The pineapple MADS-box gene family and the evolution of early monocot flower

Juan Hu1, Xiaojun Chang1, Ying Zhang1, Xianxian Yu2, Yuan Qin1, Yun Sun2 & Liangsheng Zhang1,3*

Unlike the flower of the model monocot rice, which has diverged greatly from the ancestral monocot flower, the pineapple (Ananas comosus) flower is more typical of monocot flowers. Here, we identified 43 pineapple genes containing MADS-box domains, including 11 type I and 32 type II genes. RNA-seq expression data generated from five pineapple floral organs (sepals, petals, stamens, pistils, and ovules) and quantitative real-time PCR revealed tissue-specific expression patterns for some genes.

We found that AcAGL6 and AcFUL1 were highly expressed at different stages of fruit development and have similar expression profiles, implicating these genes' role in fruit development and ripening processes.

We propose that the pineapple flower can be used as a model for studying the ancestral form of monocot flowers to investigate their development and evolutionary history.

Typical flowers are composed of sepals, petals, stamens, and pistils arranged in concentric rings around a central axis, and the molecular and evolutionary mechanisms of floral development have been studied extensively. The genetic 'ABC' model of floral organ development was originally based on the functional analysis of genes in Arabidopsis thaliana and Antirrhinum majus then subsequently extended to many other flowering plants. In Arabidopsis, A function is performed by APETALA1 (AP1) and APETALA2 (AP2), B function by APETALA3 (AP3) and PISTILLATA (PI), C function by AGAMOUS (AG), D function by SEEDSTICK (STK), also called AGAMOUS-LIKE11 (AGL11) and SHATTERPROOF (SHP), and E function by SEPALLATA1/2/3/4 (SEP1/2/3/4), also known as AGAMOUS-LIKE2/4/9/3 (AGL2/4/9/3). According to the ABC model, floral organs in each of the four whorls are specified by three classes of genes with different homeotic functions acting in a combinatorial way. Specifically, A-function genes alone specify sepals; A and B petals; B and C stamens; and C alone carpels. Moreover, A and C genes function antagonistically. The classic ABC model was modified by adding D and E functions, leading to the ABCDE model. D-function genes specify ovule identity, and E-function genes contribute to the development of all four whorls by interacting with A-, B-, C-, and D-class genes in the form of heterotetramer protein complexes.

Monocot flower morphology and the underlying developmental mechanisms have notable similarities and differences with those of early-diverging angiosperms and eudicots. The four predicted A-class genes (AP1/FUL-like subfamily) in rice (Oryza sativa) are OsMADS14, OsMADS15, OsMADS18, and OsMADS20. OsMADS14 and OsMADS15 double mutants exhibit true recessive phenotypes with defects in the first and second whorls, consistent with the phenotypes of A-function mutants in Arabidopsis. OsMADS2, OsMADS4, and OsMADS16/SPW1 are B-function genes and primarily determine the identity of lodicules and stamens in rice. OsMADS16/SPW1 is homologous to the AP3 gene of A. thaliana and has a similar expression domain and analogous function. The PI homologs OsMADS2 and OsMADS4 diverge in terms of B function. OsMADS2 is specialized for the development of lodicules and does not affect stamen morphogenesis. In contrast, OsMADS4 maintains the ancestral gene function of the B-class PI lineage, and suppression of its expression leads to ectopic development of palea.

1 Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology, College of Horticulture, Fujian Agriculture and Forestry University, Fuzhou 350002, China. 2 School of Urban-Rural Planning and Landscape Architecture, Xuchang University, Xuchang 461000, China. 3 Genomics and Genetic Engineering Laboratory of Ornamental Plants, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310058, China.

*email: zls83@zju.edu.cn
Lemma-like organs and carpels in the second and third whorls. OsMADS3 and OsMADS58 are homologous to the Arabidopsis AG gene, and these two copies show partial sub- and neo-functionalization. OsMADS3 is primarily involved in stamen and ovule development, and a deletion mutant was found to have normal carpel development. A knockdown mutation in the OsMADS58 gene led to disrupted floral meristem determinacy rather than floral organ defects. In addition to floral determinacy defects, the silencing of both OsMADS3 and OsMADS58 resulted in homeotic conversion of female and male reproductive organs into palea/lemma-like organs and lodicules, respectively, similar to the phenotypes of the ag mutant in Arabidopsis. Two D-function genes, OsMADS13 and OsMADS21, underwent functional diversification. While OsMADS13 is primarily involved in ovule formation, OsMADS21 lost its role in this process. The rice genome contains five SEP-like subfamily homologs: OsMADS1, OsMADS5, OsMADS7, OsMADS8, and OsMADS34. Among them, OsMADS1 and OsMADS34 contribute to the specification of the four whorls of floral organs and control the determination of spikelet meristems.

The expression and function of the ABCDE MADS-box genes are known to have helped establish the unique floral architecture of rice and orchid. In contrast, the genetic basis of pineapple flower development and the evolutionary history of the molecular mechanisms in ancestral monocot flowers remain poorly understood. Therefore, a comprehensive study of this gene family, especially the phylogeny and roles of MADS-box genes in flower development, is urgently needed for pineapple. To classify the MADS-box genes in pineapple and elucidate their evolutionary relationships, we utilized the pineapple genome as a reference for a systematic phylogenetic analysis. We also studied the expression patterns of pineapple MADS-box genes at different stages of flower and fruit development. We identified candidate floral ABCDE genes in pineapple, and the gene numbers and expression patterns could explain the conserved floral architecture of pineapple flower to a large extent.

Results
Pineapple flower and the ancestral monocot flower. Pineapple, orchid, and rice vary vastly in terms of flower morphology. Pineapple, a perennial monotoc of the Bromeliaceae family, is indigenous to Central and South America. Each pineapple flower, from the outside inwards, is composed of three broadly ovate and fleshy sepals in whorl 1, three long elliptic petals in whorl 2, six stamens in whorl 3, and a pistil in whorl 4 with three fused carpels (Fig. 1A). Orchids have unique labellum and gynostemium. The grass species, including rice, have the unique floral organization and morphology of florets, which comprise grass-specific peripheral organs, including a pair of bract-like organs (lemma and palea), two lodicules, and conserved sexual organs (six stamens and a pistil with a single carpel) (Fig. 1B). Conversely, individual pineapple flowers resemble ancestral monocot floral organs and carpels in the second and third whorls. OsMADS3 and OsMADS58 are homologous to the Arabidopsis AG gene, and these two copies show partial sub- and neo-functionalization. OsMADS3 is primarily involved in stamen and ovule development, and a deletion mutant was found to have normal carpel development. A knockdown mutation in the OsMADS58 gene led to disrupted floral meristem determinacy rather than floral organ defects. In addition to floral determinacy defects, the silencing of both OsMADS3 and OsMADS58 resulted in homeotic conversion of female and male reproductive organs into palea/lemma-like organs and lodicules, respectively, similar to the phenotypes of the ag mutant in Arabidopsis. Two D-function genes, OsMADS13 and OsMADS21, underwent functional diversification. While OsMADS13 is primarily involved in ovule formation, OsMADS21 lost its role in this process. The rice genome contains five SEP-like subfamily homologs: OsMADS1, OsMADS5, OsMADS7, OsMADS8, and OsMADS34. Among them, OsMADS1 and OsMADS34 contribute to the specification of the four whorls of floral organs and control the determination of spikelet meristems.

The expression and function of the ABCDE MADS-box genes are known to have helped establish the unique floral architecture of rice and orchid. In contrast, the genetic basis of pineapple flower development and the evolutionary history of the molecular mechanisms in ancestral monocot flowers remain poorly understood. Therefore, a comprehensive study of this gene family, especially the phylogeny and roles of MADS-box genes in flower development, is urgently needed for pineapple. To classify the MADS-box genes in pineapple and elucidate their evolutionary relationships, we utilized the pineapple genome as a reference for a systematic phylogenetic analysis. We also studied the expression patterns of pineapple MADS-box genes at different stages of flower and fruit development. We identified candidate floral ABCDE genes in pineapple, and the gene numbers and expression patterns could explain the conserved floral architecture of pineapple flower to a large extend.

Results
Pineapple flower and the ancestral monocot flower. Pineapple, orchid, and rice vary vastly in terms of flower morphology. Pineapple, a perennial monotoc of the Bromeliaceae family, is indigenous to Central and South America. Each pineapple flower, from the outside inwards, is composed of three broadly ovate and fleshy sepals in whorl 1, three long elliptic petals in whorl 2, six stamens in whorl 3, and a pistil in whorl 4 with three fused carpels (Fig. 1A). Orchids have unique labellum and gynostemium. The grass species, including rice, have the unique floral organization and morphology of florets, which comprise grass-specific peripheral organs, including a pair of bract-like organs (lemma and palea), two lodicules, and conserved sexual organs (six stamens and a pistil with a single carpel) (Fig. 1B). Conversely, individual pineapple flowers resemble ancestral monocot floral organs and carpels in the second and third whorls. OsMADS3 and OsMADS58 are homologous to the Arabidopsis AG gene, and these two copies show partial sub- and neo-functionalization. OsMADS3 is primarily involved in stamen and ovule development, and a deletion mutant was found to have normal carpel development. A knockdown mutation in the OsMADS58 gene led to disrupted floral meristem determinacy rather than floral organ defects. In addition to floral determinacy defects, the silencing of both OsMADS3 and OsMADS58 resulted in homeotic conversion of female and male reproductive organs into palea/lemma-like organs and lodicules, respectively, similar to the phenotypes of the ag mutant in Arabidopsis. Two D-function genes, OsMADS13 and OsMADS21, underwent functional diversification. While OsMADS13 is primarily involved in ovule formation, OsMADS21 lost its role in this process. The rice genome contains five SEP-like subfamily homologs: OsMADS1, OsMADS5, OsMADS7, OsMADS8, and OsMADS34. Among them, OsMADS1 and OsMADS34 contribute to the specification of the four whorls of floral organs and control the determination of spikelet meristems.

The expression and function of the ABCDE MADS-box genes are known to have helped establish the unique floral architecture of rice and orchid. In contrast, the genetic basis of pineapple flower development and the evolutionary history of the molecular mechanisms in ancestral monocot flowers remain poorly understood. Therefore, a comprehensive study of this gene family, especially the phylogeny and roles of MADS-box genes in flower development, is urgently needed for pineapple. To classify the MADS-box genes in pineapple and elucidate their evolutionary relationships, we utilized the pineapple genome as a reference for a systematic phylogenetic analysis. We also studied the expression patterns of pineapple MADS-box genes at different stages of flower and fruit development. We identified candidate floral ABCDE genes in pineapple, and the gene numbers and expression patterns could explain the conserved floral architecture of pineapple flower to a large extend.

Results
Pineapple flower and the ancestral monocot flower. Pineapple, orchid, and rice vary vastly in terms of flower morphology. Pineapple, a perennial monotoc of the Bromeliaceae family, is indigenous to Central and South America. Each pineapple flower, from the outside inwards, is composed of three broadly ovate and fleshy sepals in whorl 1, three long elliptic petals in whorl 2, six stamens in whorl 3, and a pistil in whorl 4 with three fused carpels (Fig. 1A). Orchids have unique labellum and gynostemium. The grass species, including rice, have the unique floral organization and morphology of florets, which comprise grass-specific peripheral organs, including a pair of bract-like organs (lemma and palea), two lodicules, and conserved sexual organs (six stamens and a pistil with a single carpel) (Fig. 1B). Conversely, individual pineapple flowers resemble ancestral monocot floral organs and carpels in the second and third whorls. OsMADS3 and OsMADS58 are homologous to the Arabidopsis AG gene, and these two copies show partial sub- and neo-functionalization. OsMADS3 is primarily involved in stamen and ovule development, and a deletion mutant was found to have normal carpel development. A knockdown mutation in the OsMADS58 gene led to disrupted floral meristem determinacy rather than floral organ defects. In addition to floral determinacy defects, the silencing of both OsMADS3 and OsMADS58 resulted in homeotic conversion of female and male reproductive organs into palea/lemma-like organs and lodicules, respectively, similar to the phenotypes of the ag mutant in Arabidopsis. Two D-function genes, OsMADS13 and OsMADS21, underwent functional diversification. While OsMADS13 is primarily involved in ovule formation, OsMADS21 lost its role in this process. The rice genome contains five SEP-like subfamily homologs: OsMADS1, OsMADS5, OsMADS7, OsMADS8, and OsMADS34. Among them, OsMADS1 and OsMADS34 contribute to the specification of the four whorls of floral organs and control the determination of spikelet meristems.

The expression and function of the ABCDE MADS-box genes are known to have helped establish the unique floral architecture of rice and orchid. In contrast, the genetic basis of pineapple flower development and the evolutionary history of the molecular mechanisms in ancestral monocot flowers remain poorly understood. Therefore, a comprehensive study of this gene family, especially the phylogeny and roles of MADS-box genes in flower development, is urgently needed for pineapple. To classify the MADS-box genes in pineapple and elucidate their evolutionary relationships, we utilized the pineapple genome as a reference for a systematic phylogenetic analysis. We also studied the expression patterns of pineapple MADS-box genes at different stages of flower and fruit development. We identified candidate floral ABCDE genes in pineapple, and the gene numbers and expression patterns could explain the conserved floral architecture of pineapple flower to a large extend.
flowers in their organization and arrangement, except that the ancestral flower has two outer whorls and is considered to have undifferentiated perianth organs rather than sepals and petals\(^{21–23}\) (Fig. 1B). Thus, unlike rice and orchids, which has diverged significantly from the ancestral flower, the pineapple flower is typical of monocot flowers and more closely resembles an ancestral state.

The identification of pineapple MADS-box genes. To identify the MADS-box genes in pineapple, we employed the following strategy: a profile hidden Markov model (HMMER) search against the pineapple genome protein database using the SRF-TF domain (PF00319) as a query dialog. Redundant and extremely short sequences were removed, and conserved MADS domains were detected by verifying sequences in the Pfam databases. Forty-four candidate pineapple MADS-box proteins with a complete M-domain were obtained and designated AcMADS1 to AcMADS43 (Tables 1, Supplementary Table S1). To clarify the evolutionary relationships of MADS-box genes in various plant species, MADS-box genes from ten other plant species were detected using the same method described above for pineapple (Fig. 2). We found that the 43-member MADS-box gene family of pineapple was relatively small compared to those of other species. The pineapple MADS-box genes were subsequently designated as type I (11) and type II (32) MADS-box genes, according to the phylogenetic relationship with well-known MADS-box gene members in *Arabidopsis*.

Phylogenetic analysis of pineapple MADS-box genes. MADS-box genes can be assigned into two phylogenetically distinct groups: type I and type II\(^8\). To further assess the phylogenetic relationship of pineapple MADS-box genes with those in other species and to assign them to a specific subfamily, two phylogenetic trees were constructed for type I and type II (Fig. 3, Supplementary Fig. S1). Tree construction was based on multiple sequence alignments of MADS-box protein sequences from pineapple and five additional species including basal angiosperm, core eudicot, and monocot species. For the type I group, the 11 pineapple genes were divided into the Mα, Mβ, and Mγ subgroups, with six, two, and three members, respectively, which is similar to other species\(^6,29\). The type II group are classified into the MIKCC and MIKC* subclades, based on structure divergence at the I domain\(^30\). The pineapple MIKCC proteins further were divided into 13 subfamilies (AGL6, SEP1/AGL2/3/4, SEP3/AGL9, AGL6) (Fig. 3, Supplementary Fig. S2). Apart from the AGL15 clade, pineapple MIKC\(^C\) genes were present in 12 of the 13 clades with their counterparts in *Arabidopsis* and rice, indicating that the genes in the AGL15 clade were lost in pineapple. SOC1 and ANR1/AGL17 were the largest clades, both with five members, whereas AGL6 and OSMADS32-like each had only one member. Compared with the MIKC\(^C\) gene numbers in rice and orchid, pineapple had fewer members in the AP1, B-PI, Bs, AG, AGL6, and SEP clades. In contrast, significant expansion was observed in the SOC1, ANR1, AGL11, and SVP clades (Fig. 3).

| Subfamily | Pineapple Gene name | Gene ID | Chromosome location | Rice Gene name | Gene ID |
|-----------|---------------------|---------|---------------------|----------------|---------|
| AP1/FUL   | AcFUL1              | Aco004839.1 | LG07 217862–228880 | OsMADS14       | LOC_Os03g54160.1 |
|           | AcFUL2              | Aco012428.1 | LG01 2127109–2132551 | OsMADS15       | LOC_Os07g01820.1 |
|           |                     |          |                     | OsMADS18       | LOC_Os07g11370.1 |
|           |                     |          |                     | OsMADS20       | LOC_Os12g31748.1 |
| AP3       | AcAP3               | Aco017598.1 | LG09 328289–335839 | OsMADS16       | LOC_Os06g49840.1 |
| PI        | AcPI                | Aco019365.1 | LG05 6423630–6427189 | OsMADS2        | LOC_Os01g66030.1 |
| Bis (Bs)  | AcBS                | Aco008539.1 | LG19 9790787–9794944 | OsMADS29       | LOC_Os02g07430.1 |
|           |                     |          |                     | OsMADS30       | LOC_Os06g45650.1 |
| AG        | AcAG                | Aco009993.1 | LG10 2019062–2029924 | OsMADS31       | LOC_Os04g52410.1 |
| AGL11     | AcAGL11a            | Aco004785.1 | LG05 578956–5790607 | OsMADS13       | LOC_Os12g10540.1 |
|           | AcAGL11b            | Aco007999.1 | LG21 8935775–8941681 | OsMADS13       | LOC_Os01g6290.1 |
|           | AcAGL11c            | Aco011341.1 | LG01 12274322–12278849 | OsMADS14       | LOC_Os03g16141.1 |
| SEP1/AGL2/3/4 | AcSEP1        | Aco017563.1 | LG09 90797–101510 | OsMADS5        | LOC_Os06g06750.1 |
|           |                     |          |                     | OsMADS34       | LOC_Os03g54170.1 |
| SEP3/AGL9 | AcSEP3              | Aco015105.1 | LG01 24291803–24304406 | OsMADS7       | LOC_Os08g19550.1 |
|           |                     |          |                     | OsMADS8        | LOC_Os09g2948.1 |
| AGL6      | AcAGL6              | Aco015487.1 | LG03 12961737–12978045 | OsMADS6       | LOC_Os02g5770.1 |
|           |                     |          |                     | OsMADS17       | LOC_Os04g49150.1 |

Table 1. List of the ABCDE genes in the pineapple and rice genomes.
Conserved ABCDE genes in pineapple. We conducted more detailed phylogenetic analyses to identify pineapple genes homologous to *Arabidopsis* and rice genes involved in the ABCDE model of floral development. For A-class genes, the monocot AP1/FUL-like group can be subdivided into two main subclades: FUL-like I and FUL-like II. The two AP1/FUL-like members from pineapple were evenly assigned to these two subclades, with *AcFUL1* closely related to the rice FUL-like I genes *OsMADS15* and *OsMADS14*, and *AcFUL2* similar to the rice FUL-like II genes *OsMADS18* and *OsMADS20* (Fig. 4A). The three B-class candidates were designated as *AcAP3* and *AcPI*. Phylogenetic analysis showed that *AcAP3* belonged to the B-AP3 clade, and only one member, *AcPI*, corresponded to the B-PI clade (Fig. 4B). Because most monocots have only one B-AP3 member, except for orchid with four copies (Fig. 4B). *AcBS*, the B-sister homolog in pineapple, formed a sister group with the genes from the AP3/PI clade.

In pineapple, the AG-like subfamily had four members, one of which (*AcAG*) consistently grouped into the C lineage (AG), while the others (*AcAGL11a-c*) were associated with D function (STK/AGL11) (Fig. 4C). According to the sequence and phylogenetic analyses, *AcAG* formed a monophyletic lineage along with two rice MADS-box genes (*OsMADS3* and *OsMADS8*), whereas *AcAGL11a*, *AcAGL11b*, and *AcAGL11c* formed another monophyletic lineage along with the D-lineage MADS-box genes from other monocot species (e.g., *OsMADS13* and *OsMADS21*). The SEP subfamily in pineapple had only two members, *AcSEP1* and *AcSEP3*, compared to five in rice and six in orchid (Fig. 4D). The topology of the phylogenetic trees grouped *AcSEP3* into a subclade (AGL9 or SEP3) with two rice genes, *OsMADS7* and *OsMADS8*. *AcSEP1* was grouped into another subclade (AGL2/3/4 or SEP1) and formed a separate monocot subgroup with three rice MADS-box genes, *OsMADS1*, *OsMADS5*, and *OsMADS34*. There was only one member from pineapple in the AGL6 lineage, compared to two homologous genes in rice, *OsMADS6* and *OsMADS17*. Overall, the phylogenetic analysis revealed that pineapple has fewer ABCE genes compared to rice and orchid (Fig. 5).

Expression patterns of MADS-box genes during vegetative and reproductive development. Using publicly available transcriptome data and the expression profiles of MADS-box genes in specific developmental processes in pineapple, (A1), we analyzed expression levels in four vegetative and reproductive organs (root, leaf, flower, and fruit) at different developmental stages (Fig. 6; Supplementary Table S2). Three major clusters (A1, A2, and A3) of expression patterns were distinguished according to the expression specificity in different tissues. The A1 cluster included two expression groups. The first group comprised *AcSEP1*, *AcSEP3*, *AcAGL6*, and *AcAG*, which were expressed primarily in leaves and flowers and accumulated at very high levels during fruit development. In the second group, all three members classified as D genes (*AcAGL11a-c*) and one AP1 member (*AcFUL1*) were strictly restricted to fruit. The A2 cluster contained almost all Type I genes (except for *AcMADS40*) and several members from MIKC subclades (e.g., *AcBS*, *AcSOC1d/c*, and *AcSVP1/3*). These genes exhibited relatively low transcript accumulation in almost all tissues except for the B-class genes.
AcAP3, which were moderately expressed in flowers and leaves, and AcANR1a expressed in roots. The genes in the A3 cluster segregated into three major expression groups. The first group contained three genes, AcFUL2, AcMADS23, and AcMADS40. AcFUL2 was mainly detected in stage S4 fruit, and AcMADS23 and AcMADS40 were expressed in flowers, leaves, and roots. The second group comprised AcSOC1c, AcSVP1, and AcAGL12a/b, which had root-specific expression. The third group comprised AcPI, AcSOC1a, and AcFLC1. AcSOC1a and AcFLC1 were evenly expressed in almost all organs, whereas AcPI exhibited tissue-specific expression, with FPKM values more than 20-fold higher in flowers and leaves compared to roots.

**Expression patterns of floral MADS-box genes in pineapple.** To determine whether certain genes were associated with specific pineapple floral organs, we analyzed RNA-seq transcriptome data for MADS-box genes in five floral organs (sepals, petals, stamens, pistils, and ovules) at different developmental stages. Four
major expression clusters (B1, B2, B3, and B4) were observed for all MADS-box genes after expression normalization and cluster analysis (Fig. 7; Supplementary Table S3). The B1 cluster included six genes, which were subdivided into two expression groups. The first group contained three genes (AcAGL6, AcAG, and AcPI) that were highly expressed in three floral organs. AcSEP3, AcSOC1a, and AcFLC1 in the second group all had abundant transcript levels in all whorls of floral organs, which indicated extensive involvement in the physiological processes of flower development and organogenesis. The B2 cluster contained three genes, a BS subfamily gene (AcBS) and two AGL11 subfamily genes (AcAGL11a and AcAGL11c). These were highly expressed throughout all developmental stages of only ovules. The B3 cluster had three expression groups, and the genes exhibited diverse expression profiles. The first expression group contained three genes, which were subdivided into two expression groups. The group contained three genes, a BS subfamily gene (AcBS) and two AGL11 subfamily genes (AcAGL11a and AcAGL11c). These were highly expressed throughout all developmental stages of only ovules. The B3 cluster had three expression groups, and the genes exhibited diverse expression profiles. The first expression group contained three genes, AcAGL11b, AcFUL1, and AcSEP1, which belonged to distinct MIKC subfamilies. AcAGL11b was mainly expressed in reproductive organs (stamens and pistils). AcFUL1 was mostly detected in whorls 1 and 4; AcSEP1 in whorls 1, 2, and 4. In the second group,
AcMADS23 and AcMADS40 were detected at low levels in all four whorls. The third group included three genes with moderate or low expression in certain stages of a specific whorl. The remaining MADS-box genes composed the B4 cluster and included most of the type I genes (except for AcMADS40) and some members from the SVP, ANR, and SOC1 subfamilies.

Consistent with the expression analysis results in floral tissues, these genes had low transcript levels or no significant expression in each of the floral organs, suggesting that they are not important for floral organ development. AcMADS43 expression was not detected at any developmental stage. The putative ABCDE model genes had typical temporal and spatial expression profiles in the five analyzed floral organs. The expression of AcFUL1, an A-class gene, was high in sepals but was also detected at lower levels in pistils and stamens. The B-PI lineage gene AcPI was highly expressed in petals, stamens, and pistil tissues, whereas AP3 lineage gene, AcAP3, showed no appreciable expression in these organs. The B-sister class gene Ac Bs was only expressed in ovules, and its transcripts accumulated during the process of ovule development. The C-class gene AcAG retained the ancestral function in female and male reproductive organs and was highly expressed in ovules, and its transcripts accumulated during the process of ovule development. The C-class gene AcAG retained the ancestral function in female and male reproductive organs and was highly expressed in ovules, and its transcripts accumulated during the process of ovule development.

Even though RNA-seq data had been proved reliable by RT-qPCR and in situ hybridization, we selected eight A, B, C, E class orthologous genes (AcFUL1, AcFUL2, AcAP3, AcPI, AcAG, AcAGL6, AcSEP1, AcSEP3) to test their expression in four whorls of floral organs (sepal, petal, stamen, pistil) by RT-qPCR in pineapple. The RT-qPCR results were consistent with our RNA-seq results (Fig. 8A) and another published dataset of pineapple tissue-specific transcriptome34 (Fig. 8B), which provided solid proofs for the subsequent propose of ABCDE gene model (Fig. 8C).

**Discussion**

Our bioinformatics analysis identified 43 MADS-box genes in the pineapple genome. The phylogenetic analysis divided the type II MADS-box genes into 13 subfamilies, among which the OsMADS32-like clade appeared to be a novel monocot-specific lineage9,33,36. Two members from the FLC clade were identified in the A. comosus genome, but their relationship to genes from other species, such as Arabidopsis, was not well resolved. The low
support value might be due to the highly divergent sequences and extremely short length of monocot FLC genes. The gene numbers in most MIKC subfamilies were less than those identified in rice and sorghum, consistent with the idea that grasses underwent a recent whole genome duplication (WGD) after the divergence from the Bromeliaceae. The reduced number of MADS-box gene copies in pineapple has important implications for elucidating the evolutionary history of the MADS-box gene family prior to the divergence of grasses and pineapple.

**Different evolutionary patterns of MADS-box genes in pineapple and rice.** The phylogenetic analysis revealed two A-class, three B-class, one C-class, three D-class, and two E-class homologs in pineapple (Table 1). Although pineapple and rice shared a WGD event near the base of monocot plants, only one pair of duplicated genes, \(AcFUL1\) and \(AcFUL2\), was retained. Another gene pair, \(AcAGL11a\) and \(AcAGL11c\), probably arose from species-specific duplication in pineapple. Unsurprisingly, no gene species-specific duplication was found in pineapple within AP3/PI lineage. Indeed, in most monocots, only one B-AP3 and B-PI copy has been reported. Therefore, gene duplication in this subclade is rare, except in orchids where gene duplication events occurred frequently, followed by expression divergence that led to the innovation of specialized labellum. Rice had at least seven pairs of genes in the AP1, PI, AG, AGL, AGL2/3/4, AGL9, and AGL6 lineages that derived from a more recent WGD event that occurred prior to the origin of the Poaceae 66–70 MYA. The significant expansion of genes in the B-AP3 and E classes in orchid may have also resulted from the WGD. In contrast, there were fewer instances of recent gene duplications in pineapple, showing that the composition of ABCDE genes in pineapple was conserved and that the reduced gene numbers in these model classes are representative of an ancestral state. The five members in the SOCI clade, five members in the ANR1 clade, and three members in the SVP clade suggest that these subfamilies expanded frequently (Fig. 5). Thus, contraction of the gap between the number of type II genes in pineapple (31) and rice (43) may have been a consequence of the expansion of these clades. Among them, the expression profiles of SOCI subfamily genes were diverse, with each member having a distinct expression pattern. Because SOCI genes are primarily involved in the flowering phase transition and stress tolerance, the expansion and diversification of the SOCI subfamily in pineapple may have contributed to its adaptation to extreme tropic environments.
Pineapple MADS-box genes involved in fruit development. In most cases, MADS-box genes within the same phylogenetic subgroup exhibited analogous expression patterns, indicating that these genes have similar biological functions in shaping regulatory networks that influence specific developmental processes. The genes clustered in the AP3/PI, AG, and SEP/AGL6 clades were highly expressed in pineapple reproductive organs, consistent with their expected roles in floral organogenesis and development. Notably, most of the genes belonging to the AP1/FUL (AcFUL1), AG/AGL11 (AcAG, AcAGL11a/b/c), SEP (AcSEP1/3), and AGL6 (AcAGL6) subfamilies in pineapple were highly expressed throughout all fruit developmental stages. In tomato, a fleshy fruit model plant, several genes including SIMADS-RIN from the SEP subfamily, TAGL1 from the AG-like subfamily, and FUL1 and FUL2 from the AP1-like subfamily regulate both early fruit expansion and later ripening. In a study in banana, two SEP-like subfamily genes, MaMADS1 and MaMADS2, were functionally characterized, and repression of either gene led to slow-ripening and prolonged shelf-life phenotypes. The present finding that AcFUL1 transcript levels were relatively high in fruits is consistent with the hypothesis that the AP1/FUL clade in monocots includes only FUL-like rather than AP1-like sequences, and that these genes may control fruit development analogously to the Arabidopsis gene FRUITFULL. In summary, MADS-box genes from these subfamilies may play a conserved and essential role during fruit development and the ripening of the non-climacteric fleshy fruit of pineapple and therefore warrant further study in the context of pineapple fruit yield and storage.

Figure 7. Expression heat map of MADS-box genes in five representative floral organs (sepals, petals, stamens, pistils, and ovules) at different developmental stages of pineapple. The different developmental stage samples comprised four sepal stages (S1–S4), three petal stages (S1–S3), five stamen stages (S1–S5), and seven stages of pistils (S1–S7) and ovules (S1–S7). The expression value was quantified as fragments per kilobase per million reads (FPKM), and relative gene expression data were gene-wise normalized. AcMADS43 was not detected at any developmental stage. Four major expression groups were marked as B1, B2, B3, and B4. The color scale bar is at the top-right corner; blue, white, and red indicate low, medium, and high expression levels, respectively.
Unlike the ABCDE genes of Arabidopsis and Antirrhinum whose expression is limited to flowers, many B-, C-, and E-class pineapple genes were also expressed in leaves. For example, AcSEP1 and AcPI were up-regulated in flowers and highly expressed in leaves, suggesting that these genes may also be critical for vegetative development. AcAGL12a and AcAGL12b had preferentially high expression levels in roots, with sparse expression in leaves and fruits, indicating that these two genes are root-specific and may be important for root development (Fig. 6). Almost all type I MADS-box genes had relatively low transcript levels or no significant expression based on their FPKM values from the RNA-seq data for different tissues. This finding is consistent with earlier reports that type I genes have relatively limited functions in plants compared to type II genes.48.

**Functional conservation of ABCDE model genes in pineapple.** The expression of A-class AcFUL1 in pineapple was notably higher in sepals than in other floral tissues, which is in line with its expected function in sepal identity specification. However, its expression was extremely low in petals, indicating that other A-function genes might be recruited to specify sepal identity in pineapple. AcFUL1 was also detected at moderate levels in pistils, a pattern also reported in orchid, in which two AP1/FUL-like members, PhaMADS1 and PhaMADS2, promote carpel and ovary development before and after pollination.49. There is evidence from a study of water lily and Nigella damascena that the role of AGL6-like subfamily genes is similar to that of AP1.50 In orchid, these genes have a specialized function in labellum formation. Although AcAGL6 was expressed in all tissues in our study, its expression was highest in sepals and petals, indicating a possible role in the A-class functions of whorl 1 and 2 specification via interaction with AP1 and SEP.

---

**Figure 8.** Floral organ-specific expression patterns of AcMADS genes and a deduced pineapple ABCDE model. (A) The left y-axis scales the relative expression of RT-qPCR result and Blue solid line indicated values of RT-qPCR. The right y-axis scales the FPKM value from RNA-seq result (stage1 sepal, petal, stamen, pistil) and orange dashed line indicated values of RT-qPCR. (B) Heatmap plot of expression patterns of pineapple A-, B-, C-, D-, E-class genes from another pineapple tissue-specific transcriptome. (C) A proposed pineapple flower model based on the expression patterns and referring to the ancestral functions of homeotic MADS-box genes. Blocks in different colors with distinct heights represent different expression levels. Gene functions for specific members are in brackets. *The gene function is difficult to predict because an expression split has occurred. The gaps are linked with grey lines.
Robust expression levels of AcAP3 and AcPI expression in petals and stamens indicated that these two genes performed B function based on the conserved expression patterns of B-class genes in angiosperms. The Bs subgroup, a phylogenetic sister group of the B-class floral homeotic genes, is specifically expressed in female reproductive organs and developing seeds. Both the sequence and ovule-specific expression profile of pineapple AcBs were highly similar to those of the rice ortholog OsMADS29, whose expression is restricted to developing seeds. In line with ancestral C function in specifying both male and female reproductive organs, high expression levels of AcAG were detected in stamens, pistils, and ovules (Fig. 8a). Regarding D function, three subfamily members (AcAGL11a, AcAGL11b, and AcAGL11c) were detected and found to be homologous to orchid MADS2 and rice OsMADS13/21. AcAGL11a/c were identified as D-function candidate genes because their expression patterns in our study were similar to those of other D-lineage genes, which are preferentially expressed in ovules. The expression of AcAGL11b in the inner whorls of flowers and ovules overlapped with the expression domain of AcAG. In rice, OsMADS21 exhibits the same expression pattern but lost its function in determining ovule identity, presumably because of its redundancy with OsMADS13, whose expression is also restricted to the female reproductive organs. We therefore hypothesized that like OsMADS21, AcAGL11b might have similarly lost its role in D function. AcSEP3 is more likely to have a major role in E function because it had higher transcript levels than AcSEP1 in all organs except sepals. Additionally, AcSEP3 orthologs, including ATAGL9 in Arabidopsis and OsMADS7/8 in rice, are more critical for E function than any other SEP-like family members in the two plants. By comparing the expression patterns of pineapple MADS-box genes with those of previously characterized orthologs, we inferred the functions of candidate pineapple genes involved in the ABCDE model. The numbers and evolutionary history of the putative ABCDE genes strongly indicated that the pineapple flower is similar to the ancestral state of monocot flowers.

Materials and methods
Data sources and sequence retrieval. The whole pineapple genome sequences that were used to identify MADS-box genes were downloaded from the Pineapple Genomics Database. Additionally, the water lily (Nymphaea colorata) genome was generated from our own genome project. The MADS-box protein sequences of Arabidopsis and rice were retrieved from the TAIR (http://www.arabidopsis.org/) and RGAP (http://rice.plantbiology.msu.edu/) databases, respectively. Non-redundant protein sequences of Amborella trichopoda, Vitis vinifera, Sorghum bicolor, Musa acuminate, and Spirodela polyrhiza were collected from Phytozome (http://www.phytozome.net/). The latest proteome release of Phalaenopsis equestris was from a recent study. The proteome of Elaeis guineensis was downloaded from the Genomsawit website (http://genomsawit.mpob.gov.my/index.php?track=30), and the Phoenix dactylifera proteome was obtained from the Date Palm Research Program (http://qatar-weill.cornell.edu/research/research-highlights/date-palm-research-program).

Genome-wide identification of MADS-box genes. To identify MADS-box gene family members in pineapple, the hidden Markov model (HMMER) profile of the SRF-TF domain (Pfam accession: PF00319) was obtained from the Pfam database (http://pfam.xfam.org/) as a query to search against pineapple proteins. The gene IDs and sequence information are provided in Table 1 and Supplementary Table S1. In addition to pineapple, MADS-box protein sequences in the following species were also collected and screened to investigate their evolutionary relationships: two basal angiosperm species (Amborella and waterlily); two core eudicots (Arabidopsis and Vitis vinifera); and seven monocot species (rice, sorghum, Phalaenopsis equestris, Musa acuminate, Musa acuminata (Musaceae), Elaeis guineensis (Arecales), Phoenix dactylifera (Arecales), and Spirodela polyrhiza (Lemnoidae)).

Classification of MADS-box genes in pineapple. To assign putative pineapple MADS-box genes to specific gene subfamilies, multiple sequence alignments were performed based on amino acid sequences using the alignment tool MAFFT with default parameter settings. MADS-box proteins from six plant species, A. trichopoda, N. colorata, A. thaliana, O. sativa, S. bicolor, and A. comosus, were used. Maximum-likelihood phylogenetic trees were constructed using FastTree software with the JTT+CAT model. Furthermore, a more detailed MIKC-protein phylogenetic tree was constructed using the same strategy with six additional species: V. vinifera (Vitaceae), P. equestris (Epidendroideae), M. acuminate (Musaceae), E. guineensis and P. dactylifera (Arecales), and S. polyrhiza (Lemnoidae). In the phylogenetic tree, bootstrap supporting values below 50 were generally regarded as unreliable and were not shown. As a matter of convenience, pineapple MIKC genes were renamed according to the phylogenetic relationship deduced by sequence comparison with proteins from the whole genomes of 11 flowering plants and their corresponding clade/subclade names. A phylogenetic tree of pineapple and ten other species (Fig. 2) was also constructed in the context of angiosperms. The phylogeny was inferred using RAxML v7.1.0 with the PROTGAMMAJTT model, 100 bootstrap replicates, and 756 single-copy genes, and the methods described above.

Expression profiles of MADS-box genes in different pineapple tissues. Two transcriptome datasets were extracted from previous studies in pineapple. The first dataset included floral, leaf, and root developmental tissues and six stages of fruit development (F1–F6, from young to mature). The other dataset was composed of different floral organ samples, including four sepal stages (S1–S4), three petal stages (S1–S3), five stamen stages (S1–S5), seven pistil stages (S1–S7), and seven ovule stages (S1–S7). The criteria for the different stages were previously described in details. The expression level of each gene was quantified as fragments per kilobase of exon model per million reads mapped (FPKM) values by featureCounts. The third dataset from previous published study provided available expression matrix including different floral organs in transcripts per kilobase million (TPM) by Stringtie. Two heat expression maps were generated by the Pheatmap package.
in R software using pre-treated FPKM values by log2 transformation. Differences in gene expression levels were represented using a color scale.

RNA isolation and RT-qPCR. Total RNA was extracted from tissues following manufacturer’s guidelines RNA extraction Kit (Omega Bio-Tek, Shanghai, China). cDNA was synthesized from 1 μg of RNA using the EasyScript One-Step gDNA Removal and cDNA Synthesis SuperMix (Transgen, Beijing, China). RT-qPCR analysis was carried out using TransStart Top Green qPCR SuperMix (Transgen, Beijing, China). The primers of MADS-box gene and internal control (Actin2) used for RT-qPCR were shown in Supplementary Table S5. To confirm results reliability, three biological replicates were conducted. 2-ΔCT method was applied to calculate the relative expression of pineapple MIKC genes. For statistical analysis, normalized relative values are given as mean value ± standard deviations.

Received: 12 April 2020; Accepted: 27 November 2020
Published online: 13 January 2021

References
1. Bowman, J. L. & Meyerowitz, E. M. Genetic control of pattern formation during flower development in Arabidopsis. Symp. Soc. Exp. Biol. 45, 89–115 (1991).
2. Ng, M. & Yanofsky, M. F. Function and evolution of the plant MADS-box gene family. Nat. Rev. Genet. 2, 186–195 (2001).
3. Yanofsky, M. F. et al. The protein encoded by the Arabidopsis homeotic gene agamous resembles transcription factors. Nature 346, 35–39 (1990).
4. Mizukami, Y. & Ma, H. Ectopic expression of the floral homeotic gene AGAMOUS in transgenic Arabidopsis plants alters floral organ identity. Cell 71, 119–131 (1992).
5. Xu, G. & Kong, H. Duplication and divergence of floral MADS-box genes in grasses: Evidence for the generation and modification of novel regulators. J. Integr. Plant Biol. 49, 927–939 (2007).
6. Wu, F. & Yanofsky, M. F. Function and evolution of the plant MADS-box gene family. Curr. Opin. Plant Biol. 16, 743–756 (2013).
7. Kater, M. M., Dreni, L. & Colombo, L. Functional conservation of MADS-box factors controlling floral organ identity in rice and Arabidopsis. J. Exp. Bot. 57, 3433–3444 (2006).
8. Wu, F. & Yanofsky, M. F. Linking plant development and evolution. Plant Physiol. 180, 3259–3271 (2016).
9. Tozawa, T. et al. Genomic duplication of MADS-box genes in rice. Plant Cell Environ. 49, 585–600 (2016).
10. Causier, B., Schwarz-Sommer, Z. & Davies, B. Floral organ identity. Semin. Cell Dev. Biol. 21, 73–79 (2010).
11. Xu, G. & Kong, H. Duplication and divergence of floral MADS-box genes in grasses: Evidence for the generation and modification of novel regulators. J. Integr. Plant Biol. 49, 927–939 (2007).
12. Wu, F. & Yanofsky, M. F. Linking plant development and evolution. Plant Physiol. 180, 3259–3271 (2016).
13. Dreni, L. et al. Functional analysis of all AGAMOUS subfamily members in rice reveals their roles in reproductive organ identity determination and meristem determinacy. Plant Cell 23, 2850–2863 (2011).
14. Lopez-Dee, Z. P. et al. OsMADS13, a novel rice MADS-box gene expressed during ovule development. Dev. Genet. 25, 237–244 (1999).
15. Dreni, L. et al. The D-lineage MADS-box gene OsMADS13 controls ovule identity in rice. Plant J. 52, 690–699 (2007).
16. Cui, R. et al. Functional conservation and diversification of class E floral homeotic genes in rice (Oryza sativa). Plant J. 61, 767–781 (2010).
17. Callens, C., Tucker, M. R., Zhang, D. & Wilson, Z. A. Dissecting the role of MADS-box genes in monocot floral development and diversity. J. Exp. Bot. 69, 2435–2459 (2018).
18. Lu, H., Liu, Z. & Lan, S. Genome sequencing reveals the role of MADS-box gene families in the floral morphology evolution of orchids. Hortic. Plant J. 5, 247–254 (2019).
19. Ming, R. et al. The pineapple genome and the evolution of CAM photosynthesis. Nat. Genet. 47, 1435–1442 (2015).
20. Bartholomew, D. P., Paull, R. E. & Rohrbach, K. G. The Pineapple: Botany, Production and Uses 1–301 (CABI Publishing, Wallingford, 2003).
21. Remizova, M. V., Sokoloff, D. D. & Rudall, P. J. Evolutionary history of the monocot flower. Ann. Mo. Bot. Gard. 97, 617–645 (2010).
22. Smyth, D. R. Evolution and genetic control of the floral ground plan. New Phytol. 20, 20 (2018).
23. Saquenet, H. et al. The ancestral flower of angiosperms and its early diversification. Nat. Commun. 8, 20 (2017).
24. Imms, D. M. & Kelly, S. OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16, 157 (2015).
25. Tang, H., Bowers, J. E., Wang, X. & Paterson, A. H. Angiosperm genome comparisons reveal early polyplody in the monocot lineage. Proc. Natl. Acad. Sci. USA 107, 472–477 (2010).
26. Paterson, A. H., Bowers, J. E. & Chapman, B. A. Ancient polyplidization predating divergence of the cereals, and its consequences for comparative genomics. Proc. Natl. Acad. Sci. USA 101, 9903–9908 (2004).
27. Zhang, L. et al. The ancient wave of polyplidization events in flowering plants and their facilitated adaptation to environmental stress. Plant Cell Environ. 20, 20 (2020).
28. Alvarez-Buylla, E. R. et al. An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. Proc. Natl. Acad. Sci. USA 97, 5328–5333 (2000).
29. Dudiuchenko, O. et al. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science 356, 92–95 (2017).
30. Theissen, G., Melzer, R. & Rumpler, F. MADS-domain transcription factors and the floral quartet model of flower development: Linking plant development and evolution. Development 143, 3259–3271 (2016).
31. Mondragon-Palomino, M. & Theissen, G. Conserved differential expression of paralogous DEFICIENS- and GLOBOSA-like MADS-box genes in the flowers of Orchidaceae: Retining the “orchid code”. Plant J. 66, 1008–1019 (2011).
32. Su, Z. et al. Genome-wide identification of auxin response factor (ARF) genes family and its tissue-specific prominent expression in pineapple (Ananas comosus). Trop. Plant Biol. 10, 86–96 (2017).
33. Wang, L. et al. Floral transcriptomes reveal gene networks in pineapple floral growth and fruit development. Commun. Biol. 3, 500 (2020).
34. Mao, Q. et al. Comprehensive tissue-specific transcriptome profiling of pineapple (Ananas comosus) and building an eFP-browser for further study. PeerJ 6, e6028 (2018).
35. Zhao, T. et al. Characterization and expression of 42 MADS-box genes in wheat (Triticum aestivum L.). Mol. Genet. Genom. 276, 334–350 (2006).
36. Arora, R. et al. MADS-box gene family in rice: Genome-wide identification, organization and expression profiling during reproductive development and stress. BMC Genom. 8, 242 (2007).
37. Ruelens, P. et al. FLOWERING LOCUS C in monocots and the tandem origin of angiosperm-specific MADS-box genes. Nat. Commun. 4, 2280 (2013).
38. Zhang, G. Q. et al. The Apostasia genome and the evolution of orchids. Nature 549, 379–383 (2017).
39. Qu, C. et al. Genome-wide association mapping and identification of candidate genes for fatty acid composition in Brassica napus L. using SNP markers. BMC Genom. 18, 232 (2017).
40. Liu, C. et al. Direct interaction of AGL24 and SOC1 integrates flowering signals in Arabidopsis. Development 135, 1481–1491 (2008).
41. Liu, J. et al. MIKC(C)-type MADS-box genes in Rosa chinensis: The remarkable expansion of ABCDE model genes and their roles in floral organogenesis. Hortic. Res. 5, 25 (2018).
42. Karlova, R. et al. Transcriptional control of fleshy fruit development and ripening. J. Exp. Bot. 65, 4527–4541 (2014).
43. Itkin, M. et al. TOMATO AGAMOUS-LIKE 1 is a component of the fruit ripening regulatory network. Plant J. 60, 1081–1095 (2009).
44. Zhong, S. et al. Single-base resolution methylomes of tomato fruit development reveal epigenome modifications associated with ripening. Nat. Biotechnol. 31, 154–159 (2013).
45. Vrebalov, J. et al. A MADS-box gene necessary for fruit ripening at the tomato ripening-inhibitor (rin) locus. Science 296, 343–346 (2002).
46. Elitzur, T. et al. Banana MaMADS transcription factors are necessary for fruit ripening and molecular tools to promote shelf-life and food security. Plant Physiol. 171, 380–391 (2016).
47. Ferrandiz, C. & Fourniquin, C. Role of the FUL-SHP network in the evolution of fruit morphology and function. J. Exp. Bot. 65, 4505–4513 (2014).
48. Kofuji, R. et al. Evolution and divergence of the MADS-box gene family based on genome-wide expression analyses. Mol. Biol. Evol. 20, 1963–1977 (2003).
49. Acrt-Nunes-Miranda, R. & Mondragon-Palomino, M. Expression of paralogous SEP-, FUL-, AG- and STK-like MADS-box genes and food security. Plant Physiol. 171, 278 (2016).
50. Wang, P. et al. Flexibility in the structure of spiral flowers and its underlying mechanisms. Nat. Plants 2, 15188 (2015).
51. Becker, A. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. Mol. Phylogenet. Evol. 29, 464–489 (2003).
52. Schilling, S. et al. Non-canonical structure, function and phylogeny of the B-sister MADS-box gene OsMADS30 of rice (Oryza sativa). Plant J. 84, 1059–1072 (2015).
53. Yang, X. et al. Live and let die—the B(sister) MADS-box gene OsMADS29 controls the degeneration of cells in maternal tissues during seed development of rice (Oryza sativa). PLoS One 7, e51435 (2012).
54. Schneitz, K. The molecular and genetic control of ovule development. Curr. Opin. Plant Biol. 2, 13–17 (1999).
55. Galinha, K. D. & Di Stilio, V. S. Sub-functionalization to ovule development following duplication of a floral organ identity gene. Dev. Biol. 405, 158–172 (2015).
56. Sozzi, V. L., Nelsen, C. D., Hewett Hazeltin, K. D. & Di Stilio, V. S. Partial redundancy and functional specialization of E-class SEPALLATA genes in an early-diverging eudicot. Dev. Biol. 419, 143–155 (2016).
57. Wu, D. et al. Loss of LOFSEP transcription factor function converts spikelet to leaf-like structures in rice. Plant Physiol. 20, 20 (2017).
58. Xu, H. et al. PGD: Pineapple genomics database. Hortic. Res. 5, 66 (2018).
59. Zhang, L. et al. The water lily genome and the early evolution of flowering plants. Nature 577, 79–84 (2020).
60. Finn, R. D. et al. Pfam: The protein families database. Nucleic Acids Res 42, D222–230 (2014).
61. Kato, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30, 772–780 (2013).
62. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. Mol. Biol. Evol. 26, 1641–1650 (2009).
63. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313 (2014).
64. Chen, P. et al. Genome-wide identification and expression profiling of ATP-binding cassette (ABC) transporter gene family in pineapple (Ananas comosus (L.) Merr.) reveal the role of AcABC G38 in pollen development. Front Plant Sci. 8, 2150 (2017).
65. Yang, S., Zhang, G. S. & Wei, S. FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930 (2014).
66. Kovaka, S. et al. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. 20, 278 (2019).

Acknowledgements
This work was supported by the National Natural Science Foundation of Fujian (2019J1423), Program for New Century Excellent Talents in Fujian Province University, an outstanding youth fund (KXQ17002) from Fujian Agriculture and Forestry University to Liangsheng Zhang.

Author contributions
L.Z. conceived the original research plans; J.H., L.Z., X.C., Y.Z., X.Y., Y.S., and Y.Q. analyzed the data; J.H. and L.Z. wrote the article with contributions from all the authors.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-020-79163-8.

Correspondence and requests for materials should be addressed to L.Z.
