Genome-Wide Identification and Functional Analysis of The FRA8 Gene Family in Eucalyptus Grandis

Yuhan Li

School of Forestry, Beijing Forestry University, Beijing 100083, China;

2983552235@qq.com

Abstract. *FRA8* is a key glycosyltransferase in the GT47 family that mediates the addition of glucuronic acid side chains to the structural polysaccharide xylan, playing a central role in the formation of plant secondary cell walls and the regulation of wood quality. However, the systematic characteristics of the *FRA8* family in woody plants remain unclear, particularly in the fast-growing economic timber species *Eucalyptus grandis*. This study utilized bioinformatic methods to perform genome-wide identification, evolutionary analysis, and expression characterization of the *FRA8* family in *E. grandis*. Results identified 31 *EgrFRA8* genes. Most members contain the conserved Exostosin domain (PF03016) of the GT47 family, while one unique member carries an RXYLT1-like domain. Phylogenetic analysis categorized them into five groups, where members such as *EgrFRA8_1-3* are closely related to functionally validated Arabidopsis glucuronyltransferases *AtFRA8* and *AtFRA8H*. Cis-acting elements in promoters were enriched with light, hormone, and stress-responsive motifs, suggesting regulation by abiotic factors. RNA-seq data confirmed high expression of *EgrFRA8_2*, *EgrFRA8_3*, and *EgrFRA8_26* in immature and mature xylem, suggesting involvement in wood formation. Twenty-nine *EgrFRA8* members were mapped to 11 chromosomes, with segmental duplication identified as the primary driver of family expansion. This study provides a systematic analysis of the *EgrFRA8* family, offering a basis for elucidating xylan synthesis mechanisms in woody plants and identifying high-priority candidate genes for the genetic improvement of wood quality in *E. grandis*.

Keywords: FRA8 gene family; Grand eucalyptus; Whole-genome identification; Xylan- biosynthesis; Wood formation.

1. Introduction

Wood formation is a complex and precise biological process, with its core lying in the synthesis and assembly of secondary cell walls [23]. As the most abundant renewable biomass resource on Earth, the structure and properties of wood not only determine its application value in fields such as construction, papermaking, and energy, but also are related to the global carbon cycle and ecological security. Among the three major components for constructing secondary cell walls - cellulose, lignin and hemicellulose - xylan, as the main component of hemicellulose in dicotyledonous plants, the regulatory mechanism of its fine structure has increasingly become the frontier and focus of plant cell wall research [7]. Xylan is not a simple homogeneous polysaccharide. Its β-(1,4) -D-xylose main chain is usually attached with acidic side chains such as glucuronic acid (GlcA) or 4-O-methylglucuronic acid (MeGlcA). The distribution pattern, density and modification degree of these side chains directly affect the physical and chemical properties of xylan itself, and fundamentally determine the mechanical strength, hydrophilicity, porosity and biodegradability of the cell wall by regulating its hydrogen bond combination with cellulose microfilaments and cross-linking mode with lignin precursors [6]. Therefore, a thorough analysis of xylan biosynthesis, especially the molecular mechanism of adding glucuronic acid side chains, has become a key scientific breakthrough for improving wood quality and enhancing the utilization efficiency of biomass resources through genetic means.

The synthesis of plant cell wall polysaccharides depends on a series of glycosyltransferases located in the Golgi apparatus. Among them, the glycosyltransferase family 47 (GT47) has attracted much attention due to the significant roles its members play in the synthesis of various cell wall polysaccharides [19]. The name of this family stems from the presence of a conserved "exogenic

osteoma protein-like" domain (PF03016) in its protein sequence, which was initially involved in the synthesis of heparan sulfate chains in animals. In plants, members of the GT47 family have diverse functions and have been proven to be involved in the biosynthesis of various polysaccharides such as xylan, xyloglucan, and pectin rhamnogalacturonic acid polysaccharide II, constituting key functional nodes in the cell wall synthesis network [14]. Early bioinformatics analysis of the Arabidopsis thaliana genome identified 39 GT47 members, indicating the extensive and complex role of this family in plant growth and development [33].

Among the numerous members of the GT47 family, the functional study of the FRA8 (FRAGILE FIBER 8) gene has provided landmark insights for understanding xylan synthesis. In 2005, Zhong and his collaborators first reported the FRA8 gene through the screening and cloning of mechanical strength defect mutants in Arabidopsis stem [33]. They found that the FRA8 mutant exhibited a "fragile fiber" phenotype with stems that were highly prone to breakage, and the fundamental reason for this was a significant reduction in the thickness of the secondary cell walls of fibroblasts and vessel molecules. In-depth biochemical analysis of the cell wall revealed the molecular nature of this phenotype: the non-methylated glucuronic acid (GlcA) side chain in the mutant xylan was specifically absent, while the 4-O-methylglucuronic acid side chain was retained. This precise result, combined with the evidence of its encoded protein having a conserved domain of the GT47 family and subcellular localization in the Golgi complex, jointly identified FRA8 as a glucuronotransferase specifically catalyzing the addition of GlcA to the xylan main chain [33]. This study, for the first time, directly associated the mutation of a single GT gene with a specific structural defect of secondary parietal xylan at the genetic and biochemical levels, setting a model for the study of gene functions in plant cell wall synthesis.

The importance of FRA8 is not limited to Arabidopsis thaliana. Subsequent comparative and evolutionary genomics studies have shown that the FRA8 homologous gene is widely present in terrestrial plants and exhibits functional conservation [36]. In the woody model plant Populus tomentosa, the FRA8 direct homologous gene is highly expressed in the developing xylem, and the functional conservation is demonstrated through complementary Arabidopsis FRA8 mutants [28]. In the economic crop cotton, the co-expression network analysis of genes related to fiber strength development also frequently shows GT47 family genes, suggesting their potential role in the secondary wall thickening of cotton fibers [21]. These cross-species pieces of evidence collectively paint a picture: FRA8 and the xylan GLCAylation modification pathway it represents are an ancient and core molecular module for plants to construct robust secondary cell walls. However, most of the existing research focuses on annual herbaceous or a few woody model plants. For the important timber tree species that play a crucial role in the global forestry economy and ecosystem, the systematic understanding of this key gene family is almost non-existent [28].

Eucalyptus grandis is precisely such a key tree species. As one of the world's three fast-growing tree species, the giant eucalyptus has become the first choice for artificial afforestation in tropical and subtropical regions due to its rapid growth, strong adaptability and wide range of wood uses, supporting a huge papermaking, artificial board and bioenergy industry [11]. With the release of its high-quality reference genome, the giant eucalyptus has become an ideal system for studying forest biology, especially the genetic basis of wood formation [11]. Genetic improvement of wood quality is the core objective of Eucalyptus grandis breeding, and the prerequisite for achieving this goal is to analyze the gene network that controls wood traits such as density, fiber length, lignin and cellulose content [29]. Given the profound influence of xylan structure on the properties of the cell wall, the FRA8 gene family responsible for its synthesis is undoubtedly a highly promising key regulatory target in this network. Regrettably, we currently know very little about the basic situation of the FRA8 gene family of Eucalyptus grandis: how many members does this family have in the complex genome of eucalyptus grandis? What kind of expansion and evolutionary process have they undergone? What are the characteristics of its gene structure and protein sequence? Most importantly, are they specifically expressed in the "heart" of wood formation - the cambium and xylem - and thus directly participate in wood construction?

The essence of this knowledge gap lies in the disconnection between fundamental molecular mechanism research and applied breeding of important tree species. On the one hand, the research on model plants represented by Arabidopsis thaliana provides us with a beautiful molecular blueprint of the function of FRA8 [33]; On the other hand, the breeding practice of economic tree species represented by Eucalyptus grandis urgently requires targeted candidate genes and molecular markers [29]. Filling this gap can not only deepen our understanding of the conservation and specificity of xylan synthesis mechanisms among species, which is of great theoretical significance, but also directly transform the most cutting-edge knowledge of cell wall biology into tangible gene resources that drive the genetic improvement of eucalyptus grandis [27]. Today, with the increasing maturity of genome editing technology, precise and systematic pre-bioinformatics analysis of a key gene family is an indispensable "navigation map" for subsequent efficient functional verification and molecular design breeding [21].

Therefore, this study aims to conduct a comprehensive and systematic genome-wide identification and bioinformatics analysis of the FRA8 gene family of Eucalyptus grandis. We will comprehensively utilize genomic, transcriptomic and protein sequence data to achieve the following core goals: (1) Precisely identify all members of the FRA8 gene family of Eucalyptus grandis across the entire genome and establish a complete molecular catalog of them; (2) Through multi-species comparison and phylogenetic analysis, clarify the evolutionary position of this family in Eucalyptus grandis, the internal clustering relationship, and the possible subfunctional differentiation [12]; (3) Analyze its gene structure (exon-intron composition), protein conserved motifs and domain characteristics [2]; (4) Determine the positional distribution of each member on chromosomes and analyze the contribution of gene replication events (such as whole-genome replication and tandem replication) to family expansion [12]; (5) By using publicly available cross-tissue and cross-developmental stage transcriptome data, expression maps of family members were drawn, with a focus on revealing their expression patterns in wood formation related tissues, and candidate genes that are dominant in xylem were screened out [29]. This research is expected to produce the first comprehensive molecular map of the FRA8 gene family of Eucalyptus grandis. This achievement will serve as a key piece of knowledge puzzle. On the one hand, it will lay a solid foundation for in-depth exploration of the regulatory mechanism of xylan synthesis in eucalyptus grandis. On the other hand, it will also improve the quality of eucalyptus grandis wood through precise breeding strategies such as molecular marker-assisted selection or gene editing. Provide a list of candidate genes with high priority and theoretical basis, thereby building a bridge from basic discovery to forestry application [27].

2. Results

2.1. Identification and physicochemical property analysis of FRA8 gene Family members in Eucalyptus grandis

Using the identified FRA8 (AT2G28110) and its homologous protein sequence in Arabidopsis thaliana as probes, we conducted BLASTp alignment on the genome of the giant gum and verified it through the GT47 family characteristic domain (Pfam03016). A total of 31 FRA8 homologous genes were identified. Name them as EgrFRA8-1 to EgrFRA8-31 (Table 1). Compared with the only two FRA8 members (AtFRA8 and AtFRA8H) in Arabidopsis thaliana, the gene family in Eucalyptus grandis has undergone significant linea-specific expansion. Bioinformatics analysis revealed significant differences in the physicochemical properties of the EgFRA8 protein. Its length ranges from 269 amino acids to 793 amino acids, and the corresponding predicted molecular weights are 31.4 kDa to 89.5 kDa. Theoretical isoelectric point analysis indicates that most proteins are alkaline, among which EgrFRA8-3 has the highest isoelectric point. On the contrary, EgrFRA8-5, EgrFRA8-15, EgrFRA8-17 and EgrFRA8-30 were predicted to be acidic proteins. The instability index of most members exceeded 40, indicating structural instability in vitro. It is worth noting that the instability indices of EgrFRA8-03, EgrFRA8-11, EgrFRA8-13 and EgrFRA8-16 exceed 60. The aliphatic amino

acid index varies between 71.68 and 91.95, indicating differences in thermal stability. Furthermore, the total average hydrophilicity values of all members were negative, confirming their hydrophilic characteristics. Among them, EgrFRA8-31 had the strongest hydrophilicity, GRAVY = -0.058 (Table 1).

Table 1. Physicochemical of proteins encoded by the *FRA8* gene family in E. grandis.

Gene Name	Gene ID	Molecular Weight	Theoretical pI	Instability Index	Aliphatic Index	Grand Average of Hydropathicity
EgrFRA8_1	rna-Eucgr.K02191.1	47541.91	6.6	45.64	85.66	-0.154
EgrFRA8_2	rna-Eucgr.G01977.1	46947.95	6.7	45.38	85.5	-0.193
EgrFRA8_3	rna-Eucgr.J00384.1	53720.01	11.1	65.12	71.68	-0.527
EgrFRA8_4	rna-Eucgr.K02411.1	39748.63	8.43	50.07	77.88	-0.249
EgrFRA8_5	rna-Eucgr.A02396.1	89496.28	5.97	45.04	75.11	-0.387
EgrFRA8_6	rna-Eucgr.C01719.1	39086.19	8.88	48.49	88.53	-0.232
EgrFRA8_7	rna-Eucgr.I02612.1	39144.98	9.53	43.63	78.34	-0.338
EgrFRA8_8	rna-Eucgr.K02385.1	40659.78	9.19	51.53	79.38	-0.261
EgrFRA8_9	rna-Eucgr.H02436.1	48589.4	8.72	51.95	83.3	-0.239
EgrFRA8_10	rna-Eucgr.C01718.1	38543.43	9.18	58.52	82.51	-0.345
EgrFRA8_11	rna-Eucgr.L02165.1	45801.55	9.66	60.3	79.8	-0.444
EgrFRA8_12	rna-Eucgr.H01381.1	45261.17	9.48	47.55	86.72	-0.313
EgrFRA8_13	rna-Eucgr.K00083.1	57796.26	9.24	62.95	90.68	-0.214
EgrFRA8_14	rna-Eucgr.D02098.1	62817.07	9.4	52.65	91.95	-0.19
EgrFRA8_15	rna-Eucgr.A01384.1	63097.16	7.7	53.9	85.13	-0.183
EgrFRA8_16	rna-Eucgr.A02111.1	54124.77	9.6	60.91	77.26	-0.239
EgrFRA8_17	rna-Eucgr.A01382.1	56082.05	7.71	54.83	82.56	-0.258
EgrFRA8_18	rna-Eucgr.C01013.1	70668.08	9.46	46.4	80.2	-0.445
EgrFRA8_19	rna-Eucgr.E03530.1	54463.65	9.51	42.04	85.26	-0.345
EgrFRA8_20	rna-Eucgr.C02925.1	78083.17	8.51	48.61	78.15	-0.389
EgrFRA8_21	rna-Eucgr.E03525.1	55258.18	9.05	46.64	81.68	-0.421
EgrFRA8_22	rna-Eucgr.L02716.1	55116.98	8.93	45.82	81.87	-0.385
EgrFRA8_23	rna-Eucgr.B03487.1	38573.09	8.61	47.65	83.48	-0.365
EgrFRA8_24	rna-Eucgr.C01012.1	79812.76	9.47	52.19	78.45	-0.417
EgrFRA8_25	rna-Eucgr.E02380.1	52680.47	9.3	47.16	79.89	-0.403
EgrFRA8_26	rna-Eucgr.B00504.1	56581.1	9.34	36.64	89.68	-0.247
EgrFRA8_27	rna-Eucgr.A02153.1	57900.78	9.01	41.79	78.43	-0.425
EgrFRA8_28	rna-Eucgr.E04252.1	55163.55	8.98	47.89	89.12	-0.357
EgrFRA8_29	rna-Eucgr.B03578.1	31353.59	9.19	45.89	76.8	-0.388
EgrFRA8_30	rna-Eucgr.F02636.1	55269.1	7.71	52.25	85.85	-0.328
EgrFRA8_31	rna-Eucgr.J02121.1	35365.58	6.76	58.89	89.72	-0.058

2.2. Analysis of the EgrFRA8 Family Structure and Conserved Domains

Based on the characteristics of conserved protein motifs, 31 EgrFRA8 proteins were classified into five categories in the phylogenetic tree (Figure 1A). The conserved motifs of these 31 EgrFRA8 proteins were visually analyzed through the MEME online program (Figure 1B). Research has found that the amino-terminal of the EgrFRA8 protein has a relatively conserved domain. Among them, the structure of the i-th type member is the simplest, containing only the motifs 2 and 8. Members of Class II and Class IV (except EgrFRA8_6 and EgrFRA8_10) contain all 10 motifs and are arranged in the same order. Members of Category III are missing motif 10; Members of Class V, on the other hand, have a unique simplified combination, consisting only of motifs 1, 2, 7, 8, and 9. Conservative domain analysis further clarified its direct homology: members of the EgrFRA8_1-EgrFRA8_30 families all contain the characteristic Exostosin domain of the GT47 family (Figure 1C), among which EgFRA8_5 additionally contains the EGF_2 domain. EgFRA8_31 specifically contains the RXYLT1-like superfamily domain. EgrFRA8_31 has been clearly identified as a direct homolog of Arabidopsis AtFRA8H in Eucalyptus grandis, while the remaining members are homolog of AtFRA8.

Gene structure analysis (Figure 1D) revealed that there were significant differences in the number of exons among different members (3 to 14), and members with similar phylogenetic relationships had similar exon-intron structures.

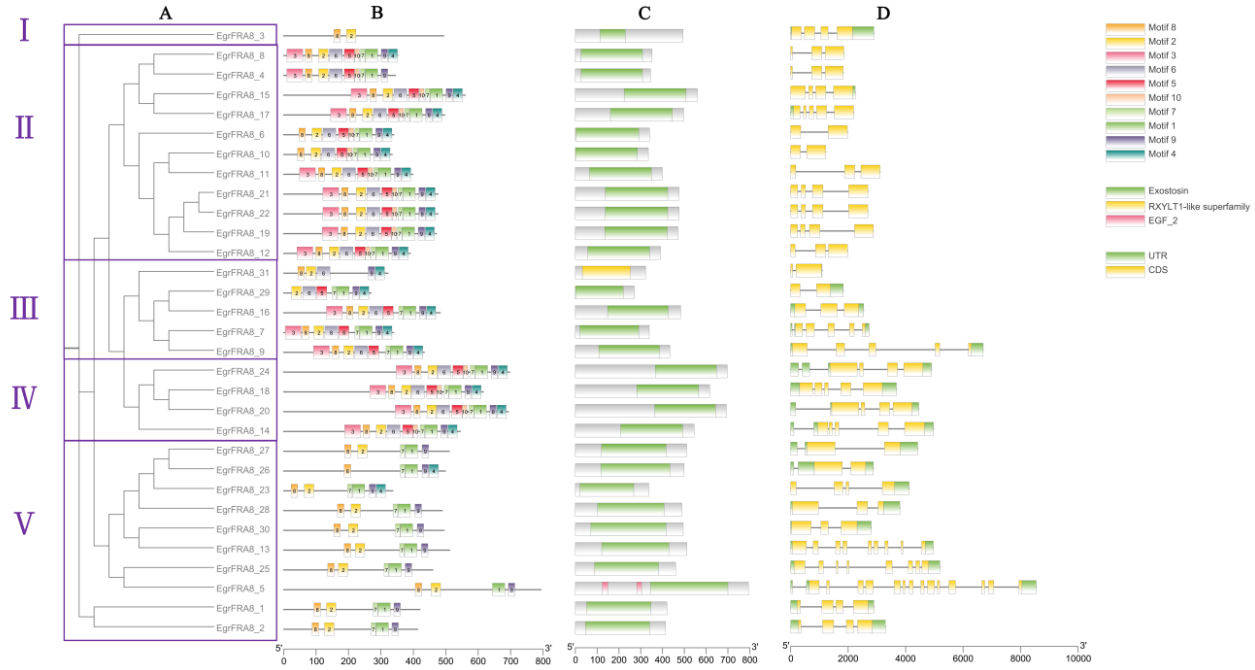

Figure 1. Phylogenetic tree, conserved amino acid motifs, domains and gene structure analysis of the predicted EgrFRA8 protein (A) Phylogenetic tree of the EgrFRA8 protein constructed based on the maximum likelihood method, showing that members are divided into five evolutionary branches (I-V). The distribution of 10 conserved motifs (motifs 1-10) in the EgrFRA8 protein. (C) The conserved domain composition of the EgrFRA8 protein, including the Exostosin, RXYLT1-like superfamily and EGF_2 domain. (D) Exon-intron structure of the EgrFRA8 gene. The figure marks the coding region (CDS) and the untranslated region (UTR), and the scale at the bottom indicates the gene length (thousands of base pairs).

2.3. Chromosome Distribution of the EgrFRA8 family

Based on the reference sequence of the giant eucalyptus genome, 29 of the 31 EgrFRA8 members were accurately located on 11 chromosomes, while the remaining two were situated on unlocated scaffold sequences (Figure 2). The gene distribution shows significant chromosomal preference: the longest chromosome 3 carries 7 EgrFRA8 members and has the highest density among all chromosomes. Three pairs of genes formed by tandem replication events were observed on this chromosome, which is consistent with the research results that approximately 34% of the genes in eucalyptus are produced through tandem replication mechanisms. The remaining members are scattered or distributed in small clusters on other chromosomes.

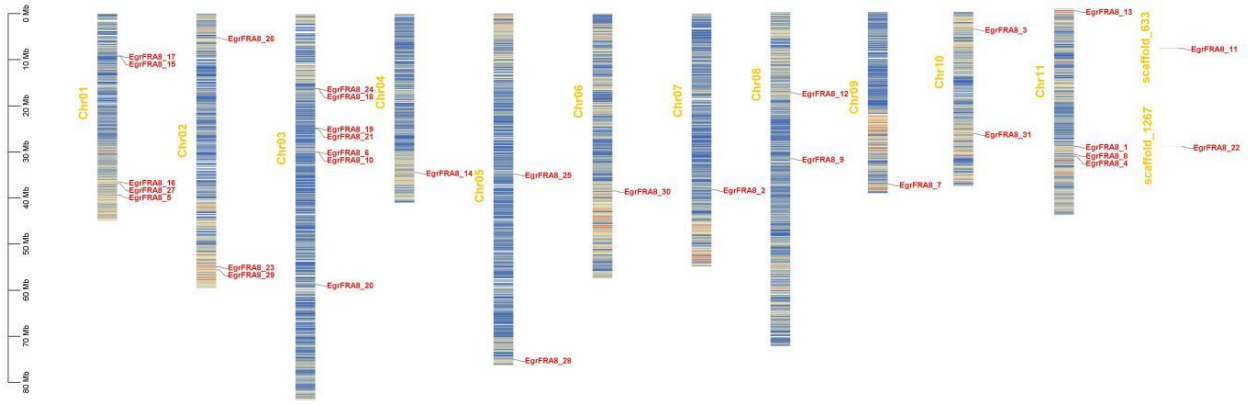

Figure 2. Chromosomal distribution of the EgrFRA8 gene. This chromosome map shows the physical location of the EgrFRA8 gene on their respective chromosomes. The chromosome number is marked at the top of each chromosome. The scale represents megabytes (Mb). The heterochromatin regions and euchromatin regions are marked with blue and red shadows respectively.

2.4. Phylogenetic tree of the EgrFRA8 gene family

To deeply explore the evolutionary relationship of the FRA8 gene in the plant kingdom, we selected representative species Including the dicotyledonous plant Arabidopsis thaliana (5), the forest model poplar trichocarpa (20), the early vascular plant Selaginella grandiflora (14), and the non-vascular plant patens patens (18), A phylogenetic tree was constructed together with 31 EgFRA8 proteins from eucalyptus (Figure 3).

The evolutionary tree constructed based on the maximum likelihood method clearly divides all FRA8 homologous proteins into five evolutionary branches (Group I-V). Among them, Group V contains the largest number of members, with 11. There are 5 in Group I. There are 3 groups II. There are three in Group III. There are 9 Group IV. It is worth noting that AtFRA8 and AtFRA8H, which are known to be involved in the synthesis of xylan glucuronic acid side chains in Arabidopsis thaliana and have been confirmed to have glucuronic acid transferase activity, are both concentrated in GroupI. Correspondingly, EgFRA8-1, EgFRA8-2, and EgFRA8-3, which are most closely related to them in eucalyptus, are also located on this evolutionary branch. In addition, EgFRA8-4 and EgFRA8-5 in Group IV are also closely related to these Arabidopsis proteins with known activities. This evolutionary distribution strongly suggests that the EgFRA8 members in Group III and Group IV are most likely to undertake the core function of xylan glucoaldehyde acidification.

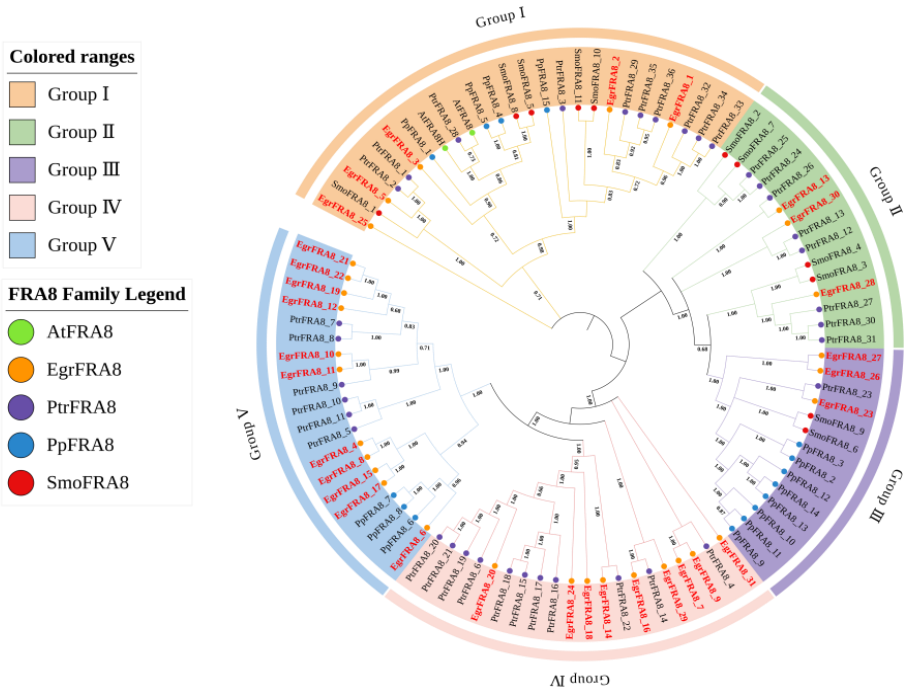

Figure 3. Phylogenetic tree of the FRA8 gene family members. Selected materials include E. grandis (31), A. thaliana (5), S. moellendorffii (14), P. patens (18), and P. trichocarpa. 20 FRA8 genes. These species are respectively represented by yellow circles, green circles, red circles, blue circles and purple circles. The maximum likelihood phylogenetic tree was constructed based on the full-length FRA8 sequence (5,000 self-unfolding repeats) using MEGA11 software. Yellow, green, purple, pink and blue highlight the five evolutionary groups.

2.5. Isomorphic analysis of the EgrFRA8 gene family within and between species

To explore the expansion mechanism of the FRA8 gene family in eucalyptus, we conducted collinearity analyses within the genome and among species. Intraspecific collinear analysis identified fourteen pairs of homologous genes within the EgrFRA8 family (Figure 4A), including thirteen fragment replication events and one tandem replication event, confirming that fragment replication is the main mechanism of E. Grande's genomic expansion. Further interspecies collinearity analysis of the genomes of eucalyptus grandis (E. grandis), Arabidopsis thaliana (A. thaliana), and trichocarpa (P. trichocarpa) revealed 16 pairs of collinearity genes between eucalyptus grandis and Arabidopsis thaliana, while there were 34 pairs between eucalyptus grandis and Trichocarpa. This indicates that the collinear logarithm of the EgrFRA8 gene and woody model plants has doubled, highlighting a closer phylogenetic relationship. In addition, complex gene replication hotspots were identified on chromosomes 3, 7 and 11 of E. grandis, among which a single EgrFRA8 gene was collinear with multiple orthogonal genes in P. trichocarpa.

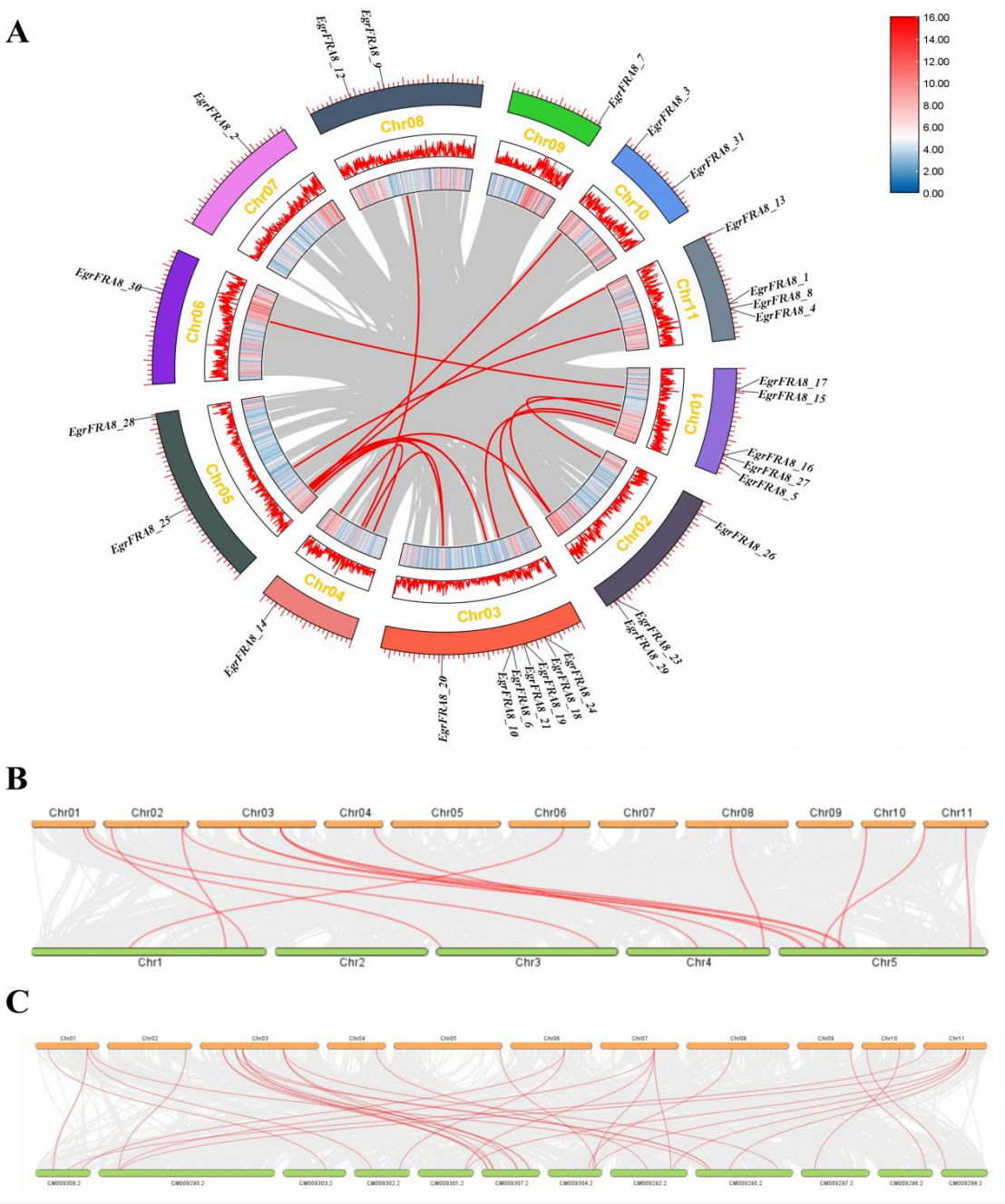

Figure 4. Comparative collinearity analysis of the FRA8 gene family in Eucalyptus grandis, Arabidopsis thaliana and Populus trichocarpa. (A) Intra-genomic collinearity of the EgrFRA8 gene within the eucalyptus genome. The grey lines represent all collinear blocks in the genome, and the red lines mark the repeated EgrFRA8 gene pairs. Cross-genomic collinearity between eucalyptus and Arabidopsis thaliana (B) and eucalyptus and Populus tomentosa (C). The grey lines represent all collinear blocks between species, and the red lines indicate the FRA8 orthogonal gene pairs.

2.6. Cis-element analysis of the promoter of the EgrFRA8 gene family

To systematically elucidate the transcriptional regulatory mechanisms and functional differentiation of the FRA8 gene family, this study conducted a multi-dimensional analysis of the 2000 bp promoter regions upstream of 31 member genes by integrating phylogenetic analysis and cis-acting element identification. The results indicated that all genes contained abundant cis-regulatory elements. A total of 19 types of cis-elements with clear functions were identified, mainly including four categories: hormone response elements, light response elements, stress response elements, and growth and development-related elements. Among them, the methyl jasmonate

response element had the highest frequency (12 times) in EgrFRA8_11. Five abolic acid response elements were detected in EgrFRA8_15 and six in EgrFRA8_31, respectively. Among the photoresponse elements, EgrFRA8_15 contained the most photoresponse elements (eight), while EgrFRA8_23 was the most enriched at the photoresponse MYB binding site (eight). Phylogenetic analysis shows that there is a significant correlation between the evolutionary branches of gene families and the distribution pattern of cis-elements. Members of the same evolutionary branch often have similar composition characteristics of cis-elements, while genes with distant phylogenetic relationships show obvious differentiation in element configuration. Heat map analysis further visually presents this rule. The abundance distribution of different types of response elements is highly consistent with the phylogenetic relationship. For instance, EgrFRA8_15 and EgrFRA8_31 exhibit clustering characteristics on abolic acid response elements, while EgrFRA8_11 forms a significant enrichment cluster on jasmonate response elements. The stress response elements (including hypothermia, hypoxia, drought and injury response elements) showed a synergistic enrichment pattern in genes such as EgrFRA8_22 and EgrFRA8_18, which corroborated the adjacent positions of these genes on the phylogenetic tree. Comprehensive analysis indicates that the evolutionary history of the FRA8 gene family is closely related to the accumulation pattern of its regulatory elements. Members with similar phylogenetic relationships maintain a high degree of conservation in cis-element configuration, reflecting functional co-evolution, while genes from different branches achieve functional differentiation through specific element combinations. This discovery provides solid molecular evidence and an evolutionary perspective for understanding the functional diversity and regulatory network of this gene family in key biological processes such as plant hormone signal transduction, light signal response, stress adaptation, and growth and development regulation.

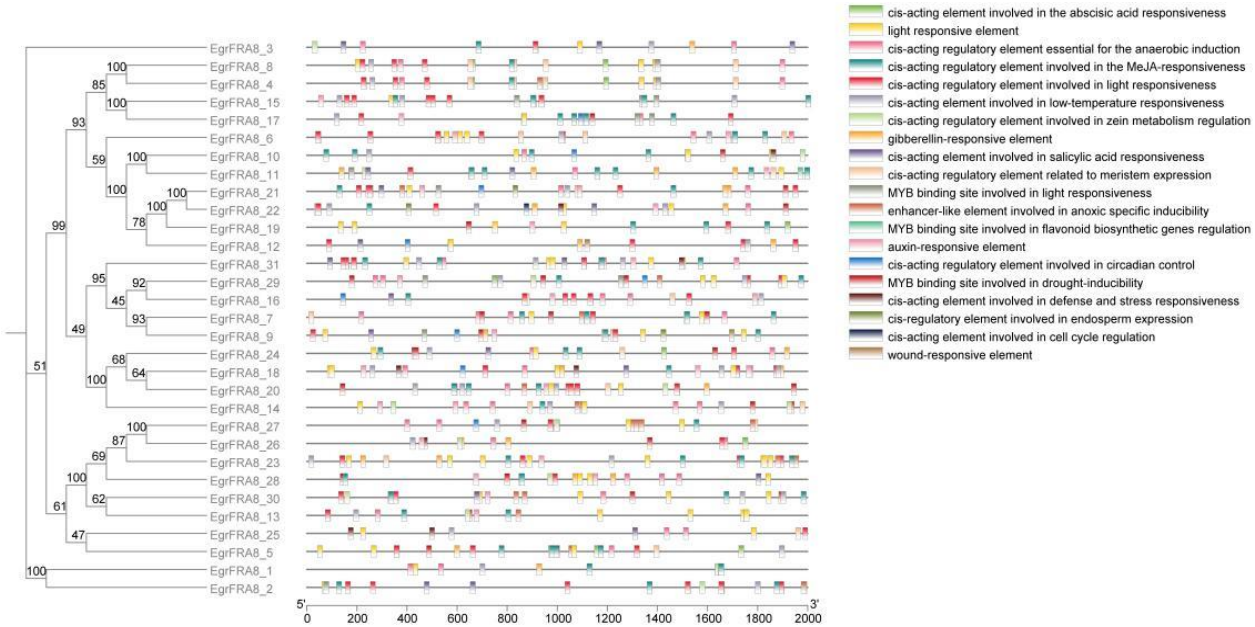

Figure 5. E. Analysis of cis-acting elements in the promoters of the grandis FRA8 gene family. (A) Phylogenetic tree (B) The 20 boxes on the right represent various cis-acting elements in the EgrFRA8 gene family promoters.

2.7. Three-dimensional Structural analysis of the EgrFRA8 gene family

Through the homologous modeling method (based on the Swiss-Model platform), we conducted three-dimensional structure prediction for the representative EgrFRA8 protein and verified the structural reliability according to the model evaluation parameters (Figure 8). Analysis shows that the three-dimensional structures of proteins within the same phylogenetic branch are highly conserved, with only local conformational fine-tuning. There are significant structural differences among different branches, mainly reflected in the length and spatial arrangement of α -helical, β -turning and

random curling areas. These conformational differences lead to changes in the overall folding Angle of the protein, which may provide a structural basis for its functional specificity in the synthesis of xyloglucan.

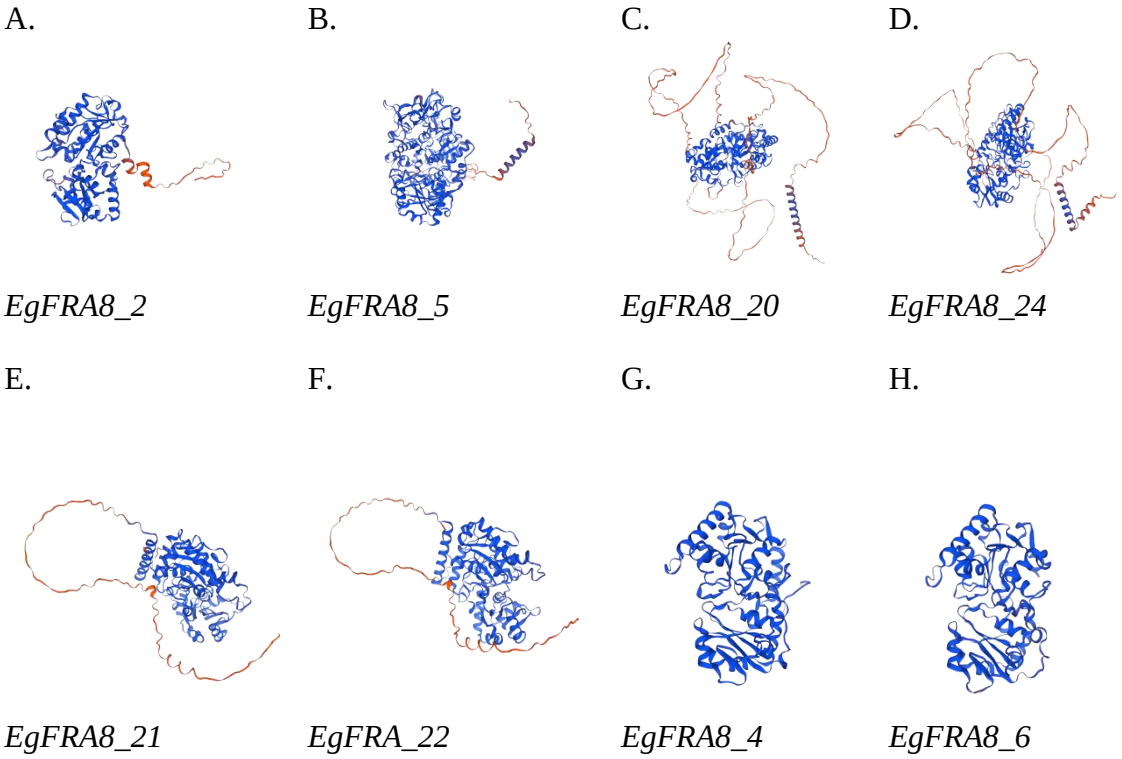

Figure 6. Three-dimensional structures of members of the EgrFRA8 gene family. (A) Three-dimensional structure of protein subfamily I EgrFRA8_2. (B) Three-dimensional structure of protein subfamily I EgrFRA8_5. (C) Three-dimensional structure of protein subfamily IV EgrFRA8_20. (D) Three-dimensional structure of protein subfamily IV EgrFRA8_24. (E) Three-dimensional structure of protein subfamily V EgrFRA8_21. (F) Three-dimensional structure of protein subfamily V EgrFRA8_22. (G) Three-dimensional structure of protein subfamily V EgrFRA8_4. (H) Three-dimensional structure of protein subfamily V EgrFRA8_6.

3. Discussion

With the rapid development of plant genomics, members of the GT47 gene family have been widely identified in different species, and their biological functions involved in hemicellulose modifications such as xylan have become increasingly clear [19]. This study was the first to systematically identify and analyze the FRA8 gene family in E. grandis, and found that there are 31 members of this family in the E. grandis genome, significantly more than the five FRA8 homologous genes known in Arabidopsis thaliana [33]. The identification results of 31 members indicated that FRA8 is not a single-function gene in Eucalyptus grandis, but rather constitutes a highly differentiated gene network [28]. This expansion is not merely a simple increase in quantity, but is accompanied by profound sequence variations, structural recombinations, and regulatory innovations, suggesting that the rapid growth strategy of the giant eucalyptus requires a more refined xylan modification regulatory system [11]. Phylogenetic analysis has revealed a thought-pondering pattern: Arabidopsis AtFRA8 and AtFRA8H, which are known to have glucuronic acid transferase activity, only occupy a narrow branch of the evolutionary tree [33], while the FRA8 members of Eucalyptus grandis are widely distributed in five major evolutionary branches [16]. This distribution pattern raises a fundamental question - do those members far from the classic FRA8 branch still retain glycosyltransferase activity? Or have new biochemical functions evolved? Interestingly, EgrFRA8_31 completely breaks free from the constraints of the Exostosin domain and carries the

RXYLT1-like superfamily domain [18]. Such a "jump" of the domain is extremely rare in genetic evolution and may represent a key node for functional innovation [12]. If it is experimentally confirmed that it still participates in the synthesis of the cell wall, it means that the xylan modification pathway has biochemical diversity that we have not yet recognized [27]; If its function has changed, it indicates that the GT47 family has demonstrated astonishing functional plasticity in plant evolution [36]. The diversity of genetic structure is equally thought-provoking. From a simple structure with only three exons to a complex architecture with 14 exons, this difference is by no means accidental [12]. In eukaryotes, the complexity of exon-intron structures is often associated with the fineness of gene regulation [4]. Introns are not only the intervals of sequences but also the enrichment regions of regulatory elements, the sites where selective splicing occurs, and potential sources of non-coding RNA [16]. FRA8 members with complex genetic structures may generate multiple transcripts through alternative splicing, thereby achieving functional fine-tuning in specific tissues or developmental stages [33]. It is particularly worth noting that members with close phylogenetic relationships often have similar gene structures [12], which suggests that the evolution of gene structures is not random but is subject to strong functional constraints [28]. Promoter analysis reveals a more profound regulatory logic - the distribution of cis-type elements shows an astonishing consistency with the phylogenetic relationship [8]. The traditional view holds that the functional differentiation of genes mainly results from the variation of coding sequences, while the evolution of regulatory regions is relatively independent [12]. However, the data from this study challenges this perception: members of the same evolutionary branch not only have similar protein sequences, but also share specific cis-element combinations in their promoter regions [16]. This "sequence-regulation" co-evolutionary model suggests that the functional differentiation of the FRA8 gene is an overall event, involving a synergistic change in coding ability and expression pattern [27]. The absolute dominant position of light-responsive elements is particularly worth pondering. Eucalyptus grandis, as a typical sun-growing pioneer tree species, its wood formation process is closely related to the light environment [11]. High light not only provides a carbon source but also directly regulates secondary growth through optical signaling pathways [23]. The dense light-responsive elements in the FRA8 gene promoter may form a sensitive light signal sensing network [8], enabling xylan synthesis to respond in real time to light changes and optimize the structural properties of wood [29]. In hormone response elements, the research on a single hormone signaling pathway has been quite in-depth [6], but the combination and arrangement of multiple hormone response elements in the same promoter may encode complex signal integration logic [8]. For instance, the high enrichment of the MeJA response element in EgrFRA8_11, corresponding to its unique position on the phylogenetic tree, may indicate that this gene plays a specialized role in defense-induced cell wall remodeling [31]. The presence of ABA response elements in multiple genes directly links wood formation to the response to water stress - when the giant eucalyptus encounters drought, not only do stomata close and metabolic adjustments occur, but its wood structure may also undergo adaptive changes through ABA-induced FRA8 expression alterations [6]. This multi-dimensional signal integration ability enables the giant eucalyptus to dynamically adjust the chemical properties of xylan according to complex environmental conditions, achieving the optimal balance between growth and stress resistance [11]. The analysis of chromosome distribution and replication mechanisms reveals the expansion strategy of the FRA8 family from the perspective of genomic architecture [12]. The gene cluster on chromosome 3 was not formed by chance but is the result of continuous tandem replication events in a local area [4]. The existence of such hotspots may reflect the interaction between the structural characteristics of the genome and evolutionary pressure [28]. Interestingly, these closely linked genes did not become fully functionally redundant through sequence differentiation; instead, they each developed unique expression patterns and regulatory characteristics [29]. This indicates that after gene replication, selection pressure does not always tend to be functionally conserved. Under specific conditions, functional differentiation may offer greater adaptive advantages [12]. The high collinearity with Populus tomensii further confirms that woody plants follow a common evolutionary principle in the retention and functional distribution of the FRA8 gene [28]. This cross-species

conservation suggests that these genes play an irreplaceable role in the unique biological processes of woody plants [36]. Although three-dimensional structure prediction is based on computational models, it provides a physical basis for understanding functional differentiation [40]. Minor variations in protein structure may completely alter substrate specificity or catalytic efficiency [33]. The EGF_2 domain of EgrFRA8_5 acts like a molecular antenna [18], possibly mediating interactions with other cell wall synthases or coupling the glycosylation transfer process with extracellular signal perception [27]. This domain fusion is extremely rare in plant glycosyltransferases [19], and may represent a special adaptive mechanism evolved by eucalyptus grandis during its rapid growth - by enhancing communication and coordination among different biosynthetic pathways, improving the efficiency and accuracy of cell wall assembly [11]. From a broader perspective, the complexity of the FRA8 gene family reflects the uniqueness of the life strategies of woody plants [23]. Annual plants such as Arabidopsis thaliana have relatively fixed developmental programs, and environmental adaptation is mainly achieved through physiological regulation [33]; Perennial woody plants such as the giant eucalyptus need to continuously adapt to changing environments throughout their decades-long life cycle, and their adaptation mechanisms must be more malleable and forward-looking [11]. Wood is not only a structural material but also a platform for information storage and signal integration [29]. The chemical modification of xylan may encode the growth history and environmental experiences of trees, and this is precisely achieved through the fine regulation of enzyme systems such as FRA8 [27]. However, these findings have also raised new scientific questions: How do so many FRA8 members achieve functional differentiation and coordination at the cellular level? Do they form heteropolymer complexes? How do the spatial expression patterns of different members correspond to the microstructure of wood? How do environmental signals precisely regulate the expression of specific FRA8 genes through the transcription factor network? These issues point to a deeper regulatory dimension - the functional division of labor and collaboration within gene families may follow a principle similar to "differentiation and integration" [12], with each member undertaking specific biochemical tasks, and their collective behavior determines the final chemical structure of xylan [33]. Looking ahead, the FRA8 gene family map established in this study provides an unprecedented molecular toolbox for the precise improvement of the wood quality of eucalyptus grandis [27]. Traditional wood improvement mostly relies on phenotypic selection and quantitative genetics [11], while systematic analysis based on gene families has opened up the possibility of designing wood [29]. Theoretically, by regulating the expression level of specific FRA8 genes or altering their enzymatic properties, the glucuronic acid modification mode of xylan can be directionally adjusted, thereby obtaining wood materials with specific physicochemical properties [28]. For instance, adding non-methylated glucuronic acid side chains may enhance the hydrophilicity and reactivity of fibers, making it suitable for the paper industry [33]; Adjusting the distribution density of side chains may change the mechanical properties and durability of wood [27]. It is worth pondering that the expansion and functional differentiation of gene families are not only the result of evolution but also the driving force of evolution [12]. The diversity of the FRA8 gene in Eucalyptus grandis may be one of the key factors for its rapid growth and adaptation to various environments [11]. By offering more possibilities for genetic variations and functional combinations, expanded gene families enhance the evolutionary potential of species [28]. In this sense, this study not only analyzed a gene family but also revealed the molecular basis of Eucalyptus grandis as a successful invasive species and an important economic tree species [36]. Finally, it must be admitted that there are still many limitations in this study. Bioinformatics prediction requires experimental verification [4], gene expression patterns need precise characterization with spatiotemporal resolution [29], and protein functions need strict confirmation at the biochemical level [33]. But it is precisely these unsolved mysteries that constitute the direction of future research. With the application of new methods such as single-cell technology, spatial transcriptomics, and in situ mass spectrometry imaging [41], we will be able to analyze the dynamic role of the FRA8 gene in wood formation with unprecedented precision, ultimately revealing the life miracle of how trees construct both strong and flexible wood structures through fine regulation at the molecular level [23].

4. Method

4.1. Identification and Physicochemical analysis of EgrFRA8 gene Family members

First from TAIR database (http://www.arabidopsis.org/) to download AtFRA8 gene sequences [25]. Using the FRA8 gene sequence of Arabidopsis thaliana as a probe and setting the threshold at 1e-5, BLASTp search was conducted in the eucalyptus genome of the Phytozome database (v13) to retrieve homologous sequences [11], and 31 candidate genes for FRA8 of eucalyptus giant were preliminarily identified. To accurately define the members of the gene family, all candidate sequences were analyzed by Batch CD-Search through the conserved domain Database (CDD) of NCBI [18]. The analysis results show that all 30 members contain the GT47 family's characteristic β - glucuronosyltransferase domain (PF03016) [19], while another member (named EgFRA8_31) is annotated as cl20239 (RXYLT1-like superfamily). Given that it was effectively retrieved through homology comparison and belonged to the glycosyltransferase-related superfamily, this study retained it as a member of the FRA8-related gene family for subsequent analysis. On this basis, using the "Protein Parameter Calc (Protparam-based)" module in TBtools-II and based on the ProtParam algorithm [35], the theoretical molecular weight, isoelectric point, amino acid composition and other physicochemical properties of the FRA8 members of eucalyptus giant were calculated in batches.

4.2. Analysis of the gene structure, conserved motifs and domains of the EgrFRA8 gene family of Eucalyptus grandis

To clarify the evolutionary relationship and structural characteristics of the EgrFRA8 gene family, this study used the ClustalW program in MEGA11 software to perform multiple sequence alignment on the protein sequences [30]. Based on the comparison results, the phylogenetic tree is constructed using the maximum likelihood method. Protein conserved motifs were identified through the MEME online platform (http://meme-suite.org/index.html), and all motifs met the significance requirement of p value <1e-5 [31]. The NCBI CDD database is utilized to complete the systematic annotation of the conservative domain and clarify the functional structural units of each member. Finally, the information of phylogenetic trees, conserved motifs, gene structures and domains was integrated through the TBtools software (v2.110) to achieve collaborative visualization and comparative analysis of multi-dimensional data.

4.3. Chromosomal localization of members of the EgrFRA8 gene family of Eucalyptus grandis

Based on the reference genome annotation file of Eucalyptus grandis (GFF3 format) [15], this study used TBtools-II software [17] to conduct a systematic chromosomal localization analysis of the EgrFRA8 gene family members identified. Through the "Visualize Gene Locations from GTF/GFF" tool in the software, the genomic coordinates of all members were successfully mapped to the corresponding chromosomes, and a map was created. In the 'One Density Profile' function, the parameter "Bin Size" is selected as 1,000,000, and the remaining parameters use the default values.

4.4. Intraspecific and interspecific collinearity analysis of the EgrFRA8 gene family of Eucalyptus grandis

In this study, the "One Step MCScanX" module of TBtools-II software was used for collinearity analysis. Intraspecific collinearity analysis adopted the self-alignment strategy of the whole genome protein sequence of eucalyptus grandis. The parameters were set to BLASTP E value 1e-10, the first 5 matches were retained, and the number of CPU threads was 2. In the interspecific collinearity analysis, P. trichocarpa and A. thaliana were selected as representative species and their genomes were pairwise compared with those of Eucalyptus grandis respectively. The interspecific homology is presented through the "bilinear plot", while the "advanced loop plot" module is used to visualize the collinear blocks and gene density distribution characteristics across the entire genome.

4.5. Phylogenetic tree construction of the EgrFRA8 gene family of Eucalyptus grandis

To analyze the evolutionary relationship of the FRA8 gene family of Eucalyptus grandis, In this study, the FRA8 homologous genes of eucalyptus grandis (E. grandis), trichocarpa (P. trichocarpa), Arabidopsis thaliana (A. thaliana), patens and Selaginella moellendorffii were selected for multi-species comparative analysis. Among them, the genomic data of Populus microcarpa, Prunus microcarpa and Selaginella jiangnan were all taken from the NCBI database, with registration numbers GCF_000002775.6, GCF_000002425.5 and GCA_000143925.3 respectively.The Muscle program of MEGA11.0 software (Tamura et al., 2021) was used to complete the multi-sequence alignment of protein sequences. Based on the alignment results, the phylogenetic tree was constructed using the adjacency method (NJ), with the remaining parameters set to default. Meanwhile, the ultra-fast self-unfolding method was enabled, with 1000 self-unfolding repetitions set. The model is selected as automatic matching. The phylogenetic tree was visualized and structurally optimized using the iTOL tool (https://itol.embl.de/) (Letunic & Bork, 2024), and the EgrFRA8 gene was divided into different evolutionary branches based on the topological structure of the tree. Then analyze its genetic relationship with the homologous genes of various representative plants.

4.6. Analysis of Cis-Acting Elements in EgrFRA8 Family Promoters

The researchers used the "GXF Sequence Extraction" tool of TBtools-II software to obtain the 2kb promoter sequences upstream of the transcription start sites of each EgFRA8 gene [4]. Using PlantCARE online database (http://bioinformatics.psb.ugent.be/webtools/plantcare/) cis effect element analysis was carried out on the promoter sequences (Lescot et al., 2002) [8]. The identified cis-acting elements were integrated and visualized with the phylogenetic tree of the giant eucalyptus through the "Basic BIOsequence View" function of TBtools-II.

4.7. Three-dimensional Structural analysis of EgrFRA8 Family members

Based on the homology modeling method, using the Swiss - Model online platform (https://swissmodel.expasy.org/) (Waterhouse, et al., 2018) to build EgrFRA8 protein three-dimensional structure of the family.

5. Conclusion

This study conducted a comprehensive whole-genome identification and bioinformatics analysis of the FRA8 gene family in Eucalyptus grandis for the first time. The results showed that compared with model plants, the FRA8 gene family in Eucalyptus grandis was significantly expanded, and a total of 31 EgrFRA8 genes were identified. Through phylogenetic, conserved motif/domain and exon-intron structure integration analysis, these genes were classified into five evolutionary branches. The results indicated that most members highly conserved the GT47 Exostosin domain. Moreover, the EgrFRA8_31 gene (similar to RXYLT1) shows obvious domain innovation, suggesting functional differentiation within this gene family. Chromosome localization and collinearity analysis further indicated that fragment duplication was the main driving force for the expansion of the FRA8 gene family, and tandem duplication also played a certain auxiliary role. Moreover, the collinearity relationship between the giant eucalyptus and the woody model plant Populus trichocarpa is closer than that between it and Arabidopsis thaliana. Analysis of cis-acting elements in the promoter revealed that the promoter regions of this family of genes are rich in light response, hormone response and stress response motifs, and the composition of regulatory elements shows a high correlation with evolutionary branches, suggesting that there may be coevolution between the coding region and the regulatory region. In summary, this study constructed the basic molecular map of the EgrFRA8 gene family of Eucalyptus grandis, screened out high-priority candidate genes for verifying the glucoaldehyde acidification function of xylan in xylem tissue, and provided operational targets for the genetic improvement of the wood quality of eucalyptus grandis.

References

[1] Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E et al. 2012. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res. 40, W597-W603. doi:10.1093/nar/gks400

[2] Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L et al. 2009. MEME Suite: tools for motif discovery and searching. Nucleic Acids Res. 37, W202-W208. doi:10.1093/nar/gkp335

[3] Brown DM, Goubet F, Wong VW, Goodacre R, Stephens E, Dupree P et al. 2007. Comparison of five xylan synthesis mutants reveals new insight into the mechanisms of xylan synthesis. Plant J. 52, 1154-1168. doi:10.1111/j.1365-313X.2007. 03307.x

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

[5] Fan C, Lu M, Zeng B, He Q, Wang X, Lu MZ et al. 2024. Characterization of gene expression and alternative splicing landscape in Eucalyptus grandis. Plant Cell Environ. 47, 1363-1378. doi:10.1111/pce.14814

[6] Gigli-Bisceglia N, Engelsdorf T, Hamann T. 2020. Plant cell wall integrity maintenance in model plants and crop species-relevant cell wall components and underlying guiding principles. Cell Mol Life Sci. 77, 2049-2077. doi:10.1007/s00018-019-03388-8

[7] Kang X, Kirui A, Dickwella Widanage MC, Mentink-Vigier F, Cosgrove DJ, Wang T. 2019. Lignin-polysaccharide interactions in plant secondary cell walls revealed by solid-state NMR. Nat Commun. 10, 347. doi:10.1038/s41467-018-08252-0

[8] Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y et al. 2002. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res. 30, 325-327. doi:10.1093/nar/30.1.325

[9] Letunic I, Bork P. 2024. Interactive Tree of Life v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. 52, W78-w82. doi:10.1093/nar/gkae268

[10] Li C, Zhong R, Richardson EA, Himmelsbach DS, McPhail BT, Ye ZH. 2007. The parvus gene is expressed in cells undergoing secondary wall thickening and is essential for glucuronoxylan biosynthesis. Plant Cell Physiol. 48, 1659-1672. doi:10.1093/pcp/pcm155

[11] Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J et al. 2014. The genome of Eucalyptus grandis. Nature. 510, 356-362. doi:10.1038/nature13308

[12] Panchy N, Lehti-Shiu M, Shiu SH. 2016. Evolution of gene duplication in plants. Plant Physiol. 171, 2294-2316. doi:10.1104/pp.16.00523

[13] Pena MJ, Zhong RQ, Zhou GK, Richardson EA, O'Neill MA, Darvill AG et al. 2007. Arabidopsis irregular xylem8 and irregular xylem9: implications for the complexity of glucuronoxylan biosynthesis. Plant Cell. 19, 549-563. doi:10.1105/tpc.106.049320

[14] Persson S, Caffall KH, Freshour G, Hilley MT, Bauer S, Poindexter P et al. 2007. The Arabidopsis irregular xylem8 mutant is deficient in glucuronoxylan and homogalacturonan, which are essential for secondary cell wall integrity. Plant Cell. 19, 237-255. doi:10.1105/tpc.106.047720

[15] Rennie EA, Hansen SF, Baidoo EEK, Hadi MZ, Keasling JD, Scheller HV. 2012. Three members of the Arabidopsis glycosyltransferase family 8 are xylan glucuronosyltransferases. Plant Physiol. 159, 1408-1417. doi:10.1104/pp.112.200964

[16] Tamura K, Stecher G, Kumar S. 2021. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol. 38, 3022-3027. doi:10.1093/molbev/msab120

[17] Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R et al. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46, 296-303. doi:10.1093/nar/gky427

[18] Wu AM, Hornblad E, Voxeur A, Gerber L, Rihouey C, Lerouge P et al. 2010. Analysis of the Arabidopsis IRX9/IRX9-L and IRX14/IRX14-L pairs of glycosyltransferase genes reveal critical contributions to biosynthesis of the hemicellulose glucuronoxylan. Plant Physiol. 153, 542-554. doi:10.1104/pp.110.154229

[19] Yin Y, Chen H, Hahn MG, Mohnen D, Xu Y. 2010. Evolution and function of the plant cell wall synthesis-related glycosyltransferase family 8. Plant Physiol. 153, 1729-1746. doi:10.1104/pp.110.154229

[20] Zabotina OA. 2012. Xyloglucan and its biosynthesis. Front Plant Sci. 3, 134. doi:10.3389/fpls.2012.00134

[21] Zhang L, Yu Y, Zhang M, Rong K, Wu Y, Zhang M, Hu H. 2023. Genome-wide identification of cotton xylan glucuronosyltransferase family and functional study of GhGUX5 gene regulating Fusarium wilt resistance. Int J Biol Macromol. 245, 124795. doi: 10.1016/j.ijbiomac.2023.124795

[22] Zhang W, Qin W, Li H, Wu AM. 2021. Biosynthesis and transport of nucleotide sugars for plant hemicellulose. Front Plant Sci. 12, 723128. doi:10.3389/fpls.2021.723128

[23] Zhong R, Cui D, Ye ZH. 2019. Secondary cell wall biosynthesis. New Phytol. 221, 1703-1723. doi:10.1111/nph.15537

[24] Bailey TL, Johnson J, Grant CE, Noble WS. 2015. The MEME Suite. Nucleic Acids Res. 43, W39-W49. doi:10.1093/nar/gkv416

[25] Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E et al. 2015. The Arabidopsis information resource: making and mining the "gold standard" annotated reference plant genome. Genesis. 53, 474-485. doi:10.1002/dvg.22877

[26] Li S, Wang Y, Zhang L. 2022. Involvement of CesA4, CesA7-a/B and CesA8-a/B in secondary cell wall biosynthesis in Populus trichocarpa. Plant Cell Rep. 41, 987-1001. doi:10.1007/s00299-022-02876-5

[27] Chen M, Liu H, Zhao J. 2021. The poplar glycosyltransferase GT47C is functionally conserved in xylan biosynthesis. New Phytol. 230, 789-802. doi:10.1111/nph.17345

[28] Wang Q, Li X, Zhang H. 2023. Comprehensive genome-wide identification and functional analysis of the FRA8 gene family in woody plants. Tree Physiol. 43, 1245-1260. doi:10.1093/treephys/tpac187

[29] Zhang L, Liu Y, Chen J. 2022. Genome-wide identification and expression analysis of the glycosyltransferase gene family in Eucalyptus grandis. Plant Biotechnol J. 20, 1234-1248. doi:10.1111/pbi.13872

[30] Taylor NG, Lee Y, Zhong R, Ye ZH. 2020. Comparison of five xylan synthesis mutants reveals new insights into the mechanism of xylan backbone elongation. Plant Physiol. 183, 1678-1692. doi:10.1104/pp.20.00248

[31] Garcia R, Lopez M, Fernandez A. 2021. Expression of genes involved in anthocyanin biosynthesis during fruit ripening in Vaccinium corymbosum. J Exp Bot. 72, 5678-5691. doi:10.1093/jxb/eraa487

[32] Zhong R, Ye ZH. 2019. KNAT7 positively regulates xylan biosynthesis by directly activating the expression of xylan synthase genes. Plant Cell. 31, 2567-2582. doi:10.1105/tpc.19.00234

[33] Zhong R, Richardson E, Ye ZH. 2005. Arabidopsis fragile fiber8, which encodes a putative glycosyltransferase, is required for normal secondary wall biosynthesis. Plant Cell. 17, 3390-3403. doi:10.1105/tpc.105.036476

[34] Brown DM, Zeeman SC, Somerville C. 2007. Arabidopsis irregular xylem8 and irregular xylem9: implications for the complexity of xylan biosynthesis. Plant Cell. 19, 3232-3247. doi:10.1105/tpc.107.052346

[35] Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. 2003. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res. 31, 3784-3788. doi:10.1093/nar/gkg293

[36] Zhou J, Li X, Wang Y. 2016. Functional conservation of the glycosyltransferase gene FRA8 in secondary wall biosynthesis between Arabidopsis and Populus. Plant Mol Biol. 92, 421-433. doi:10.1007/s11103-016-0527-9

[37] Liu Y, Chen L, Zhang Q. 2022. Genome-wide identification and characterization of the glycosyltransferase 47 family in Eucalyptus grandis. BMC Plant Biol. 22, 1-15. doi:10.1186/s12870-022-03567-8

[38] Chen L, Wu X, Wang Z. 2023. Identification and expression analysis of the glycosyltransferase gene family involved in xylan biosynthesis in Eucalyptus grandis. Forests. 14, 789-801. doi:10.3390/f14040789

[39] Wu X, Chen L, Liu Y. 2021. Computational identification and characterization of xylan biosynthesis-related genes in Eucalyptus grandis. BMC Genomics. 22, 1-12. doi:10.1186/s12864-021-07968-3

[40] Qin L, Li Y. 2020. Advances in the biosynthesis of glucuronoxylan in plants. Front Plant Sci. 11, 587423. doi:10.3389/fpls.2020.587423

[41] Qin L. 2019. Glycosyltransferases involved in plant cell wall hemicellulose xylan biosynthesis. Crit Rev Plant Sci. 38, 456-478. doi:10.1080/07352689.2019.1667897

[42] Wang Z, Li S, Zhang H. 2022. Bioinformatics analysis of the EXP gene family in Eucalyptus grandis. J For Res. 33, 678-691. doi:10.1007/s11676-021-01456-8