J. Plant Biol. (2019) 62:48-60
DOI 10.1007/s12374-018-0319-9

ORIGINAL ARTICLE

Land-plant Phylogenomic and Pomegranate Transcriptomic Analyses Reveal an Evolutionary Scenario of CYP75 Genes Subsequent to Whole Genome Duplications

Taikui Zhang[1,2], Cuiyu Liu[1,2], Xianbin Huang[1,2], Hanyao Zhang[3,*] and Zhaohe Yuan[1,2,*]

[1]*Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing, Jiangsu Province 210037, China*
[2]*College of Forestry, Nanjing Forestry University, Nanjing, Jiangsu Province 210037, China*
[3]*Key Laboratory of Biodiversity Conservation in Southwest China, the State Forestry Administration, Southwest Forestry University, Kunming, Yunnan Province 650224, China*

Received: September 6, 2018 / Accepted: October 19, 2018

Abstract Regulatory and developmental genes are retained following whole-genome duplication (WGD) events, and, thus, are central to elucidating the evolution of the gene family subsequent to WGDs. Among these genes, the CYP75 gene family is a key member of the biggest enzyme superfamily in land-plant lineages. Although the molecular genetics of the biological progress involved with CYP75 genes have been partly elucidated, the evolution after WGDs in land-plant lineages are still largely unknown. Here, we identified CYP75 orthologues in pomegranate (*Punica granatum*) and other twenty-five representative species to explore the gene evolution under WGD shaping on a broad evolutionary scale. Phylogenomic analyses identified genome-wide CYP75 candidates and suggested that a recent duplication of the CYP75 genes in seed plants occurred prior to the split of gymnosperms and angiosperms approximately 400 million years ago. Molecular evolution analyses revealed that CYP75 gene lineages evolved under a different purifying selection pressure, and slight relaxations occurred in the recent duplication groups in gymnosperms and angiosperms. The syntenic analyses showed that WGDs together with segmental duplications contributed to the CYP75 gene evolution in pomegranate. RT-PCR, qRT-PCR and RNA-Seq verification suggested that pomegranate CYP75 genes evolved through exon fusion and had a fruit-specific expression pattern. Neo- or sub-functionalization is the main fate of CYP75 genes following duplication. The expression pattern of homologous copies of CYP75 in pomegranate supports the CYP75 family evolution contributing to species reproduction that showy fruit colours attracted birds and other animals to spread seeds. Integration of the above analyses generated a putative evolutionary scenario of the CYP75 family in land plants. Our data provided a potential reference model to further elucidate the evolution of the regulatory and developmental gene families after WGDs.

Keywords: CYP75 Gene family, Expression pattern, Land plant lineage-specific evolution, Pomegranate genome, Whole-genome duplication

*Corresponding author; Zhaohe Yuan, Hanyao Zhang
Tel : +86-25-85427056; +86-871-63863540
E-mail : zhyuan88@hotmail.com; zhanghanyao@hotmail.com

Introduction

Whole-genome duplication (WGD) provides the raw material for adaptation in the form of duplicated genes (Hollister 2015). After genome duplication, most duplicate copies are lost, but a considerable fraction is retained, presumably due to neo- or sub-functionalization or through selection for dosage balance in protein complexes (Lee and Irish 2011). The strongly biased retention of regulatory and developmental genes subsequent to WGDs has far-reaching ramifications for evolution in the longer term (Van de Peer et al. 2017). These regulatory and developmental genes control a wide variety of reactions in biochemical pathways in land plants. WGD events have consequences for the evolvability or the robustness of these reactions. The regulatory and developmental genes hence bear a unique significance for elucidating the gene family evolution after WGDs.

For the regulatory and developmental gene families, we selected the cytochrome P450 (CYP) 75 gene family as a

model to explore the evolution after WGDs. As the largest enzyme superfamily in plants, CYPs catalyse a wide variety of reactions in biochemical pathways, such as the biosynthesis pathways of anthocyanins and flavonoids, the jasmonic acid and ethylene signaling pathway, modification of membrane topology and enhanced resistance to stresses (Tanaka 2006; Seitz et al. 2015; Renault et al. 2017; Wei and Chen 2018). During fruit or flower development, the CYP75 gene family controls the transformation of the biosynthetic precursors of delphinidin 3-glucoside and cyanidin 3-glucoside (Tanaka 2006; Seitz et al. 2015) that can be catalysed into anthocydin diglucosides to diversify the pigments that are central to attracting seed dispersers (Petroni and Tonelli 2011; Chen 2015). CYP75 genes have been preferentially endowed with flower/fruit-specific expression patterns and/or functional roles in this tissue (Tanaka 2006). Although the molecular genetics of the biological progress involved with CYP75 genes have been partly elucidated, the evolution after WGDs in land-plant lineages are still largely unknown.

Non-functionalization, neo- or sub-functionalization, and pseudogenization are the main fates of genes after duplication. The CYP75 gene family contains two distinct phylogenetic clades with different functions, one catalysing delphinidin 3-glucoside, while the other one catalyses cyanidin 3-glucoside (Seitz et al. 2015). It is a topic of interest as to whether neo-functionalization or another fate occurred in CYP75 genes since the duplication. An estimate of the evolutionary fate enables a better elucidation of the shaping of WGDs on CYP75 orthologs. The increasing amount of genomic sequence data available provides an opportunity to explore the evolvability of CYP75 families subsequent to WGDs in an appropriate evolutionary context. High-quality genomic sequences (Yuan et al. 2018) coupled with transcriptomic data (Zhang et al. 2017) enable pomegranate (*Punica granatum*) to be a key candidate species for plant phylogenomics. Most of the pomegranate cultivars are genetically improved for peel colour, driving its selection as an interesting system to study fruit pigments (Holland et al. 2009). Hence, we selected the pomegranate CYP75 gene family to explore whether its duplicated copies generate novel expression patterns in fruit following WGDs.

Here, we sampled twenty-six interesting species bearing reference genome sequences for land-plant phylogenomics and identified the genome-wide CYP families. We applied on phylogenomic approaches to identify and compare the CYP gene families in land-plant lineages overall. Phylogenomic analyses found a recent duplication of the CYP75 genes in seed plants that occurred prior to the split of gymnosperms and angiosperms approximately 400 million years ago (Mya). Genomic synteny analyses were subsequently done to elucidate that WGDs shaped the evolution of the CYP75 gene. After checking the lineage-specific branch evolutionary pressure on CYP75 genes through molecular evolution analyses, gene structure, genomic mapping and RNA-Seq analyses of CYP75 genes in pomegranate were applied to speculate on the evolution scenario. Our results provided an evolutionary scenario of the CYP75 gene family in land plants that are of great importance for understanding the lineage-specific evolution subsequent to WGDs.

Results

Low-copy Phylogenetic Markers are Well Developed to Reconstruct Species Tree for Phylogenomics

To reconstruct a species tree with representative taxa of land plants, we performed all-against-all BLASTP and OrthoMCL analyses among *P. granatum*, *Citrus sinensis*, *Actinidia chinensis*, *Musa acuminata*, *Ginkgo biloba* and *Marchantia polymorph*, yielding 356 single-copy gene families. Low-copy gene families shared by all 26 species were retrieved with these 356 gene families as query and filtered by removing genes that lost their sequences in more than 3 species (Note S1), generating 21 low-copy phylogenetic markers (Table S3) with high quality signals.

An ML phylogenetic tree was reconstructed with the concatenated low-copy genes (Fig. 1A). The phylogenetic tree bears robust bootstrap-supported values overall. The tree provided a representative species phylogeny for land plants, with the *M. polymorpha* as the outgroup, and clustered the pomegranate into the Myrtales clade (Bowman et al. 2017; Murat et al. 2017; Yuan et al. 2018). A divergence time analysis showed pomegranate and *E. grandis* diverged from their MRCA at ~55.8 Mya (Fig. 1A), which is in agreement with the fossil record (Graham 2013). Land plants diverged from streptophyte alga ~912 Mya but diversified into living lineages ~475 Mya (Magallón et al. 2013). The split of the seed plant lineage and the bryophyte lineage is estimated at ~400 Mya (Fig. 1A), which is consistent with recent phylogenetic analyses (Jiao et al. 2011). Hence, the reconstructed species tree could provide reference phylogeny materials for downstream phylogenomics.

Complex Lineage-specific Duplication, Loss and Co-speciation Events Occurred in the CYP75 Gene Family

A total of 3982 putative CYP candidates were identified in twenty-six species though a phylogenetic approach and formed four main phylogenetic clades, Clan71, Clan72, Clan85 and Clan86 (Fig. 1B, 1C; See detail CYP phylogenetic analyses in Note S2). The CYP75 gene family has evolved to be the ancient phylogenetic clade of the CYP76, CYP80 and CYP706 families (Fig. 1C). Compared to the CYP75 gene

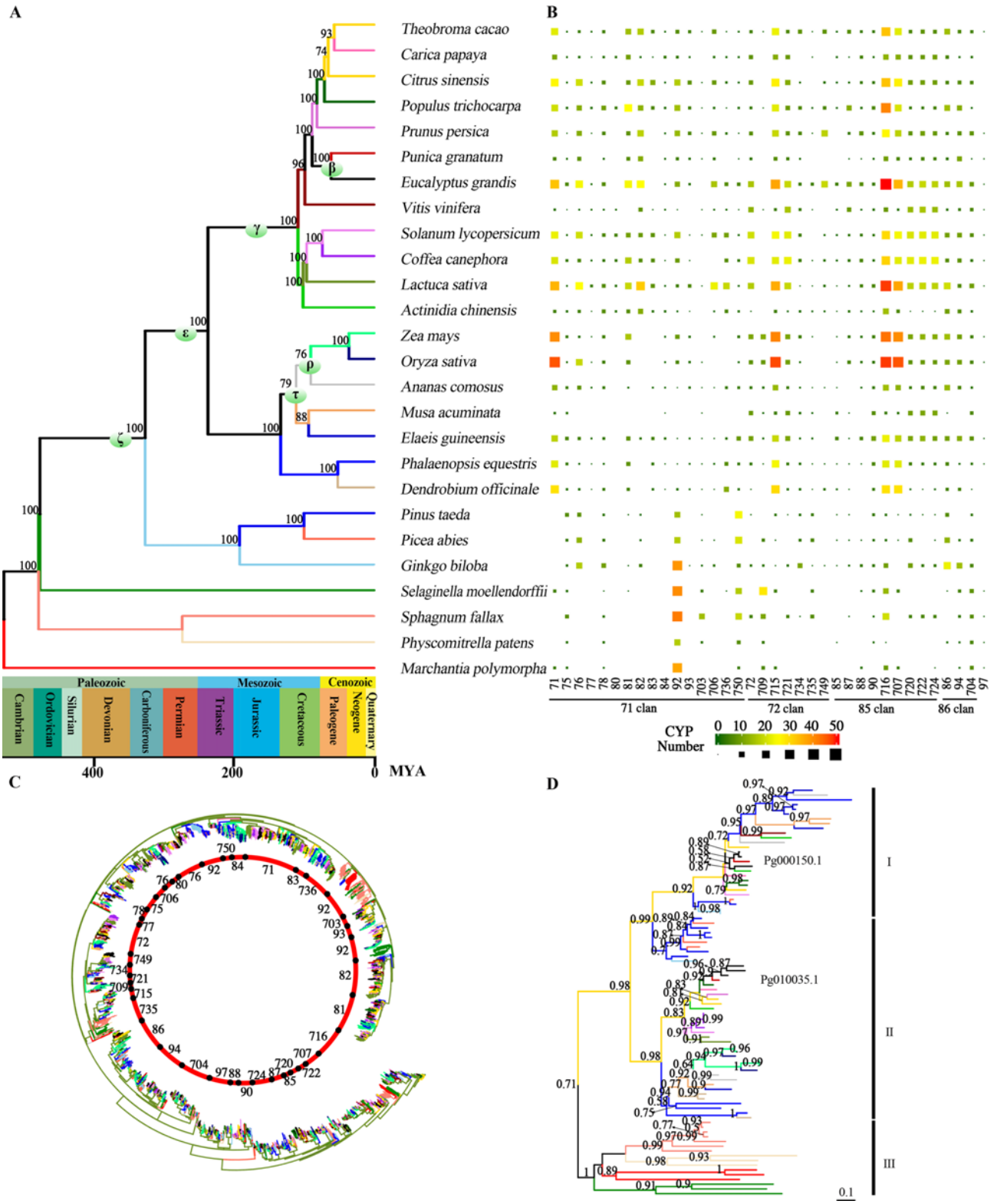

Fig. 1. Evolutionary plasticity of CYP75 gene families in plants. (A) A time-calibrated phylogeny with different branch color showing distinct species. Node support was evaluated with 1000 bootstrap pseudoreplicates. The greek letters stand for the known WGD events. The geologic timescale data is obtained from the International Commission on Stratigraphy (http://stratigraphy.org, version 2017/02). (B) CYP numbers (squares with different sizes and colors) in each land plant (each row). (C) A ML phylogenetic tree (outer circle) and group classification (inner circle) of CYP gene families in plants presented by various branch colors as same as that in Figure A. (D) CYP75 gene family phylogenetic tree in land plant lineages presented by various branch colors as same as that in Figure A. Node support (decimals) was quantified by aLRT statistics with the SH-like procedure.

family occurring in land plants, only putative CYP76, CYP80 and CYP706 candidates were identified in seed plants (Fig. 1B, C). This result suggested that a lineage-specific loss event occurred after the split of bryophyte species and vascular plants approximately 400 Mya. Altogether, the phylogenomic analyses provided novel lineage-specific duplication and loss events in CYP gene evolution history.

We split the CYP75 gene tree solely between the entire CYP phylogeny to further interpret lineage-specific evolution events (Fig. 1C, D). Node support based on aLRT statistics with SH-like procedure yielded a robust inferred phylogenetic tree of CYP75 genes overall (Fig. 1D). Eighty-five putative CYP75 genes fall into three distinct clades. The CYP75 genes from bryophyte species constitute the basal clade, which represented an ancient phylogenetic clade. Further analyses showed that these genes split into four subfamilies belonging to the *M. polymorpha*, *Physcomitrella patens*, *S. moellendorffii*, and *Sphagnum fallax* groups, respectively (Fig. 1D). Although each of the four species had more than 3 CYP75 orthologs, the topologies for the four main subfamilies were essentially identical to that of the species tree (Fig. 1A, D). This result indicated that a co-speciation event and four lineage-specific duplication events occurred during the CYP75 sequence divergence.

Seventy CYP75 orthologues formed two sister clades, I and II, both of which contained CYP75 members from angiosperm and gymnosperm species. In agreement with previous molecular phylogenetic studies (Tanaka 2006), clade I split from clade II (support =0.99; Fig. 1D). These two joint clades formed a seed plant-lineage specific clade (support=0.98; Fig. 1D), which is sister to the ancient bryophyte clade. It indicated a recent duplication (clade II) of the CYP75 genes in seed plants that occurred prior to the split of gymnosperms and angiosperms and the complex lineage-specific duplication events during the evolution of the CYP75 gene family.

The CYP75 Gene Family Underwent a Purifying Selection During Evolution

We detected the molecular evolution of the CYP75 gene family using the program CODEML in the package PAML (Yang 2007). The M0 model assumes a single ω ratio for the whole tree, and both the Nearly Neutral M1a site model and the Selection M2a site model allow the ω ratio to vary among sites. Under the M0 model, a global ω of 0.16166 (Fig. 2; Table S4) revealed a purifying selection on the CYP75 gene family in land plants. This observation was further supported by both the Nearly Neutral M1a site model following two categories (ω0 < 1 and ω1 = 1) and the Selection M2a site model following three categories (ω0 < 1, ω1 = 1 and ω2 = 1). It showed that 91.866% of CYP75 codons

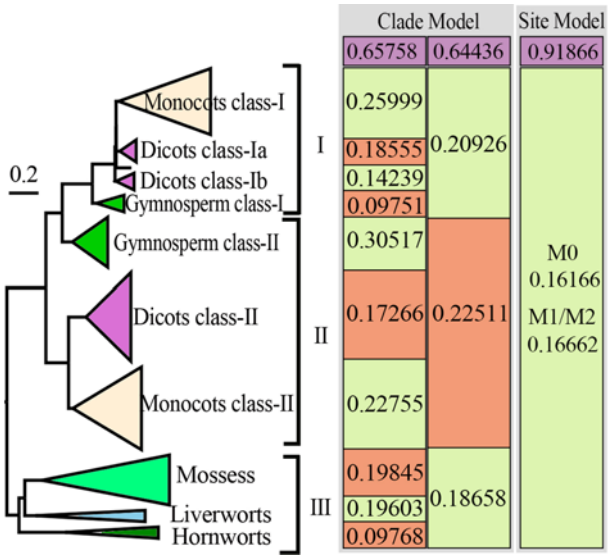

Fig. 2. Molecular evolution analysis of CYP75 gene family in plants. Left tree is a collapsed phylogenetic tree of CYP75 genes from different evolutionary lineages (different colors). The right panel shows the ω values (decimals in green and orange blocks) of different collapsed phylogenetic clades under the clade mode and site mode. The decimals in purple blocks present the proportion of the ω values.

have a ω of 0.16662 (Fig. 2; Table S4). Thus, purifying selection on the CYP75 gene family in land plants was inferred under the codon site model.

To decipher whether some CYP75 gene lineages evolved under different selection pressure, a clade model analysis was employed by assigning three different ω ratios to the three main clades. The sister clades I (ω = 0.20926), II (ω = 0.22511) and III (ω = 0.18658) evolved under overall purifying selection (Fig. 2; Table S4), revealing that slight relaxations occurred on class II in gymnosperms and angiosperms. This supported that clade II was the recent duplication groups that appeared prior to the split of gymnosperms and angiosperms.

A further clade model that allows one site category (out of three) to vary between defined branches of the tree was performed on ten clades, including the monocots class-I, dicots class-Ia, dicots class-Ib, gymnosperm class-I, gymnosperm class-II, monocots class-II, dicots class-II, mosses, liverworts and hornworts (Fig. 2; Table S4). All ten groups were found to evolve under overall strong purifying selection (Fig. 2; Table S4), although the CYP75 duplication in dicots appeared to result in a slight relaxation of the purifying selection. This relaxation is less obvious for gymnosperms for which purifying selection exerted on class I genes (ω = 0.09751) is already looser than for monocots and dicots (Fig. 2; Table S4), which is possibly linked to an additional gene duplication. A comparatively relaxed negative selection (ω = 0.19845) was found in mosses in compared with lower vascular plants (Fig. 2; Table S4). This might be related to the lineage-

specific amplification of mosses.

WGDs Shaped the Pomegranate CYP Gene Superfamily

Most flowering plant lineages underwent one or more rounds of WGDs, the pomegranate genome underwent a paleotetraploidy event shared by *E. grandis* and a paleohexaploidy event shared by all eudicots (Jiao et al. 2011; Murat et al. 2017; Yuan et al. 2018). Thus, a syntenic block ratio (1:2:2) of *V. vinifera* versus pomegranate versus *E. grandis* serves as a key clue to identify WGDs. We used MCScanX to identify syntenic blocks and found the syntenic block ratio (1:3:6:6) of ancient eudicots versus *V. vinifera* versus pomegranate versus *E. grandis* (Fig. 3A). A similar ratio of syntenic blocks was found in the CYP genes (Fig. 3A), which indicated that WGDs shaped the evolution of the major CYP genes. Despite the two *PgCYP75* genes and the six *EgCYP75* genes in Fig. 1D, a ratio of 2:3 was found among pomegranate versus *E. grandis* in *CYP75* genes (Fig. 3A). WGDs together with segmental duplications contributed to the evolution of the CYP75 gene family (Fig. 3B). It is concluded that WGDs shaped the evolution of the major CYP genes and that the WGD event is one of the key forces in the CYP75 gene evolution.

We conducted gene mapping and found that a total of 92 pomegranate CYP genes were mapped onto 44 scaffolds (Fig. 3B). The majority of the CYP genes were located in the distal regions, while a few were mapped in proximal regions (Fig. 3B). The scaffolds of s25 and s1 contained the maximum density of CYP genes. Interestingly, numerous CYP genes were mapped very close to a specific location on the scaffold. The presence of clusters of CYP genes from the same clade suggested that these genes were tandem-duplicated (Fig. 4A; Fig. 3B). Approximately 42.4% (39 out of 92) of CYP genes had the tandem locations in the pomegranate genome scaffolds (Fig. 4A; Fig. 3B), revealing that tandem duplication as an essential event for CYP gene evolution. The rest of the CYP genes, with 2-5 members belonging to the same family, were mapped on distinct scaffolds (Fig. 4A; Fig. 3B). This result indicated that most CYP genes might evolve to have one or more paralogs through WGDs. The CYP75 gene family had two members that were located on different scaffolds (Fig. 3; Fig 4A), revealing that their evolution might be related to the WGD events. Together, the gene mapping analyses indicated that the pomegranate CYP genes underwent tandem duplication and WGD events and were non-randomly distributed in the genome scaffolds.

Pomegranate CYPs Fused Their Exons to Generate New Copies

The exon and intron distribution of the pomegranate CYP genes were analysed based on their phylogenetic similarities (Fig. 4A, B). The exon and intron analyses indicated that the majority of the pomegranate CYP genes contained introns, and the intron distribution is diverse and varied from 0 to 15 (Fig. 4B). The same gene family had a similar exon-intron structure and lengths, except for a few length differences within the CYP81 and CYP93 gene families (Fig. 4B), supporting the assumption that the intron number and length contribute to the gene family differentiation. Our findings revealed that the pomegranate CYP introns have to co-evolve with their parent genes through duplication events, including tandem duplications or WGDs. A combination of the phylogenetic

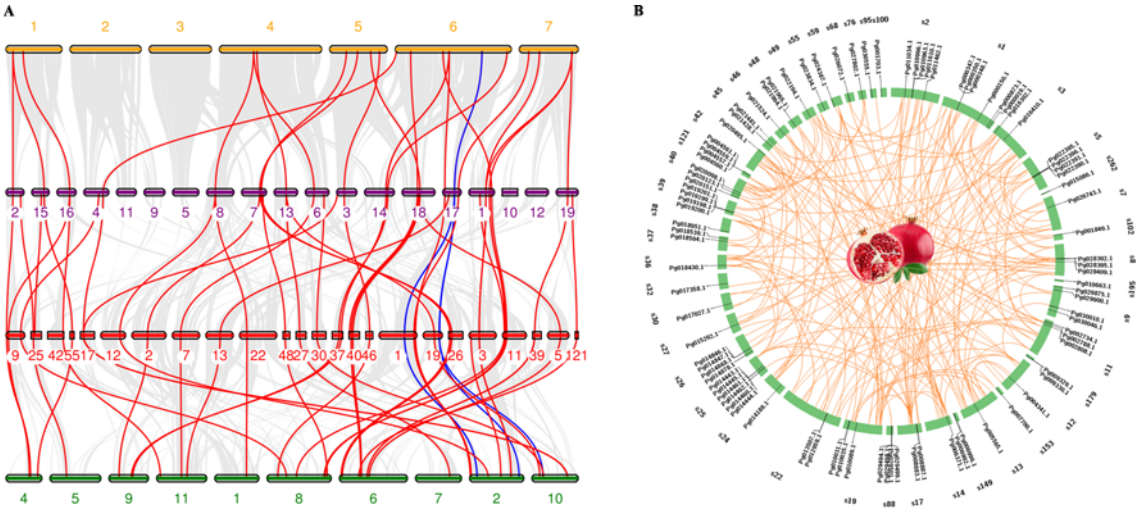

Fig. 3. Syntenic blocks of genomes. (A) Syntenic CYP75 (blue lines) and other CYP (red lines) genes and large-scale syntenic blocks (gray lines) in ancient eudicot chromosomes (yellow ideogram), *Vitis vinifera* chromosomes (purple ideogram), *Punica granatum* scaffolds (red ideogram) and *Eucalyptus grandis* chromosomes (green ideogram). (B) Syntenic genes (orange line) and CYP genes (black label) in pomegranate genome scaffolds.

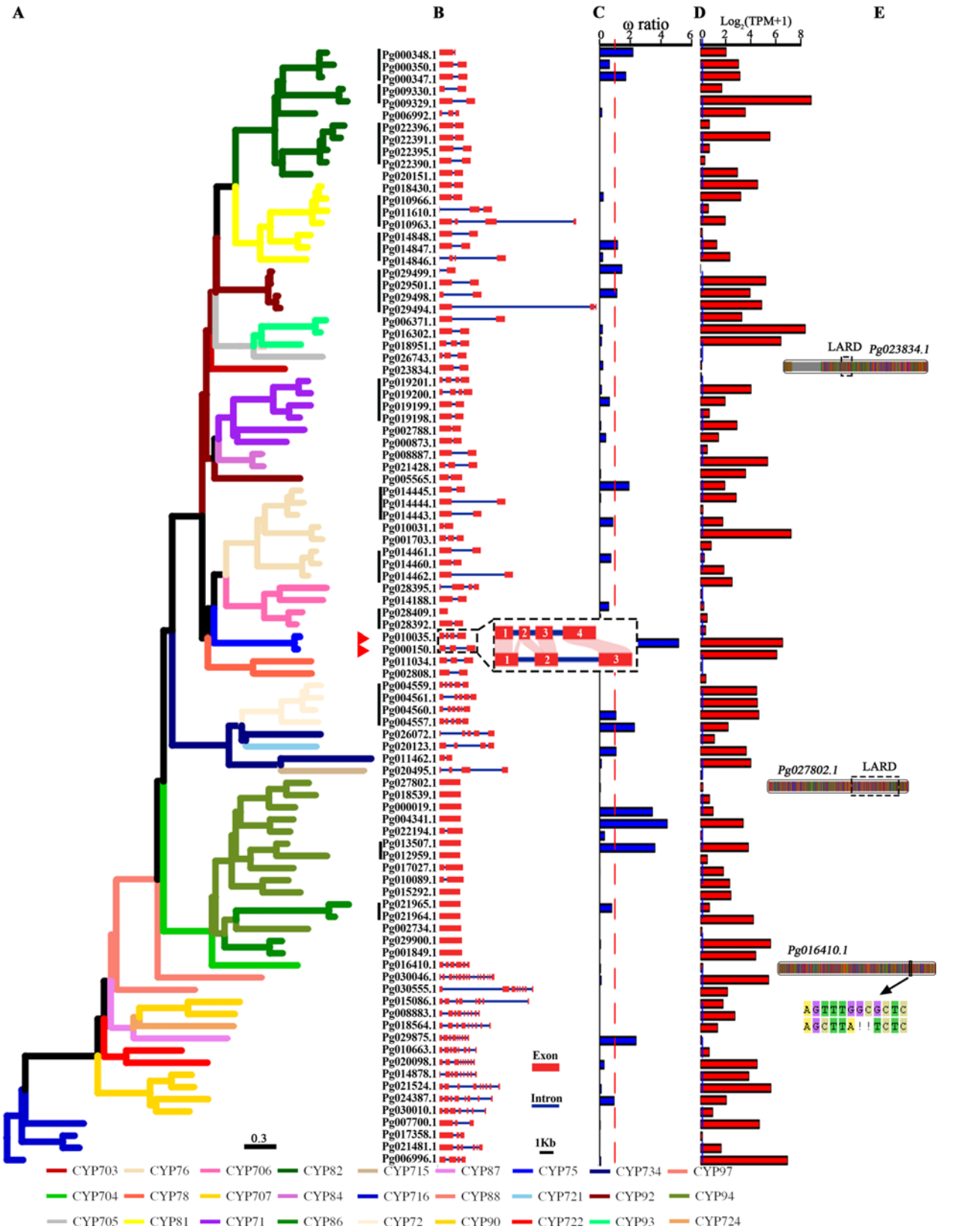

Fig. 4. Phylogenetic tree (A), gene structure (B), ω ratio (C), expression (D) and pseudogene model (E) of pomegranate CYP genes. The dotted box in Figure B presents an enlarged model of the fusion of 4 exons into 3 exons. Colour letters in Figure E present bases, above ones showing the true sequences, while below ones are collected through GeneWise analyses.

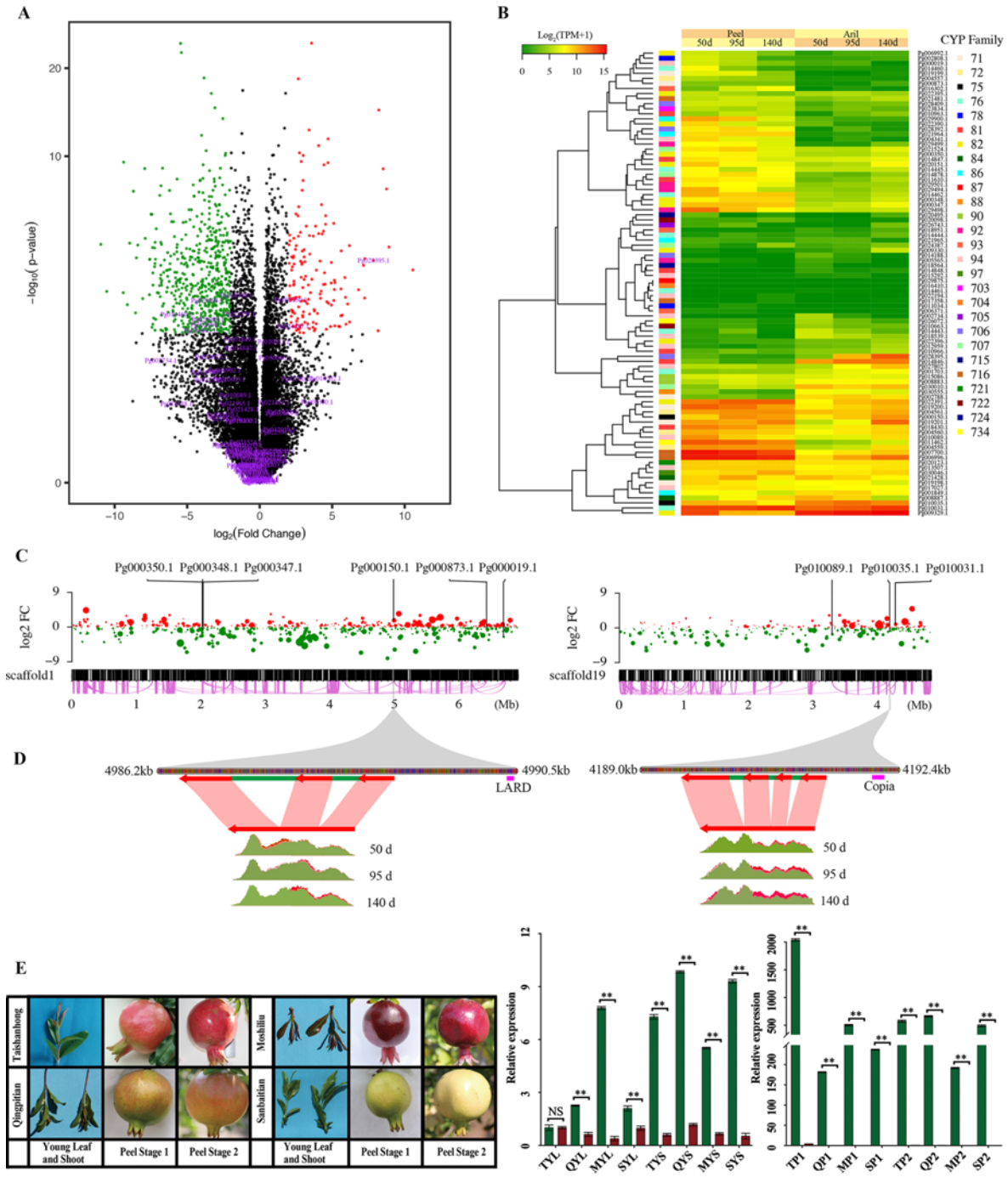

Fig. 5. Expression pattern of CYP75 gene family. (A) CYP genes (purple markers) of up-regulated (red points) and down-regulated (green points) differentially expressed genes during peel and aril development. (B) Expression heatmap of CYP genes during peel and aril development. (C) Genomic scale gene expression pattern (top points) and syntenic genes (purple lines) in scaffold 1 (top left panel) and scaffold 19 (top right panel). (D) Enlarged genome mapping (top panel), gene structure models (middle panel) and RNA-Seq reads coverage (bottom panel) of *Pg000150.1* and *Pg010035.1*. In middle panels, red arrows, green bars and purple bars present exons, introns and promoters, respectively. In bottom panels, color areas present the reads coverage in peel (green) and aril (red) from three developmental stages, including 50d, 95d and 140d after flowing. (E) Relative expression (right panel) of *Pg000150.1* (green bar) and *Pg010035.1* (red bar) quantified in four genotypes (left photos) by qRT-PCR. NS, $P > 0.05$, not significant; **, $P < 0.01$, with one-way ANOVA. TYL, Taishanhong young leaf; QYL, Qingpitian young leaf; MYL, Moshiliu young leaf; SYL, Sanbaitian young leaf; TYS, Taishanhong young shoot; QYS, Qingpitian young shoot; MYS, Moshiliu young shoot; SYS, Sanbaitian young shoot; TP1, Taishanhong peel at stage1; QT1, Qingpitian peel at stage1; MT1, Moshiliu peel at stage1; ST1, Sanbaitian peel at stage1; TP2, Taishanhong peel at stage2; QT2, Qingpitian peel at stage2; MT2, Moshiliu peel at stage2; ST2, Sanbaitian peel at stage2.

and gene structure analyses suggested that pomegranate CYPs with early duplication events (such as 704, 707, 722, 724, 87, and 90) tended to own more introns. In contrast, the other CYPs tended to contain fewer introns, and their exons tended to be longer (Fig. 4B). The gene structure evidence revealed that pomegranate CYPs fused their exons to generate new copies.

We further tested the exon fusion model in the CYP75 gene family. Either of the pomegranate CYP75 genes contained more than two exons, and the ORFs of *Pg010035.1* (NCBI accession: KY939736) and *Pg000150.1* (NCBI accession: KY939737) were isolated by RT-PCR. Through sequence alignment, we found two exon fusion occurrences. The first and partial second CDSs of *Pg010035.1* were integrated into the first CDS of *Pg000150.1*, and the third and the rest of the second CDSs of *Pg010035.1* were integrated into the second CDS of *Pg000150.1*, while the fourth CDS of *Pg010035.1* was integrated into the third CDS of *Pg000150.1* (Fig. 4B). Our data showed that the CDS recombination and large intron insertion occurred during the evolution from *Pg010035.1* to *Pg000150.1*.

Pseudogenization is a Fate of CYP Genes that Exclude CYP75 Orthologues after Duplication

Pseudogenization is a fate of genes that have undergone tandem duplications or WGDs. Three genes (*Pg023834.1*, *Pg027802.1*, and *Pg016410.1*) were identified as pseudogenes, in which a low ω ratio for the pseudogene/parent gene pairs ($\omega < 1$; Fig. 4C) and low gene expression (0~1 TPM in mix tissues; Fig. 4D) were observed. *Pg027802.1* contained only an entire exon, in which an obvious frameshift and an inserted large retrotransposon derivatives (LARD) element were found (Fig. 4E), revealing that the *Pg027802.1* was a processed pseudogene derived from a retrotransposition event of a LARD insertion in an exon. Pseudogene structure analyses for *Pg023834.1* and *Pg016410.1* indicated that the long intron insertion followed by a frameshift in CYP genes might result in pseudogenization (Fig. 4B, E).

CYP75s Exhibit Distinct Expression Patterns and Play Key Roles in Fruit Development

To elucidate the putative functions in peel and aril development, we checked the expression levels of CYP genes using the public pomegranate RNA-Seq data (Fig. 5A, B). A volcano plot analysis (Fig. 5A) indicated that among 734 DEGs, 1.39% (2 out of 144) of up-regulated genes were CYP genes (*Pg028395.1* and *Pg000150.1*), and 8.47% (5 out of 590) of significantly down-regulated genes were CYP genes (*Pg029494.1*, *Pg029498.1*, *Pg014460.1*, *Pg002808.1*, and *Pg014443.1*). This revealed that the seven CYP genes played vital roles in peel and aril development. Additionally, the expression heatmap in Fig. 5B shows that the CYP gene families had distinct tissue-specific expression patterns during fruit development. For instance, the CYP716 (*Pg007700.1* and *Pg006996.1*), CYP734 (*Pg011462.1*) and CYP75 (*Pg000150.1*) genes were highly expressed in the peel, indicating a peel-specific expression pattern. The CYP71 (*Pg002788.1*), CYP81 (*Pg014846.1*), CYP706 (*Pg028395.1*) and CYP75 (*Pg010035.1*) genes were highly expressed in aril, indicating an aril-specific expression pattern. The DEG analyses indicated that seven CYP genes played key roles in peel and aril development.

To verify the evolutionary assumption underlying the diverse expression patterns, the gene expression levels on a genomic scaffold scale were investigated (Fig. 5C). For the large-scale duplications in scaffold1 (Fig. 5C), the tandem duplication genes (*Pg000350.1*, *Pg000348.1* and *Pg000347.1*) were expressed at low levels, while the CYP75 gene (*Pg000150.1*) had a high expression fold change during the peel and aril development. CYP genes with different expression patterns might be related to the segmental large-scale duplications due to the tandem duplicates resulting in gene non-functionalization (Panchy et al. 2016). Similar results to those of the CYP duplicates were also found in AOMT (Yuan et al. 2018).

We further elucidated the links between expression patterns and gene structure in the CYP75 gene family. Two CYP75 genes bore different expression patterns. Phylogenetic evidence indicated that *Pg000150.1* diverged from *Pg010035.1* through exon fusion. RNA-Seq reads analyses revealed that CDSs of both *Pg010035.1* and *Pg000150.1* contained reads coverage, while they had distinct deep coverages (Fig. 5D). Except for the exons and introns, the promoters are the crucial control region surrounding transcription start sites in plant gene structure. Gene structure analyses showed that a LARD element occurred in the promoter blocks of *Pg000150.1*, while a Copia element occurred in the *Pg010035.1* (Fig. 5D). Together, the CYP75 gene expression pattern during fruit development might be related to the repeat sequence-induced promoter area and the exon structure variation.

The CYP75 gene family played key roles in anthocyanin biosynthesis. To further investigate the expression pattern of CYP75s, we performed the qRT-PCR verification in young leaves and stems, and the two stages of peels of four distinct cultivars. Relative expression analyses indicated that the gene expression levels of *Pg000150.1* prior to that of *Pg010035.1* in all tissue and cultivars and that the CYP75 family had a peel tissue-specific expression pattern (Fig. 5E). Notably, the expression ratio of *Pg000150.1* to *Pg010035.1* in the cultivar with red peels increased during coloration, while those with purple, green and white peels degraded. The RNA-Seq together with qRT-PCR verification supported

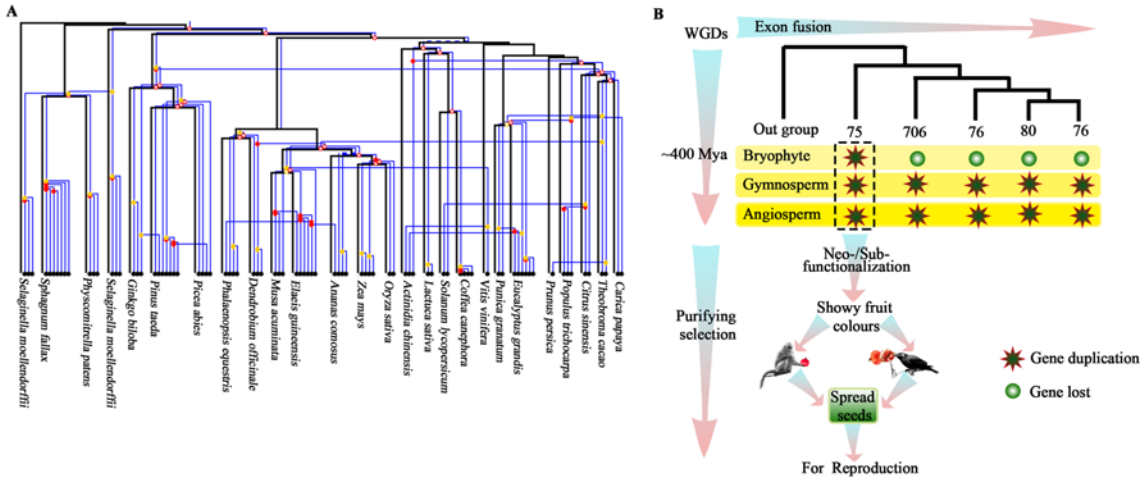

Fig. 6. Co-speciation (hollow circle), duplication (dot), loss (dotted line) and transfer (arrow) events (A) and evolutionary scenario (B) of CYP75 gene family in plants.

the conclusion that CYP75 genes had fruit-specific expression patterns.

Discussion

In this paper, phylogenomic analyses identified the putative candidates and inferred the molecular evolution of CYP75 gene families in land-plants. Then, transcriptomic analyses unveiled the gene fate of CYP75 genes after WGDs in pomegranate. Integrated above analyses generated a putative evolutionary scenario of the CYP75 family in land plants.

In the selected 26 species, a total of 3982 CYPs were identified through phylogenetic approach, of which only 92 CYP homologous genes were identified in the pomegranate genome. In plants, CYP75 is one of the gene families in a phylogenetic clade of CYP gene superfamily, CYP71 clan, that had the most gene families (Nelson et al. 2004; Zhang et al. 2017). The CYP genes from pomegranate were lower than that from species of the Malvids lineage (average > 100; Zhang et al. 2017), suggesting that there might be a pomegranate lineage-specific loss event under purifying selection. The difference might also be related to the strict filtering criteria in identification. To ensure phylogeny with strong posterior probability, we deleted the longest gene with an apparent split variant and paralogs with few amino acid residues along the entire sequence.

In our study, phylogenomic analyses showed that a recent duplication of the CYP75 genes in seed plants occurred prior to the split of gymnosperms and angiosperms approximately 400 million years ago. The observed lineage-specific duplication and loss events might result from the incomplete genomic annotation database and from the deletion of sequences with obvious variants in a phylogenetic approach. However, gene-loss events are not unusual because the non-functionalization is the major fate of gene copies after duplication (Hartmann et al. 2014). Gene duplication has been recognized as a primary mechanism underlying increased functional diversification, and the increased expression divergence in duplicated genes can substantially contribute to physiological and morphological diversification (Victor et al. 2015; Soltis and Soltis 2016). These inferred duplication events thus have a unique evolutionary significance that provides materials for functional diversification in different lineages.

Molecular evolution analyses found an overall purifying selection on the land plant lineage-specific CYP75 orthologues under the codon site model and clade model. Lineage specific gene duplication can contribute to unique evolutionary changes and novel phenotypic adaptation (Brockington et al. 2015). Compared with lower vascular plants, a relaxed negative selection ($\omega = 0.19845$) on mosses might be related to the special WGD in *S. fallax* (Devos et al. 2016). A relaxation in angiosperms might be associated with the fruit-specific expression pattern. An angiosperm-specific CYP75 homolog expansion resulted in the special evolution of anthocyanin biosynthesis, which can make fruits to be colourful for fruit/seed dispersal (Petroni and Tonelli 2011; Chen 2015).

In comparison with three other bryophyte species, *S. fallax* contained a high number of CYP75 genes. This observation is congruent with the hypothesis that the last WGD in *S. fallax* contributes to their ecological dominance in peatlands (Devos et al. 2016). The majority of CYP families was found in angiosperm and gymnosperm species compared to bryophyte species. That might be related to the ζ-WGD event that occurred prior to the split of bryophyte and seed plants (Murat et al. 2017). The evolution of family number

and gene content of CYPs after WGDs in plants has been an innovation, because WGDs contributing to the genomic variation drive key innovations during angiosperm evolution (Soltis and Soltis 2016).

Our syntenic analyses revealed that the paleotetraploidy and paleohexaploidy events shaped the evolution of the CYP75 gene family in the pomegranate genome. However, inconsistent content ratios of *PgCYP75* and *EgCYP75* genes were observed in total (2:6) and syntenic (2:3) CYP75 genes. This difference may be traced back to the WGDs shared by pomegranate and *E. grandis* following which only a small proportion of duplicated genes have survived subsequent gene loss. Due to the neo-functionalization or sub-functionalization occurring after WGDs, a considerable fraction of CYPs survive, while most duplicates are lost (Lee and Irish 2011; Kaltenegger et al. 2013; Hartmann et al. 2014).

Pseudogenization, neo- or sub-functionalization, and non-functionalization are the main fates of genes following duplication. Three CYP but not CYP75 pseudogenes *Pg023834.1*, *Pg027802.1* and *Pg016410.1* clustered into the out branch of each responding clade, supporting the conclusion that pseudogenes tended to be derived from relatively ancient duplicates (Zou et al. 2009). The *Pg027802.1* was a processed pseudogene resulted from a retrotransposition event of LARD insertion in an exon. LARD is a remnant of deletion of autonomous long terminal repeat retrotransposons. Of the two pomegranate CYP75 genes, only *Pg000150.1* with a LARD element in the promoter blocks was significantly expressed during peel and aril development. The distinct CYP75 gene expression pattern during fruit development might be related to the LARD-induced promoter area. The recent pomegranate lineage-specific radiations of LARDs could be responsible for the specific functional traits in fruit development (Yuan et al. 2018). In addition, a gene duplication event might result in the neo-functionalization of CYP75 genes. The expression ratio of *Pg000150.1* to *Pg010035.1* in the cultivar with red peels increased during colouration, while that with purple, green and white peels degraded. The different expression pattern in peels might be linked with the ratio of Dp3G to Cy3G increasing in red peel (Zhao et al. 2015; Zhu et al. 2015). These two CYP75 orthologues evolved to bear distinct expression patterns. They catalyse different substrates to determine the hydroxylation pattern of the B-ring in the anthocyanin biosynthesis pathway (Tanaka 2006).

The integration of the above analyses generated a putative evolutionary scenario of the CYP75 family in land plants (Fig. 6A, B). Duplication, as well as co-speciation events, are the predominant features in the evolutionary scenario of the CYP75 family (Fig. 6A). WGDs, as the main contribution to genome evolution and speciation, shaped the CYP gene families. The CYP75 gene family was evolved from another CYP clade under purifying selection through exon fusion. Phylogenomics and molecular evolution analyses supported the recent CYP75 duplication groups that appeared prior to the split between gymnosperms and angiosperms. Neo- or sub-functionalization is the main fate of CYP75 genes after duplication. The expression pattern of homologous copies of CYP75 in pomegranate supports the evolutionary hypothesis of plant reproduction. The fruit tissue-specific expression pattern supports the CYP75 family evolution contributing to species reproduction that showy fruit colours attracted birds and other animals to spread seeds (Tanaka 2006; Chen 2015; Seitz et al. 2015). Future sampling broader taxa in land plants and functional studies in pomegranate will help to assess the full potential of the evolutionary hypothesis of plant reproduction. The CYP75 genes, as key regulator and developmental genes, provide vital resources to explore the evolution subsequent to WGDs in land-plant lineages.

Materials and Methods

Taxon Sampling for Phylogenomics and Species Tree Inference

To elucidate the evolutionary traits of the CYP75 family subsequent to WGDs in plants, we investigated twenty-six species bearing reference genome sequences. To construct a representative phylogeny across land plants, nineteen angiosperms, three gymnosperms and four bryophytes were selected (Genomics source of 26 species and the identification of shared low-copy gene families can be found in Note S1). Protein sequences for each low-copy gene set were aligned using MAFFT v7.305b (Katoh and Standley 2016) with the iterative refinement method (L-INS-i), and the aligned blocks were converted into coding sequences (CDS). Afterward, the concatenated blocks were used to infer a species tree and estimate the divergence time. A maximum likelihood (ML) phylogenetic tree was constructed with PhyML v3.1 (Guindon et al. 2010) under a subtree pruning and regrafting (SPR) topology search, 1000 bootstrap replicates, and the JTT model. Divergence times were estimated using the Reltime method (Tamura et al. 2012) under the General Time Reversible model and published calibrate time (See detail calibrate time in Note S1, Kumar et al. 2017; Murat et al. 2017).

Identification and Phylogenetic Inference of the CYP75 Gene Family

Since CYP enzymes belong to a superfamily, the CYP75 gene family was identified using a phylogenetic approach to avoid the putative errors based on the best blast hits (Koski and Golding 2001). Briefly, all putative CYP paralogs were identified in twenty-six genome sequences using the software HMMER v3.1b1 (Finn et al. 2011) with the CYP-Pfam model (Accession no. PF00067), as described previously (Zhang et al. 2017). A total of 7609 putative CYP candidates with the cut-off E-value of $\leq 10^{-10}$ were retained. We utilized MAFFT v7.305b to align CYP candidates with the 'auto' setting. To obtain the effective phylogeny signals, the low-quality alignment regions and incorrect sequences with apparent splice variants were removed by the means in Hartmann et al. (2014). Then, an ML tree of 3982 CYPs was inferred using PhyML v3.1 with the JTT model, SPR topology search, and aLRT returning statistics (Guindon et al. 2010). All CYP family clades were identified according to the known CYPs of the *Eucalyptus grandis*, *Picea abies* and *Selaginella moellendorffii*. The

CYP75 gene tree was isolated from the total CYP phylogeny using Mesquite v3.40 (http://mesquiteproject.org). Exon and intron structures of CYP genes were illustrated using the GSDS v2.0. Additionally, we extracted the pomegranate CYP gene tree from the identified CYP ML tree using Mesquite v3.40. Simplified versions of both the CYP75 gene tree and species tree were employed using Jane v4.0 (Conow et al. 2010) with 30 generations and 200 populations to infer the pattern of CYP75 gene duplication versus loss. The trees were visualized using Ggtree v1.9.4 (Yu et al. 2017).

Molecular Evolution Analysis

The positive Darwinian selection at the DNA sequence level is generally measured in comparative studies using the nonsynonymous/synonymous substitution rate ratio (ω = dN/dS) between homologous protein-coding gene sequences, with ω =1, <1, and >1, indicating neutral evolution, purifying selection, and positive selection, respectively. The survey of the selection pressures at the molecular level is crucial to understand how the lineage-specific CYP75 gene family evolved over time. Translated CYP75 protein coding sequences were aligned using MAFFT v7.305b, and those coupled with CDSs were transformed into the PAML alignment file using PAL2NAL (Suyama et al. 2006). Those results were then used to calculate the ω value using CODEML in PAML v4.9e (Yang 2007). After columns with missing data or gaps were removed, aligned sequences with at least 180 sites were retained. Ultimately, the clade model was used to further test the lineage-specific evolution pressure on CYP75 genes.

The ω value was also calculated for the pair of putative CYP pseudogene and its parent gene. After pairwise alignment, a custom Perl script was used to remove premature termination codons and codons containing gaps or repeat sequences. Yn00 in implement PAML v4.9e was applied to calculate ω value with the Yang and Nielsen (2000) method.

Estimation of CYP Duplications Associated with WGDs

The syntenic analyses were performed only among the basal rosids, *Vitis vinifera*, pomegranate and *E. grandis*, due to their well-established paleo-genomic events (Murat et al. 2017; Yuan et al. 2018). The pomegranate genome experienced a paleotetraploidy event shared by *E. grandis* and a paleohexaploidy event shared by the eudicots (Yuan et al. 2018), both of which could affect the evolution of gene families, especially the CYP superfamily. The links between the WGDs and CYP75 duplication were detected in the pomegranate, *V. vinifera* and *E. grandis* genomes. We applied the syntenic genes among species to detect the WGDs during evolution. Seven ancient eudicot chromosomes were retrieved from the inference of Murat et al. (2017). Genome synteny for the ancient eudicots versus *V. vinifera*, *V. vinifera* versus pomegranate, and pomegranate versus *E. grandis* was identified by all-against-all BLASTP (Altschul et al. 1990) followed by MCScanX (Wang et al. 2012) as described in Yuan et al. (2018). To clearly reveal the putative WGDs related to the CYP gene duplications, we only selected a small group of pomegranate scaffolds (Table S1) and highlighted the CYP genes in the syntenic blocks. We also used Circos v0.69 (Krzywinski et al. 2009) to map all 92 CYP genes to the genome scaffolds (Table S1).

Identification of Pseudogenes

CYP pseudogenes in the pomegranate genome were identified using the BLAT (Kent 2002) and GeneWise (Birney et al. 2004) software. If the frameshift (signal '!') or pre-mature stop codons occurred in the gene alignment, the gene was identified as the putative pseudogene. Additionally, the expression levels of pseudogenes tended to be lower than those of annotated genes. We also checked the expression levels of the putative CYP pseudogenes using RNA-Seq. RNA from roots, shoots, leaves, flower, and fruit was extracted from 'Taishanhong' pomegranate and mixed. Total RNA from the mixed tissues was extracted using TRI Reagent (Sigma-Aldrich, St. Louis, MO, USA). Paired-end (PE) cDNA libraries were generated and sequenced on a HiSeq 4000 platform (Illumina, San Diego, CA, USA). The original RNA-Seq data were deposited in the NCBI Sequence Read Archive (Accession no. SRP136573). Quality controls on PE reads were implemented using NGS QC Toolkit v2.3.3 (Patel and Jain 2012) with the default parameters. The clean reads were aligned with the pomegranate reference genome sequence (Yuan et al. 2018) through HISAT2 v2.1.0 (Pertea et al. 2016) to estimate the cumulative read coverage. Gene expression levels were quantified by transcripts per million (TPM) using Kallisto v0.43.1 (Bray et al. 2016). The candidate pseudogenes with high expression levels were discarded.

Expression Pattern Analyses

The public pomegranate transcriptomes presenting the development of pericarp (SRR5279392, SRR5279393, and SRR5279394) and aril (SRR5279386, SRR5279387, and SRR5279388) were downloaded from NCBI. Quality controls for these transcriptomes were performed using the NGS QC Toolkit v2.3.3. Afterward, the Kallisto v0.43.1 was employed to quantify the gene expression. The differentially expressed genes were defined as the genes with $\log_2$(fold change) > 2 and adjust.pvalue < 0.01 and identified using DESeq2 (Love et al. 2014).

To verify the expression pattern of the CYP75 family by qRT-PCR, various tissues, including young leaves and shoots, and ripening peels were collected from distinct pomegranate cultivars. These cultivars included 'Taishanhong', with whole-red peels, 'Moshiliu', with purple peels, 'Qingpitian', with green peels and 'Sanbaitian', with white peels. First-strand cDNA was synthesized from total RNA and further used for RT-PCR using a OneTaq® RT-PCR Kit (NEB, Ipswich, MA, USA) following the kit protocol. Degenerate primers were designed for PCR amplification of the *Pg010035.1* and *Pg000150.1* genes based on the predicted genes through ePCR. The primer sequences for the open reading frame (ORF) amplification of *Pg010035.1* and *Pg000150.1* display in Table S2. The purified PCR product was cloned into the vector pUC57 using a GBclonart Seamless Assembly Kit (GBI, Suzhou, China) according to the manufacturer's instructions. Several clones of each construct were sequenced. qRT amplification was performed in triplicate using the Luna® Universal qPCR Master Mix (NEB, USA). The primer sequences for qRT-PCR amplification of *Pg010035.1* and *Pg000150.1* display in Table S2. The expression levels of these genes were assessed using the method described by Liu et al. (2018).

Availability of Data and Materials

The ORFs of pomegranate CYP75 gene family sequences were deposited into NCBI GeneBank (Accession number: KY939736 and KY939737). The transcriptome data were deposited into the SRA database of NCBI (Accession number: SRP136573).

Acknowledgements

We thank Dr. Jianqing Zhou and Mengwei Zhang for preparing samples and Weicheng Yuan for data analyses and Dr. Yujie Zhao for valuable discussions. This work was supported by the National Natural Science Foundation of China (31272143), the Initiative Project for Talents of Nanjing Forestry University (GXL2014070), the Priority Academic Program Development of Jiangsu High Education Institutions (PAPD), the Research Fund for Postgraduate

Innovation Project of Jiangsu Province (KYLX16_0857), and the Doctorate Fellowship Foundation of Nanjing Forestry University.

Authors' Contributions

TZ and ZY developed the project and designed the experiments; TZ, XH and CL performed the experiments; TZ, HZ and ZY wrote and revised the manuscript; All authors discussed the results and commented on the manuscript. All authors read and approved the final manuscript.

Supporting Information

Note S1. Taxon sampling and phylogenetic analysis
Note S2. Phylogenomic analyses for CYP gene families
Table S1. Genebank accession number of pomegranate scaffolds
Table S2. Primer sequences
Table S3. Summary of low-copy phylogenetic markers
Table S4. PAML analysis of CYP75 genes

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Birney E, Clamp M, Durbin R (2004) Genewise and genomewise. Genome Res 14:988–995

Bowman JL, Kohchi T, Yamato KT, Jenkins J, Shu S, Ishizaki K, Yamaoka S, Nishihama R, Nakamura Y, Berger F, Adam C, Aki SS, Althoff F, Araki T, Arteaga-Vazquez MA, Balasubrmanian S, Barry K, Bauer D, Boehm CR, Briginshaw L, Caballero-Perez J, Catarino B, Chen F, Chiyoda S, Chovatia M, Davies KM, Delmans M, Demura T, Dierschke T, Dolan L, Dorantes-Acosta AE, Eklund DM, Florent SN, Flores-Sandoval E, Fujiyama A, Fukuzawa H, Galik B, Grimanelli D, Grimwood J, Grossniklaus U, Hamada T, Haseloff J, Hetherington AJ, Higo A, Hirakawa Y, Hundley HN, Ikeda Y, Inoue K, Inoue S-i, Ishida S, Jia Q, Kakita M, Kanazawa T, Kawai Y, Kawashima T, Kennedy M, Kinose K, Kinoshita T, Kohara Y, Koide E, Komatsu K, Kopischke S, Kubo M, Kyozuka J, Lagercrantz U, Lin S-S, Lindquist E, Lipzen AM, Lu C-W, De Luna E, Martienssen RA, Minamino N, Mizutani M, Mizutani M, Mochizuki N, Monte I, Mosher R, Nagasaki H, Nakagami H, Naramoto S, Nishitani K, Ohtani M, Okamoto T, Okumura M, Phillips J, Pollak B, Reinders A, Rövekamp M, Sano R, Sawa S, Schmid MW, Shirakawa M, Solano R, Spunde A, Suetsugu N, Sugano S, Sugiyama A, Sun R, Suzuki Y, Takenaka M, Takezawa D, Tomogane H, Tsuzuki M, Ueda T, Umeda M, Ward JM, Watanabe Y, Yazaki K, Yokoyama R, Yoshitake Y, Yotsui I, Zachgo S, Schmutz J (2017) Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. Cell 171:287–304

Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-Seq quantification. Nat Biotechnol 34:525–527

Brockington SF, Yang Y, Gandia-Herrero F, Covshoff S, Hibberd JM, Sage RF, Wong GKS, Moore MJ, Smith SA (2015) Lineage-specific gene radiations underlie the evolution of novel betalain pigmentation in Caryophyllales. New Phytol 207:1170–1180

Chen C (2015) Overview of plant pigments, In Chen C ed, Pigments in fruits and vegetables: genomics and dietetics, Springer New York, New York, NY pp 1–7

Conow C, Fielder D, Ovadia Y, Libeskind-Hadas R (2010) Jane: a new tool for the cophylogeny reconstruction problem. Algorithms Mol Biol 5:16

Devos N, Szövényi P, Weston DJ, Rothfels CJ, Johnson MG, Shaw AJ (2016) Analyses of transcriptome sequences reveal multiple ancient large-scale duplication events in the ancestor of Sphagnopsida (Bryophyta). New Phytol 211:300–318

Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. Nucleic Acids Res 39:W29–W37

Graham SA (2013) Fossil records in the Lythraceae. Bot Rev 79:48–145

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59:307–321

Hartmann A-M, Tesch D, Nothwang HG, Bininda-Emonds ORP (2014) Evolution of the cation chloride cotransporter family: ancient origins, gene losses, and subfunctionalization through duplication. Mol Biol Evol 31:434–447

Holland D, Hatib K, Bar-Ya'akov I (2009) Pomegranate: botany, horticulture, breeding, In Janick J ed, Hort Rev, Vol 35. John Wiley & Sons, Inc., Hoboken, NJ, USA pp 127–191

Hollister JD (2015) Polyploidy: adaptation to the genomic environment. New Phytol 205:1034–1039

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, Soltis DE, Clifton SW, Schlarbaum SE, Schuster SC, Ma H, Leebens-Mack J, dePamphilis CW (2011) Ancestral polyploidy in seed plants and angiosperms. Nature 473:97

Kaltenegger E, Eich E, Ober D (2013) Evolution of homospermidine synthase in the Convolvulaceae: a story of gene duplication, gene loss, and periods of various selection pressures. The Plant Cell 25:1213–1227

Katoh K, Standley DM (2016) A simple method to control over-alignment in the MAFFT multiple sequence alignment program. Bioinformatics 32:1933–1942

Kent WJ (2002) BLAT--the BLAST-like alignment tool. Genome Res 12:656–664

Koski LB, Golding GB (2001) The closest BLAST hit is often not the nearest neighbor. J Mol Evol 52:540–542

Krzywinski MI, Schein JE, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: an information aesthetic for comparative genomics. Genome Res 19:1639–1645

Kumar S, Stecher G, Suleski M, Hedges SB (2017) TimeTree: a resource for timelines, timetrees, and divergence times. Mol Biol Evol 34:1812–1819

Lee H-L, Irish VF (2011) Gene duplication and loss in a MADS box gene transcription factor circuit. Mol Biol Evol 28:3367–3380

Liu X, Liu X, Zhang Z, Sang M, Sun X, He C, Xin P, Zhang H (2018) Functional analysis of the *FZF1* genes of *Saccharomyces uvarum*. Front Microbiol 9:96

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550

Magallón S, Hilu KW, Quandt D (2013) Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. Am J Bot 100:556–573

Murat F, Armero A, Pont C, Klopp C, Salse J (2017) Reconstructing the genome of the most recent common ancestor of flowering plants. Nat Genet 49:490–496

Nelson DR, Schuler MA, Paquette SM, Werck-Reichhart D, Bak S (2004) Comparative genomics of rice and Arabidopsis. Analysis of 727 cytochrome P450 genes and pseudogenes from a monocot

and a dicot. Plant Physiol 135:756–772
Panchy N, Lehti-Shiu M, Shiu S-H (2016) Evolution of gene duplication in Plants. Plant Physiol 171:2294–2316
Patel RK, Jain M (2012) NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLoS One 7:e30619
Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL (2016) Transcript-level expression analysis of RNA-Seq experiments with HISAT, StringTie and Ballgown. Nat Protoc 11:1650–1667
Petroni K, Tonelli C (2011) Recent advances on the regulation of anthocyanin synthesis in reproductive organs. Plant Sci 181:219–229
Renault H, De Marothy M, Jonasson G, Lara P, Nelson DR, Nilsson I, André F, von Heijne G, Werck-Reichhart D (2017) Gene duplication leads to altered membrane topology of a cytochrome P450 enzyme in seed plants. Mol Biol Evol 34:2041–2056
Seitz C, Ameres S, Schlangen K, Forkmann G, Halbwirth H (2015) Multiple evolution of flavonoid 3',5'-hydroxylase. Planta 242:561–573
Soltis PS, Soltis DE (2016) Ancient WGD events as drivers of key innovations in angiosperms. Curr Opin Plant Biol 30:159–165
Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res 34:W609–W612
Tamura K, Battistuzzi FU, Billing-Ross P, Murillo O, Filipski A, Kumar S (2012) Estimating divergence times in large molecular phylogenies. Proc Natl Acad Sci U S A 109:19333–19338
Tanaka Y (2006) Flower colour and cytochromes P450. Phytochem Rev 5:283–291
Van de Peer Y, Mizrachi E, Marchal K (2017) The evolutionary significance of polyploidy. Nat Rev Genet 18:411
Victor C, Marçal S, Charles H, Hua CW, Pedro F, A. MA, P. PJA, Jacqueline GP (2015) Genome-wide analysis of the lignin toolbox of *Eucalyptus grandis*. New Phytol 206:1297–1313
Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee T-h, Jin H, Marler B, Guo H, Kissinger JC, Paterson AH (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res 40:e49
Wei K, Chen H (2018) Global identification, structural analysis and expression characterization of cytochrome P450 monooxygenase superfamily in rice. BMC Genomics 19:35
Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591
Yang Z, Nielsen R (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol 17:32–43
Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y (2017) Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol 8:28–36
Yuan Z, Fang Y, Zhang T, Fei Z, Han F, Liu C, Liu M, Xiao W, Zhang W, Wu S, Zhang M, Ju Y, Xu H, Dai H, Liu Y, Chen Y, Wang L, Zhou J, Guan D, Yan M, Xia Y, Huang X, Liu D, Wei H, Zheng H (2018) The pomegranate (*Punica granatum* L.) genome provides insights into fruit quality and ovule developmental biology. Plant Biotechnol J 16:1363–1374
Zhang T, Liu C, Zhang H, Yuan Z (2017) An integrated approach to identify Cytochrome P450 superfamilies in plant species within the Malvids, In Zhang D ed, Proceedings of the 5th International Conference on Bioinformatics and Computational Biology, ACM, New York, NY, USA pp 11–16
Zhao X, Yuan Z, Yin Y, Feng L (2015) Patterns of pigment changes in pomegranate (*Punica granatum* L.) peel during fruit ripening. Acta Hortic 1089:83–89
Zhu FZ, Yuan ZH, Zhao XQ, Yin YL, Feng LJ (2015) Composition and contents of anthocyanins in different pomegranate cultivars. Acta Hortic 1089:35–41
Zou C, Lehti-Shiu MD, Thibaud-Nissen F, Prakash T, Buell CR, Shiu S-H (2009) Evolutionary and expression signatures of pseudogenes in Arabidopsis and rice. Plant Physiol 151:3–15