Comparative Transcriptome Analysis between Ornamental Apple Species Provides Insights into Mechanism of Double Flowering

Hera Gul 1,†, Zhaoguo Tong 1,†, Xiaolei Han 1, Iqra Nawaz 1, Safdar Ali Wahocho 1, Shumaila Khan 2, Caixia Zhang 1, Yi Tian 1, Peihua Cong 1,* and Liyi Zhang 1,*

1 Research Institute of Pomology, Chinese Academy of Agricultural Science (CAAS); Key Laboratory of Biology and Genetic Improvement of Horticultural Crops (Germplasm Resources Utilization), Ministry of Agriculture, Xinghai South Street 98, Xingcheng 125100, China; heragul30@yahoo.com (H.G.); tongzhaoguo@163.com (Z.T.); yichunxil@sina.com (X.H.); iqranawazcaas@gmail.com (I.N.); sawahoocho@yahoo.com (S.A.W.); cxzhang-bj@163.com (C.Z.); tianyi@caas.cn (Y.T.)

2 Institute of Vegetables and Flowers, Graduate School of Chinese Academy of Agricultural Sciences, Beijing 100081, China; shaila_dimple@yahoo.com

* Correspondence: congph@163.com (P.C.); zhangliyi@caas.cn (L.Z.); Tel.: +86-13998-905918 (P.C.); +86-13998-36122 (L.Z.)

† The authors have equal contribution.

Received: 26 December 2018; Accepted: 20 February 2019; Published: 26 February 2019

Abstract: Double-flower ornamental crabapples display eye-catching morphologies in comparison to single flower, but the genetic basis of double-flower development is not yet well known in apples. In order to comprehensively understand the differential expression of genes (DEGs) between single and double flower, the transcriptome of double flower crabapples Malus Kelsey, Malus micromalus, Malus Royalty, and a single flower cultivar Malus Dolgo were compared by RNA-sequencing. The results showed that there were 1854 genes in overlapped DEGs among all sample comparisons in apple single and double flower varieties. A large number of development and hormone related DEGs were also recognized on the basis of GO and KEGG annotations, and most of the genes were found to be down-regulated in double flowers. Particularly, an AGL24-MADS-box gene (MD08G1196900) and an auxin responsive gene (MD13G1137000) were putatively key candidate genes in the development of double flower by weighted gene co-expression network analysis (WGCNA). The study provides insights into the complex molecular mechanism underlying the development of the double flower in apple.

Keywords: apple; double flower; transcriptome (RNA-seq); floral organogenesis

1. Introduction

Double flower ornamental crabapples are one of the most eye-catching entity of plant landscape [1–3] but, little is known about the molecular mechanisms of double flowering in apple. The well-established ABCE model on flower development clearly explains how a few key regulatory genes, principally the ABC(E) genes, could control floral organ identity in double flowers [4,5]. In addition, researches on model plants Arabidopsis and Antirrhinum showed that the development of double flower is associated with plant hormones and environmental factors, such as temperature [6–9]. In previous studies, many specifically expressed transcription factor (TF) genes were identified during the double flowering process, besides, hormones also play a vital role in double flowering [10–12], suggesting a complex gene regulatory network underlying flower development regulating different floral forms [13]. Although, several double flowering related genes including MADS box and hormone
signaling genes have also been identified from certain apple species [14–19] but still the molecular mechanism behind these changes is not yet fully known [6,20–22].

In recent years, genome-wide transcriptomic analyses of floral organs in plants have provided a new tool to better understand the different floral forms from single to double [23]. Several studies have demonstrated the feasibility of understanding the differences in single and double flower morphologies by gene expression profiles [24,25], but no such study is found on crabapple double flowers. Therefore, to gain a deeper insight into the molecular mechanism of double flowering, we performed RNA-seq analysis for single flowers of *Malus* Dolgo, double flowers of *Malus* Kelsey, *Malus micromalus*, *Malus* Royalty, and the single flowers of *M.* Royalty, and identified the genes involved in double flower on the basis of differential expression of genes including MADS-box genes and hormone responsive genes.

2. Materials and Methods

2.1. Plant Material and Growth Conditions

Small flower buds at the very initial stage (when the floral buds appeared) from *Malus* Dolgo, *Malus* Kelsey, *Malus micromalus* and *Malus* Royalty [26] were used in this research. All the flower samples were collected in April 2016 from 10 year-old trees grafted on Borkh (*Malus baccata*) rootstock grown under natural open field conditions in The National Repository of Apple Germplasm Resources, China. The samples were labeled as SFD (Single Flower *M.* Dolgo), DFK (Double Flower *M.* Kelsey), DFM (Double Flower *M. micromalus*), DFR (Double Flower *M.* Royalty) and SFR (Single Flower *M.* Royalty) respectively. Each sample was collected in three replicates namely SFD1-3, DFK1-3, DFM1-3, DFR1-3, and SFR1-3, immediately immersed in liquid nitrogen and stored at −80 °C for future use until RNA extraction.

2.2. RNA Extraction and Library Preparation for Illumina Sequencing

All the samples were ground to powder in liquid nitrogen by constant crushing, using sterilized and chilled mortar and pestle. TRIzol Reagent kit (TIANGEN, Beijing, China) was used to isolate total RNA according to the manufacturer’s instruction. Thereafter, the samples were purified by using RNase-free DNase I (TaKaRa, Kyoto, Japan) to remove the genomic DNA. The quality of RNA was monitored by electrophoresis on 1% agarose gels. RNA purity and concentration were checked using a NanoPhotometer spectrophotometer (Implen, Westlake Village, CA, USA). RNA integrity was assessed using the Bioanalyzer 2100 system (Agilent Technologies, San Diego, CA, USA).

2.3. Construction of Library

After the extraction of total RNA, it was treated with Oligo (dt) to isolate mRNA and then mixed with the fragmentation buffer. Briefly, 3 µg of RNA was used as the input for each RNA sample preparation. The cDNA was synthesized by using the mRNA fragments as templates. Short fragments were purified and resolved with buffer EB (elution buffer) for end reparation and single nucleotide-A (adenine) addition. After that, the short fragments were connected with adapters and the suitable fragments were selected for the PCR amplification. During the QC (quality control) steps, Agilent 2100 Bioanalyzer and ABI StepOnePlus Real-Time PCR system were used in qualification and quantification of the sample library respectively. Then, the library was sequenced using Illumina HiSeq 4000 sequencer. Library construction and RNA-seq were performed by Beijing Genomics Institute (BGI) Co. Ltd. China.

2.4. Sequencing Quality Control and Reads Mapping to the Reference Genome

After sequencing, raw data in FASTQ [27] format were first processed through SOAPnuker v1.5.2 software. In this step, clean data were obtained by removing low-quality reads, and reads containing the adapter sequence (poly-N). At the same time, Q20 and Q30 of the clean data were calculated. All subsequent analyses were based on these clean datasets.
The genome was then mapped to the reference genome database of apple available at Genome Database for Rosaceae (http://www.rosaceae.org) [28]. An index of the reference genome was built using Bowtie2 [29], and paired-end clean reads were aligned to the reference genome using HISAT (Hierarchical Indexing for Spliced Alignment of Transcripts) software v2.0.4 [30]. We used StringTie [31] to reconstruct transcripts. After that, we used Cuffmerge [32] and Cuffcompare [32] respectively to reconstruct the information of all the samples together and compare the reconstructed transcripts to reference annotations available through the genome browser [33]. By using CPC software [34], protein coding potential of the new transcripts was predicted, then coding novel transcripts were merged with reference transcripts to get a complete reference for further analysis.

2.5. Differentially Expressed Gene Detection

We used RNA-seq by expectation-maximization (RSEM) [35] to calculate the expression level of genes and transcripts. After that, cor and hclust functions of R were used to calculate the Pearson correlation and hierarchical clustering between all the samples [36]. The expressional abundances in each library were normalized to transcript per million, where, the most differentially expressed genes (DEGs) were selected. The significance of the differential transcript expression was determined by using the False Discovery Rate (FDR) control method [37] to justify the p-values. The only genes with an absolute log2 fold change [38] of ≥+1 (up-regulated) and ≤−1 (down-regulated), and a FDR significance score of ≤0.01, were included in subsequent analysis. For heat maps, the log10FPKM values were calculated for all DEGs, normalized between −1.5 to +1.5, and used as an input to the heatmap function of R.

2.6. Gene Enrichment Analysis

To describe the characteristics and reaction features of DEGs that were retrieved from samples, we conducted gene ontology (GO) analysis and also performed GO functional enrichment using phyper, a function of R. FDR was calculated for each p-value, in general, the terms with FDR ≤ 0.01 were identified as significantly enriched. For pathway analysis, we mapped all the DEGs in terms of the Kyoto Encyclopedia of Genes and Genomes (KEGG) and retrieved the significantly enriched pathway with p-value ≤ 0.05 [39]. With KEGG annotation results, we classified the DEGs accordingly and performed pathway enrichment using the phyper function of R. To find the transcription factors (TFs) of DEGs, we used getorf [40], then aligned ORF to TF domain (from PlntfDB) by using hmmsearch [41], and coded according to the characteristics of the TF family.

2.7. Co-Expression Gene Models

All the genes expressed in three replicates of each sample were first used to calculate the mean FPKM values of respective samples. After that, genes having normalized reads, lower than 0.1 FPKM were considered “too lowly expressed genes” and were removed from analysis. Leftover genes after filtration, were then subjected to calculate the DEGs between DFK, DFM, DFR, and SFD. Furthermore, DEGs with log2FoldChange ≥1 and ≤−1 were used for sample comparisons of all double flowers to single flower (SFD). The weighted gene co-expression network analysis (WGCNA) R package was used to analyze and perform pairwise correlations between a limited number (~3000 genes) of most differentially expressed genes. The steps involved in analysis and module construction are described in detail by [42]. Clusters of highly interconnected genes i.e., modules, were constructed and to further determine the biological meaning of the identified co-expression module, the up/down-regulation of the gene within the modules were analyzed [43]. The co-expression network was further visualized using the free software Cytoscape [44].

2.8. Validation of RNA-seq Data

Different genes from DEGs were selected to validate the RNA-Seq data. The corresponding sequences were retrieved from the apple genome database, available at Genome Database for
Rosaceae (GDR) [45] (http://www.rosaceae.org). Primers for the selected genes were designed using Beacon Designer-8 software. Moreover, the total RNA from the samples was isolated by using TaKaRa MiniBEST universal RNA Extraction Kit (Cat#9767 TaKaRa, Shiga, Japan). First-strand cDNA was synthesized from 1 µg of total RNA, using PrimeScript RT reagent kit with gDNA Eraser (perfect real time) (Cat#RR047A TaKaRa, Shiga, Japan), and was diluted 10 folds with ddH2O before further use as a template for quantitative real-time PCR (qRT-PCR) assays. qRT-PCR was performed in 96-well plates on BIO-RAD CFX Connect Real-Time System using SYBR Premix DimerEraser (Cat#RR091A Clontech, Madison, WI, USA). Actin was used as an internal control during qRT-PCR, along with NTC (no template control), to get more accurate results [46]. The thermal cycling conditions were set at 95 °C for 3 min, followed by 30 cycles of 94 °C for 20 s, 60 °C for 30 s and 72 °C for 30 s. The relative expression level of the selected genes was calculated with the $2^{-ΔΔCT}$ method [47].

3. Results

3.1. RNA Sequencing of Different Samples

In order to have a summary of the transcriptome, RNA-seq libraries from single and double flowers were generated using RNA samples from four different species of ornamental crabapples (Figure 1a) in three replicates. We acquired in total reads ranging from 53.56 to 57.16 Mb with a ratio of total clean reads ranging from 77.73% to 83.92% (Figure 1b, Table 1), while the average ratio of gene set rate was 78.10%. The high quality and rate of genome coverage of the present data exposed that it could be used for further bioinformatics study. Afterwards, the blast tool was used to identify the transcripts based on homology to the reference genes. About 42,009 genes were detected, including 39,628 known genes, and 2381 novel genes predicted in total. The data for each sample is listed in Table S1 of the Supplementary Materials. Venn analysis for replicates of each sample provided the common genes, which were used for downstream analysis (Figure 1c). Hierarchical cluster analysis between all the samples showed a high and indistinguishable genetic similarity between DFR and SFR from the same tree (Figure 1d).

### Table 1.

| Sample ID | Total Raw Reads (Mb) | Total Clean Reads (Mb) | Total Clean Bases (Gb) | Clean Reads Q20 (%) | Clean Reads Q30 (%) | Clean Reads Ratio (%) | Total Mapping Ratio (%) | Uniquely Mapping Ratio (%) |
|-----------|---------------------|------------------------|------------------------|---------------------|---------------------|-----------------------|-------------------------|--------------------------|
| SFD1      | 55.52               | 44.06                  | 6.61                   | 97.91              | 93.36               | 79.36                 | 76.62                   | 37.36                    |
| SFD2      | 53.89               | 44.27                  | 6.64                   | 98.12              | 93.86               | 82.15                 | 76.79                   | 37.04                    |
| SFD3      | 55.53               | 44.12                  | 6.62                   | 97.94              | 93.44               | 79.45                 | 75.93                   | 37.16                    |
| DFK1      | 53.56               | 42.04                  | 6.31                   | 97.84              | 93.18               | 78.48                 | 72.53                   | 34.98                    |
| DFK2      | 57.16               | 44.60                  | 6.69                   | 97.85              | 93.22               | 78.03                 | 71.89                   | 34.2                     |
| DFK3      | 57.16               | 44.72                  | 6.71                   | 97.97              | 93.48               | 78.23                 | 73.03                   | 34.46                    |
| DFM1      | 57.16               | 44.75                  | 6.71                   | 97.89              | 93.31               | 78.29                 | 73.76                   | 34.76                    |
| DFM2      | 55.90               | 43.45                  | 6.52                   | 97.82              | 93.15               | 77.73                 | 72.37                   | 34.8                     |
| DFM3      | 57.16               | 45.20                  | 6.78                   | 97.9               | 93.36               | 79.07                 | 73.56                   | 35.32                    |
| DFR1      | 57.16               | 44.55                  | 6.68                   | 98                 | 93.6               | 77.95                 | 71.5                    | 34.17                    |
| DFR2      | 53.90               | 45.02                  | 6.75                   | 98.64              | 95.67               | 83.52                 | 72.11                   | 34.82                    |
| DFR3      | 53.90               | 44.78                  | 6.72                   | 98.66              | 95.73               | 83.09                 | 71.07                   | 34.61                    |
| SFR1      | 53.90               | 45.23                  | 6.78                   | 98.69              | 95.84               | 83.92                 | 72.06                   | 34.74                    |
| SFR2      | 55.53               | 44.43                  | 6.66                   | 98.09              | 94.09               | 80.01                 | 72.36                   | 34.93                    |
| SFR3      | 55.53               | 45.03                  | 6.75                   | 98.64              | 95.7               | 81.09                 | 69.09                   | 33.51                    |
| Average   | 55.53               | 44.42                  | 6.66                   | 98.13              | 94.07               | 80.02                 | 72.98                   | 35.13                    |
By comparing the transcript levels of each gene between single flower SFD compared to all the other samples (DFK, DFM, DFR, and SFR), we generated heat maps for DEGs of all the samples shown in Figure 2. The heat maps for SFR and DFR, showed the highest similarity of expressed genes in double flowers. In double flowers, along with SFR, most of the DEGs were observed as downregulated as compared to SFD, which indicated that in double flowers most of the expressed genes were downregulated. Afterwards, more detailed pairwise comparisons between all the samples were made to understand the pattern of expressed genes in double flowers. Analyzing on basis of \( p \)-value \( \leq 0.05 \) and log2fold change (log2FC) \( \geq 1 \) or \( \leq -1 \), a total of 71933 DEGs in all comparative transcriptomes were retrieved. In DFK-VS-SFD 5041 DEGs were down-regulated while 3279 were up-regulated, in DFM-VS-SFD 3247 DEGs were down-regulated while 2983 were up-regulated, in DFR-VS-SFD 4652 DEGs were down-regulated while 3455 were up-regulated, in SFR-VS-SFD 4950 genes were down-regulated while 3142 were up-regulated, in DFK-VS-SFR 3786 genes were down while 3878 were up-regulated, in DFM-VS-SFR 3247 DEGs were down-regulated while 4720 genes were up-regulated, in DFR-VS-SFD 2 genes were down-regulated while 4720 genes were up-regulated, in DFR-VS-SFD 2 genes were
down-regulated while 69 were up-regulated, in DFK-VS DFM 5008 were down-regulated while 3635 were up-regulated, in DFK-VS-DFR 4751 genes were down-regulated while 4291 were up-regulated and in DFR-VS-DFM 4408 genes were down-regulated while 3427 were up-regulated (Table S2). Furthermore, DEGs overlapped between all the single and double flower comparative samples were analyzed, where 1854 genes were overlapped while 2737, 1875, 841, and 612 DEGs were distinctive for DFK-VS-SFD, DFM-VS-SFD, DFR-VS-SFD, and SFR-VS-SFD, respectively (Figure 3).

Figure 2. Heat map illustrating the expression profiles of differentially expressed genes (DEGs) between all sample comparisons in apple single and double flower varieties.

Figure 3. Illustration of all differentially expressed genes (DEGs) between single and double flower cultivars using a Venn diagram.
3.3. Gene Ontology and Kyoto Encyclopedia of Genes and Genome (KEGG) analysis

According to the results of differential gene detection, we performed gene ontology (GO) classification and functional enrichment for DEGs. GO has three functional categories: Cellular component, molecular function, and biological process for which we performed functional enrichment analysis. The detailed GO analysis of DEGs between single flower and double flower was performed, and significantly enriched GO terms were obtained by using false discovery rate (FDR) ≤ 0.01 (Figure 4). All the cumulative genes of double flowers were compared to SFD and were grouped into 49 GO terms; including 21 biological process terms, 15 cellular components and 13 molecular function related terms. For biological process, metabolic process and cellular processes were dominant terms; for the cellular component, the dominant terms were cell, cell part and organelle; and for molecular function, a large percentage of genes were related to catalytic activity, binding, and transporter activity (Table S3).

![Figure 4. Gene ontology analysis between single and double flower varieties showing abundance of differentially expressed GO-terms.](image)

Numerous flower-development and flowering genes play a vital role in entire flower development cycle [48]. In total, 11 DEGs in double flowers showed similarity to known flower development...
and flowering related genes within the NCBI and Uniport databases (Table S4). According to this data, three flowering promoting factor genes (MD15G1109300, MD16G1170700, and MD08G1131700) showed higher expression levels in double flowers than in single flower. Furthermore, two early flowering protein genes (MD01G1152700 and MD11G1074300), two flowering time control genes (MD17G1123300 and MD16G1150700), one flowering locus T gene (MD12G1262000), one flowering locus C gene (MD10G1041100), one protein terminal flower-like gene (MD14G1021100), and one Protein MOTHER of FT and TF (MFT) were also identified.

Based on the GO terms (GO: 0009908 and GO: 0009909), 25 DEGs were identified as flower development and regulation of flower development-related genes (Table S5). Most of the genes were up-regulated in double flowers but interestingly one flowering gene, a predicted glycerol-3-phosphate 2-O-acyltransferase (MD04G1148700), was very highly expressed in double flowers as compared to single flower Dolgo.

The KEGG pathways signified by all the DEGs, were predicted to further set the involvement of 135 KEGG pathways in flower organogenesis (Table S6). In all the sample comparisons the highest genes representation was observed in metabolic pathways with average of 1556 genes followed by biosynthesis of secondary metabolites with an average of 870 genes, plant-pathogen interaction with average of 345 genes, and endocytosis with 284 genes. In KEGG pathways analysis the trend of regulation of pathways enrichment showed that almost all the differentially enriched pathways were down-regulated in double flowers (Figure 5).

### 3.4. Recognizing TF-Encoding Genes Associated with Floral Organ Development

TF genes seems to be involved in the regulation of various physiological systems in response to floral organogenesis and flower development [49]. All DEGs were predicted with transcription factor. In total 2923 TF genes were identified as DEGs and classified in 60 different TF families. In all the sample groups, the top largest differentially expressed TF families were: MYB (402 members), followed by MYB-related (334 members), AP2-EREBP (218 members), bHLH (191 members), NAC (177
members), WRKY (118 members) and MADS (116 members) as shown in Figure 6. Furthermore, several hormone-related TFs were also identified as DEGs, including 20 GRASs and 31 ARFs.

MADS box genes play a very pivotal role in double flowering [18,50]. In this study, we identified 116 MADS-box genes, in which 46 DEGs were significantly expressed on basis of their log2FC values (Table S7). Among these, 16 genes like AGL15, AGL18, AGL17, SOC1, AGL24, and SVP belongs to MIKCc group of MADS family, while several other MADS genes like AGL80, AGL61, AGL62, FBP24, AGL12, AGL14, SEP1/MADS8, MADS15, MADS22, PHERES2, ZMM17, CMB1, and JOINTLESS were also identified (Supplementary Figure S1).

3.5. Recognizing DEGs Related to Hormones

A huge number of DEGs involved in hormones metabolism and interactions were identified between double flowers and single flower comparison. The KEGG pathway analysis assigned the DEGs to key elements involved in numerous hormone signaling pathways. Interestingly, an outsized variety of hormone-related DEGs belonged to auxin, gibberellic acid (GA), brassinosteroid (BRs), and cytokinin (CK) signaling pathways. An overview of the different hormonal signaling networks in apple flowers is shown in Figure 7. Most hormone related DEGs were preponderantly expressed in SFD as compared to double flower.
Firstly, DEGs associated with auxin were analyzed. For auxin biosynthesis and metabolism, 8 YUCCA genes were categorized as DEGs. Seven putative auxin influx carrier-encoding AUX1/LAX genes and 17 putative GH3 genes were categorized as DEGs, while for auxin transport, 25 efflux carrier-encoding genes were identified as DEGs. For downstream response, 27 putative auxin-repressed genes AUX/IAA and 27 auxin response factor ARF-encoding genes were identified as DEGs. After that, DEGs associated with GA were identified and analyzed. For GA biosynthesis, 52 putative DEGs encoding GA20OX, GA2OX, and GA3OX were identified, while 7 putative DEGs were annotated as GA receptors GID1. For GA downstream responses, nine putative DELLA, two F-box protein GID2, nine phytochrome-interacting factors3, and eight phytochrome-interacting factors4-responsive genes were identified as differentially expressed. For ABA biosynthesis, 42 genes were identified, which were differentially expressed, encoding CYB707A, ABA2, ABA1, AOG, and NCED, while for receptors and transporters 2 PYL-encoding genes were identified. For downstream responses, 12 DEGs were annotated as PP2C, SnRK2, and ABA responsive element binding factor (ABF) genes. Furthermore, 79 putative genes were identified as DEGs and mapped to CK pathway (Figure 7a). Based on reads per kilobase per million mapped reads (RPKM) values, the expression level of the hormone-related genes were analyzed and shown in Figure 7b. The expression levels of most hormone-related genes in the double flower varieties were less than those in single flowers (Table S8).
3.6. Weighted Gene co-Expression Analysis (WGCNA)

A WGCNA Analysis was performed to identify the genes related to double flowering. The similarity of genes that were expressed across conditions (single flower and double flower), governed the double flowering response. After filtering from lowly expressed genes, the expression values from remaining 2969 DEGs were analyzed and 9 WGCNA modules were identified. The number of genes per module ranged from 5 (grey module) to 1483 (turquoise module) (Figure 8a,b).

![Weighted gene co-expression network analysis (WGCNA) of DEGs identified from double flower and single flower comparison.](image)

Figure 8. Weighted gene co-expression network analysis (WGCNA) of DEGs identified from double flower and single flower comparison. (a) Hierarchical cluster tree showing nine modules of co-expressed genes. Each of the DEGs used in this analysis is represented by a tree leaf and each of the modules by major tree branch. The panel below shows the colors representing each module. (b) Module-trait relationships (correlation) and corresponding p-values (given in parenthesis). On the left side the panel shows nine modules while the color scale on the right shows the correlations ranging from −1 (green) to 1 (red). The left panel represents double flowering as a trait while the right panel represents the single flower as a trait. (c) Cytoscape network analysis of co-expressed genes with edge weights ≥0.1 in module “turquoise”. On the right the color scale represents the interaction degree of genes among each other ranging from 10 (red) to 3 (blue). The larger the size of the circle representing the gene the more the degree of outgoing interactions with other genes, while more thicker the edge line the greater the correlation efficiency.

The analysis of the module-trait relationship revealed that the turquoise module was highly correlated with the double and single flower genes ($r = −1$ & $1; p = 2 \times 10^{-10}$). The correlation coefficients with the same expression level but opposite expression pattern (i.e., negative in the case of the double flower) indicated that most of the genes were downregulated in double flowers with respect to single flower. In this module, we found 11 structural genes involved in the double flowering mechanism (Table S9). The expression pattern of these genes were highly connected to each other, even when we used the stringent edge weight cut-off of 0.1 (Figure 8c, Table S10). Specifically, MD08G1196900 which is a MADS-box gene (AGL24), had the highest number of edges connecting to other genes, followed by MD13G1137000, which is an auxin responsive protein IAA31 gene.
3.7. Validation of the Randomly Selected Genes Expression

To verify the RNA-seq results of gene differential expression, genes MD16G1148300, MD08G1196900, MD15G1014200, MD17G1259400, MD10G1007100, MD07G1174000, MD14G1054500, MD12G1198600, MD06G1022600, MD09G1074100, MD13G1137000, and MD03G1116000 (details are listed in Table S11), were selected from the data to perform qRT-PCR assay with independent samples collected from single flower and all double flower varieties. The expression results are attached in Table S12 of the Supplementary Materials. All the genes showed expression levels in sync with RNA-seq data (Figure 9). Correlation analysis showed that qRT-PCR results were highly correlated to RNA-seq results (Figure S2). In short, the expression levels of these selected genes were basically consistent with RNA-seq results and thus, validated the RNA-seq analysis results.

![Figure 9. Real-time quantitative PCR validation profiles of 12 randomly selected genes from flowers of different varieties of apple along with single and double flowers from same tree; the data was normalized by using actin as an internal reference. Data are represented as mean ± SD for three biological replicates.](image-url)
4. Discussion

The evolution of double flower in apple is intriguing, but little is known about the molecular mechanism of double flower morphology. In our study, we carried out a comparative transcriptome analysis of flowers of three double flowered cultivars to a single flower cultivar, and acquired significant DEGs. Due to the fact that several transcription factor families, such as the MADS-box, ARF, bHLH, and NAC, were associated with floral development and organogenesis [51–54], we performed a detailed analysis about differential transcription factors. Particularly, several evidence of involvement of AGAMOUS-like C-class genes repression by transposable element (TE) insertion could enhance the petal number in ranunculid, Arabidopsis and many other species [55–57]. In apple, RNAi of apple AG-like led to showy polypetalous flowers [18]. In this study, our data extended previous findings that five AGL24-like MIKC-type MADS-box genes were down-regulated in double flowers, which may be involved in double flowering. Interestingly, among these, a MADS-box (MD08G1196900) was predicted as most significant in terms of interactions, correlations and WGCNA analysis, but its function remains to be investigated. Additionally, overexpression of PISTILLATA gene MADS8 contributes to the conversion of sepals into petal-like structures [58–60], and over expression of E-class SEPALLATA were involved in the floral organs transition from stamens to petals in Arabidopsis and some other plants [61,62]. Here, we identified three up-regulated genes (a MADS8 and two SEP3-like) in double flowers, which provide a new clue to further investigate floral transitions in double flower apple in future.

The regulation network of double flower in apple is also complex. Various endogenous environmental signals are related to an array of cellular and biochemical processes throughout floral organs formation [8]. Among these signals, the phytohormones, such as auxins, gibberellins (GAs), ethylene, abscisic acid (ABA) are endogenously occurring compounds involved in floral organ development and transitions [63]. Our enrichment analysis of KEGG pathway showed that the DEGs were considerably enriched to hormone signaling pathways associated with signal transduction and phytohormone metabolism processes throughout the double flowering [63–65]. Also, several environmental factors like low temperature, induces DNA methylation and regulates the hormones that alters the petal number in certain flowering plants [66–68]. In the present study, the AGAMOUS genes were down-regulated in double flowers and so was the ARF3, while the cytokinin was up-regulated in double flowers, which was in agreement with reports [69,70]. In addition, 52 putative GA20OX, GA2OX, and GA3OX genes and 34 putative DELLA genes were identified as DEGs suggesting the involvement of diverse regulatory mechanisms in GA biosynthesis during apple flower development [71,72]. However, the more complex problem is that there are both single flower and double flowers in the same tree, such as royalty cultivar (Figure 1). This causes us to think more deeply and discover the methylome of royalty through bisulfite sequencing in future.

5. Conclusions

In our study the RNA-Seq data analysis between a single flower and three double flower species was carried out, which led to the identification of double flower-specific expression patterns of flowering related genes focusing on morphogenesis of flower. Furthermore WGCNA analysis and subjection of genes to Cytoscape analysis led to the detection of two key genes MD08G1196900 (AGL24-which is a MADS-box gene) and MD13G1137000 (IAA31-which is an auxin responsive protein gene) indicating their strong involvement in the double flowering mechanism. Besides, these DEGs could be used for further experiments to validate the double flower mechanism in apple plants.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4395/9/3/112/s1, Table S1: statistics of predicted genes, Table S2: statistics of DEGs between samples, Table S3: GO classification of all sample comparisons, Table S4: flowering related genes differentially expressed between all sample comparisons, Table S5: the gene list of GO:0009908 & GO:0009909 differentially expressed between all sample comparisons, Table S6: the KEGG pathways represented by all the DEGs, Table S7: DEGs related to MADS-box TFs in all sample comparisons, Table S8: hormones regulation in double flower varieties-VS-single flower control variety, Table S9:
key genes involved in double flowering, Table S10: Cytoscape gene network analysis for key genes involved in double flowering, Table S11: details of genes selected for qRT-PCR assay, Table S12: gene expression values from qPCR, Figure S1: expression trends of differentially expressed MADS-box transcriptional factors regulating double flowering, Figure S2: Correlation Analysis of qRT-PCR data and RNA-seq data.

Availability of Supporting Data: All the original RNA-Seq data has been submitted to the NCBI Sequence Read Archive under the accession number of PRJNA488997.

Author Contributions: Conceptualization, P.C. and L.Z.; methodology, X.H.; software, Z.T.; validation, H.G.; formal analysis, Z.T. and Y.T.; writing—original draft preparation, H.G.; visualization, H.G. and Z.T.; writing-review and editing, I.N., S.A.W. and S.K.; supervision, L.Z. and P.C.; project administration, C.Z.

Acknowledgments: Supported by the funds China Agriculture Research System (CARS-27) and the Agricultural Science and Technology Innovation Program (CAAS-ASTIP-2016-RIP-02).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Reynold, J.; Tampion, J. Double Flowers: A Scientific Study; Polytechnic of Central London Press: Pembridge, UK, 1983; p. 41.
2. Ramirez, F.; Davenport, T.L. Apple pollination: A review. Sci. Hortic. 2013, 162, 188–203. [CrossRef]
3. Garratt, M.P.D.; Breeze, T.D.; Boreux, V.; Fountain, M.T.; Mckerchar, M.; Webber, S.M.; Coston, D.J.; Jenner, N.; Dean, R.; Westbury, D.B. Apple Pollination: Demand Depends on Variety and Supply Depends on Pollinator Identity. PLoS ONE 2016, 11, e0153889. [CrossRef] [PubMed]
4. Bowman, J.L.; Smyth, D.R.; Meyerowitz, E.M. The ABC model of flower development: then and now. Development 2012, 139, 4095–4098. [CrossRef] [PubMed]
5. Chanderbali, A.S.; Berger, B.A.; Howarth, D.G.; Soltis, P.S.; Soltis, D.E. Evolving ideas on the origin and evolution of flowers: New perspectives in the genomic era. Genetics 2016, 202, 1255–1265. [CrossRef] [PubMed]
6. Zik, M.; Irish, V.F. Global identification of target genes regulated by APETALA3 (AP3) and PISTILLATA floral homeotic gene action. Plant Cell 2003, 15, 207. [CrossRef] [PubMed]
7. Kramer, E.M.; Hall, J.C. Evolutionary dynamics of genes controlling floral development. Curr. Opin. Plant Biol. 2005, 8, 13–18. [CrossRef] [PubMed]
8. Sablowski, R. Flowering and determinacy in Arabidopsis. J. Exp. Bot. 2007, 58, 899–907. [CrossRef] [PubMed]
9. Egea-Cortines, M.; Saedler, H.; Sommer, H. Ternary complex formation between the MADS-box proteins SQUAMOSA, DEFICIENS and GLOBOSA is involved in the control of floral architecture in Antirrhinum majus. EMBO J. 1999, 18, 5370–5379. [CrossRef] [PubMed]
10. Cheng, Y.; Zhao, Y. A role for auxin in flower development. J. Integr. Plant Biol. 2007, 49, 99–104. [CrossRef]
11. Cheng, Y.; Dai, X.; Zhao, Y. Auxin biosynthesis by the YUCCA flavin monooxygenases controls the formation of floral organs and vascular tissues in Arabidopsis. Genes Dev. 2006, 20, 1790–1799. [CrossRef] [PubMed]
12. Pfluger, J.; Zambryski, P. The role of SEUSS in auxin response and floral organ patterning. Development 2004, 131, 4697–4707. [CrossRef] [PubMed]
13. Matiashernandez, L.; Aguilarjaramillo, A.E.; Cigliano, R.A.; Sanseverino, W.; Pelaz, S. Flowering and trichome development share hormonal and transcription factor regulation. J. Exp. Bot. 2016, 67, 1209–1219. [CrossRef] [PubMed]
14. Hättasch, C.; Flachowsky, H.; Hanke, M.V.; Lehmann, S.; Gau, A.; Kapturska, D. The switch to flowering: genes involved in floral induction of the apple cultivar ‘Pinova’ and the role of the flowering gene MdFT. In International Symposium on Biotechnology of Fruit Species: BIOTECHFRUIT2008; Hanke, M.-V., Dunemann, F., Flachowsky, H., Eds.; International Society for Horticultural Science: Dresden, Germany, 2009; pp. 701–705.
15. Vimolmangkang, S.; Han, Y.; Wei, G.; Korban, S.S. An apple MYB transcription factor, MdMYB3, is involved in regulation of anthocyanin biosynthesis and flower development. BMC Plant Biol. 2013, 13, 176. [CrossRef] [PubMed]
16. Der Linden, C.G.V.; Vosman, B.; Smulders, M.J.M. Cloning and characterization of four apple MADS box genes isolated from vegetative tissue. J. Exp. Bot. 2002, 53, 1025–1036. [CrossRef]
17. Sung, S.; Yu, G.; Nam, J.; Jeong, D.; An, G. Developmentally regulated expression of two MADS-box genes, MdMADS3 and MdMADS4, in the morphogenesis of flower buds and fruits in apple. Planta 2000, 210, 519–528. [CrossRef] [PubMed]
18. Klocko, A.L.; Borejsza, W.E.; Brunner, A.M.; Shevchenko, O.; Aldwinckle, H.; Strauss, S.H. Transgenic suppression of AGAMOUS genes in apple reduces fertility and increases floral attractiveness. *PLoS ONE* 2016, *11*, e0159421. [CrossRef] [PubMed]

19. Tanaka, N.; Tanaka-Moriya, Y.; Mimida, N.; Honda, C.; Iwanami, H.; Komori, S.; Wada, M. The analysis of transgenic apples with down-regulated expression of MdPISTILLATA. *Plant Biotechnol.* 2016, *33*, 395–401. [CrossRef]

20. Di Stilio, V.S.; Kramer, E.M.; Baum, D.A. Floral MADS box genes and homeotic gender dimorphism in *Thalictrum dioicum* (Ranunculaceae)—A new model for the study of dioecy. *Plant J.* 2005, *41*, 755–766. [CrossRef]

21. Thomson, B.; Zheng, B.; Wellmer, F. Floral organogenesis: when knowing your ABCs is not enough. *Plant Physiol.* 2017, *173*, 56–64. [CrossRef] [PubMed]

22. Cliff, M. Development of a model for prediction of consumer liking from visual attributes of new and established apple cultivars. *Am. Pomol. Soc.* 2002, *4*, 223–229.

23. Hong, Y.; Jackson, S.D. Floral induction and flower formation—The role and potential applications of miRNAs. *Plant Biotechnol.* 2015, *13*, 282–292. [CrossRef] [PubMed]

24. Chanderbali, A.S.; Yoo, M.-J.; Zahn, L.M.; Brockington, S.F.; Wall, P.K.; Gitzendanner, M.A.; Albert, V.A.; Leebens-Mack, J.; Altman, N.S.; Ma, H. Conservation and canalization of gene expression during angiosperm diversification accompany the origin and evolution of the flower. *Proc. Natl. Acad. Sci. USA* 2010, *107*, 22570–22575. [CrossRef] [PubMed]

25. Zahn, L.M.; Ma, X.; Altman, N.S.; Zhang, Q.; Wall, P.K.; Tian, D.; Gibas, C.J.; Gharaiibeh, R.; Leebens-Mack, J.H.; Ma, H. Comparative transcriptionomics among floral organs of the basal eudicot *Eschscholzia californica* as reference for floral evolutionary developmental studies. *Genome Biol.* 2010, *11*, R101. [CrossRef] [PubMed]

26. Romer, J.P. A Srvey of Selection Preferences for Crabapple Cultivars and Species. 2002. Available online: https://lib.dr.iastate.edu/rtd/17486/ (accessed on 26 December 2018).

27. Cock, P.J.A.; Fields, C.J.; Goto, N.; Heuer, M.L.; Rice, P.M. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res.* 2010, *38*, 1767–1771. [CrossRef] [PubMed]

28. Daccord, N.; Celton, J.-M.; Linsmith, G.; Becker, C.; Choisne, N.; Schijlen, E.; van de Geest, H.; Bianco, L.; Micheletti, D.; Velasco, R.; et al. High-quality de novo assembly of the apple genome and methylation dynamics of early fruit development. *Nat. Genet.* 2017, *49*, 1099–1106. [CrossRef] [PubMed]

29. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 2012, *9*, 357–360. [CrossRef] [PubMed]

30. Pertea, M.; Pertea, G.; Antonescu, C.; Chang, T.C.; Mendell, J.T.; Salzberg, S.L. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 2015, *33*, 290–295. [CrossRef]

31. Daccord, N.; Celton, J.-M.; Linsmith, G.; Becker, C.; Choisne, N.; Schijlen, E.; van de Geest, H.; Bianco, L.; Micheletti, D.; Velasco, R.; et al. The Apple Genome and Epigenome. Available online: https://iris.angers.inra.fr/gddh13/ (accessed on 26 December 2018).

32. Li, B.; Dewey, C.N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* 2011, *12*, 323. [CrossRef] [PubMed]

33. Eisen, M.B.; Spellman, P.T.; Brown, P.O.; Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* 1998, *95*, 14863–14868. [CrossRef] [PubMed]

34. Benjamini, Y.; Yekutieli, D. The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* 2001, *29*, 1165–1188.
38. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014, 15, 550. [CrossRef] [PubMed]
39. Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 1999, 28, 27–30. [CrossRef]
40. Rice, P.M.; Longden, I.; Bleasby, A.J. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* 2000, 16, 276–277. [CrossRef]
41. Mistry, J.; Finn, R.D.; Eddy, S.R.; Bateman, A.; Punta, M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* 2013, 41. [CrossRef] [PubMed]
42. Zhang, B.; Horvath, S. A General Framework for Weighted Gene Co-expression Network Analysis. *Stat. Appl. Genet. Mol. Biol.* 2005, 4, 1–45. [CrossRef] [PubMed]
43. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* 2008, 9, 559. [CrossRef] [PubMed]
44. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003, 13, 2498–2504. [CrossRef] [PubMed]
45. Jung, S.; Lee, T.; Cheng, C.H.; Buble, K.; Zheng, P.; Yu, J.; Humann, J.; Ficklin, S.P.; Gasic, K.; et al. 15 Years of GDR: New Data and Functionality in the Genome Database for Rosaceae. *Nucleic Acids Res.* 2019, 47, D1137–D1145. [CrossRef] [PubMed]
46. You, Y.; Xie, M.; Vasseur, L.; You, M. Selecting and validating reference genes for quantitative real-time PCR in *Plutella xylostella* (L.). *Genome* 2018, 61, 349–358. [CrossRef] [PubMed]
47. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 2001, 25, 402–408. [CrossRef] [PubMed]
48. Sui, S.; Luo, J.; Ma, J.; Zhu, Q.; Li, M. Generation and analysis of expressed sequence tags from chimonanthus praecox (wintersweet) flowers for discovering stress-responsive and floral development-related genes. *Int. J. Genom.* 2012. [CrossRef]
49. Ali, M.A.; Azeem, F.; Nawaz, M.A.; Acet, T.; Abbas, A.; Imran, Q.M.; Shah, K.H.; Rehman, H.M.; Chung, G.; Yang, S.H.; et al. Transcription factors WRKY11 and WRKY17 are involved in abiotic stress responses in *Arabidopsis*. *J. Plant Physiol.* 2018, 226, 12–21. [CrossRef] [PubMed]
50. Malcomber, S.T.; Kellogg, E.A. SEPALLATA gene diversification: Brave new whorls. *Trends Plant Sci.* 2005, 10, 427–435. [CrossRef] [PubMed]
51. Reeves, P.H.; Ellis, C.M.; Ploense, S.E.; Wu, M.F.; Yadav, V.; Tholl, D.; Chetelat, A.; Haupt, I.; Kennerley, B.J.; Hodgens, C. A Regulatory network for coordinated flower maturation. *PLoS Genet.* 2012, 8, e1002506. [CrossRef] [PubMed]
52. Huang, T.; Irish, V.F. Gene networks controlling petal organogenesis. *J. Exp. Bot.* 2016, 67, 61–68. [CrossRef] [PubMed]
53. Kim, S.; Kim, S.; Park, C. A membrane-associated NAC transcription factor regulates salt-responsive flowering via FLOWERING LOCUS T in *Arabidopsis*. *Planta* 2007, 226, 647–654. [CrossRef] [PubMed]
54. Ning, Y.; Ma, Z.; Huang, H.; Mo, H.; Zhao, T.; Li, L.; Cai, T.; Chen, S.; Ma, L.; He, X. Two novel NAC transcription factors regulate gene expression and flowering time by associating with the histone demethylase JMJD1. *Nucleic Acids Res.* 2015, 43, 1469–1484. [CrossRef] [PubMed]
55. Galimba, K.D.; Tolkin, T.R.; Sullivan, A.M.; Melzer, R.; Theißen, G.; Di Stilio, V.S. Loss of deeply conserved C-class floral homeotic gene function and C- and E-class protein interaction in a double-flowered ranunculid mutant. *Proc. Natl. Acad. Sci. USA* 2012, 109, E2267–E2275. [CrossRef] [PubMed]
56. Gregis, V.; Sessa, A.; Dorcafornell, C.; Kater, M.M. The *Arabidopsis* floral meristem identity genes AP1, AGL24 and SVP directly repress class B and C floral homeotic genes. *Plant J.* 2009, 60, 626–637. [CrossRef] [PubMed]
57. Ma, H. Regulatory genes in plant development: MADS. *Essent. Life Sci.* 2009. [CrossRef]
58. Mondragonpalomino, M. Perspectives on MADS-box expression during orchid flower evolution and development. *Front. Plant Sci.* 2013, 4, 377.
59. Chang, Y.; Kao, N.; Li, J.; Hsu, W.; Liang, Y.; Wu, J.; Yang, C. Characterization of the possible roles for B class MADS box genes in regulation of perianth formation in orchid. *Plant Physiol.* 2010, 152, 837–853. [CrossRef] [PubMed]
60. Tsai, W.C.; Pan, Z.J.; Hsiao, Y.Y.; Chen, L.; Liu, Z. Evolution and function of MADS-box genes involved in orchid floral development. *J. Syst. Evol.* 2014, 52, 397–410. [CrossRef]

61. Chang, Y.; Chiu, Y.; Wu, J.; Yang, C. Four orchid (*Oncidium Gower Ramsey*) AP1/AGL9-like MADS box genes show novel expression patterns and cause different effects on floral transition and formation in Arabidopsis thaliana. *Plant Cell Physiol.* 2009, 50, 1425–1438. [CrossRef] [PubMed]

62. Pan, Z.J.; Chen, Y.Y.; Du, J.S.; Chen, Y.Y.; Chung, M.C.; Tsai, W.C.; Wang, C.; Chen, H. Flower development of Phalaenopsis orchid involves functionally divergent SEPALLATA-like genes. *New Phytol.* 2014, 202, 1024–1042. [CrossRef] [PubMed]

63. Davis, S.J. Integrating hormones into the floral-transition pathway of Arabidopsis thaliana. *Plant Cell Physiol.* 2009, 50, 1201–1210. [CrossRef] [PubMed]

64. Goto, N.; Pharis, R.P. Role of gibberellins in the development of floral organs of the gibberellin-deficient mutant, GA1-1, of Arabidopsis thaliana. *Botany* 1999, 77, 944–954. [CrossRef]

65. Woodward, A.W.; Bartel, B. Auxin: Regulation, action, and interaction. *Ann. Bot.* 2005, 95, 707–735. [CrossRef] [PubMed]

66. Chmelnitsky, I.; Colauzzi, M.; Algorn, R.; Zieslin, N. Effects of temperature on phyllody expression and cytokinin content in floral organs of rose flowers. *Plant Growth Regul.* 2001, 35, 207–214. [CrossRef]

67. Ma, N.; Chen, W.; Fan, T.; Tian, Y.; Zhang, S.; Zeng, D.; Li, Y. Low temperature-induced DNA hypermethylation attenuates expression of RhAG, an AGAMOUS homolog, and increases petal number in rose (*Rosa hybridra*). *BMC Plant Biol.* 2015, 15, 237. [CrossRef] [PubMed]

68. Jacobsen, S.E.; Sakai, H.; Finnegan, E.J.; Cao, X.; Meyerowitz, E.M. Ectopic hypermethylation of flower-specific genes in Arabidopsis. *Curr. Biol.* 2000, 10, 179–186. [CrossRef]

69. Zhang, K.; Wang, R.; Zi, H.; Li, Y.; Cao, X.; Li, D.; Guo, L.; Tong, J.; Fan, Y.; Jiao, Y. AUXIN RESPONSE FACTOR3 regulates floral meristem determinacy by repressing cytokinin biosynthesis and signaling. *Plant Cell* 2018, 30, 324–346. [CrossRef] [PubMed]

70. Wang, H.; Hu, B.; Chen, G.; Shi, N.; Zhao, Y.; Yin, Q.; Liu, J. Application of Arabidopsis AGAMOUS second intron for the engineered ablation of flower development in transgenic tobacco. *Plant Cell Rep.* 2008, 27, 251–259. [CrossRef] [PubMed]

71. Fan, S.; Zhang, D.; Zhang, L.; Gao, C.; Xin, M.; Tahir, M.M.; Li, Y.; Ma, J.; Han, M. Comprehensive analysis of GASA family members in the *Malus domestica* genome: Identification, characterization, and their expressions in response to apple flower induction. *BMC Genom.* 2017, 18, 827. [CrossRef] [PubMed]

72. Xing, L.-B.; Zhang, D.; Li, Y.-M.; Shen, Y.-W.; Zhao, C.-P.; Ma, J.-J.; An, N.; Han, M.-Y. Transcription profiles reveal sugar and hormone signaling pathways mediating flower induction in apple (*Malus domestica* Borkh.). *Plant Cell Physiol.* 2015, 56, 2052–2068. [CrossRef] [PubMed]