New Phytologist

The floral transcriptome of *Eucalyptus grandis*

Kelly J. Vining[1], Elisson Romanel[2], Rebecca C. Jones[3], Amy Klocko[1], Marcio Alves-Ferreira[4], Charles A. Hefer[5], Vindhya Amarasinghe[1,6], Palitha Dharmawardhana[6], Sushma Naithani[6], Martin Ranik[7], James Wesley-Smith[8], Luke Solomon[9], Pankaj Jaiswal[6], Alexander A. Myburg[6] and Steven H. Strauss[10]

[1]Center for Genome Research and Biocomputing, Oregon State University, Corvallis, OR 97331, USA; [2]Departamento de Biotecnologia, Escola de Engenharia de Lorena, Universidade de São Paulo (EEL-USP), CP 116, 12602-810 São Paulo, Brazil; [3]School of Biological Sciences, University of Tasmania, Private Bag 55, Hobart, 7001 TAS, Australia; [4]Laboratório de Genética Molecular Vegetal (LGMV), Departamento de Genética, Universidade Federal do Rio de Janeiro (UFRJ), Av. Prof. Rodolpho Paulo Rocco, CCS 21949900 Rio de Janeiro, Brazil; [5]Department of Botany, University of British Columbia, 3529-6270 University Blvd, Vancouver, BC V6T 1Z4, Canada; [6]Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331, USA; [7]Department of Genetics, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Private Bag X20, Pretoria 0028, South Africa; [8]Council for Scientific and Industrial Research, 1 Meiring Naude Rd, Pretoria, South Africa; [9]Seed Technology Programme, Sappi Forests Shaw Research Center, Howick 3290, South Africa; [10]Department of Forest Ecosystems and Society, Oregon State University, Corvallis, OR 97331, USA

Author for correspondence:
Steven H. Strauss
Tel: +1 541 737 6578
Email: steve.strauss@oregonstate.edu

Received: *7 June 2014*
Accepted: *13 August 2014*

New Phytologist (2015) **206:** 1406–1422
doi: 10.1111/nph.13077

Key words: *Eucalyptus*, floral transcriptome, flowering, MADS-box, S-domain receptor-like kinase (SDRLK).

Summary

- As a step toward functional annotation of genes required for floral initiation and development within the *Eucalyptus* genome, we used short read sequencing to analyze transcriptomes of floral buds from early and late developmental stages, and compared these with transcriptomes of diverse vegetative tissues, including leaves, roots, and stems.
- A subset of 4807 genes (13% of protein-coding genes) were differentially expressed between floral buds of either stage and vegetative tissues. A similar proportion of genes were differentially expressed among all tissues. A total of 479 genes were differentially expressed between early and late stages of floral development. Gene function enrichment identified 158 gene ontology classes that were overrepresented in floral tissues, including 'pollen development' and 'aromatic compound biosynthetic process'. At least 40 floral-dominant genes lacked functional annotations and thus may be novel floral transcripts.
- We analyzed several genes and gene families in depth, including 49 putative biomarkers of floral development, the MADS-box transcription factors, 'S-domain'-receptor-like kinases, and selected gene family members with phosphatidylethanolamine-binding protein domains. Expanded MADS-box gene subfamilies in *Eucalyptus grandis* included *SUPPRESSOR OF OVEREXPRESSION OF CO 1* (*SOC1*), *SEPALLATA* (*SEP*) and *SHORT VEGETATIVE PHASE* (*SVP*) *Arabidopsis thaliana* homologs.
- These data provide a rich resource for functional and evolutionary analysis of genes controlling eucalypt floral development, and new tools for breeding and biotechnology.

Introduction

Comparative analysis of flowering pathways can provide important insights into mechanisms of evolution and gene function (Izawa *et al.*, 2003). Much of the knowledge of the genetic regulation of flowering and floral development has been obtained by studying the small herbaceous annual *Arabidopsis thaliana*. Although few plant species have the full suite of *A. thaliana*'s model plant characteristics, genomic methods can be applied to any species, greatly expanding the scope of evolutionary analysis (e.g. Zahn *et al.*, 2010). The unique structure of *Eucalyptus* flowers – as discussed in this paper – provides an excellent opportunity for comparative floral genomics.

In *A. thaliana*, flowering is controlled by five genetic pathways: three endogenous (gibberellic acid, autonomous, and age) and two environmental (photoperiod and temperature) (Wellmer & Riechmann, 2010; Posé *et al.*, 2012). The floral activators *SUPPRESSOR OF OVEREXPRESSION OF CO 1* (*SOC1*) and *FLOWERING LOCUS T* (*FT*) integrate signals from these pathways to promote the transition to flowering. *SOC1* activates floral meristem identity genes *LEAFY* (*LFY*), *APETALA1* (*AP1*) and *FRUITFUL* (*FUL*), which promote the transition from vegetative to floral meristems. Both *AP1* and *SOC1* are activated by *FT* (Lee & Lee, 2010). Ectopic expression of *FT* promotes early flowering in a variety of plant species (Pin & Nilsson, 2012). *FT*, as well as *MOTHER OF FT* (*MFT*) and *TERMINAL-FLOWER-LIKE 1* (*TFL1*), is a member of the phosphatidyl ethanolamine-binding protein (PEBP) gene family, duplicated members of which have been shown to have highly divergent/subfunctionalized expression patterns in both angiosperms and gymnosperms, and to play a role in plant adaptation and evolution (Karlgren *et al.*, 2011; Klintenäs *et al.*, 2012; Pin & Nilsson, 2012).

Once established, the floral meristem undergoes differentiation, leading to development of distinct floral organs. The prevailing ABC genetic model of floral organ identity (Coen & Meyerowitz, 1991), later expanded to ABCDE (Theissen, 2001), has been applied to a wide variety of species. This model posits that the identity of each concentric whorl of floral organs depends on expression of a specific combination of genes from particular classes (Pelaz *et al.*, 2000). Studies of diverse plant taxa have revealed a high degree of conservation of the regulation of floral development across species, including eudicots with atypical floral structures and monocots (Kramer & Hall, 2005; Thompson & Hake, 2009).

Many of the ABCDE class genes are members of the MADS-box gene family. These genes encode transcription factors that share a highly conserved *c.* 180-nucleotide DNA-binding domain, and control various processes of plant development, from root formation to fruit ripening (Gregis *et al.*, 2006). In *A. thaliana*, expression of MADS-box genes is essential for the identity of the four floral whorls, and also for late events of flower development. Mutations may result in homeotic transformations of floral organs (reviewed in Theissen, 2001; Krizek & Fletcher, 2005; Causier *et al.*, 2010). There are 109 distinct plant MADS-box genes, many of which are expressed in a tissue- or cell-specific manner, sometimes beyond floral organs and flowering pathways (Masiero *et al.*, 2011).

Eucalypt homologs of *A. thaliana* flowering genes appear to have conserved roles (Kyozuka *et al.*, 1997; Southerton *et al.*, 1998a,b; Brill & Watson, 2004; Dornelas *et al.*, 2004; Jaya *et al.*, 2010; Jones *et al.*, 2011a), making them suitable as biomarkers for eucalypt floral development. For example, a homolog of *A. thaliana APETALA1* (*API*) has been used as an indicator of both flower induction and floral organ differentiation in *Eucalyptus occidentalis* (Jaya *et al.*, 2010, 2011), and expression of the *Eucalyptus* homolog of *LFY* (Southerton *et al.*, 1998a; Dornelas *et al.*, 2004; Jaya *et al.*, 2011; Jones *et al.*, 2011a) provides a biomarker for early floral development. Eucalypt genes in the *SEPALLATA* (*SEP*) and *PISTILLATA* (*PI*) groups were specifically expressed in petals, stamens, and carpels of developing flowers (Southerton *et al.*, 1998a), and therefore provide biomarkers for late floral development. Biomarker indices comprised of several genes, enabled by genomic studies such as that presented here, should provide robust markers of developmental stages.

Genes from the receptor-like kinase (RLK) family are involved in flower development and reproduction. RLKs constitute one of the largest gene families in plants (e.g. *c.* 600 in *A. thaliana*, *c.* 1500 in rice (*Oryza sativa*), and *c.* 2100 in poplar, *Populus trichocarpa* (Shiu & Bleecker, 2001; De Smet *et al.*, 2009)). The structural diversity in the extracellular region (ECR) allows RLKs to respond to a variety of ligands, which is consistent with the expression of RLK genes in diverse tissue types (Gao & Xue, 2012; Lan *et al.*, 2013), in response to both biotic and abiotic stresses (Lan *et al.*, 2013), and in diverse cellular processes including those related to flower development and reproduction (e.g. tapetum development, microspore maturation (Colcombet *et al.*, 2005), pollen–pistil interaction (Kachroo *et al.*, 2001), pollen tube reception (Escobar-Restrepo *et al.*, 2007), pistil elongation, stigma exsertion (Tantikanjana *et al.*, 2009), and gametophyte development (Wang *et al.*, 2012)). RLKs are classified according to their ECR (e.g. *A. thaliana* has > 21 classes) (Shiu & Bleecker, 2001). S-domain RLKs (SDRLKs) share sequence and structural similarity with the ECR of the *Brassica* S-locus receptor kinase (SRK) (Naithani *et al.*, 2007) and constitute the second largest class of RLKs. One of these, SRK, is an integral plasma membrane protein of the stigma epidermis and acts as the female determinant of self-incompatibility (Kachroo *et al.*, 2001), promoting pistil elongation and stigma exsertion (Tantikanjana *et al.*, 2009). The functions of the majority of SDRLKs are not known, although most genes are expressed in more than one tissue type. However, many SDRLKs show preferential expression in the flowers and in response to stress (http://planex.plantbioinformatics.org/), (Gao & Xue, 2012; Lan *et al.*, 2013).

The *Eucalyptus grandis* W. Hill ex Maiden BRASUZ1 genome is the first sequenced genome from the Myrtaceae family of angiosperms (Myburg *et al.*, 2014). The flower buds of *Eucalyptus* and other genera in the Myrtaceae family are structurally unusual. In *Eucalyptus*, the sepal primordia fuse to each other to form an outer calycine operculum and the petal primordia fuse to form an inner corolline operculum during early flower development (Pryor & Knox, 1971) (Fig. 1). In some eucalypt species, these two operculae are fused to each other, but in *E. grandis* they are separate structures, as in other species from the subgenus *Symphyomyrtus* (Ladiges, 1997). The inflorescence of *E. grandis* contains seven to 11 flower buds arranged in an umbel, which is enclosed by floral bracts. These bracts are shed early in flower development, followed by the shedding of the outer operculum. Once the flower buds are fully developed, the inner operculum sheds to expose the reproductive structures (anthesis). Flower buds typically reach anthesis in 4 months, and fruits usually reach

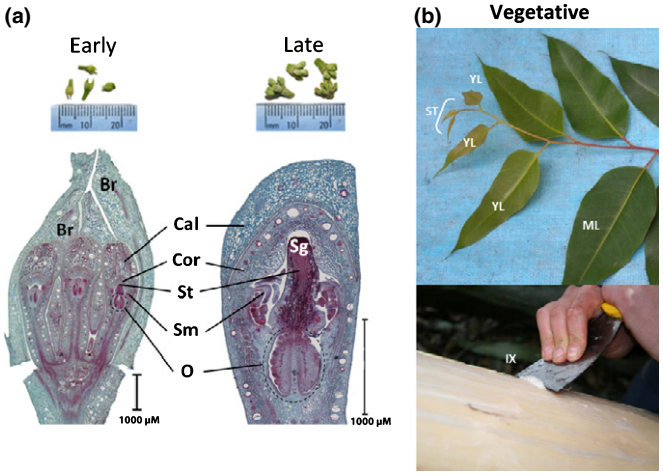


Fig. 1 Tissues sampled for RNA-seq analysis. (a) Floral tissues. Structures of early and late flowers are shown. The early image is of an inflorescence with three flower buds visible, while the late image is of a single flower bud. Cal, calycine operculum (fused sepals); Cor, corolline operculum (fused petals); Sg, stigma; St, style; O, ovary; Sm, stamen; Br, floral bract. (b) Vegetative tissues. Healthy vegetative tissues were collected from three actively growing, 5-yr-old *Eucalyptus grandis* trees in a clonal field trial. ST, shoot tips; YL, young leaves; ML, mature leaves; IX, immature xylem. Phloem and roots (not shown) were also sampled.

maturity after an additional 5–8 months (Hodgson, 1976). The unique structure of eucalypt flowers is likely to have been accompanied by substantial diversification of floral gene families and changes in the floral transcriptome.

In this study, we used an RNA-seq approach to identify the genes and developmental pathways that take part in *Eucalyptus* floral bud development. We surveyed genome-wide expression patterns in differentiated vegetative and floral tissues, and analyzed the MADS-box, SDRLK, and selected PEBP domain gene subfamilies. We also identified homologs of a number of well-known genes associated with floral initiation or floral organ identity that appear to be useful biomarkers of floral development. We report numerous variations upon established themes of floral development, laying the foundation for functional studies.

Materials and Methods

Materials for RNA-seq

Flowers were sampled from 15-yr-old grafted clonal *Eucalyptus grandis* trees grown in seed orchards owned by Sappi Forests (genotypes T1087 and T1144) and Mondi South Africa (genotype GC0004), in Pietermaritzburg, South Africa. The two breeding stations are both on the Natal Midlands (Sappi at Howick; Mondi at Hilton), and have very similar climates. The *E. grandis* flowering cycles can therefore be expected to be highly synchronized between the two sites. Tissues were sampled between December 2008 and April 2009, with several sampling dates per genotype. Flowers were sampled from three ramets (clones) of each genotype and pooled to promote sample homogeneity and decrease sampling bias. All plant material was harvested directly into liquid nitrogen and subsequently stored at −80°C.

Three stages of flower bud development were sampled. 'Stage one early' (S1E) buds were the smallest flower buds that could be distinguished from vegetative buds, and were clustered closely together with the entire inflorescence (umbel) enclosed by the floral bracts (Fig. 1a, early). 'Stage one late' (S1L) buds were larger floral buds with the entire umbel still enclosed by the bracts, the calycine and corolline operculae distinguishable and the stamens and carpels in early stages of development. 'Stage two early' (S2E) buds had separated into a distinct umbel and the floral bracts had shed, but the outer (sepaline) operculum was still in place, and the inner whorls (male and female reproductive structures) were actively developing (Fig. 1a, late).

Healthy vegetative tissues were collected from three actively growing, 5-yr-old *E. grandis* trees in a clonal field trial (Mondi Tree Improvement, Kwambonambi, South Africa) as described previously (Ranik & Myburg, 2006; Mizrachi *et al.*, 2010). The following tissues were sampled: from breast height (1.35 m) on the main stem following bark removal, immature xylem (outer glutinous 1–2-mm layer comprising developing xylem cell layers) and early developing phloem tissue (1–2-mm layer from the inner surface of the bark); from the tip of the crown after felling the trees, shoot tips (soft green termini of young crown tip branches containing apical meristems, shoots, and a small section of soft green stem below the shoot primordia and the first one or two unfolded leaves), young leaves (three to four soft, rapidly expanding leaves), and mature leaves (older, fully expanded leaves of the current growth season). Roots were collected from rooted cuttings of the same *E. grandis* clonal genotypes, grown in a glasshouse in Pretoria.

Floral tissue microscopy

Samples of floral buds were fixed in a solution of 24 : 1 ethanol : acetic acid, and dehydrated in a graded ethanol series diluted with phosphate-buffered saline (15 min in each of 30%, 50%, and 75% ethanol; 30 min in 100% and a further 60 min in 100%). Samples were subsequently infiltrated at 37°C with mixtures of absolute ethanol and Steedman's wax made up in the proportions of 2 : 1 (v/v) overnight; 1 : 1 and 1 : 2 (v/v), for 2 h at each step, followed by two changes of pure wax (2 × 15 min each) under mild vacuum. Specimens were transferred to plastic molds, covered in fresh wax at 37°C, allowed to stand overnight under vacuum, then polymerized by cooling to room temperature.

Sections were dewaxed by immersion in xylene (2 × 5 min), gradually hydrated by rinsing in decreasing concentrations of ethanol : water (100%, 70%, 50% and 30% ethanol, for 5 min per step) and stained for 1–2 h with Safranin O (1% w/v in water). Slides were subsequently washed in distilled water and dehydrated in an ethanol series (30%, 50%, 70% and 95% ethanol (v/v) for 5 min per step) before immersion in Fast Green stain (0.1% w/v in 95% ethanol for 30 s). Slides were then washed in 100% ethanol, dipped in carbol-xylene/methyl salicylate and xylene (1 : 1), cleared in xylene, air-dried, and mounted on glass slides using Permount (Fisher Scientific, Waltham, MA, USA). Slides were viewed using a Nikon 80i, and calibrated images were captured digitally using a DS-Fi1 camera and NIS-D software (Nikon, Tokyo, Japan).

RNA extraction and sequencing library preparation

Total RNA was extracted from plant tissue using a cetyltrimethylammonium bromide (CTAB)-based method (Chang *et al.*, 1993) and quantified using an RNA 6000 Pico assay on a Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA). Two polyadenylated RNA enrichments per sample (Oligotex mini kit; Qiagen, Valencia, CA, USA) resulted in *c.* 1% recovery of polyA+ RNA. RNA-seq libraries were constructed using previously described methods (Mizrachi *et al.*, 2010). Vegetative tissue libraries used nonbarcoded adapters. Floral libraries used barcoded adapters, which permitted equimolar pooling of genotypes. Libraries (10 nM) were sequenced using the Illumina GAIIx sequencer (2 × 76 bp; Illumina Inc., San Diego, CA, USA) at the Center for Genome Research and Biocomputing (CGRB) at Oregon State University (Corvallis, OR, USA).

Identification of differentially expressed genes

Based on initial analysis of several dozen biomarker genes, expression results from the two early floral bud stages were found to be

nearly identical. Thus, sequencing reads from those two libraries were pooled to constitute the 'early' floral bud sample for all subsequent analyses. Reads were aligned to the *E. grandis* reference genome version 1.1 using TopHat (Trapnell *et al.*, 2010). Normalized gene expression levels (fragments per kilobase of exon per million reads mapped (FPKM)) were calculated, and differentially expressed genes were determined using CuffDiff version 1.3 (Trapnell *et al.*, 2010).

Genes meeting criteria of at least 5 FPKM and at least four-fold difference were subjected to hierarchical clustering using a Pearson correlation metric, then plotted using the Gplots package version 2.10 in R (cran.r.-project.org/web/packages/gplots/index.html). Shoot tips were excluded from this analysis, as these were recognized to consist of a composite of differentiated tissues that may have included cryptic floral buds. The subset of 4807 genes with floral-specific expression was analyzed for enrichment of gene ontology (GO) functional categories using the agriGO Singular Enrichment tool (Du *et al.*, 2010).

Biomarker selection

Key genes in *A. thaliana* flowering time networks (Flowers *et al.*, 2009; Amasino, 2010; Higgins *et al.*, 2010) were considered as potential biomarkers for early flower development. Genes involved in development of the inner flower whorls were considered as biomarkers of late flower development, as the key anatomical difference between early and late flower buds was the stage of development of these inner whorls (Fig. 1). This list included genes predominantly expressed in the tapetum and pollen grains of *A. thaliana* (Alves-Ferreira *et al.*, 2007) supplemented by searching TAIR10 (http://www.arabidopsis.org/index.jsp) for genes with GO annotations such as GO:0048658 (tapetal layer development), GO:0048480 (stigma development), GO:0048481 (ovule development) and GO:0048440 (carpel development), and retaining only those genes with expression patterns relatively specific to these tissue types. In the case of the MADS-box and *FT* families, *E. grandis* orthologs were identified by phylogenetic analyses as described in the Methods section of this paper. In addition, the Inparanoid algorithm (Ostlund *et al.*, 2010) was used to identify orthologous and paralogous genes based on best two-way pairwise relationships between *A. thaliana* and *E. grandis* protein sequences using a similarity cut-off score of 50. Overall, 244 unique genes were considered (Supporting Information Table S1). Of these, 117 had no apparent ortholog in *E. grandis* (Table S1); this was confirmed by reciprocal BLASTp searches of the BOGAS and TAIR10 databases. For the 127 floral genes with at least one putative *E. grandis* ortholog, gene models were excluded if they: were not expressed in *E. grandis* floral tissue (normalized FPKM values < 3.0); had non-differential expression in tissue harvested late versus early in flower bud development (0.5- to 1.5-fold change), or had a pattern of expression that was inconsistent with the expected pattern (i.e. a putative ortholog of a floral initiation gene expressed late rather than early in flower development, or the inverse pattern for genes involved in floral organ development).

Quantitative reverse transcription PCR

Quantitative reverse transcription PCR (qRT-PCR) was used to confirm basic expression levels of a subset of genes (*LFY*, *AGAMOUS* (*AG*) and *SEEDSTICK* (*STK*) selected as candidate biomarkers. Floral tissues from separate collection years were used to maximize independence in relation to the original samples used for RNA-seq. Samples for qRT-PCR were harvested from genotypes T1144 and T1099 between December 2011 and April 2012, from genotype T1087 between December 2010 and April 2011, and from genotype B0133 during both of these periods.

Floral buds from each of the three genotypes were used as biological replicates for 'early' and 'late' that were comparable in morphology to the early and late samples described above in the 'Materials for RNA-seq' section for RNA-seq. Total RNA was extracted using an RNeasy kit (Qiagen), substituting a custom homogenization for that included in the kit (4 M guanidine thiocyanate, 0.2 M sodium acetate (pH 5.0), 25 mM EDTA, 2.5% w/v polyvinylpyrrolidone, and 1% v/v β-mercaptoethanol). Any DNA contamination was removed by DNaseI (Invitrogen, Carlsbad, CA, USA) treatment. One microgram of total RNA per sample was used to synthesize cDNA for qRT-PCR using SuperScript III reverse transcriptase (Invitrogen) and oligo (dT) primers. The synthesized cDNA was diluted 1 : 10 v/v with nuclease-free water, and 1 µl was used for qRT-PCR amplification. Three technical replicates per biological replicate were amplified in a StepOne Plus (Applied Biosystems, Foster City, CA, USA) thermal cycler using SyBR green I detection under the following conditions: 95°C for 10 min, then 40 cycles of 95°C for 15 s and 57°C for 1 min, followed by a melt curve stage increasing from 60 to 95°C at a rate of $+0.3° s^{-1}$. Actin (Eucgr.G02932.1) was used as a housekeeping gene to normalize the expression of the investigated genes. All targets had similar amplification efficiencies of ± 97.5%. Expression ratios were calculated with StepOne Plus software version 2.2 using the $2^{-\Delta\Delta C_t}$ method. Primer sequences are given in Table S2.

Annotation, phylogenetic analysis, and expression profiling of MADS-box genes

The identification and annotation of all MADS-box gene sequences used in this work from *E. grandis*, *A. thaliana*, *Populus trichocarpa* and *Vitis vinifera* were previously described (Myburg *et al.*, 2014). BLASTn searches of the Phytozome v9.1 database (http://www.phytozome.net/) using *A. thaliana* genes as queries were used to identify members of the *FT/TFL1* gene family in *E. grandis*, *P. trichocarpa* and *V. vinifera*. As the *FT/TFL1* family from *Populus nigra* has been annotated (Igasaki *et al.*, 2008), we conducted multiple sequence alignments to annotate all *P. trichocarpa FT/TFL1* family genes. A similar method was used with *V. vinifera*, in which the *FT/TFL1* family has also been annotated (Carmona *et al.*, 2007). All multiple sequence alignments and phylogenetic analysis methods used for MADS-box and *FT/TFL1* genes families were as previously described

(Myburg *et al.*, 2014). A 'digital northern blot' was used to represent expression profiles of the putative MADS-box genes and to identify differential expression among the diverse *E. grandis* tissues. The normalized frequency of reads for each gene in each tissue sample was calculated, then log-transformed gene expression data were plotted using CLUSTER and MAPPLE TREE (Eisen *et al.*, 1998).

Phylogenetic analysis and gene co-expression profiling of the SDRLK subfamily

We aligned peptide sequences of 253 members of the *E. grandis* SDRLK subfamily (Myburg *et al.*, 2014) using MUSCLE, then further refined the results using GBLOCKS (Talavera & Castresana, 2007). We produced and rooted the phylogenetic tree of the SDRLKs using PHYML, which was imported into the online Interactive Tree Of Life (iTOL) tool (Letunic & Bork, 2007) to show gene expression profiles. We used BIOLAYOUT EXPRESS^{3D} (Theocharidis *et al.*, 2009) to generate and visualize weighted co-expression clusters using a correlation coefficient value of 0.85. To compare and visualize expression profiles of tandem duplicates, we examined expression data for 221 genes across all tissues, sorted data based on gene order, created a heatmap using MATRIX2PNG (Pavlidis & Noble, 2003), and performed multiple sequence alignment using CLUSTALW2 (Larkin *et al.*, 2007).

Results

Differential expression among tissues

To identify differentially expressed genes, we compared transcriptomes of seven different *E. grandis* tissues and developmental stages. The number of sequencing reads that mapped to the reference genome ranged from 33 692 935 in early floral buds to 171 749 948 in young leaves (Table S3). We detected expression of > 35 000 genes, corresponding to 98% of all annotated protein-coding genes (Table S4). CUFFDIFF identified 18604 genes that were differentially expressed between any two of the sampled tissues (Table S5), representing *c.* 51% of the 36 376 annotated protein-coding genes. In pairwise comparisons, 1% (young leaf versus mature leaf) to 43% (early floral buds versus roots) of genes were differentially expressed (Table S6). A subset of 15 180 genes were differentially expressed between one or both floral bud stages and any vegetative tissue. While some floral versus vegetative gene expression differences may be a result of differences between sites and tree ages, these effects are expected to be minor compared with differences in gene expression among tissues that are vastly different developmentally and morphologically.

Differentially expressed genes were subjected to hierarchical clustering after low-expression and fold-change cutoffs had been applied. This filtering resulted in a set of 7193 genes with highly differentiated expression between pairs of tissues (Fig. 2a). Of these, 4807 genes were differentially expressed between early or late floral buds and any vegetative tissue. Further, 479 of these genes were differentially expressed between early and late floral buds (Fig. 2b), with 462 more highly expressed in late floral buds compared with early floral buds, and 17 more highly expressed in early floral buds compared with late floral buds. The late-dominant cluster included 462 genes, and was enriched in GO categories related to floral development, organ development, and secondary metabolite production (Table 1, S7; Supporting Information Fig. S1). Forty of the late-dominant genes had no functional (GO) annotation. When the 17 genes that were significantly up-regulated in early floral buds relative to late were analyzed with respect to their expression in the full set of tissues, only six of these had higher expression in floral than in vegetative tissues, and only five had expression values with an FPKM > 5 (FPKM range 5.01–129.30). Two were similar to putative chloroplast proteins *RbcX* (an assembly chaperone of ribulose-bisphosphate carboxylase/oxygenase (Rubisco)) and *Lir1* (*LIGHT-INDUCED RICE1*), while the remaining three had the GO annotation 'oxidoreductase activity'. All five genes were from very large gene families (84–110 predicted homologs) in the annotated *E. grandis* Phytozome 9.1 database.

Transcription factors

Many of the differentially expressed genes were transcription factors. Hierarchical clustering of the set of 842 differentially expressed transcription factor genes revealed 265 transcription factors that were highly expressed in early and/or late floral buds relative to vegetative tissues (Fig. S2). Ten had the MADS-box transcription factor motif, and seven of those were on the list of candidate biomarkers described in the 'Floral Biomarkers' section. The non-MADS-box genes fell into 46 diverse transcription factor categories (Table S8).

MADS-box genes

Many of the key regulators of floral development are members of the MADS-box family of transcription factors (reviewed in Causier *et al.*, 2010; Litt & Kramer, 2010). A total of 107 MADS-box genes were identified in the *E. grandis* genome, many of which had previously been identified in related eucalypt species (Myburg *et al.*, 2014). Phylogenetic analysis of these genes, along with all putative MADS-box genes from *A. thaliana*, *P. trichocarpa* and *V. vinifera*, allowed us to obtain a robust classification for proteins within each subclade (MIKC (C-terminal type MADS-box genes), Mα, Mβ, Mγ, and Mδ). This analysis revealed 70 type II and 35 type I MADS-box genes. The high number of type II MADS-box genes in *E. grandis* was mainly a result of expansion of the *SOC1* subfamily (Myburg *et al.*, 2014). Fifteen *SOC2* genes were located in close proximity on chromosome 11. Tandem duplication events were also observed for several other genes (Myburg *et al.*, 2014).

To investigate the phylogenetic context of *E. grandis* MADS-box gene expression, we generated two trees: one containing only the MIKC subclade (Fig. 3), and another with the Mα, Mβ, Mγ, and Mδ subclades (Fig. S3). The MIKC and the Mα/Mβ/Mγ/Mδ trees showed 13 and 12 subfamilies, respectively. *Eucalyptus grandis* MADS-box genes were present in 12 MIKC (except the TT16 subfamily) and nine Mα/Mβ/Mγ/Mδ subclades (except

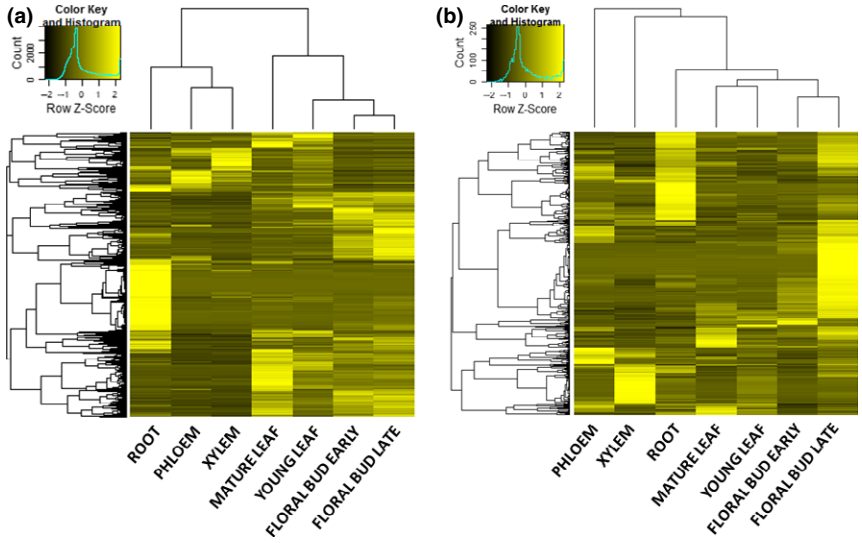


Fig. 2 Expression patterns of differentially expressed genes. Heat maps show expression profiles of differentially expressed genes. (a) 7193 genes differentially expressed between any two tissues. A low expression cutoff of five fragments per kilobase of exon per million fragments mapped (FPKM) in at least one tissue and fold-change cutoff of at least 4-fold difference between any two tissues were applied. (b) 479 genes differentially expressed between early and late floral buds. No FPKM or fold-change cutoffs were applied.

Table 1 Gene ontology (GO) categories enriched among late-floral-predominant genes

GO term	Ontology	Description	Number in input list	Number in BG/Ref	*P*-value	FDR
Floral development						
GO:0010584	P	Pollen exine formation	9	22	5.30E-11	4.40E-09
GO:0010208	P	Pollen wall assembly	9	25	1.30E-10	9.80E-09
GO:0048229	P	Gametophyte development	19	269	1.40E-09	8.90E-08
GO:0009555	P	Pollen development	17	209	1.40E-09	8.90E-08
GO:0080110	P	Sporopollenin biosynthetic process	5	5	4.80E-08	2.50E-06
GO:0048437	P	Floral organ development	10	134	7.30E-06	0.00027
GO:0048569	P	Post-embryonic organ development	10	135	7.70E-06	0.00028
GO:0009908	P	Flower development	12	297	0.00028	0.0048
GO:0048466	P	Androecium development	5	54	0.0006	0.0088
GO:0048443	P	Stamen development	5	54	0.0006	0.0088
Organ development						
GO:0071669	P	Plant-type cell wall organization or biogenesis	17	111	1.60E-13	3.10E-11
GO:0009808	P	Lignin metabolic process	14	82	6.00E-12	7.20E-10
GO:0042546	P	Cell wall biogenesis	13	69	1.20E-11	1.20E-09
GO:0048646	P	Anatomical structure formation involved in morphogenesis	15	112	2.40E-11	2.30E-09
GO:0009809	P	Lignin biosynthetic process	12	66	1.00E-10	8.20E-09
GO:0010927	P	Cellular component assembly involved in morphogenesis	9	25	1.30E-10	9.80E-09
GO:0045229	P	External encapsulating structure organization	9	36	2.00E-09	1.20E-07
GO:0007275	P	Multicellular organismal development	51	1763	4.80E-09	2.80E-07
GO:0048856	P	Anatomical structure development	45	1488	1.20E-08	6.30E-07
Secondary metabolite production						
GO:0019438	P	Aromatic compound biosynthetic process	30	318	1.70E-17	2.80E-14
GO:0006725	P	Cellular aromatic compound metabolic process	33	452	4.50E-16	2.50E-13
GO:0009698	P	Phenylpropanoid metabolic process	23	250	1.70E-13	3.10E-11
GO:0009699	P	Phenylpropanoid biosynthetic process	22	227	2.20E-13	3.30E-11
GO:0042398	P	Cellular amino acid derivative biosynthetic process	25	307	1.90E-13	3.10E-11
GO:0019748	P	Secondary metabolic process	30	545	9.00E-12	1.00E-09

GO categories were identified using the AgriGO Singular Enrichment Analysis tool. A subset of categories is shown; the full list is shown in Supporting Information Table S5. BG/Ref, background/reference; FDR, false discovery rate.

the AGL39/AGL74, AGL26/AGL103, and AGLS2 subfamilies). Floral-dominant genes were found in the AGL47/AGL82, AGLS1 and AGL29/AGL91 subfamilies (Fig. S3).

To compare MADS-box gene expression profiles in floral versus vegetative tissues, we performed hierarchical clustering and generated a digital northern blot and a heat map (Figs 4, S4). The MADS-box genes formed five clusters in which closely related genes displayed conserved expression profiles. Two clusters showed floral-dominant expression. The *EgSOC4* group was phloem dominant, and the *EgAGL57-L* group was root-dominant. Most of the *EgSOC2-L* genes were leaf-dominant. There were, however, also a number of exceptions to the general pattern of conserved expression among paralogs (Figs 3,4).

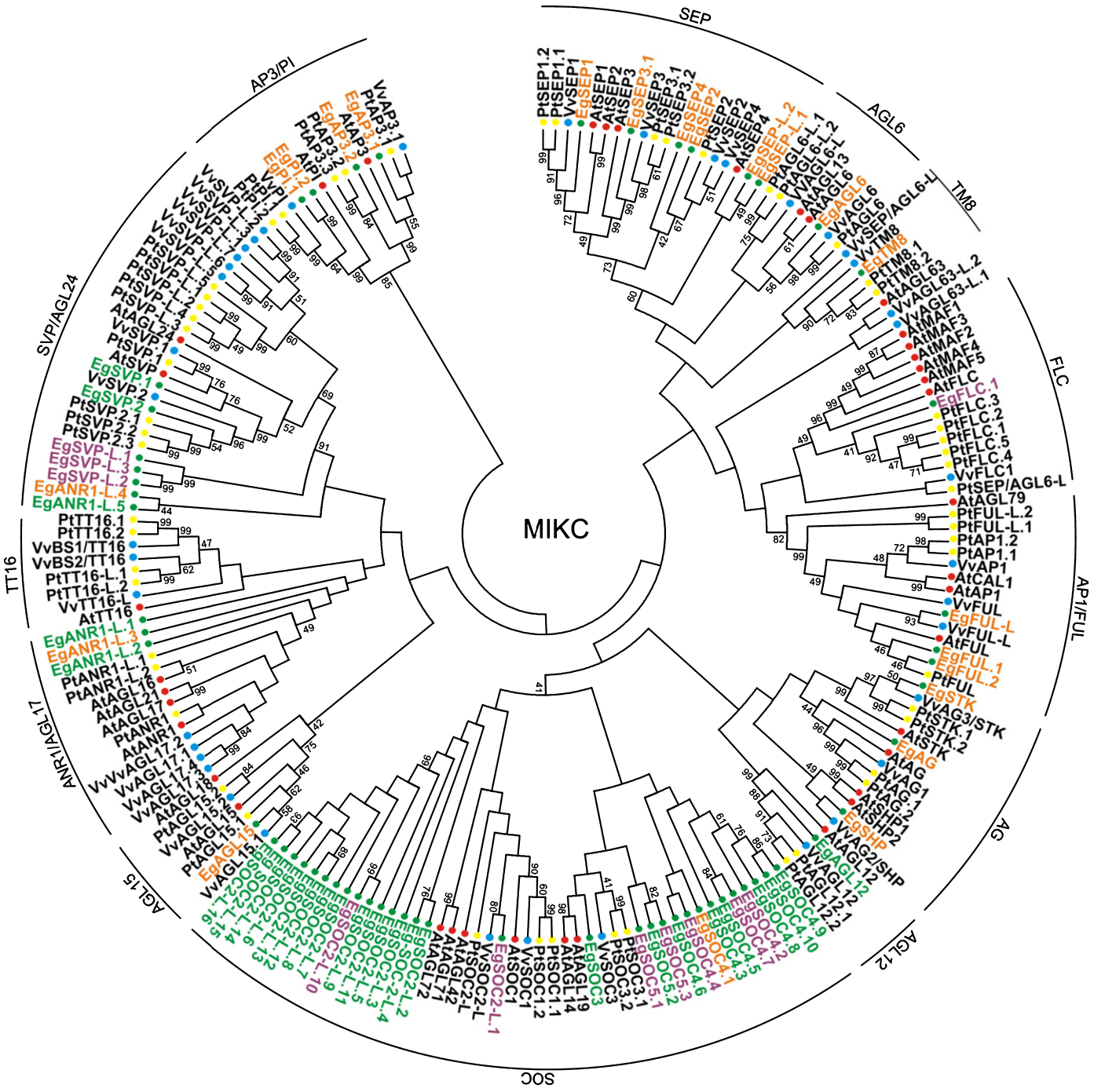


Fig. 3 Phylogenetic tree of type II MIKC MADS-box proteins of *Eucalyptus grandis*, *Arabidopsis thaliana*, poplar and grape. Neighbor-joining consensus trees using protein sequences from *E. grandis*, *A. thaliana*, *Populus trichocarpa* and *Vitis vinifera* are shown. Bootstrap values from 1000 replicates were used to assess the robustness of the tree. Bootstrap values < 40% were removed from the tree. Colors were chosen for *E. grandis* gene IDs according to a tissue predominance index calculated by comparing mean expression value over all vegetative samples (root, young leaf, mature leaf, immature xylem and phloem) to that of floral samples. Orange represents ≥ 75% total expression found in the floral sample, green represents ≥ 75% of total expression in vegetative samples, and purple represents expression < 75% of total in both floral and vegetative samples. Green dots, *E. grandis* genes; red dots, *A. thaliana* genes; yellow dots, poplar genes; blue dots, grape genes. The gene names for *E. grandis*, poplar and grape are abbreviated from 'Eucgr.' to 'Eg', from 'Potri.' to 'Pt' and from 'GSVIVT' to 'Vv', respectively, to better fit in the figure. *AP*, *APETALA*; *FUL*, *FRUITFUL*; *SEP*, *SEPALLATA*; *SOC*, *SUPPRESSOR OF OVEREXPRESSION OF CO*.

SDRLK genes

CuffDiff revealed differential expression for 233 of the 253 SDRLK genes. Co-expression analysis of these genes with BioLayout Express^{3D} revealed 11 distinct expression clusters and clear tissue predominance (clusters) that represented 212 genes (nodes) (Figs 5a–c, S5; Supporting Information Notes S1).

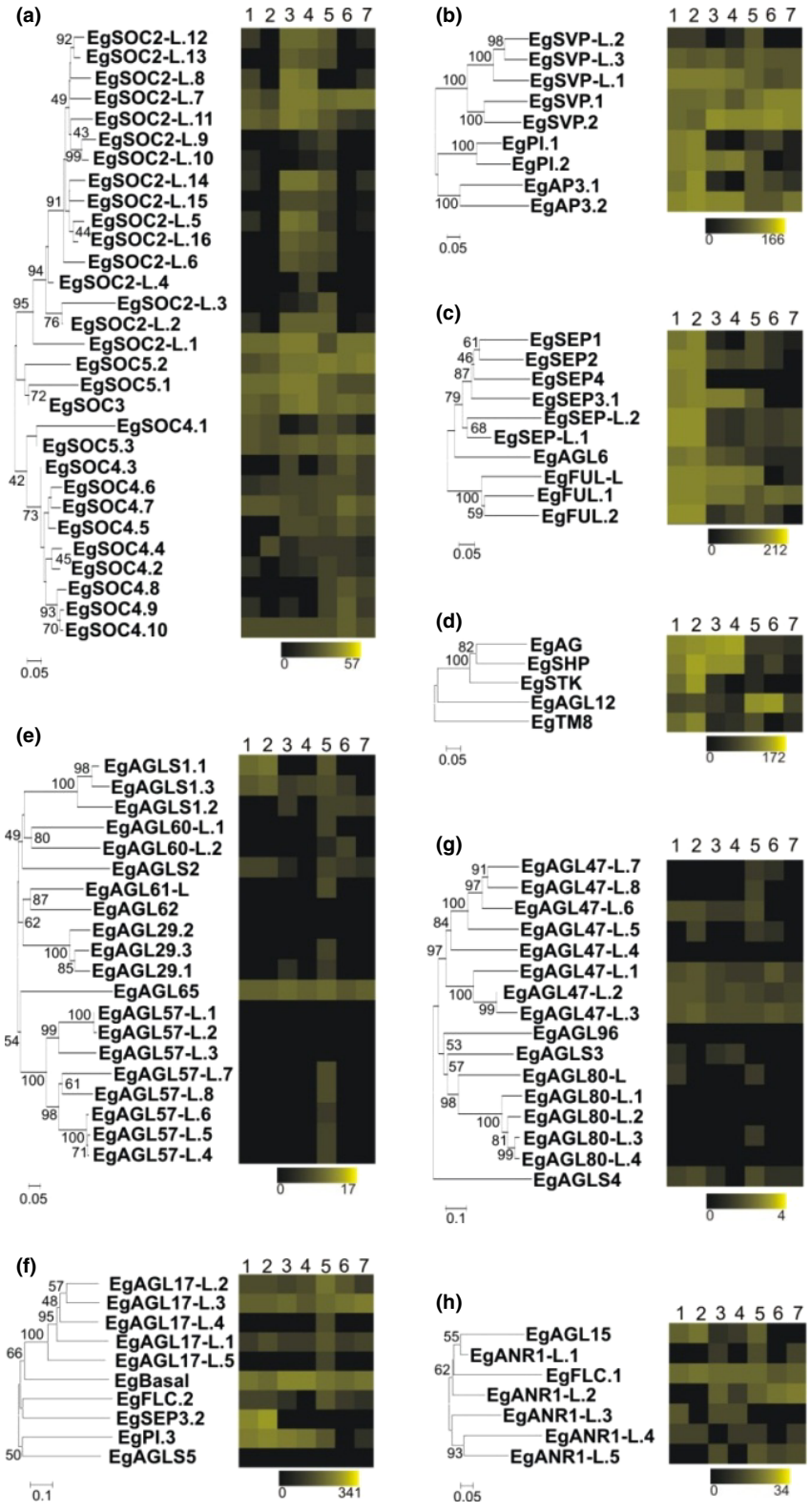


Fig. 4 Expression profiles of type II and type I MADS-box gene classes in *Eucalyptus grandis* during reproductive and vegetative development. (a–d, h) 'Digital northern' representing the expression profile of the type II MIKC MADS-box genes in *E. grandis*. (a) *SUPPRESSOR OF OVEREXPRESSION OF CO* (*SOC*) subfamilies. (b) *AP3/PI-SVP/AGL24* subfamilies. (c) *SEPALLATA* (*SEP*)-*AGL6-AP1- FRUITFUL* (*FUL*) subfamilies. (d) *AG* subfamilies. (e, g) 'Digital northerns' representing the expression profile of the type I MADS-box genes in *E. grandis*. (e) Mα and Mδ clades. (f) Putative truncated MADS-box-containing protein with only K-box domain. (g) Mβ and Mγ clades. (h) *AGL15/AGL18-FLC-ANR1/AGL17* subfamilies. Expression analyses were performed using RNA-seq data. Expression patterns were correlated to the phylogenetic neighbor-joining trees of MIKC and non-MIKC proteins. The normalized number of reads for each gene in each sample is represented by a yellow scale (below the gene expression profile). Each row corresponds to a gene, while the tissue samples are represented by the columns. Bootstrap values from 1000 replicates were used to assess the robustness of the tree. Bootstrap values < 40% were removed from the tree. The sample names are indicated by numbers: 1, early floral bud; 2, late floral bud; 3, young leaf; 4, mature leaf; 5, roots; 6, phloem; 7, immature xylem.

Approximately 55% of 212 SDRLK genes with similar expression profiles were tandem duplicates. In clusters 1–3, > 70% of genes were tandem duplicates; specifically, 33 of 47 genes in cluster 1, 29 of 41 genes in cluster 2, and 24 of 31 genes in cluster 3. Cluster 4 was an exception among large clusters, with < 50% (14 of 31 genes) being tandem duplicates. To study sequence–expression correlations, the gene expression profiles of 233 SDRLKs were plotted on the phylogenetic tree (Fig. S5). The results

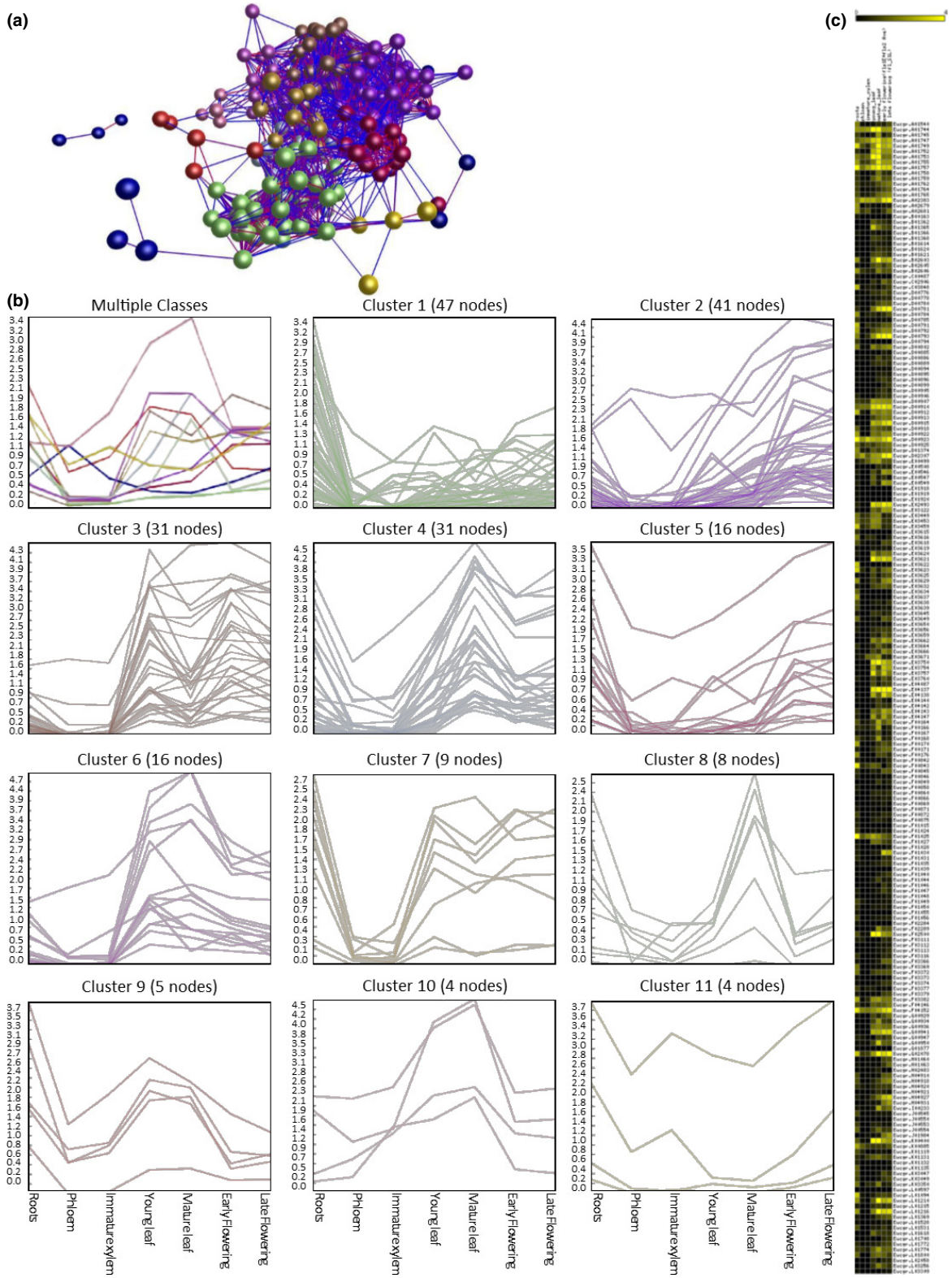


Fig. 5 Expression profiles of S-domain receptor-like kinase (SDRLK) genes in the vegetative and reproductive tissues of *Eucalyptus grandis*. A total of 233 SDRLK genes out of 253 showed differential expression in one or more tissues. (a) BioLayout Express^{3D} was used to detect co-expression patterns of SDRLK genes across the different tissues. (b) Eleven unique expression clusters representing 212 genes identified for SDRLKs are depicted. Each graph shows the number of associated genes (nodes) in the respective cluster and their expression profiles across seven tissues. (c) Heatmap shows expression of 212 SDRLKs across seven tissues and overlap in expression pattern between tandem duplicates. The rows and columns represent genes and tissues, respectively. The normalized mapped read count for an individual gene in a given tissue is represented on a black (low expression) to yellow (high expression) scale. Columns from left to right are roots, phloem, xylem, young leaf, mature leaf, early flower, and late flower.

showed that the most tandemly duplicated SDRLKs, evident as clades of recently diverged genes, shared high sequence similarity and a conserved expression profile.

Expression profiling of well-known floral genes

Expression profiles of *E. grandis* ABCDE model genes such as *AP1/FUL*, *SEP* and *AG* subfamilies were consistent with those of their putative *A. thaliana* orthologs. Most *EgFUL* genes had a wide-ranging expression pattern (Fig. 4c). All members of the *EgSEP* subfamily were expressed in early and late floral buds. In the *EgAG* subfamily, *EgAG* and *EgSHP* showed similar expression patterns to each other (Fig. 4d). *EgSTK* showed low expression in young leaves and roots, but no expression in mature leaves, phloem and immature xylem (Fig. 4d). When late and early floral buds were compared, an *AP3* ortholog (Eucgr.F01615-*EgAP3.1*) and a *SEP3* ortholog (Eucgr.B03515-*EgSEP3.2*) showed the strongest late floral dominance (Tables 2, S9).

Phylogenetic analysis of the *FT/TFL1* family revealed the presence of one locus for *TERMINAL-FLOWER-LIKE 1* (*TFL1*), *MOTHER OF FT* (*MFT*), *BROTHER OF FT AND TFL1* (*BFT*) and *FT*. By contrast, *ARABIDOPSIS THALIANA CENTRORADIALIS* (*ATC*) was apparently duplicated (Fig. 6a). These two loci were called *EgTFL3* (EgJ00485) and *EgTFL2.1* (EgB03867) based on *P. nigra* (Igasaki *et al.*, 2008) and *P. trichocarpa* annotations. In addition, our search of the *A. thaliana* genome revealed a new *FT/TFL1*-like gene (At5 g01300) and a eucalypt ortholog (EgA00595-*EgFT/TFL1-L*). *EgFT*, *EgTFL3* and *EgMFT* genes were mainly expressed in floral buds. By contrast, the *EgTFL2.1*, *EgBFT* and *EgFT/TFL1-L* genes were dominantly expressed in vegetative tissues (Fig. 6b,c).

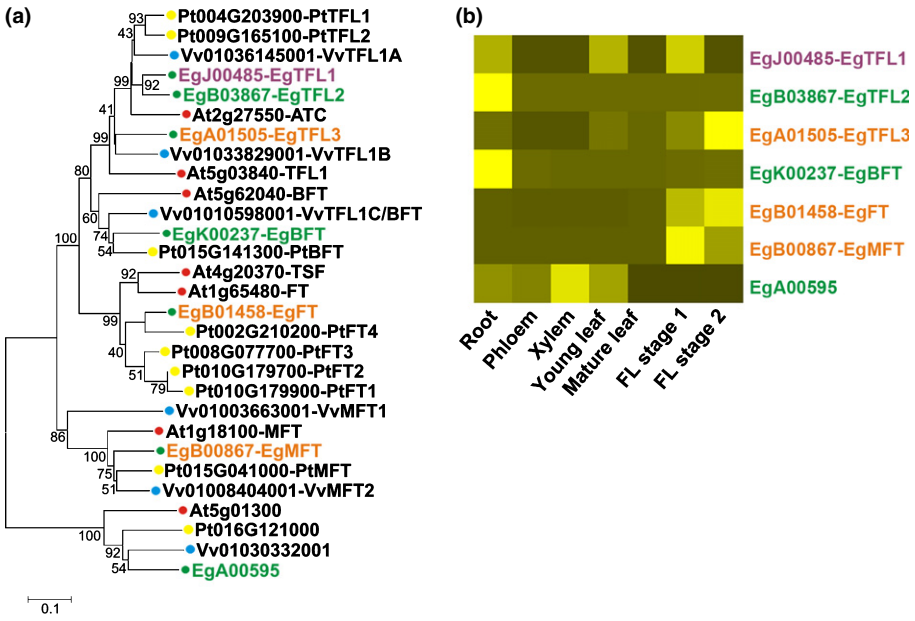
Floral biomarker genes

A set of 49 gene models were up-regulated either early or late in floral development, and thus are potential biomarkers (Table 2). Putative orthologs of many of the key regulators of flowering onset in *A. thaliana*, such as *FT*, *FD* (*bZIP transcription factor FD*), *FUL* and *FLC* (*FLOWERING LOCUS C*), had either nondifferential patterns of expression in early versus late flower buds or were not expressed in these tissue types (Table S10). Two genes, *EgSVP.1* and *EgLFY*, were up-regulated in early compared to late flower development (Table 2; Fig. 7), although the difference was not statistically significant. There were 47 gene models that could serve as biomarkers for late flower development. These represented *E. grandis* homologs of B- and C-class genes in the ABCDE model of flower development (e.g. *EgPI.1*, *EgPI.2*, *EgPI.3* and *EgAG*), genes involved in ovule development (e.g. *BEL1* (*BELL1*), *INO* (*INNER NO OUTER*) and *EgSTK*), and genes involved in anther or tapetum development (e.g. *ACOS5*, *AMS*, *CYP703A2*, *TDF1* and *WBC27*) (Fig. 7). Genes specifically involved in anther or tapetum development displayed the clearest up-regulation: transcripts of these genes were undetected or extremely low in early floral buds (Fig. 7), when stamen primordia were first visible (Fig. 1), but up-regulated in the late stage of floral bud development, coinciding with active anther and stamen development (Figs 1,7).

Table 2 Expression of *Eucalyptus grandis* putative flowering biomarkers in buds harvested late versus early in floral bud development

Gene	*E. grandis* homolog (v1.1)	Early FPKM	Late FPKM	Late: early
Early biomarkers				
LFY	Eucgr.K02192	40.8	21.3	0.5
EgSVP.2	Eucgr.F00420	11.2	5.6	0.5
Late biomarkers				
ACOS5	Eucgr.B03502	2.4	71.1	29.9*
EgAG	Eucgr.E02863	67.4	172.2	2.6
EgAGL6	Eucgr.E01330	25.9	56.8	2.2*
AMS.1	Eucgr.G01783	0.3	55.2	195.8*
AMS.2	Eucgr.H02602	10.1	66.2	6.6*
AP2	Eucgr.I00892	23.0	42.9	1.9*
EgAP3.1	Eucgr.F01615	24.1	166.5	6.9*
EgAP3.2	Eucgr.I02376	67.7	151.1	2.2
BEL1	Eucgr.D01826	11.4	26.5	2.3
CYP703A2	Eucgr.D02382	0.0	12.9	12.9/0*
CYP94B3	Eucgr.I01651	1.5	3.7	2.5
ERL1	Eucgr.K00138	8.8	14.3	1.6
GRI	Eucgr.F01147	0.0	37.1	37.1/0*
HTH	Eucgr.F03132	48.3	125.7	2.6*
HTR12	Eucgr.D00189	6.7	12.2	1.8
INO	Eucgr.B03310	0.4	3.3	8.9
JAG	Eucgr.G02316	5.1	13.7	2.7*
MYB26	Eucgr.F03143	1.5	16.0	10.5*
NAP.1	Eucgr.B03208	1.4	5.2	3.7
NAP.2	Eucgr.G02486	18.2	40.6	2.2
EgPI.1	Eucgr.E01007	31.8	74.9	2.4
EgPI.2	Eucgr.E01006	51.0	85.3	1.7
EgPI.3	Eucgr.D00249	43.9	67.9	1.5
PI4K GAMMA 1	Eucgr.K01678	2.6	5.8	2.3
RPK2	Eucgr.A00983	11.3	16.6	1.5
EgSEP1	Eucgr.I02058	23.0	69.6	3.0*
EgSEP2	Eucgr.K02546	67.4	109.7	1.6
EgSEP3.1	Eucgr.H04617	46.7	87.1	1.9
EgSEP3.2	Eucgr.B03515	68.2	341.6	5.0*
EgSEP4	Eucgr.B00633	32.1	79.1	2.5*
EgSHP	Eucgr.K01195	41.4	85.4	2.1*
STIG1	Eucgr.E04008	0.0	9.4	9.4/0
EgSTK	Eucgr.F02981	9.0	27.7	3.1*
STY1	Eucgr.I02380	15.8	24.4	1.5
TDF1	Eucgr.I02017	0.0	8.7	8.7/0
WBC27	Eucgr.F03676	0.2	11.7	58.8*
WUS	Eucgr.J02429	0.0	6.0	6.0/0
AT1G24620	Eucgr.B02672	9.0	15.9	1.8
AT1G58120	Eucgr.I02628	6.0	14.0	2.3
AT1G68540	Eucgr.G02325	16.7	43.3	2.6*
AT2G26490	Eucgr.F01502	1.0	8.5	8.6*
AT2G46210	Eucgr.C00655	35.8	65.9	1.8
AT2G46210	Eucgr.D02331	38.2	63.0	1.6
AT3G07450	Eucgr.J01258	1.8	368.8	204.6*
AT3G43120	Eucgr.C03209	7.0	14.1	2.0
AT4G30030	Eucgr.K00305	0.4	7.0	18.8*
AT5G16920	Eucgr.K00086	1.4	21.9	16.0*

Genes with 'Eg' prefix are orthologs confirmed by phylogenetic analyses (Figs 3,6); others are putative orthologs of *Arabidopsis thaliana* genes. FPKM, fragments per kilobase of exon per million fragments mapped. Genes with ratio values less than unity show increased expression during early stages of flower development; genes with ratio values greater than unity show increased expression during later stages. Values marked with an asterisk are statistically significant (CuffDiff analysis; $q < 0.05$). ACOS5, ACYL-COA SYNTHASE; AG, AGAMOUS; AGL, AGAMOUS-LIKE; AM, ABORTED MICROSPORES; AP2, APETALA2; AP3, APETALA3; BEL, BELL; CYP, CYTOCHROME P450; ERL, ERECTA-LIKE; GRI, GRAVITROPIC RESPONSE INDICATOR; HTH, HOTHEAD; HTR, HISTONE3-RELATED; INO, INNER NO OUTER; JAG, JAGGED; LFY, LEAFY; MYB, MYB-DOMAIN; NAP, NUCLEOSOME ASSEMBLY PROTEIN; PI, PISTILLATA; PI4K GAMMA 1, phosphoinositide 4-kinase gamma 1; RPK, RECEPTOR PROTEIN KINASE; SEP, SEPALLATA; SHP, SHATTERPROOF; STIG, STIGMA-SPECIFIC; STK, SEEDSTICK; STY, STYLISH; SVP, SHORT VEGETATIVE PHASE; TDF, DEFECTIVE IN MERISTEM DEVELOPMENT AND FUNCTION; WBC, White-Brown Complex homolog protein; WUS, WUSCHEL.



(c)

Gene	ROOT	PHLOEM	XYLEM	YOUNG LEAF	MATURE LEAF	FL STAGE 1	FL STAGE 2
EgJ00485-EgTFL1	0.33	0.00	0.00	0.34	0.00	0.44	0.00
EgB03867-EgTFL2	0.14	0.00	0.00	0.00	0.00	0.00	0.00
EgA01505-EgTFL3	0.72	0.00	0.00	1.00	0.52	1.65	5.38
EgK00237-EgBFT	68.84	0.59	0.07	0.04	0.00	1.93	2.04
EgB01458-EgFT	0.28	0.00	0.15	0.17	0.95	18.45	26.99
EgB00867-EgMFT	0.00	0.00	0.00	0.00	0.00	1.80	0.74
EgA00595	13.47	10.47	29.76	17.09	0.00	0.00	0.00

Fig. 6 Phylogenetic relationships among *Eucalyptus grandis*, *Arabidopsis thaliana*, poplar and grape members of the *FLOWERING LOCUS T* (*FT*)/*TFL1* gene family. (a) Neighbor-joining tree of *E. grandis*, *A. thaliana*, *Populus trichocarpa* and *Vitis vinifera FT*/*TFL1* gene family containing phosphatidylethanolamine-binding proteins (PEBPs). Bootstrap values from 1000 replicates were used to assess the robustness of the tree. The scale indicates the average number of substitutions per site. The colors chosen for the *E. grandis* gene code (ID) were based on the per cent index calculated based on a mean value over all vegetative samples (root, young leaf, mature leaf, immature xylem and phloem) and floral samples (early and late flower development). Gene names are colored according to their expression pattern. Orange color represents 75% total expression found in the floral sample, green represents 75% over-expression in vegetative samples, and purple represents expression < 75% in either floral and vegetative samples. *Eucalyptus grandis*, *A. thaliana*, poplar and grape genes are denoted by green, red, yellow and blue dots, respectively. The gene names for *E. grandis* are abbreviated from 'Eucgr.' to 'Eg'. (b) Heatmap showing relative expression of *E. grandis FT*/*TFL1* genes in the sampled tissues (yellow, high; gray, low). (c) Tabular summary of expression values used to generate the heat map, with strongly differentially expressed gene/tissue combinations highlighted in gray.

Expression of one early candidate biomarker (Eucgr.K02192, *LFY*) and two late candidate biomarkers (Eucgr.E02863, *EgAG* and Eucgr.F02981, *EgSTK*) was tested using qRT-PCR ($2^{-\Delta\Delta C_t}$ method) with cDNA from early and late floral buds. Relative expression profiles of the three genes agreed with RNA-seq results (Fig. S6, Table S11). The relative quantity of *LFY* was *c.* 2.6-fold greater in early compared with late floral buds (FPKM values were 40.8 versus 21.3). For *EgAG* and *EgSTK*, relative quantities were 4.8-fold (FPKM 172.1 in late and 67.4 in early floral buds) and 9.6-fold (FPKM 27.7 in late and 9.0 in early floral buds), respectively.

Discussion

As anticipated, there were strong similarities between the identities of floral genes and their expression profiles in *E. grandis* and other plants; however, there were also important differences. Most genes up-regulated in early flowers were also expressed in other tissues, particularly in young and mature leaves (Fig. 2), perhaps as a consequence of the presence of leaf-like bracts on the floral samples (Fig. 1). Only five genes were up-regulated specifically in early floral buds, and two of these were putative chloroplast genes. The other three were putative proteins with 'oxidoreductase activity' Pfam annotations. This result is similar to a finding in *A. thaliana*, where a study of the floral transcriptome found an overall low number (*c.* 200) of genes up-regulated in the shoot apical meristem during floral transition, and overrepresentation of genes related to oxidation-reduction.(Torti *et al.*, 2012). Members of the glutaredoxin oxidoreductase family, ROXYs, have also been shown to be required for normal anther development in *A. thaliana* (Xing *et al.*, 2011).

We found that the late-floral-dominant gene set was enriched in genes associated with secondary metabolism, specifically aromatic compound biosynthesis. Eucalypts possess diverse plant secondary metabolite profiles, which vary among organs and developmental stages as well as within and between species (Li *et al.*, 1995, 1996; Bignell *et al.*, 1998; O'Reilly-Wapstra *et al.*, 2010; McKiernan *et al.*, 2012). Moreover, eucalypt floral buds

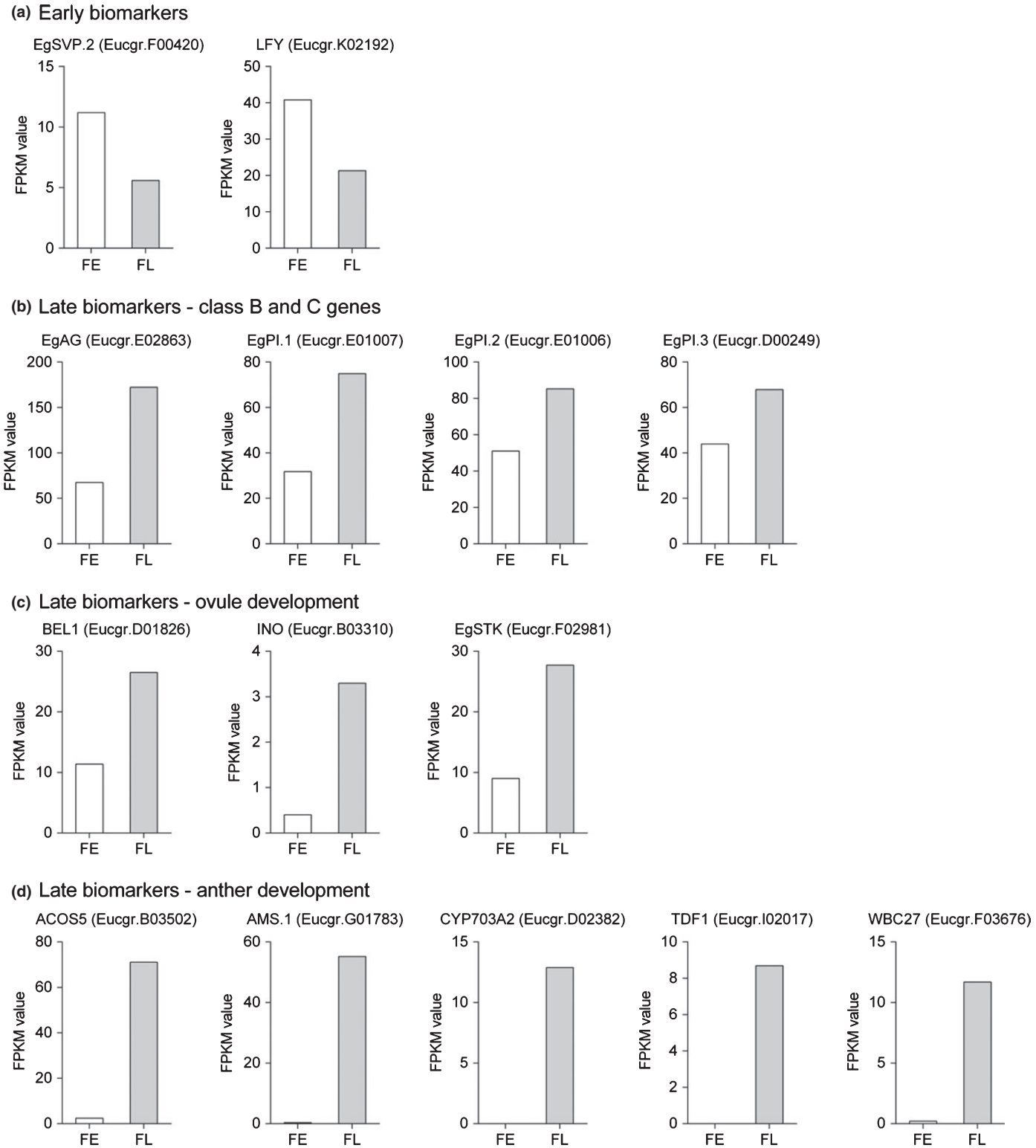


Fig. 7 Examples of expression patterns of *Eucalyptus grandis* gene models orthologous to biomarkers of (a) early and (b–d) late flower development in *Arabidopsis thaliana.* Late biomarkers are grouped into genes of similar function. Fragments per kilobase of exon per million fragments mapped (FPKM) from *E. grandis* flower buds harvested early (FE) and late (FL) in floral bud development are shown. Genes with 'Eg' prefix are orthologs confirmed by phylogenetic analyses (Fig. 3); others are putative orthologs of *A. thaliana* genes. Data for the full set of biomarkers are given in Table 2. ACOS, ACETYL COA SYNTHASE; AG, AGAMOUS; AMS, ABORTED MICROSPORES; BEL1, BELL1; CYP703A2, cytochrome P450, family 703, subfamily A, polypeptide 2; INO, INNER NO OUTER; LFY, LEAFY; PI, PISTILLATA; STK, SEEDSTICK; SVP, SHORT VEGETATIVE PHASE; TDF, DEFECTIVE IN TAPETAL DEVELOPMENT AND FUNCTION; WBC, WHITE BROWN PROTEIN COMPLEX HOMOLOG PROTEIN.

and expanded flowers are rich in oil glands; the terpenes produced there are well known to play a role in attracting pollinators as well as in defense against herbivores (Theis & Lerdau, 2003).

Five per cent of the 4807 genes that were differentially expressed between floral and vegetative tissues were predicted transcription factors. These genes were from a variety of families (Table S8), some with documented roles in floral development. For example, a member of the NAC2 (NAC DOMAIN PROTEIN) family is expressed in *A. thaliana* ovule integuments (Kunieda *et al.*, 2008), and three MYB (MYB DOMAIN PROTEIN) transcription factors (MYB21, MYB24, and MYB57) are implicated in stamen filament elongation (Cheng *et al.*, 2009); the up-regulated *E. grandis* NAC and MYB family member genes may have similar functions. Ten of the differentially expressed genes (either floral tissue versus any vegetative tissue) were members of the MADS-box family of transcription factors. These included *E. grandis* homologs of key genes such as *SEP* and *API/FUL* (Fig. 3). This MADS-box gene family is expanded in *E. grandis*, with 107 members, including 70 type II and 35 type I. Compared with *A. thaliana* (also Eurosid II) (Bremer *et al.*, 2009), poplar (Eurosid I) and grape (Rosid), the increase of type II genes in *E. grandis* was surprising.

Interestingly, while the *E. grandis* genome contains an increased number of type II MADS-box genes, we were unable to identify a single putative ortholog of *SOC1*; however, nearly half of the type II MADS-box genes belonged to the *SOC* subfamily. This *SOC* subfamily expansion, mainly attributable to tandem duplications, is a unique event among angiosperms with available genome sequences. The *SOC* subfamily was also expanded in Norway spruce (*Picea abies*), a gymnosperm (Nystedt *et al.*, 2013). Nystedt *et al.* (2013) hypothesized that the large number of *SOC*-like genes may have been an important factor in the evolution of developmental phase change in gymnosperms. Much research has focused on understanding the role of *SOC1* as an integrator of multiple flowering signals derived from different internal and external cues (Lee & Lee, 2010). In *A. thaliana*, other *SOC* subfamily genes (*AGL71*/*AGL72* and *AGL19*) are involved in control of the onset of flowering (Dorca-Fornell *et al.*, 2011; Kim *et al.*, 2012). Members of the *EgSOC2-L* group are proposed to act as functional orthologs of *AtSOC1* in eucalypts (Watson & Brill, 2004), and play a role in control of floral onset. We found that most members of the expanded *SOC2-L* subgroup were predominantly expressed in mature and young leaves (Figs 4a,S5). Similarly, *AtSOC1* is also expressed in leaves (Parcy, 2005). Eucalypts contain *SOC4* and *SOC5*, two species-specific groups of genes. The *SOC5* group had three members (*EgSOC5.1–3*), each with a distinct expression profile (Figs 4a,S5). Six members of the *SOC4* group were expressed in all tested tissues, with some members showing relatively high expression in phloem (Figs 4a,S5). The expression patterns of the distinct *SOC* groups could be a result of regulatory hypofunctionalization or neofunctionalization (Lawton-Rauh, 2003). The large number of *SOC* genes in the *Eucalyptus* genome may be important for adaptation to a subtropical climate, where reproductive timing depends on a broad range of environmental cues (Myburg *et al.*, 2014).

Members of the *FT/TFL1* family of genes are important regulators of floral timing. We found one *E. grandis FT* gene (EgB01458-*EgFT*) and seven *FT/TFL1*-like genes (Fig. 6). Many other tree species have multiple *FT*-like genes with divergent functions, for example, apple *Malus* × *domestica* Borkh. (Kotoda *et al.*, 2010), and the gymnosperm Norway spruce (Gyllenstrand *et al.*, 2007). Analysis of gene expression showed that the *E. grandis FT*-like genes displayed a variety of expression patterns, with three (EgA01505-*EgTFL3*, EgB01458-*EgFT* and EgB00867-*EgMFT*) up-regulated in floral tissues. *FT* is part of a multigene family, some members of which, like *FT* itself, promote flowering, while others, such as *TFL*, repress the onset of flowering (Ohshima *et al.*, 1997) or have roles outside of floral regulation (Pin & Nilsson, 2012). Interestingly, expression of *A. thaliana FT* (*AtFT*) accelerated flowering in hybrid *E. grandis* × *Eucalyptus urophylla* plants, while untransformed controls showed no flower formation (A. Klocko & S. H. Strauss, unpublished results), indicating conservation of an *FT*-responsive signaling pathway.

Floral development and floral organ identity are controlled by the so-called ABC model (Coen & Meyerowitz, 1991) and its further elaborations (Krizek & Fletcher, 2005; Litt & Kramer, 2010). We found that, while *E. grandis* contains homologs of genes from all of these gene classes (Fig. 3, Table S10), there were some important differences in the numbers of family members and their expression patterns. For example, we found that *E. grandis* lacks an archetypical *AP1* homolog, a class A gene important for floral meristem determination and floral organ identity, and contains a pair of *FUL*-like genes (*EgFUL.1* and *EgFUL.2*). These genes are members of two major eudicot lineages: *euAP1* (which includes *AtAP1* and *AtCAL*) and *euFUL* (which includes *AtFUL* and *EAP1*/*EAP2*), which originated from duplication events predating the divergence of the core eudicots (Litt & Irish, 2003a; Shan *et al.*, 2007). These duplications are correlated with the eudicots' diversification and fixed floral architecture (Litt & Irish, 2003a). Interestingly, the eucalypt *euFUL* gene is able to complement *A. thaliana ap1* mutants (Kyozuka *et al.*, 1997). These findings suggest that the paralogs *EgFUL.1* and/or *EgFUL.2* fulfill the functions of AP1 and FUL in eucalypts. This result is corroborated by the expression of the *FUL* homologs in the early stages of *E. grandis* flower development (Table S10). The absence of an archetypical *AP1* homolog in the *E. grandis* genome may help to explain the distinct floral external whorl structures in *Eucalyptus*, such as the calycine operculum (fused sepals) and corolline operculum (fused petals).

We also found that *E. grandis* had distinct BCE gene expression patterns from those of *A. thaliana*, and multiple copies of several B (including *AP3* and *PI*) and C (*AG*) class genes (Table S10). In contrast to the *A. thaliana* homologs (Duarte *et al.*, 2006), expression of *EgAP3*, *EgAG* and *EgSTK* was not restricted to floral buds (Table S5). *EgAG* was expressed at similar levels in mature leaves as it was in early floral buds, and only 3.8-fold more in late floral buds as in the mean of the two leaf tissues (Table S11). The broad expression of eucalypt BCE genes may

be characteristic of the ancestral condition of the angiosperms (Yoo *et al.*, 2010). The expression of the *Eucalyptus EgFUL* genes and BCE genes in leaves may also indicate an additional role in vegetative development.

We identified a number of *E. grandis* genes that could be useful floral biomarkers (Fig. 7) and used qPCR to verify the general expression changes of three selected biomarker genes (*LFY*, *AG* and *STK*) in early and late floral buds. *LFY* regulates floral initiation (Weigel *et al.*, 1992) and activates the floral homeotic genes (Weigel & Meyerowitz, 1994). As expected, the putative *E. grandis* ortholog of *LFY* was expressed early in floral bud development, making it an ideal early floral biomarker (Fig. 7). *AG* and *STK* are important for inner floral organ (stamen and carpel) and ovule development, respectively. We found that the *E. grandis* orthologs of *AG* and *STK* were more highly expressed in late floral buds (Fig. S7). Other possible late-floral biomarkers belong to the *E. grandis* orthologs of B- and C-class genes in the ABCDE model (e.g. *EgPI.1*, *EgPI.2*, *EgPI.3* and *EgAG*; see Fig. 7), and putative orthologs of genes involved in ovule development (e.g. *BEL1*, *INO* and *EgSTK*), and anther or tapetum development (e.g. *ACOS5*, *AMS*, *CYP703A2*, *TDF1* and *WBC27*). In particular, the genes involved in anther/tapetum development displayed the clearest up-regulation, and are perhaps the best late floral biomarkers. These genes were previously characterized in other species as being involved in later stages of tapetum and pollen development, such as pollen wall formation (Sorensen *et al.*, 2003; Zhu *et al.*, 2008; de Azevedo Souza *et al.*, 2009).

The 253 members of the *E. grandis* SDRLK family include RLKs and nonreceptor proteins (Myburg *et al.*, 2014) that share sequence and structural similarity with SRK, the female determinant of self-incompatibility in *Brassica* (Naithani *et al.*, 2007). The majority of SDRLK genes in *E. grandis* can be divided into 11 discrete groups based on their expression profiles (Fig. S6), including groups with predominant expression in roots, leaves, flowers, or multiple tissues. A large number of *E. grandis* SDRLK genes are expressed in floral tissue, but only a few showed floral-dominant expression (Table S5). Similar observations have been made in *A. thaliana* and rice, in which SDRLKs showed higher expression in flowers relative to vegetative tissues, but most were expressed in multiple tissue types (Yamada *et al.*, 2003; Gao & Xue, 2012). Both pre- and post-zygotic level self-incompatibilities operate in many species of *Eucalyptus*, including *E. grandis* (Sedgley & Smith, 1989). Late-acting self-incompatibility has been observed in both *E. urophylla* and *E. grandis*, where either self-pollen tubes fail to fertilize the ovules or most self-fertilized ovules degenerate and embryos are aborted (Pound *et al.*, 2002; Horsley & Johnson, 2007). Thus, SDRLKs are unlikely to act as determinants of self-incompatibility at the pollen–pistil recognition step in *E. grandis* but could participate in processes regulating its mating system.

We have presented several large data sets that together describe gene expression for the major tissue types in *E. grandis*. Further, we examined the evolution of several gene families important to vegetative and reproductive development. Within these families we identified groups of genes whose expression patterns suggest novel biological functions. These genes would be logical targets for functional studies using cytological, genetic, and transgenic approaches. The biomarker data in particular provide a guide to selection of promoters that could be useful for controlling gene expression during functional studies, as well as in efforts to modify diverse traits, including plant fertility. Expected outcomes of this work are new insights into the adaptation, evolution, and development of this extraordinarily successful and important taxon, and thus acceleration of progress in eucalypt breeding and biotechnology.

Acknowledgements

We thank Nelisha Murugan for performing tissue sectioning and microscopy and Andrew Drinnan for assisting with interpretation of floral micrographs; R. Harcourt, D. Llewellyn and E.S. Dennis for EST sequences of *AGE1* and *AGE2* from *E. globulus*; the Churchill Fellowship to R.C.J. for her working visit to OSU, fellowship (150690/2012-6 (PDJ)) and the Conselho Nacional de Pesquisa de Desenvolvimento Científico e Tecnológico (CNPq) of Brazil for financial support for E.R.; the industrial members of the Tree Biosafety and Genomics Research Cooperative (TBGRC) at Oregon State University and the National Science Foundation I/UCRC Center for Advanced Forestry for partial support of S.H.S. The contributions of S.N. and P.J. were supported by Oregon State University startup funds.

References

Alves-Ferreira M, Wellmer F, Banhara A, Kumar V, Riechmann JL, Meyerowitz EM. 2007. Global expression profiling applied to the analysis of *Arabidopsis* stamen development. *Plant Physiology* **145**: 747–762.

Amasino R. 2010. Seasonal and developmental timing of flowering. *Plant Journal* **61**: 1001–1013.

de Azevedo Souza C, Kim SS, Koch S, Kienow L, Schneider K, McKim SM, Haughn GW, Kombrink E, Douglas CJ. 2009. A novel fatty Acyl-CoA synthetase is required for pollen development and sporopollenin biosynthesis in *Arabidopsis*. *Plant Cell* **21**: 507–525.

Bignell CM, Dunlop PJ, Brophy JJ. 1998. Volatile leaf oils of some south-western and southern Australian species of the genus *Eucalyptus* (series 1). Part XIX. *Flavour and Fragrance Journal* **13**: 131–139.

Bremer B, Bremer K, Chase MW, Fay MF, Reveal JL, Soltis DE, Soltis PS, Stevens PF, Anderberg AA, Moore MJ *et al.* 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* **161**: 105–121.

Brill EM, Watson JM. 2004. Ectopic expression of a *Eucalyptus grandis* SVP orthologue alters the flowering time of *Arabidopsis thaliana*. *Functional Plant Biology* **31**: 217–224.

Carmona MJ, Calonje M, Martínez-Zapater JM. 2007. The *FT/TFL1* gene family in grapevine. *Plant Molecular Biology* **63**: 637–650.

Causier B, Schwarz-Sommer Z, Davies B. 2010. Floral organ identity: 20 years of ABCs. *Seminars in Cell & Developmental Biology* **21**: 73–79.

Chang S, Puryear J, Cairney J. 1993. A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter* **11**: 113–116.

Cheng H, Song S, Xiao L, Soo HM, Cheng Z, Xie D, Peng J. 2009. Gibberellin acts through jasmonate to control the expression of *MYB21*, *MYB24*, and *MYB57* to promote stamen filament growth in *Arabidopsis*. *PLoS Genetics* **5**: e1000440.

Coen ES, Meyerowitz EM. 1991. The war of the whorls: genetic interactions controlling flower development. *Nature* **353**: 31–37.

Colcombet J, Boisson-Dernier A, Ros-Palau R, Vera CE, Schroeder JI. 2005. *Arabidopsis* SOMATIC EMBRYOGENESIS RECEPTOR KINASES1 and 2 are essential for tapetum development and microspore maturation. *Plant Cell* **17**: 3350–3361.

De Smet I, Voss U, Jürgens G, Beeckman T. 2009. Receptor-like kinases shape the plant. *Nature Cell Biology* **11**: 1166–1173.

Dorca-Fornell C, Gregis V, Grandi V, Coupland G, Colombo L, Kater MM. 2011. The *Arabidopsis SOC1*-like genes *AGL42*, *AGL71* and *AGL72* promote flowering in the shoot apical and axillary meristems. *Plant Journal* **67**: 1006–1017.

Dornelas MC, Neves do Amaral WA, Rodriguez APM. 2004. *EgLFY*, the *Eucalyptus grandis* homolog of the *Arabidopsis* gene *LEAFY* is expressed in reproductive and vegetative tissues. *Brazilian Journal of Plant Physiology* **16**: 105–114.

Du Z, Zhou X, Ling Y, Zhang Z, Su Z. 2010. agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Research* **38**: W64–W70.

Duarte JM, Cui LY, Wall PK, Zhang Q, Zhang XH, Leebens-Mack J, Ma H, Altman N, dePamphilis CW. 2006. Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of *Arabidopsis*. *Molecular Biology and Evolution* **23**: 469–478.

Eisen MB, Spellman PT, Brown PO, Botstein D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences, USA* **95**: 14863–14868.

Escobar-Restrepo J-M, Huck N, Kessler S, Gagliardini V, Gheyselinck J, Yang W-C, Grossniklaus U. 2007. The FERONIA receptor-like kinase mediates male-female interactions during pollen tube reception. *Science* **317**: 656–660.

Flowers JM, Hanzawa Y, Hall MC, Moore RC, Purugganan MD. 2009. Population genomics of the *Arabidopsis thaliana* flowering time gene network. *Molecular Biology and Evolution* **26**: 2475–2486.

Gao L-L, Xue H-W. 2012. Global analysis of expression profiles of rice receptor-like kinase genes. *Molecular Plant* **5**: 143–153.

Gregis V, Sessa A, Colombo L, Kater MM. 2006. *AGL24*, *SHORT VEGETATIVE PHASE*, and *APETALA1* redundantly control *AGAMOUS* during early stages of flower development in *Arabidopsis*. *Plant Cell* **18**: 1373–1382.

Gyllenstrand N, Clapham D, Källman T, Lagercrantz U. 2007. A Norway spruce *FLOWERING LOCUS T* homolog is implicated in control of growth rhythm in conifers. *Plant Physiology* **144**: 248–257.

Higgins JA, Bailey PC, Laurie DA. 2010. Comparative genomics of flowering time pathways using *Brachypodium distachyon* as a model for the temperate grasses. *PLoS ONE* **5**: e10065.

Hodgson LM. 1976. Some aspects of flowering and reproductive behaviour in *Eucalyptus grandis* (Hill) Maiden at J.D.M. Keet Forest Research Station (formerly Zomerkomst Forest Research Station). *South African Forestry Journal* **97**: 18–28.

Horsley TN, Johnson SD. 2007. Is *Eucalyptus* cryptically self-incompatible? *Annals of Botany* **100**: 1373–1378.

Igasaki T, Watanabe Y, Nishiguchi M, Kotoda N. 2008. The *FLOWERING LOCUS T/TERMINAL FLOWER 1* family in Lombardy poplar. *Plant and Cell Physiology* **49**: 291–300.

Izawa T, Takahashi Y, Yano M. 2003. Comparative biology comes into bloom: genomic and genetic comparison of flowering pathways in rice and *Arabidopsis*. *Current Opinion in Plant Biology* **6**: 113–120.

Jaya E, Clemens J, Song JC, Zhang HB, Jameson PE. 2010. Quantitative expression analysis of meristem identity genes in *Eucalyptus occidentalis*: *AP1* is an expression marker for flowering. *Tree Physiology* **30**: 304–312.

Jaya E, Song J, Clemens J, Jameson PE. 2011. Effect of environment and shoot architecture on floral transition and gene expression in *Eucalyptus occidentalis* and *Metrosideros excelsa*. *Plant Growth Regulation* **64**: 53–61.

Jones RC, Hecht VFG, Potts BM, Vaillancourt RE, Weller JL. 2011a. Expression of a *FLOWERING LOCUS T* homologue is temporally associated with annual flower bud initiation in *Eucalyptus globulus* subsp. *globulus* (Myrtaceae). *Australian Journal of Botany* **59**: 756–769.

Kachroo A, Schopfer CR, Nasrallah ME, Nasrallah JB. 2001. Allele-specific receptor-ligand interactions in *Brassica* self-incompatibility. *Science* **293**: 1824–1826.

Karlgren A, Gyllenstrand N, Källman T, Sundström JF, Moore D, Lascoux M, Lagercrantz U. 2011. Evolution of the *PEBP* gene family in plants: functional diversification in seed plant evolution. *Plant Physiology* **156**: 1967–1977.

Kim W, Latrasse D, Servet C, Zhou D-X. 2012. *Arabidopsis* histone deacetylase HDA9 regulates flowering time through repression of AGL19. *Biochemical and Biophysical Research Communications* **432**: 394–398.

Klintenäs M, Pin PA, Benlloch R, Ingvarsson PK, Nilsson O. 2012. Analysis of conifer *FLOWERING LOCUS T/TERMINAL FLOWER1*-like genes provides evidence for dramatic biochemical evolution in the angiosperm *FT* lineage. *New Phytologist* **196**: 1260–1273.

Kotoda N, Hayashi H, Suzuki M, Igarashi M, Hatsuyama Y, Kidou S, Igasaki T, Nishiguchi M, Yano K, Shimizu T *et al.* 2010. Molecular characterization of *FLOWERING LOCUS T*-like genes of apple (*Malus* × *domestica* Borkh.). *Plant and Cell Physiology* **51**: 561–575.

Kramer EM, Hall JC. 2005. Evolutionary dynamics of genes controlling floral development. *Current Opinion in Plant Biology* **8**: 13–18.

Krizek BA, Fletcher JC. 2005. Molecular mechanisms of flower development: an armchair guide. *Nature Reviews Genetics* **6**: 688–698.

Kunieda T, Mitsuda N, Ohme-Takagi M, Takeda S, Aida M, Tasaka M, Kondo M, Nishimura M, Hara-Nishimura I. 2008. NAC family proteins NARS1/NAC2 and NARS2/NAM in the outer integument regulate embryogenesis in *Arabidopsis*. *Plant Cell* **20**: 2631–2642.

Kyozuka J, Harcourt R, Peacock WJ, Dennis ES. 1997. *Eucalyptus* has functional equivalents of the *Arabidopsis AP1* gene. *Plant Molecular Biology* **35**: 573–584.

Ladiges PY. 1997. Phylogenetic history and classification of eucalypts. In: Williams J, Woinarski J, eds. *Eucalypt ecology: individuals to ecosystems*. Cambridge, UK: Cambridge University Press, 16–29.

Lan P, Li W, Schmidt W. 2013. Genome-wide co-expression analysis predicts protein kinases as important regulators of phosphate deficiency-induced root hair remodeling in *Arabidopsis*. *BMC Genomics* **14**: 210.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R *et al.* 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947–2948.

Lawton-Rauh A. 2003. Evolutionary dynamics of duplicated genes in plants. *Molecular Phylogenetics and Evolution* **29**: 396–409.

Lee J, Lee I. 2010. Regulation and function of SOC1, a flowering pathway integrator. *Journal of Experimental Botany* **61**: 2247–2254.

Letunic I, Bork P. 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**: 127–128.

Li H, Madden JL, Potts BM. 1995. Variation in volatile leaf oils of the Tasmanian *Eucalyptus* species I. Subgenus *Monocalyptus*. *Biochemical Systematics and Ecology* **23**: 299–318.

Li H, Madden JL, Potts BM. 1996. Variation in volatile leaf oils of the Tasmanian *Eucalyptus* species II. Subgenus *Symphyomyrtus*. *Biochemical Systematics and Ecology* **24**: 547–569.

Litt A, Irish VF. 2003a. Duplication and diversification in the *APETALA1/FRUITFULL* floral homeotic gene lineage: implications for the evolution of floral development. *Genetics* **165**: 821–833.

Litt A, Kramer EM. 2010. The ABC model and the diversification of floral organ identity. *Seminars in Cell & Developmental Biology* **21**: 129–137.

Masiero S, Colombo L, Grini PE, Schnittger A, Kater MM. 2011. The emerging importance of Type I MADS box transcription factors for plant reproduction. *Plant Cell* **23**: 865–872.

McKiernan AB, O'Reilly-Wapstra JM, Price C, Davies NW, Potts BM, Hovenden MJ. 2012. Stability of plant defensive traits among populations in two *Eucalyptus* species under elevated carbon dioxide. *Journal of Chemical Ecology* **38**: 204–212.

Mizrachi E, Hefer CA, Ranik M, Joubert F, Myburg AA. 2010. *De novo* assembled expressed gene catalog of a fast-growing *Eucalyptus* tree produced by Illumina mRNA-Seq. *BMC Genomics* **11**: 681.

Myburg AA, Grattapaglia D, Tuskan GA, Hellsten U, Hayes RD, Grimwood J, Jenkins J, Lindquist E, Tice H, Bauer D *et al.* 2014. The genome of *Eucalyptus grandis*. *Nature* **510**: 356–362.

Naithani S, Chookajorn T, Ripoll DR, Nasrallah JB. 2007. Structural modules for receptor dimerization in the S-locus receptor kinase extracellular domain. *Proceedings of the National Academy of Sciences, USA* **104**: 12211–12216.

Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin Y-C, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A *et al.* 2013. The Norway spruce genome sequence and conifer genome evolution. *Nature* **497**: 579–584.

Ohshima S, Murata M, Sakamoto W, Ogura Y, Motoyoshi F. 1997. Cloning and molecular analysis of the *Arabidopsis* gene *Terminal Flower 1*. *Molecular Genetics and Genomics* **254**: 186–194.

O'Reilly-Wapstra JM, Bailey JK, McArthur C, Potts BM. 2010. Genetic-and chemical-based resistance to two mammalian herbivores varies across the geographic range of *Eucalyptus globulus*. *Evolutionary Ecology Research* **12**: 491–505.

Ostlund G, Schmitt T, Forslund K, Köstler T, Messina DN, Roopra S, Frings O, Sonnhammer ELL. 2010. InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Research* **38**: D196–D203.

Parcy F. 2005. Flowering: a time for integration. *The International Journal of Developmental Biology* **49**: 585–593.

Pavlidis P, Noble WS. 2003. Matrix2png: a utility for visualizing matrix data. *Bioinformatics* **19**: 295–296.

Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF. 2000. B and C floral organ identity functions require *SEPALLATA* MADS-box genes. *Nature* **405**: 200–203.

Pin PA, Nilsson O. 2012. The multifaceted roles of *FLOWERING LOCUS T* in plant development. *Plant, Cell & Environment* **35**: 1742–1755.

Posé D, Yant L, Schmid M. 2012. The end of innocence: flowering networks explode in complexity. *Current Opinion in Plant Biology* **15**: 45–50.

Pound L, Wallwork M, Potts BM, Sedgley M. 2002. Self-incompatibility in *Eucalyptus globulus* ssp. *globulus* (Myrtaceae). *Australian Journal of Botany* **50**: 365–372.

Pryor L, Knox R. 1971. Operculum development and evolution in eucalypts. *Australian Journal of Botany* **19**: 143–171.

Ranik M, Myburg AA. 2006. Six new cellulose synthase genes from *Eucalyptus* are associated with primary and secondary cell wall biosynthesis. *Tree Physiology* **26**: 545–556.

Sedgley M, Smith RM. 1989. Pistil receptivity and pollen tube growth in relation to the breeding system of *Eucalyptus woodwardii* (*Symphyomyrtus*: Myrtaceae). *Annals of Botany* **64**: 21–31.

Shan HY, Zhan N, Liu CJ, Xu GX, Zhang J, Chen ZD, Kong HZ. 2007. Patterns of gene duplication and functional diversification during the evolution of the *AP1/SQUA* subfamily of plant MADS-box genes. *Molecular Phylogenetics and Evolution* **44**: 26–41.

Shiu S-H, Bleecker AB. 2001. Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. *Proceedings of the National Academy of Sciences, USA* **98**: 10763–10768.

Sorensen A-M, Kröber S, Unte US, Huijser P, Dekker K, Saedler H. 2003. The *Arabidopsis ABORTED MICROSPORES (AMS)* gene encodes a MYC class transcription factor. *Plant Journal* **33**: 413–423.

Southerton SG, Marshall H, Mouradov A, Teasdale RD. 1998a. Eucalypt MADS-box genes expressed in developing flowers. *Plant Physiology* **118**: 365–372.

Southerton SG, Strauss SH, Olive MR, Harcourt RL, Decroocq V, Zhu XM, Llewellyn DJ, Peacock WJ, Dennis ES. 1998b. *Eucalyptus* has a functional equivalent of the Arabidopsis floral meristem identity gene *LEAFY*. *Plant Molecular Biology* **37**: 897–910.

Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology* **56**: 564–577.

Tantikanjana T, Rizvi N, Nasrallah ME, Nasrallah JB. 2009. A dual role for the S-Locus receptor kinase in self-incompatibility and pistil development revealed by an *Arabidopsis rdr6* mutation. *Plant Cell* **21**: 2642–2654.

Theis N, Lerdau M. 2003. The evolution of function in plant secondary metabolites. *International Journal of Plant Sciences* **164**: S93–S102.

Theissen G. 2001. Development of floral organ identity: stories from the MADS house. *Current Opinion in Plant Biology* **4**: 75–85.

Theocharidis A, van Dongen S, Enright AJ, Freeman TC. 2009. Network visualization and analysis of gene expression data using BioLayout Express(3D). *Nature Protocols* **4**: 1535–1550.

Thompson BE, Hake S. 2009. Translational biology: from *Arabidopsis* flowers to grass inflorescence architecture. *Plant Physiology* **149**: 38–45.

Torti S, Fornara F, Vincent C, Andres F, Nordstrom K, Gobel U, Knoll D, Schoof H, Coupland G. 2012. Analysis of the *Arabidopsis* shoot meristem transcriptome during floral transition identifies distinct regulatory patterns and a leucine-rich repeat protein that promotes flowering. *Plant Cell* **24**: 444–462.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology* **28**: 511–515.

Wang N, Huang H-J, Ren S-T, Li J-J, Sun Y, Sun D-Y, Zhang S-Q. 2012. The rice wall-associated receptor-like kinase gene *OsDEES1* plays a role in female gametophyte development. *Plant Physiology* **160**: 696–707.

Watson JM, Brill EM. 2004. *Eucalyptus grandis* has at least two functional *SOC1*-like floral activator genes. *Functional Plant Biology* **31**: 225–234.

Weigel D, Alvarez J, Smyth DR, Yanofsky MF, Meyerowitz EM. 1992. *LEAFY* controls floral meristem identity in *Arabidopsis*. *Cell* **69**: 843–859.

Weigel D, Meyerowitz EM. 1994. The ABCs of floral homeotic genes. *Cell* **78**: 203–209.

Wellmer F, Riechmann JL. 2010. Gene networks controlling the initiation of flower development. *Trends in Genetics* **26**: 519–527.

Xing S, Salinas M, Huijser P. 2011. New players unveiled in early anther development. *Plant Signaling & Behavior* **6**: 934–938.

Yamada K, Lim J, Dale JM, Chen H, Shinn P, Palm CJ, Southwick AM, Wu HC, Kim C, Nguyen M *et al.* 2003. Empirical analysis of transcriptional activity in the *Arabidopsis* genome. *Science* **302**: 842–846.

Yoo MJ, Chanderbali AS, Altman NS, Soltis PS, Soltis DE. 2010. Evolutionary trends in the floral transcriptome: insights from one of the basalmost angiosperms, the water lily *Nuphar advena* (Nymphaeaceae). *Plant Journal* **64**: 687–698.

Zahn L, Ma X, Altman N, Zhang Q, Wall PK, Tian D, Gibas C, Gharaibeh R, Leebens-Mack J, Ma H. 2010. Comparative transcriptomics among floral organs of the basal eudicot *Eschscholzia californica* as reference for floral evolutionary developmental studies. *Genome Biology* **11**: R101.

Zhu J, Chen H, Li H, Gao J-F, Jiang H, Wang C, Guan Y-F, Yang Z-N. 2008. *Defective in Tapetal Development and Function 1* is essential for anther development and tapetal function for microspore maturation in *Arabidopsis*. *Plant Journal* **55**: 266–277.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 Subset of GO categories enriched in late flower-predominant genes.

Fig. S2 Expression patterns of differentially expressed transcription factors.

Fig. S3 Phylogenetic tree of type I MADS-box-containing proteins of *Eucalyptus*, *Arabidopsis thaliana*, poplar and grape.

Fig. S4 Expression profiles of MADS-box genes in the vegetative and reproductive tissues of *Eucalyptus grandis*.

Fig. S5 Phylogenetic tree of the S-domain receptor-like kinase (SDRLK) subfamily of *Eucalyptus grandis*.

Fig. S6 Relative expression of three floral biomarker genes assayed by qRT-PCR.

Table S1 Floral-specific genes expected to be expressed early or late in flower bud development for use as flowering biomarkers, and the putative number of paralogs in *E. grandis*

Table S2 Sequences of primers used in qRT-PCR

Table S3 Number of reads mapped to the *Eucalyptus grandis* v1.1 genome assembly using TopHat

Table S4 Number of genes with detected expression (CuffDiff; 95% CI) in each of the sampled tissues

Table S5 Gene expression values of all expressed genes in sampled tissues

Table S6 Numbers of differentially expressed genes in pairwise comparisons

Table S7 Gene ontology categories enriched in the late-floral-predominant gene set

Table S8 Categories of floral-predominant, transcription factor genes

Table S9 List of MADS-box genes for *Eucalyptus grandis*, *Arabidopsis thaliana*, *Populus trichocarpa* and *Vitis vinifera*

Table S10 Expression of *Eucalyptus grandis* putative orthologs of the ABC and SEPALLATA classes of genes in flower buds harvested late versus early in floral bud development

Table S11 Raw data from qRT-PCR of selected biomarker candidate genes

Notes S1 Supporting results (phylogenetic analysis of distinct eucalypt species and expression profiles of SDRLK genes).

Please note: Wiley Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.