Protopercms and Protocorm-Like Bodies Are Molecularly Distinct from Zygotic Embryonic Tissues in Phalaenopsis aphrodite[OPEN]

Su-Chiung Fang*, Jhun-Chen Chen², and Miao-Ju Wei²

Biotechnology Center in Southern Taiwan, Academia Sinica, Tainan 741, Taiwan (S.-C.F., J.-C.C., M.-J.W.); and Agricultural Biotechnology Research Center, Academia Sinica, Taipei 115, Taiwan (S.-C.F., J.-C.C., M.-J.W.)

ORCID ID: 0000-0001-6894-5025 (S.-C.F.).

The distinct reproductive program of orchids provides a unique evolutionary model with pollination-triggered ovule development and megasporogenesis, a modified embryogenesis program resulting in seeds with immature embryos, and mycorrhiza-induced seed germination. However, the molecular mechanisms that have evolved to establish these unparalleled developmental programs are largely unclear. Here, we conducted comparative studies of genome-wide gene expression of various reproductive tissues and captured the molecular events associated with distinct reproductive programs in Phalaenopsis aphrodite. Importantly, our data provide evidence to demonstrate that protocorm-like body (PLB) regeneration (the clonal regeneration practice used in the orchid industry) does not follow the embryogenesis program. Instead, we propose that SHOOT MERISTEMLESS, a class I KNOTTED-LIKE HOMEOBOX gene, is likely to play a role in PLB regeneration. Our studies challenge the current understanding of the embryonic identity of PLBs. Taken together, the data obtained establish a fundamental framework for orchid reproductive development and provide a valuable new resource to enable the prediction of gene regulatory networks that is required for specialized developmental programs of orchid species.

Orchids have a unique reproductive program. Unlike most flowering plants, whose ovules and embryo sacs are fully developed and whose egg cells are ready to be fertilized at anthesis, in some orchids, development of the ovule and embryo sac is triggered by pollination (Withner, 1959; Arditti, 1992; Yeung and Law, 1997). Pollination and fertilization are separated by relatively long periods during which the ovule and embryo sac develop. These periods vary among species and can span from as little as 4 d in Gastrodia elata to 10 months in Vanda suavis (Arditti, 1992). After fertilization, embryogenesis of the vast majority of orchids lacks characteristic organogenesis, and the embryos are arrested at the globular stage as seeds continue to mature and desiccate (Raghavan and Goh, 1994; Yu and Goh, 2001; Kull and Arditti, 2002). As a result, matured embryos of orchids do not have the embryonic leaves (cotyledons) and embryonic roots commonly established in other seed plants (Arditti, 1992; Dressler, 1993; Burger, 1998). Despite the incomplete morphogenesis, orchid embryos possess axis polarity and complete maturation and desiccation processes (Yeung et al., 1996; Lee et al., 2007a, 2007b, 2008).

The plant embryo is a miniature sporophyte developed from the zygote after fertilization. Plant embryogenesis is a genetically defined program comprising sequential developmental processes that result in a mature embryo in which the structural and functional organization of the adult plant is established. Two distinct but overlapping phases, morphogenesis and maturation, take place during embryogenesis (Bentsink and Koornneef, 2008; Braybrook and Harada, 2008). The basic body plan, including the organization of apical-basal polarity, the formation of functionally organized domains, and cell differentiation and tissue specification, is established during morphogenesis (Steeves and Sussex, 1989). Transcriptional regulators and signaling components are important to control axis polarity and the cell division plane (Jeong et al., 2012; Ueda and Laux, 2012). As morphogenesis approaches completion, embryos cease to grow and macromolecules start to be synthesized and accumulated as storage reserves. Abscisic acid plays an important role and is coordinated with a battery of transcriptional factors, such as ABSCISIC ACID INSENSITIVE3 (ABI3), FUSCA3 (FUS3), LEAFY COTYLEDON1 (LEC1), and LEC2, to regulate

¹ This work was supported by the Development Program of Industrialization for Agricultural Biotechnology (to S.-C.F.) and by the Biotechnology Center in Southern Taiwan, Academia Sinica (to S.-C.F.).
² These authors contributed equally to the article.
* Address correspondence to scfang@gate.sinica.edu.tw.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Su-Chiung Fang (scfang@gate.sinica.edu.tw).

S.-C.F. conceived and designed the experiments; S.-C.F., J.-C.C., and M.-J.W. performed the experiments and analyzed the data; S.-C.F. wrote the article.

[OPEN] Articles can be viewed without a subscription.
Germination and early seedling development of orchid seeds is not a spontaneous process and often requires species-specific mycorrhizal colonization (Smith and Read, 2008; Waterman and Bidartondo, 2008). It has been proposed that mycorrhizal colonization provides phosphate, nitrogen, or other mineral nutrients to support seed germination (Leake, 1994; Rasmussen, 2002). Orchid seeds germinate and form small spherical tuber-like bodies referred to as protocorms. Unlike the vast majority of plants, whose growth axes (shoot and root meristems) are established early during embryo development, the protocorm has only one meristematic domain at the anterior end where new leaves and roots are formed sequentially (Nishimura, 1981). Because protocorm development initiates organogenesis that is absent during the morphogenesis program of orchid embryogenesis, protocorm development is sometimes considered to be a continuum of zygotic embryogenesis (Jones and Tisserat, 1990; Ishii et al., 1998). However, the embryonic status of protocorms remains to be determined.

Protocorm-like bodies (PLBs) resemble protocorms structurally and are triggered from explants and/or calluses (Jones and Tisserat, 1990; Chugh et al., 2009). During the initiation of PLBs, callus cells from the explant form compact regions that are composed of meristemoids (promeristems; Lee et al., 2013). Polarized growth starts from the surface cells of each compact pool of meristemoids. Continued cell division produces anterior smaller cells that give rise to the shoot pole of a PLB and posterior larger and vacuolated cells that form at the base of a PLB (Lee et al., 2013). The first leaf is formed from the shoot pole of the PLB, and the root is usually produced at the base of the first leaf (Hong et al., 2008). Sometimes roots initiate from the middle or bottom of the PLB. As the clonal propagation of PLBs can be multiplied by cutting, it has become a routine practice in the orchid floriculture industry to generate clonal plantlets (Yam and Arditti, 2009).

Because the protocorm is considered to be an extended state of the zygotic embryo, the initiation and development of PLBs is referred to as somatic embryogenesis (Begum et al., 1994; Chang and Chang, 1998; Ishii et al., 1998; Chen and Chang, 2006; Zhao et al., 2008; Lee et al., 2013). However, molecular evidence supporting such a hypothesis remains limited. Here, we describe genome-wide gene activity in ovary tissues containing developing ovules and embryos from fertilization to maturity as well as in protocorm and PLB tissues of *Phalaenopsis aphrodite*. Our study reveals that the molecular components required for plant ovule and embryo development are evolutionarily conserved in *P. aphrodite* and provides new insights into the molecular dynamics that characterize the reproductive program of *Phalaenopsis* spp. orchids. Moreover, our data show that protocorms and PLBs share similar transcriptomic signatures that are mostly different from those of zygotic embryos. Furthermore, the initiation and developmental processes of orchid PLBs do not follow the somatic embryogenesis program but instead show a distinct regeneration program. We report that *SHOOT MERISTEMLESS*, a class I KNOTTED-LIKE HOMEOBOX gene, is likely to play an important role in PLB regeneration. Our studies challenge the current understanding of the embryonic identity of PLBs and suggest an alternative pathway for PLB regeneration.

RESULTS

De Novo Transcriptome Assembly

To ensure full-spectrum coverage of transcripts whose expression is important/required for general and tissue-specific proliferation and the maintenance of meristematic tissues, a reference transcriptome library was generated by paired-end sequencing mRNA populations collected from different meristematic tissues, including developing ovaries containing developing ovules collected at 30, 40, 50, and 60 d after pollination (DAP; Fig. 1A; Supplemental Fig. S1) and developing embryos collected at 70, 80, 90, 100, 120, 140, 160, 180, and 200 DAP (Fig. 1A; Supplemental Fig. S1), developing protocorms (Fig. 1B), developing PLBs (Fig. 1C), newly emerging young leaves (less than 2 cm), emerging stalk buds (less than 2 cm), 5- to 10-cm floral stalks, root tips (less than 2 cm), and floral buds (less than 5 mm). The total size of the assembly was approximately 117 Mb. A summary of assembly statistics is given in Supplemental Table S1. This assembly yielded 112,467 unigenes that are longer than 200 nucleotides, with a mean unigene size of 1,038 nucleotides and a maximum unigene size of 15,584 nucleotides. This transcriptome provides the comprehensive transcript repertoire from various *P. aphrodite* meristems.

Different Sets of Coexpressed Genes Characterize Reproductive Development in *P. aphrodite*

To compare the transcriptome dynamics of protocorms, PLBs, ovary tissues spanning from ovule development to embryogenesis, young leaves, stalk buds, and floral stalks, 10-Mb single-end mRNA reads from each group of tissues (see “Materials and Methods”) were mapped to the assembled reference transcriptome and analyzed by principal component analysis. The mapping read analyses are summarized in
Supplemental Table S2. Four distinct groups were identified by the analysis (Fig. 2A): (1) ovaries containing developing ovules (30–40 and 50–60 DAP) and developing embryos (70–80 and 90–120 DAP); (2) ovaries containing maturating embryos (140–160 and 180–200 DAP); (3) protocorms and PLBs; and (4) young leaves, emerging stalk buds, and floral stalks. Developmental time frames identified as being closely associated are likely to share the greatest similarity in overall gene expression and, therefore, are likely to share cellular functions.

We were interested in understanding the extent to which gene expression was coordinated during developmental periods such as fertilization, embryo development, seed maturation, and protocorm and PLB development. K-mean analysis was applied to the RNA sequencing (RNA-seq) data set to identify genes whose expression was preferentially enriched in specific tissues. To reduce false-positive results, only transcripts expressed (fragments per kilobase of transcript per million mapped reads [FPKM] ≥ 5) in at least one tissue were selected. To increase the stringency, an equal to or greater than 10-fold change in at least two tissues was chosen as a cutoff. Twelve clusters representing stage-specific or tissue-specific enriched transcripts were identified (Supplemental Fig. S1B) and are listed in Supplemental Table S3. These transcripts encode a range of functional categories, although some of the encoded proteins had no known functions (Supplemental Data Set S1). Gene Ontology (GO) category enrichment analysis was applied to these clusters to identify predominant cellular processes associated with developing ovaries, protocorms, and PLBs (Fig. 2B; Supplemental Data Set S2). A detailed list of clustered transcripts and their GO terms is given in Supplemental Data Set S1.

After pollination, as the ovary started to develop at 30 to 40 DAP (Supplemental Fig. S1B, I), transcripts associated with cell cycle-associated biological processes, such as the cell cycle, cytokinesis by cell plate formation, DNA replication initiation, microtubule cytoskeleton organization, cell division, and regulation of the cell cycle, are strongly enriched (Fig. 2B). As development proceeded to 50 to 60 DAP (Supplemental
Fig. S1B, II), genes involved in metal ion transport, phosphoenolpyruvate carboxykinase activity, and Ser-type peptidase activity were represented (Fig. 2B). At around 70 to 80 DAP (Supplemental Fig. S1B, III), when fertilization had occurred and embryos started to develop in ovary tissues (Nadeau et al., 1996; Lee et al., 2008), the production of small interfering RNAs involved in RNA interference, DNA methylation, and methyltransferase activity were overrepresented (Fig. 2B). As embryos continued to develop in ovaries at 90 to 120 DAP and 140 to 160 DAP (Supplemental Fig. S1B, IV and V), transcripts encoded by genes associated with catabolism and macromolecular transport and storage such as lipid transport, carbohydrate transport, DNA catabolic processes, and sugar transmembrane transporter activity were significantly enriched (Fig. 2B). As the ovary reached maturity (Fig. 3B; Supplemental Fig. S1B, VI), transcripts of genes involved in antiporter activity, the ubiquitin ligase complex, and the oxidation reduction process were enriched (Fig. 2B).

GO analysis of the K-mean clusters VII, VIII, and IX revealed that cellular processes involved in oxidation reduction processes, heme binding, peroxidase activity, and copper ion binding were strongly enriched in both protocorms and PLBs (Fig. 2B; Supplemental Fig. S1B). Transcripts encoded by genes associated with terpene synthase activity, lyase activity, nucleoside triphosphatase activity, lipid metabolic processes, and the abscisic acid-activated signaling pathway were significantly enriched only in PLB tissues.

**Transcription Factors Enriched before, during, and after Fertilization in Ovary Tissues**

To survey the regulatory network controlling developmental transitions during orchid reproductive development, we used the GO category of regulation of transcription, DNA-templated to identify tissue- or stage-specific transcription factors (TFs) enriched in each cluster (Table I; Supplemental Data Set S3). Transcripts of Arabidopsis (Arabidopsis thaliana) orthologs previously shown to regulate carpel and ovule development, such as AGAMOUS (PaMADS1 and PaMADS61), SEPALLATA (PaMADS13), AINTEGUMENTA (PaANT), CUP-SHAPED COTYLEDON (PaCUC), and SPATULA (PaSPT; Elliott et al., 1996; Favaro et al., 2003; Pinyopich et al., 2003; Brambilla et al., 2007; Nahar et al., 2012; Galbiati et al., 2013), were overrepresented in ovary tissues collected from 30 to 60 DAP, during which ovules were developing. A homeobox TF, O39, isolated previously as an ovule-specific DNA of Phalaenopsis orchids (Nadeau et al., 1996), also was identified in our data set (PaHD-Zip4; Table I). The nomenclature of MADS domain TFs of P. aphrodite follows that of Phalaenopsis equestris (Cai et al., 2015). Three additional MADS domain factors, PaMADS61, PaMADS62, and PaMADS63, were identified in this study (Supplemental Fig. S2A).

As ovary development proceeded to 70 to 80 DAP, when fertilization occurred, mRNAs of four MADS domain factors, PaMADS61, PaMADS62, and PaMADS63, were identified in this study (Supplemental Fig. S2A).

**Figure 2.** Protocorms and PLBs are molecularly distinct from other reproductive tissues. A, Principal component analysis of developing ovaries at 30 to 40, 50 to 60, 70 to 80, 90 to 120, 140 to 160, and 180 to 200 DAP, PLBs, protocorms, young leaves, stalk buds, and floral stalks. B, Heat map showing the P-value significance of enrichment of GO terms for stage-specific mRNAs. Empty cells indicate a lack of association.
Figure 3. Expression profiles of selected genes during reproductive development. A, Expression patterns of TFs enriched in developing ovaries from 30 to 80 DAP. B, Expression patterns of genes enriched in developing ovaries from 100 to 200 DAP. Small PLBs (PLBS), medium PLBs (PLBM), large PLBs (PLBL), 10-d-old protocorms (protocorm10), 20-d-old protocorms (protocorm20), and 30-d-old protocorms (protocorm30) are indicated.
### Table 1. Comparative transcript abundances of selected TFs and overrepresented genes in reproductive tissues of *P. aphrodite* by RNA-seq analysis

Asterisks indicate that the expression pattern of the selected transcripts was confirmed by qRT-PCR.

<table>
<thead>
<tr>
<th>Transcript Identifier</th>
<th>Annotation</th>
<th>FPKM Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>orchid.id50198.</td>
<td><em>PaANTL1</em></td>
<td>30–40 DAP</td>
</tr>
<tr>
<td>tr203744*</td>
<td></td>
<td>67.2</td>
</tr>
<tr>
<td>orchid.id63296.</td>
<td><em>PaANTL2</em></td>
<td>50–60 DAP</td>
</tr>
<tr>
<td>tr203744*</td>
<td></td>
<td>80.4</td>
</tr>
<tr>
<td>orchid.id135459.</td>
<td><em>PaCUC2L</em></td>
<td>70–80 DAP</td>
</tr>
<tr>
<td>tr456992*</td>
<td></td>
<td>23.4</td>
</tr>
<tr>
<td>orchid.id124656.</td>
<td><em>PaCUC3L</em></td>
<td>90–120 DAP</td>
</tr>
<tr>
<td>tr357947</td>
<td></td>
<td>29.7</td>
</tr>
<tr>
<td>orchid.id150950.</td>
<td><em>PaHD-Zip4</em></td>
<td>140–160 DAP</td>
</tr>
<tr>
<td>tr242911*</td>
<td></td>
<td>127.1</td>
</tr>
<tr>
<td>orchid.id1211128.</td>
<td><em>PaMADS1</em></td>
<td>180–200 DAP</td>
</tr>
<tr>
<td>tr255012*</td>
<td>(AG-like)</td>
<td>138.3</td>
</tr>
<tr>
<td>orchid.id127983.</td>
<td><em>PaMADS7</em></td>
<td></td>
</tr>
<tr>
<td>tr216633*</td>
<td>(SHP/STK-like)</td>
<td>244.1</td>
</tr>
<tr>
<td>orchid.id125975.</td>
<td><em>PaMADS13</em></td>
<td></td>
</tr>
<tr>
<td>tr36909*</td>
<td>(SEP-like)</td>
<td>8.9</td>
</tr>
<tr>
<td>orchid.id5222.</td>
<td><em>PaMADS30</em></td>
<td></td>
</tr>
<tr>
<td>tr206845*</td>
<td></td>
<td>16.9</td>
</tr>
<tr>
<td>orchid.id125352.</td>
<td><em>PaMADS61</em></td>
<td></td>
</tr>
<tr>
<td>tr345834*</td>
<td>(AG-like)</td>
<td>310.5</td>
</tr>
<tr>
<td>orchid.id126107.</td>
<td><em>PaMADS62</em></td>
<td></td>
</tr>
<tr>
<td>tr333603*</td>
<td>(AGL6-like)</td>
<td>14.2</td>
</tr>
<tr>
<td>orchid.id122032.</td>
<td><em>PaSPT</em></td>
<td></td>
</tr>
<tr>
<td>tr208541*</td>
<td></td>
<td>16.1</td>
</tr>
<tr>
<td>orchid.id139127.</td>
<td><em>PaHD-Zip1</em></td>
<td>90 to 200 DAP</td>
</tr>
<tr>
<td>tr218563*</td>
<td></td>
<td>0.3</td>
</tr>
<tr>
<td>orchid.id134449.</td>
<td><em>PaHD-Zip2</em></td>
<td></td>
</tr>
<tr>
<td>tr70997*</td>
<td></td>
<td>0.8</td>
</tr>
<tr>
<td>orchid.id126646.</td>
<td><em>PaMADS39</em></td>
<td></td>
</tr>
<tr>
<td>tr143303*</td>
<td></td>
<td>1.7</td>
</tr>
<tr>
<td>orchid.id161716.</td>
<td><em>PaMADS51</em></td>
<td></td>
</tr>
<tr>
<td>tr186472*</td>
<td></td>
<td>0.1</td>
</tr>
<tr>
<td>orchid.id2086.</td>
<td><em>PaMADS52</em></td>
<td></td>
</tr>
<tr>
<td>tr469871*</td>
<td></td>
<td>0.0</td>
</tr>
<tr>
<td>orchid.id123254.</td>
<td><em>PaMADS63</em></td>
<td></td>
</tr>
<tr>
<td>tr4585*</td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>orchid.id67332.</td>
<td><em>PaABI3L1</em></td>
<td>90 to 200 DAP</td>
</tr>
<tr>
<td>tr200912*</td>
<td></td>
<td>0.0</td>
</tr>
<tr>
<td>orchid.id135554.</td>
<td><em>PaABI3L2</em></td>
<td></td>
</tr>
<tr>
<td>tr200912*</td>
<td></td>
<td>0.0</td>
</tr>
<tr>
<td>orchid.id166714.</td>
<td><em>PaBBM</em></td>
<td></td>
</tr>
<tr>
<td>tr43006*</td>
<td></td>
<td>0.0</td>
</tr>
<tr>
<td>orchid.id139573.</td>
<td><em>PaFUS3</em></td>
<td></td>
</tr>
<tr>
<td>tr88217*</td>
<td></td>
<td>0.0</td>
</tr>
<tr>
<td>orchid.id150115.</td>
<td><em>PaLEC1</em></td>
<td></td>
</tr>
<tr>
<td>tr415912*</td>
<td></td>
<td>0.0</td>
</tr>
<tr>
<td>orchid.id143578.</td>
<td><em>PaOLE1</em></td>
<td></td>
</tr>
<tr>
<td>tr366311*</td>
<td></td>
<td>7.9</td>
</tr>
<tr>
<td>orchid.id145648.</td>
<td><em>PaOLE2</em></td>
<td></td>
</tr>
<tr>
<td>tr244338*</td>
<td></td>
<td>1.9</td>
</tr>
<tr>
<td>orchid.id116824.</td>
<td><em>PaSTM</em></td>
<td></td>
</tr>
<tr>
<td>tr240339*</td>
<td></td>
<td>0.6</td>
</tr>
</tbody>
</table>

*Note: Table continues on following page.*
PaMADS39, PaMADS51, and PaMADS63) and three homeodomain-Leu zipper (HD-Zip) family TFs were found to be enriched (Table I). The expression patterns of selected transcripts listed in Table I were confirmed independently by qRT-PCR (Fig. 3A). The phylogenetic relationship of the identified orchid proteins and their Arabidopsis counterparts is shown in Supplemental Figure S2, A to C.

As ovary development proceeded from 90 to 200 DAP, many seed-specific TFs, such as BABY BOOM (PaBBM), PaLEC1, PaFLUS3, PaABI3L1, PaABI3L2, and WRINKLED1 (PaWRI1), were preferentially expressed (Table I). The identities of these embryonic marker genes were supported by phylogenetic analysis (Supplemental Fig. S2, B and D) and sequence alignment (Supplemental Fig. S3). qRT-PCR confirmed that mRNAs of PaBBM, PaLEC1, PaABI3L1, PaABI3L2, and PaFLUS3 started to accumulate in ovary tissues at 100 DAP and reached a peak at 180 DAP (Fig. 3B). The expression pattern of PaWRI1 mRNA followed a similar pattern except that its accumulation peaked at 160 DAP. The maturation status of the developing seeds was marked by the expression of two seed maturation markers, OLEOSIN1 (PaOLE1) and PaOLE2 (Table I; Fig. 3B). The preferential accumulation of PaBBM and PaLEC1 mRNAs in developing embryos was further confirmed by in situ hybridization, with both mRNAs starting to accumulate in developing embryos 80 DAP (early stage of seed development; Fig. 4A). The PaBBM mRNA level continued to increase from 110 to 180 DAP and reached a peak when embryos approached the end of maturation. PaLEC1 mRNA, on the other hand, reached a peak in embryos at 140 DAP and was maintained at a relatively high level from 160 to 180 DAP.

### Two 7S Globulins Are Preferentially Accumulated in Developing Seeds

During orchid seed development, storage proteins are synthesized and form protein bodies in developing embryos (Lee et al., 2006; Yang and Lee, 2014). We were interested to learn the compositions of the protein reserve and surveyed the expression patterns of the major seed storage protein, albumins, globulins, and

<table>
<thead>
<tr>
<th>Transcript Identifier</th>
<th>Annotation</th>
<th>30–40 DAP</th>
<th>50–60 DAP</th>
<th>70–80 DAP</th>
<th>90–120 DAP</th>
<th>140–160 DAP</th>
<th>180–200 DAP</th>
<th>PLBs</th>
<th>Protocorms</th>
<th>Young Leaves</th>
<th>Stalk Buds</th>
<th>Floral Stalks</th>
</tr>
</thead>
<tbody>
<tr>
<td>orchid.id158957.</td>
<td>PaERF1</td>
<td>0.0</td>
<td>0.6</td>
<td>0.16</td>
<td>0.06</td>
<td>4.0</td>
<td>0.0</td>
<td>47.5</td>
<td>5.8</td>
<td>0.0</td>
<td>6.5</td>
<td>0.0</td>
</tr>
<tr>
<td>tr258501*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>orchid.id152501.</td>
<td>PaERF2</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.2</td>
<td>0.3</td>
<td>0.1</td>
<td>24.0</td>
<td>4.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>tr67371*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>orchid.id130709.</td>
<td>PaGA2OX1</td>
<td>8.2</td>
<td>2.5</td>
<td>0.8</td>
<td>2.8</td>
<td>7.0</td>
<td>6.8</td>
<td>78.2</td>
<td>15.0</td>
<td>0.9</td>
<td>3.5</td>
<td>0.5</td>
</tr>
<tr>
<td>tr95546*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table I. (Continued from previous page.)**

You can find more details and analysis in the original text.
prolamins (Shewry et al., 1995). At least one 11S globulin and four 7S globulins were identified from our transcriptome database, and their transcript abundance is shown in Table I. The phylogenetic relationship of the full-length proteins encoded by the 11S globulin and two 7S globulins is shown in Supplemental Figure S4. Notably, transcripts of two 7S globulin genes, Pa7S-1 and Pa7S-2, accumulated steadily during the seed maturation process with a peak at 180 DAP (Fig. 3B), suggesting that they are the major storage proteins in orchid seeds. In addition to developing seeds, Pa7S-1 mRNA was extremely abundant in protocorms. Intriguingly, prolamin and 2S albumin, the storage proteins commonly found in endosperms, were not identified in our transcriptome database and two publicly available orchid transcriptome databases (Su et al., 2013; Cai et al., 2015).

Embryo-Specific Markers Are Not Enriched in Protocorm and PLB Tissues

The expression of embryonic markers such as BBM, LEC1, FUS3, and ABI3s is not only enriched in zygotic and somatic embryos, but some of their functions also have been demonstrated to be sufficient to establish embryonic fate during somatic embryogenesis (Braybrook and Harada, 2008; Smertenko and Bozhkov, 2014). Unlike somatic embryonic tissues observed in other plant species, PaLEC1 and PaFUS3 mRNAs were hardly detectable in protocorms and PLBs. The mRNAs of PaBBM and PaABI3L2 accumulated to low levels in 10-d-old protocorms and were hardly detectable in developing PLBs (Table I; Fig. 3B). The level of PaABIL1 mRNA was relatively high in abundance in developing protocorms and low in abundance in developing PLBs (Table I; Fig. 3B). In addition, mRNAs of the two embryonic marker genes, PaWRI1 and PaOLE1, whose functions are important for the accumulation of storage reserves during the maturation phase of embryogenesis (Cernac and Benning, 2004; Hsieh and Huang, 2004), were expressed in very low abundance or were undetectable in developing protocorms and PLBs. In situ hybridization also confirmed the lack of detectable PaLEC1 and PaBBM mRNAs in developing PLBs (Fig. 4, B and C). Therefore, the expression of the embryonic marker genes was not correlated with the initiation and development of PLBs.

The ability of PaLEC1 and PaBBM to induce somatic embryogenesis was verified in Arabidopsis. Expression of the PaBBM and PaLEC1 proteins was verified by western-blot analysis (Supplemental Fig. S5A).
Overexpression of 35S:PaBBM-eGFP or 35S:eGFP-PaLEC1/35S:PaLEC1-eGFP (Fig. 5, A and B) but not the empty vector (Supplemental Fig. S5, B and C) caused ectopic production of cotyledons from somatic tissues. The embryo-specific genes, such as OLE, COLEOSIN, CRUCIFERIN3 (CRU3), ABI3, and FUS3, were up-regulated in somatic embryonic culture tissues of the transgenic plants (Fig. 5, C and D). In addition, triacylglycerols, which accumulate in embryonic tissues and stained intensely with Sudan Red 7B, were also accumulated in embryonic-like tissues of 35S:PaLEC-eGFP and 35S:eGFP-PaBBM transgenic lines (Fig. 5E). Moreover, overexpression of eGFP-PaLEC1 rescued the lec1-1 desiccation-tolerant phenotype (Fig. 5E), supporting that PaLEC1 is the functional ortholog of Arabidopsis LEC1. The 35S:eGFP-PaBBM construct failed to be generated and, therefore, was omitted from this study. In short, PaLEC1 and PaBBM genes are evolutionarily conserved and capable of inducing the somatic embryonic program.

In addition to surveying transcriptional factors that displayed an enriched expression pattern during seed development, we also examined SOMATIC EMBRYOGENESIS RECEPTOR KINASE (SERK), whose expression has been shown to be associated with somatic embryogenesis (Schmidt et al., 1997; Hecht et al., 2001; Santos et al., 2005; Steiner et al., 2012). Three previously reported SERK-like genes, PaSERK1, PaSERK2, and PaSERK5 (Huang et al., 2014), also were present in our transcriptome database. However, the expression of PaSERK1, PaSERK2, and PaSERK5 was ubiquitous across the tissues we analyzed and did not display embryo-specific patterns (Table I). The lack of correlation between embryonic/somatic embryonic tissues and the expression of SERK genes has been reported in rice (Oryza sativa) and maize (Zea mays; Baudino et al., 2001; Ito et al., 2005).

Figure 5. PaBBM and PaLEC1 are sufficient to initiate somatic embryogenesis. A, Overexpression of PaBBM-eGFP induces embryonic culture tissue (ECT) in the wild type. B, Overexpression of eGFP-PaLEC1 induces ECT in the wild type. C, Expression of embryo-specific genes, CRU3, ABI3, and FUS3, in the ECT of PaBBM-eGFP overexpressors. D, Expression of embryo-specific genes, CRU3, ABI3, and FUS3, in the ECT of eGFP-PaLEC1 overexpressors. E, Sudan Red 7B staining indicates the accumulation of triacylglycerols in PaLEC1-eGFP and PaBBM-eGFP overexpressors. F, Overexpression of eGFP-PaLEC1 rescues the desiccation intolerance of the Arabidopsis lec1-1 mutant. Arrows indicate the complemented lec1-1 plants. Bars in A, B, and E = 500 μm.
Expression of PaSTM Is Associated with PLB Development

Knowing that PLB development does not follow the embryogenesis program, we were interested in genes whose functions may potentially contribute to PLB initiation and development and provide hints about pathway(s) from which the PLB is derived. To do that, TF genes under the GO category regulation of transcription, DNA-templated in PLB clusters (Supplemental Fig. S1B) were identified (Supplemental Data Set S3).

Among the PLB-enriched TFs, we were particularly interested in the class I KNOX TF PaSTM because ectopic expression of class I KNOX TFs has been reported to cause the formation of new tissue organization centers that are sufficient to induce shoots or organ-related structures (Sinha et al., 1993; Chuck et al., 1996; Williams-Carrier et al., 1997; Golz et al., 2002). PaSTM encodes a protein that shares 63% identity and 73% similarity with Arabidopsis STM. Phylogenetic analysis of class I KNOX TFs shows that PaSTM and Arabidopsis STM constitute a paired clade (Supplemental Fig. S6A). The expression level of PaSTM mRNA was correlated with the initiation of PLB, being significantly elevated at the early stage of PLBs (small) and declining but being maintained at moderate levels as development proceeded (30 d after germination; Fig. 6A). The level of PaSTM mRNA, on the other hand, was relatively low in 10- and 20-d-old protocorms and increased to a moderate level in 30-d-old protocorms. Coincidently, the expression pattern of PaSTM was positively correlated with its potential target, GIBBERELLIN 2-OXIDASE1 (PaGA2OX1; Fig. 6A). The maize STM homolog, KNOTTED1, has been shown to directly regulate GA 2-oxidase to maintain a low level of GA in shoot apical meristems (Bolduc and Hake, 2009). In addition, the accumulation of PaSTM in PLBs was positively correlated with a shoot generation marker (Che et al., 2006), PaERF1 (homologous to...
Arabidopsis RAP2.6L ERF/AP2 TF; Fig. 6A; Supplemental Fig. S6B), suggesting that PLBs have a shoot generation capacity. The expression of PaERF1 did not seem to be a general stress-related response, because the mRNA of PaERF2, a stress-related gene (Supplemental Fig. S6B), showed a different expression pattern. In situ hybridization further confirmed that PaSTM mRNA was evenly distributed in newly emerging and small PLBs but was undetectable in highly vacuolated callus cells (Fig. 6B). In developing protocorms, on the other hand, PaSTM mRNA was restricted to the anterior end where shoots initiate (Fig. 6C). Taken together, these results showed that the expression of PaSTM mRNA is tightly associated with the initiation of PLBs.

To exclude the possibility of the involvement of PaLEC1 and PaFUS3 during PLB initiation, the expression of PaLEC1 and PaFUS3 also was examined in newly emerging PLBs. Unlike high levels of PaLEC1 and PaFUS3 mRNAs detected in embryonic tissues at 160 DAP, PaLEC1 and PaFUS3 mRNAs were almost undetectable in developing PLBs, including newly emerging PLBs (Fig. 6D). This result confirms that the expression of PaLEC1 and PaFUS3 is not associated with PLB initiation. The preferential expression of PaSTM in newly emerging PLBs was validated by qRT-PCR analysis (Fig. 6D). WUSCHEL (WUS) is reported...
to act as a positive regulator of shoot stem cell fate (Gordon et al., 2007; Duclercq et al., 2011). *Phalaenopsis WUSCHEL* (*PaWUS*) and *WUS*-RELATED HOMEOBOX (*PaWOX*) genes were identified from our transcriptome database (Supplemental Fig. S6C). However, *PaWUS* and *PaWOX* mRNAs were not up-regulated in PLBs (Table I). It is possible that the expression of *PaWUS* is restricted to a defined area of PLB meristem, and its potential role in PLB development remains to be determined.

To test the ability of *PaSTM* to induce PLB-like organogenesis, we generated transgenic Arabidopsis plants overexpressing *PaSTM* or its fusion to GFP (GFP-*PaSTM* or *PaSTM*-GFP). Expression of the *PaSTM* proteins was confirmed by immunoblot (Fig. 7A).

Compared with the transgenic plant carrying the empty vector (Fig. 7B), most of the T1 transgenic plants overexpressing *PaSTM* grew very slowly and had severe developmental defects. Among the 237 T1 transformants we recovered, most displayed aborted normal postembryonic development and initiated unorganized cell proliferation, with 116 plants (49.4%) forming a markedly large dome-shaped structure (Fig. 7C) and 101 plants (43%) forming a brush-like disorganized structure (Fig. 7D) at the position where shoot meristems resided. The first two rosette leaves of overexpressors were severely deformed and displayed either a slender or a lobed structure (Fig. 7E). Most of these plants arrested at the seedling stage. Very few transgenic plants managed to make a few more leaves and bolted approximately 35 to 37 d after germination (Fig. 7F). The rest of the other transgenic lines (7.6%) grew like wild-type plants. To exclude the potential influence of antibiotics on transgenic plants overexpressing *PaSTM*, T1 transgenic plants that were kept off antibiotics 10 d after germination were examined. Phenotypes similar to those of transgenic plants kept on the antibiotics were observed (Supplemental Fig. S7A). Homozygous T2 lines from wild-type-looking 35:SFP-*PaSTM* overexpressors were recovered for further analysis. Unlike their heterozygous counterparts, leaf blades of cauline and rosette leaves of homozygous lines appeared to be curved and lobed and petals were shortened (Fig. 7G), similar to previous reports on transgenic plants overexpressing class I KNOX TFs (Chuck et al., 1996; Sakamoto et al., 1999; Ori et al., 2000). However, we did not observe ectopic meristem cells from *PaSTM* overexpressors, probably because of the structural divergence of STM between *PaSTM* and *Arabidopsis*. Similar to 35:S*PaSTM* transgenic lines, 35:SFP-*PaSTM* overexpressors displayed large dome-shaped or brush-like shoot apical meristems (Supplemental Fig. S7B). A less severe phenotype was observed in 35:S*PaSTM*-GFP overexpressors (Supplemental Fig. S7C). None of the described phenotypes were observed in vector-only control plants (Supplemental Fig. S7D).

The shoot regeneration capacity of the unorganized structures of *PaSTM* overexpressors was monitored by the expression of the shoot regeneration markers *RAP2.6L* and *GA2OX2* (Che et al., 2006). As shown in Figure 7H, mRNAs of these markers were increased in abundance in *PaSTM* overexpressors but not in the wild-type or transgenic plants carrying the empty vector. Additionally, the *RAP2.6L* mRNA level was positively correlated with the severity of abnormality of *PaSTM* overexpressors. Taken together, we propose that *PaSTM* may be an important factor for PLB generation.

**DISCUSSION**

**Transcriptome Dynamics of Pollination-Triggered Reproductive Development**

Here, we profiled temporal transcriptome dynamics in developing ovaries as development progressed from ovule to embryo development. The data revealed that distinct cellular functions were associated with different stages of the *Phalaenopsis* spp. reproductive program. Cell division-related gene activities were overrepresented at the early stage of ovary development, during which ovules and pollen tubes were developing. This is consistent with our previous finding that the selected core cell cycle genes are expressed preferentially in ovaries transitioning from gametophyte development to embryogenesis (Lin et al., 2014). In addition, TFs such as *AG*, *ANT*, *BEL1*, *CUC*, *SEP*, and SPT-like genes shown previously to regulate carpel and ovule development in Arabidopsis were enriched in ovary tissues before transition. It has been shown that *CUCs* work together with *SPT* to control the formation of carpel marginal structures and the development of septa and ovules (Nahar et al., 2012; Galbiati et al., 2013). *ANT* and *CUCs* are important for the initiation of the ovule primordium (Ishida et al., 2000; Galbiati et al., 2013). *BEL1* is important to control integument identity and works together with *AG* and *SEP* to regulate ovule development (Modrusan et al., 1994; Ray et al., 1994; Reiser et al., 1995; Brambilla et al., 2007). Ovule fate is further maintained by members of the MADS box family of TFs, *SHP1*, *SHP2*, *STK*, and *AG* (Favaro et al., 2003; Pinyopich et al., 2003). Taken together, our results indicate that ovule development of *Phalaenopsis* spp. orchids requires similar sets of regulatory factors and is likely to be evolutionarily conserved.

RNA-seq analysis revealed that DNA methylation, RNA interference, and Met biosynthetic processes marked the stage when fertilization occurred and embryos began to develop. The epigenetic reprogramming involved in small RNA-guided methylation and DNA methylation has been reported to maintain genome integrity and epigenetic inheritance in reproductive tissues (Calarco et al., 2012; Jullien et al., 2012) and is likely to play a role during the reproductive development of *Phalaenopsis* spp. orchids.

As embryos started to develop and seeds began to form (90–200 DAP), seed-specific TFs, including *PaLEC1*, *PaFULCA3*, *PaABI3L1*, *PaABI3L2*, *PaBBM*, and
**PaWRI1**, were identified as overrepresented transcripts. In addition, the enrichment of GO terms associated with lipid and carbohydrate mobilization and storage at the late stage of developing ovary tissues (140–160 and 180–200 DAP) confirms the concurrent seed maturation processes (Fig. 2B; Supplemental Data Set S2). The tissue-specific localization and functional studies of PaLECl and PaBBM validate their roles in activating and maintaining the embryonic program. Interestingly, the major shift in gene expression between early and late seed development (Fig. 2A) indicates maturation-specific transcriptome reprogramming, which also has been reported in Arabidopsis (Belmonte et al., 2013). Interestingly, a substantial overlap in RNA populations of ovary tissues progressing from ovule development to early embryo establishment (Figs. 2A and 3A, cluster X) suggests that similar cellular processes are likely to be shared as tissues transition from gametophytic to embryonic development in *Phalaenopsis* spp. orchids. Taken together, global comparisons of mRNA populations suggested a coordinated temporal shift in gene expression during reproductive development in ovary tissues. Although lacking the complete organogenesis and histodifferentiation, the maturation process of *Phalaenopsis* spp. orchid embryos characterized by seed-specific TFs is evolutionarily conserved.

One of the distinct features of orchid seeds is their lack of endosperms. It has been documented that triple fusion between polar nuclei and the sperm nucleus rarely occurs, and endosperm development is either aborted at a very early stage or cannot be initiated (Kull and Arditti, 2002; Batygina et al., 2003). Coincidently, transcripts of the storage proteins 2S albumin and prolamins commonly found in endosperms of monocotyledon plants (Shewry and Halford, 2002) could not be identified from our and publicly available transcriptome databases. Intriguingly, members of type I MADS box TFs, whose functions are found to be associated with endosperm development (Masiero et al., 2011), are greatly reduced in orchid genomes (Cai et al., 2015; Lin et al., 2016; Zhang et al., 2016). The concurrent loss of the molecular repertoire normally associated with endosperm development may contribute to the lack of endosperm in orchid seeds.

The Regulatory Network Controlling Genes Expressed during Embryo Development Is Different from Those of the PLB and Protocorm

Our studies provide evidence to support the notion that protocorms and PLBs share similar transcriptome dynamics. Particularly, chloroplast development and functions involved in heme binding, copper binding, and the oxidation reduction process were overly active. In addition, stress-related cellular processes including the abscisic acid-activated signaling pathway and the wounding response were overrepresented. It is likely that cutting and the culture conditions activate stress responses during tissue culture practice, as suggested previously (Fehér et al., 2003; Bai et al., 2013; Gliwicka et al., 2013; Maximova et al., 2014).

Surprisingly, protocorms and PLBs (often described as somatic embryos in the literature) shared little similarity with zygotic embryonic tissues or other meristematic tissues like emerging young leaves, emerging stalk buds, and floral stalks (Fig. 2A). The lack of embryonic identity in PLBs was further supported by the lack of RNA accumulation of two embryo-specific marker genes, PaLEcl and PaBBM. Therefore, protocorms and PLBs seem to possess a unique molecular program that is different from that of embryonic and other meristematic tissues.

Our data show that PLB and protocorm development seems to be regulated similarly but not identically. For example, PaABI3L1 mRNA was relatively abundant in protocorms but its level was low in PLBs (Fig. 3B). PaSTM, PaGA2OX1, and PaERF1 mRNAs, on the other hand, were relatively abundant in PLBs but less abundant in protocorms (Fig. 6A). Functional characterization of these genes may help in identifying gene regulatory networks unique to protocorms or PLBs.

**PLBs: Organogenesis Versus Somatic Embryogenesis**

Somatic embryogenesis is a defined developmental process that leads to the establishment of embryos independent of a fertilization event (von Arnold et al., 2002; Braybrook and Harada, 2008; Yang and Zhang, 2010). Similar to zygotic embryogenesis, specific molecular markers contributing to the formation of a basic body plan (morphogenesis) and the accumulation of seed storage macromolecules (maturation) are commonly used to define the somatic embryogenesis process. Despite the commonly accepted view in the orchid community that PLBs are of somatic embryonic origin, the comparative transcriptome and marker gene analyses presented here argue that the regeneration of PLBs does not follow the embryogenesis program. Instead, the tight correlation between PLB development and PaSTM expression suggests that the PaSTM-mediated shoot organogenesis pathway may be important for PLB initiation.

The expression of *STM* has been reported to be associated with organogenic shoot formation of *Kalanchoë daigremontiana*, *Agave tequilana*, and dodder (*Cuscuta pentagona*; Garcés et al., 2007; Abraham-Juárez et al., 2010; Alakonya et al., 2012). Up-regulation of the *STM* gene also is associated with the de novo assembly of shoot apical meristems from cultured explants (Gordon et al., 2007; Atta et al., 2009). Similarly, de novo leaf initiation of river weeds that lack typical shoot apical meristems is tightly linked to the expression of the *STM* gene (Katayama et al., 2010). The positive role of class I KNOX genes in meristem cell fate is further supported by the gain-of-function approach. In these cases, ectopic expression of the class I KNOX genes has been found to cause the formation of new tissue organization centers that are sufficient to induce ectopic shoots or
organ-related structures (Sinha et al., 1993; Müller et al., 1995; Chuck et al., 1996; Williams-Carrier et al., 1997; Brand et al., 2002; Golz et al., 2002; Lenhard et al., 2002). Consistently, PaSTM is able to increase shoot regeneration ability by promoting the expansion of shoot apical meristems and/or inducing organ-related structures with concurrent up-regulation of shoot regeneration makers, such as RAP2.6L and GA2OX2, when overexpressed in Arabidopsis. Based on these findings, we propose that PaSTM may be an important factor for PLB regeneration.

During de novo organogenesis, the initiation of cell division, establishment of auxin maxima, and specification of founder cells are commonly observed before organ development (Pernisová et al., 2009; Perianez-Rodriguez et al., 2014). In addition, the coordinated interaction of phytohormones such as auxin, cytokinin, and GAs is known to regulate a battery of genes required for the acquisition of competence and/or the organization of the shoot apical meristem (Benková et al., 2003; Gordon et al., 2009; Pernisová et al., 2009; De Smet et al., 2010; Marhány et al., 2011, 2014; Schuster et al., 2014). It is likely that similar signaling and gene regulatory activities are important in setting up PLB regeneration. In fact, the expression of PaGA2OX1, a GA-catabolizing enzyme, was associated with PLB development. This is consistent with previous reports that GA inactivation is required for the activity of the shoot apical meristem (Sakamoto et al., 2001; Jasinski et al., 2005; Bolduc and Hake, 2009). Additionally, several auxin-responsive genes were found to be up-regulated in PLB tissues (Supplemental Data Set S1). It will be of future interest to understand how phytohormone signaling is integrated with the transcription network to coordinate the initiation and development of PLBs.

Unlike the de novo organogenesis commonly observed in other plant species, where organs such as shoots or roots are directly generated from callus-derived meristematic tissues (Duclercq et al., 2011), wound-induced callus tissues of _Phalaenopsis_ spp. orchids often if not entirely take the PLB route before subsequent shoot regeneration (Ernst, 1994). This raises the question of whether PLB is a specialized form of organ with shoot meristem development. If this is the case, what are the unique factors that contribute to PLB regeneration, and how are they different from those for direct shoot regeneration? The comprehensive transcriptomic catalog of developing PLBs presented here provides a valuable resource and may help in identifying key genes to address these questions.

In summary, in this study, we profiled the genome-wide transcriptome of reproductive development in _P. aphrodite_. The temporal and spatial integration of transcriptomic dynamics in reproductive tissues at different developmental stages provides valuable insights into the gene regulatory programs that characterize the reproductive processes of _Phalaenopsis_ spp. orchids. In addition, our findings uncover a previously unrecognized tissue identity of PLBs. Furthermore, we propose that the regeneration program of PLBs is likely derived from a PaSTM-dependent regeneration process.

### MATERIALS AND METHODS

#### Plant Materials and Growth Conditions

_Phalaeonopsis aphrodite_ ssp. _foranossana_ (m1663) seedlings in 2.5- or 3-inch pots were purchased from Chain Port Orchid Nursery. Plants were grown in a growth chamber with alternating 12-h-light (23°C)/12-h-dark (18°C) cycles. Arabidopsis (_Arabidopsis thaliana_) Columbia ecotype was grown at 22°C under 16-h-light/8-h-dark cycles in a growth chamber. The _iao1-1_ mutant (Lotan et al., 1998) was purchased from the Arabidopsis Biological Resource Center (http://www.arabidopsis.org).

#### Induction of PLBs

 Mature seeds were allowed to germinate on one-quarter-strength Murashige and Skoog basal medium supplemented with 0.1% (w/v) tryptone, 1.8% (w/v) Suc, 2% (w/v) potato homogenate, 0.00001% (w/v) thiamine-HCl, 0.0001% (w/v) pyridoxine-HCl, 0.0001% (w/v) nicotinic acid, 0.01% (w/v) myoinositol, and 1% (w/v) agar, adjusted to pH 5.7. Fifty-five to 80 d after germination, protocorms grown to approximately 3 mm in diameter were excised to remove the tip and bottom portions. The protocorm segments were transferred to PLB-inducing medium containing 0.1% (w/v) tryptone, 2% (w/v) Suc, 2% (w/v) potato homogenate, 2.5% (w/v) banana homogenate, 0.01% (v/v) citric acid, 0.1% (w/v) charcoal, and 1% (w/v) agar, adjusted to pH 5.5 to induce primary PLBs. Primary PLBs were then used to induce secondary PLBs by cutting to remove the tip and bottom portions. The PLB segments were then moved to PLB-inducing medium as described above. Secondary PLBs were used for the experiments described below. Cultures were maintained at 25°C under 16-h-light/8-h-dark cycles under illumination at 45 to 55 μmol photons m⁻² s⁻¹ in a growth chamber. A thin layer of callus tissue formed on the cutting surface of explants before the appearance of emerging PLBs (Fig. S5).

#### Sample Collection and RNA Extraction

Orchid flowers were hand pollinated, and developing ovaries were harvested at the specified day. For reproductive tissues, only the interior tissues of developing capsules were scraped and pooled for RNA extraction (Fig. 1A; Supplemental Fig. S1). Because PLBs did not grow at a synchronized rate, the size and developmental stage of PLB samples were categorized and collected separately (Fig. 1B). Protocorms that germinated after 10, 20, and 30 d were collected and categorized (Fig. 1C). Because germination rate varied in different batches of capsules, only healthy-looking protocorms (light-green in color and larger than the ones that failed to develop) were individually collected for the experiments. Young leaves, emerging stalk buds, or root tips with sizes smaller than 2 cm in length, 5- to 10-cm-long floral stalks, and floral buds less than 5 mm were collected for RNA extraction. The tissue samples were flash frozen in liquid nitrogen and stored in a freezer at ~80°C. RNA was isolated using OmicZol RNA Plus extraction reagent (Omic Bios) according to the manufacturer’s instructions. The isolated total RNA was treated with RNase-free DNase (Qiagen) followed by RNaseasy mini-column purification according to the manufacturer’s instructions (Qiagen).

#### Transcriptome Assembly and RNA-Seq

RNA samples extracted from different tissues, including interior tissues of developing capsules (50, 40, 50, 60, 70, 80, 90, 100, 120, 140, 140, 160, 180, and 200 DAP), protocorms, PLBs, young leaves, emerging stalk buds, floral stalks, floral buds, and root tips, were pooled together for transcriptome sequencing. The paired-end cDNA library was synthesized and amplified using the TruSeq RNA Sample Prep Kit (version 2) following the manufacturer’s instructions (Illumina). The cDNA library was sequenced by Genome HiSeq2000 (Illumina) according to the supplier’s protocols (Wolgene). The raw sequencing data were filtered to remove low-quality (quality value ≥ 20), adapter, and repeat sequences (Smeds and Künstner, 2011). Qualified reads were de novo assembled by the Trinity program (Grabherr et al., 2011). The resulting sequences and paired-end information were connected and incorporated by SPSPACE Basic version 2 (Boetzer et al., 2011) and extended by TGICL (Pertea et al., 2003) to
Data Analysis and Coexpression Clustering

Principal component analyses were performed using GeneSpring12.6.1 software. To reduce the false-positive result caused by low-abundance transcripts and generate robust groups of coregulated genes, transcripts with an FPKM value larger than or equal to 5 in at least one tissue and that showed a 10-fold difference in at least two different tissues were selected for clustering. The identification of coregulated mRNAs was performed using Multi-Experiment Viewer software (Howe et al., 2011). K-mean support using Pearson’s correlation was used to separate 10,893 assembled transcripts into 12 clusters (groups of coregulated genes). Gene lists derived from K-mean clusters were analyzed for GO term enrichment. The cumulative probability (P value) of hypergeometric distribution was calculated to evaluate the significance of the GO enrichment from each K-mean cluster (Chen et al., 2015). TFs falling into the designated clusters were manually selected and verified by qRT-PCR analysis (Fig. 3).

Phylogenetic Tree Construction

Protein Sequences Were Aligned by ClustalW

The resulting alignments were used to construct phylogenetic trees using MEGA 5.2.2 (Tamura et al., 2011). The maximum-likelihood method was used to generate phylogenetic trees, and 1,000 replicates were used for bootstrapping. Only bootstrap values of 50% or higher were shown for each clade. The evolutionary distances were computed using the Poisson correction method, and the rate variation among sites was modeled with a g-distribution.

Plasmid Construction and Arabidopsis Transformation

Total RNA was isolated as described previously (Lin et al., 2014). Micrograms of total RNA was used for cDNA synthesis. LEC1 cDNA from various plant species (maize [Zea mays], AF410176; rice [Oryza sativa], AX262824; soybean [Glycine max], BD242763; Arabidopsis, NM_102046; and Brassica napus, EU371726) were aligned. The evolutionarily conserved region of cDNA was used to design primer 5'-CAGCGCAAGATCTCCTGGACAC-3' and reverse primer 5'-CTCACGGTATCGGTGGAGAT-3' to isolate PaLEC1 cDNA. 3' RACE-PCR was carried out using the SMARTer RACE cDNA amplification kit according to the manufacturer’s instructions (Clontech). The amplified PaLEC1 cDNA was verified by sequencing. A new primer, 5'-TCAAGGAAGCTAAACATGCAA-3', derived from the sequenced PaLEC1 cDNA was designed for 5' RACE-PCR (SMARTer RACE cDNA amplification kit; Clontech). The full-length cDNA of PaLEC1 was BP recombined into pDONR221 vector to make pDONR_PaLEC1. pDONR_PaLEC1 was then LR recombined with pK2FWG2 or pH7FWG2.0 (Karimi et al., 2002) to make 35S CaMV:PaLEC1-eGFP or 35S CaMV:eGFP-PaLEC1. Similarly, the full-length cDNA of PaSTM was PCR amplified and Bp recombined into pDONR221 vector to make pDONR_PaSTM. pDONR_PaSTM was then LR recombined with pK2FWG2.0, pK7FWG2.0, or pH7FWG2.0 to make 35S CaMV: PaSTM, 35S CaMV:eGFP-PaSTM, or 35S CaMV:PaSTM-eGFP, respectively. For the PaBBM gene, the full-length cDNA of PaBBM was PCR amplified, digested by XhoI and SalI, and cloned into a modified version of the pDONR221 vector (the ccbF gene was replaced by multiple cloning sites from NotI to KpnI of the pBluescript vector) to make pDONR_PaBBM. pDONR_PaBBM was then LR recombined with pH7FWG2.0 to make 35S CaMV:PaBBM-eGFP construct.

The resulting plasmids were transformed into Agrobacterium tumefaciens strain GV3101. The wild-type strain (Wassilewskija ecotype) was transformed by A. tumefaciens GV3101 using the floral dipping method (Clough and Bent, 1998). In addition to antibiotic selection, the transformants were further verified by PCR amplification of a 207-bp DNA fragment using forward primer 5'-CAGCGCAAGATCTCCTGGACAC-3' and reverse primer 5'-CTCACGGTATCGGTGGAGAT-3'; a 525-bp DNA fragment using forward primer 5'-AACTGCAATCTCCCTGGCATG-3' and reverse primer 5'-AAACCTGCAATCTCCCTGGATG-3' for PaLEC1; a 279-bp DNA fragment using forward primer 5'-TGTGAGAATGAGCAGGAA-3' and reverse primer 5'-ATAGAGCCCTCTTCTGTTT-3' for PaBBM. For the complementation test of PaLEC1, 35S CaMV: PaLEC1-eGFP or 35S CaMV:eGFP-PaLEC1 was transformed into the lco1-1 mutant (Lotan et al., 1998).

qRT-PCR

DNA-free RNA was reverse transcribed in the presence of a mixture of oligo(dT) and random primers (9:1 ratio) using the GoScript Reverse Transcription System (Promega) according to the manufacturer’s instructions. Ten microliters of reverse transcription-PCR reaction contained 2.5 μL of 1:20 diluted cDNA, 0.2 μM of primers, and 5 μL of 2× KAPA SYBR FAST master mix (KAPA Biosystems). The following program was used for amplification: 95°C for 1 min, and 40 cycles at 95°C for 5 s and 58°C to 62°C for 20 s. PCR was performed in triplicate, and the experiments were repeated with RNA isolated from two independent samples. Primer pairs and the specified annealing temperature used for quantitative PCR are listed in Supplemental Table S4. LUBQUITIN was used as an internal control (Lin et al., 2014).

In Situ Hybridization

Protocorms and PLBs samples were fixed in 4% paraformaldehyde, 4% DMSO, 0.25% glutaraldehyde, 0.1% Tween 20, and 0.1% Triton X-100 in diethylpyrocarbonate-treated water at 4°C overnight. Tissues were then dehydrated and infiltrated with Paraplast (Leica) using a KOS Rapid Microwave Labstation (Milestone). Tissues (10 μm thick) were sectioned using a MICROM 315R microtome (Thermo Scientific). Sections were then dehydrated and infiltrated in 70, 80, 90, 100% ethanol for 1 h at room temperature, rinsed several times in water, and then mounted in glycerol for examination with an SMZ800 stereoscopic zoom microscope (Nikon).

Sudan Red 7B Staining

Sudan Red staining has been described in detail (Harding et al., 2003). Briefly, tissues were incubated in 70% (v/v) ethanol overnight to remove chlorophyll before staining. Tissues were stained in 0.1% (w/v) Sudan Red 7B (Sigma) solution for 1 h at room temperature, rinsed several times in water, and then mounted in glycerol for examination with an SMZ800 stereoscopic zoom microscope (Nikon).

PaSTM Antibody Production

Specific antibody for PaSTM was generated in rabbits against a peptide region (LHFHPKRSMENWSGNNP) unique to PaSTM. This peptide was conjugated to ovalbumin via a Cys residue and injected into rabbits according to the supplier’s protocols (LTG BioLaboratories). The resulting antibody was affinity purified by column chromatography using PaSTM peptide.
Protein Isolation and Immunoblot Analysis

Approximately 0.1 g of tissues was homogenized in 1 mL of RIPA buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA, 0.25% [w/v] deoxycholate, 0.1% [w/v] SDS, and 1% [v/v] Nonidet P-40) or 1 mL of urea buffer (250 mM Tris-HCl, pH 6.8, 3.5% [w/v] SDS, 1 M urea, and 10% [v/v] glycerol) supplemented with 1 mM phenylmethanesulfonyl fluoride, 1× protease inhibitor cocktail (Sigma), and 10 μg/mL MG132 (Sigma). Homogenized protein extracts were centrifuged at 16,000g for 4°C for 5 min to remove cell debris. The protein concentration of the supernatant was determined by the DC Protein Assay Kit according to the manufacturer’s instructions (Bio-Rad). Approximately 50 μg (PaLEC1 and PaBBM) or 10 μg (PaSTM) of protein was resolved on a Bolt 4% or 12% or 10% Bis-Tris Plus gel (Invitrogen) and transferred to Immobilon-P (PaLEC1 and PaBBM) or 100 μm PVDF (PaSTM) membranes. The peroxidase activity was detected by a chemiluminescence assay (Advansta). The peroxidase antibody was detected by a chemiluminescence assay (Advansta).

Accession Numbers

Sequence data from this article can be found in the GenBank data libraries under accession numbers SRA261547 and GENW01000000.

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Images of interior tissues of developing ovules harvested at different stages for RNA-seq and K-mean clustering identified 12 groups of genes associated with specific developmental stages.

Supplemental Figure S2. Phylogenetic relationship of MADS, AP2, HD-Zip, and B3 domain TFs.

Supplemental Figure S3. Protein sequence alignment of LEC1 and BBM proteins.

Supplemental Figure S4. Unrooted maximum-likelihood trees of the 7S and 11/12S albumins.

Supplemental Figure S5. Confirmation of PaLEC1-cGFP and PaBBM-cGFP proteins in overexpression lines.

Supplemental Figure S6. Phylogenetic relationship of class I KNOX, AP2/ERF, and WUS/WOX TFs.

Supplemental Figure S7. Overexpression of PaSTM causes abnormal meristems.

Supplemental Table S1. Statistics of the de novo transcriptome assembly.

Supplemental Table S2. Statistics of RNA-seq reads.

Supplemental Table S3. Statistics of 12 gene clusters in Supplemental Figure S1B.

Supplemental Table S4. List of primer pairs used for qRT-PCR.

Supplemental Data Set S1. Comparative transcript abundances of preferentially regulated genes in reproductive tissues of P. aphrodite by RNA-seq analysis.

Supplemental Data Set S2. GO analysis of tissue-specific K-mean clusters.

Supplemental Data Set S3. Comparative transcript abundances of preferentially regulated TFs in reproductive tissues of P. aphrodite by RNA-seq analysis.

ACKNOWLEDGMENTS

We thank Dr. Donna Fernandez for providing the Arabidopsis seeds; Dr. Sree-Suak Ko, Yi-Chen Lien, and Hue-Ting Yang for assistance with in situ hybridization; Chi-Nga Chow and Dr. Wen-Chi Chang for assistance with GO analysis and comments on the article; Cheng-Pu Wu for database maintenance; the AS-BCST Greenhouse Core Facility for greenhouse service; and Miranda Loney for English editing.

Received May 27, 2016; accepted June 18, 2016; published June 23, 2016.

LITERATURE CITED


Copyright © 2016 American Society of Plant Biologists. All rights reserved.


De Smet I, Lau S, Vanneste S, Neettleton D, Howell SH

Fang et al.


Ref: Copyright © 2016 American Society of Plant Biologists. All rights reserved.


