Plant Cell Rep
DOI 10.1007/s00299-015-1817-6

ORIGINAL PAPER

Next-generation transcriptome analysis in transgenic birch overexpressing and suppressing APETALA1 sheds lights in reproduction development and diterpenoid biosynthesis

Haijiao Huang[1] · Su Chen[1] · Huiyu Li[1] · Jing Jiang[1]

Received: 30 January 2015 / Revised: 8 May 2015 / Accepted: 1 June 2015

Abstract

Key message **Overexpression of *BpAP1* could cause early flowering in birch. *BpAP1* affected the expression of many flowering-related unigenes and diterpenoid biosynthesis in transgenic birch, and *BpPI* was a putative target gene of *BpAP1*.**

Abstract APETALA1 (AP1) is an MADS-box transcription factor that is involved in the flowering process in plants and has been a focus of genetic studies examining flower development. Here, we carried out transcriptome analysis of birch (*Betula platyphylla* Suk.), including *BpAP1* overexpression lines, *BpAP1* suppression lines, and non-transgenic line (NT). Compared with NT, we detected 8302 and 7813 differentially expressed unigenes in 35S::*BpAP1* and 35S::*BpAP1RNAi* transgenic lines, respectively. Overexpression and suppression of *BpAP1* in birch affected diterpenoid biosynthesis and altered expression of many flowering-related unigenes. Moreover, combining information from the RNA-seq database and the birch genome, we predicted downstream target genes of *BpAP1*. Among the 166 putative target genes of *BpAP1*, there was a positive correlation between *BpAP1* and *BpPI*. These results provide references for further examining the relationship between *BpAP1* and its target genes, and reveal that *BpAP1* functions as a transcription regulator in birch.

Keywords Birch · Transcriptome · *BpAP1* · Flower development · Target gene

Abbreviations

AP1	*APETALA1*
NT	Non-transgenic line
FT	*FLOWERING LOCUS T*
SOC1	*SUPPRESSOR OF OVEREXPRESSION OF CO1*
LFY	*LEAFY*
PI	*PISTILLATA*
ChIP	Chromatin immunoprecipitation
GGDP	Geranylgeranyl diphosphate
RNA-seq	RNA sequencing
WPM	Woody plant medium
Nr	Non-redundant
KEGG	Kyoto encyclopedia of genes and genomes
COG	Clusters of orthologous groups
GO	Gene ontology
RPKM	Reads per kb per million reads
DEUs	Differentially expressed unigenes
DETGs	Differentially expressed target genes

Communicated by Z.-Y. Wang.

Electronic supplementary material The online version of this article (doi:10.1007/s00299-015-1817-6) contains supplementary material, which is available to authorized users.

✉ Jing Jiang
jiangj-cf@nefu.edu.cn

Haijiao Huang
haijiao_sea@163.com

[1] State Key Laboratory of Tree Genetics and Breeding, Northeast Forestry University, 26 Hexing Road, Harbin 150040, China

Introduction

Flower development is an important stage in tree growth and development and is affected by environmental and endogenous signals. Flowering has been extensively

Published online: 11 June 2015

studied in the model plant *Arabidopsis* (Irish and Sussex 1990; Jack 2004; Kaufmann et al. 2010; Wellmer et al. 2006; Yang et al. 2012). Flowering is an initiation of reproductive growth of plant. There are a number of genes involved in the transition from vegetative to reproductive growth. The flowering-time genes *FLOWERING LOCUS T* (*FT*) and *SUPPRESSOR OF OVEREXPRESSION OF CO1* (*SOC1*), together with the floral meristem identity gene *LEAFY* (*LFY*), are three essential regulators integrating floral signals from multiple pathways in *Arabidopsis thaliana*. *AGAMOUS-LIKE 24* (*AGL24*) also plays a role in the regulation of flowering time (Michaels et al. 2008; Yu et al. 2002). Since trees have a long juvenile phase and are tall, it is difficult to study their floral development compared with that of herbaceous plants, and few studies have focused on the transcriptional regulation of flowering in trees.

The flowering process involves a complex regulatory network consisting of many genes, including *AP1*, a multiple regulatory locus. In *Arabidopsis*, *LFY*, like *AP1*, is an important floral meristem identity gene. There is a positive–negative regulatory interaction that exists between these two genes. *AP1* expression is delayed and reduced in the *lfy* mutant, whereas constitutive ectopic expression of *LFY* results in precocious *AP1* expression. These observations indicate that *LFY* can directly regulate *AP1* in a positive manner (Wagner et al. 1999). All floral meristems do not develop into inflorescence meristems in a strong *ap1* mutant, while more floral meristems develop into inflorescence meristems in a strong *ap1 lfy* mutant, suggesting that *AP1* and *LFY* regulate each other in a positive manner, which occurs through cell–cell signaling transmission (Sessions et al. 2000).

AP1 positively regulates the expression of *AP3* and *PISTILLATA* (*PI*), functioning as a transcription activation factor to regulate petal specificity (Lamb et al. 2002; Ng and Yanofsky 2001). Conversely, chromatin immunoprecipitation (ChIP) analysis demonstrated that PI, together with AP3, can bind directly to the promoter region of AP1, suggesting that AP1 is a direct target of the AP3/PI heterodimer. *AP1* positively regulates *AP3* and *PI* (Sundstrom et al. 2006). *LFY* acts in both a direct and an indirect manner to regulate *AP3* expression; LFY binds directly to the region of the *AP3* promoter required for early expression of *AP3*, while it is dependent on *AP1* activity to regulate *AP3* expression (Lamb et al. 2002). To further investigate flower development in *Arabidopsis*, ChIP and microarray analyses were employed to identify AP1 binding sites and target genes, revealing 1942 regions bound by AP1. Most of the target genes of AP1 encode transcription factors, and most are negatively regulated by AP1 (Kaufmann et al. 2010).

Gibberellins are a class of diterpenoid plant hormones involved in regulating plant growth and development that promote seed germination, accelerate plant growth, promote flowering, and alter the ratio of male-to-female flowers (Garcia-Martinez et al. 1997). Diterpenoid biosynthesis comprises three steps: First, the precursor geranylgeranyl diphosphate (GGDP) is converted to *ent*-kaurene by a two-step cyclization; second, *ent*-kaurene is oxidized into GA12; finally, GA12 forms the intermediate product GA and bioactive GA via different metabolic pathways, i.e., 13-hydroxylation and non-13-hydroxylation, respectively (Yamaguchi 2008). KO and KAO are key enzymes in the second stage of GAs biosynthesis, and the 2-hydroxylation of GA1, GA4, GA9, and GA20 occurs via GA2ox. GA2ox can regulate the levels of bioactive GA in the cell, thereby regulating plant growth and development (Dijkstra et al. 2008; Huang et al. 2010; Olszewski et al. 2002; Thomas et al. 1999). Transfer of *PcGA2ox1* into *Solanum nigrum* and *Solanum melanocerasum* produces dwarf transgenic plants (Dijkstra et al. 2008).

The completion of the birch (*Betulaplatyphylla* Suk.) genome sequence (http://birch.genomics.cn/) has made it easier to study the molecular mechanisms underlying the flower development in this tree. *BpAP1* belongs to the *AP1/SQUA* subfamily of the MADS-box family and plays an important role in the flowering transition and floral organ development; overexpression of *BpAP1* can shorten the juvenile phase of birch (Huang et al. 2014).

RNA sequencing (RNA-seq) represents a new powerful method that can be used to elucidate the overall patterns of gene transcription shedding light on molecular mechanisms underlying the development of plants (Marioni et al. 2008; Wang et al. 2009b; Zhang et al. 2005), including trees (Zhang et al. 2011). Here, we employed high-throughput sequencing technology to carry out RNA-seq analysis of the previously described *BpAP1* overexpression (Huang et al. 2014) and suppression lines in birch. In addition, using the birch genome as a reference, we predicted *BpAP1* target genes. The results of this study will provide a reference for confirming the target genes of *BpAP1*, and for revealing the function of *BpAP1* in the flowering transition as well as growth and development in birch.

Materials and methods

Vector and plant materials

Agrobacterium tumefaciens strain EHA105 harboring the RNA interference expression vector pROKII-*AP1RNAi*, which contains the *BpAP1RNAi* gene, was employed in this study (Fig. 1a). The constitutive expression vector

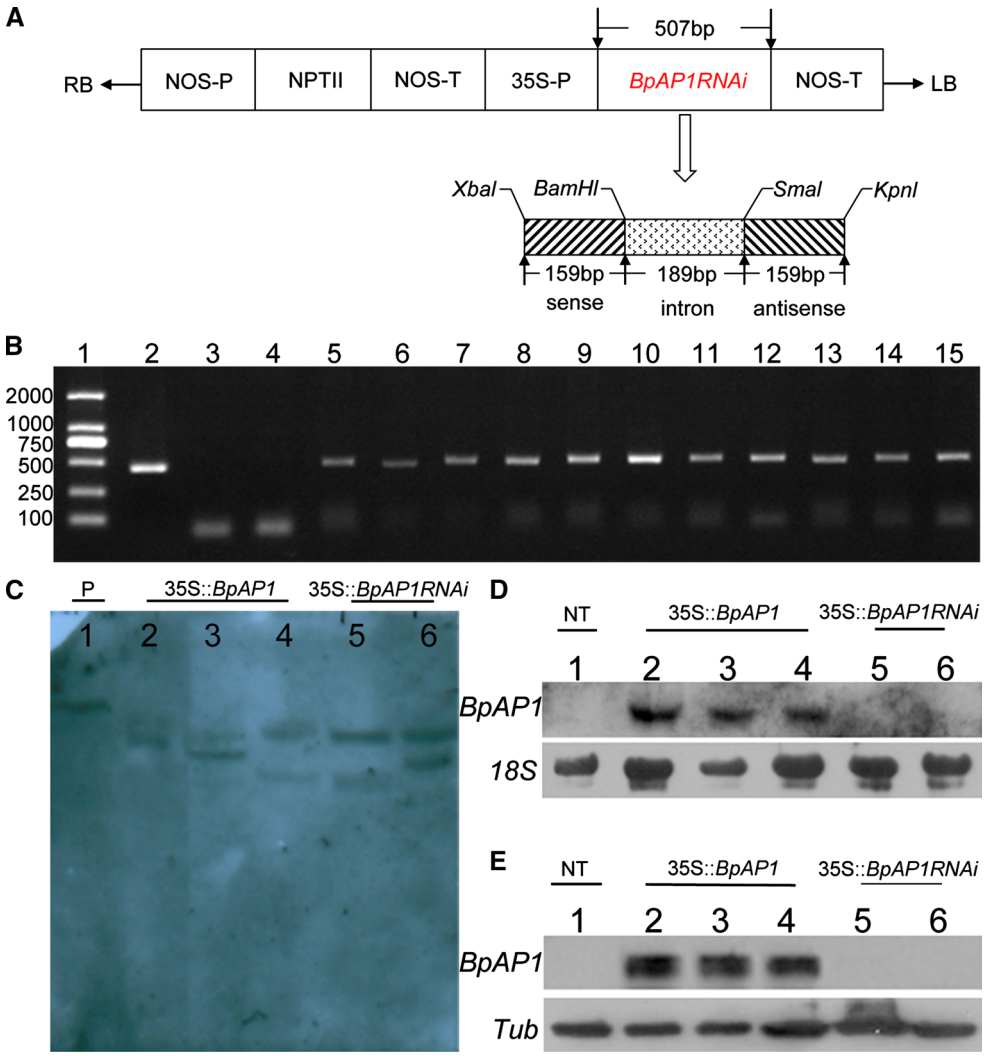

Fig. 1 Generation of 35S::*BpAP1 RNAi* transgenic birch. **a** Diagram of the plant suppression vector. *Nos-P* promoter of the nopaline synthase gene, *Nos-T* terminator of the nopaline synthase gene, *NPT II* kanamycin resistance marker gene, *35S-P* CaMV 35S promoter. **b** PCR of 35S::*BpAP1RNAi* transgenic lines. *1* DNA Marker DL2000, *2* positive control (pROKII-*AP1RNAi*), *3* water control, *4* Non-transgenic plant; 5-12. 35S::*BpAP1RNAi* transgenic lines. **c** Southern blot analysis. **d** Northern blot analysis. **e** Western blot analysis. The restriction enzyme *EcoR*I was used for Southern analysis. About 20 μg DNA was used for Southern and 20 μg RNAwas used for Northern. The full-length *BpAP1* was amplified and used as a probe for Southern and Northern. For Western blot analysis, the polyclonal antibody of BpAP1 was prepared, and 40–60 μg plant total protein was used. *P* positive control, pROKII-*AP1*, *NT* non-transgenic plant, *35S::BpAP1 BpAP1* overexpression transgenic lines, *35S::BpAP1 BpAP1* suppression transgenic lines, *18S* 18S ribosomal RNA; birch *18S rRNA* was also hybridized to verify that similar amounts of RNA were used in Northern blot analysis. *Tub* α-Tubulin, birch *tubulin* was hybridized to verify that similar amounts of protein were used in Western blot analysis

pRKOII-*AP1* and plant transformation method of transgenic birch are described as Huang et al. (2014).

Plant materials used for RNA-seq analysis included the whole plants from multiple shoot of two transgenic lines overexpressing *BpAP1* (35S::*BpAP1*, A2, A5) and two transgenic lines in which *BpAP1* was silenced (35S::*BpAP1RNAi*, FA10, FA11) and the non-transgenic (NT) line.

The seedlings were cultivated in woody plant medium (WPM) (Lloyd and McCown 1980), solid medium supplemented with 1.0 mg/L 6-BA, 0.5 mg/L GA3. After germinating from the multiple shoot, the whole plants were harvested at 20-day-old and frozen immediately in liquid nitrogen. They were stored at −80 °C for RNA extraction. The plants were grown in a cultivation room at 22–25 °C under a 16-h light/8-h dark photoperiod at illumination intensity of 1000–1500 cd × sr/m^2.

Analysis of transformants

To analyze the transgenic lines, we employed PCR, and Southern, Northern, and Western blotting techniques implemented by others (Huang et al. 2014; Lemmetyinen et al. 2001; Towbin et al. 1979). In brief, we confirmed the

integration of exogenetic *BpAP1RNAi* in transgenic lines using PCR method. The PCR primers were designed from the vector sequences, forward primer: 5′-AAGACCGG CAACAGGATTC-3′ and reverse primer: 5′-CGCA CAATCCCACTATCCT-3′. We next employed Southern blot to investigate the insertion sites of the exogenetic *BpAP1*. About 20 μg genomic DNA was digested with *Eco*RI (Takara) at 37 °C for 16 h used for Southern hybridization. The full-length *BpAP1* was used as a probe. Northern blot was then implemented to examine the transcriptional expression of exogenetic *BpAP1*. Finally, Western blot was used to examine the protein level of exogenetic BpAP1. For Western blotting analysis, the polyclonal antibody of BpAP1 was prepared, and 40-60 μg plant total protein was used.

RNA extraction

Total RNA was isolated from each sample using the CTAB method (Chang, et al. 1993) and was digested with DNaseI (Promega, Madison, USA) to remove the contaminating DNA. Based on the results of detection using a 2100 Biological Analyzer (Agilent, Palo Alto, USA), 20 μg RNA from each sample was used to construct five cDNA libraries and the remaining RNA was used for the real-time quantitative PCR verification.

cDNA library construction and RNA sequencing

Enrichment of mRNA, fragment interruption, addition of adapters, size selection and PCR amplification, and RNA-Seq were performed at the Beijing Genome Institute (BGI, Shenzhen, China).

Beads with Oligo (dT) were used to isolate poly (A) mRNA after total RNA was extracted from the samples. Fragmentation buffer was added to cleave the mRNA into short fragments. Using these short fragments as templates, random hexamer primers were used to synthesize the first-strand cDNA. Second-strand cDNA was synthesized using buffer, dNTPs, RNaseH, and DNA polymerase I. Short fragments were purified with a QiaQuick PCR extraction kit and resolved with EB buffer, followed by end repair and the addition of poly(A). Sequencing adapters were then added to the short fragments, which were then subjected to agarose gel electrophoresis. Suitable fragments were selected for use as templates for PCR amplification. Finally, the library was sequenced using an IlluminaHiSeq^{TM} 2000.

Data filtering and de novo assembly

The raw reads were first filtered by removing adapter sequences and low-quality sequences, which included reads with N percentages over 5 % and sequences containing more than 20 % nucleotides with a Q value $\leq$10. The Q value represents the sequencing quality of nucleotides. Clean reads were used in de novo assembly and read-mapping to the transcriptome.

Transcriptome de novo assembly was carried out with the short read assembly program Trinity (Grabherr et al. 2011). Trinity first combines reads with overlaps of a certain size to form longer fragments without N, which are referred to as contigs. The reads are then mapped back to contigs; with paired-end reads, contigs from the same transcript as well as the distances between these contigs can be detected. Next, Trinity connects the contigs, using N to represent unknown sequences between two contigs, and scaffolds are then constructed. Finally, sequences without Ns that cannot be extended on either end are obtained; such sequences are designated as unigenes. Unigenes from each sample's (NT, A2, A5, FA10, FA11) assembly can be further processed via sequence splicing and redundancy removal using sequence clustering software TGICL to acquire non-redundant unigenes that are as long as possible.

Functional annotation and differential expression analysis

Finally, unigenes were aligned with the NCBI non-redundant (Nr), Swiss-Prot, Kyoto Encyclopedia of Genes and Genomes (KEGG), and Clusters of Orthologous Groups (COG) protein databases using Blastx with an E value $\leq 10^{-5}$. The proteins with the highest sequence similarity were retrieved for analysis. Incongruent results from different databases were settled under a priority order of Nr, Swiss-Prot, KEGG, and COG (Huang et al. 2012). Blast2GO (Conesa et al. 2005) was used to obtain gene ontology (GO) annotation of the unigenes based on the Nr annotation. In addition, putative metabolic pathways were assigned to each unigene by performing Blastx searches against the KEGG pathway database with an E value threshold of 1×10^{-5}.

Unigene expression was calculated with the RPKM method (Reads Per kb per Million reads) (Mortazavi et al. 2008), using the following formula: $RPKM = 10^6 C/(NL/10^3)$, where RPKM represents expression of Unigene A, C is the number of reads that are uniquely aligned to Unigene A, N is the total number of reads uniquely aligned to all unigenes, and L is the base number in the CDS of Unigene A. The RPKM method is able to eliminate the influence of different gene lengths and sequencing levels on the calculation of gene expression. Therefore, the calculated gene expression value can be directly used to compare the differences in gene expression between samples.

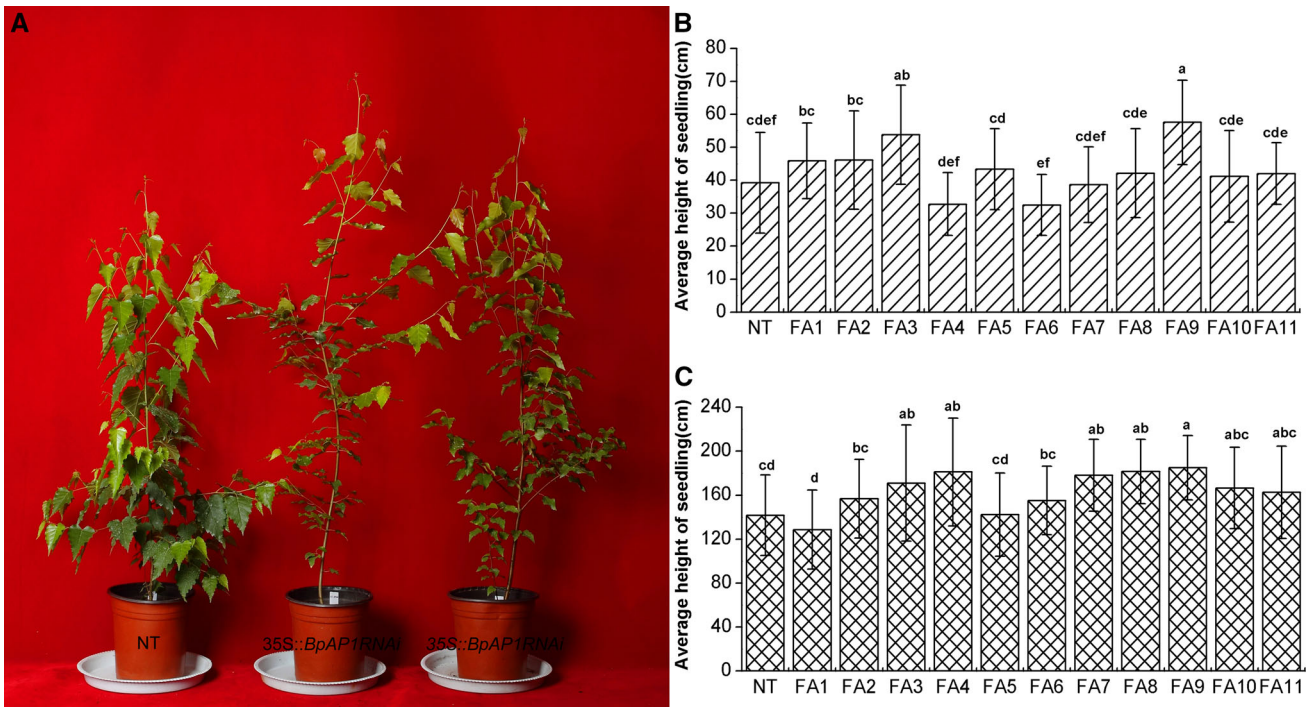

Fig. 2 Seedling heights of 35S::*BpAP1RNAi* transgenic lines and non-transgenic plants. **a** *NT* non-transgenic plant, *35S::BpAP1RNAi* 35S::*BpAP1RNAi* transgenic lines. **b** Seedling heights of 1-year-old 35S::*BpAP1RNAi* transgenic and NT plants. **c** Seedling heights of 2-year-old 35S::*BpAP1RNAi* transgenic and NT plants. Data accompanied by the *same letter* indicate no significant difference, and *different letters* indicate a significant difference ($n = 12$, $P < 0.05$; Duncan's multiple range analysis). The *error bars* represent SD

A rigorous algorithm was used to identify differentially expressed unigenes (DEUs) in comparisons between the two samples (Audic and Claverie 1997). In the current analysis, FDR $\leq$0.001 and the absolute value of log2Ratio $\geq$1 were used as the threshold to judge the significance of differential gene expression. DEUs were subjected to GO functional analysis and KEGG pathway analysis (Ashburner et al. 2000; Kanehisa and Goto 2000).

qRT-PCR verification

Total RNA was extracted from transgenic and NT lines using the CTAB method. First-strand cDNA was synthesized using a ReverTreAce®qPCR RT Kit (Toyobo, Osaka, Japan) and qRT-PCR was performed using quantitative SYBR green PCR Master Mix (Toyobo, Osaka, Japan) on an ABI 7500 real-time PCR system (45 cycles of 95 °C for 30 s, 95 °C for 15 s, and 60 °C for 1 min). *18S rRNA* (Accession number: GU476453.1) and *α-Tubulin* (Accession number: FJ228477.1) were used as housekeeping genes to analyze the expression of the selected unigenes. All experiments were conducted with three technical replicates per sample, and the results were analyzed using the $2^{-\Delta\Delta Ct}$ (Livak and Schmittgen 2001) method; for primer sequences, see Table S1.

Prediction of *BpAP1* target genes

With reference to MADS domain-containing binding sites in *Arabidopsis* (Kaufmann et al. 2010; Riechmann et al. 1996), searches for DNA regions bound by *BpAP1* in the birch genome were performed using R and Perl (Team RC 2012) to predict genes targeted by *BpAP1*. BioEdit (Altschul et al. 1997) was used to Blast the RNA-seq database of the 35S::*BpAP1* and 35S::*BpAP1RNAi* transgenic lines and NT plants against birch genome sequence to predict downstream target genes of *BpAP1*. Phylogenetic analysis of *BpAP1*, *PI*, and *AP3*-related genes from different plant species was performed using MEGA5 (Tamura et al. 2011).

Cluster analysis of the *BpAP1* target genes was performed using Cluster 3.0 (de Hoon et al. 2004).

Correlation analysis of the expression of *BpAP1* and its target genes was performed using SPSS16.0 (Norusis 2008).

Expression characteristic analysis of *BpAP1* and *BpPI*

A 10-year-old WT line from the birch breeding base was chosen as the control to study expression characteristic in inflorescences (Huang et al. 2014). Total RNA from roots, shoots, leaves, terminal buds, and male inflorescences of

Table 1 Quality of RNA-seq and de novo assembly of *BpAP1* transgenic lines and non-transgenic plant (NT)

	NT	35S::*BpAP1* transgenic line		35S::*BpAP1RNAi* transgenic line		ALL
		A2	A5	FA10	FA11	
Total reads	55,367,288	53,862,930	54,548,464	54,449,892	51,440,772	
Total nucleotides (nt)	4,983,055,920	4,847,663,700	4,909,361,760	4,900,490,280	4,629,669,480	
High quality reads	54,055,083	52,607,924	53,228,391	53,654,924	50,710,313	
Q20 percentage (%)	97.63	97.67	97.58	98.54	98.58	
GC percentage (%)	48.13	48.17	48.12	48.24	48.12	
Number of contigs	188,038	200,808	189,624	191,330	209,000	
Number of unigenes	99,671	79,753	76,593	74,339	81,611	104,505
Mean length (nt)	830	647	665	680	641	857
N50 (nt)	1566	1206	1259	1283	1176	1555

Q20 percentage indicates the proportion of nucleotides with quality values greater than 20

N50 length is defined as half of the sum of the lengths of all assembled unigenes, which are ordered from longer to shorter

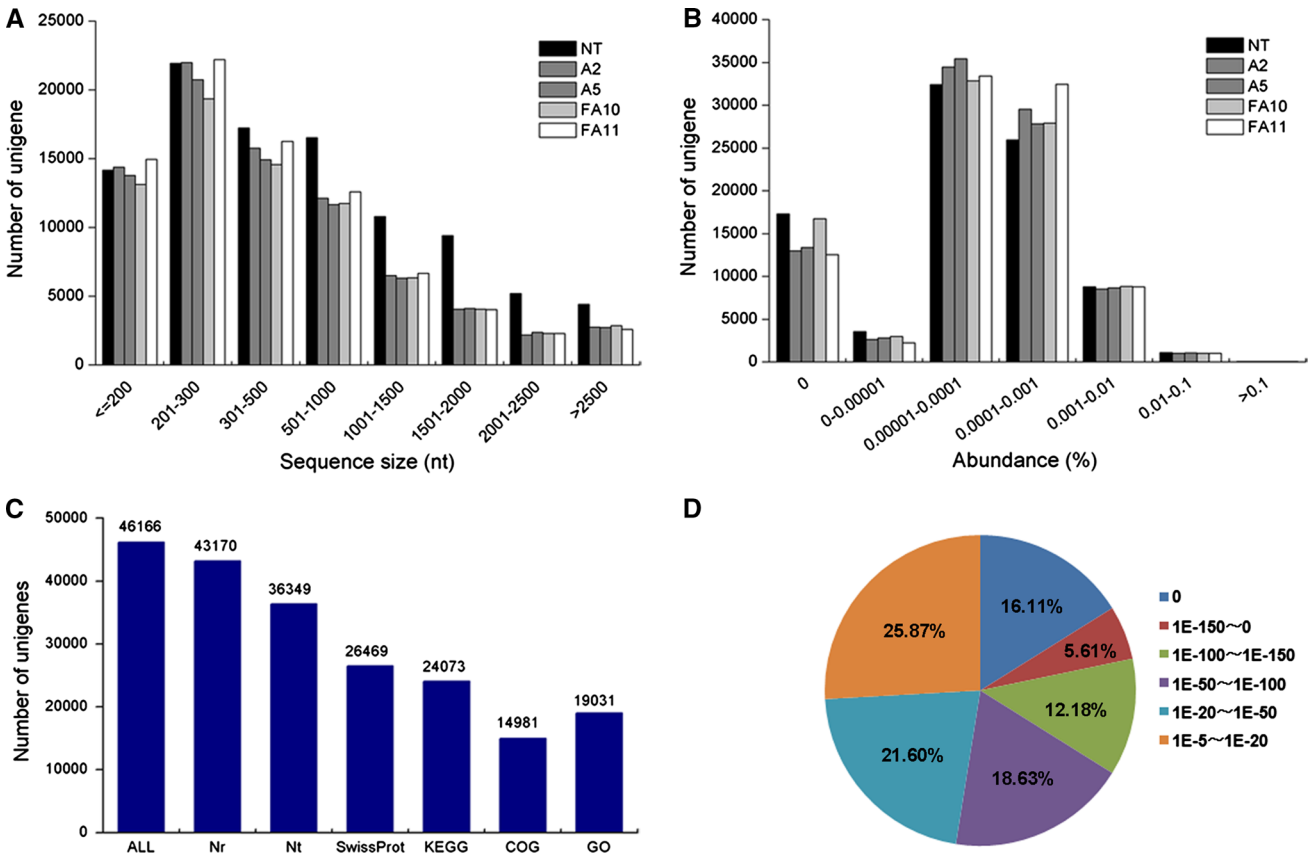

Fig. 3 Quality of RNA-seq and de novo assembly. **a** Distribution of the lengths of the assembled unigenes. **b** Distribution of the unigene abundance. The abundance of each unigene was calculated as the percentage of total unigenes in the transgenic lines and the untransformed control plants. *NT* non-transgenic plants, *A2, A5* 35S::*BpAP1* transgenic lines, *FA10 FA11* 35S::*BpAP1RNAi* transgenic lines. **c** Number of unigenes annotated by Blastx and Blastn against the databases with an *E* value threshold of 10^{-5}. The *numbers* above the columns indicate the number of annotated unigenes identified in each database. **d** *E* value distribution of the top Blastx hits against the Nr database for each unigene

the 35S::*BpAP1* transgenic lines (2-year-old plant) and the WT plant was extracted using the CTAB method. Leaves and male inflorescences of 35S::*BpAP1* transgenic lines, leaves of 35S::*BpAP1RNAi* transgenic line, and NT were sampled in early and mid-month from June to September, 2012 (these eight stages were designated ds1, ds2,…, ds8).

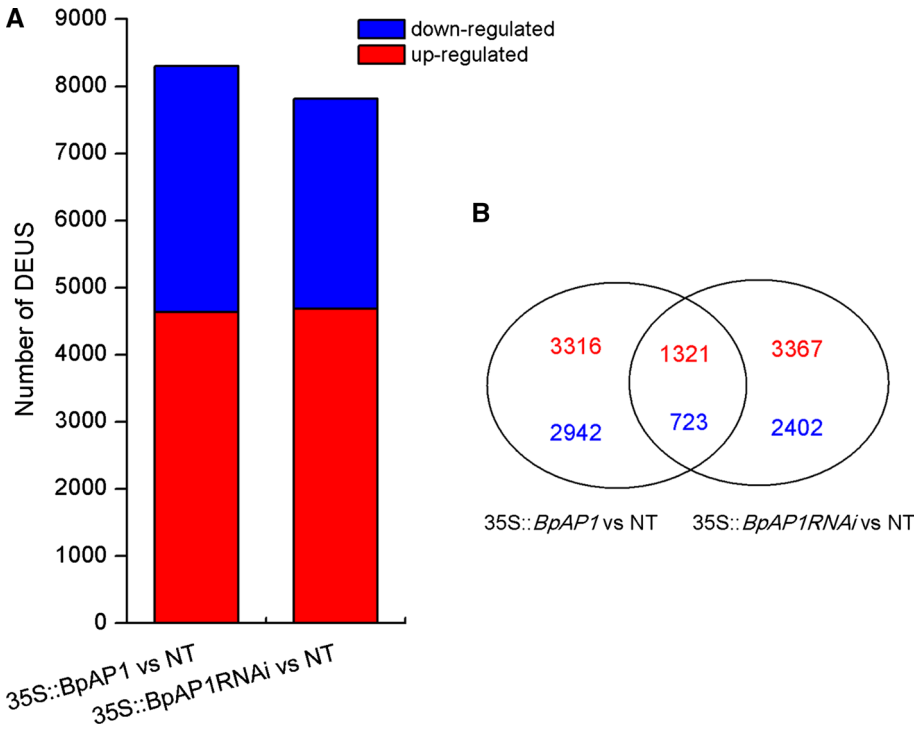

Fig. 4 Comparison of differentially expressed unigenes between transgenic lines and untransformed control plant. **a** Number of differentially expressed unigenes. **b** Comparison of differentially expressed unigenes. *NT* represents non-transgenic plants, *red font* represents up-regulation, and *blue font* represents down-regulation (color figure online)

Female and male inflorescences of the 35S::*BpAP1* transgenic lines and WT plants were sampled on 17 April, 22 April, and 27 April, 2013 [female inflorescences open around mid-April in 2013, these three stages were designated as ds9, ds10, and ds11 (Fig. 10c)]. qRT-PCR was conducted as described above; primer sequences are shown in Table S2.

Results

Generation of 35S::*BpAP1RNAi* transgenic birch

Using *Agrobacterium*-mediated transformation, we generated 11 independent 35S::*BpAP1RNAi* kanamycin-resistant lines.

PCR analysis was performed on the 35S::*BpAP1RNAi* kanamycin-resistant lines. All of the 35S::*BpAP1RNAi* kanamycin-resistant lines produced a band around 600 bp, while the negative control did not produce a band (Fig. 1b). These results provided preliminarily evidence that exogenous *BpAP1RNAi* was integrated into the genome of *B. platyphylla* × *B. pendula*. The lines were named FA1, FA2, FA3,…, FA10, and FA11. The 35S::*BpAP1RNAi* transgenic lines FA10 and FA11 were subjected to Southern blot hybridization, using pROKII-*AP1* as a positive control. Lines FA10 and FA11 both produced two bands, which showed that there were two copies of *BpAP1RNAi* in the genome of *B. platyphylla* × *B. pendula* (Fig. 1c). Northern blot and Western blot showed that each of the two 35S::*BpAP1RNAi* transgenic lines and non-transgenic plants failed to show any hybridization signal (Fig. 1d, e).

We investigated the height growth of the 11 transgenic lines. The result showed that in the 2 years of the experiment, the height of most of the transgenic plants was more than that of the non-transgenic lines ($P < 0.05$). The average height of the 1-year-old transgenic plants was 10.25 % greater than that of NT, and the average height of the 2-year-old transgenic plants was 16.01 % greater than that of NT (Fig. 2).

Quality of RNA-seq and de novo assembly

To analyze the function of *BpAP1* in birch, we constructed five libraries using RNA extracted from multiple shoot of 35S::*BpAP1* (A2, A5), 35S::*BpAP1RNAi* (FA10, FA11) transgenic lines, and NT. The sequence sets were filtered to remove low-quality reads containing ambiguous nucleotides and adaptor sequences in the five libraries, resulting in approximately 264 million clean reads. The Q20 percentages were all over 97 %. Short reads were assembled into contigs, taking the distance of paired-end reads into account; these contigs were then assembled into unigenes. Five libraries containing 99, 671, 79, 753, 76, 593, 74, 339; and 81,611 unigenes were obtained with mean lengths of 830, 647, 665, 680, and 641 nt, respectively (Table 1). The length of the assembled unigenes ranged from 200 to 3000 bp in the five databases (Fig. 3a). All of the sequences were assembled, yielding 104,505 non-

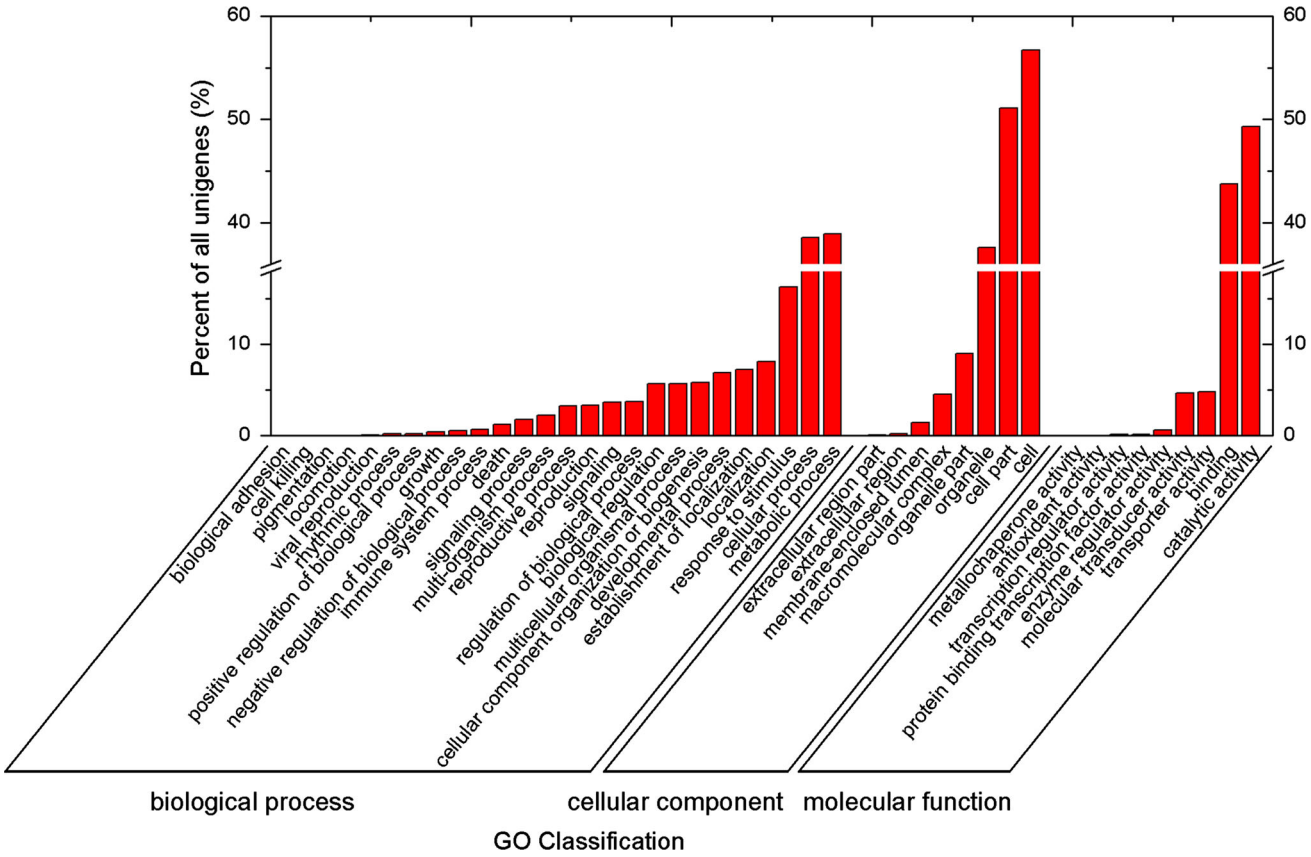

Fig. 5 Gene ontology classification of assembled unigenes. The unigenes were categorized based on gene ontology (GO) annotation, and the proportion of each category is displayed based on the ontologies biological process, cellular component and molecular function. In total, 6183 unigenes with Blast matches to known proteins were assigned to GO categories

redundant unigenes with a mean length of 857 nt, and the N50 was 1555 nt (Table 1).

Our annotation method was based on sequence homology searches and the annotations that accompanied them. The aim of this method was to capture the most informative and complete annotation possible. Figure 3c shows the hit numbers and percentages relative to those of the major public databases. These annotation statistics show data for all unigenes annotated by the Blast search against the public protein and nucleotide databases (Nr, Nt, Swiss-Prot, KEGG, COG, and GO).

We found that Illumina/Solexa was able to detect many transcripts expressed at low levels. Figure 3b shows the distribution of unigenes at different abundance levels in RPKM. The distribution of unigene abundances across the five libraries was generally quite similar. There were 41, 50, 49, 46, and 49 unigenes in A2, A5, FA10, FA11, and NT expressed at high abundance (more than 0.1 %, or >1000 RPKM). The number of unigenes increased dramatically with decreasing abundance; 56.12, 57.85, 58.14, 53.23, and 59.76 % of the unigenes in A2, A5, FA10, FA11, and NT, respectively, had an abundance of less than 0.0001 %. Moreover, of the unigenes present in a given library, the abundance of the vast majority, 88.86, 89.24, 89.08, 89.04, and 89.12 % in the A2, A5, FA10, FA11, and NT libraries, respectively, was below 0.001 % (<100 PRKM). These results illustrate the sensitivity of next-generation sequencing technology in identifying transcripts with low expression levels.

DEUs between transgenic lines and NT

Compared with the NT, there were 59.25, 56.54, 56.38, and 55.20 % unigenes that showed a twofold (ratio >2 or <0.5) or greater expression difference between the NT and the A2, A5, FA10, and FA11 libraries. There were 8302 DEUs in the 35S::*BpAP1* library; among these, 4637 DEUs were up-regulated and 3665 were down-regulated. By contrast, there were 7813 DEUs in the 35S::*BpAP1RNAi* library, including 4688 up-regulated and 3125 down-regulated DEUs (Fig. 4).

Functional classification by gene ontology

To evaluate the potential functions of unigenes with significant transcriptional changes between the transgenic

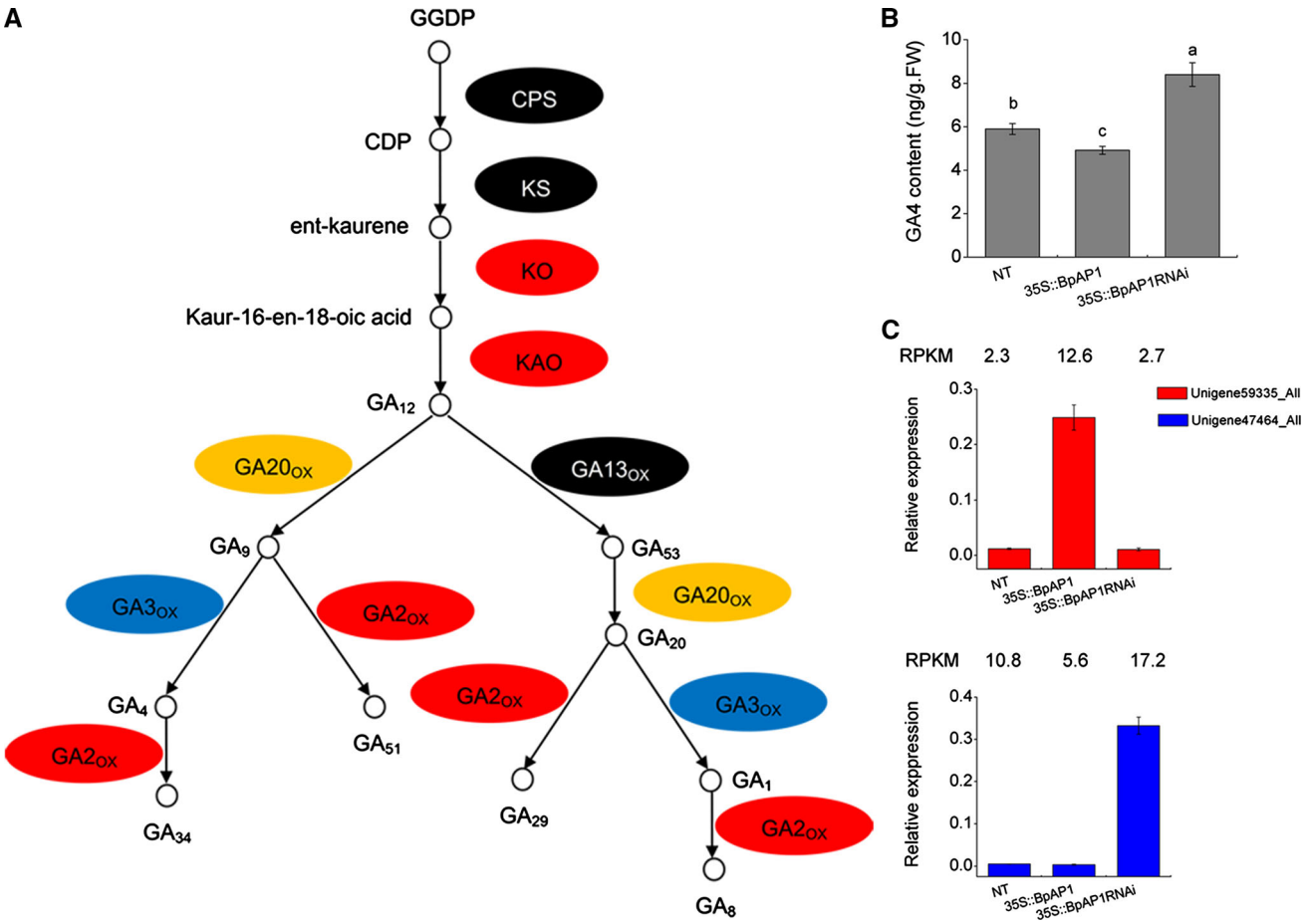

Fig. 6 Diterpenoid biosynthesis. **a** Diterpenoid biosynthesis pathway. *GGDP* geranylgeranyl diphosphate, *CPS* ent-copalyl diphosphate synthase, *CDP* ent-copalyl diphosphate, *KS* ent-kaurene synthase, *KO* ent-kaurene oxidase, *KAO* ent-kaurenoic acid oxidase, *GA20ox* gibberellin 20-oxidase, *GA13ox* gibberellin 13-oxidase, *GA3ox* gibberellin 3-beta-dioxygenase, *GA2ox* gibberellin 2-oxidase. *Red* indicates up-regulation, *blue* indicates down-regulation, *yellow* indicates that the expression of the gene encoding an enzyme was up-regulated or down-regulated and *black* indicates that there was no change in expression. **b** GA_4 content. GA_4 content in the data accompanied by different letters indicates a significant difference ($n = 18$, $P < 0.05$; Duncan's multiple range analysis). The *error bars* represent SD. **c** Real-time quantitative PCR verification of GA2ox (Unigene59335_All) and GA3ox (Unigene47464_All). The Illumina/Solexa sequencing results are indicated above each *bar*. The relative expression levels were determined by real-time quantitative PCR using *18srRNA* and *Tubulin* for internal calculations. Each PCR was repeated three times and the *error bars* represent SD (color figure online)

lines and NT, GO categories were assigned to the 6183 significant genes based on TAIR GO slim provided by Blast2GO. The categorization of DEUs according to the biological process, cellular component, and molecular function ontology is shown in Fig. 5. For cellular component, the analysis revealed a high percentage of cell (3504 unigenes, 56.67 %), cell parts (3160 unigenes, 51.11 %), and organelles (2325 unigenes, 37.60 %). For molecular function, the unigenes were classified into nine categories, as shown in Fig. 4; the most two overrepresented GO terms were catalytic activity (3048 unigenes, 49.30 %) and binding (2704 unigenes, 43.73 %). DEUs were related to 26 biological processes, including metabolic process, cellular process, response to stimulus, localization, establishment of localization, developmental process, and others (Fig. 5).

Pathway analysis of DEUs

Significant DEUs were also subjected to KEGG pathway analysis. There were 7802 DEUs mapped to 1242 KEGG pathways, and 31 pathways were significantly enriched (Q value <0.05), which included biosynthesis of secondary metabolites, phenylpropanoid biosynthesis, plant-pathogen interaction, stilbenoid, diarylheptanoid, and gingerol biosynthesis and metabolic pathways.

Diterpenoid biosynthesis

A previous study demonstrated that 35S::*BpAP1* transgenic birch exhibits a significantly shortened the juvenile phase, and 1-year-old seedlings are obviously dwarfed (Huang et al. 2014). Therefore, we analyzed diterpenoid

biosynthesis to identify growth and metabolism-related genes in transgenic birch. We identified 11 DEUs involved in diterpenoid biosynthesis, including genes encoding ent-kaurene oxidase (KO), ent-kaurenoic acid oxidase (KAO), gibberellin 20-oxidase (GA20ox), gibberellin 3-beta hydroxylase (GA3ox), and gibberellin 2-oxidase (GA2ox), key enzymes in the process of GA biosynthesis (Fig. 6a) (Kaufmann et al. 2010; Yamaguchi 2008); KO, KAO, and GA2ox genes were significantly up-regulated in the 35S::*BpAP1* transgenic line (Table 2).

The expression of genes encoding GA2ox (Unigene59335_All) and GA3ox (Unigene47464_All) was verified by qRT-PCR, revealing that GA2ox was significantly up-regulated in the 35S::*BpAP1* transgenic line; however, GA3ox was significantly up-regulated in the 35S::*BpAP1RNAi* transgenic line (Fig. 6c).

Identification of flowering-related unigenes and their differential expression

In this study, we identified 256 flowering-related unigenes, including floral integrator or floral identity genes *AP1*, *SOC1*, *FT*, and *LFY*, photoperiod genes *CO*, *FKF1*, and *GI*, vernalization pathway genes *FLC* and *PIE*, the autonomous pathway gene *FCA*, and flowering-related genes *MADS5*, *ELF4*, *SPL*, *AGL*, *TFL1*.

Compared with NT, the expression of *AP1* was significantly high in the 35S::*BpAP1* library, with 1160.07-fold greater expression than in NT. However, the expression of *AP1* was significantly low in the 35S::*BpAP1RNAi* library. Moreover, the expression of *SOC1*, *FT*, and *LFY* was up-regulated in the 35S::*BpAP1* library (Fig. 7; Table S3). We found two *FT* homologs, which were significantly up-regulated in the 35S::*BpAP1* library, suggesting that the up-regulation of *AP1* promotes the up-regulation of *FT*. By contrast, there was only one *FT* homolog in the 35S::*BpAP1RNAi* library, and its expression was lower than in NT. This result suggests that the knockout of *AP1* affected the expression of *FT*.

SOC1 encodes an MADS-box transcription factor. We found two unigenes homologous to *Citrus* genes (*CsSOC1-1*, *CsSOC1-2*) and one unigene homologous to a *Vitis* gene (*VvSOC1-1)* in this study; the expression of these genes was 8.83-, 1.62- and 7.6-fold that of NT in the 35S::*BpAP1* library, and 2.14-, 21.86-, and 2.14-fold that of NT in the 35S::*BpAP1RNAi* library, respectively (Table S3).

We also identified the number of circadian rhythm-related homologous unigenes in the photoperiod pathway, including *CO*, *ZTL*, *FKF1*, *CRY*, *LHY*, *CCA1*, *TOC1,* and so on (Table S4). *CO* encodes a zinc finger transcription factor in *Arabidopsis*, which is a flowering activator and a key regulator of the photoperiod response (Hsu et al. 2006). In this study, we found nine *CO* genes homologous with those of *Malus*, three with *Arabidopsis*, three with *Glycine*, and one with *Prunus*. These unigenes were all DEUs, including both up-regulated and down-regulated genes. These results imply that there is more than one *CO* gene in *Betula platyphylla*, and these genes function together. In the 35S::*BpAP1* library, *ZTL*, *FKF1*, *GI*, and *TOC1* had significantly up-regulated expression, while *LHY* and *CCA1* had significantly down-regulated expression. Moreover, these genes had opposite or slightly different expression patterns in the 35S::*BpAP1RNAi* library.

For genes in the vernalization pathway, the expression of *FLC* and *PIE1* was lower in the 35S::*BpAP1* library, with 0.83- and 0.23-fold expression levels compared with that in the NT library. However, in the 35S::*BpAP1RNAi* library, the expression of *FLC* was 1.27 times that of NT

Table 2 Differentially expressed genes involved in diterpenoid biosynthesis in transgenic and non-transgenic lines

Pathway	GI	*E* value	Expression $\log_2$(35S::*BpAP1*/ NT RPKM)	Expression $\log_2$(35S::*BpAP1RNAi*/ NT RPKM)	Annotation
Diterpenoid biosynthesis	293792354	0	1.1	0.2	Ent-kaurene oxidase
	197209780	1.00E−140	1.3	0.3	Ent-kaurenoic acid oxidase 1/cytochrome P450 88D6
	197209774	3.00E−36	1.2	0.3	Ent-kaurenoic acid oxidase 2/cytochrome P450 88D3
	18496057	0	2.0	0.7	Gibberellin 20-oxidase 1
	224123248	1.00E−159	1.2	0.3	Gibberellin 20 oxidase 1
	68509984	1.00E−155	−2.4	1.1	Gibberellin 20 oxidase 2
	224096043	1.00E−107	−3.2	−0.3	Gibberellin 20 oxidase 2
	61651585	1.00E−169	−2.6	0.4	Gibberellin 3-beta-dioxygenase
	224114543	1.00E−156	1.2	1.0	Gibberellin 2-oxidase
	254935149	1.00E−151	2.5	0.3	Gibberellin 2-oxidase
	224064641	1.00E−127	2.5	1.4	Gibberellin 2-oxidase

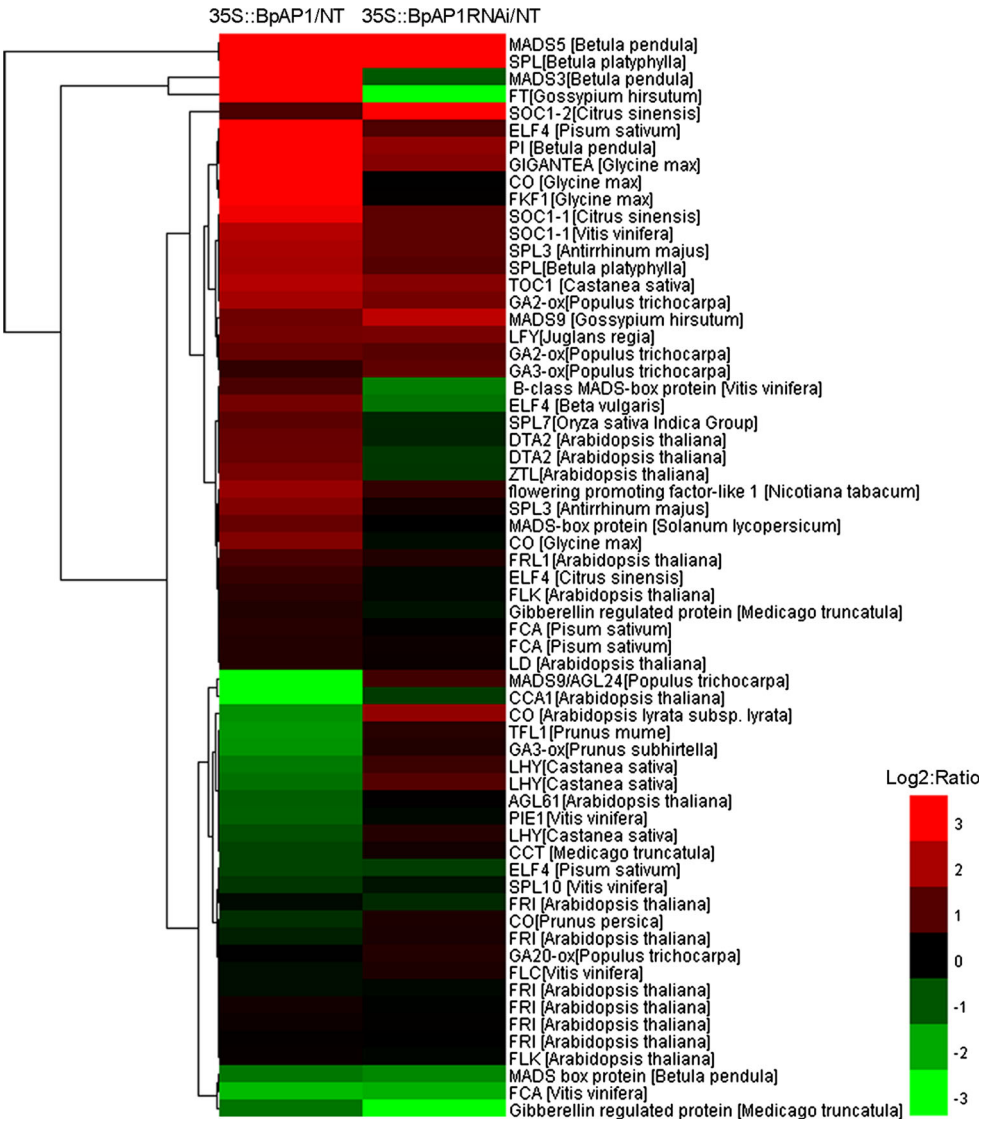

Fig. 7 Expression pattern of flowering-related genes in transgenic lines. Cluster analysis was performed using log2 ratios for 63 flowering-related genes identified by RNA-seq analysis

(Table S4). In the autonomous pathway, the expression of *FCA* and *LD* did not obviously differ between libraries (Fig. 7; Table S4). These genes encode flowering repressors, and lower expression of these genes can accelerate flowering.

Moreover, we identified some genes that have not been placed into any specific flowering pathway, such as early flowering *ELF* genes, *MADS-box* family genes, *SBP* family genes, *TFL1* and so on (Table S5). We identified two *ELF4* unigenes homologous to *Pisum sativum*; in the 35S::*BpAP1* library, their expression levels were 56.60- and 2.76-fold that of NT, while in the 35S::*BpAP1RNAi* library, their expression levels were 1.92- and 0.61-fold that of NT, respectively. *TFL1* encodes a suppressor of flowering. The expression of *TFL1* was 0.19-fold that of NT in the 35S::*BpAP1* library and 1.39-fold that of NT in the 35S::*BpAP1RNAi* library (Table S5).

Basically, most of the DEUs that we identified in transgenic and non-transgenic lines are flowering-related genes. Compared with the NT, expression of *BpAP1* was greatly up-regulated in the 35S::*BpAP1* library, and the expression of *SOC1*, *FT*, *MADS5*, *ELF4*, *SVP*, and *FKF1* was also significantly up-regulated. On the other hand, most flowering repressor genes were down-regulated in the 35S::*BpAP1* library compared with that in the NT library (Fig. 7; Tables S3, 4, 5).

qRT-PCR verification

To test the reliability of RNA-Seq, qRT-PCR was performed with specific primers (Table S1) for 20 unigenes, including 16 DEUs and four non-differentially expressed unigenes. The transcript abundance patterns of the transgenic and NT lines were compared with the RNA-Seq data.

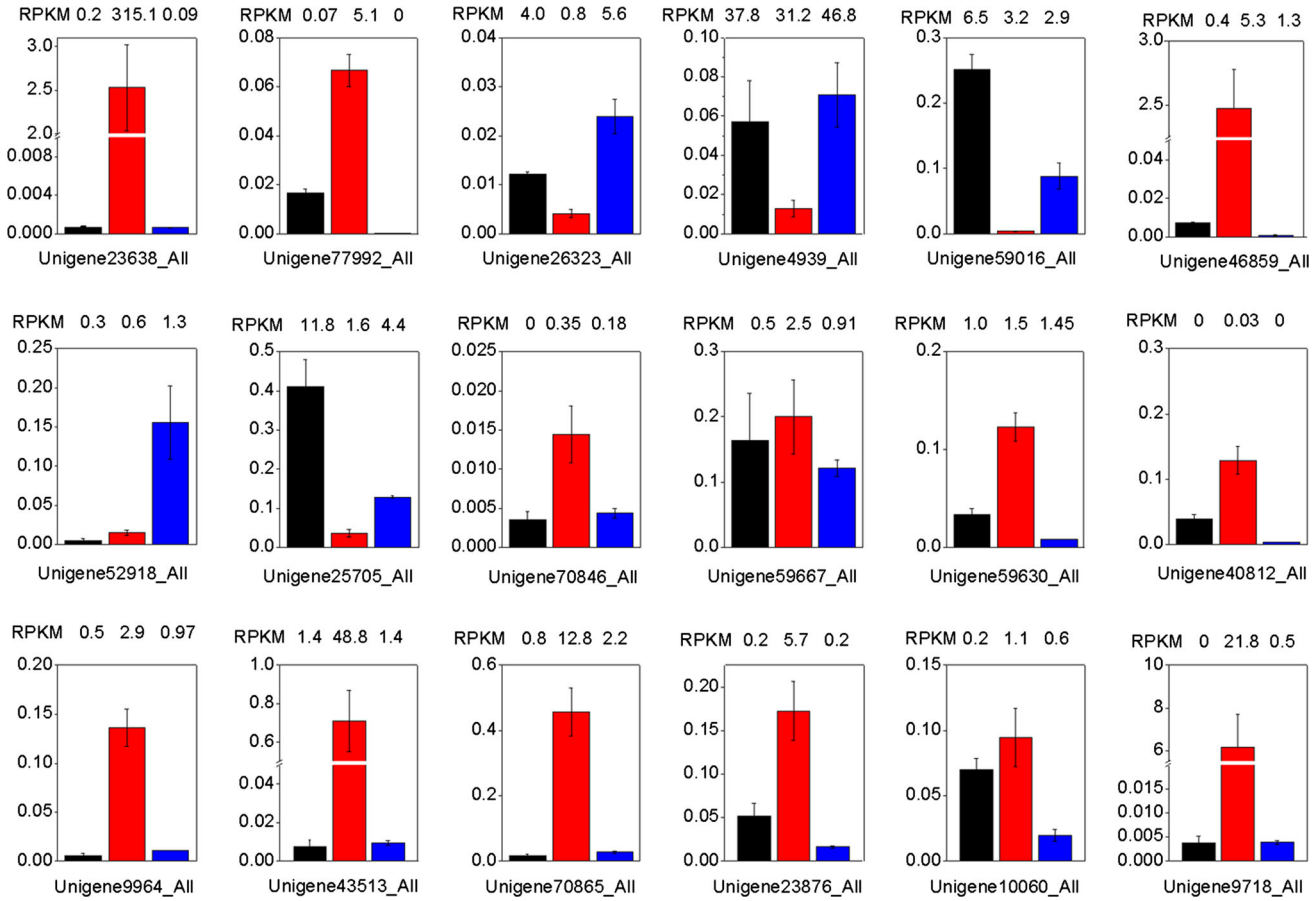

Fig. 8 Quantitative real-time PCR verification of RNA-seq. Differential expression among non-transgenic line (*black column*), 35S::*BpAP1* transgenic line (*red column*), and 35S::*BpAP1RNAi* transgenic line (*blue column*). The Illumina/Solexa sequencing results are indicated above each *bar*. The relative expression levels were determined by quantitative real-time PCR using *18srRNA* and *Tubulin* were used for internal calculations; each PCR was repeated three times and the *error bars* represent SD (color figure online)

All 18 selected genes displayed expression patterns consistent with the RNA-Seq results (Fig. 8).

Prediction of *BpAP1* target genes

Based on the birch genome and binding motif of MADS domain transcription factors, i.e., CArG box (CC[W]6GG) and CArG box-like (C[W]7GG and CC[W]7G) (Kaufmann et al. 2010; Riechmann et al. 1996), we predicted target genes of *BpAP1* and identified 468 genes as putative *BpAP1* targets. These genes were classified according to their functions (Table 3). The protein kinase category represented the largest number of target genes, and we also identified genes in the categories: zinc finger-containing, cytochrome P450, cyclin-like, MADS-box transcription factor, plant peroxidase, and protein of unknown function.

AP1 encodes a transcription factor, and its overexpression and suppression can affect the expression of target genes. To more accurately identify target genes, we performed Blast analysis of *BpAP1* target genes against the RNA-seq and birch genome databases using BioEdit. We identified 275 differentially expressed target genes (DETGs) in the 35S::*BpAP1* library, including 133 up-regulated genes and 142 down-regulated genes. Of the 181 DETGs identified in the 35S::*BpAP1RNAi* library, 84 were up-regulated and 97 were down-regulated. There were 166 common DETGs in the 35S::*BpAP1* and 35S::*BpAP1RNAi* libraries, and cluster analysis revealed 91 up-regulated genes and 75 down-regulated genes in 35S::*BpAP1* and 86 up-regulated genes and 80 down-regulated genes in 35S::*BpAP1RNAi* (Fig. 9).

Furthermore, we analyzed the expression levels of *BpAP1* DETGs and determined that 82.76 and 72.41 % of the Protein kinase DETGs were down-regulated in the 35S:*BpAP1* and 35S:*BpAP1RNAi* lines, respectively, while DETGs in the categories cyclin-like, MADS-box transcription factor, and disease resistance protein were up-regulated in the 35S:*BpAP1* line (Fig. 9).

Table 3 Predicted target genes of *BpAP1*

Classification of target genes	Gene number
Protein kinase	37
Protein of unknown function	13
Zinc finger-containing	12
Cytochrome P450	10
F-box domain, cyclin-like	7
Glycoside hydrolase	7
Ubiquitin	6
Pathogenesis-related	6
Disease resistance protein	5
Plant peroxidase	5
No apical meristem (NAM) protein	5
Auxin related	5
Peptidase	5
Ankyrin repeat	4
Glutaredoxin	3
Heavy metal transport/detoxification protein	3
ATPase	3
MADS-box transcription factor	2
Thioredoxin	2
Cellulose synthase	2
Nuclear transport factor	2
Others	…

Phylogenetic analysis of *BpAP1* and two MADS-box target genes

We searched the known MADS-box protein sequences in the NCBI and EST databases and employed ClustalX1.83 to perform the multiple alignment of the two putative *B. platyphylla* MADS-box transcription factors and 14 other MADS-box proteins. Phylogenetic tree analysis revealed that *PI* in *B. platyphylla* exhibits high identity (98 %) at the amino acid level with *BpMADS2* (GI: 28874430) and 57 % identity with *PI* in *Arabidopsis* (GI: 15241299); *AP3* in *B. platyphylla* exhibits 68 % identity at the amino acid level with *TM6* in *Vitis vinifera* (GI: 526118095) and 53 % identity with *AP3* in *Arabidopsis* (GI: 5805232; Fig. 10). According to the phylogenetic analysis results, the two MADS-box transcription factors were designated as *BpPI* and *BpAP3*.

Expression correlation of *BpAP1* vs. *BpPI* and *BpAP1* vs. *BpAP3*

Analysis of RNA-seq results and the birch genome sequence revealed that of the 166 target genes, only two encoded MADS-box transcription factors, i.e., *BpPI* and *BpAP3*. Further analysis of *BpPI* and *BpAP3* revealed that the expression of *BpPI* was significantly higher in the 35S:*BpAP1* transgenic line than in NT, with expression 16.49-fold that of NT, while the expression of this gene in the 35S:*BpAP1RNAi* transgenic line was 3.27-fold that of NT (Table 4). The expression of *BpAP3* was higher in the 35S:*BpAP1* than in NT, with expression 2.22-fold that of NT; and slightly lower in the 35S:*BpAP1RNAi* than in NT, with expression 0.36-fold that of NT (Table 4). Correlation analysis was employed to clarify the relationship between *BpAP1* and its putative targets. The results revealed a highly significant positive correlation between *BpAP1* and *BpPI* ($r = 0.939$; $P < 0.01$; $n = 10$).

Expression analysis of *BpAP1* and *BpPI*

The expression patterns of *BpPI* and *BpAP1* in the *BpAP1* overexpression and WT lines were very similar (Fig. 11a). In all of these lines, the expression of both genes was higher in male inflorescences than in other tissues, and with high levels of expression observed in male inflorescences at all eight developmental stages (Fig. 11b). Moreover, the expression of *BpPI* in male inflorescences was significantly higher than that of *BpAP1* in WT, with the expression of *BpPI* 222.90-fold that of *BpAP1*. In addition, the relative expression of *BpPI* in leaves, terminal buds, and roots was significantly higher in the 35S::*BpAP1* transgenic line than in the WT, with the expression of this gene 60.48-fold that of WT. The relative expression of *BpPI* in leaves was higher in the 35S::*BpAP1* transgenic line than in the 35S::*BpAP1 RNAi* transgenic line (Fig. 11a). Perhaps the overexpression of *BpAP1* in the 35S::*BpAP1* transgenic line caused higher expression of *BpPI*.

We analyzed relative expression of *BpAP1* and *BpPI* in the WT and 35S::*BpAP1* transgenic line at three different stages of development. The results showed that *BpAP1* expressed at a significant level during the processes of female and male inflorescence development. At the ds9 stage (17 April) and ds11 stage (27 April) of female inflorescence development, *BpAP1* expression was significantly up-regulated. As the male inflorescences developed, *BpAP1* expression increased, while *BpPI* was expressed at every stage of male inflorescence development, and there was very weak expression in female inflorescences. It appears that *BpPI* is a floral organ-specific gene in birch (Fig. 11d). However, the relative expression of *BpAP1* and *BpPI* during the three stages of female and male inflorescence development was higher than in the 35S::*BpAP1* transgenic line than in the WT (Fig. 11d). Perhaps the high level of expression of *BpAP1* led to strong expression of *BpPI*, as *BpPI* expression in male inflorescences was higher in the 35S::*BpAP1* transgenic line than in the WT at all three stages examined.

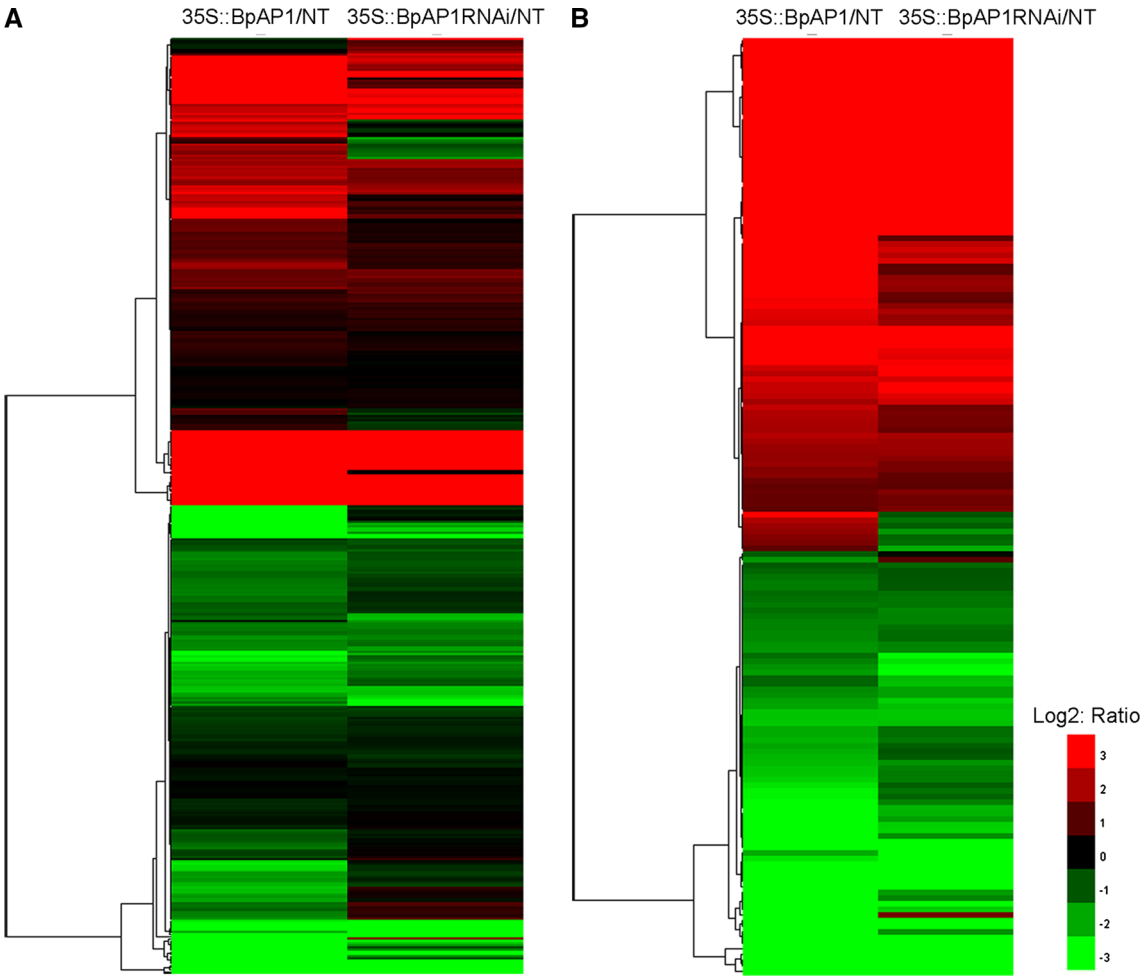

Fig. 9 Expression analysis of target genes of *BpAP1*. **a** Heat map of 424 target genes of *BpAP1* in the 35S:*BpAP1* and 35S:*BpAP1RNAi* lines; **b** heat map of 166 DETGs (differential expression target genes) of *BpAP1* in the 35S:*BpAP1* and 35S:*BpAP1RNAi* lines

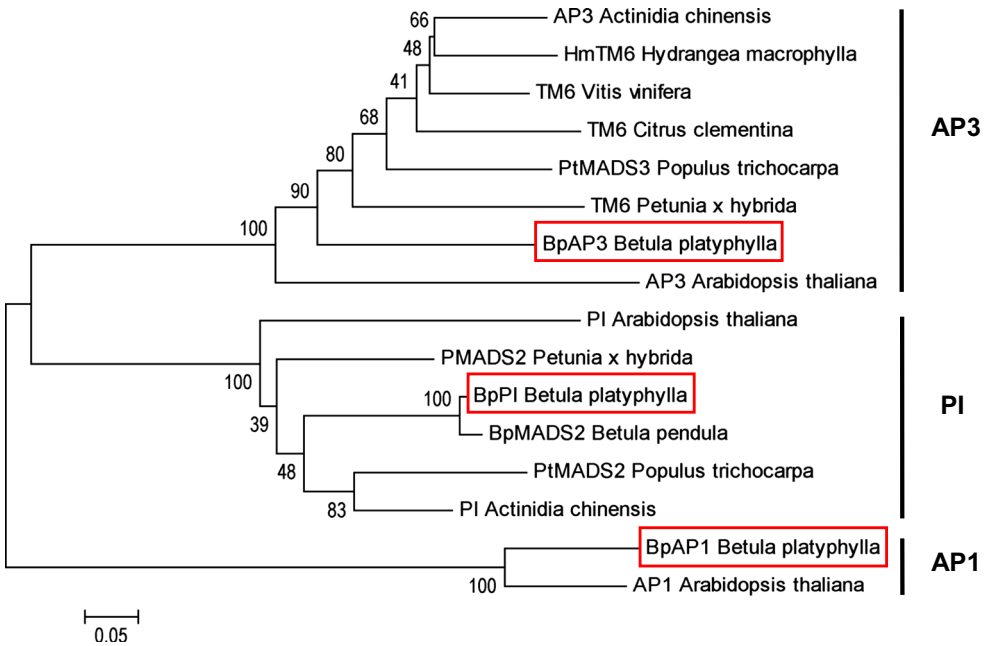

Fig. 10 Phylogenetic analysis of *PI*- and *AP3*-related genes. The tree was constructed with MEGA5 using the neighbor-joining method. The *numbers* show bootstrap values for each node

Table 4 Expression of *BpAP1*, *BpPI*, and *BpAP3* in transgenic and non-transgenic lines

Gene id	GI	Expression $\log_2$(35S:*BpAP1*/ NT RPKM)	Expression $\log_2$(35S:*BpAP1RNAi*/ NT RPKM)	Nr annotation
Unigene23638_All	1483228	10.1752	−1.0402	*MADS3* (*Betula pendula*)
Unigene46859_All	28874430	4.04391	1.71108	*PISTILLATA* homologue (*Betula pendula*)
Unigene31233_All	115492982	1.1482	−1.489	Flowering-related B-class MADS-box protein (*Vitis vinifera*)

Discussion

Diterpenoid biosynthesis is affected by *BpAP1*

Gibberellins, a class of diterpenoids, involve in regulating plant growth and development that accelerate plant growth, and promote flowering (Garcia-Martinez et al. 1997). KO, KAO, GA2ox, and GA3ox are key enzymes in diterpenoid biosynthesis. GA2ox can regulate the levels of bioactive GA in the cell, thereby regulating plant growth and development. GA3ox catalyzes the conversion of precursor GAs to their bioactive forms (Dijkstra et al. 2008; Huang et al. 2010; Olszewski et al. 2002; Thomas et al. 1999). Transfer of *PcGA2ox1* into *S. nigrum* and *S. melanocerasum* produces dwarf transgenic plants (Dijkstra et al. 2008). In the current study, analysis of RNA-seq data revealed that compared with NT, the 35S::*BpAP1* and 35S::*BpAP1RNAi* transgenic lines had significantly different levels of expression of genes involved in response to stimulus, diterpenoid biosynthesis, plant hormone signal transduction, and others. These results suggest that *GA2ox* is significantly up-regulated in the 35S::*BpAP1* transgenic line, which may lead to their dwarf phenotype. Meanwhile, *GA3ox* is significantly up-regulated in the 35S::*BpAP1* transgenic line, which may be related to increase in height.

Regulatory relationship between *BpAP1* and the major flowering-related genes

Genes involved in inflorescence meristem and floral meristem identities have rarely been studied in woody plants. In fruit trees, *AP1* and *LFY* are positively correlated with flowering, while *TFL1* is negatively correlated with flowering (Esumi et al. 2007; Pelaz et al. 2001; Pillitteri et al. 2004). Our RNA-seq data show that the expression of *AP1* up-regulated the expression of *LFY* and down-regulated the expression of *TFL1* (Tables S3, 5). These results suggest that changes in the expression of these genes may play an important role in the transition from vegetative growth to reproductive growth in birch.

Many studies have demonstrated that the FT protein is a long-distance signal that moves directly from leaves to the shoot apex (Turck et al. 2008). FT executes its role through an interaction with FD, which is expressed specifically at the shoot apex, where the FT/FD complex may activate downstream MADS-box genes such as *AP1* (Abe et al. 2005; Wigge et al. 2005). In addition, overexpression of *FT* can promote early flowering in *Arabidopsis* (Hsu et al. 2006). In the current study, expression of *FT* was significantly up-regulated in the 35S::*BpAP1* transgenic line and down-regulated in the 35S::*BpAP1RNAi* transgenic line (Table S3). These results indicate that although *AP1* functions downstream of *FT*, overexpression of this gene in birch may have an effect on *FT*. High levels of expression of both of these genes may shorten juvenile phase in birch, leading to early flowering.

SOC1 integrates multiple flowering-related signals and is regulated by the activator *CO* and the repressor *FLC*, which play roles in the photoperiod and vernalization pathways, respectively (Samach et al. 2000; Searle et al. 2006). *CO* primarily activates *SOC1* through its effect on *FT*, while *FLC* directly suppresses *SOC1* and represses *FT* by forming a complex with SVP protein (Lee et al. 2007). When *SOC1* is induced in the shoot apex, it directly activates the floral meristem identity gene *LFY* in conjunction with *AGL24*, and subsequently, *LFY* and *FT* directly regulated *AP1* and thus, the flowering transition. Additionally, miR156-SPL protein binds to the promoter of the *AP1* ortholog *SQUAMOSA* and is expressed upstream of *LFY* and *AP1*, which are then activated at the shoot apex (Cardon et al. 1997; Klein et al. 1996). *SPL* genes affect flowering, as they are downstream targets of *FT* at the shoot apex and play an important *FT*-independent role in regulating flowering. SPLs and the FT/FD complex share several direct targets, including *AP1*, and they accelerate flowering (Wang et al. 2009a).

We identified *SOC1* and *LFY* homologous genes in the RNA-seq libraries. These genes were up-regulated in the 35S::*BpAP1* transgenic line, while the flowering repressor gene *FLC* was down-regulated (Table S4). Most homologous genes of *CO* and *SPL* were significantly up-regulated in the 35S::*BpAP1* transgenic line. These results suggest that *AP1* overexpression in birch induces differential expression of these flowering-related genes, leading to early flowering in the 35S::*BpAP1* transgenic line.

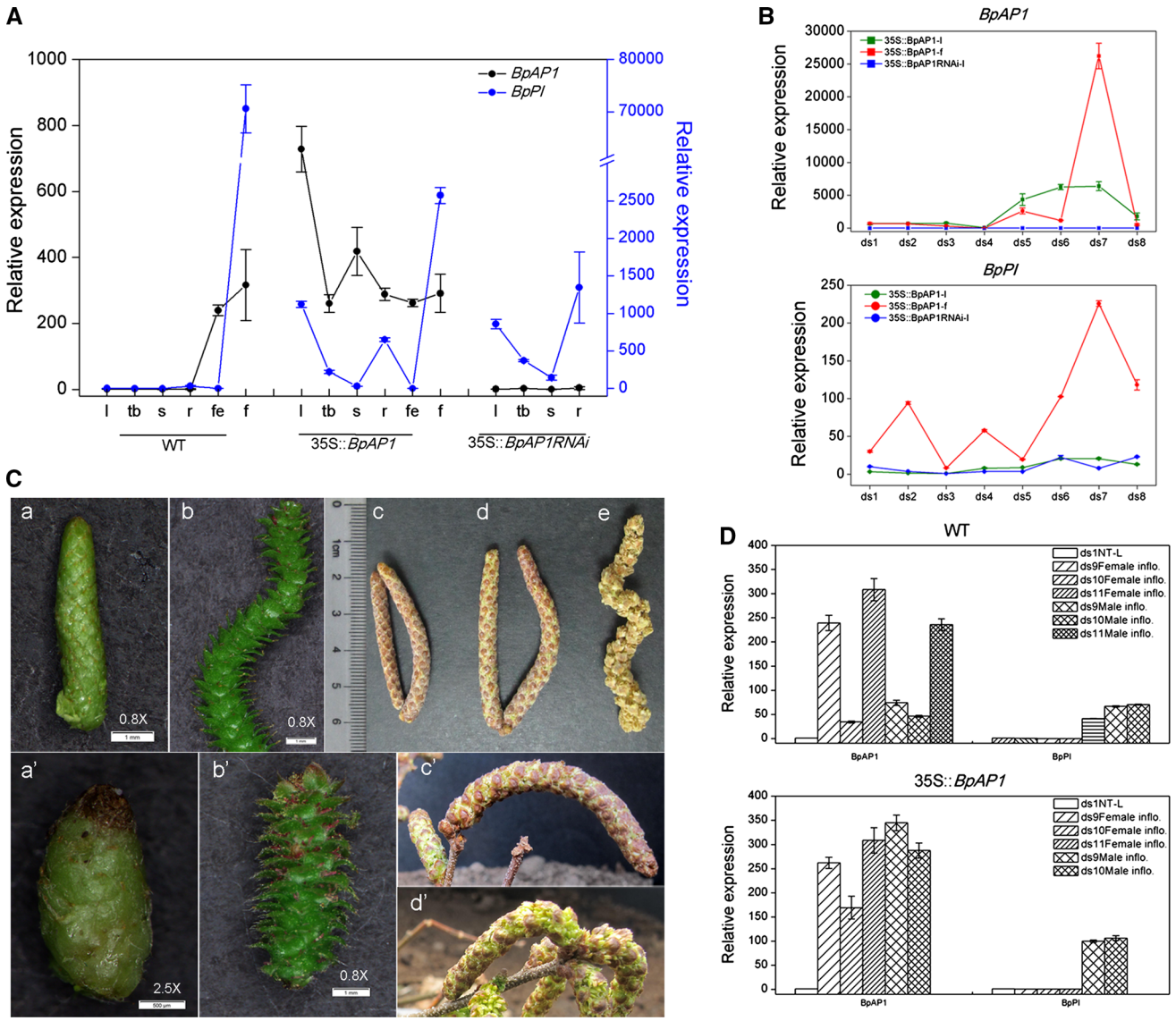

Fig. 11 Expression pattern of *BpAP1* and *BpPI*. **A** Expression of *BpAP1* and *BpPI* in different tissues of wild-type and the transgenic plants. *l* leaf, *tb* terminal bud, *s* shoot, *r* root, *fe* female inflorescence, *f* male inflorescence. The *error bars* represent SD. **B** Expression of *BpAP1* and *BpPI* in different development stages of wild-type and the transgenic plants. *35S::BpAP1-l* leaves of 35S::*BpAP1* transgenic lines, *35S::BpAP1-f* male inflorescences of 35S::*BpAP1* transgenic lines, *35S::BpAP1RNAi-l* leaves of 35S::*BpAP1RNAi* transgenic line. **C** Samples used for RNA extraction. **a** Female inflorescence of WT at ds9; **a′** female inflorescence of 35S::*BpAP1* transgenic line at ds9, **b** female inflorescence of WT at ds11, **b′** female inflorescence of 35S::*BpAP1* transgenic line at ds11, **c** male inflorescence of WT at ds9, **c′** male inflorescence of 35S::*BpAP1* transgenic line at ds9, **d** male inflorescence of WT at ds10, **d′** male inflorescence of 35S::*BpAP1* transgenic line at ds10, **e** male inflorescence of WT at ds11. **D** Expression of *BpAP1* and *BpPI* in male and female inflorescences at different developmental stages. The *error bars* represent SD

Prediction of *BpAP1* target genes

ChIP and microarray analyses were employed to identify *AP1* binding sites and target genes, revealing 1,942 regions bound by *AP1*, most of which are located in transcription factor genes. When *AP1* was activated, approximately 13 % of the target genes were up-regulated, while most of the targets were inhibited by *AP1* (Kaufmann et al. 2010). In the current study, we identified 166 putative targets of *BpAP1*, including genes encoding protein kinase, cytochrome P450, cyclin-like, plant peroxidase, and two transcription factors *BpPI* and *BpAP3*. Correlation analysis showed that a highly significant positive correlation exists between *BpAP1* and *BpPI*.

Both *AP1* and *PI* are MADS-box family genes that encode transcription factors. *AP1* and *PI* are both involved in floral organ development, and they encode proteins that control the development of petals and

stamens in dicotyledons. *PI* is a class of B-function genes in the ABC Model, which regulate the development of petals and stamens in *Arabidopsis*. In maize, the *PI* gene is expressed in seeds, leaves, and roots. *VvPI* expression is associated with inflorescences at different developmental stages, whereas it is not expressed in fruits and is expressed at very low levels in roots and leaves, Real-time RT-PCR revealed that *VvPI* expression is over 400 times higher in inflorescences than in leaves or roots (Poupin et al. 2007). *BpMADS2* is highly expressed in male inflorescences but exhibits little or no expression in female inflorescence (Jarvinen et al. 2003). Our results indicate that *BpPI* is a floral organ-specific gene in birch, and it may control the development of stamens, further supporting the notion that *BpPI* is a B-function gene resembling *PI* (Fig. 11d).

Although much is known about flower morphology and development, little is known about the mechanism underlying the transition from vegetative growth to reproductive growth in plants, especially woody plants. In the current study, we primarily analyzed gene expression based on the transcriptome; our results provide useful resources and information about *AP1* for future studies. Of course, the roles of metabolic pathways and genes in the process of plant growth and development are quite complicated. In future studies, we will employ ChIP-Seq to accurately detect the position of the binding sites of *BpAP1* throughout the entire genome, verify the target genes of *BpAP1,* and elucidate the transcriptional regulatory function of *BpAP1* in birch.

Author contribution statement Haijiao Huang extracted RNA of the samples, did the data processing, qRT-PCR verification, and wrote the article. Su Chen predicted the target genes of *BpAP1*. Huiyu Li extracted RNA of female and male inflorescences in different developmental stages. Professor Jing Jiang guided the experiment and contributed to write the article.

Acknowledgments This work was supported by the Fundamental Research Funds for the Central Universities, 2572015BA02.

Conflict of interest The authors declare that they have no conflict of interest.

References

Abe M, Kobayashi Y, Yamamoto S, Daimon Y, Yamaguchi A, Ikeda Y, Ichinoki H, Notaguchi M, Goto K, Araki T (2005) FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. Science 309:1052–1056

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. Nat Genet 25(1):25–29

Audic S, Claverie JM (1997) The significance of digital gene expression profiles. Genome Res 7:986–995

Cardon GH, Hohmann S, Nettesheim K, Saedler H, Huijser P (1997) Functional analysis of the *Arabidopsisthaliana* SBP-box gene *SPL3*: a novel gene involved in the floral transition. Plant J 12:367–377

Chang S, Puryear J, Cairney J (1993) A simple and efficient method for isolating RNA from pine trees. Plant Mol Biol Rep 11:113–116

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676

de Hoon MJ, Imoto S, Nolan J, Miyano S (2004) Open source clustering software. Bioinformatics 20(9):1453–1454

Dijkstra C, Adams E, Bhattacharya A, Page AF, Anthony P, Kourmpetli S, Power JB, Lowe KC, Thomas SG, Hedden P, Phillips AL, Davey MR (2008) Over-expression of a gibberellin 2-oxidase gene from *Phaseoluscoccineus* L. enhances gibberellin inactivation and induces dwarfism in *Solanum* species. Plant Cell Rep 27:463–470

Esumi T, Tao R, Yonemori K (2007) Relationship between floral development and transcription levels of *LEAFY* and *TERMINAL-FLOWER1* homologs in Japanese pear (*Pyruspyrifolia* Nakai) and quince (*Cydoniao blonga* Mill.). J Jpn Soc Hortic Sci 76:294–304

Garcia-Martinez JL, Lopez-Diaz I, Sanchez-Beltran MJ, Phillips AL, Ward DA, Gaskin P, Hedden P (1997) Isolation and transcript analysis of gibberellin 20-oxidase genes in pea and bean in relation to fruit development. Plant Mol Biol 33:1073–1084

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QD, Chen ZH, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652

Hsu CY, Liu Y, Luthe DS, Yuceer C (2006) Poplar *FT2* shortens the juvenile phase and promotes seasonal flowering. Plant Cell 18:1846–1861

Huang H, Wang S, Jiang J, Liu G, Li H, Chen S, Xu H (2014) Overexpression of *BpAP1* induces early flowering and produces dwarfism in *Betula platyphylla* × *Betula pendula*. Physiol Plant 151:495–506

Huang J, Lu X, Yan H, Chen S, Zhang W, Huang R, Zheng Y (2012) Transcriptome characterization and sequencing-based identification of salt-responsive genes in *Millettia pinnata*, a semi-mangrove plant. DNA Res 19:195–207

Huang J, Tang D, Shen Y, Qin B, Hong L, You A, Li M, Wang X, Yu H, Gu M, Cheng Z (2010) Activation of gibberellin 2-oxidase 6 decreases active gibberellin levels and creates a dominant semi-dwarf phenotype in rice (*Oryzasativa* L.). J Genet Genomics 37:23–36

Irish VF, Sussex IM (1990) Function of the apetala-1 gene during *Arabidopsis* floral development. Plant Cell 2:741–753

Jack T (2004) Molecular and genetic mechanisms of floral control. Plant Cell 16(suppl 1):S1–S17

Jarvinen P, Lemmetyinen J, Savolainen O, Sopanen T (2003) DNA sequence variation in *BpMADS2* gene in two populations of *Betula pendula*. Mol Ecol 12:369–384

Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27–30

Kaufmann K, Wellmer F, Muino JM, Ferrier T, Wuest SE, Kumar V, Serrano-Mislata A, Madueno F, Krajewski P, Meyerowitz EM, Angenent GC, Riechmann JL (2010) Orchestration of floral initiation by *APETALA1*. Science 328:85–89

Klein J, Saedler H, Huijser P (1996) A new family of DNA binding proteins includes putative transcriptional regulators of the *Antirrhinum majus* floral meristem identity gene *SQUAMOSA*. Mol Gen Genet 250:7–16

Lamb RS, Hill TA, Tan QKG, Irish VF (2002) Regulation of *APETALA3* floral homeotic gene expression by meristem identity genes. Development 129:2079–2086

Lee JH, Yoo SJ, Park SH, Hwang I, Lee JS, Ahn JH (2007) Role of *SVP* in the control of flowering time by ambient temperature in *Arabidopsis*. Gene Dev 21:397–402

Lemmetyinen J, Pennanen T, Lännenpää M, Sopanen T (2001) Prevention of flower formation in dicotyledons. Mol Breed 7:341–350

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. Methods 25:402–408

Lloyd G, McCown B (1980) Commercially-feasible micropropagation of mountain laurel, *Kalmia latifolia*, by use of shoot-tip culture. Combined Proc Int Plant Prop Soc 30:421–427

Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res 18:1509–1517

Michaels SD, Ditta G, Gustafson-Brown C, Pelaz S, Yanofsky M, Amasino RM (2008) *AGL24* acts as a promoter of flowering in *Arabidopsis* and is positively regulated by vernalization. Plant J 33(5):867–874

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5:621–628

Ng M, Yanofsky MF (2001) Activation of the Arabidopsis B class homeotic genes by *APETALA1*. Plant Cell 13(4):739–753

Norusis M (2008) SPSS 16.0 statistical procedures companion. Prentice Hall Press

Olszewski N, Sun TP, Gubler F (2002) Gibberellin signaling: biosynthesis, catabolism, and response pathways. Plant Cell 14(Suppl 1):S61–S80

Pelaz S, Gustafson-Brown C, Kohalmi SE, Crosby WL, Yanofsky MF (2001) *APETALA1* and *SEPALLATA3* interact to promote flower development. Plant J 26:385–394

Pillitteri LJ, Lovatt CJ, Walling LL (2004) Isolation and characterization of a *TERMINAL FLOWER* homolog and its correlation with juvenility in citrus. Plant Physiol 135(3):1540–1551

Poupin MJ, Federici F, Medina C, Matus JT, Timmermann T, Arce-Johnson P (2007) Isolation of the three grape sub-lineages of B-class MADS-box *TM6*, *PISTILLATA* and *APETALA3* genes which are differentially expressed during flower and fruit development. Gene 404:10–24

Riechmann JL, Wang M, Meyerowitz EM (1996) DNA-binding properties of *Arabidopsis* MADS domain homeotic proteins *APETALA1*, *APETALA3*, *PISTILLATA* and *AGAMOUS*. Nucleic Acids Res 24:3134–3141

Samach A, Onouchi H, Gold SE, Ditta GS, Schwarz-Sommer Z, Yanofsky MF, Coupland G (2000) Distinct roles of *CONSTANS* target genes in reproductive development of *Arabidopsis*. Science 288:1613–1616

Searle I, He YH, Turck F, Vincent C, Fornara F, Krober S, Amasino RA, Coupland G (2006) The transcription factor *FLC* confers a flowering response to vernalization by repressing meristem competence and systemic signaling in *Arabidopsis*. Gene Dev 20:898–912

Sessions A, Yanofsky MF, Weigel D (2000) Cell-cell signaling and movement by the floral transcription factors *LEAFY* and *APETALA1*. Science 289:779–782

Sundstrom JF, Nakayama N, Glimelius K, Irish VF (2006) Direct regulation of the floral homeotic *APETALA1* gene by *APETALA3* and *PISTILLATA* in *Arabidopsis*. Plant J 46:593–600

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28(10):2731–2739

Team RC (2012) R: a language and environment for statistical computing

Thomas SG, Phillips AL, Hedden P (1999) Molecular cloning and functional expression of gibberellin 2-oxidases, multifunctional enzymes involved in gibberellin deactivation. Proc Natl Acad Sci USA 96:4698–4703

Towbin H, Staehelin T, Gordon J (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc Natl Acad Sci USA 76:4350–4354

Turck F, Fornara F, Coupland G (2008) Regulation and identity of florigen: *FLOWERING LOCUS T* moves center stage. Annu Rev Plant Biol 59:573–594

Wagner D, Sablowski RW, Meyerowitz EM (1999) Transcriptional activation of *APETALA1* by *LEAFY*. Science 285:582–584

Wang JW, Czech B, Weigel D (2009a) miR156-regulated *SPL* transcription factors define an endogenous flowering pathway in *Arabidopsis thaliana*. Cell 138:738–749

Wang Z, Gerstein M, Snyder M (2009b) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63

Wellmer F, Alves-Ferreira M, Dubois A, Riechmann JL, Meyerowitz EM (2006) Genome-wide analysis of gene expression during early *Arabidopsis* flower development. PLoS Genet 2:e117

Wigge PA, Kim MC, Jaeger KE, Busch W, Schmid M, Lohmann JU, Weigel D (2005) Integration of spatial and temporal information during floral induction in *Arabidopsis*. Science 309:1056–1059

Yamaguchi S (2008) Gibberellin Metabolism and its Regulation. Annu Rev Plant Biol 59:225–251

Yang H, Mo H, Fan D, Cao Y, Cui S, Ma L (2012) Overexpression of a histone H3K4 demethylase, JMJ15, accelerates flowering time in *Arabidopsis*. Plant Cell Rep 31(7):1297–1308

Yu H, Xu Y, Tan EL, Kumar PP (2002) *AGAMOUS-LIKE 24*, a dosage-dependent mediator of the flowering signals. Proc Natl Acad Sci 99(25):16336–16341

Zhang JZ, Ai XY, Sun LM, Zhang DL, Guo WW, Deng XX, Hu CG (2011) Transcriptome profile analysis of flowering molecular processes of early flowering trifoliate orange mutant and the wild-type [*Poncirus trifoliata* (L.) Raf.] by massively parallel signature sequencing. BMC Genom 12:63

Zhang X, Feng B, Zhang Q, Zhang D, Altman N, Ma H (2005) Genome-wide expression profiling and identification of gene activities during early flower development in *Arabidopsis*. Plant Mol Biol 58:401–419