

Genome-wide analysis suggests high level of microsynteny and purifying selection affect the evolution of *EIN3/EIL* family in Rosaceae

Yunpeng Cao¹, Yahui Han², Dandan Meng¹, Dahui Li¹, Qing Jin¹, Yi Lin¹ and Yongping Cai¹

¹ College of Life Sciences, Anhui Agricultural University, Hefei, China

² State Key Laboratory of Tea Plant Biology and Utilization, Anhui Agricultural University, Hefei, China

ABSTRACT

The ethylene-insensitive3/ethylene-insensitive3-like (*EIN3/EIL*) proteins are a type of nuclear-localized protein with DNA-binding activity in plants. Although the *EIN3/EIL* gene family has been studied in several plant species, little is known about comprehensive study of the *EIN3/EIL* gene family in Rosaceae. In this study, ten, five, four, and five *EIN3/EIL* genes were identified in the genomes of pear (*Pyrus bretschneideri*), mei (*Prunus mume*), peach (*Prunus persica*) and strawberry (*Fragaria vesca*), respectively. Twenty-eight chromosomal segments of *EIL/EIN3* gene family were found in four Rosaceae species, and these segments could form seven orthologous or paralogous groups based on interspecies or intraspecies gene colinearity (microsynteny) analysis. Moreover, the highly conserved regions of microsynteny were found in four Rosaceae species. Subsequently it was found that both whole genome duplication and tandem duplication events significantly contributed to the *EIL/EIN3* gene family expansion. Gene expression analysis of the *EIL/EIN3* genes in the pear revealed subfunctionalization for several *PbEIL* genes derived from whole genome duplication. It is noteworthy that according to environmental selection pressure analysis, the strong purifying selection should dominate the maintenance of the *EIL/EIN3* gene family in four Rosaceae species. These results provided useful information on Rosaceae *EIL/EIN3* genes, as well as insights into the evolution of this gene family in four Rosaceae species. Furthermore, high level of microsynteny in the four Rosaceae plants suggested that a large-scale genome duplication event in the *EIL/EIN3* gene family was predated to speciation.

Submitted 20 January 2017

Accepted 10 May 2017

Published 31 May 2017

Corresponding author
Yongping Cai, ypcaiah@163.com

Academic editor
Marion Röder

Additional Information and
Declarations can be found on
page 14

DOI 10.7717/peerj.3400

© Copyright
2017 Cao et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Bioinformatics, Genomics, Plant Science

Keywords Rosaceae, *EIN3/EIL*, Microsynteny, Purifying selection, qRT-PCR

INTRODUCTION

Rosaceae species such as pear (*Pyrus bretschneideri*), mei (*Prunus mume*), peach (*Prunus persica*) and strawberry (*Fragaria vesca*) are important perennial trees cultivated for the commercial production of fruits available worldwide. According to previous studies, the genomes of strawberry ($X = 7$), mei ($X = 8$), peach ($X = 8$), and pear ($X = 17$) shared an ancestor, which had nine pairs of chromosomes (Shulaev et al., 2011; Verde et al., 2013; Wu et al., 2012; Zhang et al., 2012). Recently, the researchers confirmed that chromosome

inversions, fusions, and translocations played an important role in the evolution of the Rosaceae genome (Illa *et al.*, 2011). Some extant “diploid” species of Rosaceae family are originated from their polyploid ancestors, others are actually thought to be true polyploids (Wendel, 2000). These studies indicate that the diploid species in Rosaceae have evolved with a complex history. There are several gene families which share highly conserved genome sequences with each other among the related species of family Rosaceae, as well as other taxonomic families. In this study, the *EIL/EIN3* gene family was selected to investigate the specific evolutionary relationships among the related species of family Rosaceae.

The *EIN3/EIL* gene family is a relatively small one in higher plants. Some *EIN3/EIL* genes have been isolated from *Arabidopsis thaliana* (Chao *et al.*, 1997), tobacco (Kosugi & Ohashi, 2000; Rieu, Mariani & Weterings, 2003), banana (Jourda *et al.*, 2014), tomato (Tieman *et al.*, 2001; Yokotani *et al.*, 2003) and rice (Hiraga *et al.*, 2009; Mao *et al.*, 2006). These plant-specific *EIN3/EIL* proteins are located in the nuclei, with the highly conserved amino acid sequences at the N-termini, including several important structural features, such as acidic amino acid regions, proline-rich regions and 5-basic amino acid clusters (BD I–V) (Chao *et al.*, 1997). Compared to the N-terminal sequences, the conservation of their C-termini is lower. For example, it was found that although asparagine-enriched regions or glutamine-enriched regions were commonly distributed within the C-terminal sequences of *EIN3/EILs* in *Arabidopsis*, carnation and mung beans (Chao *et al.*, 1997; Lee & Kim, 2003; Waki *et al.*, 2001), they did not widely exist in other *EIN3/EIL* members, such as tobacco *NtEILs* (Rieu, Mariani & Weterings, 2003).

Functions of the *EIN3/EIL* gene family have been studied in several plants, such as *Hevea brasiliensis* (Yang *et al.*, 2015) and tomato (Tieman *et al.*, 2001; Yokotani *et al.*, 2003). Recently, research on the application of comparative genome in the analysis of evolution and function of gene family have been reported. For example, based on the comparative genomic analysis, Wang *et al.* (2015) explored the evolution and functional differences of WRKY type-III transcription factor family of poplar, grape, *Arabidopsis* and rice. Jing *et al.* (2016) explored the evolution of WRKY I subfamily in Gramineae. However, there is still lack of specific evolutionary relationships of the *EIN3/EIL* gene family in Rosaceae. To address this question, the evolutionary relationships and gene duplication events of *EIN3/EIL* genes from Rosaceae species, including pear (*Pyrus bretschneideri*), mei (*Prunus mume*), peach (*Prunus persica*) and strawberry (*Fragaria vesca*), were analyzed, based on their phylogenetic relationships, microsynteny and environmental selection pressures analysis. In addition, the expression patterns of pear *EIN3/EIL* genes were investigated on a variety of organs/tissues including fruits at several developmental stages. The results obtained from this study provided valuable information about *EIN3/EIL* genes that will aid future functional research involved in many important biological processes of this important gene family in flowering plants, especially in the pear.

MATERIALS AND METHODS

Sequence identification and collection

The genome data of four Rosaceae species were obtained from their respective genome sequence websites: *Pyrus bretschneideri* from the GigaDB database (<http://gigadb.org/site/>

[index](#)); *Prunus mume* from the Genome Database for Rosaceae (<http://www.rosaceae.org/>); *Prunus persica* from the Phytozome database (<https://phytozome.jgi.doe.gov/pz/portal.html>) and *Fragaria vesca* from the Joint Genome Institute (<http://www.jgi.doe.gov/>). The Hidden Markov Model (HMM) profiles of EIN3/EIL domain (PF04873) (*Chang & Shockey, 1999; Chao et al., 1997*) were obtained from the Pfam database (<http://pfam.xfam.org>) (*Finn et al., 2006*). The EIN3/EIL domain was used as query sequences to identify *EIN3/EIL* genes in four Rosaceae species by using DNAtools software (E -value < 0.001). To verify the *EIN3/EIL* genes in four Rosaceae genomes, all putative proteins were validated by searching for the EIN3/EIL domain using the InterPro online tool (<http://www.ebi.ac.uk/interpro/>) (*Zdobnov & Apweiler, 2001*) and SMART database (<http://smart.embl-heidelberg.de/>) (*Letunic, Doerks & Bork, 2012*). In our study, only the EIN3/EIL domain-containing sequences were retained.

Chromosomal location of *EIN3/EIL* genes

The genome annotation information was collected from GigaDB database (<http://gigadb.org/site/index>), Genome Database for Rosaceae (<http://www.rosaceae.org/>), Phytozome database (<https://phytozome.jgi.doe.gov/pz/portal.html>) and Joint Genome Institute (<http://www.jgi.doe.gov/>), respectively. Subsequently, the MapInspect software (<http://mapinspect.software.informer.com/>) was used for data visualization.

Gene structure and motif analysis

The exon-intron structure of each *EIN3/EIL* gene was determined by alignment of its CDS and genomic DNA sequence. Then a diagram was constructed using the Structure Display Server website (*Hu et al., 2014*). Subsequently, the Online MEME server was used to screen the conserved motifs encoded by *EIN3/EIL* genes. Additionally, the Pfam website (*Punta et al., 2012*) and SMART tools (*Letunic, Doerks & Bork, 2012*) were used to annotate these structural motifs.

Phylogenetic analysis of *EIN3/EIL* genes

EIN3/EIL sequences were aligned using ClustalX version 1.83 (*Thompson et al., 1997*) and evolutionary relationships were inferred by analyzing an unrooted phylogenetic tree using MEGA 5 and neighbor-joining (NJ) method (*Tamura et al., 2011*) with the following parameters: poisson correction, pairwise deletion and 1,000 bootstrap replicate.

Microsynteny analysis

In order to reveal the sequence features of the *EIN3/EIL* gene-containing regions, microsynteny analysis was performed across the four Rosaceae species using MCScanx (Multiple Collinearity Scan toolkit) (*Wang et al., 2012*) with the gene identifier file, the gene list file and the coding sequence file. Subsequently, a syntenic block was defined as a region containing three or more conserved homologs which were located within 100-kb downstream and upstream of protein-coding sequences.

Environmental selection pressure analysis

The nucleotide coding sequences from segmentally duplicated pairs were aligned by Clustal X (Thompson *et al.*, 1997). Then DnaSP (version 5.10) was used to calculate the nonsynonymous (Ka) and synonymous (Ks) substitution rates of the homologues (Librado & Rozas, 2009). For each pair of duplicated regions, we estimated the mean Ks values of the flanking conserved genes for individual homologs. To further understand the selective pressure experienced by *EIN3/EIL* genes, Ka, Ks and Ka/Ks ratios were estimated using sliding window (with parameters: window size, 150 bp; step size, 9 bp) over the entire aligned length (Cao *et al.*, 2016a; Han *et al.*, 2016).

EIN3/EIL gene expression analysis in pear different tissues

To verify the expression patterns of *EIN3/EIL* genes, qRT-PCR analysis was carried out. The first-strand cDNA was synthesized with Oligo18dT primer (Table S1) by using M-MLV reverse transcriptase (TakaRa, Japan) following the manufacture introduction. The TransStart Tip Green qPCR SuperMix (TransGen Biotech, China) with SYBR Green I as the fluorescent dye was used for the qPCR, employing a Bio-rad CFX96 Touch™ Deep Well Real-Time PCR Detection system (BioRad, USA). The transcript level relative to the *Pyrus tubulin* gene (Wu *et al.*, 2012) was estimated according to a previous workflow (Cao *et al.*, 2016a; Cao *et al.*, 2016b). For each sample, three replicates were set up in parallel experiments.

RESULTS AND DISCUSSION

Identification of *EIN3/EIL* genes in Rosaceae

The genome data of pear (*P. bretschneideri*), mei (*P. mume*), peach (*P. persica*) and strawberry (*F. vesca*) were recently published, respectively (Shulaev *et al.*, 2011; Verde *et al.*, 2013; Wu *et al.*, 2012; Zhang *et al.*, 2012). To identify the members of the *EIN3/EIL* gene families in these species, *EIN3/EIL* specific domain (PF04873) was used to perform Blastp searches of the local protein databases. Sequences identified were verified for *EIN3/EIL* domains through SMART database and InterPro online tool. In total 24 of *EIN3/EIL* genes were identified, including ten in pear, four in peach, five in mei and five in strawberry, and named as *PbEIL1-PbEIL10*, *PpEIL1-PpEIL4*, *PmEIL1-PmEIL5* and *FvEIL1-FvEIL5*, according to their locations in chromosome, respectively (Table 1 and Fig. 1). This result suggested that *EIN3/EIL* gene family was relatively small compared to other gene families in the studied species. Similar indication was also reported by the previous studies in which six, five, four, six and 17 *EIN3/EIL* genes were found in *Arabidopsis thaliana* (Chao *et al.*, 1997; Guo & Ecker, 2004), tobacco (Kosugi & Ohashi, 2000; Rieu, Mariani & Weterings, 2003), tomato (Tieman *et al.*, 2001; Yokotani *et al.*, 2003), rice (Hiraga *et al.*, 2009; Mao *et al.*, 2006) and banana (Jourda *et al.*, 2014), respectively. Furthermore, it was found that the genome sizes and number of *EIN3/EIL* gene family members appeared not to have a direct relevance. For example, although there was no significant variety in genome size of pear (271.9 Mb) and strawberry (240 Mb), the number of *EIN3/EIL* genes obviously changed. Contrarily, the number of *EIN3/EIL* genes of the peach (224.6 Mb) and strawberry (240 Mb) had a corresponding relationship with their genome size. Remarkably, compared with

Table 1 List of *EIN3/EIL* genes identified in pear, peach, mei and strawberry.

Name	Gene model	Chromosome	5' end	3' end
<i>FvEIL1</i>	mrna25474.1	Chr1	17653671	17655527
<i>FvEIL2</i>	mrna16361.1	Chr1	18891616	18892965
<i>FvEIL3</i>	mrna20650.1	Chr3	29248944	29253704
<i>FvEIL4</i>	mrna00379.1	Chr7	290967	292781
<i>FvEIL5</i>	mrna00392.1	Chr7	349495	351202
<i>PmEIL1</i>	Pm001950	Chr1	15239829	15241073
<i>PmEIL2</i>	Pm002057	Chr1	16248534	16250294
<i>PmEIL3</i>	Pm017009	Chr5	6907428	6909233
<i>PmEIL4</i>	Pm017011	Chr5	6933006	6934874
<i>PmEIL5</i>	Pm028171	scaffold103	1235430	1246520
<i>PpEIL1</i>	ppa003493m	Chr2	5516222	5518429
<i>PpEIL2</i>	ppa003550m	Chr2	5549949	5552334
<i>PpEIL3</i>	ppa003113m	Chr6	3882268	3885188
<i>PpEIL4</i>	ppa016118m	Chr6	16979366	16982360
<i>PbEIL1</i>	Pbr024739.1	Chr2	8493409	8495211
<i>PbEIL2</i>	Pbr024740.1	Chr2	8506285	8508129
<i>PbEIL3</i>	Pbr000646.1	Chr3	18718500	18721454
<i>PbEIL4</i>	Pbr026603.1	Chr8	3598382	3602224
<i>PbEIL5</i>	Pbr004535.1	Chr11	22794386	22798326
<i>PbEIL6</i>	Pbr033210.1	Chr15	31009957	31014042
<i>PbEIL7</i>	Pbr010447.1	scaffold170.2.1	239188	246113
<i>PbEIL8</i>	Pbr010448.1	scaffold170.2.1	259361	262132
<i>PbEIL9</i>	Pbr022557.1	scaffold341.0	56672	57973
<i>PbEIL10</i>	Pbr039294.1	scaffold837.0	82840	84144

Notes.

Pear gene models are found in the GigaDB Genome database; mei and peach gene models are found in the Rosaceae Genome Database; strawberry gene models are found in the Phytozome database.

those in peach, mei and strawberry, the numbers of *EIN3/EIL* genes in pear were found to be almost doubled. Moreover, the chromosome numbers of peach, mei and strawberry are 16, 16 and 14, respectively (*Shulaev et al., 2011; Verde et al., 2013; Wu et al., 2012; Zhang et al., 2012*), whereas the chromosome number of pear is 34, indicating that the *EIN3/EIL* gene family has undergone an expansion corresponding to the variation in chromosome number. However, a recent whole genome duplication event (30–45 million years ago) (*Wu et al., 2012*) that occurred in pear but not in peach, mei and strawberry probably contributed to the expansion of *EIN3/EIL* gene family in the pear.

To determine the distribution of *EIN3/EIL* genes on chromosomes among pear, peach, mei and strawberry, respectively, a chromosome map (*Fig. 1*) was drawn based on genome annotation (*Wu et al., 2012*). In pear, two *EIN3/EIL* genes were located on chromosome 2, and one gene on chromosome 3, 8, 11 and 15, respectively, with the remaining genes localized on different scaffold regions (*Fig. 1C*). In peach, two *EIN3/EIL* genes were found on chromosome 2 and 6, respectively (*Fig. 1D*). In mei, two *EIN3/EIL* genes were distributed on chromosome 1 and 5, respectively, with the remaining one localized on a scaffold region

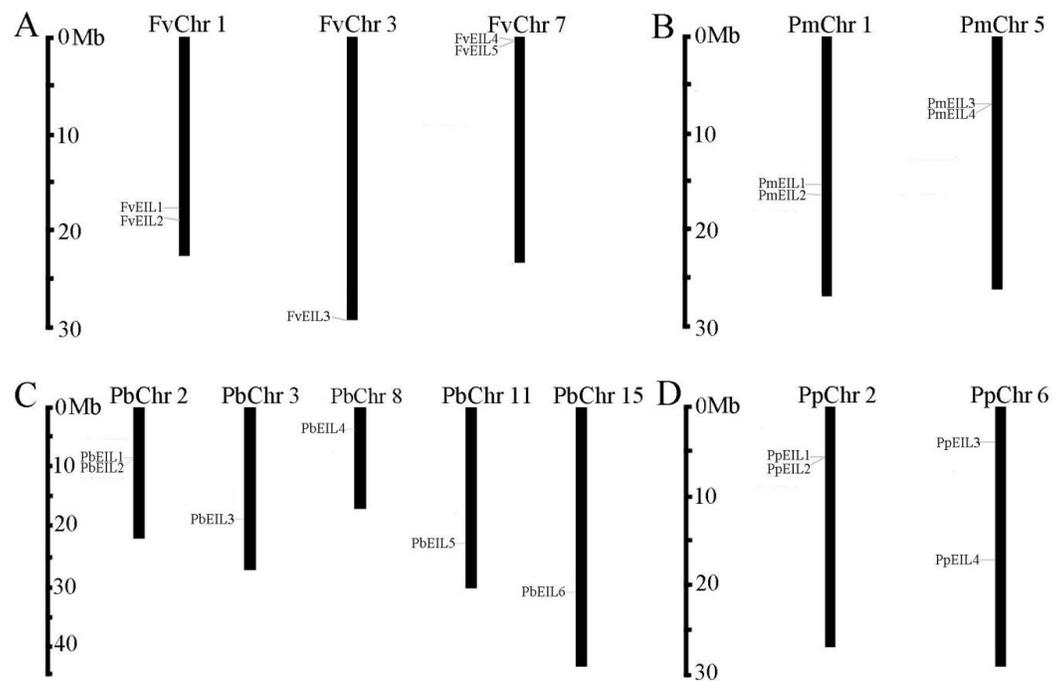


Figure 1 Chromosomal location of *EIN3/EIL* genes in the genomes of strawberry (A), mei (B), pear (C) and peach (D). The distribution of *EIN3/EIL* genes among the chromosomes in each species was diverse. The chromosome number was represented at the top of each chromosome. The left scale indicates the megabases (Mb).

(Fig. 1B). In strawberry, two *EIN3/EIL* genes were distributed on chromosome 1 and 7, respectively, with the remaining one localized on chromosome 3 (Table 1 and Fig. 1A).

Phylogenetic analysis of *EIN3/EIL* genes

The phylogenetic tree containing *EIN3/EIL* gene homologs from a variety of species, including *Arabidopsis thaliana*, rice, banana, sorghum, maize, *Brachypodium distachyon*, *Thellungiella parvula*, and four Rosaceae species, was also constructed (Fig S1). As shown in the phylogenetic tree, most *EIN3/EIL* genes from four Rosaceae species were clustered together. To further understand the evolutionary history of *EIN3/EIL* genes in Rosaceae, phylogenetic analysis was carried out using the neighbor joining (NJ) method. As shown in Fig. 2, 24 *EIN3/EIL* sequences were divided into two subfamilies, designated as A and B, which contained four classes (Classes A1, A2, B1 and B2). Classes A1, B1 and B2 were composed of the *EIN3/EIL* genes from the four species (pear, peach, mei and strawberry), while the Class A2 contained the members only from pear. According to our study, a whole genome duplication event happened 30–45 million years ago in pear, but not in peach, mei and strawberry (Wu et al., 2012). This result suggested the probable reason for occurrence of Class A2 genes in pear. Remarkably, *EIN3/EIL* genes from peach and mei showed higher similarity with each other according to genetic distance, which was consistent with a previous study reporting that the closer relationship between peach and mei versus peach and pear/strawberry (Cao et al., 2016a; Cao et al., 2016b; Cao et al., 2016d).

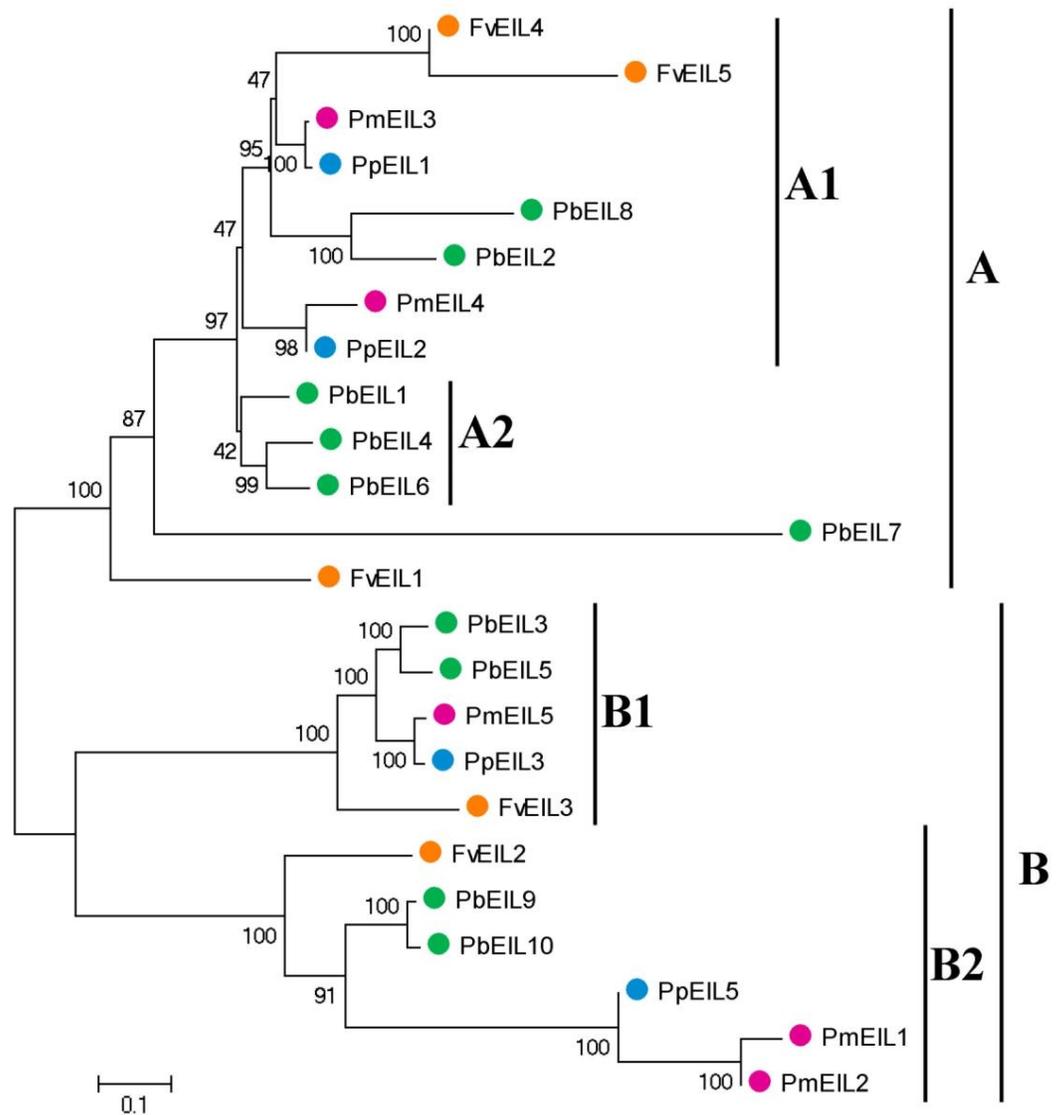


Figure 2 Phylogenetic tree of *EIN3/EIL* proteins from pear, peach, mei and strawberry. The neighbor-joining (NJ) tree was constructed by using MEGA 5 software. The neighbor-joining (NJ) tree was constructed by using MEGA 5 software (1,000 bootstrap replicates). The different colors suggest the different species background for each *EIN3/EIL* protein. Gene names are listed in Table 1. The scale bar represents 0.1 amino acid changes per site.

Each of the four Rosaceae species contributed at least one member of the *EIN3/EIL* gene to each class, with the exception for Class A2 (Fig. 2). Therefore, we deduced that *EIN3/EIL* genes had rapidly been duplicated before these dicotyledon species diverged. However, only the *EIN3/EIL* genes in class A2 revealed an internal duplication. In addition, we identified three pairs of orthologous genes among the *EIN3/EIL* genes: *PmEIL3* and *PpEIL1*, *PpEIL2* and *PmEIL4*, *PpEIL3* and *PmEIL5* based on the phylogenetic analysis.

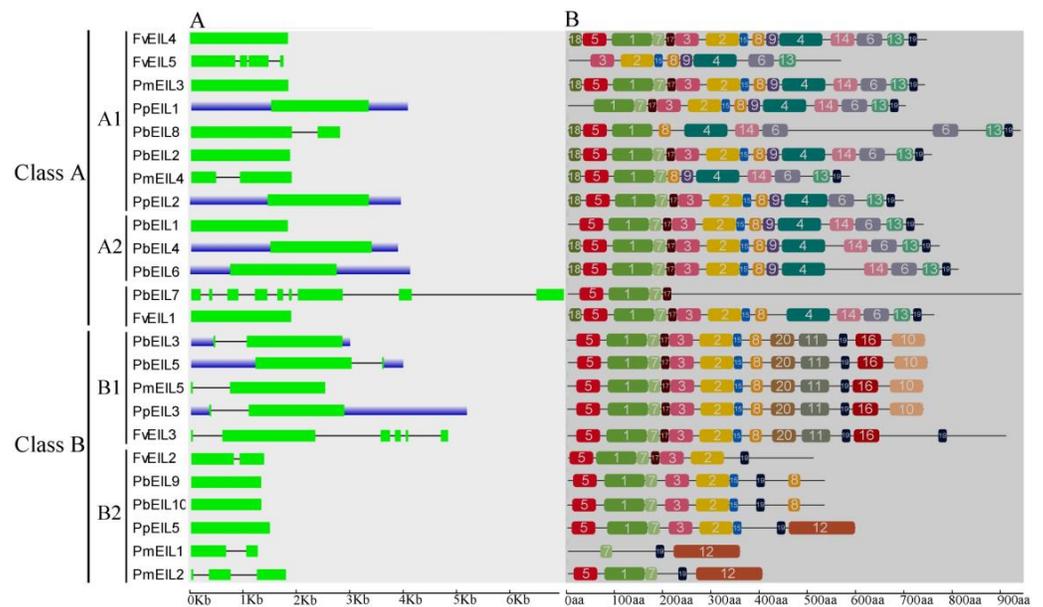


Figure 3 Gene structure (A) and conserved motif compositions (B) of *EIN3/EIL* genes in Rosaceae species. Untranslated regions (UTRs), introns and exons are represented by blue boxes, thin lines and green rectangles, respectively. Note that the gene or protein lengths can be estimated by using the scale at the bottom.

Structural analysis of *EIN3/EIL* genes

Previous studies have suggested that gene structural diversity is the primary resource for the evolution of multigene families (Cao *et al.*, 2016c; Leitch & Leitch, 2013; Mercereau-Puijalon, Barale & Bischoff, 2002). To characterize the structural diversity of the *EIN3/EIL* gene family, exon-intron organization of each *EIN3/EIL* gene was analyzed. As shown in Fig. 3A, most genes did not contain introns, such as *FvEIL1*, *PbEIL1*, *PmEIL3* and *PpEIL4* et al. Furthermore, *PbEIL7* contained eight introns, followed by *FvEIL3* (five), whereas *FvEIL5* had three introns, *PmEIL2* had two introns and eight *EIN3/EIL* genes contained one intron (Fig. 3A). These results implied that the intron/exon loss and acquire has occurred in the evolution of the *EIN3/EIL* gene family, which may be able to explain the functional divergence of closely related *EIN3/EIL* genes. In the present study, the gene structures of the *EIN3/EIL* homologous gene pairs were investigated. We found that the exon number of two gene pairs (*PbEIL2/PbEIL8* and *PmEIL1/PmEIL2*) had changed. Further analysis indicated that *PbEIL8* and *PmEIL2* obtained one exon during evolution, while *PbEIL2* and *PmEIL1* lost one exon. These diversities might be due to single intron loss or obtain events during evolution.

Because 24 *EIN3/EIL* genes did not have high similarity, MEME web server was used to find conserved motifs. Subsequently, we identified 20 conserved motifs, which were shown in Table S2 and Fig. 3B. The Pfam and SMART databases were used to annotate the individual of the putative motifs. Motif 1 and motif 2 were identified to encode a conserved *EIN3/EIL* domain (Chang & Shockey, 1999; Chao *et al.*, 1997), whereas the remaining motifs did not get function annotation. Most *EIN3/EIL* proteins have motifs 1, 2, 4, 7, 10 and 19. In addition, several proteins from clade B2 contained unique motif 12,

which might imply its specific functions. Remarkably, most of the closely related EIN3/EIL proteins in the same clade exhibited similar motif compositions (e.g., PmEIL5/PpEIL3 and PbEIL3/PbEIL5) indicating their functional similarity among these EIN3/EIL proteins. Summarily, the similarity in motif distribution and exon-intron structure of most EIN3/EIL proteins supported the results from phylogenetic analysis of the *EIN3/EIL* genes, whereas the differences of the related characteristics in the different classes indicated that their functions were diversified.

Conserved microsynteny of *EIN3/EIL* genes in the four Rosaceae species

Based on the whole-genome data, species microsynteny can be used to identify the location of orthologous genes and/or paralogous genes (Cao *et al.*, 2016d; Lin *et al.*, 2014). To identify the homologous genes (orthology or paralogy) within the *EIN3/EIL* genes from four Rosaceae species (pear, peach, mei and strawberry), as well as their evolutionary history, microsynteny analysis was performed. By pairwise comparisons of flanking sequences in the chromosomal regions containing *EIN3/EIL* genes, three or more pairs were present in this region, which were considered as either conserved microsynteny or high levels of microsynteny.

In this study, a total of 55 flanking sequences containing *EIN3/EIL* genes could be assembled into 28 regions and divided into seven microsynteny groups. It was supposed that *EIN3/EIL* genes from the same group should evolve from the most recent common ancestor. Based on this criterion, orthology and/or paralogy relationships, as well as their evolutionary origins, were detected among *EIN3/EIL* genes of the four Rosaceae species. Nine, five, four and four out of the seven microsynteny groups are from pear, peach, mei and strawberry, respectively (Fig. 4). In class B2, two gene pairs (*PbEIL1* and *PbEIL9*, *PbEIL9* and *PbEIL10*) with both a higher level of microsynteny and a noticeable inverted duplication, were identified. Interestingly, it was also found that some duplication rules in several regions were in disorder, such as *FvEIL1* and *FvEIL2*, *FvEIL1* and *PpEIL4* (Fig. 4). Similar microsynteny was identified in other classes with a concordant inverted microsynteny (Fig. 4). In addition, according to the constructed phylogenetic tree, conservation of microsynteny between different families appeared gradually. However, some flanking genes in each microsyntenic group were not conserved, indicating that they arose later than this duplication event (Fig. 4). Furthermore, we only identified four pairs of intraspecies microsynteny groups from pear (*PbEIL1* and *PbEIL9*, *PbEIL3* and *PbEIL5*, *PbEIL4* and *PbEIL6*, *PbEIL9* and *PbEIL10*), but peach, mei and strawberry were excluded (Fig. 5). This difference might result from the expansion of pear *EIN3/EIL* genes. However, no similar gene expansion was identified in peach, mei and strawberry. Some previous studies has hypothesized that transcription factors should be generally and preferentially retained after genome duplications (Blanc & Wolfe, 2004), with a lower frequency of tandem duplication events in a number of transcription factors (Freeling, 2009). Additionally, genes from whole-genome duplication events are more easily retained into genomes. With the stoichiometric relationships, these genes were strongly retained by stabilizing selection (Lynch & Conery, 2000). Our results were not only consistent with this hypothesis, but was also strong evidence for it.

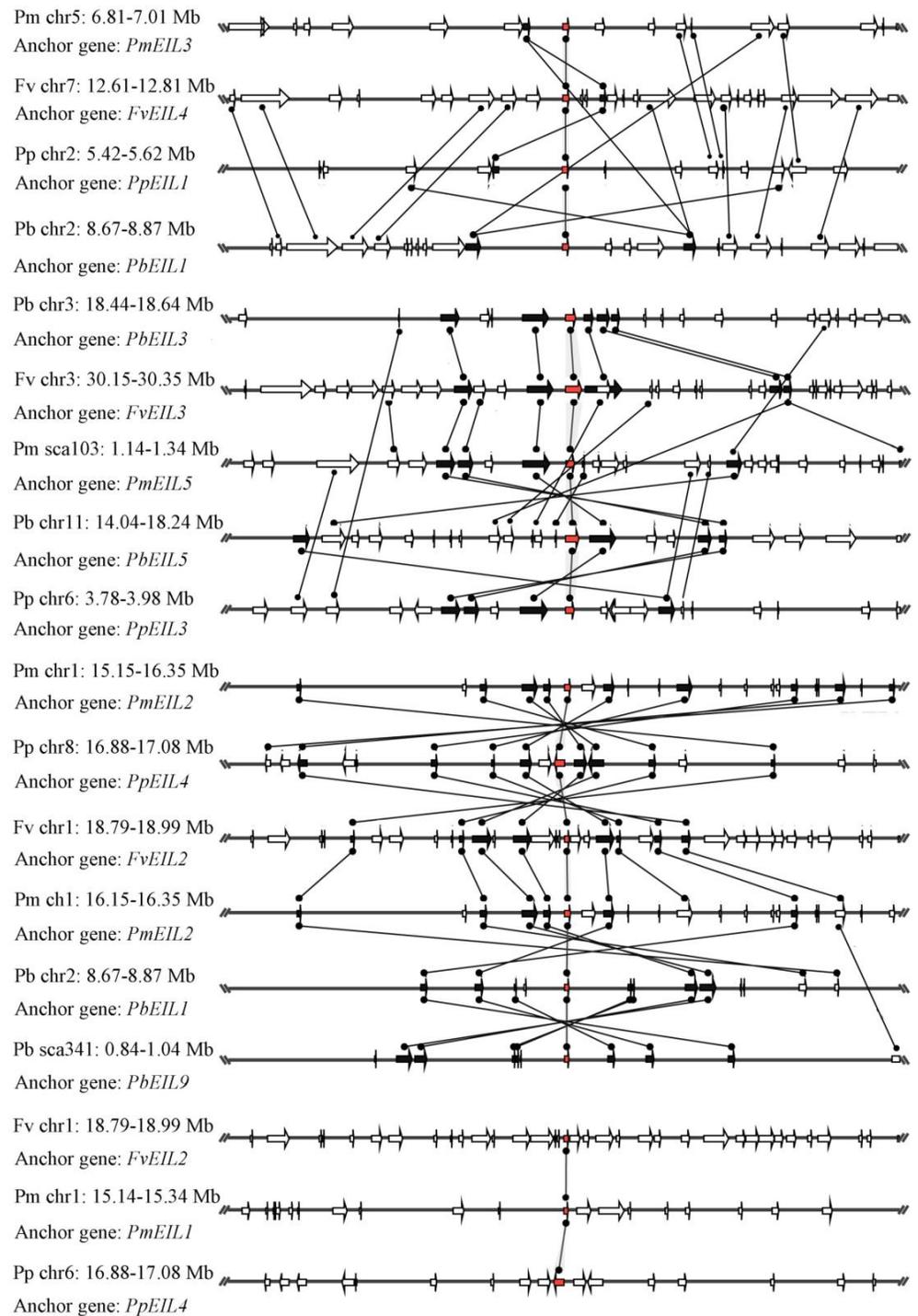


Figure 4 Interspecies microsynteny related to *EIN3/EIL* families in four Rosaceae. The relative positions of all flanking protein-coding genes were defined by anchored *EIN3/EIL* genes, highlighted in red. The gene's orientations are shown as triangle, with gray lines corresponding to chromosomal segments.

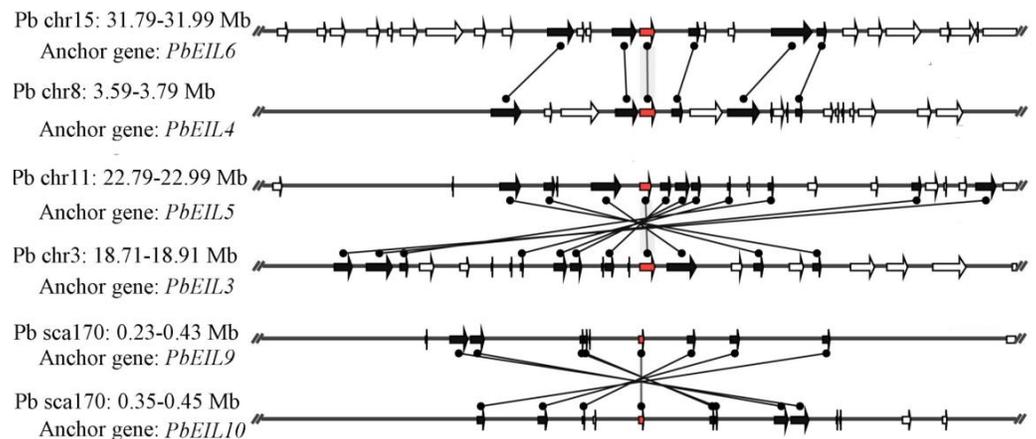


Figure 5 Intraspecific microsynteny related to *EIN3/EIL* families with the same species. The relative positions of all flanking protein-coding genes were defined by anchored *EIN3/EIL* genes, highlighted in red. The gene's orientations are shown as triangle, with gray lines corresponding to chromosomal segments.

Table 2 The relative syntenic quality of *EIN3/EIL* genes in four Rosaceae plants.

	Clade A1	Clade A2	Clade B1	Clade B2	Average
Pb-Pp			22.50%	44.07%	33.29%
Pb-Pm			21.51%	41.18%	31.35%
Pb-Fv			24.07%	28.26%	26.17%
Pp-Pm	26.67%		40.00%	29.73%	32.13%
Pp-Fv	10.26%				10.26%
Pm-Fv	10.00%		26.32%	3.77%	13.36%
					24.43%

Notes.

The relative syntenic quality was estimated as twice the number of matches divided by the sum of the total number of genes in both conserved gene regions, based on the previous methods (Cannon *et al.*, 2003; Cannon *et al.*, 2006).

Subsequently, the quality of the synteny was estimated in four Rosaceae plants based on previous research methods (Cannon *et al.*, 2003; Cannon *et al.*, 2006). As shown in Table 2, the relative synteny quality of the *EIN3/EIL* genes from these Rosaceae four species genomes was 24.43% for orthologous regions. The highest value of synteny quality found between pear and peach was 33.29%. And the lower value of synteny quality was obtained between strawberry and peach (10.26%) and mei (13.26%) The relative synteny quality in the pear/mei and pear/strawberry syntenic regions was 31.35% and 26.17%, which was substantially lower than the 32.13% found in the pear/peach syntenic blocks. Our results were essentially consistent with their evolutionary relationship (Xiang *et al.*, 2017; Zhong *et al.*, 2015).

Strong purifying selection for *EIN3/EIL* genes in four Rosaceae species

In general, K_s values can be used to estimate evolutionary data of the whole genome duplication events or segmental duplication events. Previous studies showed that pear

had experienced two whole genome duplication events, including an ancient whole genome duplication ($K_s \sim 1.5\text{--}1.8$) estimated at ~ 140 MYA (Fawcett, Maere & Van de Peer, 2009) and a recent whole genome duplication ($K_s \sim 0.15\text{--}0.3$) estimated at 30–45 MYA (Wu et al., 2012), while peach, mei and strawberry only experienced an ancient whole genome duplication event. Therefore, K_s values were applied to analyze the whole genome duplication or segmental duplication events in *EIN3/EILs* of four Rosaceae species. As shown in Table S3, the mean K_s values of each duplication pairs in the syntenic region were lists. In pear, we found the mean K_s values of *EIN3/EIL* gene pairs were 0.0363, 0.1717 and 0.2836, respectively. It was obvious that these duplications might be resulting from the latest whole genome duplication (30–45 MYA; $K_s \sim 0.15\text{--}0.3$), but an ancient whole genome duplication (~ 140 MYA; $K_s \sim 1.5\text{--}1.8$) in pear.

In addition, the K_a/K_s values are widely used to represent the gene selection pressure and evolution rate [40]: K_a/K_s value with >1 indicates positive selection with accelerated evolution, $K_a/K_s < 1$ indicates negative/purifying selection with the functional constraint, and $K_a/K_s = 1$ suggests that the genes are drifting neutrally. In this study, all paralogs was found with K_a/K_s ratios <1 (Fig. 6), indicating their purifying selection. Furthermore, to better understand the delineate regions of diversifying and purifying selection in the *EIN3/EIL* gene family, a sliding window analysis of the K_a/K_s values between paralogs was performed (Fig. 4); the *EIN3/EIL* domains in the N-termini exhibited stronger purifying selection compared with the whole gene regions (C-termini). These results suggested that the *EIN3/EIL* genes had undergone strongly purifying selection, especially for *EIN3/EIL* domains in the N-termini (Fig. 4Q). Overall, strong evolutionary constraints were involved in *EIN3/EIL* gene evolution, which may contribute to their functional stability. On the other hand, some parts of protein-coding genes had undergone positive selection, implying the generation of innovative gene functions.

Expression profiles analysis of *PbEIL* genes in different tissues

To increase our understanding of the potential functions of pear *EIN3/EIL* genes during development, qRT-PCR analysis was carried out to determine the expression profiles of ten *PbEIL* genes in different tissues. As shown in Fig. 7 and Table S4, ten pear *EIN3/EIL* genes showed significantly different tissue-specific expression patterns in eight samples from root, stem, leaves and fruits in several development stages. Among the ten pear *EIN3/EIL* genes, three (*PbEIL5*, *PbEIL6* and *PbEIL10*) showed the highest transcript accumulation in the leaves, three (*PbEIL2*, *PbEIL3* and *PbEIL9*) in 145 DAF (days after flowering), two (*PbEIL1* and *PbEIL4*) in 79 DAF, and one (*PbEIL7*) in the roots. Additionally, the duplication gene pairs showed different expression patterns; for example, *PbEIL4* was highly expressed in 79 DAF, while its duplication gene, *PbEIL6*, was expressed at a high level in the leaves. Thus, the pear *EIN3/EIL* duplicates resulting from recent whole genome duplication have different expression patterns in several different tissues, indicating subfunctionalization after duplication. At the same time, this phenomenon was also observed among other *EIN3/EIL* duplication genes (Jourda et al., 2014).

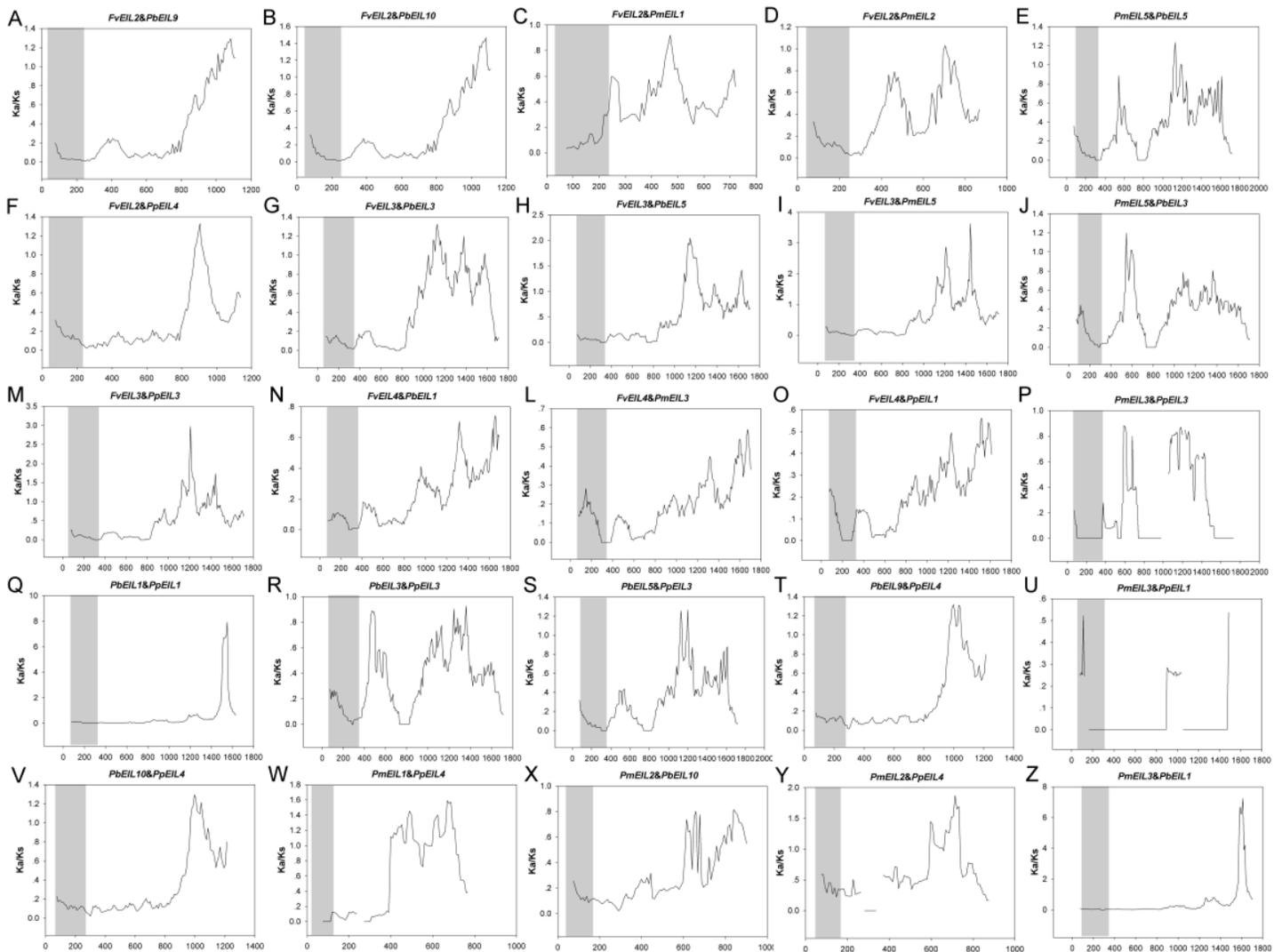


Figure 6 Sliding window plots of duplicated *EIN3/EIL* genes in Rosaceae species. The gray blocks indicate the positions of the *EIN3/EIL* domains. The window size is 150 bp, and the step size is 9 bp. The x-axis denotes the synonymous distances within each gene.

CONCLUSIONS

In this study, we identified 24 *EIN3/EIL* genes from four Rosaceae species (pear, peach, mei and strawberry). Subsequently, a systematic analysis, including their chromosomal location, evolutionary relationship, conserved microsynteny, gene structure and sliding window, was carried out. According to phylogenetic analysis, the *EIN3/EIL* genes divided into four classes. Remarkably, high level of microsynteny of the *EIL/EIN3* family in Rosaceae was found, indicating that the genome duplication plays a key role in the expansion of the *EIL/EIN3* genes in the Rosaceae. In these *EIL/EIN3* genes, all paralogs have experienced purifying selection, especially the *EIL/EIN3* domains in the Rosaceae. Furthermore, the expression profiles of the *PbEIL* genes suggested that the recent whole genome duplication

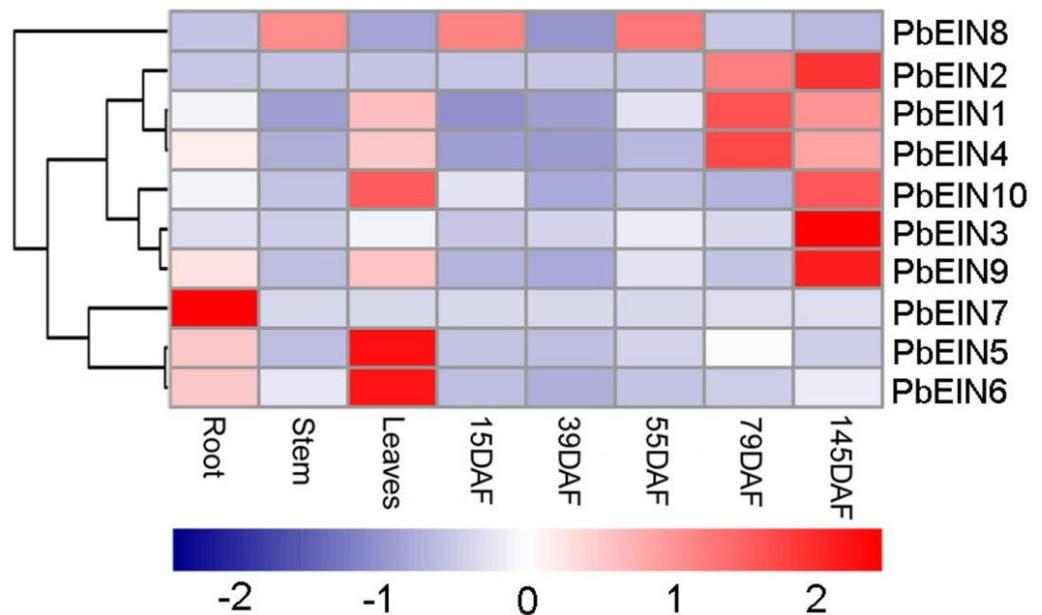


Figure 7 Expression profiling of pear *PbEIL* genes in eight samples from root, stem, leaves and fruits in several development stages. The expression profile data was obtained with qRT-PCR experiments. Blue and red colors indicate low-expression and high-expression, respectively.

derived genes show indications of subfunctionalization. These results may help promote the extrapolation of *EIL/EIN3* gene functions in future.

ACKNOWLEDGEMENTS

We would like to thank Muhammad Abdullah for his careful reading and helpful comments on this manuscript. We extend our thanks to the reviewers and editors for their careful reading and helpful comments on this manuscript.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This study was supported by the National Natural Science Foundation of China (grant 31640068) and 2017 Graduate innovation fund of Anhui Agriculture University (grant # 2017yjs-31). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

National Natural Science Foundation of China: 31640068.

2017 Graduate innovation fund of Anhui Agriculture University: #2017yjs-31.

Competing Interests

The authors declare that they have no conflict of interest.

Author Contributions

- Yunpeng Cao conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- Yahui Han conceived and designed the experiments, analyzed the data, wrote the paper.
- Dandan Meng analyzed the data, wrote the paper, reviewed drafts of the paper.
- Dahui Li conceived and designed the experiments, prepared figures and/or tables.
- Qing Jin and Yi Lin reviewed drafts of the paper.
- Yongping Cai conceived and designed the experiments, analyzed the data.

Data Availability

The following information was supplied regarding data availability:

The raw data has been supplied as [Supplementary File](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.3400#supplemental-information>.

REFERENCES

- Blanc G, Wolfe KH. 2004.** Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *The Plant Cell* **16**:1667–1678
[DOI 10.1105/tpc.021345](https://doi.org/10.1105/tpc.021345).
- Cannon SB, McCombie WR, Sato S, Tabata S, Denny R, Palmer L, Katari M, Young ND, Stacey G. 2003.** Evolution and microsynteny of the apyrase gene family in three legume genomes. *Molecular Genetics and Genomics* **270**:347–361
[DOI 10.1007/s00438-003-0928-x](https://doi.org/10.1007/s00438-003-0928-x).
- Cannon SB, Sterck L, Rombauts S, Sato S, Cheung F, Gouzy J, Wang X, Mudge J, Vasdewani J, Schiex T. 2006.** Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes. *Proceedings of the National Academy of Sciences of the United States of America* **103**:14959–14964
[DOI 10.1073/pnas.0603228103](https://doi.org/10.1073/pnas.0603228103).
- Cao YP, Han Y, Jin Q, Lin Y, Cai Y. 2016d.** Comparative genomic analysis of the GRF genes in Chinese pear (*Pyrus bretschneideri* Rehd), poplar (*Populus*), grape (*Vitis vinifera*), *Arabidopsis* and rice (*Oryza sativa*). *Frontiers in Plant Science* **7**:1750
[DOI 10.3389/fpls.2016.01750](https://doi.org/10.3389/fpls.2016.01750).
- Cao Y, Han Y, Li D, Lin Y, Cai Y. 2016a.** MYB transcription factors in chinese pear (*Pyrus bretschneideri* Rehd.): genome-wide identification, classification, and expression profiling during fruit development. *Frontiers in Plant Science* **7**:577
[DOI 10.3389/fpls.2016.00577](https://doi.org/10.3389/fpls.2016.00577).
- Cao Y, Han Y, Li D, Lin Y, Cai Y. 2016b.** Systematic analysis of the 4-Coumarate: coenzyme a ligase (4CL) related genes and expression profiling during fruit development in the Chinese pear. *Gene* **7(10)**:89 [DOI 10.3390/genes7100089](https://doi.org/10.3390/genes7100089).

- Cao Y, Han Y, Meng D, Li D, Jin Q, Lin Y, Cai Y. 2016c.** Structural, evolutionary, and functional analysis of the class III peroxidase gene family in chinese pear (*Pyrus bretschneideri*). *Frontiers in Plant Science* 7:1874 DOI [10.3389/fpls.2016.01874](https://doi.org/10.3389/fpls.2016.01874).
- Chang C, Shockey JA. 1999.** The ethylene-response pathway: signal perception to gene regulation. *Current Opinion in Plant Biology* 2:352–358 DOI [10.1016/S1369-5266\(99\)00004-7](https://doi.org/10.1016/S1369-5266(99)00004-7).
- Chao Q, Rothenberg M, Solano R, Roman G, Terzaghi W, Ecker JR. 1997.** Activation of the ethylene gas response pathway in Arabidopsis by the nuclear protein ETHYLENE-INSENSITIVE3 and related proteins. *Cell* 89:1133–1144 DOI [10.1016/S0092-8674\(00\)80300-1](https://doi.org/10.1016/S0092-8674(00)80300-1).
- Fawcett JA, Maere S, Van de Peer Y. 2009.** Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. *Proceedings of the National Academy of Sciences of the United States of America* 106:5737–5742 DOI [10.1073/pnas.0900906106](https://doi.org/10.1073/pnas.0900906106).
- Finn RD, Mistry J, Schuster-Böckler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R. 2006.** Pfam: clans, web tools and services. *Nucleic Acids Research* 34:D247–D251 DOI [10.1093/nar/gkj149](https://doi.org/10.1093/nar/gkj149).
- Freeling M. 2009.** Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annual Review of Plant Biology* 60:433–453 DOI [10.1146/annurev.arplant.043008.092122](https://doi.org/10.1146/annurev.arplant.043008.092122).
- Guo H, Ecker JR. 2004.** The ethylene signaling pathway: new insights. *Current Opinion in Plant Biology* 7:40–49 DOI [10.1016/j.pbi.2003.11.011](https://doi.org/10.1016/j.pbi.2003.11.011).
- Han Y, Ding T, Su B, Jiang H. 2016.** Genome-wide identification, characterization and expression analysis of the chalcone synthase family in maize. *International Journal of Molecular Sciences* 17(2):161 DOI [10.3390/ijms17020161](https://doi.org/10.3390/ijms17020161).
- Hiraga S, Sasaki K, Hibi T, Yoshida H, Uchida E, Kosugi S, Kato T, Mie T, Ito H, Katou S. 2009.** Involvement of two rice ETHYLENE INSENSITIVE3-LIKE genes in wound signaling. *Molecular Genetics and Genomics* 282:517–529 DOI [10.1007/s00438-009-0483-1](https://doi.org/10.1007/s00438-009-0483-1).
- Hu B, Jin J, Guo YA, Zhang H, Luo J, Gao G. 2014.** GSDB 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31(8):1296 DOI [10.1093/bioinformatics/btu817](https://doi.org/10.1093/bioinformatics/btu817).
- Illa E, Sargent DJ, Girona EL, Bushakra J, Cestaro A, Crowhurst R, Pindo M, Cabrera A, Van der Knaap E, Iezzoni A. 2011.** Comparative analysis of rosaceous genomes and the reconstruction of a putative ancestral genome for the family. *BMC Evolutionary Biology* 11:9 DOI [10.1186/1471-2148-11-1](https://doi.org/10.1186/1471-2148-11-1).
- Jing J, Kong J, Qiu J, Zhu H, Peng Y, Jiang H. 2016.** High level of microsynteny and purifying selection affect the evolution of WRKY family in Gramineae. *Development Genes & Evolution* 226:15–25 DOI [10.1007/s00427-015-0523-2](https://doi.org/10.1007/s00427-015-0523-2).
- Jourda C, Cardi C, Mbéguiéambégué D, Bocs S, Garsmeur O, D’Hont A, Yahiaoui N. 2014.** Expansion of banana (*Musa acuminata*) gene families involved in ethylene biosynthesis and signalling after lineage-specific whole-genome duplications. *New Phytologist* 202:986–1000 DOI [10.1111/nph.12710](https://doi.org/10.1111/nph.12710).

- Kosugi S, Ohashi Y. 2000.** Cloning and DNA-binding properties of a tobacco Ethylene-Insensitive3 (*EIN3*) homolog. *Nucleic Acids Research* **28**:960–967 DOI [10.1093/nar/28.4.960](https://doi.org/10.1093/nar/28.4.960).
- Lee JH, Kim WT. 2003.** Molecular and biochemical characterization of VR-EILs encoding mung bean ETHYLENE INSENSITIVE3-LIKE proteins. *Plant Physiology* **132**:1475–1488 DOI [10.1104/pp.103.022574](https://doi.org/10.1104/pp.103.022574).
- Leitch IJ, Leitch AR. 2013.** Genome size diversity and evolution in land plants. *Plant Genome Diversity* **2**:307–322 DOI [10.1007/978-3-7091-1160-4_19](https://doi.org/10.1007/978-3-7091-1160-4_19).
- Letunic I, Doerks T, Bork P. 2012.** SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Research* **40**:D302–D305 DOI [10.1093/nar/gkr931](https://doi.org/10.1093/nar/gkr931).
- Librado P, Rozas J. 2009.** DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**:1451–1452 DOI [10.1093/bioinformatics/btp187](https://doi.org/10.1093/bioinformatics/btp187).
- Lin Y, Cheng Y, Jin J, Jin X, Jiang H, Yan H, Cheng B. 2014.** Genome duplication and gene loss affect the evolution of heat shock transcription factor genes in legumes. *PLOS ONE* **9**:e102825 DOI [10.1371/journal.pone.0102825](https://doi.org/10.1371/journal.pone.0102825).
- Lynch M, Conery JS. 2000.** The evolutionary fate and consequences of duplicate genes. *Science* **290**:1151–1155 DOI [10.1126/science.290.5494.1151](https://doi.org/10.1126/science.290.5494.1151).
- Mao C, Wang S, Jia Q, Wu P. 2006.** OsEIL1, a rice homolog of the Arabidopsis *EIN3* regulates the ethylene response as a positive component. *Plant Molecular Biology* **61**:141–152 DOI [10.1007/s11103-005-6184-1](https://doi.org/10.1007/s11103-005-6184-1).
- Mercereau-Puijalon O, Barale JC, Bischoff E. 2002.** Three multigene families in Plasmodium parasites: facts and questions. *International Journal for Parasitology* **32**:1323–1344 DOI [10.1016/S0020-7519\(02\)00111-X](https://doi.org/10.1016/S0020-7519(02)00111-X).
- Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J. 2012.** The Pfam protein families database. *Nucleic Acids Research* **40**:D290–D301 DOI [10.1093/nar/gkr1065](https://doi.org/10.1093/nar/gkr1065).
- Rieu I, Mariani C, Weterings K. 2003.** Expression analysis of five tobacco *EIN3* family members in relation to tissuespecific ethylene responses. *Journal of Experimental Botany* **54**:2239–2244 DOI [10.1093/jxb/erg240](https://doi.org/10.1093/jxb/erg240).
- Shulaev V, Sargent DJ, Crowhurst RN, Mockler TC, Folkerts O, Delcher AL, Jaiswal P, Mockaitis K, Liston A, Mane SP. 2011.** The genome of woodland strawberry (*Fragaria vesca*). *Nature Genetics* **43**:109–116 DOI [10.1038/ng.740](https://doi.org/10.1038/ng.740).
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011.** MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**:2731–2739 DOI [10.1093/molbev/msr121](https://doi.org/10.1093/molbev/msr121).
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997.** The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**:4876–4882 DOI [10.1093/nar/25.24.4876](https://doi.org/10.1093/nar/25.24.4876).
- Tieman DM, Ciardi JA, Taylor MG, Klee HJ. 2001.** Members of the tomato LeEIL (*EIN3*-like) gene family are functionally redundant and regulate ethylene responses

- throughout plant development. *The Plant Journal* **26**:47–58
DOI [10.1046/j.1365-3113x.2001.01006.x](https://doi.org/10.1046/j.1365-3113x.2001.01006.x).
- Verde I, Abbott AG, Scalabrin S, Jung S, Shu S, Marroni F, Zhebentyayeva T, Dettori MT, Grimwood J, Cattonaro F. 2013.** The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nature Genetics* **45**:487–494 DOI [10.1038/ng.2586](https://doi.org/10.1038/ng.2586).
- Waki K, Shibuya K, Yoshioka T, Hashiba T, Satoh S. 2001.** Cloning of a cDNA encoding *EIN3*-like protein (DC-EIL1) and decrease in its mRNA level during senescence in carnation flower tissues. *Journal of Experimental Botany* **52**:377–379 DOI [10.1093/jexbot/52.355.377](https://doi.org/10.1093/jexbot/52.355.377).
- Wang Y, Feng L, Zhu Y, Li Y, Yan H, Xiang Y. 2015.** Comparative genomic analysis of the WRKY III gene family in populus, grape, arabidopsis and rice. *Biology Direct* **10**:48 DOI [10.1186/s13062-015-0076-3](https://doi.org/10.1186/s13062-015-0076-3).
- Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, T-h Lee, Jin H, Marler B, Guo H. 2012.** MScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research* **40**:e49 DOI [10.1093/nar/gkr1293](https://doi.org/10.1093/nar/gkr1293).
- Wendel JF. 2000.** Genome evolution in polyploids. In: *Plant molecular evolution*. Springer Netherlands, 225–249.
- Wu J, Wang Z, Shi Z, Zhang S, Ming R, Zhu S, Khan MA, Tao S, Korban SS, Wang H. 2012.** The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Research* **23**:396–408 DOI [10.1101/gr.144311.112](https://doi.org/10.1101/gr.144311.112).
- Wu T, Zhang R, Gu C, Wu J, Wan H, Zhang S, Zhang S. 2012.** Evaluation of candidate reference genes for real time quantitative PCR normalization in pear fruit. *African Journal of Agricultural Research* **7**:3701–3704 DOI [10.5897/AJAR11.1842](https://doi.org/10.5897/AJAR11.1842).
- Xiang Y, Huang CH, Yi H, Wen J, Li S, Yi T, Chen H, Xiang J, Hong M. 2017.** Evolution of Rosaceae fruit types based on nuclear phylogeny in the context of geological times and genome duplication. *Molecular Biology & Evolution* **34**:262–281 DOI [10.1093/molbev/msw242](https://doi.org/10.1093/molbev/msw242).
- Yang ZP, Li HL, Guo D, Peng SQ. 2015.** Identification and characterization of the *EIN3/EIL* gene family in *Hevea brasiliensis*. *Tree Genetics & Genomes* **11**:1–11 DOI [10.1007/s11295-015-0877-7](https://doi.org/10.1007/s11295-015-0877-7).
- Yokotani N, Tamura S, Nakano R, Inaba A, Kubo Y. 2003.** Characterization of a novel tomato *EIN3*-like gene (*LeEIL4*). *Journal of Experimental Botany* **54**:2775–2776 DOI [10.1093/jxb/erg308](https://doi.org/10.1093/jxb/erg308).
- Zdobnov EM, Apweiler R. 2001.** InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**:847–848 DOI [10.1093/bioinformatics/17.9.847](https://doi.org/10.1093/bioinformatics/17.9.847).
- Zhang Q, Chen W, Sun L, Zhao F, Huang B, Yang W, Tao Y, Wang J, Yuan Z, Fan G. 2012.** The genome of *Prunus mume*. *Nature Communications* **3**:1318 DOI [10.1038/ncomms2290](https://doi.org/10.1038/ncomms2290).
- Zhong Y, Yin H, Sargent DJ, Malnoy M, Cheng ZM. 2015.** Species-specific duplications driving the recent expansion of NBS-LRR genes in five Rosaceae species. *BMC Genomics* **16**:77 DOI [10.1186/s12864-015-1291-0](https://doi.org/10.1186/s12864-015-1291-0).