Genome-wide analysis of the citrus B3 superfamily and their association with somatic embryogenesis

Zheng Liu
Fruit and Tea Research Institute, Hubei Academy of Agricultural Sciences
Xiao-Xia Ge
Center of Applied Biotechnology, Wuhan University of Bioengineering
Xiao-Meng Wu (wuxm@mail.hzau.edu.cn)
https://orcid.org/0000-0001-5798-5734
Wen-Wu Guo
Key Laboratory of Horticultural Plant Biology (Ministry of Education), Huazhong Agricultural University

Research article

Keywords: Citrus, B3 surperfamily, Phylogenetic analysis, Somatic embryogenesis, Callus initiation, Expression profile

Posted Date: July 30th, 2019

DOI: https://doi.org/10.21203/rs.2.12055/v1

Version of Record: A version of this preprint was published at BMC Genomics on April 16th, 2020. See the published version at https://doi.org/10.1186/s12864-020-6715-9.

Abstract

Background In citrus, genetic improvement via biotechnology is hindered by the obstacle of in vitro regeneration via somatic embryogenesis (SE). Although a few of B3 transcription factors are reported to regulate embryogenesis, little is known about the possible roles of B3 superfamily during SE especially in citrus. Results In this study, a total of 72 (CsB3) and 69 (CgB3) putative B3 superfamily members were identified in the sweet orange (Citrus sinensis) and pummelo (C. grandis) genomes, respectively, each comprised four gene families and 14 phylogenetic classes. The B3 genes were unevenly distributed over citrus chromosomes and other non-anchored scaffolds. Genome duplication analysis indicated that the segmental and tandem duplication events have significantly contributed to the expansion of the citrus B3 superfamily. The evolutionary relationships among the B3 family members and their putative functions were deduced based on the results of phylogenetic analysis. Furthermore, transcriptomic analysis showed that citrus B3 genes have differential expression levels in various tissues, suggesting distinct biological roles of different members. Expression analysis revealed that the B3 superfamily members showed four types of expression profiles during SE in citrus and may play functional roles during SE, especially at late SE stages. Of them, CsARF19 is specifically expressed in sweet orange and at markedly higher levels in the embryogenic callus (EC), implying its possible involvement in EC initiation. Conclusions This study provides a genome-wide analysis of citrus B3 superfamily, including its genome organization, evolutionary features and expression profiles, which contributes to a better understanding of the B3 genes in citrus and their association with SE.

Background

B3 transcription factors (TFs), which contain at least one B3 DNA-binding domain, constitute one of the plant-specific superfamilies [1, 2]. The B3 domain was initially named according to its position in the third basic region of *VIVIPAROUS1 (VP1)* from maize [3]. The conserved B3 domain comprises approximately 110 amino acid residues for DNA recognition, consisting of seven β-barrels and two short α helices [1, 2]. According to the domain structures and phylogenetic analysis, the B3 superfamily is divided into four major families, namely the LAV (*L*EAFY COTYLEDON2-*A*BSCISIC ACID INSENSITIVE3-*V*AL), RAV (*R*ELATED TO *A*BI3/ *V*P1), ARF (*A*UXIN *R*ESPONSE *F*ACTOR) and REM (*RE*PRODUCTIVE *M*ERISTEM) families [1]. The B3 superfamily has been identified in a few model plants and crops, including *Arabidopsis,* rice, poplar, *Brassica rapa,* castor bean, cocoa, soybean, maize, moss and algae [1, 4, 5], but not yet in fruit trees like citrus.

It is reported that B3 TFs from distinct families regulate and control different aspects of plant growth and development. LAV family members, including *LEC2* (*LEAFY COTYLEDON2)*,, *FUS3* (*FUSCA3)*,,*ABI3* (*ABSCISIC ACID INSENSITIVE3)*,,*VAL1 (VP1/ABI3-LIKE 1)*,, *VAL2* and *VAL3,* of whicheach possesses a single B3 domain, regulate callus induction, embryo development and phase transition [6–14]. For instance, overexpression of *AtLEC2* in transgenic plants induced the formation of callus and somatic embryos [7]. The hypocotyls of double knockout mutants of *AtVAL1* and *AtVAL2* developed into yellow callus-like structures of embryonic activities, revealing the function of *AtVAL1* and *AtVAL2* in preventing

the expression of embryonic traits after seed germination [14]. The LAV family generally consists of two subgroups: the LEC2-ABI3 subgroup (*LEC2, ABI3* and *FUS3)* recognizes the Sph/RY motif (CATGCA) in the promoters of seed-specific genes [4, 5, 15], whereas the other subgroup VAL (*VAL1, VAL2* and *VAL3)* are expressed in many organs throughout plant development and have central roles in mediating repression of the *LEC1/LEC2-ABI3* subgroup network during seed germination [1, 9, 16]. The RAV family proteins contain a C-terminal B3 domain that recognize the consensus sequence CACCTG [17]. Some members of the RAV family also possess an N-terminal AP2/ERF domain that recognizes the consensus sequence CAACA. RAV family controls flowering, organ growth, leaf senescence, hormone signaling and responses to various stresses [17–24]. The ARF family proteins have an N-terminal B3 domain that recognizes the auxin response element TGTCTC in the promoter of gene responsive to auxin, followed by a highly divergent middle region that determines whether the *ARFs* acts as an activator or repressor, and end with a conserved carboxyl-terminal interaction domain containing motifs III and IV [25, 26]. *ARF* genes have been widely implicated in auxin-mediated response during various developmental processes from embryogenesis to flowering, and fruit development [27–33]. The REM family members contain at least one copy of B3 domain, sometimes up to seven repeats. However, it is not clear whether the B3 domain of REM protein binds to a specific recognition sequence [34]. The functions of REM genes are not well understood as the other three families of B3 superfamily, except that some genes including *REM1, VRN1* and *VOD* were validated to be involved in floral meristems formation, vernalization and female gametophyte development [35–37].

Citrus is one of the most important fruit crops in the world. However, conventional breeding of citrus is largely hindered by its inherent characteristics, such as nucellar polyembryony, long juvenility and male/female sterility [38]. The genetic improvement via biotechnology could be an effective approach, but it is hindered by the barrier of plant regeneration through somatic embryogenesis (SE). Empirically, the embryogenic callus (EC) can only be induced from the aborted seeds of polyembryonic (apomictic) citrus genotypes, but not from the monoembryonic (sexual) genotypes. In addition, the embryonic potential of EC gradually decreases during callus subculture. To understand the mechanisms of SE and overcome the obstacle of citrus SE, we have conducted series of studies and identified a few genes, proteins and miRNAs involved in citrus SE [39–41]. We found that the B3 domain regulatory network genes *CsFUS3, CsABI3* and another B3 gene (CS_P006_E_03) exhibited increased expression during citrus SE induction and formation [39], whereas *CsFUS3* was validated to promote citrus SE partially by regulating SE-related TFs and hormone pathways, especially ABA and GA pathways [42]. To understand the regulatory roles of B3 superfamily genes in citrus SE, in this study, we performed a genome-wide analysis of the B3 superfamily in the polyembryonic sweet orange and the monoembryonic pummelo [43, 44]. The chromosome localization, gene structure, domain architecture and phylogenetic relationship were systematically analyzed to infer the functional features of citrus B3 superfamily. The expression profiles ofcitrusB3 TFs were analyzed among different tissues and the developmental stages during SE. To explore the potential *B3* genes associated with EC initiation in SE, we identified the B3 genes specific to the genome of the polyembryonic sweet orange compare to that of the monoembryonic pummelo, and

analyzed their tissue-specific expression patterns. The comprehensive study of the B3 surperfamily should enhance our understanding of possible roles of B3 genes in citrus development, especially in SE.

Results

Identification and genomic distribution of B3 superfamily in citrus

A total of 72 (*CsB3)* and 69 (*CgB3)* B3 superfamily TFs were identified in the sweet orange and pummelo genomes, respectively (Additional file 1). As reported in *Arabidopsis* [1], the B3 superfamily members could be classified into LAV, RAV, ARF and REM family according to their sequence similarity. We further named these genes based on the family categories. In the present study, REM was found to be the biggest B3 family, with 52.8% (38 *CsREMs)* and 55.1% (38 *CgREMs)* of the total *B3* genes identified in sweet orange and pummelo, respectivelyAdditional file 1). The ARF family constituted the second largest group of B3 superfamily, consisting of 26.4% (19 *CsARFs)* and 24.6% (17 *CgARFs)* of the total B3 genes of sweet orange and pummelo, respectively. On contrary, LAV and RAV are two relatively small families, with 11.1% (8 *CsLAVs)* and 9.7% (7 *CsRAVs)* of *B3* genes identified in sweet orange, and 11.6% (8 *CgLAVs)* and 8.7% (6 *CgRAVs)* of *B3* genes identified in pummelo.

CsB3 TFs were distributed over eight of the nine sweet orange chromosomes. None of the *CsB3* genes was located on chromosome 9 (Fig. 1A). The *CsB3* gene density per chromosome was patchy, with only three genes (4.2%) i.e. *CsRAV5, CsARF11* and *CsARF17* on chromosome 4, but up to 17 (23.6%) of the 72 members on chromosome 5. Relatively high densities of *CsB3* genes were observed at the chromosome ends, of which the highest was located at the bottom of chromosome 5. However, the chromosomal locations for 10 *CsB3* genes were not defined because of the incompleteness of sweet orange physical genome map. Meanwhile, the distribution and density of the *CgB3* TFs were also not uniform on the nine chromosomes of pummelo (Fig. 1B). Chromosome 8 encompassed the largest number of 19 (27.5% of) *CgB3* genes, whereas on chromosome 1 there were only three (4.3% of) *CgB3* genes. The orthologous genes of B3 superfamily between sweet orange and pummelo were not located consistently on the same citrus chromosomes. For example, *CsLAV7* was on chromosome 1 of sweet orange (Fig. 1A), whereas its orthologous gene *CgLAV7* was on chromosome 2 of pummelo (Fig. 1B). These different locations of B3 TFs on chromosomes between citrus species indicated that genetic recombination have extensively occurred in citrus varieties. Among all identified *CsB3* genes, a total of 10 chromosomal segmental duplication events and 4 tandem duplication events were identified in the sweet orange genomes, whereas in the pummelo genome the corresponding events were 11 and 9 respectively (Fig.1 and Additional file 2), indicating that segmental and tandem duplications may contribute to the expansion of citrus B3 superfamily. Segmentally duplicated gene pairs (average Ka/Ks = 0.22, Ka/Ks also known as non-synonymous/synonymous substitution ratio) appeared to have undergone extensive intense purifying selection compared to tandemly duplicated gene pairs (average Ka/Ks = 0.52). The Ka/Ks ratios for the majority (82.4%) of duplicated pairs were less than 0.5, suggesting that citrus *B3* superfamily had

evolved under the effect of purifying selection. However, the other two tandemly duplicated gene pairs (*CgREM28–1/CgREM28–2* and *CgREM6–1/CgREM29–2)* seemed to be under neutral selection, as their Ka/Ks ratios were close to 1.0.

To further explore the phylogenetic relationship of *B3* superfamily genes between citrus and other plant species, comparative syntenic analyses were conducted in a pairwise manner (Fig. 2), with 37 and 24 collinear *B3* gene pairs identified in the sweet orange/*Arabidopsis* and sweet orange/rice pairs, respectively (Additional file 3). For pummelo/*Arabidopsis* and pummelo/ricepairs the corresponding gene pair numbers were 39 and 24. The number of orthologous events of *CsB3/CgB3-AtB3* was higher than that of *CsB3/CgB3-OsB3,* indicating that the divergence between citrus and *Arabidopsis* occurred after the divergence of the rice and their common ancestor of dicotyledons. Note that some *B3* collinear gene pairs of citrus/*Arabidopsis* were anchored to the highly conserved syntenic blocks, in which the number of syntenic gene pairs was up to 246, whereas none of syntenic blocks of citrus/*Oryza sativa* pair contained more than 20 genes (Additional file 3). The high level of syntenic conservation between the citrus and *Arabidopsis* indicated that B3 TFs in citrus might share the similar structure and function with orthologs in *Arabidopsis.*

Characterization of B3 proteins in citrus

To understand the molecular characteristics of B3 proteins in citrus, their physiochemical properties were analyzed. The amino acids length of putative citrus B3 proteins varied widely, ranging from 93 to 1134. The molecular weights and theoretical isoelectric points were also diverse (Additional file 1).

The majority of B3 TFs contained only one B3 domain except for some REM family members in citrus (Fig. 3D and 4D). Seven β-barrels and two short α helices of the known core structure were present among the B3 domains (Additional file 4 and 5). Amino acid sequences alignments showed that the B3 domain sequences were highly conserved in LAV, RAV and ARF families (Additional file 4), whereas the B3 domains of REM family exhibited a higher degree of divergence (Additional file 5). A total of 20, 38, and 24 highly conserved amino acid residues were identical among the B3 domains of all the LAV, RAV, and ARF family members, respectively (Additional file 4). For REM family members, only some conserved amino acid residues including one proline (position 27, P), two tryptophans (position 52 and 69, W), three glycines (position 49, 68 and 81, G) and three phenylalanines (position 30, 72 and 86, F) were observed in the B3 domains (Additional file 5), which indicated that the B3 domain might have been evolved independently in REM family.

In total, the five conserved motifs, viz. B3, AP2, AUX/IAA, ARF and CW-type zine finger, were identified in the B3 members (Fig. 3D and 4D). The number of the conserved motifs in each B3 protein varied from one to three. Each family of B3 proteins specifically shared some other conserved motifs, in addition to the B3 domain. For example, motifs ARF and AUX/IAA were specifically shared by ARF family, and the motif CW-type zinc finger and AP2 exclusively appeared in the LAV family and RAV family, respectively.

Although most of these conserved motifs remain to be functional elucidated, it is likely that these motifs were evolutionarily conserved and functionally diversified in the specific families.

Phylogenetic analyses of B3 genes

To explore the phylogenetic relationships of B3 superfamily, an unrooted phylogenetic tree was constructed among the *B3* genes of citrus (sweet orange and pummelo) and the model plant *Arabidopsis* (Additional file 6). According to the classification criteria in *Arabidopsis,* we further divided the four family members into fourteen major classes (Fig. 3A and 4A).

In detail, the LAV family could be subdivided into two classes, i.e. LEC2-ABI3classI) and VALclass (II). Four *CsLAVs* in sweet orange (*CsLAV1, CsLAV2, CsLAV6* and *CsLAV8)* and theircounterparts in pummelo (*CgLAV1, CgLAV2, CgLAV6* and *CgLAV8)* were clustered with *Arabidopsis* LEC2-ABI3subgroup. The VAL subgroup of four citrus *LAV* genes (*CsLAV3/CgLAV3, CsLAV4/CgLAV4, CsLAV5/CgLAV5* and *CsLAV7/CgLAV7),*, which had conserved B3 domain and CW-type zinc finger, were clustered with three *Arabidopsis* VAL proteins (Fig. 3 and Additional file 6).

The RAV family was grouped into two main classes based on their phylogenetic relationship. The Class I comprise three citrus RAV genes (*CsRAV1/CgRAV1, CsRAV2/CgRAV2* and *CsRAV4/CgRAV4)* that clustered with four *AtNGA* genes and three *AtRAV-like* genes from the same branch (Fig. 3A and Additional file 6). These genes commonly had the conserved B3 domain and contained no more than one intron (Fig. 3C and 3D). Classes II was comprised of four *CsRAV* genes (*CsRAV3, CsRAV5, CsRAV6* and *CsRAV7)* and three *CgRAV* genes (*CgRAV3, CgRAV5* and *CgRAV6),*, featured by a B3 domain with an upstream AP2 domain (Fig. 4D), which have no intron except *CgRAV5 (*Fig. 3C).

All citrusARF geneswere classified into four major classes. Class I and II belonged to the same branch, and contained 6 members (*CsARF1/CgARF1, CsARF3/ CgARF3, CsARF5/CgARF5, CsARF11/CgARF11, CsARF17/CgARF17* and *CsARF18)* and 5 members (*CsARF2/CgARF2, CsARF7/CgARF7, CsARF8/CgARF8, CsARF15/CgARF15* and *CsARF16/CgARF16),*, respectively (Fig. 3A and Additional file 6). Most of them were characterized by the B3 DNA binding domain, ARF and AUX/IAA (Fig. 3D). Class III (*CsARF4/CgARF4, CsARF6/CgARF6, CsARF10/CgARF10* and *CsARF19)* and Class IV (*CsARF9/CgARF9, CsARF12/CgARF12-CsARF14/CgARF14)* only had the B3 and ARF domains. All the coding sequences of ARF geneswere disrupted by introns, the number of which ranges from 2 to 15 (Fig. 3C).

As most of *REMs* in citrus possessed multiple B3 domains and had low sequence similarity with each other (Fig. 4D and Additional file 5), we decided to perform the phylogenetic analysis within each class of REM family. The first step of phylogenetic analysis was the comparison of the AtREM sequences with CsREM/CgREM sequences according to the previous study [4] (Additional file 6). After this initial analysis, six common REM types (REM I and REM VI to REM X) were identified between citrus and *Arabidopsis,* whereas REM V type (*AtREM5)* was exclusively identified in *Arabidopsis.* The vast majority of class I and class II genes contained one B3 domain, and shared homology with the AtREM I and VII type genes,

respectively (Fig. 4 and Additional file 6). The classes III and IV genes belonged to the AtREM IX and X type, respectively, which possessed only one B3 domain and presented a relatively low expression level among the most detected tissues. Class V (AtREM VI) and class VI (AtREM VIII) genes contained several members, the majority of which had more than one B3 domain.

Expression profiles of B3 genes in different tissues and during somatic embryogenesis

To understand the tissue expression profiles of the *B3* genes in citrus, we compared their transcript abundance based on the previously published RNA-seq data of different tissues including leaf, fruit callus, flower, ovule and seed (Fig. 3B and 4B). The hierarchical cluster analysis showed that many citrus B3 genes exhibited high transcript abundance level in all the tissues. However, LEC2-ABI3subgroup and two REM classes (REM IX type and REM X type) exhibited relatively lower expression level compared with other *CsB3* genes. In addition, some of the B3 TFs exhibited tissue-specific expression. For example, *CsLAV1/2/6/7, CsARF9/19, CsREM3/4/6/7/9/13/14/17/27/28/29* showed the highest transcript abundance in the embryogenic callus (EC), whereas *CsREM24* was expressed predominantly in fruit. These genes may be involved in certain biological processes that occurred in the corresponding tissues. Some duplicated gene pairs also showed divergent expression profiles. For example, *CgARF13* showed a low expression level (RPKM = 2.76; RPKM: reads per kilobase per million mapped reads) in fruit; whereas its duplicated gene, *CgARF14,* was highly expressed (RPKM = 56.13) in fruit. These results may suggest that duplicated genes may evolve to have diverse functions. Some clustered citrus *B3* genes, which were considered as orthologous genes between sweet orange and pummelo species, showed different expression profiles. For example, *CgARF17* was mainly expressed in leaf (RPKM = 59.06) and ovule (RPKM = 57.40) of pummelo, whereas its orthologous gene*CsARF17)* of sweet orange showed relatively low expression in all detected citrus tissues, with RPKM values ranged from 4.16 to 7.57. These species-specific expression differences suggest that novel functional roles of B3 genes might have been generated during citrus domestication.

To explore the possible involvement of *CsB3* genes during citrus SE, the expression profile of *23 CsB3* genes was investigated by qRT-PCR in the six SE stages of ‘Valencia’ orange, a citrus variety with strong SE capability. These genes were carefully selected based on their relatively high transcript abundance or specifically higher expression level in EC according to RNA-seq data. Based on their expression profiles, these genes could be classified into four types (Fig. 5). The expression of Type I genes was up-regulated during differentiation and showed a relative high peak value at E2 stage (embryogenic callus induced for somatic embryos for 2 weeks *CsARF1, CsARF14, CsREM17* and *CsREM18)* or E4 stage (embryogenic callus induced for somatic embryos for 4 week *CsLAV1, CsREM4, CsREM5 CsREM13* and *CsREM29)*, and then down-regulated at the early embryo morphogenesis stage (GE, globular embryos), whereas they showed another high peak at late embryo morphogenesis stage (CE, cotyledon embryos). Type II genes comprise five *CsLAVs* (*CsLAV2, CsLAV3, CsLAV5, CsLAV6* and *CsLAV7)*, one *CsRAV* (*CsRAV3)*, two

CsARFs (*CsARF5* and *CsARF19)* and one *CsREM* (*CsREM27)*,, and specifically expressed highly at CE stage, some of which also showed high transcript abundance in one other stage. For Type III genes (*CsLAV4, CsARF12* and *CsREM6)*,, the mRNA abundance was down-regulated during differentiation stages (E0-E4, embryogenic callus induced for somatic embryos for 0–4 weeks), but was higher at the subsequent stages of embryo morphogenesis (GE or CE). However, genes in type IV (*CsARF7* and *CsREM9)* increased progressively throughout the whole SE process.

Candidate B3 TFs potentially involved in embryogenesis and callus initiation

To identify the B3 regulatory factors potentially involved in embryogenesis and callus initiation, protein sequence and expression pattern were compared among the B3 genes of sweet orange and pummelo (Fig. 3 and 4). A total of 15 *CsB3* genes which were specifically accumulated in EC were retrieved from the RNA-seq data, including five *CsLAVs* (*CsLAV1* to *CsLAV4* and *CsLAV7)*,, two *CsARFs* (*CsARF12* and *CsARF19)* and eight *CsREMs (CsREM4* to *CsREM7, CsREM9, CsREM13, CsREM27, CsREM29)* (Fig. 3B and 4B). Among their orthologous genes, eight (five *CgLAVs, CgREM13, CgREM27* and *CgREM29–1)* were preferentially expressed in the seeds of pummelo, suggesting that these genes may associated with embryogenesis in *vivo* and in *vitro.* Meanwhile, eight B3 genes were identified in the genome of sweet orange, but not in that of pummelo, including *CsRAV7, CsARF18, CsARF19, CsREM24, CsREM25, CsREM33, CsREM37* and *CsREM38.* Among them, *CsARF19* (Cs7g02210) showed markedly high expression levels (≥6-fold) in EC compared with the other tissues (Fig. 3B), indicating its potential association with callus initiation, because empirically, EC can only be induced from the seeds of the polyembryonic citrus genotypes. With the availability of the citrus genome sequences [43–47], two orthologs of *CsARF19,* MSYJ162170.1 (amino acids sequence identity of 99.36%) and Ciclev10030751m (amino acids sequence identity of 99.87%), were identified in Mangshan mandarin (*Citrus reticulate,* a wild mandarin) and Clementine mandarin (*C. clementina*, which is believed to be a chance hybrid of mandarin and sweet orange) [45, 47, 48], respectively, but not in atalantia (*Atalantia buxifolia,* a primitive citrus), Ichang papeda (*C. ichangensis,* a wild citrus) and three relative genera of citrus, *viz.* hongkong kumquat (*Fortunella hindsii)*,, trifoliate orange (*Poncitrus trifoliate)* and citron (*C. medica)*..

Discussion

The role of B3 TFs in the regulation of embryo maturation in the seed and transition from late embryo development to germination has been studied previously [6, 12, 16, 49–51]. Some of the advances have expanded our understanding of B3 network that may regulate SE [7, 39, 42]. It is believed that multiple members of a specific gene family could form a large regulative network to control complicated developmental processes. To better understand the function of citrus *B3* genes during SE, we identified the *B3* superfamily genes from the genomes of the polyembryonic (sweet orange) and monoembryonic (pummelo) citrus, and portrayed their structural and phylogenetic features. Expression profiles derived

from transcriptome data and qRT-PCR analysis indicated that a few *CsB3* genes were regulated during citrus SE. *CsARF19* showed preferential expression in EC of the polyembryonic sweet orange but not present in monoembryonic pummelo, which indicated its association with EC initiation and SE. The study provides promising B3 candidates to investigate their functions in citrus SE.

The colinearity and duplication of B3 superfamily reflected evolutionary imprint of citrus genomes

The B3 superfamily is one of the largest and most diverse gene families in plants [1]. The evolution of the B3 superfamily has a long history, which can be traced back to the single-celled green algaes *Chlamydomonas reinhardtii* and *Ulva linza,* which possess a single B3 gene, strongly suggesting that the B3 domain arose before the development of multicellularity in the plant lineage [1]. In this study, we performed a comprehensive search for B3 superfamily genes throughout citrus genomes. A total of 72 *CsB3* and 69 *CgB3* genes were identified, accounting for 0.24% (29,445 predicted genes in sweet orange) and 0.23% (30,123 predicted genes in pummelo) of all predicted protein-coding genes [43, 44], which was lower compared to the result in *Arabidopsis thaliana* (110 *AtB3* genes out of the 25,498 predicted genes, accounting for 0.43%) [52] but higher than that in *Oryza sativa* (87 *Os*B3 genes out of the 53398 predicted gene, accounting for 0.16%) [53] (Additional file 7). These results showed that the B3 TFs unequally expand along with the increase of genome size, since the sizes of the sweet orange and pummelo genomes (367 Mb and 380.76 Mb, respectively [43, 44]) are about triple of that of *Arabidopsis (*125 Mb) [52]. However, there were more identified *AtB3* genes in *Arabidopsis* than in citrus (Additional file 1 and 7), which probably was because of the lack of recent whole-genome duplication (WGD) event in citrus, whereas *Arabidopsis* has experienced two additional rounds of recent WGD events [43]. The phylogenetic trees showed that most of the clades contained citrus and *Arabidopsis* B3 proteins, indicating the substantial conservation of the *B3* gene families between these two species (Additional file 6). In addition, the syntenic conservation was higher among the *B3* genes in citrus/*Arabidopsis* pair compared to that in citrus/rice pair (Fig. 2 and Additional file 3). These suggest a possible link between the existence of conserved syntenic blocks and the evolution of genomic imprint. *CsB3* genes in sweet orange share significant homology with their orthologous *CgB3* genes in pummelo (Fig. 3A and 4A), which supports the theory of broad sexual compatibility among citrus species [44]. Sweet orange has been recognized as a backcross hybrid between pummelo and mandarin [43]. In this study, the eight orange-specific *B3* genes, of which the orthologs are not identified in pummelo, suggested their origination from the mandarin parent and /or additional genetic changes that might have occurred afterward, and may have contributed to orange/mandarin-specific biological features.

Superfamilies generally originate from gene/genome duplication and follow different evolutionary process, which leave an evolutionary imprint [54]. It is believed that the emergence of the duplicated genes not only is a way of genomic rearrangement and expansion, but also diversifies gene function to ensure optimal adaptability during evolution processes [55, 56]. We found that 14 (10 segmental

duplications and 4 tandem duplications) out of 72 *CsB3* genes (19.4%) and 20 (11 segmental duplications and 9 tandem duplications) out of 69 *CgB3* genes (29.0%) were duplicated genes in sweet orange and pummelo, respectively (Additional file 2), which should contribute to the expansion of citrus B3 superfamily. It is known that the segmental duplications multiply genes through chromosome rearrangements derived from WGDs [57]. More than half of the segmental duplication events were shared by pummelo (basic citrus species) and sweet orange (interspecific citrus hybrid) (Additional file 2), which further verified our hypothesis that most of the segmentally duplicated citrus *B3* genes are resulted from the ancient WGD of citrus. Tandem duplications were characterized as multiple members resulted from unequal crossing-over during meiosis [58]. Although only a limited number of genes are affected, tandem duplication can happen relatively frequently [59]. However, functional redundancy generated by frequent gene duplication is often not advantageous [60]. For instance, two tandemly duplicated genes pairs (*CgREM28–1/CgREM28–2* and *CgREM28–2/CgREM28–3)*,, in which the Ka/Ks ratios were 1.06 and 0.69, respectively, had relatively short coding sequence length and showed very low expression levels in all detected samples (RPKM<1 by RNA-Seq) (Fig. 4), indicating that they may be not under strong selection and probably become pseudogenes during long-term evolution. On the other hand, the rapid functional divergence and the biased expression of duplicated genes appear to be responsible for their retention in the genome [57]. We noted that many duplicated genes displayed divergent expression profiles and structural features (Fig. 3B and 4B). For example, *CsLAV5* showed constitutively high expression (RPKM >10 by RNA-Seq) in all tissue samples, whereas its duplicated paralogs, *CsLAV7,* was expressed at a relatively high level (RPKM = 4.64) in callus but significantly repressed in other tissues including flower, leaf and fruit (RPKM <1), showing biased expression. Cs*LAV7* only contains one B3 domain, whereas *CsLAV5* possesses other CW-type zinc fingers in addition to the B3 domain. In plants, CW-type zinc finger domain is considered as a histone recognition module that plays a pivotal role in transcriptional regulation during plant development [61], which might have conferred additional biological functions on *CsLAV5* in development of flower, leaf and fruit (Fig. 3D). As expected, almost all of the duplicated genes belong to the same classes, except for one pair, *viz. CsREM1* of class II (AtREM VII type) duplicated with *CsREM2* of class VI (AtREM VIII type) (Fig. 4). AtREM VII and AtREM VIII genes were supposed to be functional redundant for their partially overlapping expression patterns and high sequence similarity during *Arabidopsis* flower development [4]. It would be interesting to investigate whether *CsREM1* and *CsREM2* have acquired diversified functions, because they exhibited distinct expression patterns across different citrus tissues (Fig. 4B).

Possible roles of *CsB3* genes during SE of citrus

It has been noted previously that a few of *B3* genes promote SE of *Arabidopsis* and citrus [7, 39, 42]. The expression analysis of *CsB3* genes could provide clues for selecting the potential regulators of SE. Type I genes are preferentially expressed during initiation of citrus SE (E0-E4) (Fig.5). E0-E4 stages are critical stages of differentiation when yellow-green proembryos generated from the white-yellow EC. It has been demonstrated that *CsFUS3* gene (Cs2g14320), a type I *LAV* family member (*CsLAV1)*,, can enhance SE

competence of the citrus EC partially by regulating SE-related TFs and hormone pathways, especially ABA and GA pathways [42]. In *Arabidopsis,* an ortholog (i.e., *AtARF5)* of another type I gene *CsARF1* (Cs3g25860) is known to promote *de novo* shoot formation from *Arabidopsis* callus by pathways involving the downstream functions of *STM* and *CRF2 [*62]. Thus, the elevated expression of the type I *B3* genes during SE suggested their possible involvement in citrus SE.

After the formation of GE, the embryoids develop into CE each with two well-developed cotyledons, which denote the end of embryo morphogenesis. The CEs undergo phase transition to generate a plantlet through a germination-like process. The majority, but not all, of the type II and type III *CsB3* genes showed progressively decreased expression or maintained low levels at EC differentiation stages (E0-E4), but had a dramatically high expression in CE (Fig.5). The well-studied B3 member *AtABI3,* which is the putative ortholog of type II gene *CsLAV2* (Cs5g34660) in *Arabidopsis,* has been demonstrated to regulate abscisic acid-responsive gene*s* in phase transition from late embryo development to germination, implying that *CsLAV2* may also be functional in late embryogenesis [6, 63]. In addition, *AtLEC2* gene is the best recognized regulator of plant SE [7, 49]. However, in citrus SE, *CsLAV6* (Cs2g05780), the type II ortholog of*AtLEC2,* showed constantly low expression during early stages of SE (E0-GE), but accumulated specifically in CE (Fig.5). Likewise, *CsLAV6* was not present in suppression subtractive hybridization (SSH) libraries of citrus SE tissues in our previous studies [39], and the expression of *CsLAV6* is not increased in *CsFUS3* overexpressed EC lines of which the SE competence was enhanced [42]. Thus, we suppose *CsLAV6* may not promote SE initiation like its ortholog *AtLEC2,* but have function in late SE.

Another three type II B3 genes including *CsLAV3 (*Cs6g10020), *CsLAV5* (Cs2g06770) and *CsLAV7* (Cs1g06390) were clustered together and shared a high level of sequence similarity with the three repressors of embryonic pathways in *Arabidopsis* (i.e., *AtVAL1, AtVAL2* and *AtVAL3)* (Additional file 6) [9]. AtVAL 1/2/3 proteins were required for repression of the LAFL TFs (i.e., *AtLEC1, AtABI3, AtFUS3* and *AtLEC2)* during germination, which is necessary for the transition from seed to seedling development [9, 64, 65]. *CsLAV3/5/7* expressed specifically at late embryogenesis stages of citrus, i.e. CE and (or) E4 stages, suggesting their possible involvement also in repression of the SE pathway, for transition to vegetative development.

Another type II B3 gene *CsARF5* (Cs2g09440), the ortholog of *AtARF6,* showed constant expression levels in most the detected tissues, but the accumulation peak also emerged in CE (Fig. 5 and Additional file 6). As the cleavage target of miR167, *AtARF6* was reported to be required for SE formation, and *arf6* mutant is severely inhibited for SE production in *Arabidopsis* [31]. In citrus, *CsARF5* was also identified as the target of miR167 in leaf and fruit by degradome sequencing [66]. MiR167 showed low or undetectable expression levels in both EC and NEC (non-embryogenic callus), but accumulated in GE, and reached its peak also in subsequently formed CE [38], with similar expression pattern as *CsARF5 (*Fig. 5). The accumulation of *CsARF5* in CE may suggest its involvement in late embryogenesis, whereas the non-antagonistic expression patterns between miR167 and *CsARF5* may be resulted from the fine-tune regulation of miRNA and/or post-transcriptional regulations.

During SE process, the type III B3 gene *CsARF12* was down-regulated progressively during the differentiation process (E0-E4), but up-regulated in CE (Fig. 5). It was reported that *AtARF16,* an ortholog of *CsARF12* in *Arabidopsis,* regulates the expression of *AtABI3* in enhancement of seed dormancy and ABA mediated inhibition of seed germination [67]. The prominent expression level elevation of *CsABI3* (*CsLAV2)* and *CsARF12* at CE stage might be involved in inhibition of germination at late embryogenesis stages.

The expression level of two B3 genes (*CsARF7* and *CsREM9),,* which were classified into the type IV, showed the progressively elevated expression during citrus SE process (Fig. 5). *AtARF1,* which is homologous to *CsARF7* (Cs3g01570), binds to auxin response elements and confer auxin responsiveness in development [68]. In citrus, the endogenous IAA levels likewise gradually increased during SE and reached a peak in CE [42]. However, the level of endogenous IAA was relatively lower in EC with greater potential for SE, suggesting that auxin may not be a key factor for determination of SE competence [42]. In addition to auxin, a high ratio of ABA to GA was shown to remarkably contribute to citrus SE. The pooling of auxin was previously reported to modulate the levels of ABA and GA in regions of future organogenesis [8]. Thus, the association of *CsARF7* and the plant hormone auxin, ABA and GA in citrus SE remains to be elucidated. *CsREM9* gene belongs to the poorly characterized REM family. Expression and genetic analyses showed that one *REM* gene (*AtVDD)* is required for cell differentiation in the female gametophyte and highly expressed during early stages of seed formation [69]. SE shares morphological, cytological, and molecular similarities with zygotic embryogenesis (ZE) [42, 70]. The increased expression of *CsREM9* during the early stages of SE suggests that *CsREM9* may be functional in early embryogenesis (Fig. 5).

CsARF19 might be involved in citrus EC initiation

In citrus, EC can be induced *in vitro* from the undeveloped ovules/ aborted seeds of the polyembryonic genotypes, but not from the monoembryonic genotypes, which suggests that the regenerative EC might be derived from the pluripotent nucellar embryo initiation (NEI) cells localized in the apomictic nucellus tissues [71]. In this study, we identified a candidate B3 TF possibly involved in EC initiation. *CsARF19* was identified in mandarin and sweet orange, but not in pummelo or the relative genera of citrus. In addition, *CsARF19* was expressed at a relatively higher level in EC compared to other tissues in sweet orange (Fig. 3B). In our previous study, *ARF19* was moderately expressed (RPKM values ranged from 15.69 to 34.50) in the ovules of two mandarin cultivars (the monoembryonic ‘Nour’ Clementine and the polyembryonic ‘Huagan No.2’ Ponkan) during nucellar embryo initiation, but was expressed at slightly higher levels in the polyembryonic cultivar [72]. However, *ARF19* was not expressed (RPKM = 0) in ovules of the monoembryonic ‘Huanong red’ pummelo, but accumulated in ovules of the polyembryonic ‘Cocktail’ grapefruit, with RPKM values of 1.34 and 3.58 prior to and at stage during nucellar embryo initiation [72]. Based on the fact that sweet orange and grapefruit were derived from hybridizations among mandarin and pummelo [43, 72, 73], we suggest that *CsARF19* was originated in mandarin, and has introgression into the hybrids pool. SE initiation is believed to require an induction signal that causes somatic cells to

change identity [74]. Previous study showed that two *ARF* genes (*AtARF7* and *AtARF19)* directly or indirectly target four auxin-responsive LBD (*L*ATERAL ORGAN *B*OUNDARIES *D*OMAIN) genes to regulate callus formation in *Arabidopsis* regeneration [75]. Similar to that of *Arabidopsis,* callus-induction medium containing a high concentration of 2,4-dicholorophenoxy acetic acid (2,4-D) were also shown to promote citrus callus induction [76, 77]. However, the link between auxin signaling and citrus callus initiation has not yet been established. Our analysis implies that *CsARF19* is derived from mandarin, a basic species of citrus, and may be involved in callus initiation process from the nucellus tissues of polyembryonic citrus cultivars.

Conclusions

This study provided the genome-wide identification and characterization of the B3 superfamily in citrus and their possible involvement in SE was speculated based on the expression patterns. All B3 genes were phylogenetically classified into four families, as supported by the domain composition, gene structure and phylogenetic relationship. The expression analysis of these genes was investigated in distinct tissues, including EC and SE process. The expression profiles analysis showed that most of the *B3* genes were highly expressed in E4 and CE, indicating the involvement of B3 TFs at the late SE stages, which coincides with the important roles of B3s in late embryogenesis of *Arabidopsis.* One B3 gene, *CsARF19,* was indicated to be associated with nucellar-derived callus initiation of polyembryonic citrus cultivars. The understanding of B3 genes is expected to provide information for future research on functional elucidation of the B3 regulatory network in citrus SE.

Methods

Plant materials

Non-embryogenic callus (NEC) and embryogenic callus (EC) of ‘Valencia’ sweet orange were induced, cultured and preserved as depicted previously [78]. In brief, NEC was recently induced from epicotyl segments, whereas EC was induced from the aborted seeds and preserved in tissue culture for years. EC was transferred to glycerol medium to induce SE. We collected NEC, EC, E2/4 (EC induced for somatic embryos for 2 or 4 weeks), GE and CE. All these samples were immediately frozen in liquid nitrogen and stored at −80°C for further analysis.

Genome-wide identification of B3 superfamily genes

HMM (Hidden Markov Model) profile of B3 DNA binding domain (PF02362) was downloaded from Pfam database (http://pfam.xfam.org/family/PF02362), and subsequently exploited for the comprehensive identification of sweet orange and pummelo B3 superfamily genes from *Citrus sinensis* Annotation

Project (http://citrus.hzau.edu.cn/orange/download/index.php) using HMMER program (version 3.1b2) with a threshold of e-value < 0.01 [43, 44]. Using the same criterion, B3 family sequences were obtained from *Arabidopsis thaliana* database (ftp://ftp.ensemblgenomes.org/pub/plants/release–38/fasta/arabidopsis_thaliana/) and *Oryza sativa* database (ftp://ftp.ensemblgenomes.org/pub/plants/release–38/fasta/oryza_sativa/). Furthermore, the conserved domains of all putative candidates were confirmed using InterProScan software package (version 5.25–64.0). Finally, a self-blast of these protein sequences was performed to remove the redundancy whereas the alternative splice variants were not considered. A total of seven draft genomes of citrus species, including Mangshan mandarin, Clementine mandarin, atalantia, Ichang papeda, kumquat, trifoliate orange (unpublished data) and citron, were used to search the orthologs of *CsARF19* [44–47].

Analysis of chromosomal locations, synteny relationship, protein properties, gene structure and conserved motifs

The physical locations of citrus B3 genes were obtained from the database of sweet orange and pummelo genomes (http://citrus.hzau.edu.cn/orange/download/index.php). MapChart software (https://www.wur.nl/en/show/Mapchart.htm) was applied to visualize the distribution of the B3 genes on the citrus chromosomes. To detect the gene duplication events, the Multiple Collinearity Scan toolkit (MCScan X) was applied [79]. The Dual Synteny Plotter software (https://github.com/CJ-Chen/TBtools) was adopted to exhibit the synteny relationship of the orthologous B3 genes between citrus and *Arabidopsis* as well as that between citrus and rice. The Ks and Ka were calculated using KaKs_Calculator 2.0 [80].

The theoretical isoelectric points and molecular weights of the citrus B3 proteins were predicted by the compute pI/Mw tool in the ExPASY server (https://web.expasy.org/compute_pi/). Gene Structure Display Server (GSDS, http://gsds.cbi.pku.edu.cn/) program was exploited to illustrate exon/intron organization according to cDNA and genomic DNA sequences. InterProScan program (http://www.ebi.ac.uk/interpro/) was used to characterize the domains and motifs of the citrus B3 superfamily.

Multiple sequence alignments and phylogenetic analysis

Multiple sequence alignments of the B3 domain sequences of citrus B3 proteins were performed using ClustalW tool [81]. The alignment was visualized with ESpript 3.0 (http://espript.ibcp.fr/ESPript/cgi-bin/ESPript.cgi) and presented along with the corresponding secondary structure elements.

To investigate the phylogenetic relationship between citrus and *Arabidopsis,* the neighbor-joining (NJ) tree was constructed by MEGA7 software based on the full-length of the B3 protein sequences [82]. The same method was adopted to construct the NJ phylogenetic trees for the four families of citrus B3 superfamily.

Expression analysis of *CsB3s*

To investigate the expression patterns of all *B3* genes in different citrus tissues, the normalized RPKM values of citrus *B3* genes were extracted from the previously published RNA-Seq data of leaf, fruit, callus and flower of sweet orange [43] and that of leaf, fruit, ovule and seed of pummelo [44]. The results were visualized by the heat map with transformed $\log_2$ (RPKM+1) values using 'pheatmap' R package (https://cran.r-project.org/web/packages/pheatmap/index.html).

In order to gain an insight of *CsB3s* roles in citrus SE, the genes that specifically accumulated in callus with low expression level (1 < RPKM values < 10) or highly expressed (RPKM values higher than 10) in callus were selected to further analyze their expression profiles during SE using qRT-PCR. Total RNA was extracted using the Trizol reagent from EC and somatic embryos of 'Valencia' sweet orange [78], followed by RNA integrity examination using 1.0% agarose gel electrophoresis stained with ethidium bromide. The first strand cDNA was synthesized using RevertAid^{TM} First Strand cDNA Synthesis Kit (Fermentas, USA). qRT-PCR was performed as described previously [78]. qRT-PCR primer pairs were designed by Primer Premier 5.0 software (Additional file 8). The primers were further confirmed with a melting curve analysis after amplification of each tested genes. Each PCR pattern was verified by four replicate experiments. Two reference genes, i.e. *CiteIF–1A* and *CitUBL5,* which proved to be stably expressed during citrus SE, were used as internal controls to normalize the qRT-PCR data [78]. Mixtures without template were employed as the negative control. Data was processed using the Ct method ($2^{-\triangle\triangle CT}$) for relative quantification. Statistical analyses were performed using the IBM SPSS Statistics 19 software as described previously [42].

Abbreviations

SE: somatic embryogenesis; EC: embryogenic callus; TFs: transcription factors; *VP1: VIVIPAROUS1;* LAV: LEAFY COTYLEDON2-ABSCISIC ACID INSENSITIVE3-VAL; RAV: RELATED TO ABI3/VP1; ARF: AUXIN RESPONSE FACTOR; REM: REPRODUCTIVE MERISTEM; *LEC2: LEAFY COTYLEDON2; FUS3: FUSCA3; ABI3: ABSCISIC ACID INSENSITIVE3; VAL: VP1/ABI3-LIKE; HIS: HIGH-LEVEL EXPRESSION OF SUGAR-INDUCLBLE;* RPKM: reads per kilobase per million mapped reads; Ka/Ks: non-synonymous/synonymous substitution ratio; NEC: non-embryogenic callus; E0/2/4: EC induced for somatic embryos for 0, 2, 4 weeks; GE: globular embryos; CE: cotyledon embryos; WGD: whole-genome duplication; NJ*:* neighbor-joining.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National NSF of China [nos. 31701906, 31872051, 31872084], the Fundamental Research Funds for Central Universities (nos. 2662018PY007, 2662018PY013). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' contributions

ZL, XMW and WWG conceived and designed the research. ZL and XXG performed the experiments. ZL analyzed the data. ZL and XMW wrote the paper. All authors have read and approved the manuscript for publication.

Acknowledgements

We thank Prof. Qiang Xu, from the Key Laboratory of Horticultural Plant Biology (Ministry of Education), Huazhong Agricultural University, for providing the gene expression datasets of sweet orange and pummelo, as well as the amino acids sequences of B3 superfamily of trifoliate orange.

References

1. Swaminathan K, Peterson K, Jack T. The plant B3 superfamily. Trends Plant Sci. 2008;13:647-655. 2. Yamasaki K, Kigawa T, Seki M, Shinozaki K, Yokoyama S. DNA-binding domains of plant-specific transcription factors: structure, function, and evolution. Trends Plant Sci. 2013;18:267-276. 3. Suzuki M, Kao CY, McCarty DR. The conserved B3 domain of VIVIPAROUS1 has a cooperative DNA binding activity. Plant Cell. 1997;9:799-807. 4. Romanel EAC, Schrago CG, Counago RM, Russo CAM, Alves-Ferreira M. Evolution of the B3 DNA binding superfamily: new insights into REM family gene diversification. PloS ONE, 2009;4(6):e5791. 5. Peng FY, Weselake RJ. Genome-wide identification and analysis of the B3 superfamily of transcription factors in Brassicaceae and major crop plants. Theor Appl Genet, 2013;126:1305-1319. 6. Nambara E, Hayama R, Tsuchiya Y, Nishimura M, Kawaide H, Kamiya Y, et al. The role of ABI3 and FUS3 loci in Arabidopsis thaliana on phase transition from late embryo development to germination. Dev Biol. 2000;220:412-423. 7. Stone SL, Kwong LW, Yee KM, Pelletier J, Lepiniec L, Fischer RL, et al. LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. Proc Natl Acad Sci. 2001;98:11806-11811. 8. Gazzarrini S, Tsuchiya Y, Lumba S, Okamoto M, McCourt P. The transcription factor FUSCA3 controls developmental timing in Arabidopsis through the hormones gibberellin and abscisic acid. Dev Cell. 2004;7:373-385. 9. Suzuki M, Wang HHY, McCarty DR. Repression of the LEAFY COTYLEDON 1/B3 regulatory network in plant embryo development by VP1/ABSCISIC ACID INSENSITIVE 3-LIKE B3 genes. Plant Physiol. 2007;143:902-911. 10. Veerappan V, Wang J, Kang M, Lee J, Tang Y, Jha AK, et al. A novel HSI2 mutation in Arabidopsis affects the PHD-like domain and leads to derepression of seed-specific gene expression. Planta. 2012;236:1-17. 11. Wojcikowska B, Jaskota K, Gasiorek P, Meus M, Nowak K, Gaj MD. LEAFY COTYLEDON2 (LEC2) promotes embryogenic induction in somatic tissues of Arabidopsis, via YUCCA-mediated auxin biosynthesis. Planta. 2013;238:425-440. 12. Suzuki M, McCarty DR. Functional symmetry of the B3 network controlling seed development. Curr Opin Plant Biol. 2008;11:548-553. 13. Carbonero P, Iglesias-Fernandez R, Vicente-Carbajosa J. The AFL subfamily of B3 transcription factors: evolution and function in angiosperm seeds. J Exp Bot. 2016; 203:817-829. 14. Tsukagoshi H, Morikami A, Nakamura K. Two B3 domain transcriptional repressors prevent sugar-inducible expression of seed maturation genes in Arabidopsis seedlings. Proc Natl Acad Sci. 2007;104:2543-2547. 15. Monke G, Altschmied L, Tewes A, Reidt W, Mock HP, Baumlein H, et al. Seed-specific transcription factors ABI3 and FUS3: molecular interaction with DNA. Planta, 2004;219: 158-166. 16. Jia H, Suzuki M, McCarty RM. Regulation of the seed to seedling developmental phase transition by the LAFL and VAL transcription factor networks. Wires Dev Biol, 2013;3:135-145. 17. Matias-Hernandez L, Aguilar-Jaramillo AE, Marin-Gonzalez E, Suarez-Lopez P, Pelaz S. RAV genes: regulation of floral induction and beyond. Ann Bot. 2014;114:1459-1470. 18. Hu YX, Wang YH, Liu XF, Li JY. Arabidopsis RAV1 is down-regulated by brassinosteroid and may act as a negative regulator during plant development. Cell Res, 2004;14:8-15. 19. Woo HR, Kim JH, Kim J, Kim J, Lee U, Song IJ, et al. The RAV1 transcription factor positively regulates leaf senescence in Arabidopsis. J Exp Bot. 2010;61:3947-3957. 20. Li CW, Su RC, Cheng CP, Sanjaya, You SJ, Hsieh TH, et al. Tomato RAV transcription factor is a pivotal modulator involved in the AP2/EREBP-mediated defense pathway. Plant Physiol. 2011;156:213-227. 21. Fu M, Kang HK, Son SH, Kim SK, Nam KH. A subset of Arabidopsis RAV transcription factors modulates drought and salt stress responses independent of ABA. Plant Cell Physiol. 2014;55:1892-1904. 22. Lu Q, Zhao L, Li D, Hao D, Zhan Y, Li W. A GmRAV ortholog is involved in

photoperiod and sucrose control of flowering time in soybean. PLoS One. 2014;9:e89145. 23. Lee BH, Kwon SH, Lee SJ, Park SK, Song JT, Lee S, et al. The Arabidopsis thaliana NGATHA transcription factors negatively regulate cell proliferation of lateral organs. Plant Mol Biol. 2015;89:529-538. 24. Zhao SP, Xu ZS, Zheng WJ, Zhao W, Wang YX, Yu TF, et al. Genome-wide analysis of the RAV family in soybean and functional identification of GmRAV-03 involvement in salt and drought stresses and exogenous ABA treatment. Front Plant Sci. 2017;8:905. 25. Guilfoyle TJ, Hagen G. Getting a grasp on domain III/IV responsible for Auxin Response Factor-IAA protein interactions. Plant Sci. 2012;190:82-88. 26. Boer DR, Freire-Rios A, van den Berg WAM, Saaki T, Manfield IW, Kepinski S, et al. Structural basis for DNA binding specificity by the Auxin-dependent ARF transcription factors. Cell. 2014;156:577-589. 27. Kumar R, Tyagi AK, Sharma AK. Genome-wide analysis of auxin response factor (ARF) gene family from tomato and analysis of their role in flower and fruit development. Mol Genet Genom. 2011; 285:245-260. 28. Breitel DA, Chappell-Maor L, Meir S, Panizel I, Puig CP, Hao Y, et al. AXUIN RESPONSE FACTOR2 intersects hormonal signals in the regulation of tomato fruit repening. PLoS Genet. 2016;12(3):e1005903. 29. Ge J, Li B, Shen D, Xie J, Long J, Dong H. Tobacco TTG2 regulates vegetative growth and seed production via the predominant role of ARF8 in cooperation with ARF17 and ARF19. BMC Plant Biol. 2016;16:126. 30. Krogan NT, Marcos D, Weiner AI, Berleth T. The auxin response factor MONOPTEROS controls meristem function and organogenesis in both the shoot and root through the direct regulation of PIN genes. New Phytol. 2016;212:42-50. 31. Su YH, Liu YB, Zhou C, Li XM, Zhang XS. The microRNA167 controls somatic embryogenesis in Arabidopsis through regulating its target genes ARF6 and ARF8. Plant Cell Tiss Organ Cult. 2016;124:405-417. 32. Ren Z, Liu R, Gu W, Dong X. The Solanum lycopersicum auxin response factor SIARF2 participates in regulating lateral root formation and flower organ senescence. Plant Sci. 2017;256:103-111. 33. Wojcikowska B, Gaj MD. Expression profiling of AUXIN RESPONSE FACTOR genes during somatic embryogenesis induction in Arabidopsis. Plant Cell Rep. 2017;36:843-858. 34. Mantegazza O, Gregis V, Mendes MA, Morandini P, Alves-Ferreria M, Patreze CM, et al. Analysis of the arabidopsis REM gene family predicts functions during flower development. Ann Bot. 2014;114:1507-1515. 35. Franco-Zorrilla JM, Cubas P, Jarillo JA, Fernandez-Calvin B, Salinas J, Martinez-Zapater JM. AtREM1, a member of a new family of B3 domain-containing genes, is preferentially expressed in reproductive meristems. Plant Physiol. 2002;128:418-427. 36. Levy YY, Mesnage S, Mylne JS, Gendall AR, Dean C. Multiple roles of Arabidopsis VRN1 in vernalization and flowering time control. Science. 2002;297: 243-246. 37. Matias-Hernandez L, Battaglia R, Galbiati F, Rubes M, Eichenberger C, Grossniklaus U, et al. VERDAND1 is a direct target of the MADS domain ovule identity complex and affects embryo sac differentiation in Arabidopsis. Plant Cell. 2010;22:1702-1715. 38. Wu XM, Liu MY, Ge XX, Xu Q, Guo WW. Stage and tissue-specific modulation of ten conserved miRNAs and their targets during somatic embryogenesis of Valencia sweet organe. Planta. 2011;233:495-505. 39. Ge XX, Chai LJ, Liu Z, Wu XM, Deng XX, Guo WW. Transcriptional profiling of genes involved in embryogenic, non-embryogenic calluses and somatic embryogenesis of Valencia sweet orange by SSH-based microarray. Planta. 2012;236: 1107-1124. 40. Pan ZY, Guan R, Zhu SP, Deng XX. Proteomic analysis of somatic embryogenesis in Valencia sweet orange (Citrus sinensis Osbeck). Plant Cell Rep. 2009;28:281-289. 41. Wu XM, Kou SJ, Liu YL, Fang YN, Xu Q, Guo WW. Genome-wide analysis of small RNAs in non-embryogenic and embryogenic tissues of citrus: microRNA and siRNA mediated transcript cleavage

involved in somatic embryogenesis. Plant Biotechnol J. 2015;13:383-394. 42. Liu Z, Ge XX, Qiu WM, Long JM, Jia HH, Yang W, et al. Overexpression of the CsFUS3 gene encoding a B3 transcription factor promotes soamtic embryogenesis in Citrus. Plant Sci. 2018;227:121-131. 43. Xu Q, Chen LL, Ruan XA, Chen DJ, Zhu AD, Chen CL, et al. The draft genome of sweet orange (Citrus sinensis), Nat Genet. 2013;45:59-66. 44. Wang X, Xu Y, Zhang S, Cao L, Huang Y, Cheng J, et al. Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. Nat Genet. 2017;49:765-772. 45. Wang L, He F, Huang Y, He J, Yang S, Zeng J, et al. Genome of wild mandarin and domestication history of mandarin. Mol Plant. 2018;11:1024-1037. 46. Zhu C, Zheng X, Huang Y, Ye J, Chen P, Zhang C, et al. Genome sequencing and CRISPR/Cas 9 gene editing of an early flowering Mini-Citrus (Fortunella hindsii). Plant Biotechnol J. 2019; https://doi.org/10.1111/pbi.13132 47. Wu GA, Prochnik S, Jenkins J, Salse J, Hellsten U, Murat F, et al. Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. Nat Biotechnol. 2014;32:656-662. 48. Wu GA, Terol J, Ibanez V, López-García A, Pérez-Román E, Borredá C, et al. Genomics of the origin and evolution of Citrus. Nature. 2018;554:311-316. 49. Braybrook SA, Stone SL, Park S, Bui AQ, Le BH, Fischer RL, et al. Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo maturation and somatic embryogenesis. Proc Natl Acad Sci. 2006;103:3468-3473. 50. Gutierrez L, Van Wuytswinkel O, Castelain M, Bellini C. Combined networks regulating seed maturation. Trends Plant Sci. 2007;12:294-300. 51. Wang F, Perry SE. Identification of direct targets of FUSCA3, a key regulator of Arabidopsis seed development. Plant Physiol. 2013;161:1251-1264. 52. The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature. 2000;408:796-815. 53. Yu J, Hu S, Wang J, Wong GKS, Li S, Liu B, et al. A draft sequence of the rice genome (Oryza sativa L. ssp. indica). Science. 2002;296:79-92. 54. Kumar A, Sreedharan SP, Giridhar P. Combining bioinformatics and conventional PCR optimization strategy for one-time design of high-specificity primers for WRKY gene family using unigene database. Mol Biol Rep. 2019;46:3461-3475. 55. Liu C, Xie T, Chen C, Luan A, Long J, Li C, et al. Genome-wide organization and expression profiling of the R2R3-MYB transcription factor family in pineapple (Ananas comosus). BMC Genomics. 2017;18:503. 56. Dong C, Hu H, Jue D, Zhao Q, Chen H, Xie J, et al. The banana E2 gene family: Genomic identification, characterization, expression profiling analysis. Plant Sci. 2016;245:11-24. 57. Zhu Y, Wu N, Song W, Yin G, Qin Y, Yan Y, et al. Soybean (Glycine max) expansin gene superfamily origins: segmental and tandem duplication events followed by divergent selection among subfamilies. BMC Plant Biol. 2014;14:93. 58. Ober D. Gene duplications and the time thereafter-examples from plant secondary metabolism. Plant Biology. 2010;12:570-577. 59. Panchy N, Lehti-Shiu M, Shiu SH. Evolution of gene duplication in plants. Plant Physiol. 2016;171:2294-2316. 60. Zhang J. Evolution by gene duplication: an update. Trends Ecol Evol. 2003;18:292-298. 61. Zhang Z, Zhang F, Cheng ZJ, Liu LL, Lin QB, Wu FQ, et al. Functional characterization of rice CW-domain containing zinc finger proteins involved in histone recognition. Plant Sci. 2017;263:168-176. 62. Ckurshumova W, Smirnova T, Marcos D, Zayed Y, Berleth T. Irrepressible MONOPTEROS/ARF5 promotes de novo shoot formation. New Phytol. 2014;204:556-566. 63. Parcy F, Valon C, Raynal M, Gaubier-Comella P, Delseny M, Giraudat J. Regulation of gene expression programs during Arabidopsis seed development: roles of the ABI3 locus and of endogenous abscisic acid. Plant Cell. 1994;6:1567-1582. 64. Schneider A, Aghamirzaie D, Elmarakeby H, Poudel AN, Koo AJ, Heath LS, et al. Potential targets of

VIVIPAROUS1/ABI3-LIKE1 (VAL1) repression in developing Arabidopsis thaliana embryos. Plant J. 2016;85:305-319. 65. Jia H, McCarty DR, Suzuki M. Distinct roles of LAFL network genes in promoting the embryonic seedling fate in the absence of VAL repression. Plant Physiol. 2013;163:1293-1305. 66. Liu Y, Wang L, Chen D, Wu X, Huang D, Chen L, et al. Genome-wide comparison of microRNAs and their targeted transcripts among leaf, flower and fruit of sweet orange. BMC Genomics. 2014;15:695. 67. Liu X, Zhang H, Zhao Y, Feng Z, Li Q, Yang HQ, et al. Auxin controls seed dormancy through stimulation of abscisic acid signaling by inducing ARF-mediated ABI3 activation in Arabidopsis. Proc Natl Acad Sci. 2013;110:15485-15490. 68. Ulmasov T, Gretchen H, Guilfoyle TJ. ARF1, a transcription factor that binds to auxin response elements. Science. 1997;276:1865-1868. 69. Matias-Hernandez L, Battaglia R, Galbiati F, Rubes M, Eichenberger C, Grossniklaus U, et al. VERDAND1 is a direct target of the MADS domain ovule identity complex and affects embryo sac differentiation in Arabidopsis. Plant Cell. 2010;22:1702-1715. 70. Rupps A, Raschke J, Rümmler M, Linke B, Zoglauer K, Identification of putative homologs of Larix decidua to BABYBOOM (BBM), LEAFY COTYLEDON1 (LEC1), WUSCHEL-related HOMEOBOX2 (WOX2) and SOMATIC EMBRYOGENESIS RECEPTOR-like KINASE (SERK) during somatic embryogenesis, Planta. 2016;243 :473-488. 71. Miah MN, Islam S, Hadiuzzaman S. Regeneration of plantlets through somatic embryogenesis from nucellus tissue of Citrus macroptera Mont. var. anammensis ('Sat Kara'). Plant Tissue Cult. 2002;12:167-172. 72. Long JM, Liu Z, Wu XM, Fang YN, Jia HH, Xie ZZ, et al. Genome-scale mRNA and small RNA transcriptomic insights into initiation of citrus apomixis. J Exp Bot. 2016;67:5743-5756. 73. Wu GA, Terol J, Ibanez V, López-García A, Pérez-Román E, Borredá C, et al. Genomics of the origin and evolution of Citrus. Nature. 2018;554:311-316. 74. Uddenberg D, Abrahamsson M, von Arnold S, Overexpression of PaHAP3A stimulates differentiation of ectopic embryos from maturing somatic embryos of Norway spruce, Tree Genet. Genome. 2016;12:18. 75. Fan M, Xu C, Xu K, Hu Y. LATERAL ORGAN BOUNDARIES DOMAIN transcription factors direct callus formation in Arabidopsis regeneration. Cell Res. 2012;22:1169-1180. 76. Savita, Singh B, Virk GS, Nagpal AK. An efficient plant regeneration protocol from callus cultures of Citrus jambhiri Lush. Physiol Mol Biol Pla. 2011;17:161-169. 77. Hussain M, Raja NI, Iqbal M, Iftikhar A, Sadaf HM, Sabir S, et al. Plantlets regeneration via somatic embryogenesis from the nucellus tissues of Kinnow mandarin (Citrus reticulate L.). American J Plant Sci. 2016;7:798-805. 78. Liu Z, Ge XX, Wu XM, Kou SJ, Chai LJ, Guo WW. Selection and validation of suitable reference genes for mRNA qRT-PCR analysis using somatic embryogenic cultures, floral and vegetative tissues in citrus. Plant Cell Tiss Org. 2013;113:469-481. 79. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, et al. MCScan X: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40:e49. 80. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series Methods and sliding window strategies. Genomics Proteomics Bioinformatics. 2010;8:77-80. 81. Thompson JD, Glbson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. Curr Protoc Bioinformatics. 2002;Chapter 2:Unit 2 3. 82. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33:1870-1874.

Figures

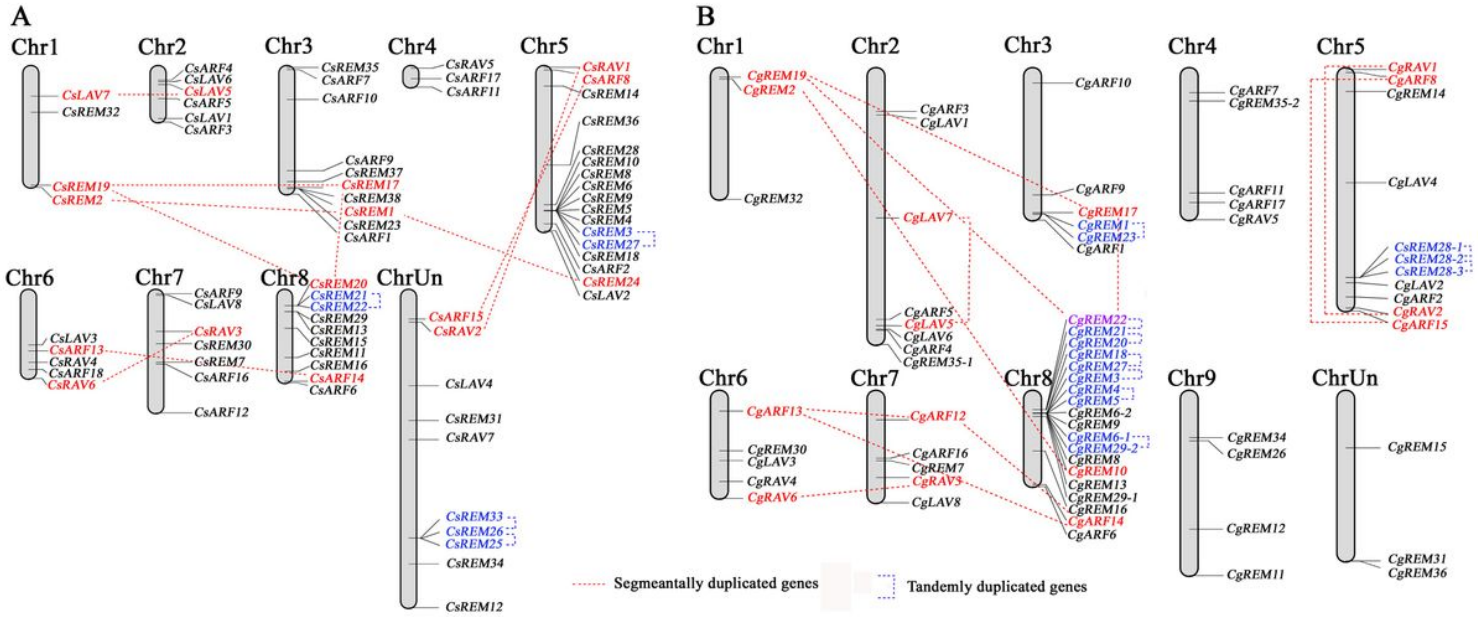

Figure 1

Chromosomal locations and regional duplication of citrus B3 genes. The chromosomal position of each B3 genes was mapped according to the sweet orange (A) and pummelo (B) genomes. The chromosome number is indicated at the top of each chromosome. Segmentally duplicated gene pairs were linked by red dotted lines, whereas tandemly duplicated gene pairs were linked by blue dotted lines.

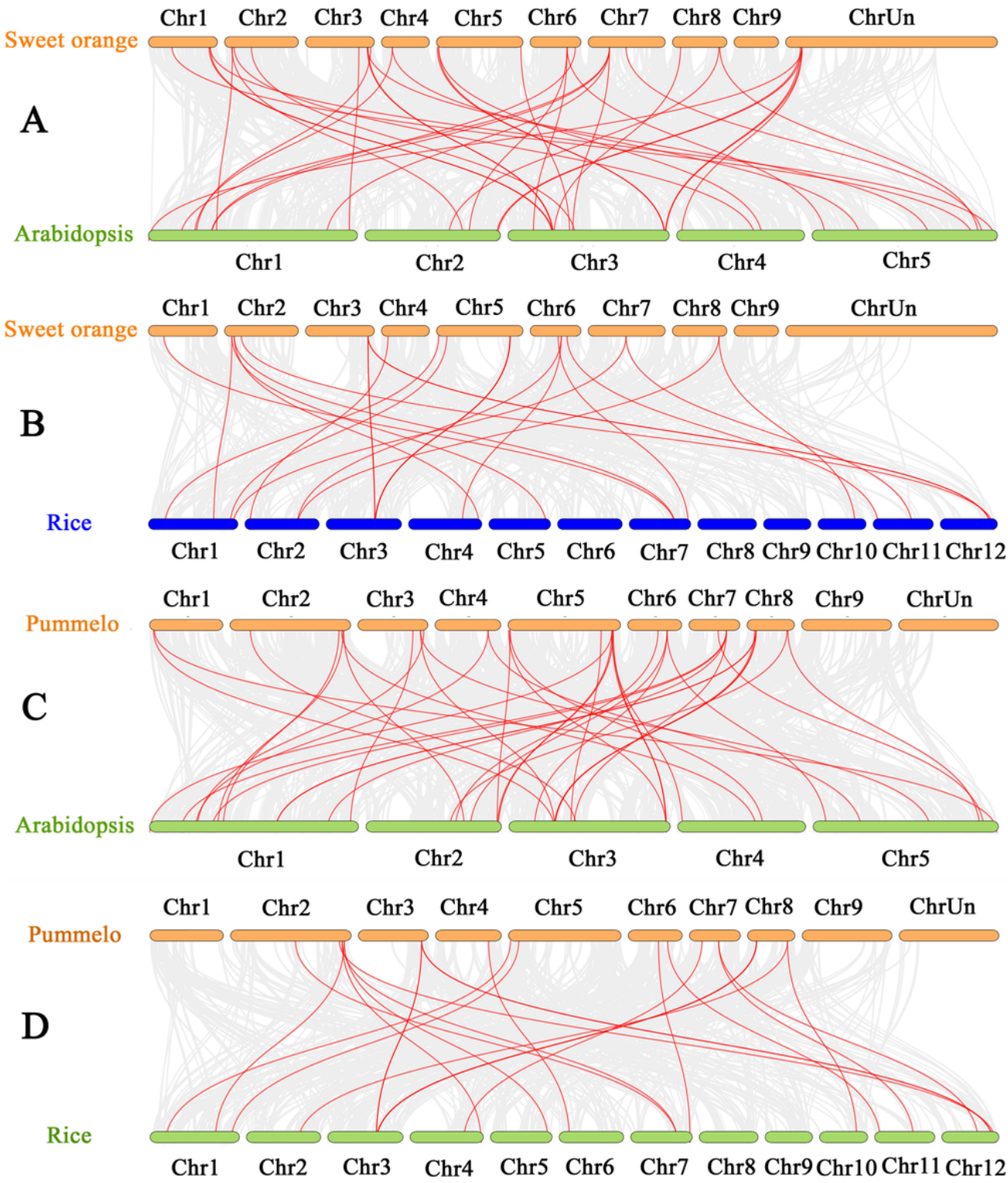

Figure 2

Gene duplication and synteny analysis of the B3 genes between sweet orange/pummelo and Arabidopsis/rice. Gray lines in the background indicated the collinear blocks within sweet orange/Arabidopsis genomes (A), sweet orange/rice genomes (B), pummelo/Arabidopsis genomes (C) and pummelo/rice genomes (D), respectively. The red lines highlight the syntenic B3 gene pairs.

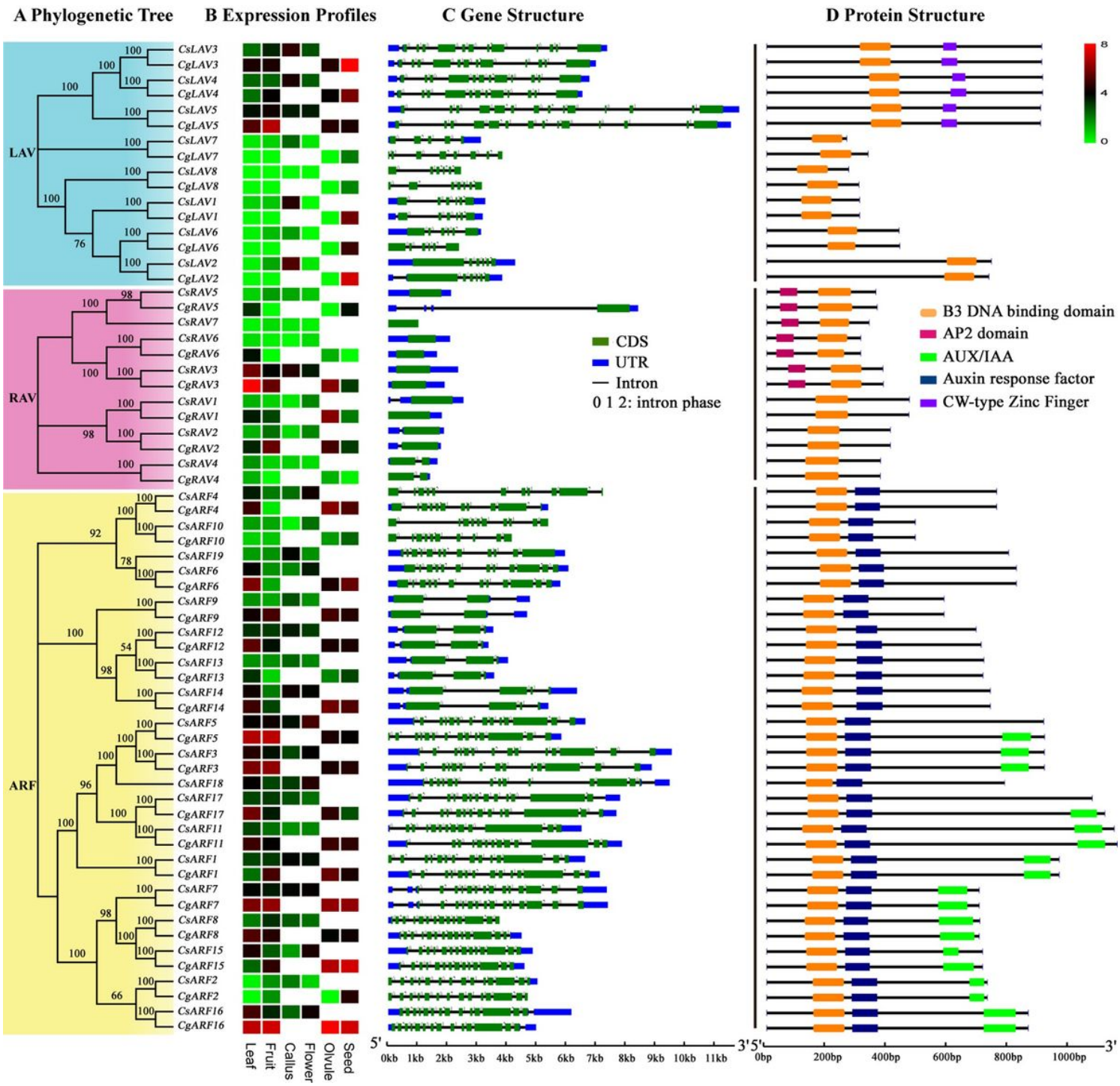

Figure 3

Phylogenetic relationships, expression profiles, gene structure and protein structure of citrus B3 genes from LAV, RAV and ARF families. (A) The neighbor-joining trees were constructed for B3 genes from each family, respectively. (B) Heatmap showed the expression profiles of B3 genes in different tissues, including four tissues (leaf, fruit, callus and flower) of sweet orange and four tissues (leaf, fruit, ovule and seed) of pummelo. (C) The gene structures were presented by green exon(s), blue UTR regions, whereas the solid lines between the colored boxes indicate introns. The number indicated the phases of the corresponding introns. (D) To present the protein structures, the B3 DNA binding domains were

highlighted by orange boxes, whereas the AP2 domain, AUX/IAA, Auxin response factor and CW-type zine finger domains were represented by red, green, blue and purple boxes, respectively.

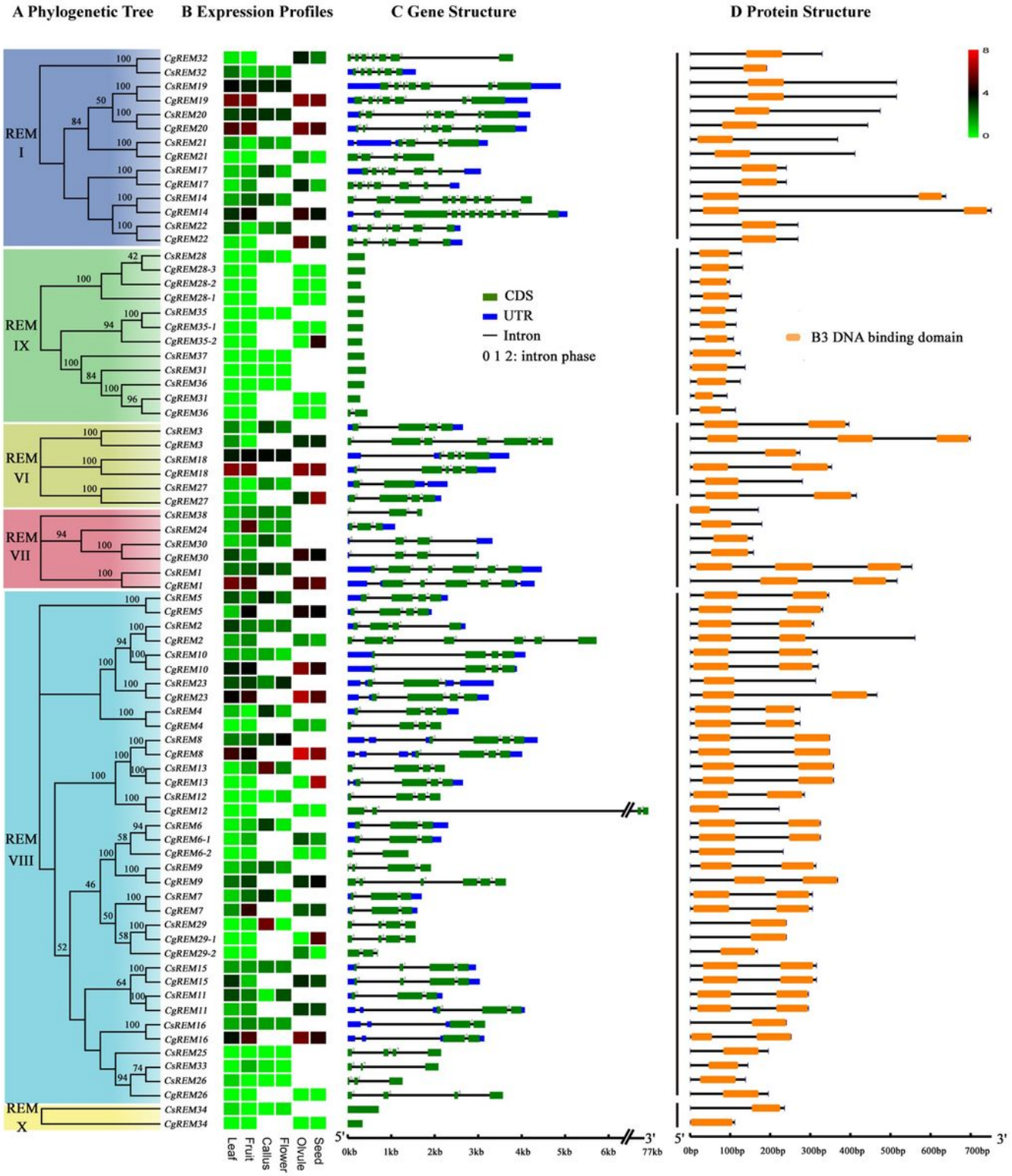

Figure 4

Phylogenetic relationships, expression profiles, gene structure and protein structure of citrus B3 genes from REM family. (A) The neighbor-joining trees were constructed for B3 genes from REM family. (B) Heatmap showed the expression of B3 genes in different tissues, including four tissues (leaf, fruit, callus

and flower) of sweet orange and four tissues (leaf, fruit, ovule and seed) of pummelo. (C) The gene structures were presented by green exon(s), blue UTR regions, whereas the solid lines between the colored boxes indicate introns. The number indicated the phases of the corresponding introns. (D) To present the protein structures, the B3 DNA binding domains were highlighted by orange boxes.

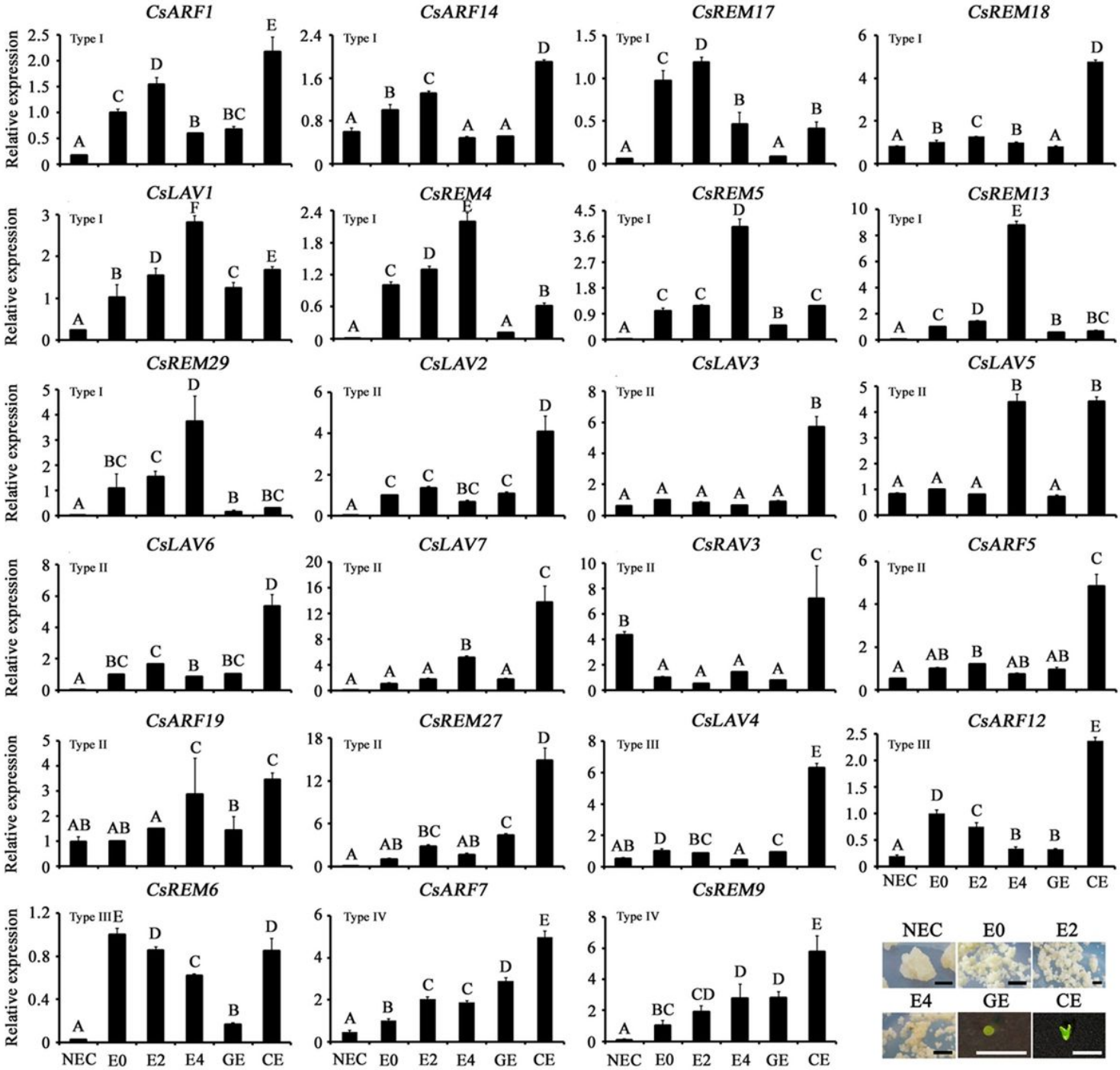

Figure 5

Expression profiles of 23 selected CsB3 genes during somatic embryogenesis of 'Valencia' orange. The capital letters above the bars indicted significant difference (P<0.01). Non-embryogenic callus (NEC),

embryogenic callus (EC) induced for somatic embryos for 0, 2, 4 weeks (E0, E2, E4), globular embryos (GE), cotyledon embryos (CE). Scale bar =5 mm.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplement1.xls
- supplement2.doc
- supplement3.xls
- supplement4.xls
- supplement5.xls
- supplement6.tif
- supplement7.tif
- supplement8.tif