; Jin et al., 2020; Liu et al., 2022). By combining this with GS, which utilizes genome-wide markers to predict breeding values, breeders can capture the effects of multiple minor alleles, thus improving the accuracy and efficiency of selection (Goddard and Hayes, 2007; Grattapaglia and Resende, 2011). This dual approach optimizes breeding programs by reorganizing field designs, increasing the number of lines evaluated, and leveraging extensive genomic and phenotypic data across diverse environments to maximize genetic gain (Merrick et al., 2022).

### 8.2 Economic and ecological impact

The economic and ecological impacts of integrating QTL mapping and GS in the breeding of *E. ulmoides* are significant. *E. ulmoides* is highly regarded for its medicinal properties, rubber production, and ecological advantages, such as wind sheltering and sand fixation (Liu et al., 2022; Du et al., 2023). Enhanced breeding strategies can yield superior varieties characterized by higher productivity, enhanced quality, and greater resistance to environmental stresses, thereby increasing economic returns for farmers and industries (Li et al., 2014; Li et al., 2015). Additionally, the ecological advantages of *E. ulmoides*, such as soil stabilization and biodiversity preservation, can be further enhanced through the cultivation of robust and resilient varieties, contributing to sustainable environmental management (Jin et al., 2020; Liu et al., 2022).

### 8.3 Policy and regulatory considerations

The implementation of advanced breeding techniques like QTL mapping and GS in *E. ulmoides* breeding programs requires careful consideration of policy and regulatory frameworks. Policymakers must ensure that regulations facilitate the use of molecular breeding methods while safeguarding genetic diversity and preventing the monopolization of genetic resources (Merrick et al., 2022). Furthermore, guidelines should be established for the ethical use of genomic data and the protection of intellectual property rights associated with newly developed varieties (Grattapaglia and Resende, 2011). Collaboration among researchers, breeders, and regulatory bodies is crucial to create an environment conducive to the adoption of these technologies, ensuring that the benefits of improved *E. ulmoides* varieties are widely accessible and contribute to both economic and ecological sustainability (Goddard and Hayes, 2007; Grattapaglia and Resende, 2011; Merrick et al., 2022).

## 9 Concluding Remarks

In this study, we integrated QTL mapping and GS to enhance the breeding efficiency of *E. ulmoides*. The construction of high-density genetic maps using SNP markers and GBS has established a robust foundation for identifying QTLs associated with growth traits. Specifically, a high-density genetic map was constructed, covering 90% of the *E. ulmoides* genome and identifying 44 QTLs related to growth traits. Furthermore, an updated genetic linkage map revealed 89 QTLs based on 10 years of growth trait measurements, further elucidating the genetic mechanisms underlying these traits. The integration of these maps with GS techniques, which utilize genome-wide markers to predict breeding values, shows promise in improving the accuracy and efficiency of breeding programs.

Future research should focus on several key areas to further enhance the integration of QTL mapping and GS in the breeding of *E. ulmoides*. Although numerous QTLs have been well identified, their validation across different populations and environments is imperative to ensure their reliability and applicability in diverse breeding programs. It will be essential to develop and optimize GS models that incorporate multi-trait and multi-environment data. This includes utilizing high-throughput phenotyping and deep learning approaches to improve prediction accuracy. Additionally, identifying and characterizing candidate genes within QTL regions will provide insights into the biological mechanisms controlling important traits, which can be achieved through transcriptomic and proteomic analyses. Traditional breeding programs need to be restructured to effectively implement genomic selection. This includes reorganizing field designs, increasing the number of lines evaluated, and optimizing training populations.

The integration of QTL mapping and GS represents a significant advancement in the breeding of *E. ulmoides*. By combining high-density genetic maps with sophisticated GS techniques, we can achieve more accurate and efficient selection of desirable traits. This approach not only accelerates the breeding process but also enhances the genetic gain and overall quality of *E. ulmoides* cultivars. Ongoing research and optimization in this field will undoubtedly contribute to the sustainable development and economic viability of this important tree species.

## 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
Aabidine A., Charafi J., Grout C., Doligez A., Santoni S., Moukhli A., Jay-Allemand C., Modafar C., and Khadari B., 2010, Construction of a genetic linkage map for the olive based on AFLP and SSR markers, Crop Science, 50: 2291-2302.
https://doi.org/10.2135/CROPSCI2009.10.0632

Clark S., Hickey J., and Werf J., 2011, Different models of genetic variation and their effect on genomic evaluation, Genetics Selection Evolution, 43: 18.
https://doi.org/10.1186/1297-9686-43-18

Desta Z., and Ortiz R., 2014, Genomic selection: genome-wide prediction in plant improvement, Trends in Plant Science, 19(9): 592-601.
https://doi.org/10.1016/j.tplants.2014.05.006

Du Q., Wu Z., Liu P., Qing J., He F., Du L., Sun Z., Zhu L., Zheng H., Sun Z., Yang L., Wang L., and Du H., 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

Eeuwijk F., Boer M., Totir L., Bink M., Wright D., Winkler C., Podlich D., Boldman K., Baumgarten A., Smalley M., Arbelbide M., Braak C., and Cooper M., 2009, Mixed model approaches for the identification of QTLs within a maize hybrid breeding program, Theoretical and Applied Genetics, 120: 429-440.
https://doi.org/10.1007/s00122-009-1205-0

Goddard M., and Hayes B., 2007, Genomic selection, Journal of Animal Breeding and Genetics, 124(6): 323-330.
https://doi.org/10.1111/J.1439-0388.2007.00702.X

Grattapaglia D., and Resende M., 2011, Genomic selection in forest tree breeding, Tree Genetics & Genomes, 7: 241-255.
https://doi.org/10.1007/s11295-010-0328-4

Habier D., Fernando R., and Dekkers J., 2007, The impact of genetic relationship information on genome-assisted breeding values, Genetics, 177: 2389-2397.
https://doi.org/10.1534/genetics.107.081190

Heffner E., Sorrells M., and Jannink J., 2009, Genomic selection for crop improvement, Crop Science, 49(1): 1-12.

https://doi.org/10.2135/CROPSCI2008.08.0512

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.

https://doi.org/10.3390/f11030311

Li H., Su G., Jiang L., and Bao Z., 2017, An efficient unified model for genome-wide association studies and genomic selection, Genetics Selection Evolution, 49: 64.

https://doi.org/10.1186/s12711-017-0338-x

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: 2053-2074.

https://doi.org/10.3390/ijms15022053

Li Y., Wei J., Li Z., Lei A., and Liu M., 2015, Quantitative trait locus analysis of leaf morphological characters, yield-related traits, and secondary metabolite contents in *Eucommia ulmoides*, Genetics and Molecular Research, 14(4): 17871-17884.

https://doi.org/10.4238/2015.December.22.12

Li Z., and Sillanpää M., 2012, Overview of LASSO-related penalized regression methods for quantitative trait mapping and genomic selection, Theoretical and Applied Genetics, 125: 419-435.

https://doi.org/10.1007/s00122-012-1892-9

Liu C., Wang L., Lu W., Zhong J., Du H., Liu P., Du Q., Du L., 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* Oliver, Forests, 13(9): 1479.

https://doi.org/10.3390/f13091479

Merrick L., Herr A., Sandhu K., Lozada D., and Carter A., 2022, Optimizing plant breeding programs for genomic selection, Agronomy, 12(3): 714.

https://doi.org/10.3390/agronomy12030714

Meuwissen T., Berg I., and Goddard M., 2021, On the use of whole-genome sequence data for across-breed genomic prediction and fine-scale mapping of QTL, Genetics Selection Evolution, 53: 19.

https://doi.org/10.1186/s12711-021-00607-4

Mora J., Micheletti D., Bink M., Weg E., Cantín C., Nazzicari N., Caprera A., Dettori M., Micali S., Banchi E., Campoy J., Dirlewanger E., Lambert P., Pascal T., Troggio M., Bassi D., Rossini L., Verde I., Quilot-Turion B., Laurens F., Arús P., and Aranzana M., 2017, Integrated QTL detection for key breeding traits in multiple peach progenies, BMC Genomics, 18: 404.

https://doi.org/10.1186/s12864-017-3783-6

Rabier C., Barre P., Asp T., Charmet G., and Mangin B., 2016, On the accuracy of genomic selection, PLoS One, 11(6): e0156086.

https://doi.org/10.1371/journal.pone.0156086

Resende M., Resende M., Sansaloni C., Petroli C., Missiaggia A., Aguiar A., Abad J., Takahashi E., Rosado A., Faria D., Pappas G., Kilian A., and Grattapaglia D., 2012, Genomic selection for growth and wood quality in *Eucalyptus*: capturing the missing heritability and accelerating breeding for complex traits in forest trees, The New Phytologist, 194(1): 116-128.

https://doi.org/10.1111/j.1469-8137.2011.04038.x

Takagi H., Abe A., Yoshida K., Kosugi S., Natsume S., Mitsuoka C., Uemura A., Utsushi H., Tamiru M., Takuno S., Innan H., Cano L., Kamoun S., and Terauchi R., 2013, QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations, The Plant Journal, 74(1): 174-183.

https://doi.org/10.1111/tpj.12105

Wang C., Gong H., Feng M., and Tian C., 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 D., Li Y., Li L., Wei Y., and Li Z., 2014, The first genetic linkage map of *Eucommia ulmoides*, J. Genet., 93: 13-20.

https://doi.org/10.1007/s12041-014-0322-y

Wang W., Yang G., Deng X., Shao F., Li Y., Guo W., Liang H., and Zhang X., 2020, Molecular sex identification in the hardy rubber tree (*Eucommia ulmoides* Oliver) via ddRAD markers, International Journal of Genomics, 1: 2420976.

https://doi.org/10.1155/2020/2420976

Wang X., Yang Z., and Xu C., 2015, A comparison of genomic selection methods for breeding value prediction, Science Bulletin, 60: 925-935.

https://doi.org/10.1007/S11434-015-0791-2

Wimmer V., Lehermeier C., Albrecht T., Auinger H., Wang Y., and Schön C., 2013, Genome-wide prediction of traits with different genetic architecture through efficient variable selection, Genetics, 195: 573-587.

https://doi.org/10.1534/genetics.113.150078

Wuyun T., Wang L., Liu H., Wang X., Zhang L., Bennetzen J., Li T., Yang L., Liu P., Du L., Wang L., Huang M., Qing J., Zhu L., Bao W., Li H., Du Q., Zhu J., Yang H., Yang S., Liu H., Yue H., Hu J., Yu G., Tian Y., Liang F., Hu J., Wang D., Gao R., Li D., and Du H., 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

Xie S., Si H., Sun H., Zhao Q., Li X., Wang S., Niu J., and Wang Z., 2023, Predicting the potential distribution of the endangered plant *Eucommia ulmoides* in China under the background of climate change, Sustainability, 15(6): 5349.

https://doi.org/ 10.3390/su15065349

Zhang H., Wang X., Pan Q., Li P., Liu Y., Lu X., Zhong W., Li M., Han L., Li J., Wang P., Li D., Liu Y., Li Q., Yang F., Zhang Y., Wang G., and Li L., 2019, QTG-seq accelerates QTL fine mapping through QTL partitioning and whole-genome sequencing of bulked segregant samples, Molecular Plant, 12(3): 426-437.

https://doi.org/10.1016/j.molp.2018.12.018

Zhang J., and Panthee D., 2020, PyBSASeq: a simple and effective algorithm for bulked segregant analysis with whole-genome sequencing data, BMC Bioinformatics, 21(99): 1-10.

https://doi.org/10.1186/s12859-020-3435-8

Zhang S., Xu T., Ren Y., Song L., Liu Z., Kang X., and Li Y., 2023, The NAC transcription factor family in *Eucommia ulmoides*: genome-wide identification, characterization, and network analysis in relation to the rubber biosynthetic genes, Front. Plant Sci., 14: 1030298.

https://doi.org/10.3389/fpls.2023.1030298