BMC Plant Biology

RESEARCH

Open Access

The dynamic regulatory network of stamens and pistils in papaya

Tao Xiang[1,2], Yating Zhu[1,2], Yang Wang[1,2], Xi Chen[2], Zhibin Zhang[2], Juan Lai[2], Ping Zhou[3], Ray Ming[2] and Jingjing Yue[2*]

Abstract

Background Papaya exhibits three sex types: female (XX), male (XY), and hermaphrodite (XY^h), making it an unusual trioecious model for studying sex determination. A critical aspect of papaya sex determination is the pistil abortion in male flowers. However, the regulatory networks that control the development of pistils and stamens in papaya remain incompletely understood.

Results In this study, we identified three organ-specific clusters involved in papaya pistils and stamens development. We found that pistil development is primarily characterized by the significant expression of auxin-related genes, while the pistil abortion genes in males is mainly associated with cytokinin, gibberellin, and auxin pathways. Additionally, we constructed expression regulatory networks for the development of female pistils, aborted pistils and stamens in male flowers, revealing key regulatory genes and signaling pathways involved in papaya organ development. Furthermore, we systematically identified 65 members of the MADS-box gene family and 10 ABCDE subfamily MADS-box genes in papaya. By constructing a phylogenetic tree of the ABCDE subfamily, we uncovered gene contraction and expansion in papaya, providing an improved understanding of the developmental mechanisms and evolutionary history of papaya floral organs.

Conclusions These findings provide a robust framework for identifying candidate sex-determining genes and constructing the sex determination regulatory network in papaya, providing insights and genomic resources for papaya breeding.

Keywords Gene regulatory network, ABCDE module, Pistil and stamen development, Hormone signaling pathway, Evolution

Introduction

Flowers are specialized structures unique in flowering plants that serve as reproductive organs. While most angiosperms develop hermaphroditic flowers with functional pistils and stamens, unisexual flowers are also widely distributed throughout the Plant Kingdom. Approximately 7% of the angiosperm genera, amounting to 959 out of a total of 13,500 genera, have at least one dioecious species. Furthermore, approximately 6% of all angiosperm species, or 14,620 of 240,000, exhibit dioecious characteristics [1, 2]. Although less common compared to hermaphroditism, dioecy occurs in approximately 40% of angiosperm families [2, 3]. Research

*Correspondence:
Jingjing Yue
jingjingyue11@126.com
[1] College of Life Sciences, Fujian Agriculture and Forestry University, Fuzhou 350002, China
[2] Center for Genomics and Biotechnology, Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology, Key Laboratory of Genetics, Breeding and Multiple Utilization of Crops, Ministry of Education, Fujian Agriculture and Forestry University, Fuzhou 350002, China
[3] Fruit Research Institute, Fujian Academy of Agricultural Sciences, Fuzhou 350013, China

suggests that angiosperms have undergone several independent transitions from hermaphroditism to dioecy, the majority of which are relatively recent [2]. Thus, unisexuality in plants likely originated from a hermaphroditic ancestor. The shift from hermaphroditism to dioecy in angiosperms is thought to occur via two primary evolutionary routes: one involving a gynodioecious intermediate stage, characterized by the presence of female and hermaphrodite individuals, referred to as the gynodioecy route, and the other traversing a monoecious intermediate stage, where an individual possesses distinct female and male flowers, known as the monoecy route [4, 5]. Despite the diverse evolutionary forces and intermediates involved in the transition from hermaphroditism to unisexuality, these evolutionary shifts resulted in the same outcome: functional unisexuality.

The *Caricaceae* family encompasses six genera and 35 species, of which 32 are dioecious, two are trioecious, and one species is monoecious. The predominance of dioecious species suggests that dioecy is ancestral in *Caricaceae*, with trioecious and monoecious species evolving more recently [6]. Thus, an evolutionary shift from hermaphroditic to unisexual flowers occurred in the forebears of the *Caricaceae* family, while the appearance of trioecious species is probably an outcome of human cultivation [7]. *Carica papaya* L.*,* commonly known as papaya, is the only species of the genus *Carica* L. [8]. Papaya is a trioecious plant that exhibits three distinct sex forms: male (XY), female (XX), and hermaphrodite (XY^h). Sex determination in papaya is controlled by an XY chromosomal system that is still in the early stages of evolution [9]. Male and hermaphrodite papaya plants are regulated by two Y chromosomes, Y and Y^h, which diverged approximately 4,000 years ago [7]. Y chromosome contains a region responsible for the development of male (Y) and hermaphroditic (Y^h) flower forms [10]. The region that governs male flower development is known as the male-specific region (MSY) and is approximately 8 Mb long and constitutes only 10% of the Y chromosome. While the region that determines hermaphroditic flower development is known as the hermaphroditic-specific region (HSY) [11]. In papayas, females are XX, whereas males and hermaphrodites are heterozygous for X, with the XY and XY^h genotypes, respectively. All combinations of Y and/or Y^h chromosomes are lethal, preventing the pairing and recombination of the YY, YY^h, and Y^hY^h chromosomes [12]. Consequently, no breeding system for papaya can simultaneously produce all three sex types, indicating that both Y chromosomes lack at least one gene critical for development [6]. In addition, the sex types of different papaya varieties vary, and the plants can be either dioecious (populations containing unisexual plants that produce either male or female flowers) or gynodioecious (populations containing female or hermaphroditic plants). Typically, the development of distinct sexes in plants necessitates two mutations: the first leads to male infertility, and the second leads to female infertility [13, 14]. Intermediates to complete dioecy are found in many plants, such as in spinach, which predominantly consists of male and female individuals but also occasionally includes hermaphrodites [15], and strawberries, where males, females, hermaphrodites, and neuter individuals coexist because of the incomplete linkage of the determining factors [16]. However, this was not the case for papaya, as the ancestors of the *Caricaceae* family were dioecious, and hermaphroditic plants emerged due to human domestication. Thus, gynodioecy in domesticated papaya populations represents a regression from dioecy rather than a transitional phase in the evolutionary progression towards dioecy.

Papaya inflorescences arise from the leaf axils, with the type of inflorescence depending on the sex of the plant. Morphologically, female and male plants exhibited distinct inflorescences. Male plants have long, pendulous, multi-flowered inflorescences with slender male flowers that retain a rudimentary pistil at the center, a feature also observed in melons [17]. Female plants possess shorter inflorescences with fewer flowers, and are characterized by large functional pistils. Compared with male flowers, female flowers have larger bracts with a more rounded base and are devoid of stamens. The hermaphrodite plants exhibit short inflorescences with perfect flowers and have slender bracts compared to that in female flowers. Sex reversal is a common phenomenon in natural settings. Environmental factors such as temperature can induce the transformation between hermaphroditic and male flowers in papaya. The pistil of hermaphroditic can undergo abortion, transforming the flowers into a male unisexual flower, thereby changing the plant from hermaphroditic to andro-monoecious. This results in males and hermaphroditic flowers coexisting within the same individual. Conversely, in male plants, the pistil redevelopment within male flowers can convert them into hermaphroditic flowers that are capable of bearing fruit. Additionally, hermaphroditic plants under stress caused by high temperatures, water scarcity, or nitrogen deficiency may aggravate female sterility, potentially resulting in pistil abortion [18–20]. In contrast to hermaphroditic and male plants, female plants are highly stable; their blossoms remain unaffected by environmental variations and do not undergo sexual transformation. The frequent sex reversal observed in papaya suggests that the floral organ development in this species is governed by complex mechanisms.

The process of floral organ development in higher plants is dynamic and involves complex physiological and morphological changes characterized by the interaction and coordination between internal plant signals and external environmental factors. The genetic control of flower development has long intrigued plant developmental geneticists and remains one of the extensively studied areas in plant developmental molecular biology. Currently, a relatively straightforward model known as the ABCDE model has been established using model species such as *Arabidopsis thaliana*, *Antirrhinum majus* (snapdragon), and *Petunia hybrida* to explain the regulatory mechanisms of floral organ development in hermaphroditic flowers. The early ABC model [21] divided the flowers of dicotyledonous plants into four concentric whorls from outside to inside: sepals, petals, stamens, and carpels. This model posits that in dicotyledonous plants, the development of each floral organ whorl is regulated by three classes of homeotic genes, known as the A, B, and C genes which functions either individually or in combination. In *Arabidopsis*, the A-class genes are *APETALA1* (*AP1*) and *AP2*; the B-class genes include *AP3* and *PISTILLATA* (*PI*); and the C-class gene is *AGAMOUS* (*AG*). The A-class genes solely determine the formation of the first whorls (sepals). The interaction between A- and B-class genes dictates the development of the second whorl (petals), whereas the joint action of B- and C-class genes specifies the third whorl (stamens). Additionally, C-class genes independently regulate the fourth whorl (carpel). The loss of A-class gene function transforms the first whorl (sepals) into carpels and the second whorl (petals) into stamens. The loss of B-class gene function transforms the second whorl (petals) into sepals and the third whorl (stamens) into carpels. The loss of C-class gene function transforms the third whorl (stamens) into petals and the fourth whorl (carpels) into sepals. Thus, the same class of genes regulates the development of adjacent floral organ whorls, whereas A- and C-class genes mutually repress each other. [21–23].

With the advancement of research, additional floral homeotic genes have been identified, revealing phenomena that the ABC model fails to explain. In 1995, researchers identified *FBP7* and *FBP11* during ovule development in *Petunia* [24]. *FBP11* is specifically expressed in the ovule primordium, integuments, and funiculus, and is considered a master regulator of ovule development. These genes determining the characteristics of the placenta and ovules, are classified as D-class genes and are responsible for ovule development. In 2000, researchers discovered that the AGAMOUS-like genes *AGL2*, *AGL4*, and *AGL9* are involved in the specification of floral organ identity [25]. The single or double mutants of these genes did not exhibit obvious phenotypes, while only the triple mutants showed a clear phenotype, with all floral organs developing into sepals, resembling the BC double mutant phenotype. These three genes were later termed *SEP1*, *SEP2*, and *SEP3* and were observed to participate in the specification of the inner three floral organ whorls. The discovery of the D-class and *SEP* genes enriched and refined the previously proposed ABC model, leading to its revision. The *SEP* genes were designated as E-class genes, and, together with the D-class genes, they led to the expansion of the ABC model into the ABCDE model. *SEP4* was identified as being expressed in all four floral organ whorls. Additionally, in the *sep-1*, *2*, *3*, *4* quadruple mutants, all floral organs were transformed into leaf-like structures [26].

The establishment of an ABCDE floral development model represents a milestone in plant developmental molecular biology. In the classical ABCDE model, all genes except the A-class gene *AP2* belong to the MADS-box gene family, specifically the MIKC subfamily. This underscores the vital function of MADS-box genes, especially the MIKC type, in controlling floral organ growth. Plant genomes harbor an extensive array of MADS-box genes, greatly exceeding those implicated in the ABCDE floral development model [27, 28].

The term "MADS-box" is derived from the initials of several transcription factors that were initially identified, including *MINICHROMOSOME MAINTENANCE1* (*MCM1*) in *Saccharomyces cerevisiae* [29], *AGAMOUS* (*AG*) in *Arabidopsis thaliana* [30], *DEFICENS* (*DEF*) in *Antirrhinum majus* [31], and serum response factor (*SRF*) in humans [32]. MADS-box genes control various aspects of plant growth and development, such as flowering time, specification of floral organ identity, maintenance of meristem activity, and fruit development and maturation [33, 34]. These genes are key regulators of floral organ identity and development [35] and widely distributed among seed plants [34, 36–41]. A defining characteristic of all MADS-box genes is a highly conserved DNA sequence of approximately 180 nucleotides known as the MADS-box [42], which encodes a MADS protein domain (M-domain), typically 56–60 amino acids in length, involved in recognizing and binding to the CC-A rich-ment-GG (CArG) motif of target genes [43]. Ancestral MADS-box genes are hypothesized to have duplicated before the most recent common ancestor (MRCA) of eukaryotes, evolving into two major branches: SRF-like (Type I) and MEF2-like (Type II) MADS-box genes [44]. Type I genes can be further divided into M-alpha, M-beta, and M-gamma subfamilies, whereas Type II genes feature a distinct MIKC structure, consisting of MADS-box (M), intervening (I), keratin-like (K), and variable C-terminal (C) transcription activation

domains [45, 46]. In land plants, Type II transcription factors form two major subclades, MIKC* and MIKC^C types [47]. However, recent studies have provided greater support for the monophyletic clade formed by MIKC* and M-type [48]. The emergence of MIKC^C is closely linked to the successful evolution of flowering plants [49]. Genes that exhibit similar functions or expression patterns are often closely related to, or clustered within, the same MADS-box family. Furthermore, MADS-box proteins demonstrated the ability to form multimeric complexes, potentially providing a molecular basis for the interaction of different floral homeotic gene combinations [50]. Consequently, a floral quartet model was proposed [51], providing deeper insights into the regulatory mechanisms of floral organ development. The floral quartet model is applicable beyond core eudicots, to monocots such as rice [51–55]. Thus, studying the MADS-box transcription factor gene family is essential for understanding the development and evolution of floral organs in plants.

Arabidopsis contains 107 members of the MADS-box gene family, with 68 classified as Type I and 39 as Type II [36]. In *Vitis vinifera* L. (grape), 90 MADS-box genes have been identified, including 48 Type I and 42 Type II [56]. In *Oryza sativa* L. (rice), 75 *MADS-box* genes have been identified, 16 of which are classified as Type II [57]. In *Pyrus communis* (European pears) and *Pyrus bretschneideri* (Chinese pears), 73 and 74 MADS-box genes, respectively, have been identified, with M-type genes further subdivided into M-alpha, M-beta, and M-gamma subfamilies [58]. However, the MADS-box family has not been comprehensively studied in papaya [59–61], and the identification and evolution of homologous genes related to the ABCDE model of flower development have not yet been systematically explored. Moreover, the limitations of early genome sequencing and assembly technologies have resulted in notable inaccuracies in the previously published draft of the papaya genome [62] and its corresponding annotation files.

Therefore, utilizing the recently published papaya reference genome [63], this study aims to clarify the regulatory pathways involved in stamen and pistil development, while also identifying critical factors affecting the pistil formation. Notably, the sex determination in papaya is not governed by a single gene. As of now, the specific genes responsible for this process have not been identified, and the mechanism behind pistil abortion in hermaphroditic papaya flowers remains unclear. Developing a regulatory network for the early development of papaya stamens and pistils will offer essential insights into identifying the genes involved in papaya's sex determination.

Results

Early stage pistil degeneration in male flowers

To investigate the dynamic gene expression during the development of reproductive organs in female and male papaya flowers, we collected 17 samples from the pistils of female flowers (FP), rudimentary pistils of male flowers (MP), and stamens of male flowers (MS) across five developmental stages, as well as from the apical meristems of female (FAM) and male plants (MAM). Each sample had three biological replicates and was subjected to RNA sequencing (RNA-seq). The morphological characteristics of female flowers with a functional gynoecium and male flowers with rudimentary pistils and functional stamens were observed at five stages (Fig. 1a, Table S1). Morphological observations revealed that pistil degeneration occurs at a very early stage in male flowers, confirming the classification of male flowers as type I unisexual flowers.

To explore the relationships among the 17 samples, we calculated their correlation coefficients. Correlation analysis of the RNA-seq datasets indicated a strong linear correlation between different floral organs during the early stages of flower development. Notably, the third (M3S) and fourth (M4S) stages of stamen development in male flowers generally showed lower correlations with other developmental stages (Fig. S1a). Conversely, a strong correlation was observed among the different stages of aborted pistil development in male flowers. Principal component analysis (PCA) of the RNA-seq datasets of the 17 samples (including three replicates for each sample) showed clear distinctions at different stages and a compact distribution of biological replicates (Fig. S1b-S1d). In the PCA of the MP, PC1 and PC2 explained 39.84% and 18.37% of the variation, respectively (Fig. S1c). The proximity of sample distributions across different groups reflected the similarity in gene expression patterns between stages (Fig. S1c), which is consistent with previous correlation analyses. Thus, the early determination of aborted pistils in male flowers causes functional loss during development, resulting in the non-specific expression of related functional genes.

GO and KEGG enrichment analysis reveals the critical role of plant hormone signaling pathways in papaya floral organ development

To understand differential gene expression, we divided the transcriptome datasets of the 17 samples into three groups based on the floral organ type: functional pistils, rudimentary pistils, and stamens. Each group used the earliest stage (the first period) as a control for differential analysis (Fig. 1b and 1c, Supplementary Data 1). The number of differentially expressed genes (DEGs) in the

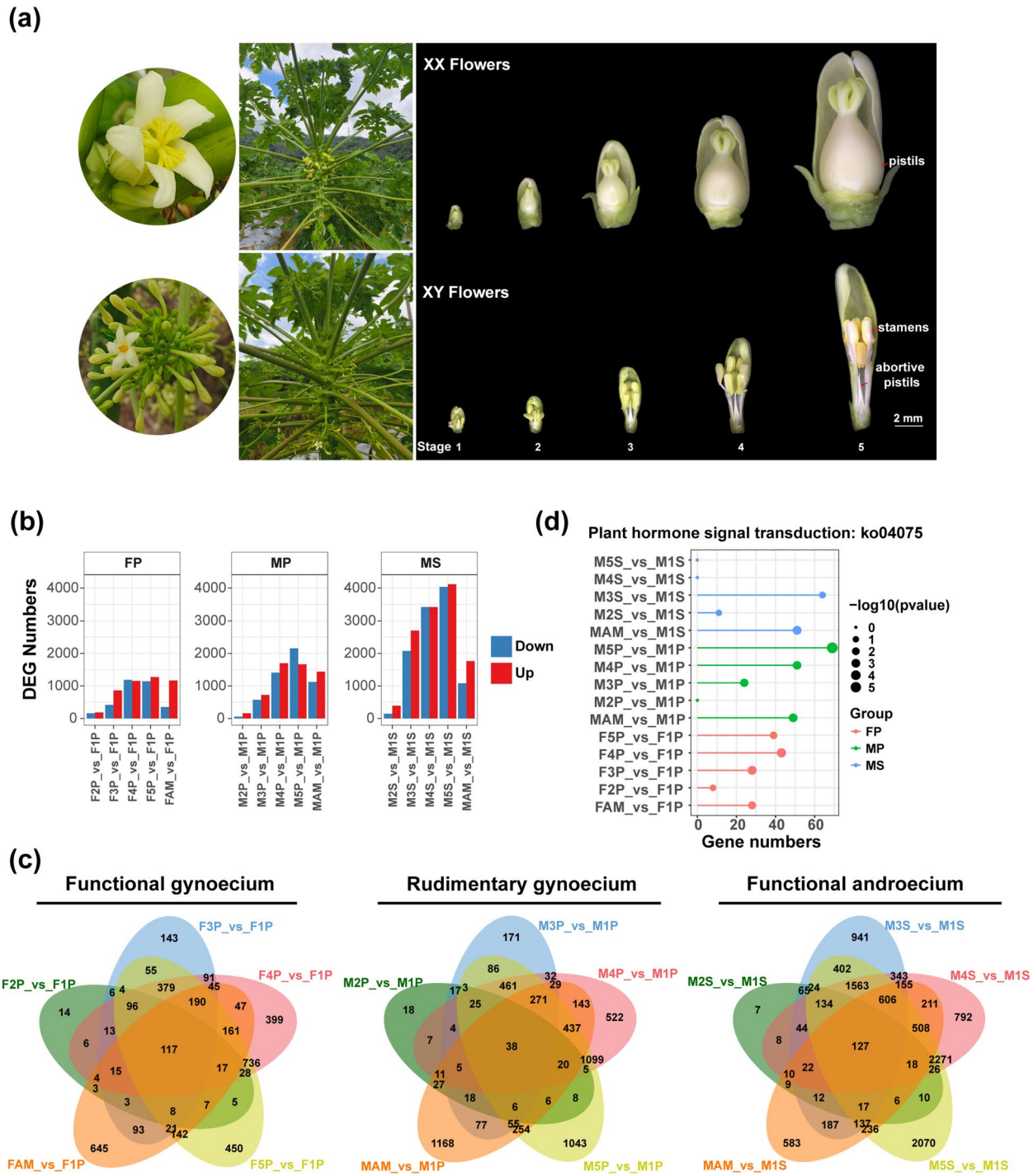

Fig. 1 Morphological characteristics of papaya floral organs and dynamic expression profiles at different developmental stages. (**a**) Morphological features of female and male papaya plants, and female pistils (FP), aborted pistils in male flowers (MP), and stamens in male flowers (MS) at five developmental stages. Flower lengths are 1.5–2.5 mm, 2.5–4.5 mm, 4.5–6 mm, 7–12 mm and 15–20 mm from Stage1 to Stage5, respectively. (**b**) Number of differentially expressed genes (DEGs) in apical meristem (AM) and stages 2–5 compared to stage 1 across female pistils (FP), aborted pistils in male flowers (MP), and stamens (MS), totaling 15 groups. (**c**) Venn diagram of DEGs across five DEG sets in female pistils, aborted pistils in male flowers, and stamens. FAM and MAM represent the apical meristems of female and male papaya plants, respectively. (**d**) Enrichment of plant hormone signal transduction pathways across 15 different combinations. For each group, the top 25 pathways with p-value < 0.05 were selected for statistical analysis

pistils of female flowers increased to 346 (F2P vs. F1P), 1,279 (F3P vs. F1P), 2,344 (F4P vs. F1P), 2,416 (F5P vs. F1P), and 1,518 (FAM vs. F1P). In the rudimentary pistils of male flowers, the number of DEGs was 218 (M2P vs. M1P), 1,298 (M3P vs. M1P), 3,109 (M4P vs. M1P), 3,817 (M5P vs. M1P), and 2,565 (MAM vs. M1P). The number of DEGs during different pistil stages of female flowers was similar to that of the rudimentary pistils of male flowers (Fig. 1b, Table S2). The number of DEGs increased progressively during the development of pistils in female flowers, rudimentary pistils in male flowers, and stamens in male flowers, indicating clear differences in gene expression patterns between the early and late stages of floral organ development, with only a few key genes being likely involved in regulating floral organ identity in the early stages. Notably, the number of DEGs increased dramatically during the middle to late stages of stamen development (M3S-M5S) (Fig. 1b, Table S2), suggesting that stamen development in papaya may be subjected to more extensive gene regulation. Among the female pistils, aborted pistils in male flowers, and stamens, there were 117, 38, and 127 DEGs, respectively, which were common to all five DEG groups (Fig. 1c). This indicates that these genes are differentially expressed throughout the developmental period and likely play crucial roles in the overall development of the pistils and stamens. The number of common DEGs (38) was lower in aborted pistils of male flowers than in the other two organs (117 and 127). Moreover, the DEGs between different developmental stages gradually increased, starting from the first stage, in the normal pistils of female flowers and abortive pistils of male flowers. However, the DGEs in MAM sharply decreased (MAM vs. M1S) compared to M5S vs. M1S (Fig. 1c). Additionally, in the three comparisons, only a small number of stage-specific DEGs were present in stage2 vs. stage1 comparisons (14 DEGs in F2P vs. F1P, 18 DEGs in M2P vs. M1P, and 7 DEGs in M2S vs. M1S) (Fig. 1c).

Furthermore, Gene Ontology (GO) enrichment analysis of DEGs in the pistils of female flowers (FPs vs. F1P) revealed significant enrichment in the top 25 GO terms related to responses to biotic and abiotic stimuli and hormone responses throughout pistil development (Fig. S2a-S2e). Specifically, there were 46, 158, 236, 218 DEGs in the F2P vs. F1P, F3P vs. F1P, F4P vs. F1P and F5P vs. F1P groups enriched in the GO term "response to hormone." The number of DEGs markedly increased during the F2P developmental stages, highlighting the critical regulatory role of hormones in early pistil development. In contrast, in the later stages (F4P vs. F1P and F5P vs. F1P), 188 and 213 DEGs were significantly enriched in response to light stimulus, respectively (Fig. S2d-S2e). In the rudimentary pistils of male flowers (MPs vs. M1P), DEGs showed significant enrichment in hormone responses, which gradually increased over time (Fig. S3a-S3e). Unlike the pistils of female flowers, DEGs during the early development of rudimentary pistils in male flowers (M2P vs. M1P group) were significantly enriched in GO terms related to the transport of energy substances, including sugars and amino acids, and the response to abscisic acid (Fig. S2b and S3b). From the M2P stage, GO terms related to light stimulation or photosystem (including GO:0009416, GO:0019684, and GO:0042550) were significantly enriched (Fig. S3b-S3e). In the stamens of male flowers, DEGs during the early stages of development (M2S vs. M1S and M3S vs. M1S) were significantly enriched in GO terms related to pollen tissue structure and shoot system development, including pollen wall assembly, pollen exine formation, and shoot system development (Fig. S4b-S4c). In addition, single-organism processes were significantly enriched throughout stamen development (Fig. S4a-S4e).

Additionally, we conducted Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis on the aforementioned sets of DEGs (FPs vs. F1P, MPs vs. M1P, and MSs vs. M1S) and selected the top 30 KEGG pathways from each group for further analysis. These results suggest that the role of the plant hormone signal transduction pathway varies among female pistils, aborted pistils in male flowers, and stamens (Fig. 1d). In female pistils, the plant hormone signal transduction pathway was significantly enriched throughout development. As the developmental stages progressed, the number of DEGs in the pathway increased: 8 DEGs in F2P vs. F1P, 28 in F3P vs. F1P, 43 in F4P vs. F1P, and 39 in F5P vs. F1P, indicating its critical role in pistil development. The enrichment trend in the rudimentary pistils of male flowers was similar to that in the pistils of female flowers. However, no enrichment was observed in the M2P vs. M1P group. In the stamens of male flowers, significant enrichment was observed only in the first two comparisons (M2S vs. M1S, and M3S vs. M1S). Significant differences were observed in the enrichment of DEGs in plant hormone signal transduction pathways during the development of different floral organs, suggesting different roles of plant hormones during stamen and pistil development.

Next, we merged and removed redundancies from the DEGs enriched in plant hormone signal transduction pathways in female pistils, aborted pistils in male flowers, and stamens, resulting in three sets of DEGs specific to each tissue. A Venn diagram was constructed using these three DEG sets, which identified 8, 24, and 20 genes specific to female pistils, aborted pistils, and stamens, respectively (Fig. S1e). Based on the expression patterns and potential functions of homologous

genes, we identified three highly expressed genes in the female pistil-specific enriched gene set: one involved in cytokinin regulation and two in gibberellin regulation. In the aborted-pistil-specific gene set of male flowers, five auxin-related genes were identified, all of which showed lower expression levels in aborted pistils than in normal pistils. Additionally, five genes involved in cytokinin regulation were identified in the stamen-specific gene set, all of which showed higher expression in stamens. The differential enrichment of plant hormone signal transduction pathways across female flowers, aborted pistils, and male flowers, along with the specific enrichment of hormone-regulating genes, suggests that hormone pathways are potentially involved in the regulation of sex differentiation in papaya, with certain hormones potentially playing key roles in this process.

Identification of MADS-box gene family and ABCDE subfamily members in papaya

MADS-box members are predominantly involved in floral organ development and play critical roles in floral organ formation. Using the MADS-box protein sequences of *Arabidopsis thaliana* and *Vitis vinifera*, we identified 65 MADS-box gene family members in papaya (Table S3). To further understand the evolutionary relationships among members of the MADS-box gene family, we constructed a phylogenetic tree combining the MADS-box gene families of the three aforementioned species (Fig. 2a). In papaya, MIKC^{C} subfamily members (29 genes) and M-type subfamily members (30 genes) were almost equally abundant, whereas in *Arabidopsis*, MIKC^{C} subfamily members were fewer than in the M-type members.

The phylogenetic tree was distinctly separated into five evolutionary clades: M-alpha, M-beta, M-gamma,

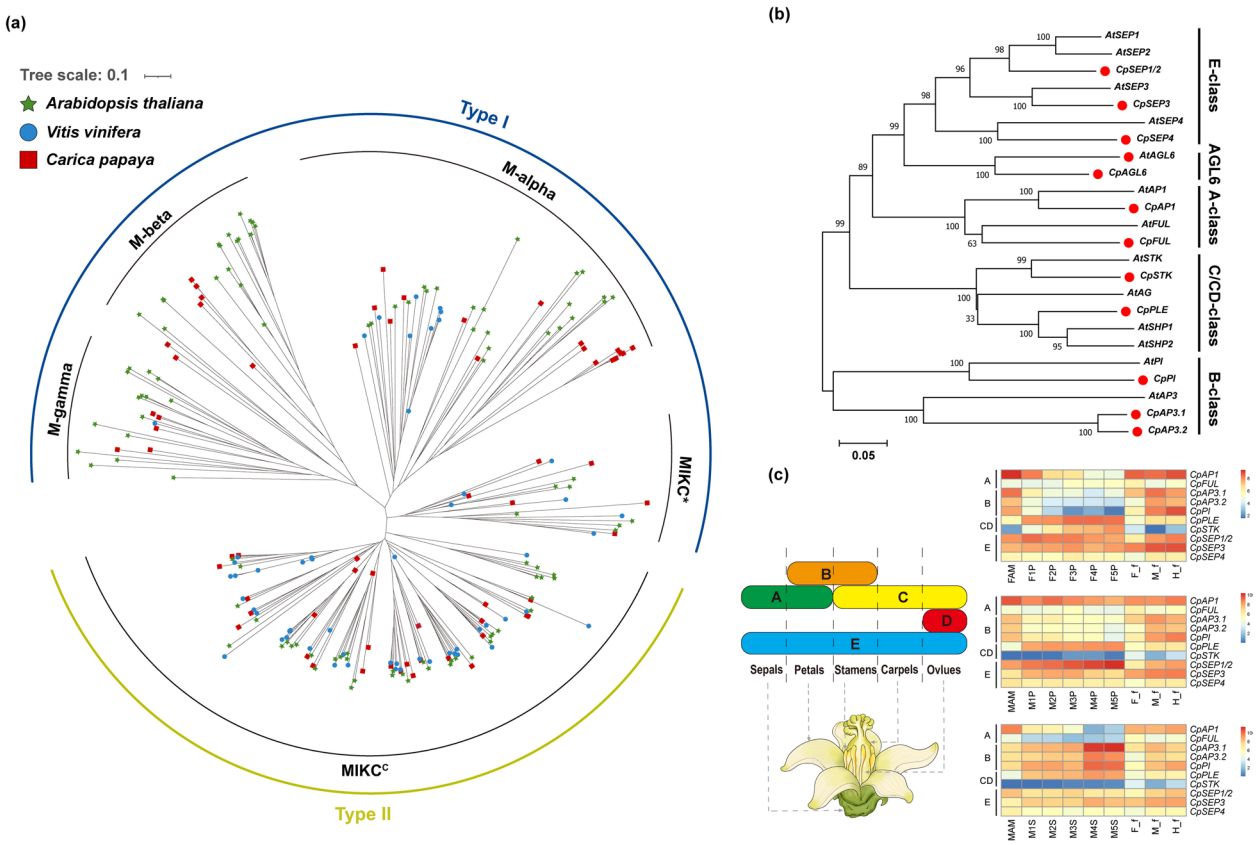

Fig. 2 Phylogenetic analysis of MADS-box transcription factor gene family members and identification of ABCDE homologs in papaya. (**a**) Phylogenetic tree constructed using *Arabidopsis* (103 members), grape (54 members), and 65 MADS-box members identified in this study, with 1000 bootstrap replicates, using the neighbor-joining (NJ) method. (**b**) Identification of ABCDE homologous genes in papaya using *Arabidopsis* ABCDE genes as references, and construction of a phylogenetic tree. (**c**) Schematic diagram of the ABCDE flower development model and heatmap of the expression patterns of the 10 identified ABCDE subfamily MADS-box genes in papaya. In addition to RNA-seq data from the apical meristems of female (FAM) and male plants (MAM), female pistils (FP), aborted pistils in male flowers (MP), and stamens (MS) used in this study, RNA-seq data from female (F_f), male (M_f), and hermaphroditic flowers (H_f) were included to show the expression patterns of the ABCDE subfamily MADS-box genes. The TPM values were $\log_2$-transformed before being used as input data for heatmap visualization

MIKC*, and MIKC^C (Fig. 2a, Table S4). A recent study suggested that the monophyletic phylogeny that groups the previously defined Type II MIKC* subfamily with the Type I M-type subfamily is significantly more robust than the previously proposed hypotheses of monophyletic MIKC* and MIKC^C division [48]. Therefore, classifying MIKC* members as part of the Type I group rather than Type II is recommended. In the present study, we classified 65 MADS-box members in papaya according to this revised model, assigning them to Type I (including M-alpha, M-beta, M-gamma, and MIKC*) and Type II (containing only MIKC^C). The gene structure, conserved motifs, and domains of the MADS-box genes in papaya were analyzed to further understand the MADS-box genes in papaya. MEME predicted 20 motifs, of which motifs 1 and 10 were widely distributed among the 65 MADS-box members, whereas motif 15 appeared to be specific to Type II members (Fig. S5b). Analysis of the conserved protein domains revealed that all Type II genes contained the MADS_MEF2_like domain, and the K-box domain was exclusively present in Type II genes (Fig. S5c). Gene structure analysis indicated clear differences between the M-type and MIKC^C-type members. Most M-type members contain a single exon, whereas the MIKC^C-type members typically exhibit longer gene structures with numerous introns (Fig. S5d).

MADS-box genes are widely involved in the morphogenesis and development of floral organs, particularly the Type II members. Following the establishment of the ABCDE flower development model, the MADS-box gene family is recognized as essential for floral organ development. Moreover, in the ABCDE model of flower development, all MADS-box genes belong to the Type II group. In our study, we identified 10 ABCDE homologous genes from 65 MADS-box gene family members in papaya, including *CpAPETALA1* (*CpAP1*), *CpFRUITFULL* (*CpFUL*), *CpAPETALA3.1* (*CpAP3.1*), *APETALA3.2* (*CpAP3.2*), *CpPISTILLATA* (*CpPI*), *CpPLENA* (*CpPLE*), *CpSEEDSTICK* (*CpSTK*), *CpSEPALLATA1/2* (*CpSEP1/2*), *CpSEPALLATA3* (*CpSEP3*), and *CpSEPALLATA4* (*CpSEP4*) (Fig. 2b, Table S5).

To further investigate the transcript levels of ABCDE genes in papaya floral organs, we analyzed the expression patterns of these genes at different developmental stages in female pistils, abortive pistils in male flowers, and stamens (Fig. 2c, Table S6). The A-class gene *CpAP1* exhibited a broad expression pattern with high expression levels in female, male, and hermaphroditic flowers. Its expression is also elevated during the early stages of pistil development and in aborted pistils in male flowers and stamens. B-class genes, including *CpAP3.1*, *CpAP3.2*, and *CpPI*, were highly expressed in the stamens of male papaya flowers, with sharp upregulation during the later stages of stamen development (M4S and M5S). In addition, B-class genes were more highly expressed in male and hermaphroditic flowers than in female flowers. The expression patterns of the CD class genes *CpSTK* and *CpPLE* were notably different. *CpPLE* was broadly expressed throughout the development of female pistils, rudimentary pistils in male flowers, and stamens, with gradually increasing expression levels. It was also expressed at higher levels in male and hermaphroditic flowers than in female flowers. In contrast, *CpSTK* was almost exclusively expressed in the female pistils, which is consistent with the results of recent studies [24, 64, 65], with its expression levels sharply increasing during development. The E-class genes, *CpSEP1/2*, *CpSEP3*, and *CpSEP4*, were also broadly expressed throughout the development of female pistils, rudimentary pistils, and stamens of male flowers. Notably, *CpSEP1/2* was highly expressed in the rudimentary stamens of male flowers, with gradual upregulation. However, *SEP* genes show differential expression patterns between flowers of different sexes in papaya, with higher expression observed in male and hermaphroditic flowers, particularly pronounced for *CpSEP1/2* and *CpSEP3*.

In summary, the expression patterns of the ABCDE genes in papaya strongly support the ABCDE model of floral development, providing substantial evidence for its applicability to angiosperms, especially core eudicots.

Evolutionary insights to ABCDE subfamily MADS-box genes reveal gene contraction and expansion in papaya

To further understand the evolution of the ABCDE subfamily of MADS-box genes, we selected the ABCDE subfamily of MADS-box genes and *AGL6* from gymnosperms (*Ginkgo biloba*, *Gnetum gnemon*, *Picea abies,* and *Cycas panzhihuaensis*), basal angiosperms (*Amborella trichopoda* and *Nymphaea colorata*), monocots (*Oryza sativa* and *Zea mays*), and eudicots (*Arabidopsis thaliana* and *Carica papaya*) to construct a phylogenetic tree. The phylogenetic tree clearly clustered 100 ABCDE and AGL6 proteins from 10 seed plants into five subgroups: *AP1/FUL* (A-class genes, 17), *AP3* and *PI* (B-class genes, 20), *AGAMOUS* (CD- or C-class genes, 24), *SEP* (E-class genes, 24), and *AGAMOUS-LIKE 6* (*AGL6* genes, 15) (Fig. 3a, Table S5).

The A-class genes are involved in the development of sepals and petals. Our results indicate that A-class genes are specific to angiosperms, which is consistent with previous reports [49, 66, 67]. Additionally, in monocots, A-class genes seem to have undergone duplication, resulting in more paralogous genes than in dicots (Fig. 3a and 3b, Table S5). The emergence and subsequent specialization of A-class genes in angiosperms may have

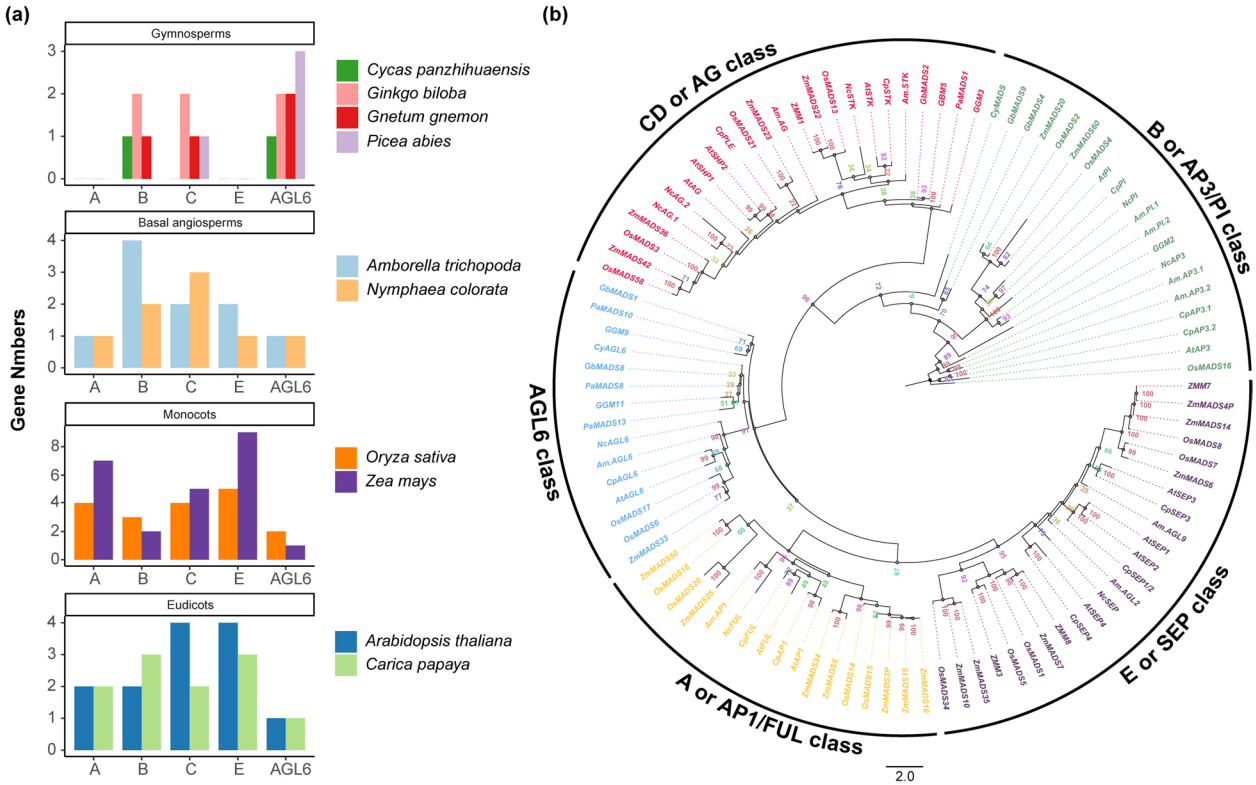

Fig. 3 Evolutionary analysis of ABCDE and *AGL6* genes in 10 seed plants. (**a**) Statistics of A, B, CD, E, and *AGL6* class genes in gymnosperms (*Cycas panzhihuaensis*, *Ginkgo biloba*, *Gnetum gnemon* and *Picea abies*), basal angiosperms (*Amborella trichopoda*, and *Nymphaea colorata*), monocots (*rice* and *maize*), and dicots (*Arabidopsis* and papaya). (**b**) Phylogenetic tree of ABCDE and *AGL6* genes in 10 seed plants. Best-fit model: JTT + F + R6, constructed with 1000 bootstrap replicates using the maximum likelihood (ML) method

substantially contributed to the diversity of floral organs in these plants.

B-class genes control the development of petals and stamens. In gymnosperms, this subfamily of genes has been identified in *Ginkgo biloba* (two B-class genes), *Cycas panzhihuaensis,* and *Gnetum gnemon* (one B-class gene), but not in *Picea abies*. In angiosperms, two B-class genes were identified in *Nymphaea colorata*, *Zea mays*, and *Arabidopsis thaliana*, three B-class genes were obtained in *Oryza sativa* and *Carica papaya*; and *Amborella trichopoda* had four B-class genes, indicating a clear gene duplication event (Fig. 3a and 3b, Table S5). In the present study, B-class genes were generally more abundant in angiosperms than in gymnosperms. Notably, in papaya, despite the relatively small number of genes, two paralogous genes, *CpAP3.1* and *CpAP3.2*, were present within the AP3 gene family. The greater abundance of B-class genes in flowering plants likely contributes to the formation of more complex floral organs, which is consistent with the conclusions of Shen et al. [49].

CD-class genes are involved in controlling stamen and pistil development. In the phylogenetic tree, the CD-class genes formed a distinct clade with 96% bootstrap support, indicating high confidence in this branch (Fig. 3b). The number of CD-class genes is higher in angiosperms than gymnosperms, suggesting that the increase in CD-class genes may be closely related to the evolution of complex floral organs in angiosperms. Notably, in papaya, we identified only two CD-class genes, *CpSTK* and *CpPLE*, which distinctly differed from other angiosperms such as rice with four CD-class genes, maize with five, and *Arabidopsis* with four (Fig. 3a, Table S5). In the more primitive basal angiosperms *Amborella* and *Nymphaea*, two and three CD class genes were observed, respectively, with *Nymphaea* having two copies of the *AG* gene (Table S5).

E-class genes interact with ABCD genes [68] and are involved in the development of floral organs. Phylogenetic analysis suggests that E-class genes, similar to A-class genes, are specific to angiosperms. Further analysis revealed that *SEP1* and *SEP2* were more closely related to *SEP4* than to *SEP3*, which is consistent with previous research [49, 69]. Additionally, in monocots, the *LOFSEP* subclass of E class genes (also known as the *SEP1/2/4* clade) formed a distinct branch, separating E class genes into two clades, along with those from core

eudicots and basal angiosperms, with a bootstrap support of 95% (Fig. 3b). Phylogenetic analysis indicated that the *LOFSEP* subclass genes in monocots were more specialized than the E-class genes in the core eudicots and basal angiosperms (Fig. 3b). The emergence of these new organ forms provide protection to monocot flowers and represents an evolutionary advancement in floral organ development.

AGL6 is involved in floral development in angiosperms [38, 70] and cone formation in gymnosperms [71]. We identified *AGL6* across all gymnosperm samples (Fig. 3a and 3b, Table S5). Additionally, multiple copies of *AGL6* were present in *Ginkgo biloba*, *Gnetum gnemon,* and *Picea abies*, highlighting the importance of *AGL6* in the development of gymnosperm reproductive organs. The phylogenetic tree showed a clear separation between the *AGL6* subfamilies in angiosperms and gymnosperms, with the angiosperm clade exhibiting longer branches, suggesting potential evolutionary diversification of the *AGL6* gene in angiosperms (Fig. 3b).

In our phylogenetic analysis, the MADS-box genes of the ABCDE subfamily and *AGL6* were clearly distinguished in both gymnosperms and angiosperms. In particular, as a core eudicot, the ABCDE and *AGL6* genes showed the closest phylogenetic relationships with those of *Arabidopsis*, which is a core eudicot. This finding is consistent with the taxonomic classifications and has been validated using molecular evolutionary analyses [9, 62]. Additionally, we observed that CD-class and E-class genes within the ABCDE subfamily, appeared to be lost in papaya, in contrast to other angiosperms, especially grasses (*Poaceae*) [38].

Clustering analysis of dynamic gene expression unveils flower organ-specific clusters in papaya pistil and stamen development

To explore the dynamic expression changes in floral organs across different developmental stages, we performed cluster analysis on the mRNA datasets. Cluster analysis of DEGs from F5P vs. F1P, M5P vs. M1P, and M5S vs. M1S was conducted using the Gap Statistic method to determine optimal clusters (Fig. S6a-S6c). We identified 12 clusters in the female pistils (Fig. S7, Table S7), 19 clusters in male abortive pistils (Fig. S8, Table S8) and 23 clusters in male stamens (Fig. S9, Table S9).

In female pistils, cluster 8 genes showed increased expression over time, whereas their expression remained low in male abortive pistils (Fig. 4a and 4b). Expression pattern and functional annotation analyses revealed that many genes involved in plant hormone signal transduction, especially auxin-related genes such as *CpIAA32* (*INDOLE-3-ACETIC ACID INDUCIBLE 32*), *CpGH3*.1, and *CpEOD1* (*ENHANCER1 OF DA1*), were enriched in cluster 8 (Fig. 4d, Table S10). In contrast, cluster 16 genes in male abortive pistils exhibited the opposite expression patterns (Fig. 4d and 4e), indicating their potential role in the negative regulation of female reproductive function. Several genes involved in plant hormone regulation were also identified in cluster 16. For instance, *CpLOG5* (*LONELY GUY 5*) and *CpCKX1* (*CYTOKININ OXIDASE/DEHYDROGENASE 1*) are directly involved in the synthesis and degradation pathways of cytokinins, while *CpGASA1* (*GAST1 PROTEIN HOMOLOG 1*), *CpSCL15* (*SCARECROW-LIKE 15*), and *CpARF4* (*AUXIN RESPONSE FACTOR 4*) participate in hormone signaling pathways such as gibberellin and auxin. Additionally, many genes involved in oxidative stress response and sucrose transport were enriched in cluster 16, including *CpMT2A* (*METALLOTHIONEIN 2A*), *CpPER12* (*Peroxidase 12*), *CpLETM2* (*LEUCINE ZIPPER-EF-HAND-CONTAINING TRANSMEMBRANE PROTEIN 2*), *CpF3H* (*FLAVANONE 3-HYDROXYLASE*), *CpAPX1* (*ASCORBATE PEROXIDASE 1*), *CpUGT84A1*, and *CpSTP8* (*SUGAR TRANSPORT PROTEIN 8*) (Fig. 4f, Table S10).

Moreover, cluster 8 genes in female pistils revealed significant enrichment in carpel and organ morphogenesis (Fig. 4g), as well as in plant hormone signal transduction and diterpenoid biosynthesis pathways (Fig. 4h). This suggests their role in the occurrence and formation of female organs, or the execution of female-specific functions. Conversely, in cluster 16, male abortive pistil genes were enriched in phloem development and amino acid transport (Fig. 4g) as well as in secondary metabolism and starch and sucrose metabolic pathways (Fig. 4h).

Similarly, cluster 20 genes were upregulated only during the early stages of stamen development and peaked

(See figure on next page.)

Fig. 4 Expression pattern and enrichment analysis of clusters with specific expression in female pistils and aborted pistils in male flowers. (**a**-**b**) Dynamic expression trends of DEGs in cluster 8 in female pistils. (**c**) Schematic of papaya female flower and expression patterns of genes specifically expressed in cluster 8. (**d**-**e**) Dynamic expression trends of DEGs in cluster 16 in aborted pistils in male flowers. (**f**) Schematic of papaya male flower and expression patterns of genes specifically expressed in cluster 16. The TPM values were row-normalized before being used as input data for heatmap visualization. (**g**) GO enrichment analysis for DEGs in cluster 8 in FP and in cluster 16 in MP. (**h**) KEGG pathway enrichment analysis for DEGs i in cluster 8 in FP and in cluster 16 in MP. Top 25 GO terms or KEGG pathways are shown

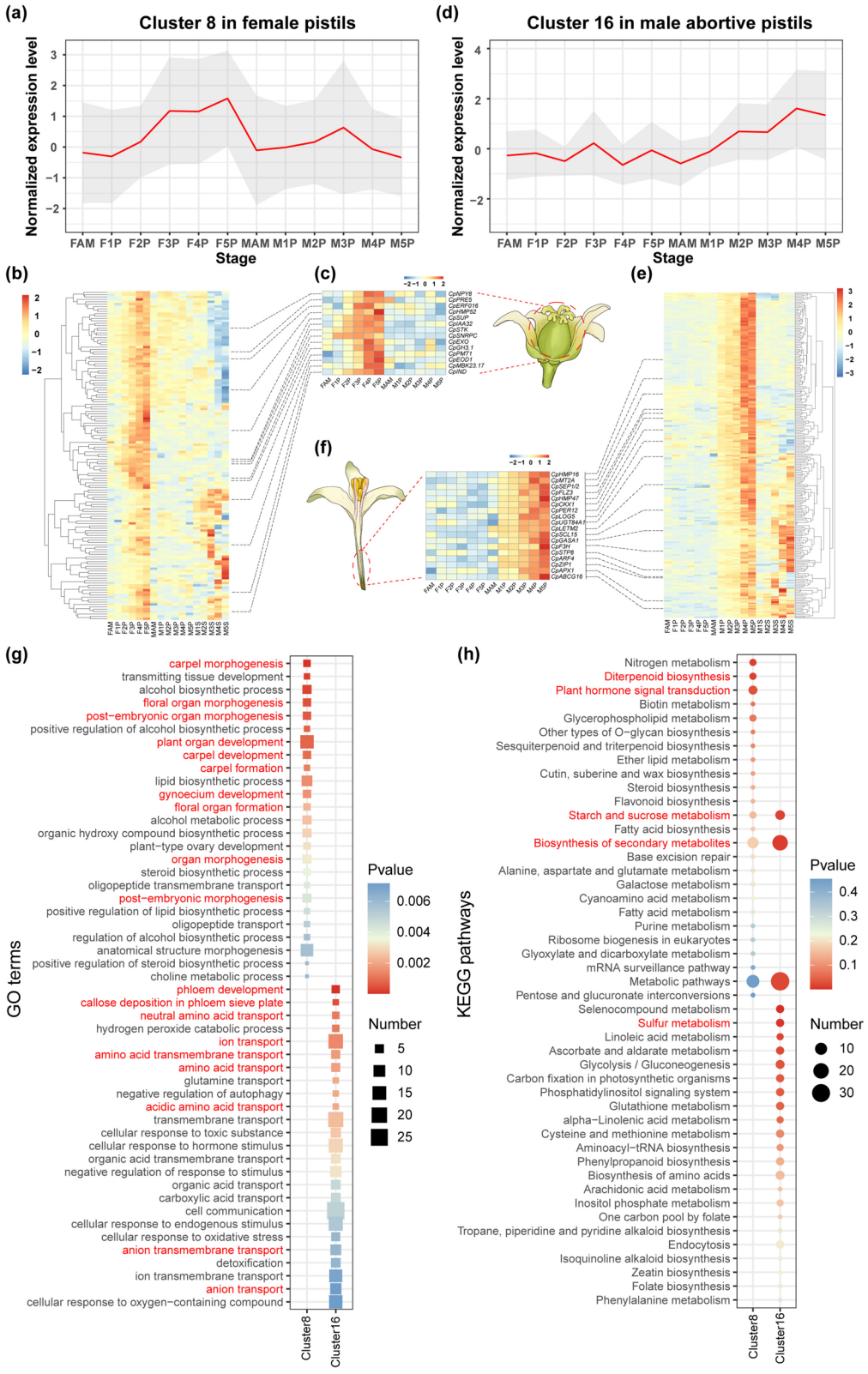

Fig. 4 (See legend on previous page.)

in the middle stage (M3S) in the stamens of male flowers (Fig. S10a). Moreover, many genes were involved in the formation of the pollen outer wall, glycosyl and lipid hydrolysis, and transport were found in cluster 20, including *CpACOS5* (*ACYL-COA SYNTHETASE 5*), *CpSWEET15*, *CpENGASE85B* (*ENDO-BETA-N-ACETYGLUCOSAMINIDASE 85B*), *CpGELP77* (*GDSL-TYPE ESTERASE/LIPASE 77*), *CpLPTG10* (*GLYCOSYLPHOSPHATIDYLINOSITOL-ANCHORED LIPID PROTEIN TRANSFER 10*), and *CpCYP704B1* (*CYTOCHROME P450*) (Fig. S10b, Table S10). These genes are closely related to cell wall remodeling, which is crucial for floral organ morphogenesis. Additionally, most genes in *Arabidopsis* are crucial for stamen and pollen development and are essential for maintaining male reproduction [72–74]. Enrichment analysis showed that the genes in cluster 20 were significantly enriched in GO terms such as anther morphogenesis and response to external stimuli (Fig. S10c and S10d), suggesting their potential role in promoting stamen morphogenesis and anther maturation.

Gene regulatory networks revealed the key roles of CD-class and E-class subfamily MADS-box genes in papaya pistil development

To preliminarily analyze the regulatory network of pistil and stamen development in papaya, we constructed gene regulatory networks (GRNs) for floral organ development. GRNs for female pistils (using DEGs of F5P versus F1P), male abortive pistils (using DEGs of M5P versus M1P), and male stamens (using DEGs of M5S versus M1S) were constructed using Mutual Rank (MR) analysis.

The GRN for female pistils contained 1,841 potential regulatory relationships, involving 1,063 genes (Table S11). Analysis of the gene expression patterns within the network revealed a gradual increase in the expression levels of *CpSTK* and *CpSUP* (*SUPERMAN*) in the pistils of female flowers, whereas they were almost unexpressed in the abortive pistils and stamens of male flowers (Fig. S11a and Table S12). This pistil-specific expression pattern suggests that *CpSTK* and *CpSUP* play important roles in pistil development and morphogenesis. In *Arabidopsis*, *STK* encodes a MADS-box transcription factor that is expressed within ovules [64, 65]. *STK* is widely recognized as a D-class gene that specifies ovule identity [46, 75] and is essential for ovule specification. Additionally, *CpSTK* and *CpSUP* both belonged to cluster 8 in the pistil of female flowers (Fig. 4a and 4c). Therefore, we extracted a subnetwork with *CpSTK* and *CpSUP* as core genes and constructed a pistil regulatory network (PRN) for pistil development (Fig. 5a). Analyzing the gene expression patterns in the PRN revealed that *CpIND* (*INDEHISCENT*) and *CpZFP11* show expression patterns similar to those of the core genes (Fig. S11a, Table S12). Both *ZFP11* and *SUP* are members of the zinc finger protein family, and studies in *Arabidopsis* revealed that the ERF repressor domain is essential for *ZFP11* to function as a transcriptional repressor, consistent with *SUP* [76]. *IND* is the closest paralog of the *HEC* gene in *Arabidopsis*. Although *IND* mainly functions in the development of the fruit dehiscence zone, *ind* mutants exhibit minor defects in stigma elongation [77, 78], suggesting the role of *CpZFP11* and *CpIND* in papaya pistil development and are regulated by core genes (*CpSTK* and *CpSUP*) within the network. Enrichment analysis of genes in the PRN highlighted significant enrichment in the biological process (BP) category, particularly in GO terms associated with flower organ formation, floral organ identity specification, and anatomical boundary formation (Fig. S11b, Table S17). This further suggests that the genes in the subnetwork centered on *CpSTK* and *CpSUP* are closely related to the initiation and development of floral organs and likely play important roles in pistil development.

In male abortive pistils, the GRN included 6,882 potential regulatory relationships involving 2,609 genes (Table S13). Further analysis revealed that the expression of *CpSEP1/2*, an E-class gene, markedly increased during the development of abortive pistils in male flowers (Fig. S11c, Table S14). Flowering plants require angiosperm-specific E-class genes to facilitate the interaction of B- and C-class genes, which is necessary for the specification of third-whorl floral organs (stamens). Similarly, the development of fourth whorl floral organs (carpels and ovules) in angiosperms requires E-class and C-class genes [79]. Furthermore, *CpSEP1/2* belonged to cluster 16 of the abortive pistils in male flowers (Fig. 4d and 4f). Therefore, we extracted a subnetwork using *CpSEP1/2* as the core gene and constructed the abortive pistil regulatory network (APRN) (Fig. 5b). By analyzing the expression patterns of genes in the APRN, we found that *CpARF4*, *CpABIG1*, *CpTPR3*, *CpIAA16*, and *CpGASA1* had expression patterns similar to those of the core genes *SEP1/2* (Fig. S11c, Table S14). Subsequently, we performed an enrichment analysis of the APRN genes. In the cell component (CC) category, these genes were significantly enriched in GO terms such as cell–cell adherens junction (GO:0005913), apical junction complex (GO:0043296), and adherens junction (GO:0005912) (Fig. S11d, Table S18). This suggests that APRN plays a role in the regulation of cell polarity. In the biological processes (BP) category, these genes were significantly enriched in the single-organism processes (25/29, 86.21%, $p < 0.05$), responses to abscisic acid (8/29, 27.59%, $p < 0.05$), and photoinhibition GO terms

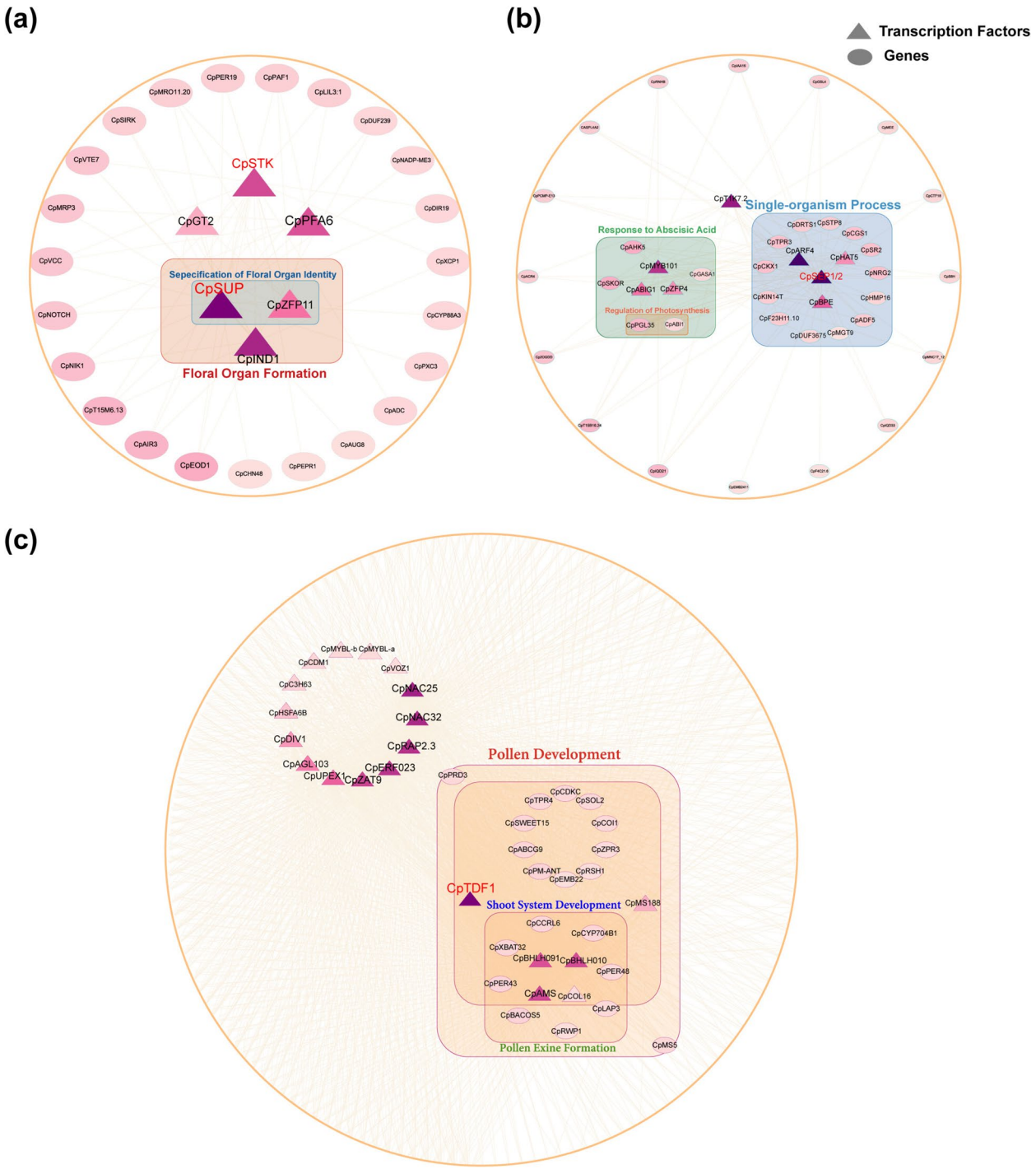

Fig. 5 Construction of regulatory networks for the development of female pistils (FP), aborted pistils in male flowers (MP), and stamens (MS). (**a**) Construction of a regulatory network for female pistil development (PRN) with *CpSTK* and *CpSUP* as core genes. (**b**) Construction of a regulatory network for aborted pistil development in male flowers (APRN) with *CpSEP1/2* as core genes. (**c**) Construction of a regulatory network for stamen development (SRN) with *CpTDF1* as the core gene. Genes in each network were classified based on significant GO term enrichment results. Genes within boxed regions in the network are enriched in the same GO term

(Table S18). "Single-organism process" refers to all physiological and developmental activities that occur within an individual organism that do not involve interactions with other organisms. The majority of APRN genes were enriched in this GO term, suggesting that they may be closely related to pistil abortion. However, the specific biological functions of the GO terms remain unclear. Additionally, the significant enrichment of genes responding to abscisic acid indicated that abscisic acid may be involved in pistil abortion.

In male stamens, GRN included 35,862 potential regulatory relationships involving 6,718 genes (Table S15). By analyzing gene expression levels, functional annotations, and previous clustering analysis results, we selected a subnetwork with *CpTDF1* as the core gene to construct the stamen regulatory network (SRN) for male flower stamen development (Fig. 5c). In *Arabidopsis*, *DEFECTIVE IN MERISTEM DEVELOPMENT AND FUNCTION 1* (*TDF1*) primarily regulates the differentiation and function of tapetal cells [80]. In papaya stamens, *CpTDF1* is expressed almost exclusively during the early stages of stamen development (M1S-M3S), peaking at the M3S stage, and then remaining almost completely unexpressed in the later stages. The specific expression pattern of *CpTDF1* highlights its important role during the early stages of stamen development. Additionally, in the *CpTDF1* subnetwork, many genes exhibit similar expression patterns, and their homologs in *Arabidopsis* have been reported to be involved in pollen formation and maturation. (Fig. S11e, Table S16). For example, in *Arabidopsis*, the *BHLH010* and *BHLH091* transcription factors jointly regulate anther development and are essential for pollen wall formation [81]. In the papaya SRN, *CpBHLH010* and *CpBHLH091* exhibit high expression levels exclusively during stamen development, peaking at the M3S stage, indicating their involvement in stamen development and pollen wall formation. Within the SRN, *CpAMS*, a member of the bHLH family, is specifically expressed during stamen development. The *AMS* homolog in *Arabidopsis* is a key regulator of sporopollenin biosynthesis, secretion, and pollen wall formation, and plays a complex role in the early stages of pollen development [82]. Additionally, *AMS* is directly regulated by *TDF1* and is essential for pollen development [83]. *CpMS188*, another transcription factor from the MYB family, is upregulated during stamen development. Its counterpart in *Arabidopsis*, *MS188*, regulates tapetum development and is vital for pollen exine synthesis [84]. Furthermore, GO enrichment analysis of genes within the SRN showed notable enrichment in stamen and anther development, particularly in the anther wall tapetum, in the biological process (BP) category (Fig. S11f, Table S19). The expression profiles and enrichment findings for the genes in the SRN suggest their potential role in regulating the identity of male flower stamens and the early development of the anther tapetum.

To further validate the expression of the ABCDE subfamily MADS-box and core genes in the papaya regulatory network, we conducted quantitative real-time PCR (qPCR). The quantitative expression data for these 12 genes were generally consistent with the trends observed in the transcriptome analysis (Fig. 6). B-class genes exhibited higher expression levels in the stamens of male flowers, highlighting their crucial role in the development and maintenance of stamen identity in papaya. The pistil-specific expression pattern of the CD class gene *CpSTK* indicates its crucial role in the normal development of papaya pistils. E-class genes showed broad expression in female pistils, aborted pistils, and stamens during development, consistent with their role in regulating the development of all floral organs as described in the ABCDE model. The specific high expression of *CpSEP1/2* in aborted pistils suggests a possible functional specialization in papaya. Quantitative data for ABCDE gene expression reinforces the significance of the ABCDE model in understanding papaya flower development. Similarly, the quantitative expression of the core gene *CpTDF1* within the SRN matched the transcriptomic findings. The targeted expression of *CpTDF1* in stamens, along with the known function of its *Arabidopsis* counterpart, underscores the significance and evolutionary conservation of this gene during stamen development.

Discussion

Papaya is a trioecious species whose sex determination is governed by a recently evolved XY chromosomal system. Two Y chromosome variants (with only 0.4% nucleotide divergence) determine the development of male (Y) and hermaphrodite (Y^h). The short divergence time along with high sequence similarity between the Y and Y^h chromosomes present an opportunity to identify candidate genes responsible for sex differentiation. These genes are strong candidates as female suppressors during the early stages of papaya chromosomal evolution [7]. Additionally, the presence of abortive pistils in male flowers highlights the crucial role of female suppressor genes in sexual differentiation. In the present study, normal pistils from female flowers, aborted pistils from male flowers, and stamens were used to construct potential developmental regulatory networks for these organs in papaya to identify candidate sex-determining genes and construct a regulatory network for sex determination.

Discrepancies in the anatomical structures of male and female flowers and the expression profiles of floral organs across various stages suggest that the development of male and female reproductive organs in papaya

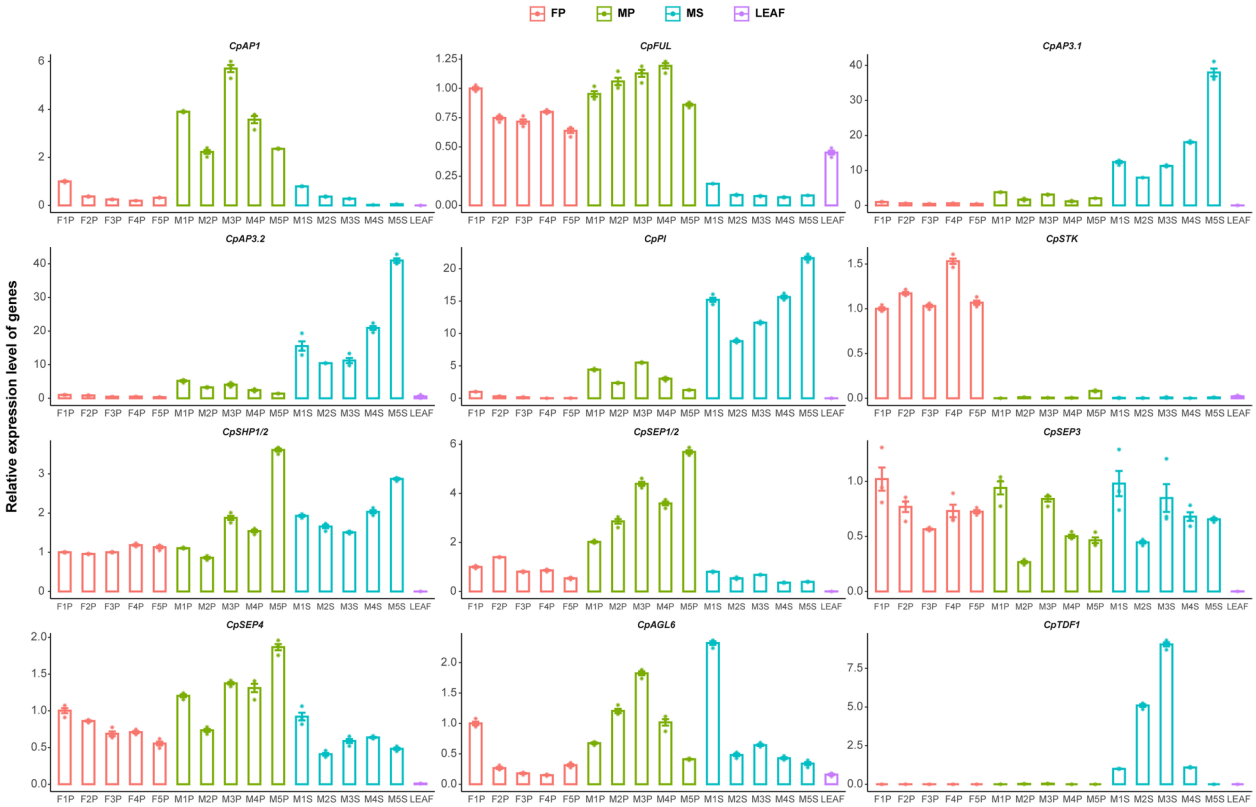

Fig. 6 Expression patterns of 12 genes during the development of papaya female pistils (FP), aborted pistils in male flowers (MP), and stamens (MS). Relative quantification of 10 ABCDE subfamily MADS-box genes and *CpAGL6* was calculated based on the expression level in the F1P sample, while *CpTDF1* is relatively quantified using the M1S sample due to its specific expression in stamens. Using female pistils (FP), abortive pistils from male flowers (MP), stamens (MS) at five developmental stages, and leaf tissue to confirm the expression levels of 10 ABCDE subfamily MADS-box genes, *CpAGL6*, and *CpTDF1*

may follow different regulatory mechanisms. This divergence is particularly evident during the later stages of development, when functional specialization across different floral organ identities (pistils, aborted pistils, and stamens) becomes more pronounced (Fig. 1b, S1b). Plant hormones, particularly auxins, are vital for the growth of papaya pistils and stamens, and the degeneration of pistils might be associated with an excess of auxin levels [85]. Our results showed that plant hormone signaling pathways were significantly enriched during papaya pistil and stamen development (Fig. 1d). Several auxin response factors (ARFs), such as *CpARF2*, *CpARF5*, and *CpARF11*, as well as auxin pathway-related genes (*CpIAA9*, *CpIAA11*, *CpIAA32*, and *CpGH3.1*), exhibit high expression levels during pistil development but are downregulated in abortive pistils. This suggests that auxin positively regulates pistil development. Notably, *CpARF5* and *CpGH3.1* maintain consistently high expression throughout all developmental stages of pistils in female flowers, whereas their expression is significantly reduced in abortive pistils of male flowers across five stages (Fig. S1e, Table S10). The coordinated regulation of these genes likely plays a crucial role in maintaining auxin homeostasis in pistils. Additionally, *CpARR9* is significantly upregulated during the late stages of pistil development but remains at low levels in abortive pistils. In contrast, *CpLOG5* is highly expressed only in abortive pistils, suggesting a potential negative feedback regulation of pistil development by cytokinin-related genes (Fig. S1e and Table S10). Cluster analysis of the dynamic gene expression patterns also revealed a significant enrichment of hormone-related genes with tissue-specific expression patterns, further highlighting the importance of plant hormones in the development of papaya pistils and stamens (Fig. 4, Table S10).

MADS-box proteins are crucial for floral organ development. In this study, 65 members of the MADS-box family were identified in papaya plants. According to the revised classification, the Type I subfamily contained slightly more genes (36 members) than the Type II subfamily (29 members). However, Gramzow and Theissen [60], identified 262 MADS-box family members

in papaya, with 229 belonging to the Type I subfamily, markedly more than the Type II subfamily. Additionally, they suggested an expansion of Type I members in papaya, which is uncommon among most angiosperms, although the exact reason for this remains unclear. Our findings differ markedly from those of the previous studies, with discrepancies primarily attributable to the limitations in the sequencing technology and assembly and annotation software available at the time of drafting of earlier genome assemblies, resulting in the low quality of the previous genome assembly and annotation [62]. In contrast, the present study used the recently published papaya reference genome [63], which features significantly improved completeness and annotation than earlier draft genomes, thus providing a solid foundation for our research. In this study, Type II subfamily members in papaya generally had more exons and longer gene structures, which is consistent with previous findings [60]. Additionally, in female pistils, aborted male pistils, and stamens, Type I subfamily members exhibited low or almost no expression, whereas Type II subfamily members generally exhibited higher and more specific expression levels (Table S20). This may be because Type I members are less essential to the plant genome or are active only under specific conditions [43]. Additionally, they suggested that Type I genes may exhibit transposon-like activity, potentially derived from (retro)transposition. Thus, Type I genes in plants may represent a novel and unprecedented class of transposable elements lacking established sequence markers [43], similar to certain transposons carrying MADS-box genes found in maize and its relatives [86].

Compared to Type I subfamily members, Type II members have been extensively studied and are known to govern the development of floral organs. The most well-known example is the ABCDE model of floral organ development established for core eudicots, which is also applicable to papaya. In the ABCDE model, almost all genes belonged to the MADS-box family, specifically from the Type II subfamily. In this study, we identified 10 ABCDE homologous genes in papaya, which differs from the findings of Shen et al. [49] who identified 12 genes (one additional A-class gene, *CpMADS20*, and one additional E-class gene, *CpMADS6*, compared to our study) (Table S21). Upon further comparison and validation, we found that *CpMADS20* and *CpMADS6* correspond to *CpMADS43* (*sunup01G0007110*, which aligns with *FUL*) and *CpMADS48* (*sunup09G0003500*, which aligns with *SEP1/2*) in our study, respectively. Although both genes belonged to the Type II subfamily, they were not identified as the best matches (Fig. 2b, Table S5, Table S6, Table S21). Additionally, as a core eudicot, the ABCDE genes of papaya are most closely related to those of *Arabidopsis*, which is consistent with its evolutionary position [9]. However, the abundance of ABCDE genes in papaya differed from that in other angiosperms. For instance, the number of CD- and E-class genes in papaya was lower than that in other angiosperms (Fig. 2b and 3b), suggesting the lack of duplication of CD- and E-class genes in the papaya lineage. This phenomenon is not uncommon in papaya; since it diverged from *Arabidopsis* approximately 72 million years ago, no genome duplication has occurred in the papaya lineage [62]. However, the presence of two paralogs of B-class gene *AP3* in papaya highlights its crucial role for the development of papaya floral organs.

Through the identification and evolutionary analysis of ABCDE homologous genes in ten seed plants, we found that A- and E-class genes were present only in angiosperms. Some studies suggest that gymnosperms have genes similar to the A or E class genes of angiosperms [79, 87, 88]. Additionally, we hypothesized that the *LOF-SEP* subfamily of E-class genes is highly evolved in monocots. One example supporting this hypothesis is that in rice, *OsMADS1*, *OsMADS5*, and *OsMADS34* collectively regulate floral meristem determinacy and specify spikelet organs, which are typical inflorescence structures of *Poaceae* plants, by positively regulating other MADS-box floral homeotic genes [55]. This evolution appears to be consistent with the complex structure of floral organs in monocots. B and CD class genes, the earliest diverging genes in the ABCDE model, are widely present in both gymnosperms and angiosperms, indicating that the roles of B and CD class genes in floral organ development may be highly conserved. *AGL6* genes closely related to E-class genes, are widely present in gymnosperms and play an important role in the development of their floral organs.

In conclusion, this study preliminarily constructed a regulatory network for the development of papaya floral organs, including female pistils, aborted pistils in male flowers, and stamens. The stamen development network in male flowers had the highest number of DEGs and the greatest potential for regulatory interactions compared to the networks governing female pistil development and the development of aborted pistils in male flowers. This suggests that the regulatory architecture underlying stamen development is more intricate, exhibiting more pronounced variations in regulatory circuits between the initial and late phases. Initially, we constructed a regulatory network for female pistil development, focusing on the core genes, *CpSTK* and *CpSUP*. As a D-class gene, *CpSTK* is specifically expressed in female papaya pistils. In *stk* mutants, ovule development is abnormal, whereas double mutants of other D-class genes (*SHP1/SHP2*) produce normal ovules [89]. This suggests that

STK is involved in determining pistil identity. In *Arabidopsis*, the *SHORT VEGETATIVE PHASE* (*SVP*) gene acts upstream of *STK* and directly targets its regulation [65, 90, 91]. Studies by Lee et al. [92] and Chae et al. [93] suggested that *SVP* functions as a repressor of pistil development in papaya. Consequently, the male-specific *CpSVP-Y* (*CpMADS64*) allele in papaya may modulate *CpSTK* activity, thereby governing pistil formation. The expression patterns of *CpSVP-Y* and *CpSTK* also suggested a potential regulatory relationship (Fig. 2c, Table S6, Table S20). Another core gene, *CpSUP*, also exhibited pistil-specific expression patterns. In *Arabidopsis*, *SUPERMAN* controls the specification and maintenance of the boundary separating the stamens in the third whorl from the carpels in the fourth whorl [94–96]. Previous studies have shown that *SUP* can inhibit the expression of B-class genes in the fourth whorl, acting as a boundary gene. The *sup* recessive mutation causes extra stamens to form within the normal third whorl stamens, causing developmental defects in the carpels of the fourth whorl [94]. Subsequent studies discovered that *SUP* can function as active inhibitors of auxin biosynthesis, and a lack of auxin biosynthesis leads to flowers with supernumerary stamens [96]. Therefore, *CpSUP* may be involved in regulating auxin homeostasis within the inner whorls of papaya floral organs, preventing auxin overaccumulation, and ensuring normal pistil development. In the network, we also identified two genes, *CpZFP11* and *CpIND*, with expression patterns resembling those of the core genes, suggesting their regulation by these core genes, particularly *CpIND*. Research on *Arabidopsis* identified that *IND* is crucial for the development of the stigma and the appropriate distribution of auxin, where the *IND-SPT-NGA* functional complex directs the morphogenesis of the apical pistil and stigma development [97]. Additionally, a regulatory network for the development of aborted pistils in male flowers was constructed, centered on the core gene *CpSEP1/2*. In *Arabidopsis*, *sep1-sep2-sep3* triple mutants exhibit severe defects (all floral organs are converted to sepals), whereas *sep1-sep2-sep4* triple mutants are phenotypically normal [25, 26, 69]. This suggests that *SEP1/SEP2* genes are not crucial in determining floral organ identity. However, the reduced copy number of E-class genes in papaya compared to that in other angiosperms implies their distinct role in papaya floral organ development. Additionally, *CpSEP1/2* genes maintained high expression levels in the aborted pistils of male papaya flowers compared to that in female pistils and male stamens, with levels increasing over time. This expression pattern indicates that *CpSEP1/2* may act as inhibitors of pistil development. Additionally, in *Arabidopsis,* many downstream target genes in the network have been reported to be involved in the regulation of plant hormones, such as abscisic acid (*ABIG1* [98], *TPR3* [99]), auxin (*ARF4* [100], *IAA16* [101]), and gibberellin (*GASA1* [102]). *CpARF4* and *CpIAA16* exhibit significantly higher expression levels in abortive pistils compared to normal pistils (Table S14), suggesting their potential role in pistil abortion. These findings highlight the complex regulatory mechanisms of auxin pathway-related genes in papaya pistil development. Therefore, we consider that excessively high and stable expression of E-class genes may affect the development of the fourth whorl floral organs. Finally, we constructed a regulatory network for stamen development in male flowers centered around the core gene *CpTDF1*. The core gene (*TDF1*) and its potential downstream target genes (*BHLH010*, *BHLH091*, *CpAMS*, *MS188*, and *CYP704B1* [74], *ABCG9* [103], and *ACOS5* [72] have all been reported to be involved in pollen and stamen development in *Arabidopsis*. These genes also exhibited specific expression patterns during papaya stamen development, further supporting the credibility of our constructed regulatory network and providing a basis for further studies on the mechanisms of stamen development in papaya.

In brief, this study analyzed the dynamic changes in the development of papaya pistils and stamens, and preliminarily constructed a gene regulatory network for the development of papaya floral organs (female pistils, aborted pistils in male flowers, and stamens in male flowers). The network provides a benchmark for further research into the mechanisms of sex determination in papaya, as well as the regulatory processes of flower organ development and pistil degeneration in papaya. We also hypothesized that the different developmental fates of pistils in male and female flowers may result from changes in the expression of key genes during the early stages, affecting the synthesis and transport of plant hormones (primarily auxins), and ultimately leading to pistil abortion in male flowers (Fig. 7). Additionally, through the identification and phylogenetic analysis of MADS-box members in papaya, we revised the number of MADS-box family members and characterized the gene structure and expression pattern differences between the Type I and Type II subfamily members, providing a theoretical basis for future research. Furthermore, we focused on the identification of ABCDE subfamily MADS-box genes in papaya and conducted a deep analysis of their evolutionary relationships in seed plants, thereby providing new insights into the expression patterns and evolutionary relationships of the ABCDE model in papaya floral development.

Upon identifying the gene that inhibits pistil development, we aim to tackle the issue of hermaphroditic papaya plants reverting to male under high temperatures or nutrient deficiencies, which results in fruit drop and

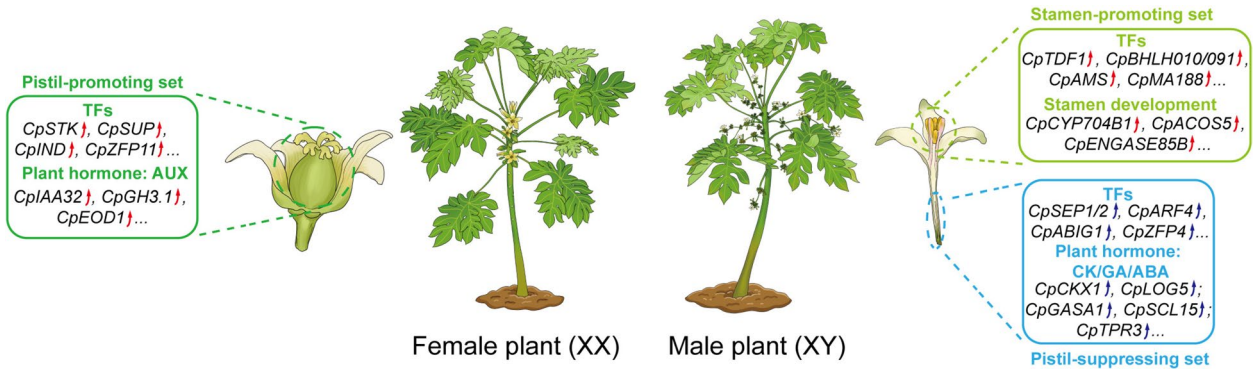

Fig. 7 The gene sets regulating the development of female and male floral organs in papaya. The three gene sets include the pistil-promoting set, pistil-suppressing set, and stamen-promoting set, each associated with the regulation of pistil or stamen development. These sets contain potential transcription factors involved in the regulation of floral organ development, as well as genes involved in the synthesis or regulation of relevant hormones. Arrows following the genes indicate their upregulation in the corresponding tissues

significantly reduced yield. By applying different hormones to hermaphroditic papaya flowers, we seek to identify hormones capable of suppressing the gene that inhibits pistil development. This approach would enable us to artificially regulate the gene's expression under high-temperature conditions, thereby mitigating the conversion of hermaphroditic flowers to male. Furthermore, by knocking out the gene that inhibits pistil development in hermaphroditic papaya plants, we can achieve sex-stable hermaphroditic papaya plants.

This study not only enhances our understanding of the genetic mechanisms underlying sex differentiation and morphological variation in papaya but also provides a theoretical foundation for future research on the mechanisms of sex determination and the role of floral organs in shaping sex-specific traits. Ultimately, this could contribute to the development of hermaphroditic papaya cultivars with stable ovary, thereby reducing resource waste in papaya cultivation.

Materials and Methods

Plant materials

The papaya flower samples used in this study, including those from the ZH, AU9, and SunUp varieties, were collected from the experimental greenhouse and cultivation base at the Center for Genomics and Biotechnology, Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology, Fujian Agriculture and Forestry University, Fuzhou, China. For the ZH variety, pistils (FP) from female flowers, rudimentary pistils (MP), and stamens (MS) from male flowers were collected from the field and immediately dissected. Additionally, shoot apical meristems from both female and male plants were collected, as obtaining flower apical meristems proved challenging. The samples were immediately frozen in liquid nitrogen and stored at −80 °C. Flowers with lengths ranging from 1.5–2.5 mm and 2.5–4.5 mm, due to the small and delicate nature of their pistils, rudimentary pistils, or stamens, were dissected under a stereo microscope. For the AU9 variety, whole male flowers (M_f) and female flowers (F_f) were collected from the field, selecting those with a flower length of less than 6 mm, and then immediately frozen in liquid nitrogen and stored at −80 °C. For the SunUp variety, whole hermaphroditic flowers (H_f) were collected under similar conditions, with flowers smaller than 6 mm in length, and also frozen in liquid nitrogen and stored at −80 °C.

Generation of datasets for papaya floral organ development

To establish datasets of gene expression patterns during different stages of floral organ development in papaya, we collected functional pistils from female flowers (FP), rudimentary pistils (MP), and functional stamens from male flowers (MS) at five developmental stages. Additionally, apical meristems of male (MAM) and female (FAM) papaya plants were collected as references. In total, 17 samples representing various stages of floral organ development were obtained to construct the expression profile of papaya reproductive organs. Each sample had three biological replicates (see Table S1 for sample information).

Correlation analysis of samples in papaya flower organ development datasets

To explore differences among the 17 samples, Pearson correlation was used to calculate the correlation coefficients between samples, and the R package corrplot (version 0.92) was employed to create a correlation heatmap. To verify the sampling quality of the three biological

replicates of each sample, principal component analysis (PCA) was performed to reduce the dimensionality of the gene expression matrix (transcripts per million, TPM).

Gene expression and differential expression analysis of floral organ development in papaya

The Fastp software (version 0.23.2, https://github.com/OpenGene/fastp) was used for quality control and filtering of raw mRNA sequencing data to obtain clean data with parameters: -z 4, -q 20, -u 30, -n 5, -w 8. Subsequently, the HISAT2 software (version 2.1.0, https://daehwankimlab.github.io/hisat2/download/) was used to map the clean data to the reference genome. The custom R script RunFeatureCounts was used to quantify the read counts per gene, and the quantitative results were standardized to TPM using default parameters. Differential gene expression between groups was analyzed using the Trinity software (version 2.11.0, https://github.com/trinityrnaseq/trinityrnaseq/wiki/Installing-Trinity) and the DESeq2 package in R. Differentially expressed genes (DEGs) were identified based on log2FoldChange (|log2FoldChange|> 1) and p-value (< 0.05). Blastp software (version 2.13.0, https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/) was used to compare sequences with the TAIR11 database (https://www.arabidopsis.org/) to determine the potential function of each gene. The online tool jvenn (https://jvenn.toulouse.inrae.fr/app/example.html) was used to create Venn diagrams. Then, The EGGNOG-MAPPER website (http://eggnog-mapper.embl.de/) was used for the functional annotation of the whole genome protein sequences of papaya. The OmicShare platform (https://www.omicshare.com/tools/) was used for enrichment analysis of the DEG sets, and the results were visualized. Heatmaps of gene expression levels were generated using the R package pheatmap (version 1.0.12, https://cran.r-project.org/web/packages/pheatmap/index.html). The TPM values were $\log_2$-transformed before being used as input data for heatmap visualization.

Identification of the MADS-box gene family in papaya

To identify members of the MADS-box family in *Carica papaya*, we used protein sequences from *Arabidopsis thaliana* and *Vitis vinifera*. The sequences for *Arabidopsis* were obtained from the TAIR database (http://www.arabidopsis.org/), while those for grape were sourced from the grape genome database (http://www.genoscope.cns.fr). We employed the MADS-box HMM (Hidden Markov Model: PF00319), downloaded from the Pfam database, to perform an HMM search within the papaya genome. To ensure the accuracy of the predicted sequences, we verified the protein sequence domains using the NCBI Conserved Domain Search (CD-Search) (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). Next, we constructed a phylogenetic tree of the MADS-box gene family across different species using the neighbor-joining method of MEGA11 (https://www.megasoftware.net/), with the following parameters: 1,000 bootstrap replicates, Poisson model, and partial deletion of gaps. We further enhanced the visualization of the phylogenetic tree using the iTOL website (https://itol.embl.de/). Additionally, the Expasy website (https://web.expasy.org/compute_pi/) was used to predict the physicochemical properties of the papaya proteins. Finally, we performed motif prediction using the MEME tool (https://meme-suite.org/meme/) with the protein sequences. The gene structure and domains of the MADS-box gene family in papaya were visualized using TBtools-II software (https://github.com/CJ-Chen/TBtools-II/releases) for comprehensive analysis. The gene structures of certain papaya MADS-box family members were manually corrected using the IGV-GSAman software (https://gitee.com/CJchen/IGV-sRNA), based on next-generation transcriptome data (Supplementary Data2).

Homologous alignment and evolutionary analysis of ABCDE subfamily MADS-box genes in papaya

To identify ABCDE genes in papaya, homologous comparisons with known ABCDE genes in *Arabidopsis* were performed. Alignment results were filtered based on criteria such as alignment length, mismatches, and gap numbers to obtain the best alignments. ABCDE genes in papaya were preliminarily determined based on the presence and integrity of conserved protein domains. A systematic phylogenetic analysis was performed using the neighbor-joining method, comparing the protein sequences of ABCDE genes in *Arabidopsis* and candidate genes in papaya.

To explore the evolution of ABCDE genes, we collected protein sequences of ABCDE genes from gymnosperms, basal angiosperms, monocots, and eudicots. These sequences, along with identified ABCDE homologs in papaya, were used to construct a phylogenetic tree. Multiple sequence alignment of the protein sequences was performed using ClustalW (version 2.1, http://www.clustal.org/clustal2/) with default parameters. The optimal evolutionary model was determined using the ModelFinder module of IQ-TREE (version 2.3.5, http://www.iqtree.org/), with parameters: -m MF -mtree -T AUTO. Subsequently, a phylogenetic tree was constructed using RAxML-NG (version 1.2.2, https://github.com/amkozlov/raxml-ng) with the following parameters: –all, –model JTT+F+R6, –bs-trees 1000, –threads AUTO. Evolutionary history was inferred using the Maximum Likelihood (ML) method. Bootstrap tests with 1000 replicates were conducted to show the percentage of replicate trees in which associated taxa clustered together.

The final dataset included 100 protein sequences and 1232 positions, with analyses conducted in RAxML-NG. Figtree software (version 1.4.4, https://github.com/rambaut/figtree/releases) was used to visualized the phylogenetic trees.

Clustering analysis of RNA datasets of different developmental stages of floral organs

To investigate dynamic changes in gene expression profiles during floral organ development, clustering analysis of DEGs expression datasets was performed. The optimal number of clusters was determined using the Gap Statistic Method, utilizing the kmeans function in R. The Mfuzz package in R was then used to cluster the gene expression dataset based on the fuzzy c-means algorithm, and the results were visualized.

Construction of gene regulatory networks for different floral organs

The custom R script cor.test.mk.R calculated the correlation between genes using Pearson correlation. Significant correlations (coefficients > 0.9 and test values < 0.01) were retained. Another custom R script, PCC_rank.R, ranked Pearson correlations between genes, and the correlation network was filtered based on ranking. The upstream 2000 bp sequences from the transcription start sites of the gene set were used to predict transcription factor binding site motifs using AME tools in the MEME Suite (https://meme-suite.org/meme/tools/ame). Transcription factor families were predicted using the PlantTFDB website (https://planttfdb.gao-lab.org/prediction.php). A gene regulatory network was built using the custom R script get_reg_file.R, and Cytoscape (version 3.10.2, https://cytoscape.org/) was used to visualize the network.

Isolation of total RNA and Quantitative real-time PCR analysis

To extract total RNA from papaya pistils, abortive pistils, and stamens, frozen tissues were ground in liquid nitrogen using a mortar and pestle. RNA extraction was performed using the RNeasy Plant Mini Kit (QIAGEN, Hilden, Germany) according to the manufacturer's instructions. Approximately 1 μg of total RNA was used for reverse transcription. First-strand cDNA synthesis was carried out using the HiScript® II 1st Strand cDNA Synthesis Kit (+ gDNA wiper) (Vazyme, Nanjing, China). Quantitative real-time PCR was performed with the TB Green® Premix Ex Taq™ II (Tli RNaseH Plus) kit (TaKaRa, Dalian, China), following the manufacturer's protocol. Gene expression was normalized using the $2^{-\Delta\Delta CT}$ method, with *ACTIN2* as the internal reference for papaya. Expression levels of 10 ABCDE genes, *CpAGL6*, and *CpTDF1*, a stamen-specific expression gene, were measured. The experiment was conducted in triplicate. Primer sequences for quantitative expression analysis are listed in Table S22.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-06242-1.

Supplementary Material 1.

Acknowledgements
Not applicable.

Authors' contributions
TX was responsible for the investigation and writing of the original draft. YTZ and YW contributed to the investigation. XC and ZBZ provided resources and conducted formal analysis. JL and PZ contributed resources. RM was involved in writing the revision. JJY handled writing review and editing, project administration, and supervision. All authors reviewed the manuscript.

Funding
This work was supported by the National Natural Science Foundation of China grants (32472697), Natural Science Foundation of Fujian Province of China (2024J09026), the "Eagle Plan" Young Talent Project in Fujian Province, the Foreign Cooperation Project of the Fujian Provincial Department of Science and Technology(2021I0009) and the Special fund for scientific and technological innovation of Fujian Agricultural and Forestry University (KFb22114XA).

Data availability
The mRNA sequencing data for floral organ development and sex type flowers in papaya have been submitted to the National Center for Biotechnology Information under accession number PRJNA1196855. Other data is provided within the manuscript or supplementary information files.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 10 December 2024 Accepted: 11 February 2025
Published online: 25 February 2025

References

1. Renner SS, Ricklefs RE. Dioecy and Its Correlates in the Flowering Plants. Am J Bot. 1995;82:596–606.
2. Renner SS. The relative and absolute frequencies of angiosperm sexual systems: Dioecy, monoecy, gynodioecy, and an updated online database. Am J Bot. 2014;101:1588–96.
3. Muyle A, Martin H, Zemp N, Mollion M, Gallina S, Tavares R, et al. Dioecy Is Associated with High Genetic Diversity and Adaptation Rates in the Plant Genus Silene. Mol Biol Evol. 2020;38:805–18.
4. Barrett SCH. The evolution of plant sexual diversity. Nat Rev Genet. 2002;3:274–84.
5. Henry IM, Akagi T, Tao R, Comai L. One Hundred Ways to Invent the Sexes: Theoretical and Observed Paths to Dioecy in Plants. Annu Rev Plant Biol. 2018;69:553–75.

6. Gschwend AR, Yu Q, Tong EJ, Zeng F, Han J, VanBuren R, et al. Rapid divergence and expansion of the X chromosome in papaya. Proc Natl Acad Sci. 2012;109:13716–21.
7. VanBuren R, Zeng F, Chen C, Zhang J, Wai CM, Han J, et al. Origin and domestication of papaya Y^h chromosome. Genome Res. 2015;25:524–33.
8. Ming R, Moore PH, editors. Genetics and Genomics of Papaya. New York, NY: Springer New York; 2014.
9. Aryal R, Ming R. Sex determination in flowering plants: Papaya as a model system. Plant Sci. 2014;217–218:56–62.
10. Ming R, Yu Q, Moore PH. Sex determination in papaya. Semin Cell Dev Biol. 2007;18:401–8.
11. Wang J, Na J-K, Yu Q, Gschwend AR, Han J, Zeng F, et al. Sequencing papaya X and Y^h chromosomes reveals molecular basis of incipient sex chromosome evolution. Proc Natl Acad Sci. 2012;109:13710–5.
12. Yu Q, Navajas-Pérez R, Tong E, Robertson J, Moore PH, Paterson AH, et al. Recent Origin of Dioecious and Gynodioecious Y Chromosomes in Papaya. Trop Plant Biol. 2008;1:49–57.
13. Ming R, Bendahmane A, Renner SS. Sex chromosomes in land plants. Annu Rev Plant Biol. 2011;62:485–514.
14. Bachtrog D, Mank JE, Peichel CL, Kirkpatrick M, Otto SP, Ashman T-L, et al. Sex determination: why so many ways of doing it? PLoS Biol. 2014;12: e1001899.
15. Bemis WP, Wilson GB. A NEW HYPOTHESIS EXPLAINING THE GENETICS OF SEX DETERMINATION. In Spinacia oleracea L J Hered. 1953;44:91–5.
16. Spigler RB, Lewers KS, Main DS, Ashman T-L. Genetic mapping of sex determination in a wild strawberry, *Fragaria virginiana*, reveals earliest form of sex chromosome. Heredity. 2008;101:507–17.
17. Zhang K, Li J, Lin T, Ma G, Zhang K, Li J, et al. A new regulation mechanism of bisexual flower development in cucurbitaceous sex determination. Trop Plants. 2023;2:17.
18. Arkle TDJ, Nakasone HY. Floral differentiation in the hermaphroditic Papaya. HortScience. 1984;19:832–4.
19. Alvarez-Buylla ER, Benítez M, Corvera-Poiré A, Chaos Cador Á, de Folter S,Gamboa de Buen A, et al. Flower Development Arab Book. 2010;8: e0127.
20. Li L, Ma H, Xing J, Liu F, Wang Y. Effects of water deficit and nitrogen application on leaf gas exchange, phytohormone signaling, biomass and water use efficiency of oat plants. J Plant Nutr Soil Sci. 2020;183:695–704.
21. Coen ES, Meyerowitz EM. The war of the whorls: genetic interactions controlling flower development. Nature. 1991;353:31–7.
22. Bowman JL. Evolutionary conservation of angiosperm flower development at the molecular and genetic levels. J Biosci. 1997;22:515–27.
23. Li G, Meng Z, Kong H, Chen Z, Lu A. ABC model and floral evolution. Chin Sci Bull. 2003;48:2651–7.
24. Angenent GC, Franken J, Busscher M, van Dijken A, van Went JL, Dons HJ, et al. A novel class of MADS box genes is involved in ovule development in petunia. Plant Cell. 1995;7:1569–82.
25. Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF. B and C floral organ identity functions require *SEPALLATA* MADS-box genes. Nature. 2000;405:200–3.
26. Ditta G, Pinyopich A, Robles P, Pelaz S, Yanofsky MF. The *SEP4* gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. Curr Biol CB. 2004;14:1935–40.
27. Ma H, Yanofsky MF, Meyerowitz EM. AGL1-AGL6, an *Arabidopsis* gene family with similarity to floral homeotic and transcription factor genes. Genes Dev. 1991;5:484–95.
28. Pnueli L, Abu-Abeid M, Zamir D, Nacken W, Schwarz-Sommer Z, Lifschitz E. The MADS box gene family in tomato: temporal expression during floral development, conserved secondary structures and homology with homeotic genes from *Antirrhinum* and *Arabidopsis*. Plant J Cell Mol Biol. 1991;1:255–66.
29. Passmore S, Elble R, Tye BK. A protein involved in minichromosome maintenance in yeast binds a transcriptional enhancer conserved in eukaryotes. Genes Dev. 1989;3:921–35.
30. Yanofsky MF, Ma H, Bowman JL, Drews GN, Feldmann KA, Meyerowitz EM. The protein encoded by the *Arabidopsis* homeotic gene agamous resembles transcription factors. Nature. 1990;346:35–9.
31. Sommer H, Beltrán JP, Huijser P, Pape H, Lönnig WE, Saedler H, et al. Deficiens, a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus*: the protein shows homology to transcription factors. EMBO J. 1990;9:605–13.
32. Norman C, Runswick M, Pollock R, Treisman R. Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element. Cell. 1988;55:989–1003.
33. Purugganan MD. The MADS-box floral homeotic gene lineages predate the origin of seed plants: Phylogenetic and molecular clock estimates. J Mol Evol. 1997;45:392–6.
34. Becker A, Theissen G. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. Mol Phylogenet Evol. 2003;29:464–89.
35. Irish V. The ABC model of floral development. Curr Biol CB. 2017;27:R887–90.
36. Pařenicová L, de Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, et al. Molecular and Phylogenetic Analyses of the Complete MADS-Box Transcription Factor Family in *Arabidopsis*. Plant Cell. 2003;15:1538–51.
37. Englund M, Carlsbecker A, Engström P, Vergara-Silva F. Morphological, "primary homology" and expression of *AG*-subfamily MADS-box genes in pines, podocarps, and yews. Evol Dev. 2011;13:171–81.
38. Li H, Liang W, Hu Y, Zhu L, Yin C, Xu J, et al. Rice MADS6 Interacts with the Floral Homeotic Genes *SUPERWOMAN1*, *MADS3*, *MADS58*, *MADS13*, and *DROOPING LEAF* in Specifying Floral Organ Identities and Meristem Fate. Plant Cell. 2011;23:2536–52.
39. Amborella Genome Project. The *Amborella* genome and the evolution of flowering plants. Science. 2013;342:1241089.
40. Zhang L, Chen F, Zhang X, Li Z, Zhao Y, Lohaus R, et al. The water lily genome and the early evolution of flowering plants. Nature. 2020;577:79–84.
41. Liu Y, Wang S, Li L, Yang T, Dong S, Wei T, et al. The Cycas genome and the early evolution of seed plants. Nat Plants. 2022;8:389–401.
42. Henschel K, Kofuji R, Hasebe M, Saedler H, Münster T, Theissen G. Two ancient classes of MIKC-type MADS-box genes are present in the moss Physcomitrella patens. Mol Biol Evol. 2002;19:801–14.
43. De Bodt S, Raes J, Van de Peer Y, Theissen G. And then there were many: MADS goes genomic. Trends Plant Sci. 2003;8:475–83.
44. Gramzow L, Ritz MS, Theissen G. On the origin of MADS-domain transcription factors. Trends Genet TIG. 2010;26:149–53.
45. Bowman JL, Smyth DR, Meyerowitz EM. The ABC model of flower development: then and now. Development. 2012;139:4095–8.
46. Yu Y, Chu X, Ma X, Hu Z, Wang M, Li J, et al. Genome-Wide Analysis of MADS-Box Gene Family Reveals *CjSTK* as a Key Regulator of Seed Abortion in Camellia japonica. Int J Mol Sci. 2024;25:5770.
47. Gramzow L, Theissen G. A hitchhiker's guide to the MADS world of plants. Genome Biol. 2010;11:214.
48. Qiu Y, Li Z, Köhler C. Ancestral duplication of MADS-box genes in land plants empowered the functional divergence between sporophytes and gametophytes. New Phytol. 2024;244:358–63.
49. Shen G, Yang C-H, Shen C-Y, Huang K-S. Origination and selection of ABCDE and AGL6 subfamily MADS-box genes in gymnosperms and angiosperms. Biol Res. 2019;52:1–15.
50. Honma T, Goto K. Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. Nature. 2001;409:525–9.
51. Theißen G, Saedler H. Floral quartets. Nature. 2001;409:469–71.
52. West AG, Causier BE, Davies B, Sharrocks AD. DNA binding and dimerisation determinants of Antirrhinum majus MADS-box transcription factors. Nucleic Acids Res. 1998;26:5277–87.
53. Egea-Cortines M, Saedler H, Sommer H. Ternary complex formation between the MADS-box proteins *SQUAMOSA*, *DEFICIENS* and *GLOBOSA* is involved in the control of floral architecture in Antirrhinum majus. EMBO J. 1999;18:5370–9.
54. West AG, Sharrocks AD. MADS-box transcription factors adopt alternative mechanisms for bending DNA. J Mol Biol. 1999;286:1311–23.
55. Wu D, Liang W, Zhu W, Chen M, Ferrándiz C, Burton RA, et al. Loss of *LOFSEP* Transcription Factor Function Converts Spikelet to Leaf-Like Structures in Rice. Plant Physiol. 2018;176:1646–64.
56. Grimplet J, Martínez-Zapater JM, Carmona MJ. Structural and functional annotation of the MADS-box transcription factor family in grapevine. BMC Genomics. 2016;17:80.
57. Arora R, Agarwal P, Ray S, Singh AK, Singh VP, Tyagi AK, et al. MADS-box gene family in rice: genome-wide identification, organization and

expression profiling during reproductive development and stress. BMC Genomics. 2007;8:242.
58. Meng D, Cao Y, Chen T, Abdullah M, Jin Q, Fan H, et al. Evolution and functional divergence of MADS-box genes in Pyrus. Sci Rep. 2019;9:1266.
59. Yu Q, Steiger D, Kramer EM, Moore PH, Ming R. Floral MADS-box Genes in Trioecious Papaya: Characterization of *AG* and *AP1* Subfamily Genes Revealed a Sex-type-specific Gene. Trop Plant Biol. 2008;1:97–107.
60. Gramzow L, Theißen G. Phylogenomics of MADS-Box Genes in Plants — Two Opposing Life Styles in One Gene Family. Biology. 2013;2:1150–64.
61. Dai Y, Wang Y, Zeng L, Jia R, He L, Huang X, et al. Genomic and Transcriptomic Insights into the Evolution and Divergence of MIKC-Type MADS-Box Genes in Carica papaya. Int J Mol Sci. 2023;24:14039.
62. Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, et al. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). Nature. 2008;452:991–6.
63. Yue J, VanBuren R, Liu J, Fang J, Zhang X, Liao Z, et al. SunUp and Sunset genomes revealed impact of particle bombardment mediated transformation and domestication history in papaya. Nat Genet. 2022;54:715–24.
64. Pinyopich A, Ditta GS, Savidge B, Liljegren SJ, Baumann E, Wisman E, et al. Assessing the redundancy of MADS-box genes during carpel and ovule development. Nature. 2003;424:85–8.
65. Petrella R, Caselli F, Roig-Villanova I, Vignati V, Chiara M, Ezquer I, et al. BPC transcription factors and a Polycomb Group protein confine the expression of the ovule identity gene *SEEDSTICK* in *Arabidopsis*. Plant J. 2020;102:582–99.
66. Nam J, dePamphilis CW, Ma H, Nei M. Antiquity and evolution of the MADS-box gene family controlling flower development in plants. Mol Biol Evol. 2003;20:1435–47.
67. Bowman JL, Moyroud E. Reflections on the ABC model of flower development. Plant Cell. 2024;36:1334–57.
68. Dreni L, Ferrándiz C. Tracing the Evolution of the *SEPALLATA* Subfamily across Angiosperms Associated with Neo- and Sub-Functionalization for Reproductive and Agronomically Relevant Traits. Plants. 2022;11:2934.
69. Jetha K, Theißen G, Melzer R. *Arabidopsis* SEPALLATA proteins differ in cooperative DNA-binding during the formation of floral quartet-like complexes. Nucleic Acids Res. 2014;42:10927–42.
70. Kim S, Koh J, Yoo M, Kong H, Hu Y, Ma H, et al. Expression of floral MADS-box genes in basal angiosperms: implications for the evolution of floral regulators. Plant J. 2005;43:724–44.
71. Katahata S-I, Futamura N, Igasaki T, Shinohara K. Functional analysis of *SOC1-like* and *AGL6-like* MADS-box genes of the gymnosperm Cryptomeria japonica. Tree Genet Genomes. 2014;10:317–27.
72. de Azevedo SC, Kim SS, Koch S, Kienow L, Schneider K, McKim SM, et al. A Novel Fatty Acyl-CoA Synthetase Is Required for Pollen Development and Sporopollenin Biosynthesis in *Arabidopsis*. Plant Cell. 2009;21:507–25.
73. Tsugama D, Fujino K, Liu S, Takano T. A GDSL-type esterase/lipase gene, *GELP77*, is necessary for pollen dissociation and fertility in *Arabidopsis*. Biochem Biophys Res Commun. 2020;526:1036–41.
74. Kobayashi K, Akita K, Suzuki M, Ohta D, Nagata N. Fertile *Arabidopsis cyp704b1* mutant, defective in sporopollenin biosynthesis, has a normal pollen coat and lipidic organelles in the tapetum. Plant Biotechnol Tokyo Jpn. 2021;38:109–16.
75. Tsai W-C, Pan Z-J, Hsiao Y-Y, Chen L-J, Liu Z-J. Evolution and function of MADS-box genes involved in orchid floral development. J Syst Evol. 2014;52:397–410.
76. Dinkins RD, Pflipsen C, Collins GB. Expression and deletion analysis of an *Arabidopsis SUPERMAN-like* zinc finger gene. Plant Sci. 2003;165:33–41.
77. Liljegren SJ, Roeder AHK, Kempin SA, Gremski K, Østergaard L, Guimil S, et al. Control of fruit patterning in *Arabidopsis* by *INDEHISCENT*. Cell. 2004;116:843–53.
78. Girin T, Paicu T, Stephenson P, Fuentes S, Körner E, O'Brien M, et al. *INDEHISCENT* and *SPATULA* interact to specify carpel and valve margin tissue and thus promote seed dispersal in *Arabidopsis*. Plant Cell. 2011;23:3641–53.
79. Hugouvieux V, Blanc-Mathieu R, Janeau A, Paul M, Lucas J, Xu X, et al. *SEPALLATA*-driven MADS transcription factor tetramerization is required for inner whorl floral organ development. Plant Cell. 2024;36:3435–50.
80. Zhu J, Chen H, Li H, Gao J-F, Jiang H, Wang C, et al. Defective in Tapetal Development and Function 1 is essential for anther development and tapetal function for microspore maturation in *Arabidopsis*. Plant J. 2008;55:266–77.
81. Lai Z, Wang J, Peng S-Q, Chang F. bHLH010/089 Transcription Factors Control Pollen Wall Development via Specific Transcriptional and Metabolic Networks in *Arabidopsis thaliana*. Int J Mol Sci. 2022;23:11683.
82. Ferguson AC, Pearce S, Band LR, Yang C, Ferjentsikova I, King J, et al. Biphasic regulation of the transcription factor ABORTED MICROSPORES (AMS) is essential for tapetum and pollen development in *Arabidopsis*. New Phytol. 2017;213:778–90.
83. Lou Y, Zhou H-S, Han Y, Zeng Q-Y, Zhu J, Yang Z-N. Positive regulation of *AMS* by TDF1 and the formation of a TDF1–AMS complex are required for anther development in *Arabidopsis thaliana*. New Phytol. 2018;217:378–91.
84. Ghelli R, Brunetti P, Marzi D, Cecchetti V, Costantini M, Lanzoni-Rossi M, et al. The full-length Auxin Response Factor 8 isoform ARF8.1 controls pollen cell wall formation and directly regulates *TDF1*, *AMS* and *MS188* expression. Plant J. 2023;113:851–65.
85. Zhou P, Fatima M, Ma X, Liu J, Ming R. Auxin regulation involved in gynoecium morphogenesis of papaya flowers. Hortic Res. 2019;6:119.
86. Fischer A, Baum N, Saedler H, Theiβen G. Chromosomal mapping of the MADS-box multigene family in Zea mays reveals dispersed distribution of allelic genes as well as transposed copies. Nucleic Acids Res. 1995;23:1901–11.
87. Carlsbecker A, Sundström JF, Englund M, Uddenberg D, Izquierdo L, Kvarnheden A, et al. Molecular control of normal and acrocona mutant seed cone development in Norway spruce (*Picea abies*) and the evolution of conifer ovule-bearing organs. New Phytol. 2013;200:261–75.
88. Gramzow L, Weilandt L, Theißen G. MADS goes genomic in conifers: towards determining the ancestral set of MADS-box genes in seed plants. Ann Bot. 2014;114:1407–29.
89. Lan J, Wang N, Wang Y, Jiang Y, Yu H, Cao X, et al. *Arabidopsis TCP4* transcription factor inhibits high temperature-induced homeotic conversion of ovules. Nat Commun. 2023;14:5673.
90. Simonini S, Roig-Villanova I, Gregis V, Colombo B, Colombo L, Kater MM. BASIC PENTACYSTEINE Proteins Mediate MADS Domain Complex Binding to the DNA for Tissue-Specific Expression of Target Genes in *Arabidopsis*. Plant Cell. 2012;24:4163–72.
91. Gregis V, Andrés F, Sessa A, Guerra RF, Simonini S, Mateos JL, et al. Identification of pathways directly regulated by *SHORT VEGETATIVE PHASE* during vegetative and reproductive development in *Arabidopsis*. Genome Biol. 2013;14:R56.
92. Lee C-Y, Lin H-J, Viswanath KK, Lin C-P, Chang BC-H, Chiu P-H, et al. The development of functional mapping by three sex-related loci on the third whorl of different sex types of *Carica papaya* L. PloS One. 2018;13:e0194605.
93. Chae T, Harkess A, Moore RC. Sex-linked gene expression and the emergence of hermaphrodites in *Carica papaya*. Am J Bot. 2021;108:1029–41.
94. Sakai H, Medrano LJ, Meyerowitz EM. Role of *SUPERMAN* in maintaining *Arabidopsis* floral whorl boundaries. Nature. 1995;378:199–203.
95. Prunet N, Yang W, Das P, Meyerowitz EM, Jack TP. *SUPERMAN* prevents class B gene expression and promotes stem cell termination in the fourth whorl of *Arabidopsis* thaliana flowers. Proc Natl Acad Sci U S A. 2017;114:7166–71.
96. Xu Y, Prunet N, Gan E, Wang Y, Stewart D, Wellmer F, et al. *SUPERMAN* regulates floral whorl boundaries through control of auxin biosynthesis. EMBO J. 2018;37: e97499.
97. Ballester P, Martínez-Godoy MA, Ezquerro M, Navarrete-Gómez M, Trigueros M, Rodríguez-Concepción M, et al. A transcriptional complex of *NGATHA* and bHLH transcription factors directs stigma development in *Arabidopsis*. Plant Cell. 2021;33:3645–57.
98. Liu T, Longhurst AD, Talavera-Rauh F, Hokin SA, Barton MK. The *Arabidopsis* transcription factor *ABIG1* relays ABA signaled growth inhibition and drought induced senescence. eLife. 2016;5:e13768.
99. Lynch TJ, Erickson BJ, Miller DR, Finkelstein RR. ABI5-binding proteins (AFPs) alter transcription of ABA-induced genes via a variety of interactions with chromatin modifiers. Plant Mol Biol. 2017;93:403–18.

100. Jiang Y-T, Zheng J-X, Li R-H, Wang Y-C, Shi J, Ferjani A, et al. Tonoplast proton pumps regulate nuclear spacing of female gametophytes via mediating polar auxin transport in *Arabidopsis*. Front Plant Sci. 2022;13:1006735.
101. Rinaldi MA, Liu J, Enders TA, Bartel B, Strader LC. A gain-of-function mutation in *IAA16* confers reduced responses to auxin and abscisic acid and impedes plant growth and fertility. Plant Mol Biol. 2012;79:359–73.
102. Cao D, Cheng H, Wu W, Soo HM, Peng J. Gibberellin Mobilizes Distinct DELLA-Dependent Transcriptomes to Regulate Seed Germination and Floral Development in *Arabidopsis*. Plant Physiol. 2006;142:509–25.
103. Choi H, Ohyama K, Kim Y-Y, Jin J-Y, Lee SB, Yamaoka Y, et al. The Role of *Arabidopsis ABCG9* and *ABCG31* ATP Binding Cassette Transporters in Pollen Fitness and the Deposition of Steryl Glycosides on the Pollen Coat. Plant Cell. 2014;26:310–24.

Publisher's Note