Comparative genomic analysis of the IDD genes in five Rosaceae species and expression analysis in Chinese white pear (*Pyrus bretschneideri*) (#31623)

First revision

Editor guidance

Please submit by **15 Jan 2019** for the benefit of the authors (and your $200 publishing discount).

Structure and Criteria
Please read the 'Structure and Criteria' page for general guidance.

Author notes
Have you read the author notes on the guidance page?

Raw data check
Review the raw data. Download from the materials page.

Image check
Check that figures and images have not been inappropriately manipulated.

Privacy reminder: If uploading an annotated PDF, remove identifiable information to remain anonymous.

Files

Download and review all files from the materials page.

1 Tracked changes manuscript(s)
1 Rebuttal letter(s)
17 Figure file(s)
9 Table file(s)
1 Other file(s)

Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

When ready submit online.

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see PeerJ policy).

EXPERIMENTAL DESIGN

- Original primary research within Scope of the journal.
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. Negative/inconclusive results accepted. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- Data is robust, statistically sound, & controlled.
- Speculation is welcome, but should be identified as such.
- Conclusions are well stated, linked to original research question & limited to supporting results.

Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

Comparative genomic analysis of the IDD genes in five Rosaceae species and expression analysis in Chinese white pear (*Pyrus bretschneideri*)

Xueqiang Su [1], **Tiankai Meng** [2], **Yu Zhao** [1], **Guohui Li** [1], **Xi Cheng** [1], **Muhammad Abdullah** [1], **Xu Sun** [1], **Yongping Cai** [Corresp., 1], **Yi Lin** [Corresp. 1]

[1] School of Life Science, Anhui Agricultural University, Hefei, China

[2] School of Life Sciences and Technology, TongJi University, Shanghai, China

Corresponding Authors: Yongping Cai, Yi Lin
Email address: ypcaiah@163.com, linyi992547404@163.com

The INDETERMINATE DOMAIN (IDD) gene family encodes hybrid transcription factors with distinct zinc finger motifs and appears to be found in all higher plant genomes. In the model plants *Arabidopsis thaliana* and *Oryza sativa*, IDD genes have been identified throughout the genome, and the functions of many members have been studied. However, few studies have been done on the IDD gene family in Rosaceae species (only *Malus domestica* has a completed genome-wide identification of the IDD gene family). This study focuses on a comparative genomic analysis of the IDD gene family in five Rosaceae species (*Pyrus bretschneideri*, *Fragaria vesca*, *Prunus mume*, *Rubus occidentalis*, and *Prunus avium*). We identified a total of 68 IDD genes: 16 genes in Chinese white *Pyrus bretschneideri*, 14 genes in *Fragaria vesca*, 13 genes in *Prunus mume*, 14 genes in *Rubus occidentalis* and 11 genes in *Prunus avium*. The evolution of the IDD genes of these five Rosaceae species was revealed by constructing a phylogenetic tree, tracking gene duplication events, and performing a sliding window analysis and a conserved microsynteny analysis. The expression analysis of different tissues showed that most of the pear IDD genes had a very high transcription level in fruit, flower, and bud. Combining our results with previous research, we speculate that *PbIDD2* and *PbIDD8* may participate in flowering induction in pear. Temporal expression analysis showed that the expression patterns of *PbIDD3* and *PbIDD5* were completely opposite to the accumulation pattern of fruit lignin and stone cell content. Combining the results of the composite phylogenetic tree and expression pattern analysis showed that *PbIDD3* and *PbIDD5* might be involved in the metabolism of lignin and secondary cell wall (SCW) formation. In summary, we provide basic information about the IDD genes of five Rosaceae species, thus providing a theoretical basis for studying the function of these IDD genes.

Comparative genomic analysis of the IDD genes in five Rosaceae species and expression analysis in Chinese white pear (*Pyrus bretschneideri*)

Xueqiang Su[#], Tiankai Meng[#], Yu Zhao, Guohui Li, Xi Cheng, Muhammad Abdullah, Xu Sun, Yongping Cai*, Yi Lin*

1 School of Life Science, Anhui Agricultural University, Hefei, Anhui, China

2 School of Life Sciences and Technology, TongJi University, Shanghai, China

# This authors have contributed equally to this work.

*Corresponding author: ypcaiah@163.com (Yongping Cai);

Co-Corresponding author: linyi1957@126.com (Yi Lin);

Abstract

The INDETERMINATE DOMAIN (IDD) gene family encodes hybrid transcription factors with distinct zinc finger motifs and appears to be found in all higher plant genomes. In the model plants *Arabidopsis thaliana* and *Oryza sativa*, IDD genes have been identified throughout the genome, and the functions of many members have been studied. However, few studies have been done on the IDD gene family in Rosaceae species (only *Malus domestica* has a completed genome-wide identification of the IDD gene family). This study focuses on a comparative genomic analysis of the IDD gene family in five Rosaceae species (*Pyrus bretschneideri*, *Fragaria vesca*, *Prunus mume*, *Rubus occidentalis*, and *Prunus avium*). We identified a total of 68 IDD genes: 16 genes in Chinese white *Pyrus bretschneideri*, 14 genes in *Fragaria vesca*, 13 genes in *Prunus mume*, 14 genes in *Rubus occidentalis* and 11 genes in *Prunus avium*. The evolution of the IDD genes of these five Rosaceae species was revealed by constructing a phylogenetic tree, tracking gene duplication events, and performing a sliding window analysis and a conserved microsynteny analysis. The expression analysis of different tissues showed that most of the pear IDD genes had a very high transcription level in fruit, flower, and bud. Combining our results with previous research, we speculate that *PbIDD2* and *PbIDD8* may participate in flowering induction in pear. Temporal expression analysis showed that the expression patterns of *PbIDD3* and *PbIDD5* were completely opposite to the accumulation pattern of fruit lignin and stone cell content. Combining the results of the composite phylogenetic tree and expression pattern analysis showed that *PbIDD3* and *PbIDD5* might be involved in the metabolism of lignin and secondary cell wall (SCW) formation. In summary, we provide basic information about the IDD genes of five Rosaceae species, thus providing a theoretical basis for studying the function of these IDD genes.

INTRODUCTION

Zinc finger proteins are transcription factors with a finger-like domain. They are widely distributed in animals and microorganisms, as well as the plant kingdom (Miller et al., 1985). The zinc finger proteins found in all plant eukaryotic lineages,

which means they may be derived from a common eukaryotic ancestor. Zinc finger proteins have been identified and functionally analysed in several plant species, such as *Jatropha curcas* (Shi et al., 2018), *Oryza sativa* (Zhang et al., 2018), *Musa acuminata* (Chen et al., 2014), *Gossypium hirsutum* (Ren et al., 2018). One group of this large family of proteins, the INDETERMINATE DOMAIN (IDD) proteins, have a highly conserved ID domain (Colasanti et al., 1998), which has typical C2H2 and C2HC zinc finger motifs (Wu et al., 2008). C2H2 zinc finger transcription factors are one of the most thoroughly studied transcription factor families (Agarwal et al., 2007; Wei et al., 2016). The C2H2 zinc finger transcription factors contain tandem repeat segments of approximately 30 amino acids, all of which have a highly conserved amino acid sequence: (F/Y)-XC-X2-5-C-X3-(F/Y)-X5-psi-X2-H-X3-5-H (wherein C and H represent cysteine and histidine, respectively, X means any amino acid at this position, and psi means a hydrophobic residue) (Parraga et al., 1988). This particular sequence structure can bind to Zn^{2+} and fold to form a structure with two β-sheets and an α-helix. Zn^{2+} mainly plays a role in linking individual amino acid chains and is also the key to the function of zinc finger protein (Islam et al., 2009; Tian et al., 2010).

The INDETERMINATE DOMAIN (IDD) protein family is a special type of C2H2 zinc finger subgroup. These proteins have a highly conserved ID domain with an N-terminal nuclear localization signal (NLS) and four distinct zinc finger motifs (ZF1, ZF2, ZF3, and ZF4) (Colasanti et al., 2006; Kozaki et al., 2004). In addition to having two canonical zinc fingers in the IDD gene, these proteins also have two unusual CCHC fingers, one of which is also associated with the zinc finger domain in the *Saccharomyces cerevisiae* SW15 protein (Kozaki et al., 2004). Among the species having a complete genome-wide identification of the IDD gene family, *Malus domestica* has a higher number of IDD genes than *Arabidopsis thaliana*, *Oryza sativa*, and *Zea mays* (Colasanti et al., 2006; Fan et al., 2017). Thus far, the functions of some IDD genes have been identified, mainly those from *Arabidopsis thaliana* (IDD genes with similar functions have also been reported in *Oryza sativa* and *Zea mays*). The function of the IDD gene involves the seed germination (*AtIDD1*, *2*, *ZmIDD9*, *ZmIDDveg9*) (Feurtado et al., 2011; Yi et al., 2015), leaf and flower development (*AtIDD4*, *14*, *15*, *16*) (Cui et al., 2013), flowering regulation (*ZmID1*, *AtIDD8*, *OsID1*) (Colasanti et al., 1998; Seo et al., 2011; Park et al., 2008), starch metabolism (*AtIDD5*) (Ingkasuwan et al., 2012), the regulation of plant gravitropism (*AtIDD14*, *15*, *16*, *LPA1*, *OsIDD16*, *18*), root development and nitrogen metabolism (*AtIDD3*, *8*, *9*, *10*) (Jeong et al., 2015; Coelho et al., 2018).

Pear, belonging to the Rosaceae family, is widely cultivated around the world and is popular because of its unique flavour. There is a long history of eating pears in China, and the variety of pear cultivars grown is very wide. 'Dangshan Su' pear (*Pyrus bretschneideri* cv. Dangshan Su) is a white pear cultivar, a characteristic pear type produced in Dangshan County, Anhui Province, China. Because of its good quality and high medicinal efficacy, the fruit has a very high market value (Hamauzu et al., 2007; Huang et al., 2010; Zhang et al., 2017). The stone cells in the pear flesh are an important source of this pear's unique taste, and the content of stone cells is a key factor in determining the quality of pear fruit (Cai et al., 2010; Yan et al., 2014). However, there are few reports on fruit development; the relationship between the IDD gene and fruit development is only reported for *Musa acuminata* (Chen et al., 2014). In *Musa acuminata*, the expression of MaIDD is related to development and to fruit ripening. In *Malus domestica*, the IDD gene also appears to play a role in regulating flower induction. In addition, we also found that some zinc finger proteins are involved in plant SCW formation. For example, the CCCH-tandem zinc finger protein IIP4 (Zhang et al., 2018), *Arabidopsis thaliana* C3H14, C3H15 (Chai et al., 2015) and *OsIDD2* in *Oryza sativa* (Huang et al., 2018). Although the function of several IDD genes has been studied, it remains unknown whether the IDD gene participates in regulating flower induction, fruit development, lignin biosynthesis, and SCW formation in *Pyrus bretschneideri*, although their basic function can be illustrated by the evolution of pear IDD family genes.

Among Rosaceae species, genome-wide identification of the IDD gene was performed only in *Malus domestica* (Fan et al., 2017). Currently, the genome-wide identification of this gene family has not been reported in *Pyrus bretschneideri*. The function

of these IDD genes in different tissues of Chinese white pear (*Pyrus bretschneideri*) is not clear. In this study, we identified the IDD genes of five Rosaceae Plants and performed a comparative genomic analysis. Woodland strawberry (*Fragaria vesca*) is a perennial herb with edible fruits and grows well in most parts of the Northern Hemisphere. Genome-wide sequencing was first reported in 2011 (Shulaev et al., 2011). Japanese apricot (*Prunus mume*) has great ornamental value, and in some Asian countries, the fruit of this tree are used for cooking and flavouring. *S*weet cherry (*Prunus avium*), which originates in Europe, is an important economic tree species (Tavaud et al., 2004). A new report on the de-genomic sequence of *Prunus avium* was reported in 2017 (Shirasawa et al., 2017). Finally, black raspberry (*Rubus occidentalis* L.) is a specialty fruit crop that is grown in the Pacific Northwest of the United States, with a unique flavour and rich nutritional value. Whole-genome sequencing work for this species was reported in 2017 (VanBuren et al., 2017). These four species belong to the same family and have a close evolutionary relationship. We attempted to understand the evolution of the pear IDD gene family by comparative genomic analysis to further predict and analyse the function of the Chinese white pear IDD gene family members.

MATERIALS AND METHODS

Identification of IDD Genes in Five Rosaceae Plants

In this study, the Chinese white pear genome database was obtained from the Pear Genome Project. (http://peargenome.njau.edu.cn/). The genome databases of three Rosaceae plants (*Fragaria vesca*, *Rubus occidentalis*, *Prunus avium*) were downloaded from GDR (https://www.rosaceae.org/). The genome sequence data of *Prunus mume* is downloaded from the *P. mume* genome project (http://prunusmumegenome.bjfu.edu.cn/). DNATOOLS software was used to create a local database containing the amino acid sequences of the five Rosaceae plants' IDD genes (Cao et al., 2016a). To identify the IDD genes of these five Rosaceae species, we used the following methods. First, the *Arabidopsis thaliana* (16) and *Malus domestica* (20) IDD gene sequences were collected for use as query sequences in TBlastN with the default E-value (Supplementary Table S1). We compared these sequences with the local database sequences of the above five Rosaceae species. Then, the IDD candidate gene sequences initially screened by BLAST were tested using SMART online software to test whether they contained a zinc finger domain (http://smart.embl-heidelberg.de/) (Letunic et al., 2012). Finally, each candidate sequence containing a zinc finger domain was manually checked for an IDD domain. We discarded the protein sequences lacking a full IDD domain and redundant sequences. Finally, we isolated the candidate IDD gene sequences. The ExPASy online site was used for investigating the molecular weights of IDD proteins (http://web.expasy.org/protparam/) (Artimo et al., 2012). Prediction of subcellular localization results for all IDD proteins was performed on the WoLFPSORT website (http://www.g enscript.com/wolf-psort.html) (Horton et al., 2007).

Phylogenetic Analysis

Sequence alignment of all IDD proteins was done using the ClustalW tool in MEGA 7.0 software. The phylogenetic tree was constructed using MEGA 7.0 software using the maximum likelihood (ML) (bootstrap = 1000) (Kumar et al., 2016). The *Malus domestica* IDD gene sequence used in phylogenetic tree was derived from the results of Fan et al (Fan et al., 2017). The sequences of *Arabidopsis thaliana*, *Oryza sativa* and *Zea mays* were obtained from the article by Colasanti (Colasanti et al., 2006).

IDD gene Structures and Conserved Motif Prediction

IDD gene structures were compared using Gene Structure Server (http://gsds.cbi.pku.edu.cn) (Guo et al., 2007). The motifs

of the IDD genes in these five Rosaceae species were analysed using MEME online analysis software (http://meme.sdsc.edu/meme4_3_0/intro.html) (Bailey et al., 2015). Parameters for the conserved motif prediction were motif width greater than 6 and less than 200. The number of identified motifs was 20.

Chromosomal Location, Gene Duplication and Ka/Ks Ratio Analysis

Five Rosaceae species chromosome start positions and other relevant information about the IDD genes were obtained from the public genome database. The chromosomal physical locations of the IDD genes of all five Rosaceae species were mapped using MapInspect software (http://mapinspect.software.informer.com) (Hu et al., 2011; Lin et al., 2011; Zhu et al., 2015). To define gene duplication events, we mainly relied on the following principles: 1, The similarity of the two gene-coding sequences (CDS) was more than 80%. 2, If the two genes were located on the same chromosome and separated by at least 200 kb, we considered these two genes tandem-duplicated genes. 3, If the two genes were located on different chromosomes, they were called segmentally duplicated genes (Long & Thornton, 2001). DnasP v5.0 software was used to calculate non-synonymous (Ka) and synonymous substitution (Ks) values and perform sliding window analysis. Parameters for sliding window analysis were window size 150 bp and step size 9 bp (Librado & Rozas, 2009).

Microsynteny analysis and Chinese white pear IDD gene promoter *Cis*-acting element analysis

We obtained the promoter sequence of each IDD gene from the Chinese white pear genome database, including the DNA sequence of the initiation codon (ATG) located 1500 bp upstream of each IDD gene. The online software Plantcare was used to analyse the cis-acting elements of the promoter region (http://bioinformatics.psb.ugent.be/webtools/plantcarere/html/) (Rombauts et al., 1999). The microsynteny of 6 Rosaceae species (we added the IDD genes of *Malus domestica*) was identified using the Multiple Collinearity Scan toolkit (MCScanX) (Abdullah et al., 2018).

RNA extraction and qRT-PCR

The plant material used in this experiment was the 'Dangshan Su' pear, which grows in the Yuanyichang agricultural park (Dangshan County, Anhui Province, China). 'Dangshan Su' pear belongs to the white pear cultivar. In April 2017, we began collecting materials of various tissues of pear. The fruits were picked at 15 days after pollination (DAP), 39DAP, 47DAP, 55DAP, 63DAP, 79DAP and 145DAP. Treatment buds of 'Dangshan Su' pear tissue-cultured seedlings with Gibberellin (GA) (700 mg•L^{-1}) and sucrose (20,000 mg•L^{-1}) (Fan et al., 2017). Samples collected were selected for 0 h post-treatment (HPT), 2HPT, 4HPT, 6HPT, 8HPT, 12HPT, 24HPT. RNA was extracted from plant material (including fruits and various tissues) using the Tiangen (Beijing, China) plant RNA extraction kit. Reverse transcription was performed using a PrimeScriptTM RT reagent kit with gDNA Eraser (Takara, Japan). Each reaction used 1 μg of RNA. The qRT-PCR primers for the pear IDD genes were designed using Beacon Designer 7 software (Supplementary Table S8). The qRT-PCR system consisted of 20 μL: 10 μL of SYBR® Premix Ex Taq™ II (2×) (Takara, Japan), 2 μL of cDNA and 0.8 μL of PbIDD-F and PbIDD-R. The reaction procedure followed the instruction manual, and it was run in three repetitions. We used the $2^{-\Delta\Delta Ct}$ method for the calculation of relative expression (Livak & Schmittgen, 2001).

RESULTS

1, Identification, Characterization and Phylogenetic Analysis of IDD Genes

A total of 68 IDD proteins were identified and used for further analysis (Table 1; Supplementary Table S1). In Chinese

white pear, we identified 16 IDD proteins (*PbIDD1-PbIDD16*). We identified a total of 52 IDD proteins in the other four species, including 14 in *Fragaria vesca* (*FvIDD1-FvIDD14*), 13 in *Prunus mume* (*PmIDD1-PmIDD13*), 14 in *Rubus occidentalis* (*RoIDD1-RoIDD14*), and 11 in *Prunus avium* (*PaIDD1-PaIDD11*). Although only 11 IDD proteins were identified in *Prunus avium*, the numbers of IDD proteins in the other species were very similar but fewer than the number of IDD proteins in *Malus domestica* (20). We used ExPASy software to calculate the physicochemical parameters of the IDD gene (Supplementary Table S2). We identified 68 IDD genes with isoelectric points (pIs) greater than 7. The lowest pI value is 7.76 (*PaIDD8*), while the highest pI value is 9.45 (*FvIDD4*) in these Rosaceae species. The molecular weights of these IDD genes are quite different, ranging from 42.0 kDa to 79.6 kDa for *PaIDD10* and *PmIDD8*. The number of amino acids in the IDD proteins is also quite different. *PbIDD14* contains the fewest (only 381) amino acids, while *PmIDD8* has the largest number (728). The grand average of hydropathicity (GRAVY) of all 68 IDD genes was less than 1, and the GRAVY value of *PbIDD15* was the smallest (-0.914).

A phylogenetic tree of a total of 98 IDD proteins was built using the maximum likelihood (ML) method (Figure 1). Based on the phylogenetic tree, we identified four phylogenetic groups (groups 1-4). Group 1 had the most members, containing all six Rosaceae species and 44 IDD genes: *Pyrus bretschneideri* (8), *Fragaria vesca* (7), *Prunus mume* (6), *Rubus occidentalis* (7), *Prunus avium* (6), and *Malus domestica* (10). In groups 3 and 2, we identified 21 and 14 IDD genes, respectively. Group 4 had the fewest members; in *Pyrus bretschneideri*, *Fragaria vesca*, *Prunus mume*, and *Rubus occidentalis,* only two IDD genes were found (in each); 1 IDD gene was found in *Prunus avium*; and no IDD gene was found in *Malus domestica*. *ZmID1* was grouped into a class by itself. We found that most *PbIDD* genes were more tightly grouped with *Malus domestica*. In addition, the IDD genes of some Rosaceae species were closely related to those of *Arabidopsis thaliana* and *Oryza sativa*; *PbIDD3*, *PbIDD5*, *FvIDD3*, *RoIDD4*, *PmIDD3*, and *PaIDD4* were included in group 1 and were closely related to *OsIDD2*; *FvIDD1*2, *RoIDD9*, *PmIDD11*, *PaIDD5* and *AtIDD3*, *8* were closer; and *AtIDD1* was in group 3, clustered with 7 Rosaceae IDD genes.

2, Structural and Conserved Motif Analysis of IDD Proteins

To more comprehensively analyse the structural diversity of the five Rosaceae species' IDD genes, we created exon-intron pattern maps of all 68 IDD genes (Figure 2). Of the 34 IDD genes in group 1, 25 genes had 3 exons, and 6 genes had two exons. Additionally, *FvIDD7* (4), *RoIDD6* (5), *PmIDD8* (6) had a large number of exons. In group 2, *FvIDD11* had two exons, *PbIDD7* had four exons, and all other members had three exons. Group 3 was a subfamily with a complex gene structure: eleven genes had four exons, and four genes had three exons. All members of group 4 had a gene structure comprising 3 exons. The lack of high similarity between the 68 IDD genes allowed us to better understand the conserved motifs of these IDD genes. We used MEME software to identify 20 conserved motifs (Figure 2; Supplementary Table S3). Motif 2, motif 3, motif 1 and motif 4 were identified to encode the ID domain, while the remaining motifs did not have functional annotations. As shown in figure 3, motif 2, motif 3, motif 1 and motif 4 represented zinc fingers ZF1, ZF2, ZF3, and ZF4, respectively. The four conserved motifs constituted a conserved ID domain, while these four conserved motifs represented two C2H2 and two C2HC structures (ZF1 and ZF2 belonged to C2H2, and ZF3 and ZF4 belonged to C2HC). Two C2H2 and two C2HC structures were identified by sequence alignment of the conserved ID domains of five species and the results showed that the ID domain was highly conserved. Cysteine (C) and histidine (H) in each zinc finger domain were conserved in each species, consistent with previous studies (Supplementary Figure S1-5). All 68 IDD genes had motifs 1-4. These four motifs were thus the most conserved. At the same time, we found that some members of the same group and the more closely related members had highly similar motif compositions (e.g., *FvIDD7*, *RoIDD6* in group 1 and *PmIDD10*, *PaIDD9* in group 2). These genes were closely related and had exactly the same motif compositions. We also identified certain group-specific motifs, such as motif 12, motif 17, motif 20 in members of group 4 and motif 19 in some members of group 1.

We also identified nuclear localization signal (NLS) in the N-terminal border of these IDD genes (Supplementary Figure 6). According to previous studies, this NLS is usually KKKR or KRKR (Colasanti et al., 2006). Of the five Rosaceae species, not all members have NLS. For example: *PbIDD13*, *14*; *FvIDD10*, *11*, *12*; *PmIDD11*, etc. In addition to the conserved IDD domain, we also found two C-terminal domains (MSATALLQKAA and TRDFLG) in some members (Figure 2; Supplementary Figure S1-5). These two C-terminal domains are encoded by motif6 and motif7 respectively. All 9 members of group 4 lacked both motif 6 and motif 7. In group 1, *PaIDD10* did not have motif 6, and *PbIDD14* and *PmIDD8* lacked motif 7. All members of group 2, 3 contained these two motifs. The lack of these two conserved sequences at the C-terminus of some IDD genes may be the reason that the members of group 4 were clustered into a single category and may also underlie differences in the functions of these proteins (Figure 1).

3, Chromosomal Location and Duplication Events of IDD Family Genes in Rosaceae

Based on the complete genome sequences of the five Rosaceae species, the exact chromosomal locations of all 68 IDD genes were identified (Figure 4). In *Pyrus bretschneideri*, the 16 IDD genes were distributed on chromosomes 3, 4, 8, 9, 11, 12, 14, 15, 16, 17. In *Fragaria vesca*, the IDD genes were distributed on chromosomes 1, 2, 3, 4, 5, 6. In *Rubus occidentalis*, the IDD genes were mainly distributed on chromosomes 2, 4 and 6. Several other IDD genes were distributed on chromosomes 1, 3 and 5. In *Prunus mume*, except *PaIDD5*, which could not be mapped to any chromosome, the IDD genes were distributed on chromosomes 1, 2, 4, 6, 7, 8. However, in *Prunus avium*, 4 IDD genes were distributed on chromosome 1 and the remaining 7 IDD genes were distributed on chromosomes 3, 5, 6, 7, 8. Only 4 pairs of duplicated genes were identified in pear (Figure 4). All duplicated genes identified in *Pyrus bretschneideri* were segmental duplications.

To clarify the driving forces of gene duplication and explore the impact of these genes on evolutionary processes, we calculated the Ka, Ks and Ka/Ks ratio for the 4 duplicated gene pairs in *Pyrus bretschneideri*. Ka/Ks=1 is the cut-off value that indicates neutral selection, Ka/Ks<1 represents negative selection and Ka/Ks>1 represents positive selection (Bitocchi et al., 2017). For all four duplicated gene pairs identified in *Pyrus bretschneideri*, the Ka/Ks values were less than 0.2726 (Supplementary Table S4). In the evolutionary processes of genes, positive selection may be overshadowed by strong negative selection (Han et al., 2016). Therefore, to comprehensively explain the selection pressure of IDD genes, we performed a sliding window analysis of the 4 pairs of IDD paralogues in pears (Supplementary Figure S7). Most coding site Ka/Ks ratios were less than one, with exceptions for one or several distinct peaks (Ka/Ks > 1). The Ka/Ks ratios of the conserved IDD domains were less than 1.

4, Microsynteny and Evolutionary Analysis of the IDD Gene Family

To identify interspecific microsynteny, we visualized the location of homologous and orthologous genes. A total of 30 orthologous gene pairs were identified in five Rosaceae species (Figure 5; Supplementary Table S5). These include 1, 8, 3, 6, 12 collinear gene pair identified between *Pyrus bretschneideri* and *Fragaria vesca*, *Prunus mume*, *Rubus occidentalis*, *Prunus avium*, *Malus domestica*, respectively. It is worth noting that *PbIDD2* also had a collinear relationship with the members of the *MdIDD*, *PmIDD,* and *PaIDD* families. On the other hand, two IDD genes from pear matched with one pair of *Prunus mume* or *Rubus occidentalis* genes. For example, *PbIDD9* and *PbIDD10* were orthologous to *PmIDD9*, while *PbIDD9* and *PbIDD10* were to orthologous to *RoIDD7*.

However, we did not identify any orthologous gene pair between *Pyrus bretschneideri* and *Oryza sativa*, *Zea mays*, *Arabidopsis thaliana*. In addition, we compared the other four Rosaceae species with these three species, and the results did not identify any orthologous gene pair. To further study the evolution of IDD genes, we collected IDD genes from four

monocotyledons (*Musa acuminata*, *Sorghum bicolor*, *Zea mays*, *Oryza sativa*) and two dicotyledons (*Vitis vinifera*, *Arabidopsis thaliana*). A phylogenetic tree of the 12 species genomes used for evolutionary analysis of IDD gene family (Supplementary Table S8A). Several Rosaceae species have a closer relationship and are more distant from monocots. This result is basically consistent with the results of the phylogenetic tree constructed by all IDD genes of 12 species (Supplementary Table S8B). We have identified 12 classes (classI-XII). *Pyrus bretschneideri* and *Malus domestica*, *Prunus avium* and *Prunus mume* IDD genes often cluster into one class. The IDD genes of monocotyledons are closely related, such as classX and classXI, which are only composed of IDD genes of monocotyledons. *ZmID1*, *OsID1*, *SbID1* are closely related, but we have not found orthologous gene of ID1 gene in several other species. In class III, we identified all species except *Malus domestica*. The IDD genes in this subfamily did not have complete MSATALLQKAA and TRDFLG domains.

5, Analysis of *Cis*-acting Elements in the Promoters of the IDD Genes in Chinese White Pear

Plant IDD genes may be involved in multiple biological processes, and hormones affect their expression. We predicted the cis-acting elements of 16 Chinese white pear IDD gene promoters (Figure 6; Supplementary Table S7). Thirteen IDD genes (except *PbIDD7*, *9*, *10*) contained at least one G-Box, which is a light-responsive element, indicating that the expression of these genes may be regulated by light. Eight *Pyrus bretschneideri* IDD genes had HSE, and 11 members contained MBS components. TC-rich repeats, which are cis-acting elements involved in defence and stress responsiveness, were identified in PbIDD1, 6, 7, 9, 10, 12, 14, and 16. In addition, there were many cis-acting elements related to the responses to hormones, including responses to abscisic acid (ABRE), methyl jasmonate (MeJA) (CGTCA-motif, TGACG-motif), gibberellin (P-box, GARE-motif), auxin (TGA-element), and salicylic acid (TCA-element). The auxin-related TGA-element was identified in only five members (*PbIDD1*, *2*, *4*, *8*, 13), which was the least number of elements components found. A total of 41 *C*is-acting elements associated with MeJA were found in 9 members, and these were the most abundant *C*is-acting elements. *C*is-acting elements responding to gibberellin were the most widespread, and twelve of the *Pyrus bretschneideri* 16 IDD genes had this element. The related *C*is-acting elements of abscisic acid and salicylic acid were identified in 10 and 9 members, respectively. We found that *PbIDD1* has *C*is-acting elements related to the above five hormones, whereas *PbIDD9* only has cis-acting elements that are responsive to MeJA. Of the 16 pear IDD genes, 12 contained 19 CGTCA-motifs, and TCA-elements appeared in the promoter regions of 10 members, for a total of 22 genes. In addition, the CAT-box and the CCGTCC-box were found in 11 members (*PbIDD1*, *2*, *3*, *4*, *5*, *6*, *7*, *8*, *9*, *12*, 14); these are *C*is-acting elements related to meristem expression.

6, Expression Characteristics of Pear IDD Genes

To deepen our understanding of the Chinese white pear IDD gene family, we investigated the expression patterns of *PbIDDs* in different tissues (Figure 7). Among the 16 members, *PbIDD4*, *9*, *11* were highly expressed in bud; however, their expression was extremely low in other tissues. *PbIDD1* and *10* were mainly expressed in flower. *PbIDD2*, *6*, *8, 15* were mainly expressed in fruit, flower, and bud. Among them, *PbIDD2* and *8* had the highest expression levels in bud, while *PbIDD6* and *15* had the highest expression levels in fruit. *PbIDD5*, *12*, *16* were detected in multiple tissues. The main difference was found in the expression levels in leaves: *PbIDD12* was highly expressed in leaves, while the expression level of *PbIDD5* in leaves was the lowest among the 5 tissues in which it is expressed. *PbIDD16* had almost no expression in the leaves. *PbIDD3*, *7*, *13* were mainly expressed in the fruit of Chinese white pear, and *PbIDD3* and *7* showed almost no expression in other tissues.

We investigated the expression patterns of 16 IDD genes at 7 developmental stages of Chinese white pear fruit (Figure 7). No expression of *PbIDD14* was detected in all tissues of Chinese white pear. The expression level of *PbIDD1* peaked at 39DAP and was low during the other six fruit developmental stages. The peaks of *PbIDD2*, *7*, *8*, *11*, *13*, and *16* expression also appeared at 39DAP, but high expression was also detected in the early developmental stages (15, 47DAP). In the late developmental stages

of fruit (79, 145DAP), these genes were expressed at a medium level. *PbIDD6*, *15* had a similar expression pattern to the above genes, except that *PbIDD6*, *15* were detected in the early developmental stages of fruit but were expressed at extremely low levels in the late developmental stages of fruit. *PbIDD9* had a high transcription level at 63, 145DAP and *PbIDD12* is mainly expressed at 79DAP. *PbIDD10* had two expression peaks, one at 39DAP and another at 55DAP. *PbIDD3*, *5* had the same expression pattern; these two genes are mainly expressed in the early stage of fruit development; thereafter, the expression level is gradually decreased.

7, Gibberellin and Sugar Response Pattern Analysis of *PbIDDs*

Chinese white pear *PbIDD2, 5, 6, 8, 9, 12, 16* are mainly expressed in bud and flower. We treated the bud of Chinese white pear with exogenous GA and sucrose to investigate the expression pattern of *PbIDD2*, *5*, *6*, *8*, *9*, *12*, *16* under GA and sucrose treatment (Supplementary figure S9). Under exogenous GA treatment, *PbIDD2*, *5*, *6*, *8*, *9*, *12*, *16* showed three different expression patterns (activation, inhibition and invariance). There was no significant change in the transcriptional level of *PbIDD5*, *9* under exogenous GA treatment. The expression levels of *PbIDD2*, *8* continued to decrease after exogenous GA treatment, indicating that the expression of *PbIDD2*, *8* was inhibited by GA. *PbIDD6*, *12* showed a significant increase in expression after GA treatment (reaching a peak at 3HPT); thereafter, the expression level gradually decreased to untreated levels. *PbIDD16* reached the maximal expression levels at 2, 4HPT. Overall, the expression of *PbIDD6*, *12*, *16* showed activation by GA. The transcription levels of *PbIDD2*, *8* increased gradually after sucrose treatment in Chinese white pear bud. The expressions of *PbIDD5*, *12*, *16* were not always induced by sucrose. The lowest expression levels of *PbIDD5*, *12* were found at 4HTP, while the lowest expression level of *PbIDD16* was observed at 1HTP. The expressions of *PbIDD6*, *9* were inhibited by sucrose.

DISCUSSION

IDD gene family encode hybrid transcription factors that have four zinc finger structures and one nuclear positioning signal. IDD proteins play important roles in regulating plant flowering, development, stress resistance, secondary metabolism, and other processes (Wong & Colasanti, 2007; Matsubara et al., 2008). The plant IDD gene can also regulate the expression of the key enzyme genes of lignin synthesis, which affects the biosynthesis of lignin and regulates the biosynthesis of SCW formation (Huang et al., 2018).

In this study, we identified 68 IDD genes in five Rosaceae species (Table 1; Supplementary Table S2). All IDD protein pI values were greater than 7, indicating that all these proteins are alkaline proteins. The proteins were all hydrophilic (their GRAVY values were less than 1). Among the Rosaceae species, *Pyrus bretschneideri* had the second greatest number of IDD gene family members after *Malus domestica* (Fan et al., 2017). *Pyrus bretschneideri* underwent two whole-genome duplications (WGDs) during its evolution, and this amplified the chromosome number from 9 to 17 (Velasco et al., 2010; Cao et al., 2017); the other four species experienced only one old WGDs. We also identified 4 pairs of duplicated genes in *Pyrus bretschneideri* and did not identify gene duplication events in the four other species. Therefore, the number of IDD genes in *Pyrus bretschneideri* is greater than that in the other four Rosaceae species; this may be the result of the interaction of recent WGD and gene duplication events. However, *Pyrus bretschneideri* and *Malus domestica* both experienced two WGDs and had 4 pairs of duplicated genes, while *Malus domestica* has more IDD genes than *Pyrus bretschneideri*. This difference may be traced back to the origin of the IDD gene family, during which time the numbers of IDD genes in *Pyrus bretschneideri* and *Malus domestica* were different. According to the phylogenetic tree, 68 IDD genes were divided into four groups (Figure 1). The *ZmID1* gene clustered into its own class in the phylogenetic tree. No homologous genes were found in *Arabidopsis thaliana* and the 6

Rosaceae species, suggesting that the ID1 gene may be unique to gramineous plants. Interestingly, five Rosaceae species had at least one IDD gene present in each clade and had a strong bootstrap. These results show that the rapid duplication of IDD genes occurred before these Rosaceae species diverged.

The gene structure and conserved motif composition may be closely related to the diversity of gene functions (Swarbreck et al., 2008). Genes in the same subfamily tend to have very similar gene structures, reflecting that these genes may have similar functions (Figure 2). For example, *PbIDD12* and *PmIDD10* in group 2 had the same gene structure (3 exons and 2 introns) and their exon lengths were basically the same. Furthermore, based on MEME analysis (Figure 2), we found that members of the same subfamily had roughly the same conserved protein motifs, but there were differences in motif composition between different subfamilies. N-terminal NLS is absent in many Rosaceae IDD genes (Supplementary Figure 6). In *Arabidopsis thaliana*, all IDD genes have NLS. In the *Oryza sativa* IDD gene family, *OsIDD10* also lacks this NLS (Colasanti et al., 2006). Apparently, members of the IDD gene family that lack the NLS sequence do so not because of mutations or the deletion of this entire characteristic sequence but because of an extensive deletion of the N-terminal sequence. The ancestors of the special IDD genes in these Rosaceae species may have appeared after the differentiation of these Rosaceae species from the common ancestors of *Arabidopsis thaliana*, and some of the N-terminal sequences (this part of the sequence includes NLS) are lost after the evolutionary process. In addition to the highly conserved ID domain, we found two small domains (Supplementary Figure 1-5) in the C-terminal region (except for members of group 4). The MSATALLQKAA and TRDFLG motif are conserved regions outside the ID domain but are not characteristically conserved domains of the IDD gene. Obviously, these two domains are preserved in the lineages, conferring special functions on IDD genes, such as protein-protein interactions. We find that members that lack these two domains are closely related to each other; perhaps these members lost these two domains and some of their functions during evolution. This finding suggests that these genes may be involved in specific biological processes in the evolutionary process, regardless of the biological functions determined by these two domains. Therefore, members that lack these two domains are clustered into a single subfamily (group 4). This finding explains why group 4 does not have *MdIDDs*, because *MdIDDs* have both domains.

The functional divergence of genes often occurs after gene duplication and is accompanied by the introduction of new functions or the loss of original functions (Chao et al., 2017). At the same time, gene duplication events are the main driving force for gene family expansion, an important way in which plants adapt to changing climates and environments (Tang et al., 2016). Gene duplication events have been identified in multiple gene families, such as the PKS gene family in *Gossypium hirsutum* (Su et al., 2017), the HSF gene family in *Sesamum indicum* (Dossa et al., 2016), and the WRKY gene family in *Musa acuminate* (Goel et al., 2016). A total of 4 *PbIDD* gene pairs were identified as duplications in this study, and these duplicated gene pairs were all created by segmental duplication (Supplementary table 4) . Four pairs of segmentally duplicated genes were also identified in *Malus domestica*, verifying the reliability of our results. However, we did not identify gene duplication events in *Fragaria vesca*, *Prunus mume*, *Rubus occidentalis* or *Prunus avium*. This may be because *Fragaria vesca*, *Prunus mume*, *Rubus occidentalis* and *Prunus avium* underwent only one WGDs, but *Pyrus bretschneideri* has had a recent WGDs in addition to an old WGDs (Zhang et al., 2012; Wu et al., 2013). The four pairs of duplicated genes that appeared in pears had a very positive effect on the expansion of the IDD gene family and contributed to the diversification of the function of the IDD gene. We then performed a sliding window analysis of the four pairs of segmentally duplicated genes in Chinese white pear. The results showed that these genes have experienced intense purifying selection to maintain functional stability.

We did not identify the orthologous gene pair of IDD genes between several dicotyledons and monocotyledons. Therefore, we preliminarily speculate that IDD genes may occur after the common ancestors of these species differentiated. Based on the

results, we drew a hypothetical evolutionary model map of IDD genes (Supplementary figure S10). After the differentiation of monocotyledons and dicotyledons, two canonical zinc fingers and two unusual CCHC fingers formed a special zinc finger domain (ID domain), which defines IDD genes in all plants. Subsequently, ID1 genes specific to Gramineae plants were isolated from those of monocotyledons. The ancestors of dicotyledons evolved through complex evolutionary processes to produce Rosaceae species. After an old WGDs, *Fragaria vesca*, *Rubus occidentalis*, *Prunus avium* and *Prunus mume* formed the current IDD gene family. They then differentiated into two major IDD genes with or without the MSATALLQKAA and TRDFLG domains. *Malus domestica* and *Pyrus bretschneideri* formed the IDD gene family after two WGDs. The *Pyrus bretschneideri* gene differentiated from the IDD gene without the MSATALLQKAA and TRDFLG domains, but the *Malus domestica* gene did not.

Previous research has shown that IDD genes are most highly expressed in the leaf and bud (Fan et al., 2017), but we found the highest pear IDD gene expression in fruit, flower, and bud. *PbIDD7* and *MdIDD19* represent a pair of homologous genes that are closely related. *MdIDD19* was mainly expressed in leaf, fruit, bud, while *PbIDD7* was mainly expressed in fruit, bud and almost no expression was detected in leaf (Figure 7). The difference between the two homologous gene expression patterns is mainly concentrated in the leaf expression pattern. Many IDD genes are associated with leaf development and photosynthesis. However, we did not identify *C*is-acting elements related to the light response in the promoter region of *PbIDD7* (Figure 6), while there are abundant light-responding elements in the promoter region of *MdIDD19* (Fan et al., 2017). This may be the main reason for the large difference in expression patterns between *PbIDD7* and *MdIDD19* in leaf. Interestingly, we found that *PbIDD2*, *5*, *6*, *8*, *9*, *12*, *16* showed higher transcription levels in lower and bud, and the expression in bud was higher than that in flower. In *Arabidopsis thaliana*, *Oryza sativa,* and *Zea mays*, IDD genes are involved in plant flower induction and are regulated by hormones and sugars (Matsubara et al., 2008; Wu et al., 2008; Jeong et al., 2015). For example: *AtIDD8*, *ZmID1*, *OsID1*. GA can regulate plant flowering by inhibiting flowering genes (Zhang et al.2016; Galvao et al. 2012). The promoter regions of *PbIDD2*, *6*, *8*, *12*, *16* have the GA-response *C*is-acting element. Exogenous GA was used to treat Chinese white pear buds, and the resulting expression patterns of *PbIDD2*, *5*, *6*, *8*, *9*, *12*, 16 were detected; these genes were mainly expressed in the bud and flower of Chinese white pears (Supplementary figure S9). *PbIDD5*, *9* were not induced by GA, probably because they did not have GA-related *C*is-acting elements. After exogenous GA treatment, the expression level of *PbIDD2*, *8* decreased continuously, indicating that these were inhibited by GA. However, the expression levels of these two genes increased gradually under sucrose treatment, indicating that *PbIDD2*, *8* were induced by sucrose. At the same time, we identified a large number of light-responsive *C*is-acting elements in the promoter region of *PbIDD2*, *8*, suggesting that these two genes may respond to photoperiodic signals. Similar results have been found in *Malus domestica*, and *MdIDD7* is also inhibited by GA in response to sucrose. In summary, it is suggested that *PbIDD2*, *8* may participate in the GA flowering pathway during physiological flower bud differentiation and regulate the plant flowering process in response to sugar and photoperiod signals. *PbIDD* and *10* were mainly expressed in the flowers of Chinese white pears, probably because these two genes are involved in the synthesis of bioactive substances in flowers.

The stone cells and lignin in 'Dangshan Su' pear fruit are mainly formed during the period from 15 to 63DAP and reach their peak levels at 55DAP (Zhang et al., 2017). The expression analysis of these 16 *PbIDDs* at seven developmental stages of fruit showed that the expression pattern of any one gene was not consistent with the developmental law of 'Dangshan Su' pear fruit stone cells and lignin (Figure 7). However, there were two more specific genes, *PbIDD3* and *PbIDD5*. The expression patterns of these two genes showed a completely opposite rule to the accumulation of 'Dangshan Su' pear stone cells and lignin. *PbIDD3*, *5* showed extremely high expression levels at 15DAP. At 15 to 47DAP, a stage of massive accumulation of lignin and stone cells, the expression of these two genes gradually decreased. At the peak of lignin accumulation (47, 55, 63DAP), *PbIDD3*, *5* showed very low levels of transcription. In the late stage of fruit development (79, 145DAP), the expression level increased slightly.

According to the phylogenetic tree (Figure 1), *PbIDD3*, *PbIDD5* and *OsIDD2* were the most closely related genes. In *Oryza sativa*, *OsIDD2* regulates SCW formation while negatively regulating the expression of the genes of key lignin synthesis enzymes, such as cinnamyl alcohol dehydrogenase (CAD), to inhibit lignin biosynthesis (Huang et al., 2018). In Chinese white pear, CAD is predicted to regulate the biosynthesis of fruit lignin. The expression pattern of *PbCAD* is exactly the same as the accumulation pattern of stone cells and lignin (Cheng et al., 2017). The expression patterns of *PbIDD3* and *PbIDD5* were opposite to those of *PbCAD*. Moreover, *PbIDD3* and *PbIDD5* expression was low during the high expression period of *PbCAD* and was higher during the low expression period of *PbCAD*. In addition, *PbIDD3* and *PbIDD5* exhibited extremely high levels of transcription in fruit. We speculate that *PbIDD3* and *PbIDD5* are mainly expressed in fruit and have similar functions to *OsIDD2*, regulating SCW formation in pear fruit cells and inhibiting lignin biosynthesis by inhibiting the expression of the genes of key lignin synthesis enzymes.

CONCLUSION

We identified 68 IDD genes in five Rosaceae species (*Pyrus bretschneideri*, *Fragaria vesca*, *Prunus mume*, *Rubus occidentalis* and *Prunus avium*), which we systematically assessed by bioinformatic analysis. According to the phylogenetic tree, the 68 IDD genes were divided into 4 groups. In each class, we found that the structures of the genes and the compositions of the conserved motifs were very similar. Through a series of bioinformatics analyses, we explained the possible evolutionary patterns of IDD genes in these Rosaceae species. According to qRT-PCR, Chinese white pear IDD genes have high expression levels in fruit, flower and bud. Further, *PbIDD2*, *8* are mainly expressed in the buds of Chinese white pears and are responsive to the induction of GA and sucrose. That *PbIDD3*, *5* are may be involved in the regulation of SCW formation and lignin biosynthesis in Chinese white pear fruit. All in all, this study reveal the basic information of the five Rosaceae species' IDD genes and predicts the potential functions of some pear IDD proteins. These results will provide an important theoretical basis for improving the quality of 'Dangshan Su' pear.

ACKNOWLEDGEMENTS

The authors are deeply grateful to Prof. Yongping Cai and Prof. Yi Lin, who provided the sample used in the study and very effective direction. The authors also thank Dr. Xu Sun, Dr. Muhammad Abdullah, Dr. Xi Cheng, Dr. Guohui Li, Dr. Yu Zhao for providing valuable suggestions and comments. Finally, author would like to thank Dr. Meng for his help in revising the paper.

References

Abdullah, M., Cao, Y. P., Cheng, X., Shakoor, A., Su, X. Q., Gao, J. S., Cai, Y. P. (2018). Genome-Wide Analysis Characterization and Evolution of SBP Genes in *Fragaria vesca*, *Pyrus bretschneideri*, *Prunus persica* and *Prunus mume*. *FRONT. GENET.* **9**, 64.

Agarwal, P., Arora, R., Ray, S., Singh, A. K., Singh, V. P., Takatsuji, H., Kapoor, S., Tyagi, A. K. (2007). Genome-wide identification of C2H2 zinc-finger gene family in rice and their phylogeny and expression analysis. *PLANT. MOL. BIOL.* **65**, 467-485.

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., Castro, E., Duvaud, E., Flegel, V., Fortier, A.,

Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I., Stockinger, H. (2012). ExPASy: SIB bioinformatics resource portal. *NUCLEIC. ACIDS. RES.* **40**, 597-603.

Bailey, T. L., Johnson, J., Grant, C. E., Noble, W. S. (2015). The MEME Suite. *NUCLEIC. ACIDS. RES.* **43**, 39-49.

Bitocchi, E., Rau, D., Benazzo, A., Bellucci, E., Goretti, D., Biagetti, E., Panziera, A., Laidò, G., Rodriguez, M., Gioia, T., Attene, G., McClean, P., Lee, R. K., Jackson, S. A., Bertorelle, G., Papa, R. (2017). High Level of Nonsynonymous Changes in Common Bean Suggests That Selection under Domestication Increased Functional Diversity at Target Traits. *FRONT PLANT SCI.* **7**, 2005.

Cai, Y. P., Li, G. Q., Nie, J. Q., Lin, Y., Nie, F., Zhang, J. Y. and Xu, Y. L. (2010). Study of the structure and biosynthetic pathway of lignin in stone cells of pear. *Sci Hortic-Amsterdam.* **125**, 374-379.

Cao, Y. P., Han, Y. H., Li, D. H., Lin, Y., Cai, Y. P. (2016a). Systematic Analysis of the 4-Coumarate: Coenzyme A Ligase (4CL) Related Genes and Expression Profiling during Fruit Development in the Chinese Pear. *GENES.* **7**, 89-106.

Chai, G. H., Kong, Y. Z., Zhu, M., Yu, L., Qi, G., Tang, X. F., Wang, Z. G., Cao, Y. P., Yu, C. J., Zhou, G. K. (2015). Arabidopsis C3H14 and C3H15 have overlapping roles in the regulation of secondary wall thickening and anther development. *J. EXP. BOT.* **66**, 2595-2609.

Cheng, X., Li, M. L., Li, D. H., Zhang, J. Y., Jin, Q., Sheng, L. L., Cai, Y. P., Lin, Y. (2017). Characterization and analysis of CCR and CAD gene families at the whole-genome level for lignin synthesis of stone cells in pear (Pyrus bretschneideri) fruit. *BIOL. OPEN.* **6**, 1602-1613.

Chao, N., Li, N., Qi, Q., Li, S., Lv, T., Jiang, X. N., Gai, Y. (2017). Characterization of the cinnamoyl-CoA reductase (CCR) gene family in Populus tomentosa reveals the enzymatic active sites and evolution of CCR. *PLANTA.* **245**, 61-75.

Chen, J., Wang, J. N., Kuang, J. F., Shan, W., Chen, J. Y., Lu, W. J. (2014). Molecular characterisation of a fruit ripening-related INDETERMINATE DOMAIN transcription factor in banana (*Musa acuminata*). *J. HORTIC. SCI. BIOTECH.* **89**, 373-380.

Coelho, C. P., Huang, P., Lee, D. Y., Brutnell, T. P. (2018). Making Roots, Shoots, and Seeds: IDD Gene Family Diversification in Plants. *TRENDS. PLANT. SCI.* **23,** 66-78.

Colasanti, J., Tremblay, R., Wong, A. Y. M., Coneva, V., Kozaki, A., Mable, B. K. (2006). The maize INDETERMINATEI flowering time regulator defines a highly conserved zinc finger protein family in higher plants. *BMC. GENOMICS.* **7**, 158-175.

Colasanti, J., Yuan, Z. and Sundaresan, V. (1998). The indeterminate Gene Encodes a Zinc Finger Protein and Regulates a Leaf-Generated Signal Required for the Transition to Flowering in Maize. *Cell*. **93**, 593-603.

Cui, D. Y., Zhao, J. B., Jing, Y. J., Fan, M. Z., Liu, J., Wang, Z. C., Xin, W., Hu, Y. X. (2013). The Arabidopsis IDD14, IDD15, and IDD16 Cooperatively Regulate Lateral Organ Morphogenesis and Gravitropism by Promoting Auxin Biosynthesis and Transport. *PLOS. GENET.* **9**, e1003759.

Dossa, K., Diouf, D., Cisse, N. (2016). Genome-Wide Investigation of Hsf Genes in Sesame Reveals Their Segmental Duplication Expansion and Their Active Role in Drought Stress Response. *FRONT. PLANT. SCI.* **7**, 1522.

Fan, S., Zhang, D., Xing, L. B., Qi, S. Y., Du, L. S., Wu, H. Q., Shao, H. X., Li, Y. M., Ma, J. J., Han, M. Y. (2017). Phylogenetic analysis of IDD gene family and characterization of its expression in response to flower induction in Malus. *MOL. GENET. GENOMICS.* **292**, 755-771.

Feurtado, J. A., Huang, D., Wicki-Stordeur, L., Hemstock, L. E., Potentier, M. S., Tsang. E. W., Cutler, A. J. (2011). The Arabidopsis C2H2 zinc finger INDETERMINATE DOMAIN1/ENHYDROUS promotes the transition to germination by regulating light and hormonal signaling during seed maturation. *Plant. Cell.* **23**, 1772-1794.

Galvao V. C., Horrer D., Kuttner F., Schmid M. (2012). Spatial control of flowering by DELLA proteins in Arabidopsis thaliana. *Development.* **139**, 4072–4082.

Goel, R., Pandey, A., Trivedi, P. K., Asif, M. H. (2016). Genome-Wide Analysis of the Musa WRKY Gene Family: Evolution and Differential Expression during Development and Stress. *FRONT. PLANT. SCI.* **7**, 299.

Guo, A. Y., Zhu, Q. H., Chen, X., Luo, J. C. (2007). [GSDS: a gene structure display server]. *Yi. Chuan.* **29**, 1023-1026.

Han, Y. H., Ding, T., Su, B.,Jiang, H. Y. (2016). Genome-Wide Identification, Characterization and Expression Analysis of the Chalcone Synthase Family in Maize. *INT. J. MOL. SCI.* **17**, 161.

Hamauzu, Y., Forest, F., Hiramatsu, K., Sugimoto, M. (2007). Effect of pear (Pyrus communis, L.) procyanidins on gastric lesions induced by HCl/ethanol in rats. *FOOD. CHEM.* **99**, 255-263.

Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., Nakai, K. (2007). WoLF PSORT: protein localization predictor. *NUCLEIC. ACIDS. RES.* **35**, 585-587.

Hu, L. F., Liu, S. Q. (2011). Genome-wide identification and phylogenetic analysis of the ERF gene family in cucumbers. *GENET. MOL. BIOL.* **34**, 624-U201.

Huang, L. J., Gao, W. Y., Li, X., Zhao, W. S., Huang, L. Q., Liu, C. X. (2010). valuation of the in Vivo Anti-inflammatory Effects of Extracts from Pyrus bretschneideri Rehd. *J. AGR. FOOD. CHEM.* **58**, 8983-8987.

Huang, P., Yoshida, H., Yano, K., Kinoshita, S., Kawai, K., Koketsu, E., Hattori, M., Takehara, S ., Huang, J., Hirano, K., Ordonio, R. L., Matsuoka, M., Ueguchi-Tanaka, M. (2018). OsIDD2, a zinc finger and INDETERMINATE DOMAIN protein, regulates secondary cell wall formation. *J. INTEGR. PLANT. BIOL.* **60**, 130-143.

Jeong, E. Y., Seo, P. J., Woo, J. C., Park, C. M. (2015). AKIN10 delays flowering by inactivating IDD8 transcription factor through protein phosphorylation in Arabidopsis. *BMC. PLANT. BIOL.* **15**, 110.

Kozaki, A., Hake, S., Colasanti, J. (2004). The maize ID1 flowering time regulator is a zinc finger protein with novel DNA binding properties. *Nucleic. Acids. Res.* **32**, 1710-1720.

Kumar S, Stecher G, Tamura K. (2016). MEGA 7.0: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *MOL. BIOL. EVOL.* **33**, 1870-1874.

Letunic, I., Doerks, T., Bork, P. (2012). SMART 7: recent updates to the protein domain annotation resource. *NUCLEIC. ACIDS. RES.* **40**, 302-305.

Lin, Y. X., Jiang, H. Y., Chu, Z. X., Tang, X. L., Zhu, S. W., Cheng, B. J. (2011). Genome-wide identification, classification and analysis of heat shock transcription factor family in maize. *BMC. GENOMICS.* **12**, 76.

Librado, P., Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *BIOINFORMATICS.* **25**, 1451-1452.

Livak, K. J., Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods.* **25**, 402-408.

Long M, Thornton K. (2001). Gene duplication and evolution. *Science.* **293**, 1551.

Matsubara, K., Yamanouchi, U., Wang, Z. X., Minobe, Y., Izawa, T., Yano, M. (2008). Ehd2, a Rice Ortholog of the Maize INDETERMINATE1 Gene, Promotes Flowering by Up-Regulating Ehd1. *PLANT. PHYSIOL.* **148**, 1425-1435.

Miller, J., Mclachlan, A. D., Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *EMBO. J.* **4**, 1609-1614.

Park, S. J., Kim, S. L., Lee, S., Je, B. I., Piao, H. L., Park, S. H., Kim, C. M., Ryu, C. H., Park, S. H., Xuan, Y. H., Colasanti, J., An, G., Han, C. D. (2008). Rice Indeterminate 1 (*OsId1*) is necessary for the expression of Ehd1 (Early heading date 1) regardless of photoperiod. *PLANT. J.* **56**, 1018-1029.

Parraga, G., Horvath, S. J., Eisen, A., Taylor, W. E., Hood, L., Young, E. T., Klevit, R. E. (1988). Zinc-dependent structure of a single-finger domain of yeast ADR1. *Science.* **241**, 1489-1492.

Ingkasuwan, P., Netrphan, S., Prasitwattanaseree, S., Tanticharoen, M., Bhumiratana, S ., Meechai, A., Chaijaruwanich, J., Takahashi, H., Cheevadhanarak, S. (2012). Inferring transcriptional gene regulation network of starch metabolism in Arabidopsis thaliana leaves using graphical Gaussian model. *BMC. SYST. BIOL.* **6**, 100-121.

Islam, M. S., Hur, J. H., Wang, M. . (2009). The Influence of Abiotic Stresses on Expression of Zinc Finger Protein Gene in Rice. *RUSS. J. PLANT. PHYSL.* **56**, 695-701.

Rombauts, S, Déhais P, Van Montagu, M, Rouzé, P. (1999). PlantCARE, a plant cis-acting regulatory element database. *NUCLEIC. ACIDS. RES.* **27**, 295-296.

Seo, P. J., Ryu, J., Kang, S. K., Park, C. M. (2011). Modulation of sugar metabolism by an INDETERMINATE DOMAIN transcription factor contributes to photoperiodic flowering in Arabidopsis. *PLANT. J.* **65**, 418-429.

Shi, X. D., Gu, Y. X., Dai, T. W., Wu, Y., Wu, P., Xu, Y., Chen, F. (2018). Regulation of trichome development in tobacco by *JcZFP8*, a C2H2 zinc finger protein gene from Jatropha curcas L. *GENE.* **658**, 47-53.

Shirasawa, K., Isuzugawa, K., Ikenaga, M., Saito, Y., Yamamoto, T., Hirakawa, H., Isobe, S. (2017). The genome sequence of sweet cherry (*Prunus avium*) for use in genomics-assisted breeding. *DNA. RES.* **24**, 499-508.

Shulaev, V., Sargent, D. J., Crowhurst, R. N., Mockler, T. C., Folkerts, O., Delcher, A. L., Jaiswal, P., Mockaitis, K.,

Liston, A., Mane, S. P., Burns, P., Davis, T. M., Slovin, J. P., Bassil, N., Hellens, R. P., Evans, C., Harkins, T., Kodira, C., Desany, B., Crasta, O. R., Jensen, R. V., Allan, A. C., Michael, T. P., Setubal, J. C., Celton, J. M., Rees, D. J. G., Williams, K. P., Holt, S. H., Ruiz Rojas, J. J. R., Chatterjee, M., Liu, B., Silva, H., Meisel, L., Adato, A., Filichkin, S. A., Troggio, M., Viola, R., Ashman, T. L., Wang, H., Dharmawardhana, P., Elser, J., Raja, R., Priest, H. D., Bryant, D. W., Fox, S. E., Givan, S. A., Wilhelm, L. J., Naithani, S., Christoffels, A., Salama, D., Carter, J., Girona, E. L., Zdepski, A., Wang, W., Kerstetter, R. A., Schwab, W., Korban, S. S., Davik, J., Monfort, A., Denoyes-Rothan, B., Arus, P., Mittler, R., Flinn, B., Aharoni, A., Bennetzen, J., Salzberg, S. L., Dickerman, A. W., Velasco, R., Borodovsky, M., Veilleux, R. E., Folta, K. M. (2011). The genome of woodland strawberry (*Fragaria vesca*). *NAT. GENET.* **43**, 109-116.

Su, X. Q., Sun, X., Cheng, X., Wang, Y. N., Abdullah, M., Li, M. L., Li, D. H., Gao, J. S., Cai, Y. P., Lin, Y. (2017). Comparative genomic analysis of the PKS genes in five species and expression analysis in upland cotton. *PEERJ.* **5**, e3974.

Swarbreck, D., Wilks, C., Lamesch, P., Berardini, T. Z., Garcia-Hernandez, M., Foerster, H., Li, D., Meyer, T., Muller, R., Ploetz, L., Radenbaugh, A., Singh, S., Swing, V., Tissier, C., Zhang, P., Huala, E. (2008). The Arabidopsis Information Resource (TAIR): gene structure and function annotation. *NUCLEIC. ACIDS. RES.* **36**, 1009-1014.

Tavaud, M., Zanetto, A., David, J. L., Laigret, F., Dirlewanger, E. (2004). Genetic relationships between diploid and allotetraploid cherry species (*Prunus avium*, *Prunus x gondouinii* and *Prunus cerasus*). *HEREDITY.* **93**, 631-638.

Tang, R. M., Zhu, W. J., Song, X. Y., Lin, X. Z., Cai, J. H., Wang, M., Yang, Q. (2016). Genome-Wide Identification and Function Analyses of Heat Shock Transcription Factors in Potato. *FRONT. PLANT. SCI.* **7**, 490.

Tian, Z. D., Zhang, Y., Liu, J., Xie, C. H., (2010). Novel potato C2H2-type zinc finger protein gene, *StZFP1*, which responds to biotic and abiotic stress, plays a role in salt tolerance. *PLANT. BIOLOGY.* **12**, 689-697.

VanBuren, R., Bryant, D., Bushakra, J. M., Vining, K. J., Edger, P. P., Rowley, E. R., Priest, H. D., Michael, T. P., Lyons, E., Filichkin, S. A., Dossett, M., Finn, C. E., Bassil, N. V., Mockler, T. C. (2016). The genome of black raspberry (*Rubus occidentalis*). *PLANT. J.* **87**, 535-547.

Wei, K., Pan, S., Li, Y. (2016). Functional Characterization of Maize C2H2, Zinc-Finger Gene Family. *PLANT. MOL. BIOL. REP.* **34**, 761-776.

Wong, A. Y., Colasanti, J. (2007). Maize floral regulator protein INDETERMINATE1 is localized to developing leaves and is not altered by light or the sink/source transition. *J. EXP. BOT.* **58**, 403-414.

Wu, J., Wang, Z., Shi, Z., Zhang, S., Ming, R., Zhu, S., Khan, M. A., Tao, S., Korban, S. S., Wang, H., Chen, N. J., Nishio, T., Xu, X., Cong, L., Qi, K., Huang, X., Wang, Y., Zhao, X., Wu, J., Deng, C., Gou, C., Zhou, W., Yin, H., Qin, G., Sha, Y., Tao, Y., Chen, H., Yang, Y., Song, Y., Zhan, D., Wang, J., Li, L., Dai, M., Gu, C., Wang, Y., Shi, D., Wang, X., Zhang, H., Zeng, L., Zheng, D., Wang, C., Chen, M., Wang, G., Xie, L., Sovero, V., Sha, S., Huang, W., Zhang, S., Zhang, M., Sun, J., Xu, L., Li, Y., Liu, X., Li, Q., Shen, J., Wang, J., Paull, R. E., Bennetzen, J. L., Wang, J., Zhang, S. (2013). The genome of the pear (Pyrus bretschneideri Rehd). *Genome. Res.* 23, 396–408.

Wu, C. Y., You, C. J., Li, C. S., Long, T., Chen, G. X., Byrne, M. E., Zhang, Q. F. (2008). *RID1*, encoding a Cys2/His2-type zinc finger transcription factor, acts as a master switch from vegetative to floral development in rice. *P. NATL. ACAD. SCI.*

USA. **105**, 12915-12920.

Yan, C. C., Yin, M., Zhang, N., Jin, Q., Fang, Z., Lin, Y., Cai, Y.P. (2014). Stone cell distribution and lignin structure in various pear varieties. *SCI. HORTIC-AMSTERDAM.* **174**, 142-150.

Yi, G., Neelakandan, A. K., Gontarek, B. C., Vollbrecht, E., Becraft, P. W. (2015). The naked endosperm Genes Encode Duplicate INDETERMINATE Domain Transcription Factors Required for Maize Endosperm Cell Patterning and Differentiation. *PLANT. PHYSIOL.* **167**, 443-456.

Zhang, D. M., Xu, Z. P., Cao, S. X., Chen, K. L., Li, S., Liu, X. L., Gao, C. X., Zhang, B. C., Zhou, Y. H. (2018). An Uncanonical CCCH-Tandem Zinc-Finger Protein Represses Secondary Wall Synthesis and Controls Mechanical Strength in Rice. *MOL. PLANT.* **11**, 163-174.

Zhang, J. Y., Cheng, X., Jin, Q., Su, X. Q., Li, M. L., Yan, C. C., Jiao, X. Y., Li, D. H., Lin, Y., Cai, Y. P. (2017). Comparison of the transcriptomic analysis between two Chinese white pear (Pyrus bretschneideri Rehd.) genotypes of different stone cells contents. *PLOS. ONE.* **12**, e0187114.

Zhang, Q., Chen, W., Sun, L., Zhao, F., Huang, B., Yang, W., Tao, Y., Wang, J., Yuan, Z., Fan, G., Xing, Z., Han, C., Pan, H., Zhong, X., Shi, W., Liang, X., Du, D., Sun, F., Xu, Z., Hao, R., Lv, T., Lv, Y., Zheng, Z., Sun, M., Luo, L., Cai, M., Gao, Y., Wang, J., Yin, Y., Xu, X., Cheng, T., Wang, J. (2012). The genome of Prunus mume. *Nat. Commun.* **3**, 1318.

Zhang, S. W., Zhang, D., Fan, S., Du, L. S., Shen, Y. W., Xing, L. B., Li, Y. M., Ma, J. J., Han, M. Y. (2016). Effect of exogenous GA(3) and its inhibitor paclobutrazol on floral formation, endogenous hormones, and flowering-associated genes in 'Fuji' apple (*Malus domestica Borkh.*). *PLANT. PHYSIOL. BIOCH.* **107**, 178-186.

Zhu, G. R., Chen, G. X., Zhu, J. T., Zhu, Y., Lu, X. B., Li, X. H., Hu, Y. K., Yan, Y. M. (2015). Molecular Characterization and Expression Profiling of NAC Transcription Factors in Brachypodium distachyon L. *PLOS. ONE.* **10**, e0139794.

Table 1. The IDD genes of *Pyrus bretschneideri* identified in this study are listed.

Figure 1. Phylogenetic relationships and subfamily designations in IDD proteins from *Pyrus bretschneideri*, *Fragaria vesca*, *Prunus mume*, *Rubus occidentalis*, *Prunus avium*, *Malus domestica*, *Arabidopsis thaliana*, *Oryza sativa* and *Zea mays*. Groups 1-4 are represented by shades of red, green, blue and cyan, respectively.

Figure 2. Predicted *Pyrus bretschneideri*, *Fragaria vesca*, *Prunus mume*, *Rubus occidentalis* and *Prunus avium* IDD protein conserved motifs and exon-intron structures. (A) Gene structures of the IDD genes. Black wedge indicates exon, black line indicates intron and red wedge indicates UTR. (B) Distribution of MEME motifs in IDD genes. (C) The colour and corresponding number of each motif box.

Figure 3. Conserved ID domain composition. All 68 IDD genes had a characteristic ID domain. The ID domain consists of four zinc fingers (Z1, Z2, Z3 and Z4). Alignment analysis of the 68 IDD gene sequences using the ClustalW tool in MEGA 7.0 software. These domain diagrams were plotted using the online WebLogo tool.

Figure 4. Chromosomal locations of IDD genes in *Pyrus bretschneideri* (A), *Fragaria vesca* (B), *Rubus occidentalis* (C), *Prunus*

mume (D) and *Prunus avium* (E). Duplicated gene pairs are connected with coloured lines.

Figure 5. Microsynteny of regions among *Pyrus bretschneideri*, *Fragaria vesca*, *Prunus mume*, *Rubus occidentalis*, *Prunus avium* and *Malus domestica*. The chromosome numbers are indicated by differently coloured boxes and are labelled by Pb, Fv, Pm, Ro, Pa, and Md. The differently coloured boxes also represent the sequence lengths of chromosomes in megabases. The black line indicates the syntenic relationship among the IDD regions.

Figure 6. Potential cis-elements in the 5' regulatory sequences of the 16 PbIDD genes.

Figure 7. Expression patterns of IDD genes of Chinese white pear in different tissues and in fruit at different developmental stages. Expression patterns of IDD genes in Chinese white pear in different tissues (A-O). Expression patterns of IDD genes in Chinese white pear at different developmental stages (P-DD). *significant difference at $P < 0.05$, **significant difference at $P < 0.01$.

Supplemental Figure S1. *Pyrus bretschneideri* IDD protein sequence alignment. Black underline indicates zinc finger domain (Z1, Z2, Z3 and Z4). Red triangle indicates a conserved C residue, and blue triangle indicates a conserved H residue. The yellow underline indicates the NLS sequence in the N-terminal region of the IDD gene. Green box means the MSATALLQKAA domain, and purple box indicates the TRDFLG domain.

Supplemental Figure S2. *Fragaria vesca* IDD protein sequence alignment. Black underline indicates zinc finger domain (Z1, Z2, Z3 and Z4). Red triangle indicates a conserved C residue, and blue triangle indicates a conserved H residue. The yellow underline indicates the NLS sequence in the N-terminal region of the IDD gene. Green box means the MSATALLQKAA domain, and purple box indicates the TRDFLG domain.

Supplemental Figure S3. *Prunus mume* IDD protein sequence alignment. Black underline indicates zinc finger domain (Z1, Z2, Z3 and Z4). Red triangle indicates a conserved C residue, and blue triangle indicates a conserved H residue. The yellow underline indicates the NLS sequence in the N-terminal region of the IDD gene. Green box means the MSATALLQKAA domain, and purple box indicates the TRDFLG domain.

Supplemental Figure S4. *Rubus occidentalis* IDD protein sequence alignment. Black underline indicates zinc finger domain (Z1, Z2, Z3 and Z4). Red triangle indicates a conserved C residue, and blue triangle indicates a conserved H residue. The yellow underline indicates the NLS sequence in the N-terminal region of the IDD gene. Green box means the MSATALLQKAA domain, and purple box indicates the TRDFLG domain.

Supplemental Figure S5. *Prunus avium* IDD protein sequence alignment. Black underline indicates zinc finger domain (Z1, Z2, Z3 and Z4). Red triangle indicates a conserved C residue, and blue triangle indicates a conserved H residue. The yellow underline indicates the NLS sequence in the N-terminal region of the IDD gene. Green box means the MSATALLQKAA domain, and purple box indicates the TRDFLG domain.

Supplemental Figure S6. N-terminal region of the ID-domain shows the putative NLS sequence. The yellow underline indicates the NLS sequence in the N-terminal region of the IDD gene.

Supplemental Figure S7. Sliding window plots of duplicated IDD genes in Chinese white pear. The grey shaded portion

604 indicates conserved ID domain. The X-axis indicates the synonymous distance within each gene.

605 **Supplemental Figure S8.** A phylogenetic tree of the 12 species genomes (A). Phylogenetic relationships and subfamily
606 designations in IDD proteins from 12 species (B).

607 **Supplemental Figure S9.** Expression modes of candidate *PbIDD2*, *5*, *6*, *8*, *9*, *12*, *16* in Chinese white Pear buds treated with
608 gibberellin (A-G) and sucrose (H-N). *significant difference at $P < 0.05$, **significant difference at $P < 0.01$.

609 **Supplemental Figure S10.** A hypothetical evolutionary model map of IDD genes.

610 **Supplementary Table S1.** Gene sequence involved in the article.

611 **Supplemental Table S2.** The IDD genes of *Fragaria vesca*, *Prunus mume*, *Rubus occidentalis* and *Prunus avium* identified in
612 this study are listed.

613 **Supplementary Table S3.** Detailed information of the 20 motifs in the 68 IDD proteins.

614 **Supplementary Table S4.** Ka/Ks analysis of the duplicated IDD paralogues from Chinese white pear.

615 **Supplemental Table S5.** Synteny data in *Pyrus bretschneideri*, *Fragaria vesca*, *Prunus mume*, *Rubus occidentalis*, *Prunus*
616 *avium*, *Malus domestica*.

617 **Supplementary Table S6.** Promoter sequence of 16 IDD genes in Chinese white pear.

618 **Supplementary Table S7.** Complete information on *C*is-acting elements of the 16 *PbIDD* genes.

619 **Supplementary Table S8.** Primers used in qRT-PCR.

Table 1(on next page)

Basic Information of IDD Gene in *Pyrus bretschneideri.*

The IDD genes of *Pyrus bretschneideri* identified in this study are listed.

1 **Table 1.** The IDD genes of pear identified in this study are listed.

Gene name	Gene ID	AA	KD	pI	GRAVY	Preditced subcellular localization
PbIDD1	Pbr029706.1	489	51.5	8.17	-0.460	nucl
PbIDD2	Pbr006167.1	535	55.7	8.98	-0.427	nucl
PbIDD3	Pbr032492.1	533	58.0	8.98	-0.659	nucl
PbIDD4	Pbr021137.1	553	60.4	9.31	-0.634	nucl
PbIDD5	Pbr028264.1	526	57.4	8.89	-0.693	nucl
PbIDD6	Pbr019853.1	537	59.1	8.67	-0.807	nucl
PbIDD7	Pbr012193.2	567	61.7	8.92	-0.772	nucl
PbIDD8	Pbr020403.1	524	57.3	9.07	-0.740	nucl
PbIDD9	Pbr009954.1	465	50.9	9.19	-0.664	nucl
PbIDD10	Pbr012907.1	465	51.1	8.99	-0.629	nucl
PbIDD11	Pbr008330.1	464	51.0	9.35	-0.727	nucl
PbIDD12	Pbr012192.1	607	64.7	9.13	-0.723	nucl
PbIDD13	Pbr025606.1	426	47.2	9.25	-0.646	nucl
PbIDD14	Pbr029012.1	381	42.7	9.36	-0.655	nucl
PbIDD15	Pbr038802.1	502	56.4	9.02	-0.914	nucl
PbIDD16	Pbr012170.1	492	54.4	8.93	-0.805	nucl

2

Figure 1

Phylogenetic relationships and subfamily designations in IDD proteins from *Pyrus bretschneideri*, *Fragaria vesca*, *Prunus mume*, *Rubus occidentalis*, *Prunus avium*, *Malus domestica*, *Arabidopsis thaliana*, Oryza sativ an

Phylogenetic relationships and subfamily designations in IDD proteins from *Pyrus bretschneideri*, *Fragaria vesca*, *Prunus mume*, *Rubus occidentalis*, *Prunus avium*, *Malus domestica*, *Arabidopsis thaliana*, *Oryza sativa* and *Zea mays*. Groups 1-4 are represented by shades of red, green, blue and cyan, respectively.

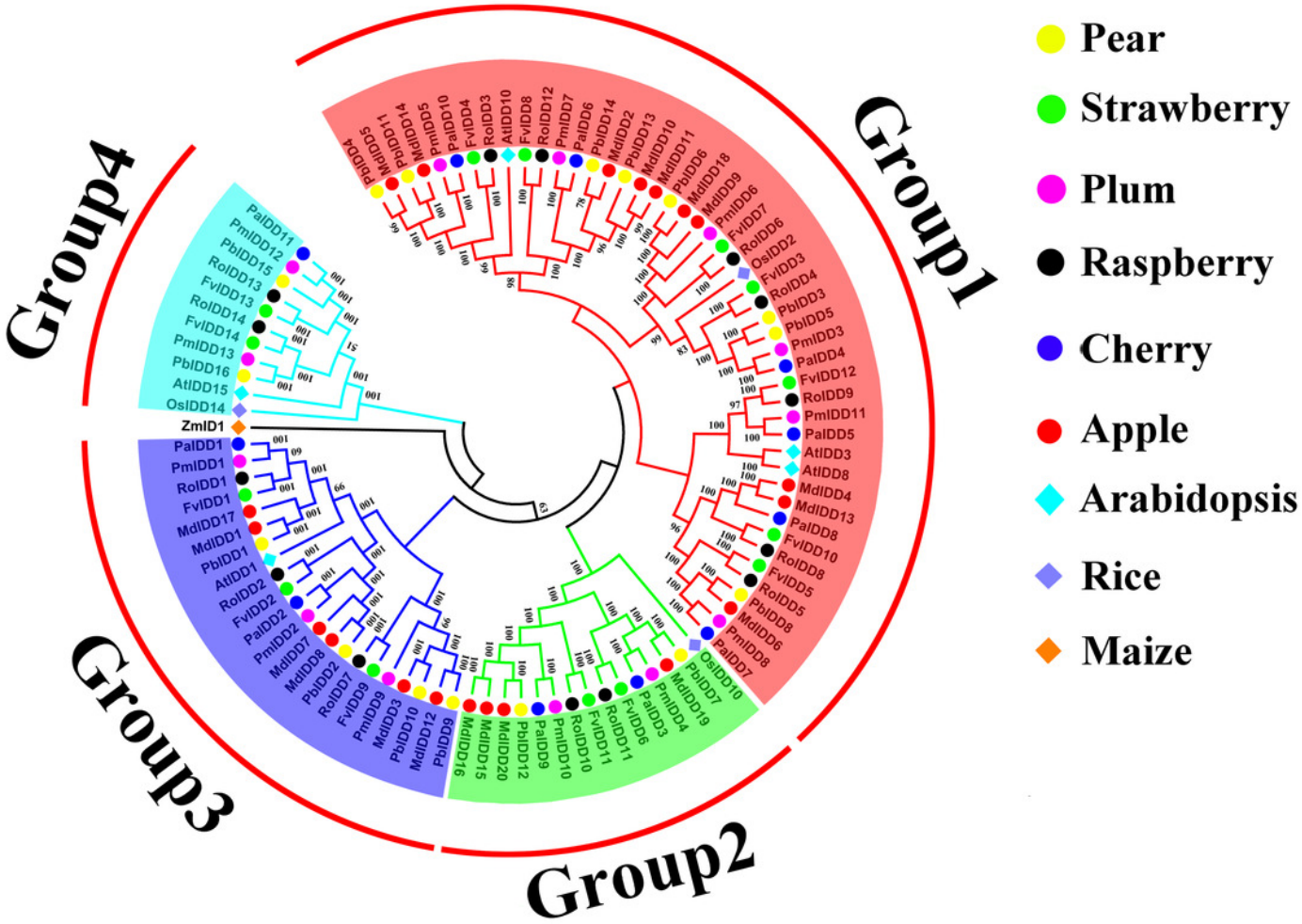


Figure 2

Predicted *Pyrus bretschneideri*, *Fragaria vesca*, *Prunus mume*, *Rubus occidentalis* and *Prunus avium* IDD protein conserved motifs and exon-intron structures.

Predicted *Pyrus bretschneideri*, *Fragaria vesca*, *Prunus mume*, *Rubus occidentalis* and *Prunus avium* IDD protein conserved motifs and exon-intron structures. (A) Gene structures of the IDD genes. Black wedge indicates exon, black line indicates intron and red wedge indicates UTR. (B) Distribution of MEME motifs in IDD genes. (C) The colour and corresponding number of each motif box.

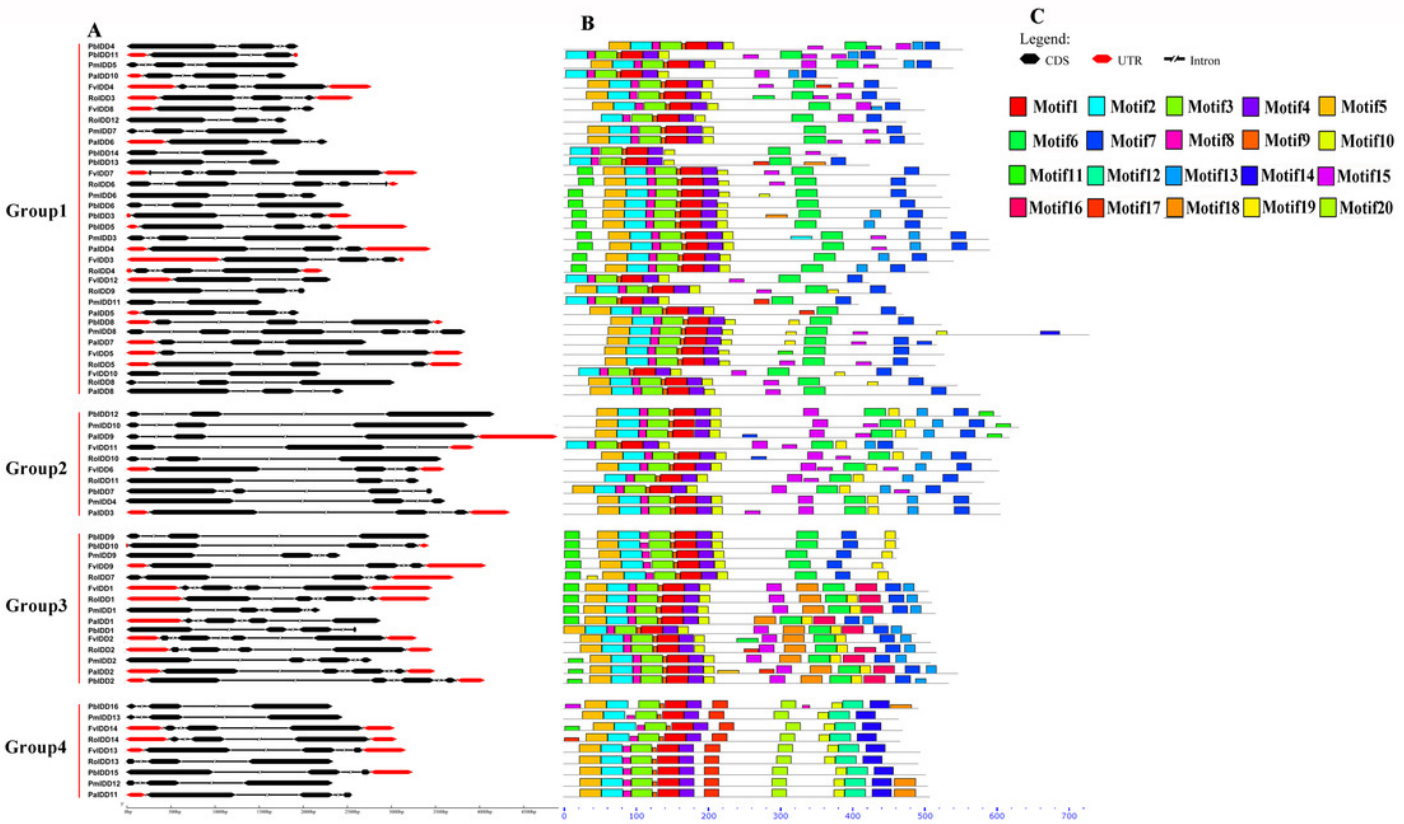


Figure 3

Conserved ID domain composition.

Conserved ID domain composition. All 68 IDD genes had a characteristic ID domain. The ID domain consists of four zinc fingers (Z1, Z2, Z3 and Z4). Alignment analysis of the 68 IDD gene sequences using the ClustalW tool in MEGA 7.0 software. These domain diagrams were plotted using the online WebLogo tool.

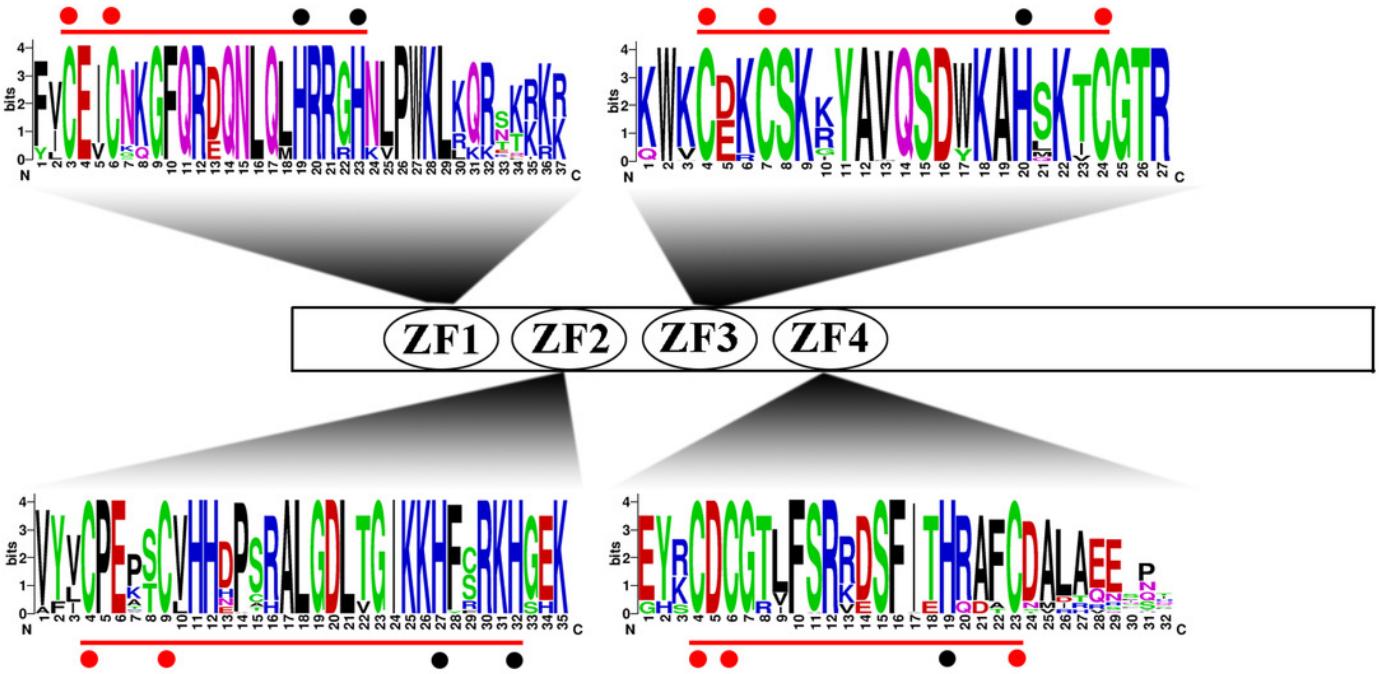


Figure 4

Chromosomal locations of five Rosaceae species.

Chromosomal locations of IDD genes in *Pyrus bretschneideri* (A), *Fragaria vesca* (B), *Rubus occidentalis* (C), *Prunus mume* (D) and *Prunus avium* (E). Duplicated gene pairs are connected with coloured lines.

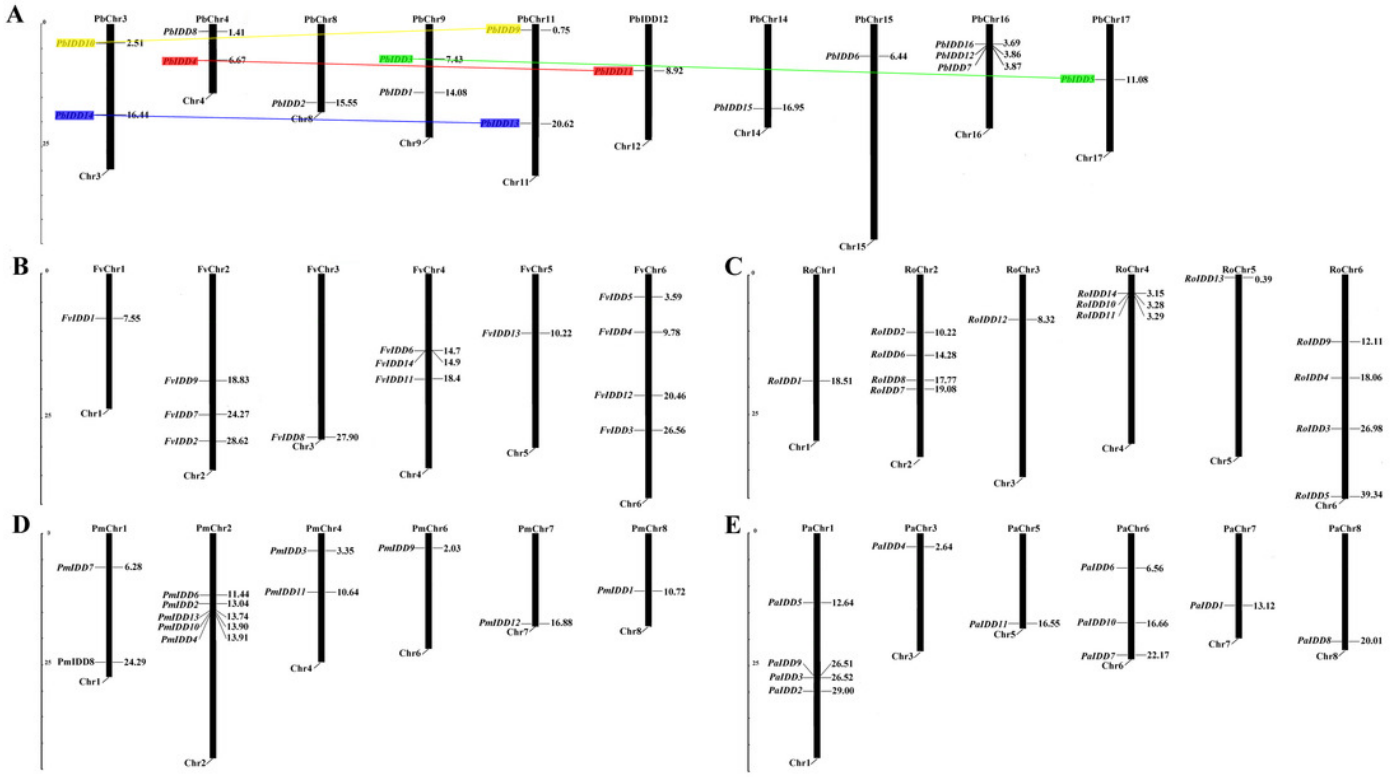


Figure 5

Microsynteny of regions among *Pyrus bretschneideri*, *Fragaria vesca*, *Prunus mume*, *Rubus occidentalis*, *Prunus avium* and *Malus domestica*.

Microsynteny of regions among *Pyrus bretschneideri*, *Fragaria vesca*, *Prunus mume*, *Rubus occidentalis*, *Prunus avium* and *Malus domestica*. The chromosome numbers are indicated by differently coloured boxes and are labelled by Pb, Fv, Pm, Ro, Pa, and Md. The differently coloured boxes also represent the sequence lengths of chromosomes in megabases. The black line indicates the syntenic relationship among the IDD regions.

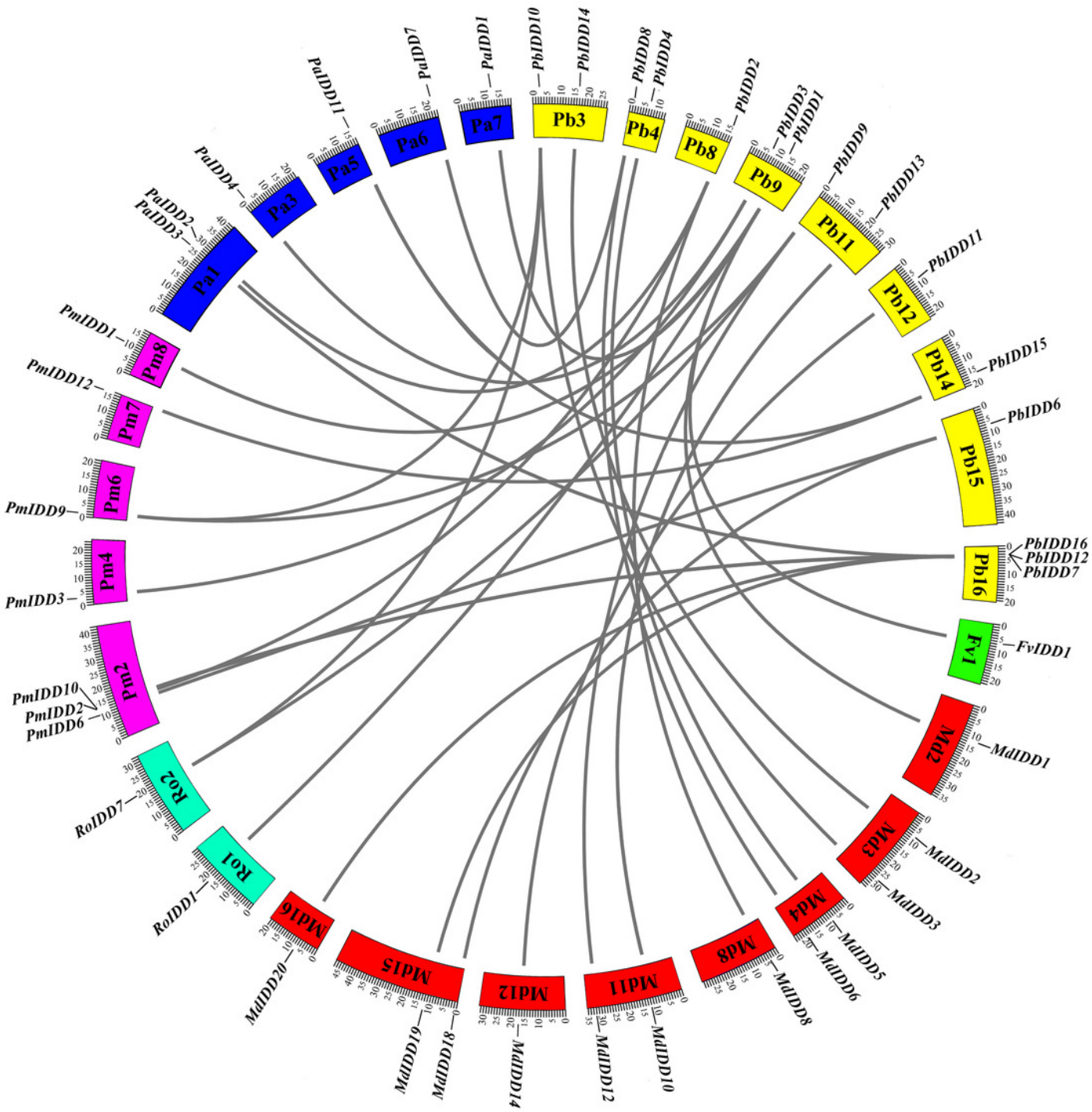

Pa1
Pa3
Pa5
Pa6
Pa7
Pb3
Pb4
Pb8
Pb9
Pb11
Pb12
Pb14
Pb15
Pb16
Fv1
Md2
Md3
Md4
Md8
Md11
Md12
Md15
Md16
R01
R02
Pm2
Pm4
Pm6
Pm7
Pm8
PaIDD2
PaIDD3
PaIDD4
PaIDD11
PaIDD7
PaIDD1
PbIDD10
PbIDD14
PbIDD8
PbIDD4
PbIDD2
PbIDD3
PbIDD1
PbIDD9
PbIDD13
PbIDD11
PbIDD15
PbIDD6
PbIDD16
PbIDD12
PbIDD7
FvIDD1
MdIDD1
MdIDD2
MdIDD3
MdIDD5
MdIDD6
MdIDD8
MdIDD10
MdIDD12
MdIDD14
MdIDD18
MdIDD19
MdIDD20
RoIDD1
RoIDD7
PmIDD10
PmIDD2
PmIDD6
PmIDD3
PmIDD9
PmIDD12
PmIDD1

Figure 6

Promoter *C*is-elements of the 16 *PbIDDs.*

Potential *C*is-elements in the 5’ regulatory sequences of the 16 *PbIDDs*.

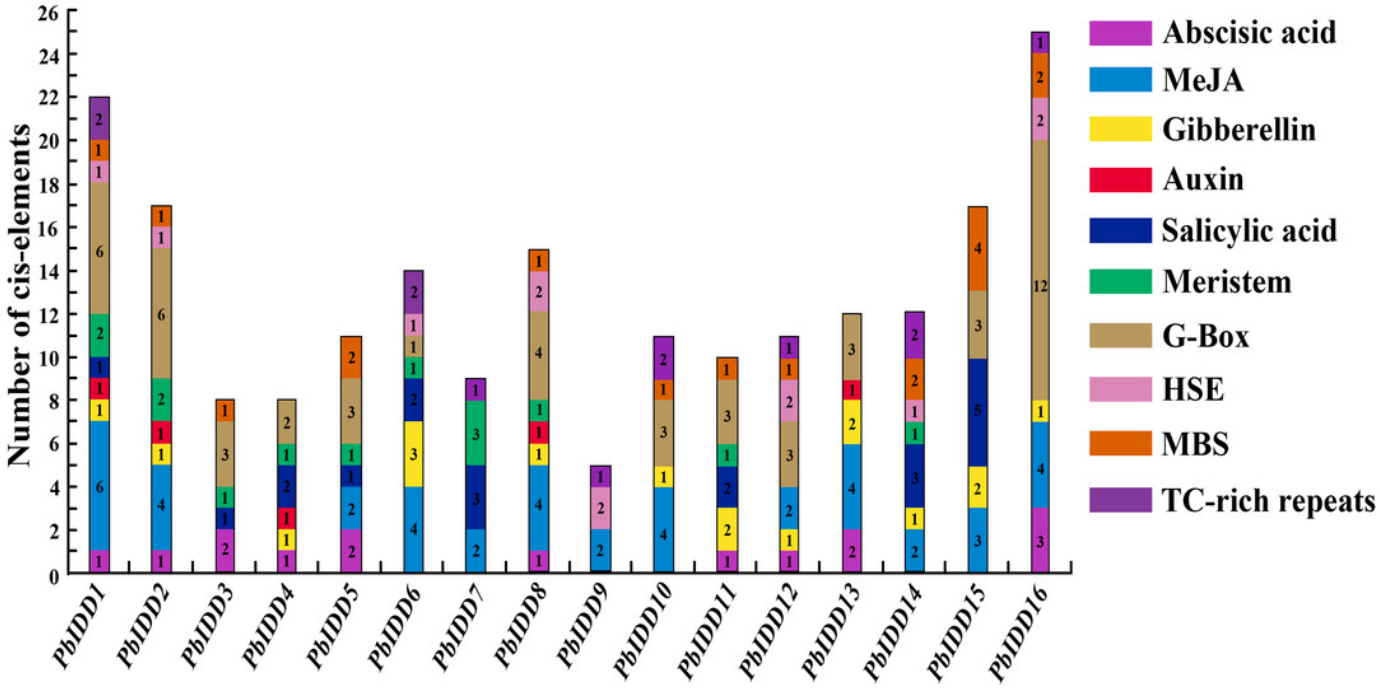
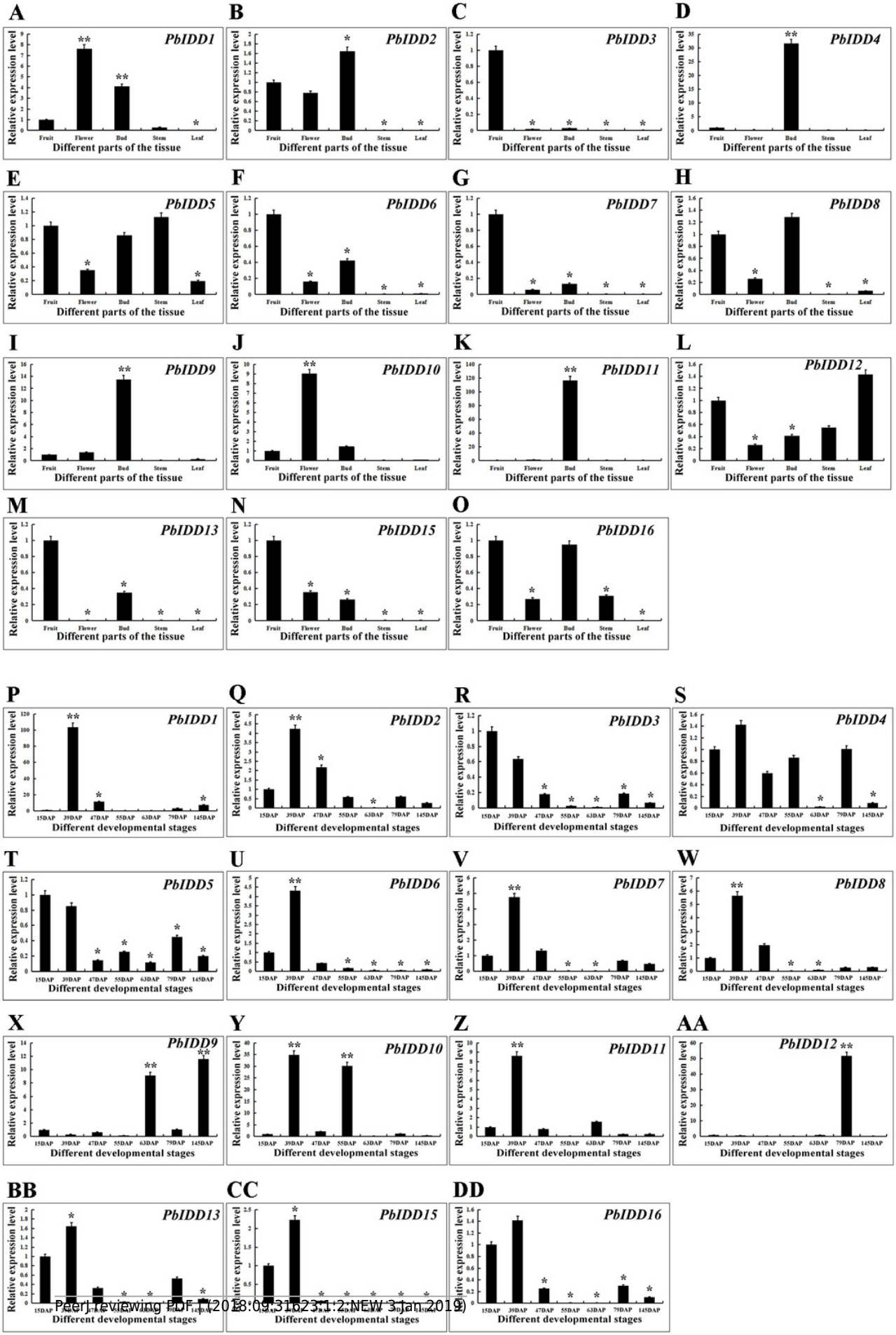


Figure 7

Expression patterns of IDD genes of Chinese white pear in different tissues and in fruit at different developmental stages.

Expression patterns of IDD genes of Chinese white pear in different tissues and in fruit at different developmental stages. Expression patterns of IDD genes in Chinese white pear in different tissues (A-O). Expression patterns of IDD genes in Chinese white pear at different developmental stages (P-DD). *significant difference at $P < 0.05$, **significant difference at $P < 0.01$.


A
PbIDD1
B
PbIDD2
C
PbIDD3
D
PbIDD4
E
PbIDD5
F
PbIDD6
G
PbIDD7
H
PbIDD8
I
PbIDD9
J
PbIDD10
K
PbIDD11
L
PbIDD12
M
PbIDD13
N
PbIDD15
O
PbIDD16
Relative expression level
Different parts of the tissue
Fruit
Flower
Bud
Stem
Leaf
P
PbIDD1
Q
PbIDD2
R
PbIDD3
S
PbIDD4
T
PbIDD5
U
PbIDD6
V
PbIDD7
W
PbIDD8
X
PbIDD9
Y
PbIDD10
Z
PbIDD11
AA
PbIDD12
BB
PbIDD13
CC
PbIDD15
DD
PbIDD16
Different developmental stages
15DAP
39DAP
47DAP
55DAP
63DAP
79DAP
145DAP

My tips

Line 17: " Chinese White *Pyrus bretschneideri*" should be " Chinese White pear *Pyrus Bretschneideri*".

Latin scientific names of a species appearing for the second time in the full text should be abbreviated. (*P. bretschneideri*) Check the whole text.

Line 60: "However, there are few reports on fruit development". There are many papers on fruit development and stone cell development in pears.

Line 78-80: I can't see the Logical Relations between the sentences.

Line 129-130: Rather than citing a reference, it should be clearly described about WHEN (dormant bud?) and WHICH bud (Flower bud ?) were treated? And OW was the bud treated (directly being sprayed or added in the solution)?

Line 134-135: As I noted in the PDF version in the first revision, " The concentration of PCR reaction is not correctly described. The reaction procedure should be described in detail. The standard gene used for control is not shown" .

Line 221: I did not find Table S8A and S8B.

Line 254: "No expression of *PbIDD14* was detected in all (any) tissues of Chinese white pear" should be in the above paragraph. Besides, tissues shown in Fig. 7(A-O) were not clearly described in the Materials section (Line 128). Flower bud or leaf bud? Fruit of which stage?

Line 253: IDD4 is missing in this paragraph. And the putative function (maybe there is no function) of IDD4 is not mentioned in the Discussion.

Line 290-292: Again, I did not see the reasonable explanation about the number of IDD genes in apple and pear. You should do some comparison by investigating the phylogenetic tree in Fig1, and also read and cite some published papers on gene family evolutionary histories.

Line 293-294: "No homologous genes were found in *Arabidopsis thaliana* and the 6 Rosaceae species, suggesting that the ID1(?) gene may be unique to gramineous plants." Please check this sentence based on Fig.1.

Line 313-316: Again, the author still tried to explain why 1+1=2 by saying because 2=1+1.

As a DISCUSSION, the authors need to read and cite papers to do some reasonable explanations in terms of gene evolution.
For this instance, the author should read papers on the function of IDDS in group 4 (AtIDD15, OsIDD14) and try to explain why no member of MdIDDs was found in group 4? What are the “Specific biologic processes”?

Lime339-340: Please check and correct this sentence (also this paragraph) according to the results.

In the last paragraph about PbIDD3 and IDD5, the author madeabundant discussion based on the expression pattern compared with CAD to strongly speculate the possible function of these two genes. Why not analyze the sequence characteristics of these two genes found in this study to strengthen the speculatio? Besides, it is also possible to analyze the interaction between the CAD and IDD using bioinformatics methods. It will be also better to propose what the authors are going to do next(using laboratory methods) as an ending of the paragraph.

Line331-341 (Fig.S10) is a little rough without drawing the evolutionary processes of orthologs and paraologs and possible function divergence. Besides, the specie scale and gap in the figure is two large. What I suggested before is to draw such kind of picture about IDDS in Rosaceae (subfamily) with detail about the sequence characteristics, putative function or pathways and gene duplication and loss.

By the way, it is not necessary to accept all of my suggestions, because it depends largely on the authors’ concern and effort on the quality of the paper and understanding of the evolutionary history of gene family