



RESEARCH PAPER

Open Access



De novo transcriptome assembly reveals characteristics of flower sex determination of *Excoecaria agallocha*

Yan Zhou¹, Lulu Hao^{1,2}, Lexiang Huang¹, Xiaoming Tang¹, Danting Zhuo¹, Li Yun Wang^{1*} and Ying Zhang^{1*}

Abstract

Key message: MADS-box genes family may play important roles in the flower sex determination in *Excoecaria agallocha*. A total of 73 MADS-box genes were identified in *E. agallocha*. De novo transcriptome assembly analysis suggested that *AP1/FUL*, *AP3/PI*, *AGL104*, and *SOC1* plays potential roles in *E. agallocha* flower sex determination.

Context: *Excoecaria agallocha* is a dioecious species containing both male and female individuals producing unisexual flowers. The underlying mechanisms regulating inflorescence differentiation in these species remains poorly understood.

Aims: Flower development influences reproduction and breeding in *E. agallocha*, which contributes to ecological restoration in the intertidal zone.

Methods: We performed de novo transcriptome assembly analysis on male and female flowers and leaves from *E. agallocha*.

Results: We identified a total of 73 MADS-box genes in the *E. agallocha* genome, which we grouped into five distinct classes (MIKC^C, Ma, Mb, My, MIKC*) after phylogenetic comparisons with *J. curcas* homologs. We analyzed expression profiles and suggested *AP1/FUL*, *AP3/PI*, *AGL104*, and *SOC1* as candidate regulators of sex determination. In addition, several *EaMADS* genes were significantly upregulated in flowers compared to leaves.

Conclusion: Our study represents the first detailed, comprehensive evaluation of the regulation of MADS-box genes associated with sex determination in *E. agallocha*. The assembled transcriptomic profiles increase the genetic information available for this species and constitute an important contribution to functional studies of inflorescence genes. In turn, this will help exploring the molecular mechanisms underlying the evolution of *E. agallocha*.

Keywords: *E. agallocha*, MADS-box, Sex determination

1 Introduction

Euphorbiaceae is a family of plants with significant medicinal and economic relevance, containing ~ 7500 species belonging to 300 genera and 5 different subfamilies

(Webster 1994). The genus *Excoecaria* comprises nearly 40 species that are distributed across the mangrove regions of the old-world, and encompasses dioecious plants containing fruits with a 3-lobe schizocarp (Webster 1994; Mondal et al. 2016). *Excoecaria agallocha* L., often referred to as “milk mangrove”, has relevant medicinal properties and belongs to the few non-viviparous mangrove species distributed across East Africa, India, and Ceylon to China and the Ryu-Kyu Islands (Tomlinson 2016). This plant can be found in mangrove regions

Handling editor: Véronique Jorje.

*Correspondence: 2368581774@qq.com; zhangyingred@lingnan.edu.cn

¹ School of Life Sciences and Technology, Lingnan Normal University, Zhanjiang, China

Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

and occasionally inland. Unlike the majority of mangrove species, *E. agallocha* does not contain specialized aerial roots extending above the soil surface in order to ensure the underground roots are supplied with oxygen (Mondal et al. 2016; Tomlinson 2016). This dioecious mangrove species has long male inflorescence and shorter female inflorescence (Duke 2006). Previous studies on *E. agallocha* essential focused on the hereditary patterns, medicinal properties and ecological adaptation of this plant to new environments (Chen and Ye 2014; Mondal et al. 2016; Liu et al. 2021). Hence, the several aspects behind the molecular evolution of this dioecious mangrove species remain unknown.

In flowering plants, sex determination represents a developmental process that increases fitness and survival by promoting allogamy. In the case of unisexual flowers, sex determination is achieved through selective abortion or arrest of reproductive organs (either male or female) (Hernández-Cruz et al. 2019). Recently, several studies have identified and characterized MADS-domain proteins by employing genome-wide expression analyses across woody plant species, including *Jatropha curcas* (Tang et al. 2020), *Quercus suber* (Rocheta et al. 2014), *Camellia sinensis* (Pan et al. 2018), *Malus × domestica* (Moser et al. 2020), *Populus trichocarpa* (Leseberg et al. 2006), and *Eriobotrya japonica* (Xu et al. 2021). Studies focusing on overexpression assays and mutant plants suggest MADS-box genes are essential for the specification in floral organs (Krizek and Fletcher 2005; Kater et al. 2006). Importantly, MADS-box transcription factors include a conserved DNA-binding motif (known as the MADS domain) at their N-terminus region (Bai et al. 2019). This domain is approximately 58–60 amino acids long and able to bind CArG boxes (CC[A/T]6GG) (Wei et al. 2015). An ancestral MADS-box gene duplication event separates two main lineages, type I and type II (Svensson et al. 2000). The former genes can be further subdivided into three distinct groups, specifically M α , M β , M γ , according to the genetic sequence of MADS domain and the presence of other motifs (Zhang et al. 2021). The proteins belonging to type II plants are named MIKC and, depending on the sequence divergence at domain I, can be divided into two different subgroups, MIKC^c and MIKC* (Lin et al. 2016). The number of type I genes is highly variable among flowering plants (Grimplet et al. 2016). While MIKC-type genes mostly evolved as a result of genome duplications, type I MADS-box genes mostly arose via segmental duplications. The latter play an essential role in the formation of the female gametophyte and the early development of seeds (Masiello et al. 2011).

MIKC^c-type have been identified as floral genes (Rounsley et al. 1995) that are also involved in flower

organogenesis. These genes can be grouped into five distinct classes: A, B, C, D, and E. Each of these classes plays different roles, in particular determining the identity of sepals (A + E), petals (A + B + E), stamens (B + C + E), carpels (C + E) and ovules (D + E). In *Arabidopsis*, these functional classes include genes such as *APETALA1* (*AP1*) in class A, *PISTILATA* (*PI*) and *APETALA3* (*AP3*) in class B, *AGAMOUS* (*AG*) in class C (Hsu 2002), *SEEDSTICK/AGAMOUS-LIKE 1* (*STK/AGL11*) and *SHATTERPROOF* (*SHP*) in class D (Mendes et al. 2013), and *SEPALATA* (*SEP1*, *SEP2*, *SEP3*, *SEP4*) in class E (Sun et al. 2014). Several MIKC^c genes in the *AG* and *APETALA1/FRUITFULL* (*AP1/FUL*) subfamilies further participate in the development of fruits and seeds (Meng et al. 2019), while others are seemingly involved in regulatory networks that control flowering time and initiation. Specifically, the genes *FLOWERING LOCUS C* (*FLC*), *SUPPRESSOR OF OVEREXPRESSION OF CONSTANTS 1* (*SOC1*) and *SHORT VEGETATIVE PHASE* (*SVP*) integrating signals from different flowering time regulatory pathways to regulate flowering transition (Yuan et al. 2009; Zhang et al. 2020a; Won et al. 2021). These genes can either be positive (*SOC1*, *AGL24*) or negative regulators (*FLC*, *SVP*) of genes regulating the identity of the flower meristem (Wang et al. 2015; Schilling et al. 2020).

We provide a reference transcriptome dataset for *E. agallocha* that can be used in future studies. To our knowledge, we performed an unprecedented and detailed transcriptomic analysis on MADS-box genes associated with floral development or sex determination in *E. agallocha*. This dataset expands the available genetic resources on *E. agallocha* and opens the door to future functional studies seeking to explore the mechanisms behind the molecular evolution of *E. agallocha*.

2 Material and methods

2.1 Plant materials

We sampled *E. agallocha* specimens from Zhanjiang Jinsha Bay, Guangdong, China (21.2771°N, 110.3899°E). Leaves and flowers were collected from one male and female trees separately at 9:00 am on July 20, 2020. The sampled materials were then used for transcriptomic analysis. In addition, we collected leaves and flowers from the same male and female trees, and fruits from the same female tree for real-time PCR analysis (Fig. 1). The plant materials were immersed in liquid nitrogen and stored at – 80 °C before downstream analyses were performed. Each samples had three biological replicates.

2.2 RNA extraction

We extracted total RNA from leaf and flower tissues using the TRIzol[®] Reagent (Plant RNA Purification Reagent) following manufacturer's instructions (Invitrogen,

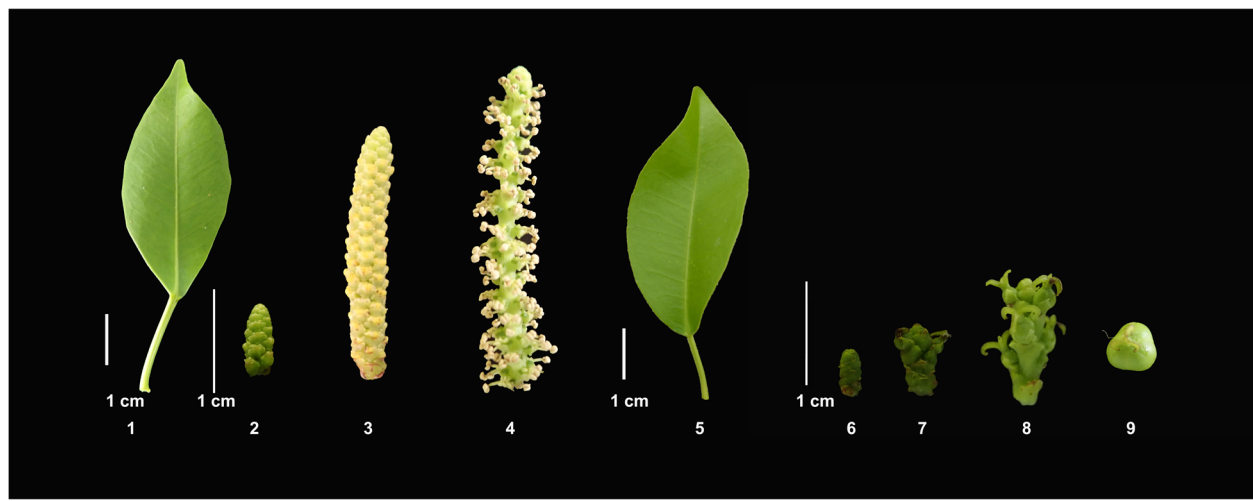


Fig. 1 Plant material used for qRT-PCR analysis. Note: **1** Male leaf. **2** Male flower bud (small). **3** Male flower bud (big), **4** Male flower. **5** Female leaf. **6** Female flower bud (small). **7** Female flower bud (big). **8** Female flower. **9** Fruit

Carlsbad, CA, USA). Genomic DNA was removed using DNase I (TaKaRa Bio, Inc., USA). The integrity and purity of the RNA were measured using the 2100 Bioanalyser (Agilent Technologies, Inc., Santa Clara CA, USA). RNA was quantified using the ND-2000 (NanoDrop Thermo Scientific, Wilmington, DE, USA). Only high-quality RNA samples ($OD_{260/280} = 1.8 \sim 2.2$, $OD_{260/230} \geq 2.0$, $RIN \geq 8.0$, $28S:18S \geq 1.0$, $> 1 \mu\text{g}$) were used to construct the sequencing library. Three biological replicates were performed for each sample.

2.3 Library preparation and Illumina NovaSeq 6000 sequencing

We performed the RNA purification, reverse transcription, library construction and sequencing experiments at the Shanghai Majorbio Bio-pharm Biotechnology Co., Ltd. (Shanghai, China). Each of these analyses followed manufacturer's instructions (Illumina, San Diego, CA). The RNA-seq transcriptome libraries (i.e., Ea_ff1, Ea_ff2, Ea_ff3, Ea_fl1, Ea_fl2, Ea_fl3, Ea_mf1, Ea_mf2, Ea_mf3, Ea_mL1, Ea_mL2, and Ea_mL3) were prepared using the Illumina sample preparation kit TruSeqTM RNA (Illumina, San Diego, CA). Three biological replicates were performed for each sample. We obtained purified poly(A) mRNA using oligo-dT-attached magnetic beads and a fragmentation buffer. We then used these short fragments as templates and synthesized double-stranded cDNA with SuperScript double-stranded cDNA synthesis kit (Invitrogen, CA) and random hexamer primers (Illumina). We performed end-repair on the synthesized cDNA, followed by phosphorylation and the addition of 'A' bases as per Illumina's library construction protocol. The libraries we selected for cDNA target fragments of 200–300 bp

using 2% Low Range Ultra Agarose and PCR amplification (15 cycles) with Phusion DNA polymerase (New England Biolabs, Boston, MA). After quantification by TBS380, we sequenced 12 RNAseq libraries on an Illumina NovaSeq 6000 sequencer (Illumina, San Diego, CA) using one lane and 2×150 bp paired-end reads. Correlation test was performed to verify the reliability of the repeated experiments. The raw reads generated in this study were deposited in the NCBI database under accession number PRJNA760657. See also Table S7 in dataset Zhou et al. 2022.

2.4 De novo assembly and annotation

The paired-end reads were trimmed and filtered for quality controls using SeqPrep (<https://github.com/jstjohn/SeqPrep>) and Sickle (<https://github.com/najos/sickle>) with default parameters. We used Trinity (<http://trinityrnaseq.sourceforge.net/>) to perform de novo assembly on the clean data (Ea_ff1, Ea_ff2, Ea_ff3, Ea_fl1, Ea_fl2, Ea_fl3, Ea_mf1, Ea_mf2, Ea_mf3, Ea_mL1, Ea_mL2, and Ea_mL3) (Grabherr et al. 2011). The assembled transcripts were functionally annotated against the NCBI protein nonredundant (NR), NCBI nucleotide sequence database (NT), SwissProt, Pfam, and Clusters of orthologous groups for eukaryotic complete genomes (KOG) databases using BLASTX with an E -value cut-off 1.0×10^{-5} . In addition, the BLAST2GO (<http://www.blast2go.com/b2gohome>) (Conesa et al. 2005) program was used to obtain GO annotations of uniquely assembled transcripts to describe the associated biological processes, molecular functions and cellular components. Metabolic pathway analysis was performed using KEGG (<http://www.genome.jp/kegg/>) (Ogata et al. 1999).

2.5 Differential expression analysis and functional enrichment

The clean data of each sample was mapped to the assembled reference transcriptome using *Jatropha curcas* L. (Seesangboon et al. 2018). The abundance of genes was calculated using RSEM (<http://deweylab.biostat.wisc.edu/rsem/>) (Li and Dewey 2011) and normalized using the transcripts per million reads (TPM) method (Table S8 dataset Zhou et al. 2022). Differential expression analysis was implemented using EdgeR (Robinson et al. 2010) that genes with $|\log_2FC| > 1$ and Q value ≤ 0.05 were considered to be significantly differentially expressed (DEGs). GO and KEGG enrichment analyses were performed on DEGs to identify relevant pathways and functions using Goatools (<https://github.com/tanghaibao/Goatools>) and KOBAS (<http://kobas.cbi.pku.edu.cn/home.do>), respectively (Xie et al. 2011).

2.6 Phylogenetic analysis

To identify the MADS-box from *E. agallocha* transcriptomes, we searched all significant MADS-box-related genes against a database composed of plant TFs (PlantTFDB 4.0; <http://planttfdb.gao-lab.org/>). We downloaded MADS protein sequences belonging to *Jatropha curcas* from PlantTFDB 4.0. Sequence alignments were performed using ClustalX (1.83). The phylogenetic trees comparing *E. agallocha* and *J. curcas* MADS-box proteins were built following a Maximum likelihood approach based on the similarity between full-length amino acid sequences using IQ-TREE (Nguyen et al. 2015).

2.7 Quantitative real-time PCR analysis

We selected a total of 14 DEGs (MADS-box gene family) detected in Ea_ff vs Ea_fl, Ea_ff vs Ea_mf, Ea_fl vs Ea_mL, and Ea_mf vs Ea_mL for quantitative real-time PCR (qRT-PCR) using male leaves, male flower buds (small), male flower buds (big), male flowers, female leaves, female flower buds (small), female flower buds (big), female flowers, and fruits. This was done in order to validate the transcriptome results. The genes were selected at random based on the changes in expression and function identified in this study (Table S1 dataset Zhou et al. 2022). A total of three biological and technical replicates were performed for each organ and gene, respectively (Fig. 1). We performed cDNA synthesis with the PrimeScript[®] II First Strand cDNA Synthesis Kit (TaKaRa, Japan). The primers were designed using the software Primer Premier 5.0. The amplified PCR product lengths varied between 80 and 300 bp (Table S2 dataset Zhou et al. 2022). qRT-PCR was performed in a CFX96Touch Real-Time PCR Detection system (Bio-Rad, America) with SYBR Premix Ex Taq[™] II (TaKaRa, Japan). We used the $2^{-\Delta\Delta C_t}$ method to estimate the relative expression

levels of the genes in different samples. The data analysis by SPSS version 13.0 (SPSS Inc., Chicago, IL, USA) was used for one-way analysis of variance (ANOVA) and post hoc multiple comparisons. We used male leaves as controls to calculate the relative expression of DEGs associated with MADS-box. Differences were considered to be significant when $P < 0.05$. The figures were done with GraphPad Prism 8.

3 Results

3.1 Transcriptome sequencing and de novo assembly of *E. agallocha*

We isolated total RNA from male leaf (mL), male flower (mf), female leaf (fL), and female flower (ff) to build the transcriptome of *E. agallocha*. After performing adapters trimming and filtering low-quality bases, we obtained 6.91–11.79 Gb of sequencing data from 12 cDNA libraries (Table S3 dataset Zhou et al. 2022). The estimated error rate of the RNA-seq was $\sim 0.02\%$. All Q30 values were higher than 90.45%, while the GC content in each sample was exceeded 42%. We obtained an assembly of 427,506 transcripts (mean sequence length 584 bp) for the *E. agallocha* transcriptome (Figure S1a dataset Zhou et al. 2022). We chose the longest transcript in each gene as the candidate unigene, and obtained a set composed by a total of 152,412 unigenes (Figure S1b dataset Zhou et al. 2022). Unigene length ranged from 201 to 16,447 bp (with a mean of 575 bp and $N_{50} = 738$ bp). The correlation of gene expression levels between samples indicates a high degree of similarity of expression patterns between samples (Figure S2 dataset Zhou et al. 2022). We identified total of 66,975 simple sequence repeats (SSRs) from the transcriptome of *E. agallocha* using the MISA software (Figure S3 and Table S4 dataset Zhou et al. 2022). These results suggest genetic high variation in *E. agallocha*.

3.2 Gene functional annotation

After assembling, we blasted the unigenes against seven public databases, including NR, NT, SwissProt, KO, Pfam, GO, and KOG. This allowed us to identify gene functions and associated pathways (Figure S4 dataset Zhou et al. 2022; Table 1). The highest annotation was found with NR (102,678; 41.40%), followed by Pfam (66,309; 26.73%) and KOG (24,375; 9.82%). The *E*-value distribution of the top-scoring BLASTX hits against the NR database showed that 25.5% of the mapped sequences exhibited high homology ($< 1e-45$). In contrast, the *E*-values for 73.80% of the homologous sequences ranged between $1e-5$ and $1e-45$ (Figure S3a dataset Zhou et al. 2022). In addition, 41.10% of the mapped sequences has a similarity higher than 80%, while 16.20% showed a similarity lower than 60%

Table 1 Summary of the annotation about the *E. agallocha* transcriptome

| Type | Number of unigenes | Percentage (%) |
|----------------------------|--------------------|----------------|
| Annotated in NR | 102678 | 41.40 |
| Annotated in NT | 69953 | 28.20 |
| Annotated in KO | 34874 | 14.06 |
| Annotated in SwissProt | 71538 | 28.84 |
| Annotated in Pfam | 66309 | 26.73 |
| Annotated in GO | 71565 | 28.85 |
| Annotated in KOG | 24375 | 9.82 |
| Annotated in all databases | 8954 | 3.61 |

(Figure S4b). Furthermore, the NR BLASTX matches showed that the top two species were *J. curcas* (26.90%) and *R. communis* (21.60%) (Figure S4d dataset Zhou et al. 2022). The above results highlight the highest homology obtained in our study.

We further annotated 71,565 (28.85%) unigenes using 55 Gene Ontology (GO) terms. This included 25, 20, 10 and GO terms associated with biological processes, cellular components, and molecular functions, respectively (Figure S5 dataset Zhou et al. 2022). The main GO terms found in each of three categories cellular process (GO: 0009987) and metabolic process (GO: 0008152); cell (GO: 0005623) and cell part (GO: 0044464); and binding (GO: 0005488) and catalytic activity (GO: 0003824), respectively. To further characterize the

biological functions and interaction between unigenes, these groups were then classified into different metabolic pathways with KEGG (Table S5 dataset Zhou et al. 2022). We assigned a total of 34,874 unigenes (14.06%) to 19 categories divided into five clusters, including cellular processes, environmental information processing, genetic information processing, metabolism and organismal systems (Figure S6 dataset Zhou et al. 2022). The top 5 KEGG pathways were translation (3719 unigenes), carbohydrate metabolism (3475 unigenes), folding and degradation (2749 unigenes), overview (2622 unigenes), and amino acid metabolism (2199 unigenes) (Table S5 dataset Zhou et al. 2022). These observations demonstrate these samples were characterized by active cell development and differentiation.

3.3 Sexuality-specific genes between female and male of *E. agallocha*

We identified specific DEGs in female and male *E. agallocha* plants based on the sexuality-specific unigenes present in leaf and flower samples (p value < 0.05 and $|\log_2(\text{fold change})| > 5$). The results were presented in Fig. 2. We found a total of 4440 and 12,812 DEGs upregulated in the Ea_ff vs Ea_fl and Ea_mf vs Ea_ml group, respectively; while 4471 and 9655 DEGs were respectively down-regulated in the same groups. We uncovered 7923 DEGs between female and male flowers, of which 3370 were upregulated and 4553 were downregulated in the Ea_ff vs Ea_mf group. A further 9396 DEGs existed

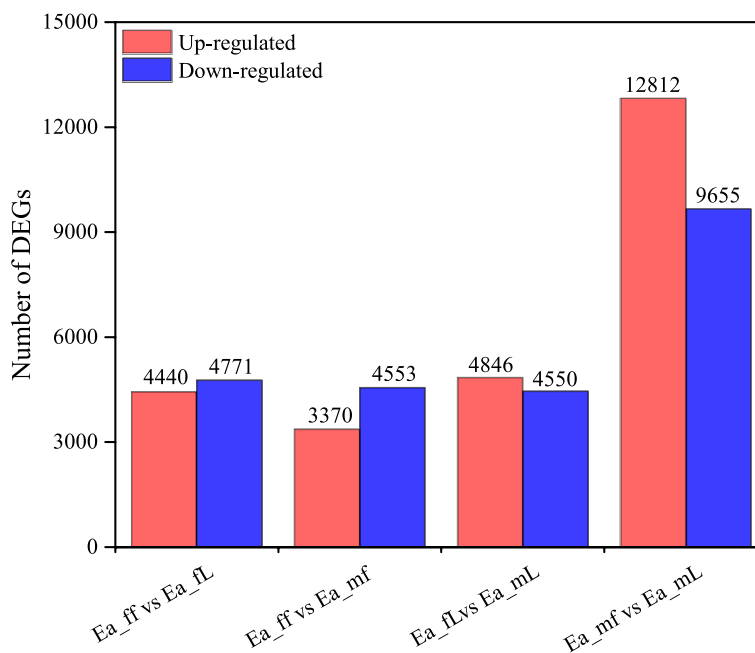


Fig. 2 Differentially expression gene (DEGs) analysis in *E. agallocha*

between female and male leaves, of which 4846 were upregulated and 4550 were downregulated in the Ea_fl vs Ea_mL group.

We functionally annotated the *E. agallocha* female and male transcriptomes by performing a GO analysis on DEGs. Our results (Fig. 3) showed that under the biological process, cellular component and molecular function categories, the most significantly enriched terms were

associated with the regulation of nitrogen compound metabolic process (GO:0051171), transcription factor complex (GO:0005667) and nuclear/ DNA-directed RNA polymerase complex (GO:0055029/GO:0000428) in flowers (Ea_ff vs Ea_mf group); metabolic process (GO: 0008152), transcription factor complex (GO:0005667) and catalytic activity (GO: 0003824) in females (Ea_ff vs Ea_fl group); metabolic process (GO: 0008152), membrane

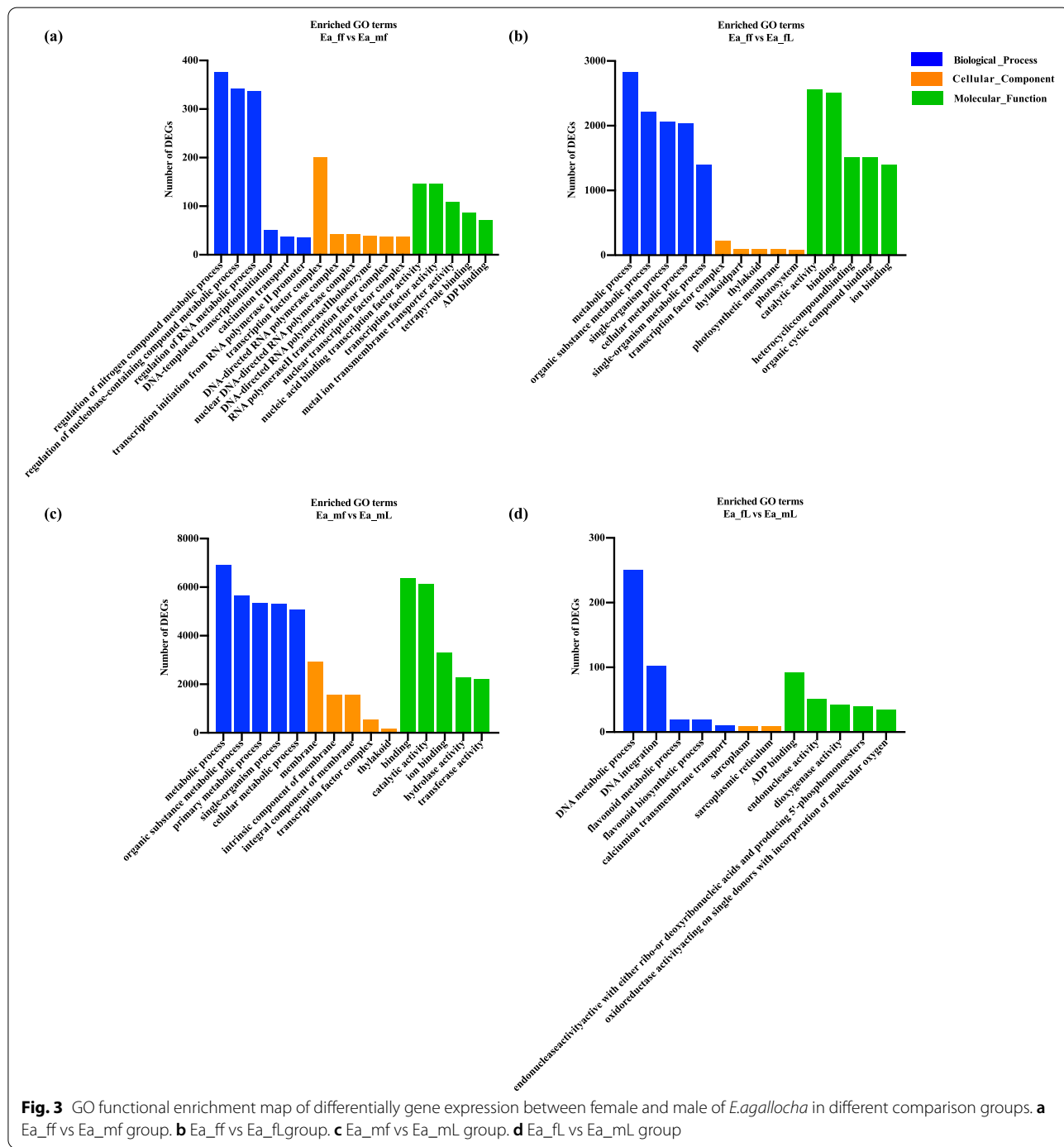


Fig. 3 GO functional enrichment map of differentially gene expression between female and male of *E. agallocha* in different comparison groups. **a** Ea_ff vs Ea_mf group. **b** Ea_ff vs Ea_fl group. **c** Ea_mf vs Ea_mL group. **d** Ea_fl vs Ea_mL group

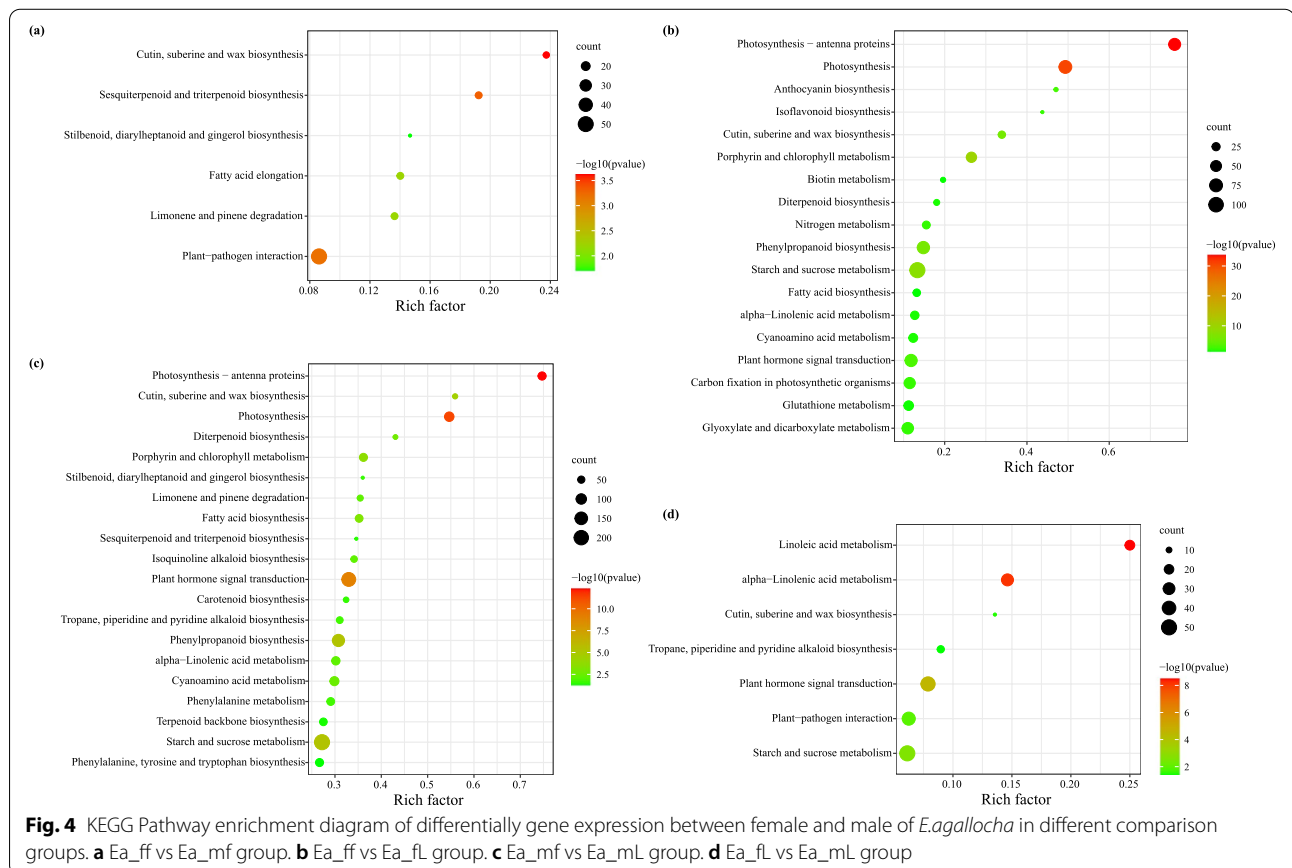
(GO:0016020) and catalytic activity (GO: 0003824) in males (Ea_mf vs Ea_mL group); and DNA metabolic process (GO:0006259), sarcoplasm and sarcoplasmic reticulum (GO:0016528 and GO:0016529) and ADP binding (GO:0043531) in leaves (Ea_fl vs Ea_mL group).

To identify which unigenes were significantly associated with either metabolic or signal transduction pathways, we used the DEGs to query the KEGG database and compared the results to the whole-transcriptome data. This pathway enrichment analysis showed that those unigenes were associated with 116, 115, 124, and 119 pathways in the Ea_ff vs Ea_mf, Ea_ff vs Ea_fl, Ea_mf vs Ea_mL, and Ea_flvs Ea_mL groups, respectively (Table S6 dataset Zhou et al. 2022). The estimated top-3 significant pathways were “plant-pathogen interaction” (ko04626), “fatty acid elongation” (ko00062) and “sesquiterpenoid and triterpenoid biosynthesis” (ko00909) in flowers (Ea_ff vs Ea_mf group) (Fig. 4a); “starch and sucrose metabolism” (ko00500), “photosynthesis” (ko00195) and “phenylpropanoid biosynthesis” (ko00940) in females (Ea_ff vs Ea_fl group) (Fig. 4b); “starch and sucrose metabolism” (ko00500), “plant hormone signal transduction” (ko04075) and “phenylpropanoid biosynthesis” (ko00940) in males (Ea_mf vs Ea_mL group) (Fig. 4c); and “starch and sucrose metabolism” (ko00500),

“plant hormone signal transduction” (ko04075) and “plant-pathogen interaction” (ko04626) in leaves (Ea_flvs Ea_mL group) (Fig. 4d).

3.4 MADS-box transcription factors in *E. agallocha*

Using the method described above, we obtained 73 unique MADS-box genes from *E. agallocha* and divided these genes into several branches (Fig. 5). At the same time, we compared the results of MADS-box protein domains in *E. agallocha* and *J. curcas* and performed phylogenetic analysis. EaMADS-box classification was based on a previous study (Tang et al. 2020). Specifically, we identified a total of 9, 7, 4, and 3 genes with M α , M β , M γ , and M δ MIKC*-type genes. The remaining 50 MADS-box proteins belonged to the MIKC^C subgroup (Fig. 5). Notably, we unveiled 7 MIKC^C clades in *E. agallocha* belonging to the MIKC^C subgroup class in *J. curcas*. These 7 MIKC^C clades included B, SVP and SOC1 clades with 4 EaMADS-box sequences; the A clade included 3 EaMADS-box sequences; the E and C/D clades included 2 EaMADS-box sequences; and the B clade contained a single EaMADS-box sequence. We identified 17, 38, 61, and 11 MADS-box related DEGs in the Ea_ff vs Ea_mf, Ea_ff vs Ea_fl, Ea_mf vs Ea_mL, and Ea_fl vs Ea_mL groups, respectively

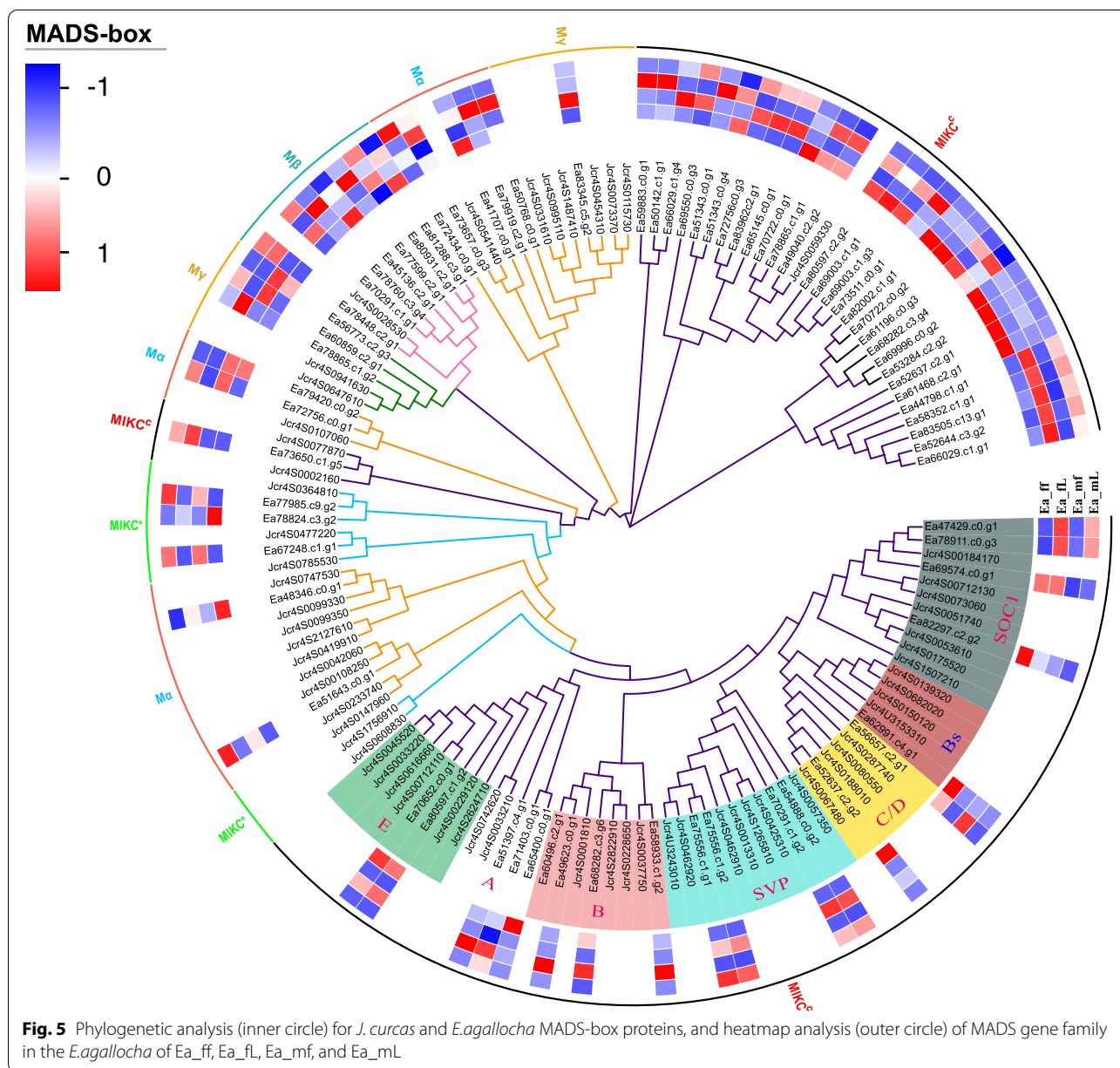


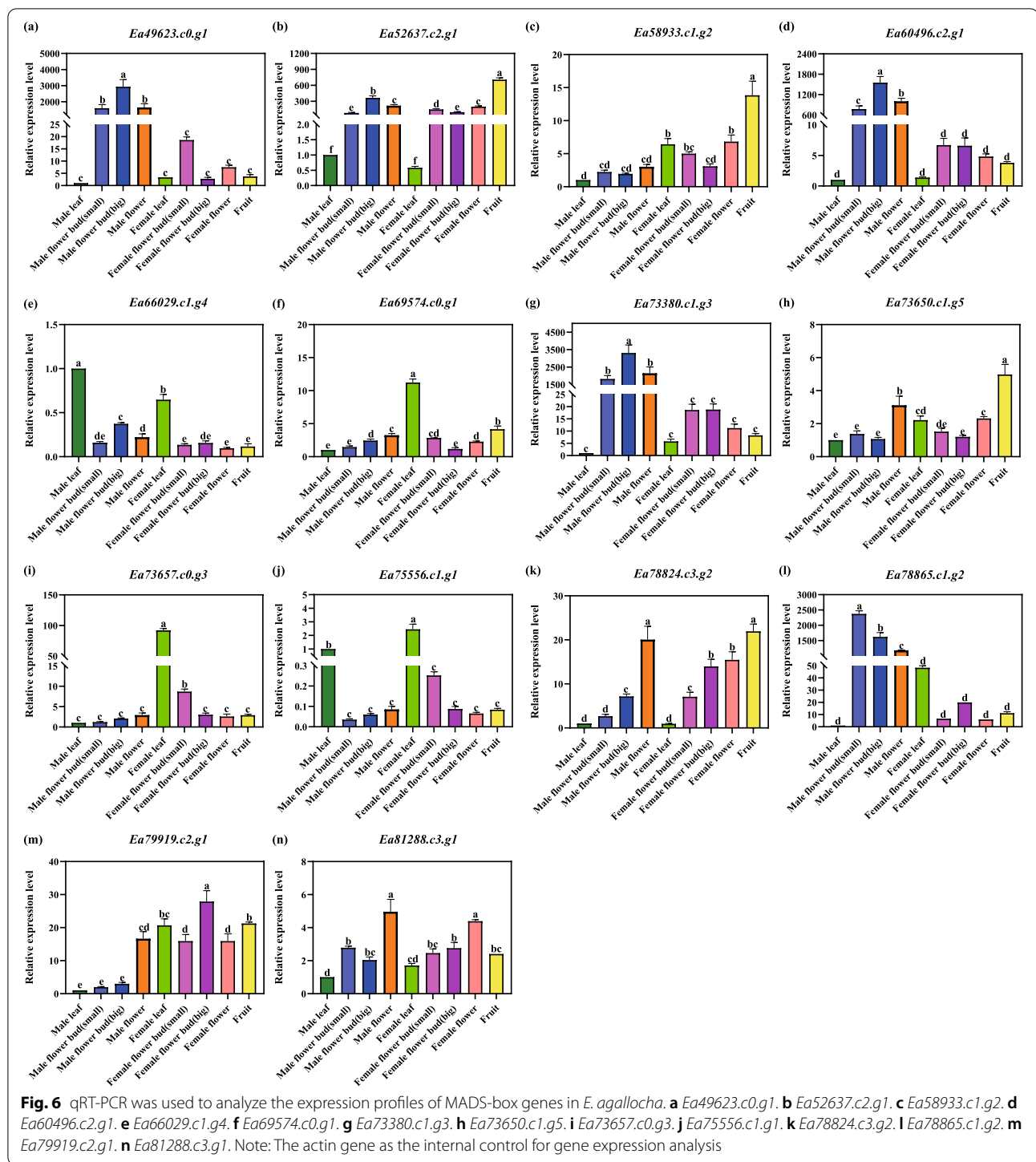
(Fig. 6; Table S1 dataset Zhou et al. 2022). In flowers (Ea_ff vs Ea_mf group), we found a total of 3 DEGs with $M\alpha$, 1 with $M\gamma$, 1 with MIKC*, and 11 with the MIKC^C subgroup. In females (Ea_ff vs Ea_fl group), we identified 3 genes with $M\alpha$, 2 with $M\beta$, 1 with $M\gamma$, 2 with MIKC*, and 33 with the MIKC^C subgroup. In males (Ea_mf vs Ea_mL group), we uncovered 6 genes with $M\alpha$, 6 with $M\beta$, 2 with $M\gamma$, 3 with MIKC*, and 39 with the MIKC^C subgroup. Finally, we found 3 genes with $M\alpha$, 1 with $M\beta$, 1 with $M\gamma$, 6 with the MIKC^C subgroup in leaves (Ea_fl vs Ea_mL group). Consequently, the phylogenetic and DEGs analyses showed that different kinds of female and male *E. agallocha* organs contained MADS proteins in each putative functional group.

However, the functional relevance of these predicted genes needs to be further determined.

3.5 Sexuality-specific expression patterns of MADS-box genes in *E. agallocha*

Because no MADS-box TFs had been previously identified in *E. agallocha*, we chose 14 EaMADS-box genes belonging to each of the MADS-box subgroups and evaluated the expression patterns of these genes (Fig. 5). Our results showed differential gene expression in 14 MADS-box genes in *E. agallocha* (Fig. 6; Figure S7 dataset Zhou et al. 2022). The pattern of RNA-seq and qRT-PCR were almost the same, specifically when comparing leaves to flowers (Figure S7 dataset Zhou et al. 2022). This indicated that





4 EaMADS-box genes were highly expressed in the male flower and flower bud tissues. These include *Ea49623.c0.g1*, *Ea60496.c2.g1*, *Ea73380.c1.g3* and *Ea78865.c1.g2*. Importantly, *Ea69574.c0.g1*, *Ea73657.c0.g3*, and *Ea75556.c1.g1* were also highly expressed in the female leaves, *Ea52637.c2.g1*, *Ea58933.c1.g2*, and *Ea73650.c1.g5* in fruits,

Ea78824.c3.g2 in male flowers, and *Ea66029.c1.g4* in male leaves. The floral expression profiles of EaMADS-box genes provide relevant information for investigating flower development in *E. agallocha*. Taken together, these results suggest the transcriptome data agreed with the observed expression patterns of most *E. agallocha* genes.

4 Discussion

This study provides the first de novo transcriptome assembly for the flower development of a dioecious mangrove. The uncovered unigenes were functionally annotated using seven public databases. The species distribution of the NR BLASTx matches revealed *E. agallocha* is closely related to *J. curcas* species. In fact, both *J. curcas* and *E. agallocha* belong to Euphorbiaceae (Satyan et al. 2009; Abdelgadir and Van Staden 2013).

We identified a large number of sex-specific unigenes significantly associated with development and other biological processes. GO analysis revealed a large number of genes associated with various transcription factor complexes were enriched in females, “membrane” was significantly abundant in males, which affecting sex determination of *E. agallocha*. KEGG pathway analysis showed much of the identified DEGs are involved in starch and sucrose metabolism in the floral and leaf tissues of *E. agallocha*. We found a higher number of DEGs involved in the ribosome and plant hormone signal transduction pathway affecting male flowers. Plant hormones can regulate the transition from vegetative development to the first inflorescence phase of reproductive development (Xing et al. 2015). In addition, phytohormones play an essential role in sex determination, and the constitutive expression of MADS-box genes leads to precocious flower differentiation in *A. thaliana* (Tanurdzic 2004; Amasino 2010). Accordingly, the present study screened additional DEGs associated with sex determination between female and male flowers in *E. agallocha*. This approach identified 73 DEGs belonging to the MADS-box gene family, and demonstrate this family plays a key regulator of floral development in *E. agallocha*. The nine harvested samples allowed further identification of the expression profiles during flower and fruit development. A total of 14 candidate genes were chosen for qRT-PCR analysis, which showed they have distinct expression profiles in different organs.

We analyzed the phylogeny of MADS-box protein families by searching for orthologous genes involved in flower development in *E. agallocha*. This allowed us to validate the identity of the different subfamilies of these genes. We performed the phylogenetic analysis on MADS-box protein sequences from *J. curcas*. The MADS-box proteins regulate floral organogenesis and are essential for flower development and sex determination (De Bodt et al. 2003; Pan et al. 2018; Bar-Lev et al. 2021). These proteins can be subdivided into two lineages, the M-type (type I) and MIKC (type II). The former can be further grouped into M α , M β , and My subgroups based on the respective sequence similarities between their MADS-box regions (Pan et al. 2018). We annotated 73 transcripts in the *E. agallocha* transcriptome that

aligned to the MADS-box family and/or contained the MADS domain. Our phylogenetic analysis identified several orthologous genes between *E. agallocha* and *J. curcas* in almost all MADS-box protein clades (Fig. 5; Table S1 dataset Zhou et al. 2022). Similar to other plants, *E. agallocha* also contains a higher number of M α and MIKC^c than My or MIKC* genes (Fig. 5; Table S1 dataset Zhou et al. 2022). The distribution of MADS-box subfamily genes in *E. agallocha* is similar to those observed genes in most plants, indicating this gene family is evolutionarily conserved (Tian et al. 2015; Wei et al. 2015; Ma et al. 2017). Several studies showed the MIKC^c gene family plays a central role in flower development and the regulation of flowering time (Yuan et al. 2009; Wang et al. 2015; Zhang et al. 2020b). Interestingly, we found 50 MIKC^c in *E. agallocha* (Fig. 5; Table S1 dataset Zhou et al. 2022). This might indicate that there may be a complex regulatory mechanism for flowering in *E. agallocha*. Moreover, a reduction in the number of subfamilies, including M β , My, and MIKC*, indicates some of these families may gradually disappear during evolution.

In MIKC^c group, we found class A, B, C, D, and E genes in *E. agallocha*. Of these, those that are significantly upregulated in male or female flowers did not show significant differences in expression between flowers and leaves. The A-class genes, *AP1/FUL* (*Ea51397.c4.g1*, *Ea65400.c0.g1*, and *Ea71403.c0.g1*) were upregulated in the male flower. In addition, the B-class genes *AP3/PI* (*Ea49623.c0.g1* and *Ea60496.c2.g1*) were also upregulated in the male flower and may influence male organ development. Similar results have been proposed in *Quercus suber* (Rocheta et al. 2014). As reported in the Norway spruce, B-type MADS-box genes are active in male organ primordia (Sundstrom and Engstrom 2002). C and D-class genes were significantly upregulated in male and female flowers in *E. agallocha*. *STK* (*Ea56657.c2.g1*) was only expressed in floral organs and is likely involved in flower organ identity (Soltis 2002). *AGL104* (*Ea52637.c2.g2*) was upregulated in the female flowers compared to male flowers. In *Arabidopsis*, *AGL104* is involved in pollen maturation and tube growth (Adamczyk and Fernandez 2009). The E-class genes, *SEP2/AGL4* (*Ea80597.c1.g2* and *Ea70652.c0.g1*) were significantly up-regulated in female and male flowers. *SEP2/AGL4* is required for the specification of floral organs (Wang et al. 2016). However, these genes did not show significant differences in expression between female and male flowers. Accordingly, we hypothesize E-class genes might not be necessary for sex determination. In apples, SUPPRESSOR OF CONSTANS 1 (SOC1) regulates the induction of the flower bud (Xing et al. 2015). *SOC1* (*Ea47429.c0.g1*, *Ea78911.c0.g3*, *Ea69574.c0.g1* and *Ea82297.c2.g2*) was significantly down-regulated in floral

organs compared to leaves, but significantly expressed in flower buds. The *SVP* gene represses floral transition in *B. juncea* (Li et al. 2019), and is highly expressed in both male and female leaves in *E. agallocha*. It would be very interesting to evaluate whether this gene may be similarly inhibited in *E. agallocha*, restricting leaf development. We propose *AP1/FUL*, *AP3/PI*, *AGL104*, and *SOC1* should be considered candidate regulators of sex determination in *E. agallocha*.

5 Conclusion

We report the transcriptome of *E. agallocha*. The data provides important information for understanding the evolutionary history of mangrove. The high number of transcript sequences can be used to discover novel genes, specifically those associated with flower development in *E. agallocha*. A total of 7923 DEGs were identified between female and male flowers, 73 of which were MADS-box genes (50 belonged to the MIKC^c subgroup). Our results show that *AP1/FUL*, *AP3/PI*, *AGL104*, and *SOC1* genes play an important role during the formation of floral organs in *E. agallocha*.

Code availability

Not applicable

Authors' contributions

Conceptualization: Ying Zhang, Yan Zhou; methodology: Lulu Hao, Lexiang Huang; formal analysis and investigation: Xiaoming Tang, DanTing Zhou; writing—original draft preparation: Yan Zhou; writing—review and editing: Yan Zhou, Liyun Wang; funding acquisition: Ying Zhang, Liyun Wang; resources: Ying Zhang; supervision: Liyun Wang. All authors read and approved the final manuscript.

Funding

This study was funded by the National Natural Science Foundation of China (No. 41776148 and No. 31360173), Basic and Applied Basic Research Fund of Guangdong Province (No. 2019A1515110138), and School-level Talents Project of Lingnan Normal University (No. ZL2003 and No. ZL2032).

Availability of data and materials

The datasets generated during the current study are available in the NCBI database: <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA760657>. Supplementary dataset is also available at: <https://doi.org/10.6084/m9.figshare.20362563.v5>

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

All authors gave their informed consent to this publication and its content.

Competing interests

The authors declare that they have no competing interests.

Author details

¹School of Life Sciences and Technology, Lingnan Normal University, Zhanjiang, China. ²National and Local Joint Engineering Research Center of Ecological Treatment Technology for Urban Water Pollution, College of Life and Environmental Science, Wenzhou University, Wenzhou, China.

Received: 7 March 2022 Accepted: 26 July 2022

Published online: 24 August 2022

References

- Abdelgadir HA, Van Staden J (2013) Ethnobotany, ethnopharmacology and toxicity of *Jatropha curcas* L. (Euphorbiaceae): a review. *South Afr J Bot* 88:204–218. <https://doi.org/10.1016/j.sajb.2013.07.021>
- Adamczyk BJ, Fernandez DE (2009) MIKC* MADS domain heterodimers are required for pollen maturation and tube growth in *Arabidopsis*. *Plant Physiol* 149:1713–1723. <https://doi.org/10.1104/pp.109.135806>
- Amasino R (2010) Seasonal and developmental timing of flowering. *Plant J* 61:1001–1013. <https://doi.org/10.1111/j.1365-3113.2010.04148.x>
- Bai G, Yang DH, Cao PJ et al (2019) Genome-wide identification, gene structure and expression analysis of the MADS-box gene family indicate their function in the development of tobacco (*Nicotiana tabacum* L.). *Int J Mol Sci* 20:5043. <https://doi.org/10.3390/ijms20205043>
- Bar-Lev Y, Senden E, Pasmanik-Chor M, Sapir Y (2021) De novo transcriptome characterization of *Iris atropurpurea* (the Royal Iris) and phylogenetic analysis of MADS-box and *R2R3-MYB* gene families. *Sci Rep* 11:16246. <https://doi.org/10.1038/s41598-021-95085-5>
- Chen Y, Ye Y (2014) Effects of salinity and nutrient addition on mangrove *Excoecaria agallocha*. *PLoS One* 9:e93337. <https://doi.org/10.1371/journal.pone.0093337>
- Conesa A, Gotz S, Garcia-Gomez JM et al (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674–3676. <https://doi.org/10.1093/bioinformatics/bti610>
- De Bodt S, Raes J, Florquin K et al (2003) Genomewide structural annotation and evolutionary analysis of the type I MADS-box genes in plants. *J Mol Evol* 56:573–586. <https://doi.org/10.1007/s00239-002-2426-x>
- Duke NC (2006) Australia's mangroves: the authoritative guide to Australia's mangrove plants. University of Queensland, St. Lucia
- Gabherr MG, Haas BJ, Yassour M et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29:644–652. <https://doi.org/10.1038/nbt.1883>
- Grimplet J, Martínez-Zapater JM, Carmona MJ (2016) Structural and functional annotation of the MADS-box transcription factor family in grapevine. *BMC Genomics* 17:80. <https://doi.org/10.1186/s12864-016-2398-7>
- Hernández-Cruz R, Silva-Martínez J, García-Campusano F, et al (2019) Comparative development of staminate and pistillate flowers in the dioecious cactus *Opuntia robusta*. *Plant Reprod* 32:257–273. <https://doi.org/10.1007/s00497-019-00365-w>
- Hsu H-F (2002) An orchid (*Oncidium Gower Ramsey*) AP3-like MADS gene regulates floral formation and initiation. *Plant Cell Physiol* 43:1198–1209. <https://doi.org/10.1093/pcp/pcf143>
- Kater MM, Dreni L, Colombo L (2006) Functional conservation of MADS-box factors controlling floral organ identity in rice and *Arabidopsis*. *J Exp Bot* 57:3433–3444. <https://doi.org/10.1093/jxb/erl097>
- Krizek BA, Fletcher JC (2005) Molecular mechanisms of flower development: an armchair guide. *Nat Rev Genet* 6:688–698. <https://doi.org/10.1038/nrg1675>
- Leseberg CH, Li A, Kang H et al (2006) Genome-wide analysis of the MADS-box gene family in *Populus trichocarpa*. *Gene* 378:84–94. <https://doi.org/10.1016/j.gene.2006.05.022>
- Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323. <https://doi.org/10.1186/1471-2105-12-323>
- Li S, Li Z, Zhang J et al (2019) Flowering signal integrator AGL24 interacts with K domain of AGL18 in *Brassica juncea*. *Biochem Biophys Res Commun* 518:148–153. <https://doi.org/10.1016/j.bbrc.2019.08.023>
- Lin C-S, Hsu C-T, Liao D-C et al (2016) Transcriptome-wide analysis of the MADS-box gene family in the orchid *Erycina pusilla*. *Plant Biotechnol J* 14:284–298. <https://doi.org/10.1111/pbi.12383>
- Liu G, Zhang Z, Wang Y, Li X (2021) Highly oxygenated ent-atisane and podocarpane diterpenoids from *Excoecaria agallocha*. *Nat Prod Res* 1–7. <https://doi.org/10.1080/14786419.2021.1900177>
- Ma J, Yang Y, Luo W et al (2017) Genome-wide identification and analysis of the MADS-box gene family in bread wheat (*Triticum aestivum* L.). *PLoS One* 12:e0181443. <https://doi.org/10.1371/journal.pone.0181443>

- Masiero S, Colombo L, Grini PE et al (2011) The emerging importance of type I MADS box transcription factors for plant reproduction. *Plant Cell* 23:865–872. <https://doi.org/10.1105/tpc.110.081737>
- Mendes MA, Guerra RF, Berns MC et al (2013) MADS domain transcription factors mediate short-range DNA looping that is essential for target gene expression in *Arabidopsis*. *Plant Cell* 25:2560–2572. <https://doi.org/10.1105/tpc.112.108688>
- Meng D, Cao Y, Chen T et al (2019) Evolution and functional divergence of MADS-box genes in *Pyrus*. *Sci Rep* 9:1266. <https://doi.org/10.1038/s41598-018-37897-6>
- Mondal S, Ghosh D, Ramakrishna K (2016) A complete profile on blind-your-eye mangrove *Excoecaria agallocha* L. (*Euphorbiaceae*): Ethnobotany, phytochemistry, and pharmacological aspects. *Pharmacogn Rev* 10(123). <https://doi.org/10.4103/0973-7847.194049>
- Moser M, Asquini E, Miolli GV et al (2020) The MADS-box gene *MdDAM1* controls growth cessation and bud dormancy in apple. *Front Plant Sci* 11:1003. <https://doi.org/10.3389/fpls.2020.01003>
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274. <https://doi.org/10.1093/molbev/msu300>
- Ogata H, Goto S, Sato K et al (1999) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 27:29–34. <https://doi.org/10.1093/nar/27.1.29>
- Pan J, Chang P, Ye X et al (2018) Transcriptome-wide analysis of MADS-box family genes involved in aluminum and fluoride assimilation in *Camellia sinensis*. *Plant Biotechnol* 35:313–324. <https://doi.org/10.5511/plantbiotechnology.18.0621a>
- Rocheta M, Sobral R et al (2014) Comparative transcriptomic analysis of male and female flowers of monoecious *Quercus suber*. *Front Plant Sci* 5. <https://doi.org/10.3389/fpls.2014.00599>
- Robinson MD, McCarthy DJ, Smyth GK (2010) EdgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140. <https://doi.org/10.1093/bioinformatics/btp616>
- Rounsley SD, Ditta GS, Yanofsky MF (1995) Diverse roles for MADS box genes in *Arabidopsis* development. *Plant Cell* 7:1259–1269. <https://doi.org/10.1105/tpc.7.8.1259>
- Satyan R, Aveek N, Eganathan P, Parida A (2009) Comparative histochemical localization of secondary metabolites in seed-raised and in vitro propagated plants of *Excoecaria agallocha* Linn. (*Euphorbiaceae*), the milky mangrove tree of historical significance. *Biotech Histochem*:1–9. <https://doi.org/10.1080/1052090903202478>
- Schilling S, Kennedy A, Pan S et al (2020) Genome-wide analysis of MIKC-type MADS-box genes in wheat: pervasive duplications, functional conservation and putative neofunctionalization. *New Phytol* 225:511–529. <https://doi.org/10.1111/nph.16122>
- Seesangboon A, Grunec L, Pokawattana T et al (2018) Transcriptome analysis of *Jatropha curcas* L. flower buds responded to the paclobutrazol treatment. *Plant Physiol Biochem* 127:276–286. <https://doi.org/10.1016/j.plaphy.2018.03.035>
- Soltis D (2002) Missing links: the genetic architecture of flower and floral diversification. *Trends Plant Sci* 7:22–31. [https://doi.org/10.1016/S1360-1385\(01\)02098-2](https://doi.org/10.1016/S1360-1385(01)02098-2)
- Sun W, Huang W, Li Z et al (2014) Functional and evolutionary analysis of the *AP1/SEP/AGL6* superclade of MADS-box genes in the basal eudicot *Epimedium sagittatum*. *Ann Bot* 113:653–668. <https://doi.org/10.1093/aob/mct301>
- Sundstrom J, Engstrom P (2002) Conifer reproductive development involves B-type MADS-box genes with distinct and different activities in male organ primordia. *Plant J* 31:161–169. <https://doi.org/10.1046/j.1365-313X.2002.01343.x>
- Svensson ME, Johannesson H, Engström P (2000) The *LAMB1* gene from the clubmoss, *Lycopodium annotinum*, is a divergent MADS-box gene, expressed specifically in sporogenic structures. *Gene* 253:31–43. [https://doi.org/10.1016/S0378-1119\(00\)00243-2](https://doi.org/10.1016/S0378-1119(00)00243-2)
- Tang Y, Wang J, Bao X et al (2020) Genome-wide analysis of *Jatropha curcas* MADS-box gene family and functional characterization of the *JcMADS40* gene in transgenic rice. *BMC Genomics* 21:325. <https://doi.org/10.1186/s12864-020-6741-7>
- Tanurdzic M (2004) Sex-determining mechanisms in land plants. *Plant Cell Online* 16:61–71. <https://doi.org/10.1105/tpc.016667>
- Tian Y, Dong Q, Ji Z et al (2015) Genome-wide identification and analysis of the MADS-box gene family in apple. *Gene* 555:277–290. <https://doi.org/10.1016/j.gene.2014.11.018>
- Tomlinson PB (2016) The botany of mangroves, second edition. Cambridge University Press, Cambridge
- Wang D, Chen X, Zhang Z et al (2015) A MADS-box gene *NtSVP* regulates pedicel elongation by directly suppressing a KNAT1-like KNOX gene *NtBPL* in tobacco (*Nicotiana tabacum* L.). *J Exp Bot* 66:6233–6244. <https://doi.org/10.1093/jxb/erv332>
- Wang X, Gao D, Sun J et al (2016) An exon skipping in a SEPALLATA-like gene is associated with perturbed floral and fruits development in cucumber: *CsSEP2* and cucumber development. *J Integr Plant Biol* 58:766–771. <https://doi.org/10.1111/jipb.12472>
- Webster GL (1994) Classification of the Euphorbiaceae. *Ann Mo Bot Gard* 81:3. <https://doi.org/10.2307/2399908>
- Wei X, Wang L, Yu J et al (2015) Genome-wide identification and analysis of the MADS-box gene family in sesame. *Gene* 569:66–76. <https://doi.org/10.1016/j.gene.2015.05.018>
- Won SY, Jung J-A, Kim JS (2021) Genome-wide analysis of the MADS-box gene family in *chrysanthemum*. *Comput Biol Chem* 90:107424. <https://doi.org/10.1016/j.compbiolchem.2020.107424>
- Xie C, Mao X, Huang J et al (2011) KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res* 39:316–322. <https://doi.org/10.1093/nar/gkr483>
- Xing L-B, Zhang D, Li Y-M et al (2015) Transcription profiles reveal sugar and hormone signaling pathways mediating flower induction in apple (*Malus domestica* Borkh.). *Plant Cell Physiol* 56:2052–2068. <https://doi.org/10.1093/pcpp/pcv124>
- Xu H, Chen T, Qi M et al (2021) Sorbitol in loquat promotes early flower bud differentiation via the MADS-box transcription factor EJCAL. *bioRxiv* 2021.08.17.456699. <https://doi.org/10.1101/2021.08.17.456699>
- Yuan Y-X, Wu J, Sun R-F et al (2009) A naturally occurring splicing site mutation in the *Brassica rapa* *FLC1* gene is associated with variation in flowering time. *J Exp Bot* 60:1299–1308. <https://doi.org/10.1093/jxb/erp010>
- Zhang X, Fatima M, Zhou P et al (2020a) Analysis of MADS-box genes revealed modified flowering gene network and diurnal expression in pineapple. *BMC Genomics* 21:8. <https://doi.org/10.1186/s12864-019-6421-7>
- Zhou Y, Hao L, Huang L, Tang X, Zhuo D, Wang LY, Zhang Y (2022) De novo transcriptome assembly reveals characteristics of flower sex determination of *Excoecaria agallocha*. [dataset]. Figshare repository V5. <https://doi.org/10.6084/m9.figshare.20362563.v5>
- Zhang Y, Chen Y, Zhou Y et al (2020b) Comparative transcriptome reveals the genes' adaption to herkogamy of *Lumnitzera littorea* (Jack) Voigt. *Front Genet* 11:584817. <https://doi.org/10.3389/fgene.2020.584817>
- Zhang Z-B, Jin Y-J, Wan H-H et al (2021) Genome-wide identification and expression analysis of the MADS-box transcription factor family in *Camellia sinensis*. *J Appl Genet* 62:249–264. <https://doi.org/10.1007/s13353-021-00621-8>

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

