frontiers in
PLANT SCIENCE

ORIGINAL RESEARCH ARTICLE
published: 20 January 2015
doi: 10.3389/fpls.2014.00802

Analysis of *Arabidopsis* floral transcriptome: detection of new florally expressed genes and expansion of Brassicaceae-specific gene families

Liangsheng Zhang*[1,2], *Lei Wang*[3,4], *Yulin Yang*[1], *Jie Cui*[3], *Fang Chang*[3], *Yingxiang Wang*[3]* and *Hong Ma*[3,4]

[1] *Department of Pharmacy, Shanghai Tenth People's Hospital, School of Life Sciences and Technology, Tongji University, Shanghai, China*
[2] *Advanced Institute of Translational Medicine, Tongji University, Shanghai, China*
[3] *State Key Laboratory of Genetic Engineering and Collaborative Innovation Center for Genetics and Development, Ministry of Education Key Laboratory of Biodiversity Science and Ecological Engineering and Institute of Biodiversity Sciences, Institute of Plants Biology, Center for Evolutionary Biology, School of Life Sciences, Fudan University, Shanghai, China*
[4] *Institutes of Biomedical Sciences, Fudan University, Shanghai, China*

Edited by:
Dazhong Dave Zhao, University of Wisconsin-Milwaukee, USA

Reviewed by:
Xuelin Wu, Harvey Mudd College, USA
Yuling Jiao, Chinese Academy of Sciences, China

****Correspondence:***
Yingxiang Wang and Hong Ma, School of Life Sciences, Fudan University, 2005 Songhu Road, Shanghai 200433, China
e-mail: yx_wang@fudan.edu.cn; hongma@fudan.edu.cn

The flower is essential for sexual reproduction of flowering plants and has been extensively studied. However, it is still not clear how many genes are expressed in the flower. Here, we performed RNA-seq analysis as a highly sensitive approach to investigate the *Arabidopsis* floral transcriptome at three developmental stages. We provide evidence that at least 23, 961 genes are active in the *Arabidopsis* flower, including 8512 genes that have not been reported as florally expressed previously. We compared gene expression at different stages and found that many genes encoding transcription factors are preferentially expressed in early flower development. Other genes with expression at distinct developmental stages included *DUF577* in meiotic cells and *DUF220*, *DUF1216*, and *Oleosin* in stage 12 flowers. *DUF1216* and *DUF577* are Brassicaceae specific, and together with other families experienced expansion within the Brassicaceae lineage, suggesting novel/greater roles in Brassicaceae floral development than other plants. The large dataset from this study can serve as a resource for expression analysis of genes involved in flower development in *Arabidopsis* and for comparison with other species. Together, this work provides clues regarding molecular networks underlying flower development.

Keywords: *Arabidopsis thaliana*, RNA-Seq, differentially expressed genes, floral development, gene families

INTRODUCTION

Flower is one of the most complex structures of the angiosperms (flowering plants), and is thought to make great contribution to sexual reproduction in either developmental or evolutionary aspects (Alvarez-Buylla et al., 2010). The basic floral architecture is highly conserved among the core eudicots, including *Arabidopsis thaliana*, which is an important model plant for studying flower development. Over the past three decades, extensive molecular genetic analyses have identified a large number of key floral regulators controlling flower development (O'Maoileidigh et al., 2014), making it one of the best-understood aspects of plant development. However, the present knowledge in understanding gene regulatory network in flower development is incomplete, such as information on genes with low expression levels.

Genome-wide approaches have become valuable tools in characterizing gene expression and in elucidating the genetic networks of flower development at a global level. In the past, large-scale analyses of transcript enrichment among *Arabidopsis* floral organs largely depends on hybridization, such as cDNA and oligonucleotide arrays (Hennig et al., 2004; Wellmer et al., 2004, 2006; Zhang et al., 2005; Alves-Ferreira et al., 2007; Benedito et al., 2008) and represents a major step in the spatial characterization of floral transcriptome, resulting in identifying many genes important for flower development (Alvarez-Buylla et al., 2010; Irish, 2010). However, array analyses and other hybridization-based approaches have several limitations, including knowledge of genes for probe design, non-specific hybridization, and difficulty in detecting low level expression (Marioni et al., 2008). On the other hand, more recently developed RNA sequencing (RNA-seq) technologies can overcome such limitations of hybridization-based approaches and other conventional large-scale gene expression analysis methods (Marioni et al., 2008; Xiong et al., 2010). It also has great sensitivity, allowing the detection of transcripts with lower expression levels, such as those of many transcription factors (Marioni et al., 2008; Chen et al., 2010). In the last few years, RNA-seq has been extensively applied in the characterization of transcriptome regarding developmental stage, organ, even specific cell types or single cell level, from yeast to human, including several plant species (Jiao et al., 2009; Zhang et al., 2010; Yang et al., 2011). To date, RNA-seq has been used for cell-specific analysis of actively translated mRNAs

associated with polyribosomes in developing flowers, providing insights and resources to further study flower development (Jiao and Meyerowitz, 2010).

To further explore the *Arabidopsis* flower transcriptome, we employed RNA-seq for three developmental periods. We detected 8512 additional genes that are not present on previously used microarray experiments, and provide evidence that at least 23,961 genes are truly expressed in the *Arabidopsis* flower. We also identified differentially and specifically expressed genes and gene families during flower development.

MATERIALS AND METHODS

SEQUENCING DATASETS

The inflorescent meristem (IM), stage 1–9 flowers (F1–9) and stage 12 flowers (F12) samples for RNA-seq were collected in our lab, and the three samples were subjected to 50 bp single-end sequencing on a SOLiD 3 platform; details for the methods were recently described in a study for alternative splicing (Wang et al., 2014a). All sequenced short reads were submitted to NCBI Short Read Archive under accession number SRP035230. The datasets for seeding and stage 4 flowers were from previously studies, which generated 36-bp and 42-bp long reads, respectively, using the Illumina genome analyzer (Filichkin et al., 2010; Jiao and Meyerowitz, 2010). The meiocyte datasets were from our previous study, which included two runs (36 and 50 bp) using Life Technologies' SOLiD sequencing platform (Yang et al., 2011).

ALIGNMENT OF SEQUENCING READS

Sequence reads from the three sample plus the three floral samples were mapped using PerM (Chen et al., 2009) to the *Arabidopsis* genome (release 9) from the *Arabidopsis* Information Resource (TAIR) database (TAIR9; www.arabidopsis.org) allowing 5, 4, and 3 mismatches per 50, 42, and 36-bp read, respectively.

DIGITAL GENE EXPRESSION AND EXPRESSION ARRAYS

For the RNA-seq experiments, we used at least 10 reads mapped to a gene as the threshold for being expressed. The raw digital gene expression counts were normalized using the reads per kilo-base of mRNA length per million of mapped reads (RPKM) method. The equation was used:

$$RPKM = \frac{10^9 * C}{N * L}$$

Where C is the uniquely mapped counts determined from mapping results, L is the length of the cDNA for the longest splice variant for a particular gene model and N is the total reads that were mapped to the genome. Log2-transformation of this normalized value was performed as in other analyses.

To test differential expression with mapping data DEGseq was used (Wang et al., 2010). Fisher's Exact Test ($P < 0.01$) method was selected. Microarray results were obtained from a previous study (Zhang et al., 2005). The Microarray experiments have a background value, which was 5 (log value of base 2) as previously described (Zhang et al., 2005) for the evaluation of "expressed" or "unexpressed" genes. Identification of differentially expressed genes according to the microarray data also used the Fisher's Exact Test method.

Z-SCORE

Calculation of the Z-score was based on the log2-transformed RPKM-normalized transcript levels as follows:

$$Z = (X - \mu)/\sigma$$

X is the RPKM of a gene for a specific tissue/developmental stage. μ is the mean RPKM of a gene across all tissues/developmental stages and σ is the RPKM standard deviation of a gene across all tissues/developmental stages. All calculations and plotting were performed by Perl and excel, respectively.

GENE FAMILY AND FUNCTIONAL ANNOTATION

The protein domain annotations were obtained from the Pfam database (http://pfam.sanger.ac.uk) (Punta et al., 2011). *Arabidopsis* protein sequences were then searched against protein family models in the Pfam-A database, resulting in 21102 *Arabidopsis* proteins identified as having at least one Pfam domain. Transcription factor family annotations were from The Database of *Arabidopsis* Transcription Factors (http://datf.cbi.pku.edu.cn/) (Guo et al., 2005), which contains 1922 transcription factors in *Arabidopsis*. Gene ontology (GO) enrichment analysis was performed with the agriGO browser (http://bioinfo.cau.edu.cn/agriGO/) (Du et al., 2010) using Singular Enrichment Analysis.

Multiple sequence alignment was performed in MUSCLE (http://www.drive5.com/muscle/) using the default parameters. Maximum likelihood (ML) trees were constructed by FastTree (www.microbesonline.org/fasttree) with the approximate likelihood ratio test method.

RESULTS AND DISCUSSION

GLOBAL GENE EXPRESSION OF FLOWER TRANSCRIPTOMES IN *ARABIDOPSIS*

To obtain more insights about the overall transcriptome landscape during flower development, we analyzed RNA-seq datasets of the *Arabidopsis* flower at three developmental stages recently generated in our laboratory; these datasets were analyzed for alternative splicing in a separate study (Wang et al., 2014a): inflorescent meristem (IM), stage 1–9 flowers (F1–9) and stage 12 flowers (F12), detecting 21,181 (IM), 22,137 (F1–9), and 22,827 (F12) reliably expressed genes (**Table S1**, **Figure S1**). A recent report summarized that a total of 126 *Arabidopsis* genes have been demonstrated genetically to have a role during flower development (Alvarez-Buylla et al., 2010), 122 of which were also detected as expressed in our dataset (**Table S2**), indicating that our data were very reliable, and can be used for further analysis. To compare gene expression during flower development, besides the three datasets described above, we also included data for *Arabidopsis* male meiocytes that we had generated previously (Yang et al., 2011), and two other public datasets of *Arabidopsis* seedlings and stage 4 flowers (Filichkin et al., 2010; Jiao and Meyerowitz, 2010), the latter of which were from isolated polysomic.

To further explore how many genes are truly expressed in *Arabidopsis* flower, we searched The *Arabidopsis* Information Resource (TAIR) database and obtained a total of 24,570 genes,

which are supported by at least one EST. Then, we searched the present tiling array database to find available probes for 30,228 genes. 4734 genes were found to be tiling array-specific compared with ESTs and RNA-seq data. Among them, 2634 and 276 are transposons and pseudogenes, respectively. The other 1824 genes seem to be expressed at very low levels. The average value of the 4734 genes is 5.4, which is regarded as a threshold in this study for the evaluation of "expressed" or "unexpressed" genes. Based on this criterion, we believe that tiling array can detect at least 27, 617 genes. As described previously, RNA-seq detected 24,769 genes in flowers. Comparison of the detected genes among EST, tiling array and RNA-seq found that 22,440 genes were detected by three data sets and 1521 genes were detected by RNA-seq and either ESTs/tiling array (**Figure 1A**). The results suggest that at least 23,961 genes are reliably detected as expressed in the *Arabidopsis* flower. In addition, 621 genes were only detected by RNA-seq; most of these are low abundance genes that are nearly undetectable by arrays and the others are likely to be stage-specific genes.

Characterization of stage or cell-specific genes provides a foundation for unraveling their molecular mechanisms. Previous studies in multiple plants demonstrated that each stage or tissue

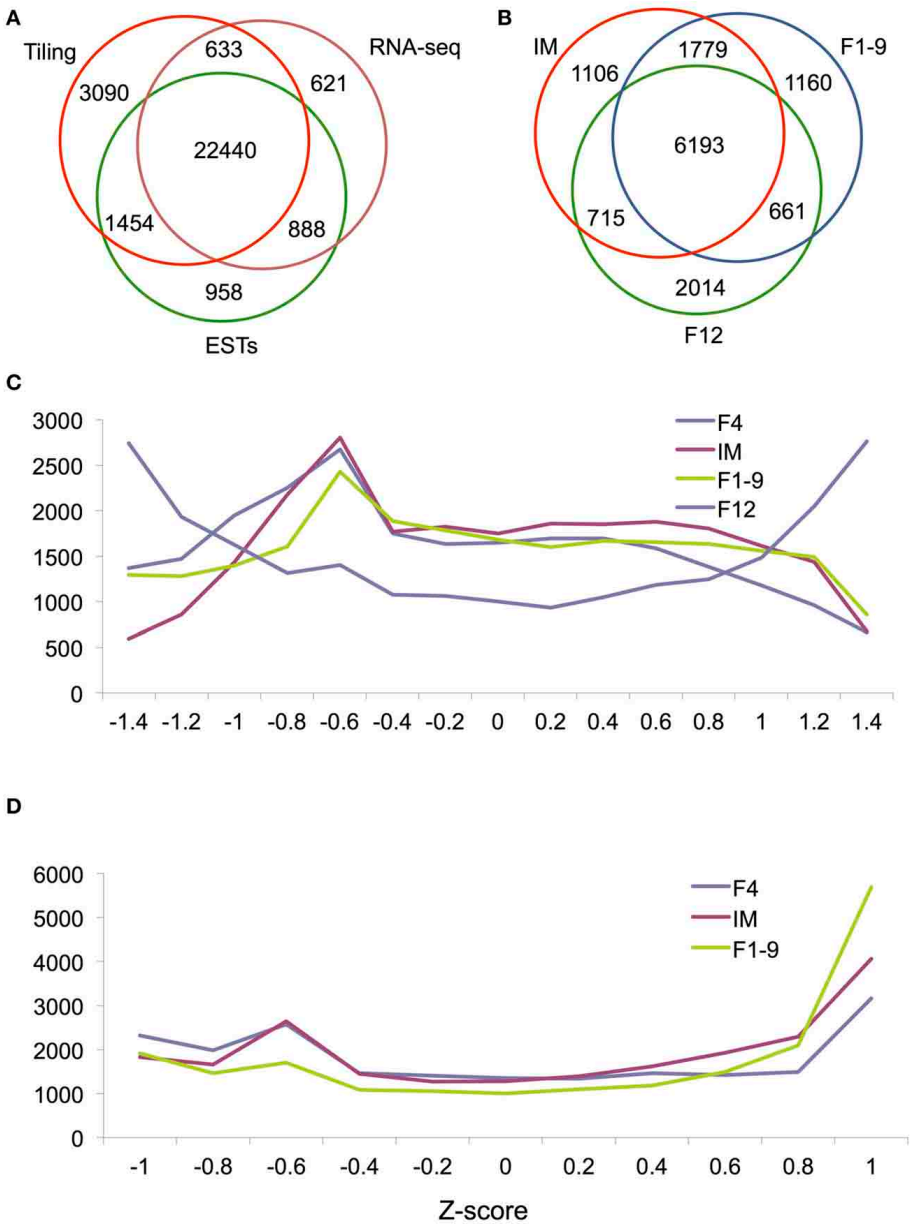

FIGURE 1 | Global gene expression during flower development. (A) A Venn diagram showing the overlap in detected genes between three technologies: the tilling array, RNA-seq and ESTs. **(B)** A Venn diagram showing the overlap between IM, F1–9, and F12 for genes called as DEG by RNA-seq. Histograms of relative expression levels (measured by Z-scores) in four **(C)** and three **(D)** organs. For easy visualization, we plotted Z-score on the x-axis and gene numbers on the y-axis.

has specific transcripts (Jiao et al., 2009; Jiao and Meyerowitz, 2010; Yang et al., 2011; Liu et al., 2014). To better establish the genome-wide gene expression pattern of flower development, we conducted a Z-score analysis to assess the extent of differential gene expression for florally expressed genes. Results showed that the Z-score distribution of gene expression in F12 was dramatically different from that for early flower development (F1–9, F4, and IM) (**Figure 1C**), suggesting that nearly mature flowers requires many more specifically or differentially expressed genes than early flowers. In contrast, Z-score distributions were very similar between F1–9, F4, and IM (**Figure 1D**), further supporting the idea that the developmental programs of these stage/organ are similar.

DETECTION OF EXPRESSION OF 8512 GENES IN THE *ARABIDOPSIS* FLOWER NOT REPORTED FROM MICROARRAY ANALYSIS

We first compared the F1–9 and F12 RNA-seq data with the Affymetrix ATH1 array data at similar stages (Zhang et al., 2005). We compared the number of sequencing reads mapped to each gene with the corresponding (normalized) absolute intensities from the array (**Figures 2A,B**), and found that the correlations between the two platforms were high, with Spearman correlation coefficients of 0.82 (F1–9; **Figure 2A**) and 0.80 (F12; **Figure 2B**). Thus, comparison of RNA-seq and microarray identified 15,180 overlapping genes with relatively high expression levels (**Figure 2C**), covering 96% of genes detected using microarray. In addition, our RNA-seq identified additional 8512 genes that were undetected by microarray (Zhang et al., 2005), whose expression levels were obviously lower; the average expression level is 56.12 (F1–9) and 57.82 (F12) RPKM (**Figure S2b**), compared to the average expression level about 250 for the 15,180 genes. Also, the curvature of the comparison toward the microarray axis suggested that microarray possibly underestimated the expression level of genes relative to RNA-seq (**Figures 2A,B**). Together, these results suggest that RNA-seq has a great advantage over microarray in detecting low-abundance transcripts, consistent with previous reports (Marioni et al., 2008; Yang et al., 2011).

To investigate further regarding the 8512 genes, we analyzed the enrichment of protein family (PFAM) domains (gene families) among these genes, and identified several enriched gene families that were not reported previously as enriched, including *F-box*, *NB-ARC*, *C1_3*, *PPR*, *LRR_1*, *Myb*, *bHLH,* and *AP2* gene families (**Table 1**). Previously, many *F-box* genes were reported as unexpressed or undetectable by microarray analysis (Schmid

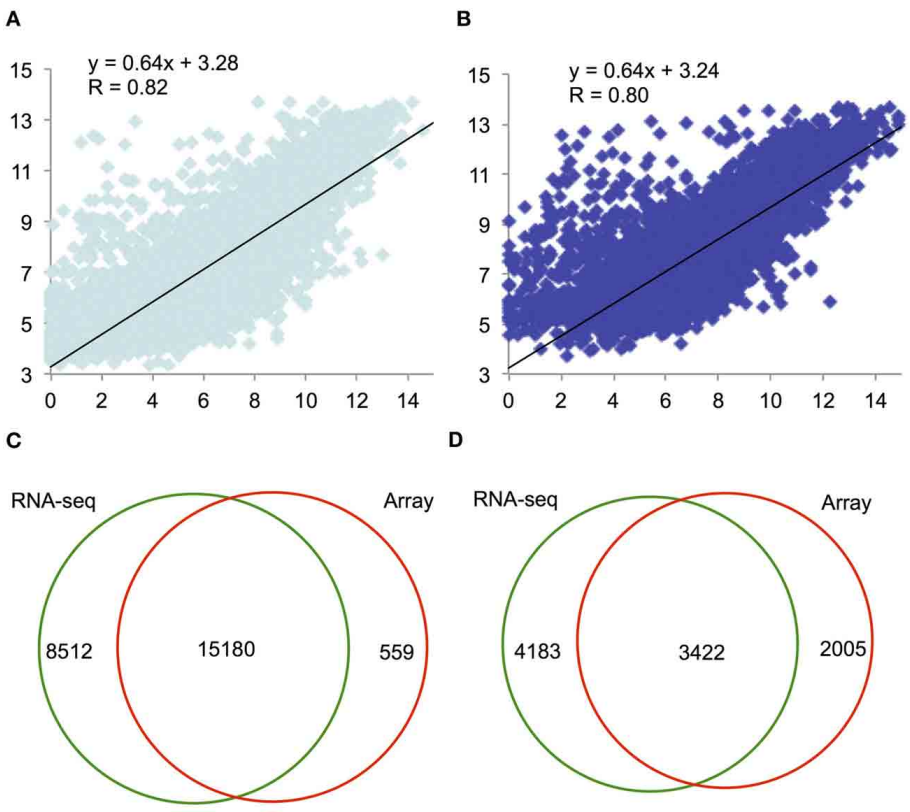

FIGURE 2 | A scatter plot of relative expression values obtained by RNA-seq and microarray for F1–9 and F12. (A) Comparison of expression levels between RNA-seq and microarray for F1–9, RNA-seq and microarray; gene expression levels transformed with log2 were plotted. **(B)** Comparison of expression between RNA-seq and microarray for F12. **(C)** A Venn diagram illustrating genes detected by RNA-seq (left) and microarray (right) analyses. **(D)** A Venn diagram presenting the overlap of differentially expressed genes between F1–9 and F12 from RNA-seq (left) and from microarray analysis (right).

Table 1 | Enriched protein family (Pfam) among 8512 genes that are undetected by microarray analysis.

Family	Total	RNA-seq	Percent	Family	Total	RNA-seq	Percent
NB-ARC	167	108	0.65	bHLH	134	59	0.44
FBD	115	71	0.62	AP2	145	62	0.43
TIR	132	80	0.61	UDPGT	115	49	0.43
SCRL	25	14	0.56	peroxidase	82	33	0.40
U-box	61	31	0.51	Myb	256	103	0.40
MATH	60	30	0.50	Malectin_like	78	31	0.40
DUF26	97	47	0.48	LRR_1	391	154	0.39
Auxin	79	38	0.48	PMEI	122	46	0.38
C1_3	146	70	0.48	SLR1-BP	41	15	0.37
DUF295	78	36	0.46	p450	249	91	0.37
PPR_1	301	138	0.46	zf-rbx1	174	63	0.36
PPR	465	211	0.45	ABC_tran	117	42	0.36
NAC	113	51	0.45	Kelch_1	110	38	0.35
F-box	522	233	0.45	Ank_2	106	36	0.34
FBA_1	176	78	0.44	zf-C3HC4	306	96	0.31

et al., 2005), further suggesting that microarray is not as sensitive as RNA-seq for detecting low-abundance transcripts. In addition, we also detected some enriched gene families that belongs to the highly expressed genes; for instance, Plant self-incompatibility response (SCRL) and S locus-related glycoprotein 1 binding pollen coat protein (SLR1-BP) are specifically enriched in F12 (**Table 1** and **Table S3**), suggesting a potential role at this stage. In contrast, 559 genes detected by microarray were not found by RNA-seq, possibly due to difference in growth conditions.

We further employed a widely used Fisher's Exact Test method to identify differentially expressed genes (DEGs) between F1–9 and F12 in RNA-seq and microarray data. Altogether, 7605 and 5327 DEGs were identified in each dataset. Among them, 3422 genes were detected by both platforms (**Figure 2D**), 1272 and 84 DEGs were only detected by RNA-seq and microarray, respectively (**Figure 2D**), and consistent with the fact that RNA-seq is more sensitive for detection and comparison of gene expression. Taken together, these results indicate that deep sequencing can greatly increase the sensitivity of transcriptome analysis.

IDENTIFICATION OF STAGE-DIFFERENTIALLY EXPRESSED GENES DURING FLOWER DEVELOPMENT

Floral organ identity and cell fate determination are highly regulated by the temporal and spatial gene expression, with each organ or cell type having distinct transcriptomes (Jiao et al., 2009; Yang et al., 2011; Wang et al., 2014b). To investigate DEGs between one of the floral stages with seedlings, we compared the flower transcriptomes of IM, F1–9, or F12 with that of seedlings, and identified IM with 9793 DEGs, F1–9 with 9583 DEGs, and F12 with 9340 DEGs (**Table 2**). Furthermore, the intersection between these three sets contained 6193 genes (**Figure 1B**), indicating these three samples are quite similar regarding differentially gene expression compared with seedlings. GO annotation showed that these genes were enriched for categories such as "histone modification" and "methylation" ($p = 3.2E\text{-}11$ and 4.7E-9), suggesting

Table 2 | Differentially expressed genes (DEGs) for each floral sample compared with seedling.

	Seedling	IM	F4	F1–9	F12	Meiocytes
Seeding		9793	8703	9583	9340	4966
IM	6627		3747	6866	6943	3900
F4	5812	5354		6016	6460	3449
F1–9	6866	7228	4834		7697	3773
F12	5987	8960	6724	8109		3820
Meiocytes	5110	7167	6154	6628	6401	

these genes are involved in the establishment of transcription regulation during flower development. Likewise, 2014 genes were specifically expressed in F12 and showed significant enrichment for genes in reproduction ($p = 2.2E\text{-}47$), flower development ($p = 2.9E\text{-}22$) and post-embryonic development ($p = 1.3E\text{-}194$), which might suggest that genes expressed during gametophyte development can function in later stages.

To further examine the combined set (13,628 genes) of the above floral DEGs, we compared these genes with 4505 genes identified as potential targets of the SEP3 and/or AP1 proteins by ChIP-seq (Immink et al., 2009; Kaufmann et al., 2010). The results showed that 2506 genes overlapped between the floral DEGs and the SEP3/AP1 targets with significance (Fisher's test, $p = 7.03e\text{-}08$). It is likely that some of these genes are involved in the regulation of flower development, but the role of these genes in flower development needs to be determined using molecular genetic analyses.

We then analyzed the enrichment of protein domains as defined in the PFAM database, and found several enriched domains ($P \leq 0.01$; **Table 3**), including ATPase, Helicase_C, DEAD box, WD40, SET, and PHD domains, suggesting that chromatin associated transcriptional regulation might be one of the major features underlying flower development. In addition, proteins with "UCH," "hydrolase," and "IQ" domains were also significantly over-represented, although their functions in flower development are largely unknown. Interestingly, we also identified "PPR," "Mito_carr" and "Miro" domains as significantly enriched; members of these genes are involved in gene expression and other functions in mitochondria and plastid, suggesting that such organellar functions might be important for flower development.

DISTINCT ENRICHMENT OF TRANSCRIPTION FACTORS IN EARLY FLOWER DEVELOPMENT

Identification of transcription factors (TFs) expressed in a specificstage provides a foundation for understanding the transcriptional regulatory networks underlying the development, structure and function of thestage. To investigatethe expressed TFs during flower development, we examined the TFs among IM, F1–9, and F12 and identified a total of 1667 transcription factors, 927 of which showed differential expression compared with the seedling (**Figure 3A**). Among the 927 TFs, 70% showed highest expression in IM (designated as D1), whereas 14 and 16% showed highest expression in F1–9 (designated as D2) and F12 (designated as D3), respectively (**Figure 3B**).

Table 3 | Significance of enriched Pfam domains in differentially expressed genes during flower development.

Pfam domain	Total	Num.	Percent	*P*-vaule	Pfam domain	Total	Num.	Percent	*P*-vaule
Helicase_C	149	133	0.89	4.00E-07	Proteasome	24	23	0.96	0.02
WD40	234	184	0.79	1.00E-06	Cyclin_C	30	28	0.93	0.02
DEAD	114	97	0.85	5.00E-05	HATPase_c	35	30	0.86	0.02
RRM_1	245	171	0.70	0.0003	AAA_5	45	37	0.82	0.02
Kinesin	61	55	0.90	0.001	Mito_carr	59	46	0.78	0.02
PHD	52	47	0.90	0.002	Galactosyl_T	21	20	0.95	0.03
SNF2_N	45	40	0.89	0.005	HA2	21	20	0.95	0.03
PPR_1	301	188	0.62	0.006	Cyclin_N	52	40	0.77	0.03
IQ	56	46	0.82	0.008	Hydrolase	60	45	0.75	0.03
PPR	465	275	0.59	0.008	Miro	114	76	0.67	0.03
AAA	144	98	0.68	0.009	OB_NTP_bind	20	19	0.95	0.04
LSM	26	26	1.00	0.01	Cpn60_TCP1	23	21	0.91	0.04
UCH	45	39	0.87	0.01	KH_1	25	22	0.88	0.04
ResIII	48	40	0.83	0.01	SET	46	35	0.76	0.04
RuvB_N	46	38	0.83	0.01	Histone	67	48	0.72	0.04

D1 mainly contained members of the homeobox domain (HB), MADS, MYB, AP2, and NAC families, suggesting that floral meristem development largely requires those transcription factor (**Figure 3B**). For example, *homeobox* genes encode transcription factors that contain a classic DNA binding domain with about 60 amino acids and regulate gene expression via Polycomb-dependent modulation of chromatin structure, thereby controlling development in animals, fungi and plants (Zhong and Holland, 2011). Several known members of *HB* (**Figure 3C**) identified in IM support that early floral development requires active *HB* genes, consistent with the finding that epigenetic reprogramming of gene expression is important for the establishment of initial floral identity (Mukherjee et al., 2009). The co-expressed pattern between *HB* genes and chromatin factors in IM is in agreement with previous studies that a number of floral genes with similar expression patterns and/or associated with each other regulate the expression of downstream genes to ensure proper flower development (Kaufmann et al., 2009, 2010; Deng et al., 2011).

D2 included *MADS-box*, *MYB*, *AS2*, *C2H2*, *bZIP*, *ABI3*, and *bHLH* families (**Figure 3C**). *MADS-box* genes encode not only key repressors or activators for flowering transition, but also master regulators of reproductive organ identities (Alvarez-Buylla et al., 2010). Our data detected expression of most MADS-box genes known to be involved in flower development (**Figure 3C**), such as *FLC, SHP1/2, AP3, AG, AGL11/15/77, TT16, SEP1-3, STK, AT5G49420*; *AG*, *AP3*, and *SEP1-3* are genes for the ABCE model, consistent with their known function in floral organ identity (Smaczniak et al., 2012). In addition, genes coding for transcription factors important for microsporogenesis were also uncovered, such as *AMS, MS1, MYB35*, and *MYB99* (Chang et al., 2011), as well as *MMD1* required for meiosis (Yang et al., 2003).

D3 was enriched in *MADS*, *MYB*, *AP2*, *C2H2*, *C2C2-CO-like*, *NAC*, *AUX-IAA-ARF*, and *bHLH* families (**Figure 3C**). Previous studies showed that auxin-dependent transcriptional regulation requires the auxin/indole-3-acetic acid (Aux/IAA) and auxin response factor (ARF) families of TFs and formation of Aux/IAA-ARFs heterodimers represses auxin signaling (Reed, 2001), which has been demonstrated to participate in pollen development, pollination and fertilization (Sundberg and Ostergaard, 2009), as well as female gametophyte specification (Pagnussat et al., 2009). Indeed, our data identified several known and unknown ARFs and IAAs factors in G3, suggesting that the Aux/IAA-ARF regulatory pathway is vital for late reproductive development. However, the function of other enriched TFs in flowers is still largely unknown. Together, these results demonstrate that flower development at different stages requires common and distinct transcription factor families.

IDENTIFICATION OF SPECIFIC GENE FAMILIES AT DISTINCT STAGES OF FLOWER DEVELOPMENT

We sought to identify stage-specific genes, which were defined as those genes that were differentially expressed (>4-fold change) at one stage over all other stages studied here using DEG seq. The largest numbers of stage specific genes were identified in the seedling, F12 and meiocytes (1083, 552, and 652 genes, respectively; **Table 4**). Given the lack of correlation in overall gene expression between the floral transcriptome (F12) and the other stages sampled (**Figure 2C**), it was not surprising to identify this stage as having the largest number of organ-specific genes. These genes are strong candidates for determining the specific functional components of the nearly mature flower.

Interestingly, the F1–9 flower-specific genes with 8-fold changes had 26 genes, including 9 transposons and 5 snoRNAs (**Table S4**), consistent with the previous finding that transposons and small RNAs were enriched among genes expressed in male meiocytes (Chen et al., 2010; Yang et al., 2011). There are also 12 coding genes, one of which (*AT5G09780*) codes for a transcription factor of the B3 family and two (*AT1G48700* and *AT4G03050*) are for iron binding proteins.

For meiocyte-specific genes, 424 genes were found with ∼8-fold changes and showed enrichment for genes in an insertion of mitochondrial origin on chromosome II, as supported by similar preferential expression in meiocytes reported previously

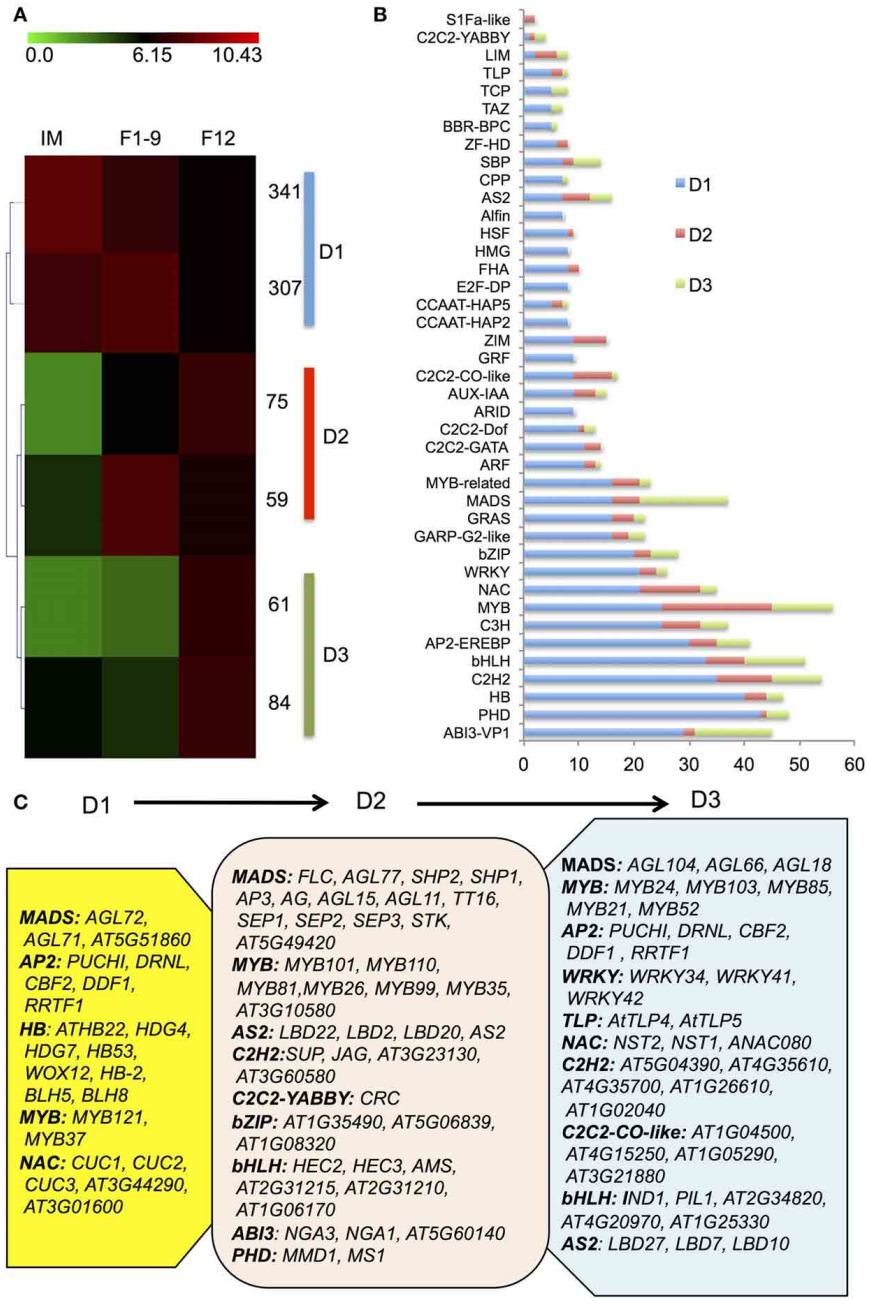

FIGURE 3 | The transcription factor accumulation profiles. (A) A heatmap for the transcription factors. 937 significantly differentially expressed transcription factors from IM, F1–9, and F12 were clustered into three groups (D1–D3) using the Self Organization Tree Algorithm. **(B)** Distribution of transcription factor families among D1–D3. **(C)** Representative functions and genes showing different transcription factor families for developmental stage-dependent groups D1 (IM), D2 (F1–9), and D3 (F12).

Table 4 | The specifically expressed genes in one sample compared with others.

Fold change	Seedling	IM	F4	F1–9	F12	Meiocyte
1	2656	1280	686	1740	1632	1636
2	1871	157	118	110	817	1200
4	1083	27	33	26	552	652
8	695	7	16	13	418	424

(Chen et al., 2010). The enriched genes also included 45 mitochondrial and 28 chloroplast genes, respectively. Moreover, in addition to previously reported gene families (Yang et al., 2011), we also detected several other enriched gene families, such as *Oxidored_q1*, *Oxidored_q2*, *Oxidored_q3*, *Oxidored_q*4, and *NADHdh*. In particular, most members of the *DUF577* family showed specific expression in meiocytes (**Table S5**). To further investigate this gene family, we performed phylogenetic analyses

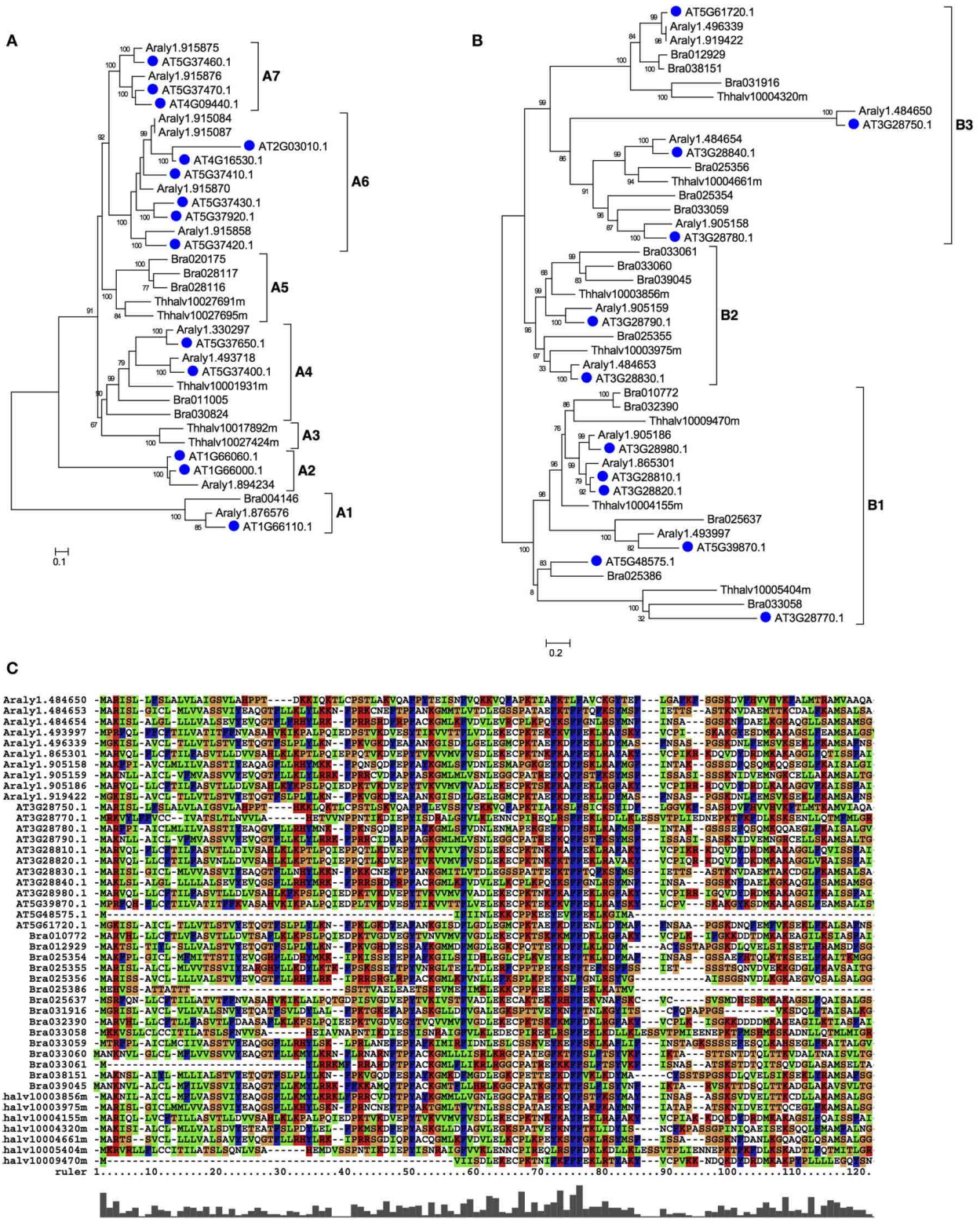

FIGURE 4 | The phylogenetic tree of *DUF577* and DUF1216 gene families in plant. The *DUF577* and *DUF1216* gene families were found only in plants, but not in other eukaryotic groups. **(A)** A Maximum likelihood (ML) tree of the *DUF577* gene family using representative species in eudicots and it can be divided into 8 clades, A1∼A7. **(B)** An ML tree of *DUF1216* gene family using representative species in eudicots and it can be divided into 3 clades, B1∼B3. **(C)** An alignment of the N-terminal regions of DUF1216 proteins. Species names are abbreviated as below: AT-*Arabidopsis thaliana*, Araly-*Arabidopsis lyrata*, Bra-*Brassica rapa*, Thhalv-*Eutrema salsugineuma*.

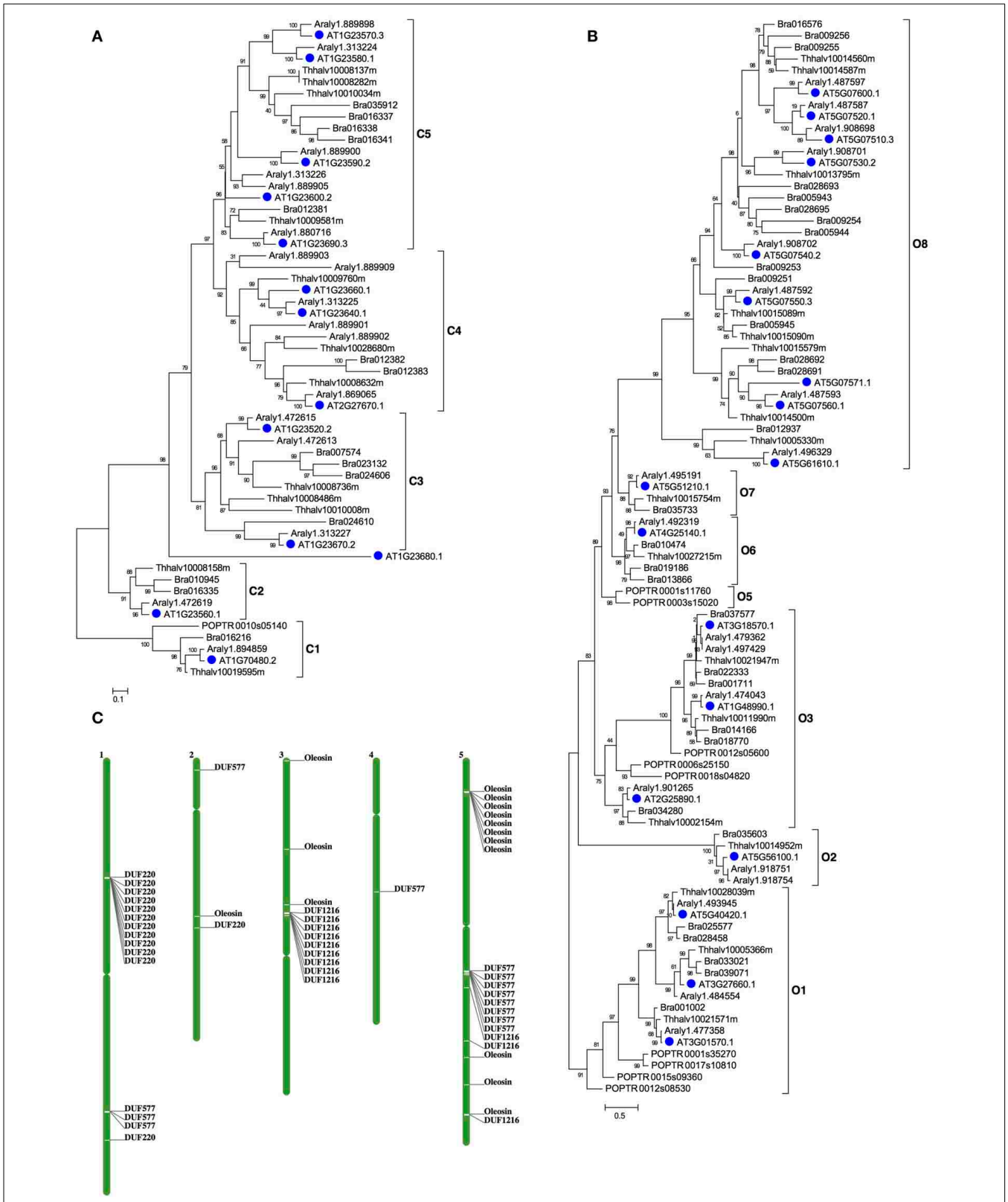

FIGURE 5 | The phylogenetic tree of *DUF220* and *Oleosin* gene families in plant. (A) An ML tree of *DUF220* gene family can be divided into 5 clades, C1∼C5. **(B)** An ML tree of *Oleosin* gene family can be divided into 8 clades, O1∼O8. Species names are abbreviated as below: AT-*Arabidopsis thaliana*, Araly-*Arabidopsis lyrata*, Bra-*Brassica rapa*, Thhalv-*Eutrema salsugineum*; POPTR-*Populus trichocarpa*. **(C)** The chromosomal positions of the *Arabidopsis DUF577*, *DUF1216*, *DUF220*, and *Oleosin* genes. The names of genes refer to locus ID as listed in **Table S5**.

of this gene family with members from several representative plant species, including *Arabidopsis lyrata*, *Eutrema salsugineum*, *Brassica rapa*. As shown in **Figure 4A**, this family can be divided into seven subfamilies, designated as A1-A7. The tree supported that this gene family have experienced expansion and origin in Brassicaceae (**Figure 4A**). The A6 and A7 subfamilies only included *Arabidopsis lyrata* and *Arabidopsis thaliana*, suggesting an expansion that occurred since the divergence of *Arabidopsis* and other Brassicaceae species. Besides, functions of the *DUF577* and other enriched family genes in meiosis need to be tested.

Similarly for F12, the enriched families included *DUF1216*, *Oleosin* and *DUF220*. Most members of the three gene families had specific expression in F12; these gene families contain lineages that originated and expanded within Brassicaceae (**Table S5** and **Figures 4**, **5**). Phylogenetic analyses of the *DUF1216* family suggested that this gene family is specific to Brassicaceae, without homologs in other plants, and experienced gene duplication during Brassicaceae history (**Figure 4B**). Interestingly, the N-terminal region of DUF1216 proteins had putative signal peptides with similar sequences, according to the SignalP prediction (http://www.cbs.dtu.dk/services/SignalP/). The predicted signal peptide contains a large number of hydrophobic amino acids, a conserved basic amino acid and a conserved cysteine at the ninth position (**Figure 4C**). *At5g07750* of the *Oleosin* family was reported to have experienced positive selection (Schein et al., 2004). However, expression of each of three tandem duplicated genes in the Oleosin family (*AT5G07510*: 10081.76, *AT5G07550*: 29536.45, *AT5G07560*: 10942.38) had extraordinarily high levels of more than RPKM of 10, 000, suggesting that such high expression levels are important for F12 for later functions. Analysis of the *Arabidopsis* genome indicates that tandem duplication contributed to the expansion of *DUF1216* in Brassicaceae, as well as the expansion of the *Oleosin*, *DUF220* and *DUF577* families (**Figure 5C**). This pattern is different to those of *SET*, *JmjC,* and *Rhomboid* gene families, which are more likely to be retained after whole genome duplication events (Zhou and Ma, 2008; Zhang and Ma, 2012; Li et al., 2014).

CONCLUSIONS

The analysis of *Arabidopsis* floral transcriptome datasets presented here provides a valuable resource of candidate genes for further studies to understand the flower development program. We provided evidence for at least 23,961 genes that are expressed in the *Arabidopsis* flower. Compared with seedling, over 10,000 DEGs were identified, revealing novel and different molecular characteristics in the developing flower such as regulatory genes, genes for high-energy production, and transposable elements. These results showed that flower development at different stages requires common and distinct transcription factor families. The gene expression in F12 was dramatically different from that for early flower development (F1–9, F4, and IM).

In addition to identifying floral developmental gene candidates, we found many genes or gene families specifically expressed at one stage. Many transposable element genes, at least 45 mitochondrial and 28 chloroplast genes showed specific expression in meiocytes. The *SCRL*, *SLR1-BP*, *DUF1216*, *Oleosin,* and *DUF220* gene families showed specific expression in F12 and *DUF577* genes were detected to have specific expression meiosis. These specifically expressed genes have functions that are closely related to reproductive development, showed that mature flowers require many more specifically or differentially expressed genes than early flowers. These gene families expanded dramatically within the Brassicaceae lineage, suggesting novel functions that are possibly important for the origin and evolution of Brassicaceae. This dataset can be useful for discovering functional genes at different stages of the flower development and provide clues for the molecular and regulatory relationships between different stages.

ACKNOWLEDGMENTS

We would like to thank Qi Li for comments on the manuscript and helpful discussions. This work was supported by the National Natural Science Foundation of China (91131007) and Chinese Ministry of Science and Technology (2011CB944600). Liangsheng Zhang was supported by funds from Tongji University (2013KJ052).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fpls.2014.00802/abstract

Figure S1 | Biological replicates were highly reproducible.

Figure S2 | Comparison between transcriptomes from RNA-Seq and microarray.

Table S1 | The expression of all genes in six samples.

Table S2 | The 126 known *Arabidopsis* flower development genes.

Table S3 | The expression of *SCRL* and *SLR1-BP* genes in different samples.

Table S4 | The expression of the 26 specific genes in F1–9.

Table S5 | The expression of *DUF577, DUF1216, DUF220* and *Oleosin* genes in different samples.

REFERENCES

Alvarez-Buylla, E. R., Benitez, M., Corvera-Poire, A., Chaos Cador, A., De Folter, S., Gamboa De Buen, A., et al. (2010). Flower development. *Arabidopsis Book* 8:e0127. doi: 10.1199/tab.0127

Alves-Ferreira, M., Wellmer, F., Banhara, A., Kumar, V., Riechmann, J. L., and Meyerowitz, E. M. (2007). Global expression profiling applied to the analysis of *Arabidopsis* stamen development. *Plant Physiol.* 145, 747–762. doi: 10.1104/pp.107.104422

Benedito, V. A., Torres-Jerez, I., Murray, J. D., Andriankaja, A., Allen, S., Kakar, K., et al. (2008). A gene expression atlas of the model legume Medicago truncatula. *Plant J.* 55, 504–513. doi: 10.1111/j.1365-313X.2008.03519.x

Chang, F., Wang, Y., Wang, S., and Ma, H. (2011). Molecular control of microsporogenesis in *Arabidopsis*. *Curr. Opin. Plant Biol.* 14, 66–73. doi: 10.1016/j.pbi.2010.11.001

Chen, C., Farmer, A. D., Langley, R. J., Mudge, J., Crow, J. A., May, G. D., et al. (2010). Meiosis-specific gene discovery in plants: RNA-Seq applied to isolated *Arabidopsis* male meiocytes. *BMC Plant Biol.* 10:280. doi: 10.1186/1471-2229-10-280

Chen, Y., Souaiaia, T., and Chen, T. (2009). PerM: efficient mapping of short sequencing reads with periodic full sensitive spaced seeds. *Bioinformatics* 25, 2514–2521. doi: 10.1093/bioinformatics/btp486

Deng, W., Ying, H., Helliwell, C. A., Taylor, J. M., Peacock, W. J., and Dennis, E. S. (2011). FLOWERING LOCUS C (FLC) regulates development pathways

throughout the life cycle of *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 108, 6680–6685. doi: 10.1073/pnas.1103175108

Du, Z., Zhou, X., Ling, Y., Zhang, Z., and Su, Z. (2010). agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* 38, W64–W70. doi: 10.1093/nar/gkq310

Filichkin, S. A., Priest, H. D., Givan, S. A., Shen, R., Bryant, D. W., Fox, S. E., et al. (2010). Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome Res.* 20, 45–58. doi: 10.1101/gr.093302.109

Guo, A., He, K., Liu, D., Bai, S., Gu, X., Wei, L., et al. (2005). DATF: a database of *Arabidopsis* transcription factors. *Bioinformatics* 21, 2568–2569. doi: 10.1093/bioinformatics/bti334

Hennig, L., Gruissem, W., Grossniklaus, U., and Kohler, C. (2004). Transcriptional programs of early reproductive stages in *Arabidopsis*. *Plant Physiol.* 135, 1765–1775. doi: 10.1104/pp.104.043182

Immink, R. G., Tonaco, I. A., De Folter, S., Shchennikova, A., Van Dijk, A. D., Busscher-Lange, J., et al. (2009). SEPALLATA3: the "glue" for MADS box transcription factor complex formation. *Genome Biol.* 10:R24. doi: 10.1186/gb-2009-10-2-r24

Irish, V. F. (2010). The flowering of *Arabidopsis* flower development. *Plant J.* 61, 1014–1028. doi: 10.1111/j.1365-313X.2009.04065.x

Jiao, Y., and Meyerowitz, E. M. (2010). Cell-type specific analysis of translating RNAs in developing flowers reveals new levels of control. *Mol. Syst. Biol.* 6, 419. doi: 10.1038/msb.2010.76

Jiao, Y., Tausta, S. L., Gandotra, N., Sun, N., Liu, T., Clay, N. K., et al. (2009). A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies. *Nat. Genet.* 41, 258–263. doi: 10.1038/ng.282

Kaufmann, K., Muino, J. M., Jauregui, R., Airoldi, C. A., Smaczniak, C., Krajewski, P., et al. (2009). Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the *Arabidopsis* flower. *PLoS Biol.* 7:e1000090. doi: 10.1371/journal.pbio.1000090

Kaufmann, K., Wellmer, F., Muino, J. M., Ferrier, T., Wuest, S. E., Kumar, V., et al. (2010). Orchestration of floral initiation by APETALA1. *Science* 328, 85–89. doi: 10.1126/science.1185244

Li, Q., Zhang, N., Zhang, L., and Ma, H. (2014). Differential evolution of members of the rhomboid gene family with conservative and divergent patterns. *New Phytol.* doi: 10.1111/nph.13174. [Epub ahead of print].

Liu, D., Sui, S., Ma, J., Li, Z., Guo, Y., Luo, D., et al. (2014). Transcriptomic analysis of flower development in wintersweet (*Chimonanthus praecox*). *PLoS ONE* 9:e86976. doi: 10.1371/journal.pone.0086976

Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M., and Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* 18, 1509–1517. doi: 10.1101/gr.079558.108

Mukherjee, K., Brocchieri, L., and Burglin, T. R. (2009). A comprehensive classification and evolutionary analysis of plant homeobox genes. *Mol. Biol. Evol.* 26, 2775–2794. doi: 10.1093/molbev/msp201

O'Maoileidigh, D. S., Graciet, E., and Wellmer, F. (2014). Gene networks controlling *Arabidopsis* thaliana flower development. *New Phytol.* 201, 16–30. doi: 10.1111/nph.12444

Pagnussat, G. C., Alandete-Saez, M., Bowman, J. L., and Sundaresan, V. (2009). Auxin-Dependent Patterning and Gamete Specification in the *Arabidopsis* Female Gametophyte. *Science* 324, 1684–1689. doi: 10.1126/science.1167324

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., et al. (2011). The Pfam protein families database. *Nucleic Acids Res.* 40, D290–D301. doi: 10.1093/nar/gkr1065

Reed, J. W. (2001). Roles and activities of Aux/IAA proteins in *Arabidopsis*. *Trends Plant Sci.* 6, 420–425. doi: 10.1016/S1360-1385(01)02042-8

Schein, M., Yang, Z., Mitchell-Olds, T., and Schmid, K. J. (2004). Rapid evolution of a pollen-specific oleosin-like gene family from *Arabidopsis thaliana* and closely related species. *Mol. Biol. Evol.* 21, 659–669. doi: 10.1093/molbev/msh059

Schmid, M., Davison, T. S., Henz, S. R., Pape, U. J., Demar, M., Vingron, M., et al. (2005). A gene expression map of *Arabidopsis* thaliana development. *Nat. Genet.* 37, 501–506. doi: 10.1038/ng1543

Smaczniak, C., Immink, R. G., Muino, J. M., Blanvillain, R., Busscher, M., Busscher-Lange, J., et al. (2012). Characterization of MADS-domain transcription factor complexes in *Arabidopsis* flower development. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1560–1565. doi: 10.1073/pnas.1112871109

Sundberg, E., and Ostergaard, L. (2009). Distinct and dynamic auxin activities during reproductive development. *Cold Spring Harb. Perspect. Biol.* 1:a001628. doi: 10.1101/cshperspect.a001628

Wang, H., You, C., Chang, F., Wang, Y., Wang, L., Qi, J., et al. (2014a). Alternative splicing during *Arabidopsis* flower development results in constitutive and stage-regulated isoforms. *Front. Genet.* 5:25. doi: 10.3389/fgene.2014.00025

Wang, L., Cao, C., Ma, Q., Zeng, Q., Wang, H., Cheng, Z., et al. (2014b). RNA-seq analyses of multiple meristems of soybean: novel and alternative transcripts, evolutionary and functional implications. *BMC Plant Biol.* 14:169. doi: 10.1186/1471-2229-14-169

Wang, L., Feng, Z., Wang, X., and Zhang, X. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26, 136–138. doi: 10.1093/bioinformatics/btp612

Wellmer, F., Alves-Ferreira, M., Dubois, A., Riechmann, J. L., and Meyerowitz, E. M. (2006). Genome-wide analysis of gene expression during early *Arabidopsis* flower development. *PLoS Genet.* 2:e117. doi: 10.1371/journal.pgen.0020117

Wellmer, F., Riechmann, J. L., Alves-Ferreira, M., and Meyerowitz, E. M. (2004). Genome-wide analysis of spatial gene expression in *Arabidopsis* flowers. *Plant Cell* 16, 1314–1326. doi: 10.1105/tpc.021741

Xiong, Y., Chen, X., Chen, Z., Wang, X., Shi, S., Zhang, J., et al. (2010). RNA sequencing shows no dosage compensation of the active X-chromosome. *Nat. Genet.* 42, 1043–1047. doi: 10.1038/ng.711

Yang, H., Lu, P., Wang, Y., and Ma, H. (2011). The transcriptome landscape of *Arabidopsis* male meiocytes from high-throughput sequencing: the complexity and evolution of the meiotic process. *Plant J.* 65, 503–516. doi: 10.1111/j.1365-313X.2010.04439.x

Yang, X. H., Makaroff, C. A., and Ma, H. (2003). The *Arabidopsis MALE MEIOCYTE DEATH1* gene encodes a PHD-finger protein that is required for male meiosis. *Plant Cell* 15, 1281–1295. doi: 10.1105/tpc.010447

Zhang, G. J., Guo, G. W., Hu, X. D., Zhang, Y., Li, Q. Y., Li, R. Q., et al. (2010). Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. *Genome Res.* 20, 646–654. doi: 10.1101/gr.100677.109

Zhang, L., and Ma, H. (2012). Complex evolutionary history and diverse domain organization of SET proteins suggest divergent regulatory interactions. *New Phytol.* 195, 248–263. doi: 10.1111/j.1469-8137.2012.04143.x

Zhang, X., Feng, B., Zhang, Q., Zhang, D., Altman, N., and Ma, H. (2005). Genome-wide expression profiling and identification of gene activities during early flower development in *Arabidopsis*. *Plant Mol. Biol.* 58, 401–419. doi: 10.1007/s11103-005-5434-6

Zhong, Y. F., and Holland, P. W. (2011). HomeoDB2: functional expansion of a comparative homeobox gene database for evolutionary developmental biology. *Evol. Dev.* 13, 567–568. doi: 10.1111/j.1525-142X.2011.00513.x

Zhou, X., and Ma, H. (2008). Evolutionary history of histone demethylase families: distinct evolutionary patterns suggest functional divergence. *BMC Evol. Biol.* 8, 294. doi: 10.1186/1471-2148-8-294

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 31 October 2014; accepted: 22 December 2014; published online: 20 January 2015.

*Citation: Zhang L, Wang L, Yang Y, Cui J, Chang F, Wang Y and Ma H (2015) Analysis of Arabidopsis floral transcriptome: detection of new florally expressed genes and expansion of Brassicaceae-specific gene families. Front. Plant Sci. **5**:802. doi: 10.3389/fpls.2014.00802*

This article was submitted to Plant Evolution and Development, a section of the journal Frontiers in Plant Science.