Article

Variation burst during dedifferentiation and increased CHH-type DNA methylation after 30 years of in vitro culture of sweet orange

Xia Wang, Lili Ke, Shuting Wang, Jialing Fu, Jidi Xu, Yujin Hao, Chunying Kang, Wenwu Guo, Xiuxin Deng and Qiang Xu*

Key Laboratory of Horticultural Plant Biology (Ministry of Education), Huazhong Agricultural University, No. 1, Shizishan Street, Wuhan 430070, China
*Corresponding author: E-mail: xuqiang@mail.hzau.edu.cn

Abstract

Somaclonal variation arising from tissue culture may provide a valuable resource for the selection of new germplasm, but may not preserve true-to-type characteristics, which is a major concern for germplasm conservation or genome editing. The genomic changes associated with dedifferentiation and somaclonal variation during long-term in vitro culture are largely unknown. Sweet orange was one of the earliest plant species to be cultured in vitro and induced via somatic embryogenesis. We compared four sweet orange callus lines after 30 years of constant tissue culture with newly induced calli by comprehensively determining the single-nucleotide polymorphisms, copy number variations, transposable element insertions, methlyomic and transcriptomic changes. We identified a burst of variation during early dedifferentiation, including a retrotransposon outbreak, followed by a variation purge during long-term in vitro culture. Notably, CHH methylation showed a dynamic pattern, initially disappearing during dedifferentiation and then more than recovering after 30 years of in vitro culture. We also analyzed the effects of somaclonal variation on transcriptional reprogramming, and indicated subgenome dominance was evident in the tetraploid callus. We identified a retrotransposon insertion and DNA modification alternations in the potential regeneration-related gene CLAVATA3/EMBRYO SURROUNDING REGION-RELATED 16. This study provides the foundation to harness in vitro variation and offers a deeper understanding of the variation introduced by tissue culture during germplasm conservation, somatic embryogenesis, gene editing, and breeding programs.

Introduction

Plant cell dedifferentiation is the transformation of cells from a differentiated state to a less differentiated state, and has long been a research hotspot as a means of understanding the ability of plant cells to turn into any differentiated cell and the striking plasticity of plant development [1, 2]. Calli are predominantly the result of cell dedifferentiation, indicating that cell dedifferentiation may play important roles in callus formation.

Somaclonal variation refers to all genetic or epigenetic alterations that may occur in the culture of calli and regenerated plants by in vitro tissue culture [3]. As it is often the only available method of propagation for many plants and is an integral part of many breeding programs, in vitro tissue culture has been widely applied in germplasm preservation, artificial seed production, genetic transformation, and metabolite engineering. On the one hand, somaclonal variation may constitute a largely untapped pool of induced polymorphisms for new genotypes (and presumably new phenotypes) and may be of great value in the selection of new varieties. Through in vitro selection, variants with desirable agronomic traits, such as stress tolerance or disease resistance, may be quickly and efficiently selected [4]. On the other hand, the high frequency of somaclonal variations in regenerated plants may also have negative effects on the true-to-type characteristics, purity, and quality of germplasm produced by in vitro culture, which has become one of the main limitations for the practical applications of micropropagation and genetic transformation, as well as for studies involving genome editing [4, 5]. In vitro calli from diverse plant species are routinely used as starting materials for gene editing and functional studies, such as clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated nuclease editing approaches. Several studies have reported that sequence alterations during CRISPR-mediated genome editing may be due in part to somaclonal variation caused by tissue culture even more often than to off-target editing [5]. Characterizing the scale of the genetic variation that arises during tissue culture will therefore provide foundational knowledge to help tease apart somaclonal variation and off-target mutations caused by gene editing [5, 6]. Furthermore, somaclonal variations...
constitute a pool of polymorphisms that is distinct from natural variations and specific to the ever-changing growth environments experienced during tissue culture, thus making in vitro selection a possible road for breeding traits that might not be accessible to natural genetic variation.

Citrus is an important horticultural crop of high economic value characterized by nucellar polyembryony that prevents the production of hybrid offspring. Citrus is thus propagated mainly via asexual means, such as grafting and cuttings [7, 8]. Sweet orange (Citrus sinensis), one of the most popular citrus cultivars worldwide, was produced from multiple crosses between mandarin (Citrus reticulata) and pummelo (Citrus grandis), resulting in a genome with high heterozygosity [9, 10]. Citrus genetic improvement and breeding programs rely heavily on somatic mutations. Therefore, studying somaclonal variation patterns in citrus is of great potential value for producing useful cultivars. Although the first use of callus induction for clonal propagation was reported in 1974 [11], this experimental system has not been widely exploited to uncover dynamic somaclonal variation during cell dedifferentiation and later developmental reprogramming, but may offer much promise in this regard [12].

Somaclonal variations manifest as frequent phenotypic changes, cytological abnormalities, sequence alterations, and gene activation or silencing [13]. At the DNA level, somaclonal variation may consist of single-nucleotide polymorphisms (SNPs), copy number variations (CNVs), and/or structural variations (SVs) that may affect chromosome structure and gene function [13, 14]. Deep sequencing techniques have recently allowed the comprehensive detection and analysis of genetic variation, and the development of longer, single-molecule sequencing methods is now paving the way to identifying large-scale structural variation that was out of reach of techniques based on shorter reads [14–16]. Although allelic imbalances might be caused by somaclonal variation, few studies have focused on detecting these during plant tissue culture. Allele-specific expression (ASE) is the preferential expression of a parental allele in a hybrid or polyploid [17]. The RNA-sequencing (RNA-seq) technologies are widely applied to reveal the transcriptome and quantify relative expression levels, therefore, RNA-seq based ASE analysis has been successfully employed to identify SNPs and genes displaying ASE [18]. The features of ASE have been demonstrated in the natural allopolyploid plants, such as monkeyflower (Mimulus sp.) [19] and cotton (Gossypium hirsutum) [20]. ASE analysis is also possible in the model diploid plants, such as rice (Oryza sativa) [21, 22], maize (Zea mays) [23, 24], and Arabidopsis (Arabidopsis thaliana) [25, 26], as a means to elucidate genome dominance between alleles and how it may shape phenotypic diversity, stress responses, heterosis, and hybrid incompatibility.

Previously quiescent transposable elements (TEs) mobilize frequently during plant in vitro culture. In fact, the activation of Ac element in the regenerated maize plant was derived from tissue culture [27]. The rice retrotransposon Tos17 may integrate into and excise itself from the genome during tissue culture, with its copy number tending to increase with culture time [28, 29]. Similarly, the MITE mPing is mobilized in regenerated rice seedlings induced from anthers [30], and the LINE-type retrotransposon Karma is activated by tissue culture and remains active in the resulting regenerated plants [31]. Transposon activation constitutes a rich source of genetic variation and might contribute to genome evolution, transcriptional regulation, and stress responses [30].

Epigenetic variation has also been reported during in vitro tissue culture, likely as a consequence of the stress imposed by the growth conditions and the intrinsic developmental program of callus cells [32, 33]. Analyses of the methylome of rice revealed a stable loss of DNA methylation after in vitro culture, and hotspots for the methylation alterations [34]. One example of somaclonal variation comes from oil palm, in which the mantled floral abnormality arises with some frequency after tissue culture and greatly reduces fruit yield and thus oil production. Transcription from the retrotransposon Karma appeared to be activated in cells afflicted with the mantled abnormality, due to DNA hypomethylation [35].

To explore the genomic changes associated with dedifferentiation and somaclonal variation during long-term in vitro culture, we performed a comprehensive investigation of the genomes, methylomes, and transcriptomes from newly induced calli and sweet orange calli maintained in vitro culture for 30 years. These multi-omics analyses offer a detailed view of genomic and epigenomic dynamics during dedifferentiation and in vitro culture. We also compared the variation observed for polyploids in response to in vitro culture, which revealed different patterns of subgenome dominance between in vitro and in vivo polyploids, and elucidated the effects of somaclonal variation on somatic embryogenesis during in vitro culture.

Results

Multi-omics sequencing of sweet orange calli newly induced and cultured in vitro for 30 years

To investigate the dedifferentiation and somaclonal variations stemming from long-term tissue culture, we investigated sweet orange callus lines maintained in vitro for 30 years and newly induced calli (referred to hereafter as old and nascent calli, respectively; Fig. 1a). Old calli were induced from immature unfertilized ovules of Valencia sweet orange (C. sinensis Valencia) in 1989 and then continuously subcultured for almost 30 years. Nascent calli were induced from the immature unfertilized ovules of the same genotype in 2019. The explants (immature unfertilized ovules) used in this study were also from the same genotype of sweet orange plant. We determined the ploidy levels of
independent long-term cultured calli by flow cytometry of isolated nuclei, which revealed that these calli were variously diploid, tetraploid, and hexaploid (Supplementary Fig. S1). We selected two diploid callus lines, one tetraploid callus, and one hexaploid callus from which to start independent cultures for comparative analyses. We then sequenced the genomes, epigenomes, and transcriptomes of all callus and explant materials described above and performed a systematic analysis of the nascent and old calli (Supplementary Tables S1–4).

Dramatic variation during nascent dedifferentiation relative to long-term in vitro culture

An average of 46.48-fold genome coverage were sequenced for explants, nascent calli, and old calli (Supplementary Table S1). We identified all SNPs for all callus samples and evaluated their mutation rates. A total of 1,460,391 SNPs were detected between explant and callus lines cultured in vitro. We estimated the in vitro somatic mutation rate to be $1.7 \times 10^{-7}$ per base per year based on the number of private SNPs in old calli, a rate approximately 100-fold higher than in vivo somatic mutation rates ($1-2 \times 10^{-9}$ per base per year) [9]. A principal component analysis (PCA) of all SNP data separated samples into three major groups (Fig. 1b). Calli maintained in long-term culture grouped together irrespective of their ploidy levels, reflecting their similar genetic backgrounds (0.64% SNP rate, on average; Supplementary Fig. S2a). Unexpectedly, nascent calli grouped separately from the explant and old calli (5.51% and 3.96%, respectively, SNP rate, on average; Supplementary Fig. S2a). The genomic variations between independent nascent calli were with high reproducibility (Supplementary Fig. S3), and the reproducibility rate of SNP was estimated to be 87.34%. Indeed, the SNP rate caused by nascent dedifferentiation was $1.45 \times 10^{-5}$ per base, when comparing nascent calli and the diploid explant, which is 2.84 times higher than the corresponding SNP rate in old calli ($5.1 \times 10^{-6}$ per base). Similarly, PacBio long reads allowed the identification of 2804 SVs in old calli and 19,220 SVs in nascent calli, or a 6.85-fold difference (Supplementary Fig. S2b). Both SNP and SV data illustrated the dramatic variation associated with recent dedifferentiation.

SNP density was not uniform across the genome (Fig. 1c). We observed hotspots for genomic variation resulting from recent dedifferentiation, with SNP rate
greater than 1%, included 24.9–29.3 Mb and 30.9 Mb to the end of chromosome 2, 29.1 Mb to the end of chromosome 3, 11.4–18.2 Mb on chromosome 4, 5.7–24.5 Mb on chromosome 5, and 14.1–23 Mb on chromosome 9 (Fig. 1c). These hotspot regions harbored 4182 genes and were significantly enriched in genes related to cytokinin metabolism ($P$ = 0.036). Variation hotspots associated with long-term in vitro culture were distinct and restricted to two genomic regions: 1–4.7 Mb on chromosome 7 and 1–2.6 Mb on chromosome 9 (Fig. 1c). Consistent with the variation burst specifically during dedifferentiation, there existed much more variation hotspots during dedifferentiation than that after long-term in vitro culture.

**Transposon outbreak during dedifferentiation**

We generated PacBio long reads with an average 17.67-fold genome coverage for each sample and an average read length of 7.2 kb for the precise detection of insertion/deletion variations (Supplementary Table S2). We also cataloged all instances of TE insertions by aligning the inserted sequences to the TE database and checking manually for the presence of TEs. We validated a number of randomly selected predicted TE insertions by polymerase chain reaction (PCR) amplification and sequencing; in all cases, we confirmed the identity of the inserted TE and its integration site, highlighting the robustness of our TE insertion detection pipeline, which had an accuracy close to 100% (Supplementary Fig. S4).

When compared with explants, a total of 723 TE insertions were detected in nascent calli, among which 631 TE insertions were specific in nascent calli, and 92 TE insertions were shared by nascent and old calli (Supplementary Table S5), which indicated the occurrence of a burst of transposon activation during dedifferentiation. Of these TE insertions in the nascent calli, 51.59% (or 373 TEs) had integrated into the gene body and/or promoter region (Supplementary Table S6). The vast majority of retrotransposon insertions in the nascent calli belonged to the long terminal repeat (LTR)/Gypsy (47%) or LTR/Copia (41%) families (Fig. 2a). We also discovered 188 instances of repeated transposition events of the same TE, indicating a high transposition frequency during dedifferentiation (Supplementary Table S5). For example, we saw evidence for the transposition of an LTR/Gypsy TE originally mapping to chromosome 4 (from 7 022 365 bp to 7 022 588 bp) into 23 separate loci across the genome (Fig. 2b).

We then turned to our methylation dataset to define features of retrotransposon insertions specifically detected in nascent calli. We established that DNA methylation levels were particularly low near the sites of retrotransposon integration relative to flanking regions, and that this pattern was much more pronounced in nascent calli relative to explants and old calli (Fig. 2c and Supplementary Fig. S5). These results implied that TEs preferentially inserted into hypomethylated chromatin, as changes in DNA methylation might modify chromatin accessibility and consequently the ability of a TE to insert.

**Variation purge after long-term in vitro culture**

We next investigated the pattern of SNPs after 30 years spent in in vitro culture. Indeed, our results above suggested that old calli had a much lower mutation rate than nascent calli, even though they must have undergone the same variation burst during their initial dedifferentiation. We determined the complements of SNPs in nascent and old calli as a prerequisite to identifying SNPs shared by both sets of calli (born SNPs), those lost during the long-term culture of calli (purged SNPs), and those specific to long-term calli (born SNPs) (Fig. 3a). We assigned 47 694 SNPs out of 1 460 391 polymorphic sites (or 3% of the total) to the category of retained SNPs, and another 82 906 SNPs (or 6% of total) to the category of born SNPs found only in old calli (Fig. 3a). Purged SNPs accounted for the remaining 1 329 791 SNPs (or 91%, Fig. 3a).

To estimate the mutation burden imposed by in vitro culture, we looked at the molecular consequences of all SNPs and identified all possible associated deleterious mutations based on genomic evolution and amino acid conservation modelling. Relative to explants (with an average of 4013 deleterious mutations), nascent calli harbored approximately 1000 more such mutations (4938 on average across nascent calli), and the long-term cultured calli seemed to have experienced a mild purge (with 3805 deleterious mutations on average) (Fig. 3b). These results indicated that the dedifferentiation process massively increased the genetic variation load, as evidenced by the newly acquired SNPs and SVs in nascent calli, but this was followed by a massive purge of most polymorphisms during long-term in vitro culture to lower the mutation burden.

**Loss and gain of CHH-type DNA methylation during dedifferentiation and long-term in vitro culture**

We next conducted genome-wide DNA methylation analyses to investigate epigenetic variation during dedifferentiation and long-term in vitro culture by whole genome bisulfite sequencing (WGBS) with an average 41.79-fold genome coverage for each sample (Supplementary Table S3). The average measured DNA methylation rates were 71.35%, 54.14%, and 8.52% for CG-, CHG-, and CHH-type in control explants, respectively, and 53.21%, 28.51%, and 4.60% for CG-, CHG-, and CHH-type in nascent calli, respectively (where H is A, C, or T). We concluded that CG, CHG, and CHH methylation simultaneously decreased during dedifferentiation. In addition, lower methylation was prevalent across the entire genome upon dedifferentiation and was not restricted to small genomic islands (Fig. 4a). In particular, CHH methylation levels were markedly lower in nascent calli, with a 46% decrease upon recent dedifferentiation (Fig. 4a).

After long-term in vitro culture, overall CG and CHG DNA methylation levels were comparable to those from...
nascent calli (Supplementary Table S7). Interestingly, in contrast to the loss of CHH-type DNA methylation during dedifferentiation, old diploid calli exhibited average global CHH methylation levels higher than those seen in nascent calli and control explants, indicating that old diploid calli gained net CHH methylation relative to
Figure 3. Variation transfer and mutational load after long-term in vitro culture. (a) Pie chart of the percentages of SNPs purged, born, and retained after long-term in vitro culture. (b) Numbers of deleterious mutations in explants and callus lines. EX1 and EX2: two explants; NI1 and NI2: two independent nascent calli; Di1 and Di2: two independent diploid calli from long-term culture.

The expression pattern of 29 genes involved in RNA-directed DNA methylation (RdDM) and small RNA biogenesis in different types of calli and different tissues of sweet orange showed that nine genes (for instance, CMT2, RDM1, RDR2, and AGO4) had higher expression in old calli than nascent calli and other tissues of sweet orange (Supplementary Fig. S6), suggesting the possible important role of small RNAs in maintaining genome stability for long-term in vitro culture.

We then determined the pattern of DNA methylation on protein-coding genes to detect potential effects on gene function due to changes in DNA methylation. The most noticeable changes were concentrated in the upstream and downstream regions of protein-coding genes, while DNA methylation rates in the gene body changed little, suggesting that the reprogramming of DNA methylation during in vitro culture mainly targeted regulatory elements acting on transcription (Fig. 4b). Consistent with the CHH methylation changes described earlier, the regions flanking the gene body, especially the upstream regions, exhibited a much lower CHH methylation level in nascent calli (Fig. 4b). The same regions, however, gained CHH methylation after long-term culture even as compared to those in the control explants (6.43%, 4.79%, and 11.60% in the control explant, nascent calli, and old calli, respectively; Fig. 4b). We also noted islands of high CHH methylation acting as epigenetic insulators from TE interference for nearby protein-coding genes within the 1-kb region upstream of the genes (Fig. 4b). We hypothesize that these solid CHH islands with high methylation level may protect genes from neighboring TEs in long-term cultured diploid calli [36].

Reshaping of subgenome dominance during in vitro culture and whole-genome duplication

Because sweet orange is a hybrid between mandarin and pummelo, we assessed subgenome dominance before and after in vitro culture. As we identified tetraploid calli in the starting material subjected to long-term in vitro culture (Supplementary Fig. S1), we investigated the role of genome duplication generated in vitro and in vivo by using both diploid and tetraploid callus lines as well as diploid and tetraploid plants.

Based on the genotypes of accessions from mandarin population and pummelo population (originally sequenced in the citrus germplasm collection) [7], the underlying subgenome composition (namely mandarin-origin alleles and pummelo-origin alleles) of all lines were identified for all heterozygous sites. We thus assembled a list of 84639 SNP sites with a clear parent of origin. We then combined this information with ASE information from RNA-seq data to determine which allele (i.e. the allele from which parent of origin) exhibited a dominant expression pattern for all materials used here. We also identified CNV for mandarin-origin and pummelo-origin by combining the information of allele frequency and total DNA copy number from the whole-genome sequencing data.

The diploid explants displayed a symmetrical distribution of relative parent-of-origin expression, whereas nascent calli had a skewed distribution toward dominant expression originating from the pummelo subgenome (Fig. 5a). By contrast, old diploid calli recovered the symmetrical distribution seen in the diploid explant (skewness coefficient, 0.00, −0.24, and 0.00 in the explant, nascent calli, and old diploid calli, respectively; Fig. 5a). From an estimated genome size of 380 Mb in diploid sweet orange, we detected regions covering 95.30 Mb and 30.96 Mb with parent-of-origin-specific CNV in nascent calli and in old diploid calli, respectively (Supplementary Fig. S7a and b). The pummelo subgenome contributed 79.56% (nascent calli) and 36.95% (old diploid calli) to these regions, based on pummelo-specific copy number gain (Fig. 5c), which was in agreement with the observed subgenome dominant expression pattern and suggested a possible causal relationship.
Figure 4. Dynamic changes in DNA methylation after dedifferentiation and long-term in vitro culture. (a) Circos plot of changes in methylation rates across the genome during the dedifferentiation and long-term in vitro culture. Tracks represent: changes in CG methylation ratio in NI relative to EX (I) and Di relative to NI (II); changes in CHG methylation ratio in NI relative to EX (III) and Di relative to NI (IV); changes in CHH methylation ratio in NI relative to EX (V) and Di relative to NI (VI); and chromosome ideogram of sweet orange (VII). The values of methylation rates changes were log2-transformed. (b) Patterns of DNA methylation levels along protein-coding genes in calli. Protein-coding genes were scaled from their transcriptional start sites (TSS) to their transcriptional termination sites (TTS). Average CG, CHG, and CHH methylation levels along the gene body and flanking 2 kb upstream and downstream regions were calculated in 40 intervals. EX: explant; NI: nascent callus; Di1 and Di2: two independent diploid calli from long-term in vitro cultures.

We then repeated the same analysis with polyploid calli, which identified regions with subgenome bias accounting for 54.44 Mb (tetraploid plant, Supplementary Fig. S7c), 36.97 Mb (tetraploid calli, Supplementary Fig. S7d), and 31.02 Mb (hexaploid calli) of their genomes. The degree of subgenome dominance was more obvious in the tetraploid plant than in tetraploid calli, as evidenced by a new peak for the mandarin-specific dominant expression in the tetraploid plant; by contrast, the distribution of relative expression in tetraploid calli was roughly symmetric (Fig. 5b). Looking at parent-specific copy number gain, 83.94% of duplicated regions came from the mandarin subgenome in the tetraploid plant against only 45.98% in tetraploid calli (Fig. 5d).

As an example, on chromosome 7, the reference haplotype of the haploid genome of sweet orange mostly originated from mandarin, and the alternate haplotype-genome mostly originated from pummelo (Fig. 5e). Most of the mandarin-origin haplotypes of chromosome 7 were duplicated in tetraploid calli, as its mean genome copy number rose to six, whereas the reference allele frequency fluctuated around 0.75 (Fig. 5f). The total genome copy number for chromosome 7 in the tetraploid plant was exactly four, as expected, while the frequency of the reference allele rose somewhat to ~0.80 (Fig. 5g), indicating that there were large segmental duplications of mandarin origin and/or deletions of pummelo origin in the tetraploid plants. Corresponding to the extent of chromosome imbalance derived from duplication of chromosome 7 with mandarin-specific allele frequency of 0.80, 0.75, and 0.5 in the tetraploid plant, tetraploid callus, and explant, respectively, the dominant expression of mandarin subgenome was most evident in the tetraploid plant, then in the tetraploid callus, and least in the diploid explants of sweet orange (Fig. 5e–g).

Effects of somaclonal variation on transcriptome reprogramming

To investigate the changes in the transcriptome caused by somaclonal variation, we performed comprehensive transcriptomic profiling. We generated libraries for six calli (one nascent callus, two independent diploid old calli, one double-haploid callus, one tetraploid callus, and one hexaploid callus) and six libraries from sweet orange tissues at various developmental stages (leaves, seeds, young fruits, ripe fruits, early-stage ovules, and late-stage ovules) with an average of 9.36 Gb for each library (Supplementary Table S4). Using weighted gene co-expression network analysis (WGCNA), 24 modules
A co-expression module comprising 884 genes (MEpink module) was specifically expressed in the nascent callus (correlation coefficient $R = 0.99; P = 4e-10$) (Supplementary Fig. S8a). Gene Ontology (GO) enrichment analysis revealed that these genes were enriched in cell wall macromolecule catabolism (for biological process category), and encoded proteins acting at the mitochondrial membrane and in the extracellular space or associating with ribosomes (for cellular component category) (Supplementary Fig. S8b), which was in agreement with the increased oxygen uptake of nascent calli [37].

Nascent calli exhibit a remarkable capability for embryogenesis under in vitro culture conditions (Fig. 6a), which diminishes with long-term culture [38]. Among the 884 genes specifically expressed in nascent callus, we identified 147 hub genes with high connectivity to other specifically expressed genes as candidate genes probably involved in regeneration capability (Supplementary Table S8). We noticed several genes related to somatic embryogenesis (Fig. 6b), including Cs5g_pb010580, a homolog of KNOTTED1-LIKE HOMEOBOX GENE 6 (KNAT6) from A. thaliana [39]; Cs3g_pb011040, encoding HISTONE DEACETYLASE6 (HDA6) [40, 41]; and Cs5g_pb013350 and Cs5g_pb013360, encoding LATE EMBRYOGENESIS

Figure 5. Subgenome dominance after in vitro culture and whole-genome duplication. (a, b) Density plots of difference of allele-specific expression (ASE) between mandarin and pummelo in diploid calli (a) and in polyploid calli and plants (b). (c, d) Pie charts of the percentages of parent-of-origin-specific variation in copy number in diploid calli (c) and in polyploid calli and plants (d). (e) Haplotype constitution and difference of allele-specific expression between mandarin and pummelo (ΔASE) pattern along chromosome 7 of sweet orange plant (EX). Top panel: schematics of the contribution of mandarin (red) and pummelo (blue) in the reference haplotype of the haploid genome of sweet orange (Ref) and the alternate haplotype-genome (Alt) in sweet orange; bottom panel: difference in expression levels of alleles of mandarin to pummelo origin. Red indicates mandarin-origin dominant expression, while blue indicates pummelo-origin dominant expression. (f, g) Allele-specific copy number variation and difference of allele specific expression along chromosome 7 in the tetraploid callus (Tet, f) and tetraploid plants (TetP, g). Top panels: total copy number variation; middle panels, frequency of the R allele (the allele of the Ref haplotype, whose frequency is a measure of relative copy number in the Ref and Alt haplotypes); bottom panels: difference in expression levels of alleles of mandarin to pummelo origin. Chromosome 7 (Chr7) is shown here, remaining chromosomes are shown in Supplementary Fig. S7.
ABUNDANT (LEA) proteins, which were considered as the markers of somatic embryogenesis [42].

Furthermore, we identified many genes apparently affected by the genetic variation or/and epigenetic instability resulting from in vitro culture (Fig. 6b). CLAVATA3/EMBRYO SURROUNDING REGION-RELATED (CLE) peptides are a group of small secreted signaling peptides that are primarily involved in the regulation of stem cell homeostasis in plant meristems [43]. In the nascent callus, we detected an LTR insertion in the promoter of a gene homologous to CLE16 (Cs2g_pb010960 in sweet orange), which we experimentally validated by PCR amplification and sequencing (Fig. 6c and d). In sweet orange explants and old calli, this promoter region was heterozygous, as evidenced by the amplification of two PCR products of different sizes: one allele included a 238-bp LTR insertion (P allele), while the other allele lacked this insertion (p allele) (Fig. 6c and d). However, the p allele was absent from nascent calli as a result of the insertion of a novel 238-bp DNA fragment into the CLE16 promoter (Fig. 6c and d). Sequence analysis revealed that the newly inserted LTR fragment was of the same length but was not identical to the P allele, as it carried one A-to-G transition, prompting us to refer to this allele as P'. (Fig. 6c). The LTR fragment might have firstly inserted in CLE16 and then acquired a polymorphism relative to the p and P alleles. CLE16 transcript levels were high only in the nascent callus (Fig. 6b), while the CLE16 locus was hypermethylated in old calli (Fig. 6e), implying the probable effects of the TE insertion during dedifferentiation or/and epigenetic alternation from long-term in vitro culture on this somatic embryogenesis-related gene.

Discussion

Dedifferentiation is a process to acquisition of the stem-cell-like tissues. Somaclonal variation, a kind of somatic mutation occurring during tissue culture, has been less investigated than meiotical or mitotical variations. However, somaclonal variation has recently drawn much attention due to the extensive use of genome editing technologies, which rely heavily on tissue culture for most plant species.
Our study provides the comprehensive quantitative evaluation of somatic variation during dedifferentiation and long-term in vitro culture, revealing the dramatic acquisition and later purging of new genetic and epigenetic changes (Fig. 7). During dedifferentiation, a burst of multi-direction mutations and “genome shock” in response to the newly in vitro culture environments might have happened as evaluated by the genome-wide SNPs and the TE outbreak (Fig. 7). On the basis of much high expression of genes involved in cell division, DNA replication as well as oxidative stress and different types of variations (including SNPs, InDels, and TE transpositions) occurred in the nascent callus, the striking variations during the dedifferentiation might be caused by aberrant genome replication due to the rapid and exuberant cell division, as well as DNA damage incurred from the severe stress caused by tissue culture [37, 44]. The massive mobilization of LTR retrotransposons after dedifferentiation may contribute to genome plasticity and may also account for the high frequency of somatic mutations.

By contrast, calli maintained in in vitro culture for 30 years were characterized by lower frequencies of genomic variation. These calli have been repeatedly selected for vigorous growth in a tissue culture setting, with one generation taking about two weeks, and are therefore the product of constant artificial selection. The purifying selection during the process of long-term in vitro culture might act by retaining the cells with fewer or irrelevant mutations, which intrinsically underwent fewer mutations or had experienced the DNA repairs, thus eliminating the mutated cells that could not survive normally. We established that the vast majority (91%) of variations that appeared during dedifferentiation were purged during long-term in vitro culture. The number of TE insertions also decreased, and those remaining were largely inserted in intergenic regions, possibly having little damage to the hosts [45, 46]. We also observed an initial decrease in CHH methylation in nascent calli, consistent with TE upheaval and their massive activation and insertion, followed by an increase in CHH methylation after 30 years of in vitro culture (Fig. 7). CHH-type DNA methylation was likely to be an important factor controlling retrotransposon activity in Citrus by hypermethylation on the CHH islands. The higher CHH methylation levels seen in old calli mainly affected the 5' or 3' regulatory regions of protein-coding genes. Broadly, our results suggest a possible interaction between environmental stresses brought upon by in vitro culture, changes in DNA methylation, and TE activation, hinting at a reversible and flexible epigenetic regulation process during tissue responses and adaptation to in vitro culture.

The calli used for our studies were from sweet orange, which is a hybrid descended from a cross between two citrus species, thus providing the possibility of analyzing parent-of-origin variation and gene expression patterns during in vitro culture. We isolated diploid and tetraploid calli grown in tissue culture for 30 years: tetraploid calli exhibited an enhanced subgenome dominance compared with diploid calli. However, subgenome dominance in the in vitro tetraploid was weaker than that in the in vivo tetraploid, which might reflect intrinsic differences between the biological activities of tetraploid calli and tetraploid plants. An enhanced subgenome-dominant expression of genes from mandarin or pummelo origin in a polyploid genome would reshape the phenotypic performance of this hybrid species, and may be a promising target for breeding new varieties.
We propose that changes in TE activity, dynamic DNA methylation, and the reprogramming of gene expression may ultimately affect the ability of somatic embryogenesis during in vitro culture, as summarized in Fig. 7. Our study may guide researchers to select the appropriate duration of in vitro culture to avoid or gain additional variation as a function of the desired goals. Perhaps counterintuitively, we found that a lower mutation load can be accomplished by maintaining plants in tissue culture longer after dedifferentiation, to allow the purging of most acquired variation. This could be especially critical for clonal propagation and experiments relying on the generation of transgenic plants, and would be consistent with findings in cocoa (*Theobroma cacao*) [13] and rice [6]. Conversely, newly induced calli would constitute a desirable material with many new polymorphisms that might be harnessed during breeding. Finally, we identified a number of candidate genes associated with plant somatic embryogenesis potential, providing key target genes for future studies.

Materials and methods

**Materials and sequencing**

The calli induced from immature unfertilized ovules of Valencia sweet orange in 1989 were subcultured twice a month, and the conditions for subculture were as following: 25°C for temperature, and 33 μmol/m²·s 14 h/d for light condition. After the measurement for ploidy levels by flow cytometry, two independent lines of diploid calli, one tetraploid callus line, and one hexaploid callus line were selected for the subsequent analyses. The nascent calli were newly induced from the immature unfertilized ovules, and grew in the tissue culture media to meet the requirements of quantity for subsequent library construction and sequencing. The nascent calli, immature unfertilized ovules (explants) of Valencia sweet orange, and leaves of tetraploid sweet orange were also used in this study.

The total DNA for each of the materials was extracted by improved CTAB method. The DNA sequencing was conducted on the Illumina platform and the PacBio platform with an average depth of 40-fold and 15-fold genome coverage, respectively. For each of the materials, the whole genome bisulfite sequencing was conducted on the Illumina platform with an average depth of 40-fold genome coverage with three biological replicates. For each material, total RNA was extracted using the TRIzol reagent (Takara). The RNA sequencing was conducted on the Illumina platform with three biological replicates and an average of 10 Gb sequencing data were generated for each sample. The published RNA sequencing data for the seeds, young fruits, ripe fruits, the early-stage ovules, and the late-stage ovules of sweet orange were collected [47]. The overall data information was listed in the Supplementary Tables S1–S4.

The constitution analysis of pummelo and mandarin in sweet orange

The published DNA sequencing data of 108 representative citrus accessions were collected [7]. The paired-end data were aligned to the updated reference genome of sweet orange (URLs, http://citrus.hzau.edu.cn) using BWA [48]. Population-based SNP calling was performed using SAMtools [49]. The SNPs in the regions of single-copy orthologous genes were used to conduct structure analysis using Admixture [50]. The optimum K value for the population structure was set as 6, then 17 mandarin accessions and 19 pummelo accessions with the percentage of the corresponding constituent greater than 99.99% were identified as representative mandarins and pummelos with homogeneous genetic backgrounds, respectively.

To determine the mandarin-origin allele and pummelo-origin allele for the heterozygous SNPs in sweet orange, the SNPs in the leaves of sweet orange were detected and the allele frequencies in the aforementioned representative mandarins and pummelos were calculated. The steps and criteria for determining the mandarin-origin allele and pummelo-origin allele were as following: (i) for the heterozygous sites (R/A) in sweet orange, the genotypes in the 17 mandarins and 19 pummelos could only be R/R, A/A, or R/A; (ii) the R allele frequency in the mandarins and pummelos was greater than 0.2, and the absolute value of the difference of R allele frequency between mandarins and pummelos was greater than or equal to 0.8; (iii) if the R allele frequency in the mandarin was greater than 0.8, the R allele was assigned as the mandarin-origin allele and A allele was assigned as the pummelo-origin allele; otherwise, if the R allele frequency in the pummelo was greater than 0.8, the R allele was assigned as the pummelo-origin allele and A allele was assigned as the mandarin-origin allele.

SNPs and CNVs analyses

For each of the materials, the reads from the Illumina platform were mapped against the reference genome of updated sweet orange using BWA. Then, SAMtools software was used to detect SNPs with parameters MQ greater than 40 and DP greater than five to guarantee the high quality of the SNPs. To estimate the SNP rates between each two of the materials across the genome, a sliding-window approach (100-kb windows sliding in 30-kb steps) was employed.

To estimate the total copy number variation for each of the materials, FREEC software was used with the window size of 30 kb and the setting of their corresponding ploidy number [51].

Identification of deleterious mutations

A comparative genomic analysis was performed for eight species in the Malvids clade of the plant kingdom, including the sweet orange, *Carica papaya*, *A. thaliana*, *Eucalyptus grandis*, *T. cacao*, *Gossypium arboreum*, *Durio zibethinus*, and *Brassica rapa*. Seven runs of pairwise alignment by
aligning the corresponding genome sequences to the sweet orange genome sequences were conducted by the LASTZ software [52]. An eight-way multiple alignment was performed using the MULTIZ software [53]. Based on 4-fold degenerate sites of sweet orange genome, the phylogenetic tree of the eight species was constructed by the phyFit program of the PHAST software [54]. On the basis of the multiple alignments and estimated phylogenetic tree, conservation scores of individual sites were computed with the phyloP program of the PHAST software, and genomic sites of sweet orange with a conservation score above one were defined as conserved.

A set of 11 Citrus ichangensis accessions were used as outgroup, and the sequencing data of them were used to predict the ancestral allele and derived allele for poly-morphic sites in sweet orange accessions. For each biallelic site, the allele was defined as ancestral allele if 10 to 11 C. ichangensis had the same homozygous genotype, and the other allele was defined as derived allele. For the SNPs from ancestral allele to derived allele, analysis of variant effect was performed using the SnpEff software [56]. The functional effects of non-synonymous amino acid substitutions were predicted using the PROVEAN software [57]. For the conserved genomic sites of sweet orange, the amino acid substitution with score from the PROVEAN software of less than −2.5 was predicted to be deleterious.

The detection of structure variations and TE insertions
To detect the structure variations for each of the materials, the reads generated from the PacBio platform were corrected with the high-quality paired-end reads from the Illumina platform using proovread [57]. The corrected reads were aligned to the updated sweet orange genome using the aligner NGM-LR [16]. Then, the Sniffles software was employed to perform the structure variations calling and genotyping with the number of supporting reads for SVs greater than five [16]. For the further comparisons of structure variations between each two of the materials, the merge program in the SURVIVOR software was utilized with the following parameters: maximal distance between breakpoints, 1000 bp; minimal number of supporting caller, one; take the type and the strands into account; do not estimate distance based on the size of SV; minimal size of SVs to be taken into account, 30 bp (URLs, https://github.com/fritzsedlazeck/SURVIVOR/wiki).

For the identification of TE insertions, the results of insertions were used for further analysis. The insertion events were manually checked using IGV to screen the false detections of insertions [58]. To identify the insertions of TEs and determine the types of the inserted TEs, the inserted DNA sequences were retrieved from the mapped long reads, and then aligned against the whole-genome TEs database of sweet orange using BLASTN software [59].

Experimental validation for SNPs and TE insertions
A total of 62 SNPs between five pairs of materials (e.g. the nascent callus vs. the explant leaf, two pairs of the long-term cultured diploid callus vs. the nascent callus, the long-term cultured tetraploid callus vs. the nascent callus, and the long-term cultured hexaploid callus vs. the nascent callus) were randomly selected. The approximately 200–800 bp DNA segments embedding the SNP sites were cloned in the corresponding materials by PCR amplification and the PCR products were sequenced by Sanger sequencing to validate the genotypes on the target sites. PCR was carried out in 50-μl reaction volumes containing 5 μl of dNTP Mix, 1 μl of Phanta Max Super-Fidelity DNA Polymerase, 25 μl of 2× Phanta Max Buffer, 400 ng of genome template, and 2 μl of each forward and reverse primer. The amplification was performed at 95°C for 3 min, followed by 33 cycles of 15 s at 95°C, 15 s at 55°C, and 30 s at 72°C, and finally 5 min at 72°C.

The PCR-based screen was performed to validate the exact TE integration sites and inserted sequences. Based on the flanking regions of TE integration sites with few sequences variations among materials, the primers were designed that would generate the longer band of a longer PCR product containing TE insertion and the shorter band of a shorter PCR product without TE insertion (Supplementary Table S9). After Sanger sequencing of PCR products or E.coli containing PCR products, the extra sequences in the longer PCR product relative to the shorter PCR product were compared with the predicted inserted TE sequences.

Transcriptomic analyses and weighted gene coexpression network analysis
The genome index of updated sweet orange was constructed using Bowtie2 [60], and the reads from RNAseq were aligned to the sweet orange genome sequences using TopHat [61]. The gene expression level in each RNAseq library was calculated as the FPKM with cuffquant and cuffnorm programs in Cufflinks [62]. The significantly differentially expressed genes between each pair of materials were identified by cuffdiff program in Cufflinks with fold change greater than 2, P-value less than 0.05, and FDR value less than 0.05. Gene Ontology (GO) enrichment analysis was conducted using the agriGO program with P-value less than 0.05 and FDR value less than 0.05 [63].

The expression profiling data of 31 samples from calli lines, leaves, fruits, seeds, and ovules were used for coexpression networks construction using WGCNA package in R [64]. A total of 19708 genes with the maximal FPKM value across the samples higher than 0.1 were used for the WGCNA analysis. The modules were generated using the automatic network construction function blockwise Modules with the optimal soft-thresholding power of nine. For the genes in the interesting modules, gene functional annotations were conducted and GO enrichment analysis was conducted using the agriGO program.
The networks were analyzed and visualized using Cytoscape [65].

Allele specific expression analyses
To identify the SNP sites with allele-specific expression pattern, the SNPs of high confidence (DP > 10 and MQ > 40) and the alignment file of RNA-seq were used as inputs into the ASEQ software [66]. The parameters of mbq was set to 30, mrq was set to 30, and mdc was set to 10.

Methylome analyses
The reads from WGBS were firstly trimmed by 6 bp from the 5’ end using Trimmomatic to avoid the bisulfite conversion failures [67]. Then, the reads trimming based on base quality was conducted using Trimmomatic.

The high-quality reads were mapped to the updated sweet orange genome using the bismark program in the Bismark software [68]. The deduplicate_bismark program in the Bismark software was used to remove duplications. The bismark_methylation_extractor program in the Bismark software was used to calculate the DNA methylation ratios for each of cytosine sites. The DNA methylation patterns across the genome were plotted using Circos [69].

To identify the differentially methylated regions, methylKit package in R was utilized with the following settings: lo.count = 10 and hi.perc = 99.9 for sites filtering; win.size = 1000 and step.size = 1000 for sliding windows setting; difference = 25 and q-value = 0.01 for the significance criteria of differential methylation [70].

To measure the DNA methylation pattern along protein-coding genes, the average DNA methylation levels of CG-, CHG-, and CHH-type in 40 equally long bins separately in the 2-kb upstream regions, 2-kb downstream regions, and the gene body regions were calculated. As to the DNA methylation pattern on the TE regions, the TEs were classified into different types based on the TE annotation, and the DNA methylation pattern along the TE regions were calculated as described earlier.

Acknowledgments
We are grateful to J. Ma for providing the citrus samples. We thank B. Hu, S. Liu, H. Zuo, M. Miao and X. Liu for help on manual check of TE insertion. This work was supported by National Key Research and Development Program of China (2018YFD1000101 to Q. Xu), the National Natural Science Foundation of China (31925034, 31872052, 32001998 to Q. Xu and X. Wang), and National Postdoctoral Program for Innovative Talents (BX201900134) (X. Wang).

Author contributions
Q.Xu conceived and designed the project and the strategy. X.Wang performed the whole bioinformatics analyses and interpretation of data. L.Ke, J.Fu, and J.Xu prepared materials for sequencing. L.Ke and S.Wang performed experimental validations of genetic variations. Y.Hao, X.Deng, and W.Guo took care of the callus materials. X.Wang wrote the manuscript, with contributions from Q.Xu and C.Kang. All authors read and approved the final manuscript.

Data availability
All data used in this study are publicly available. Whole-genome sequencing data have been deposited in BioProject under accession numbers PRJNA573566. Whole-genome bisulfite sequencing data have been deposited in BioProject under accession numbers PRJNA573590. The RNA-seq data have been deposited in BioProject under accession codes PRJNA573594. The data for genome assembly and annotation of updated C. sinensis are available through our website at http://citrus.hzau.edu.cn.

Conflict of interest
The authors declare no competing interests.

Supplementary data
Supplementary data is available at Horticulture Research Journal online.

References
1. Jiang F, Feng Z, Liu H et al. Involvement of plant stem cells or stem cell-like cells in dedifferentiation. Front Plant Sci. 2015;6:1–6.
2. Sugiyama M. Historical review of research on plant cell dedifferentiation. J Plant Res. 2015;128:349–59.
3. Larkin PJ, Scowcroft WR. Somaclonal variation—a novel source of variability from cell cultures for plant improvement. Theor Appl Genet. 1981;60:197–214.
4. Krishna H, Alizadeh M, Singh D et al. Somaclonal variations and their applications in horticultural crops improvement. 3 Biotech. 2016;6:54–71.
5. Li J, Manghwar H, Sun L et al. Whole genome sequencing reveals rare off-target mutations and considerable inherent genetic or/and somaclonal variations in CRISPR/Cas9-edited cotton plants. Plant Biotechnol J. 2019;17:858–68.
6. Wei FJ, Kuang L-Y, Oung H-M et al. Somaclonal variation does not preclude the use of rice transformants for genetic screening. Plant J. 2016;85:648–59.
7. Wang X, Xu Y, Zhang S et al. Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. Nat Genet. 2017;49:765–72.
8. Spillane C, Curtis MD, Grossniklaus U. Apomixis technology development—virgin births in farmers’ fields? Nat Biotechnol. 2004;22:687–91.
9. Wu GA, Prochnik S, Jenkins J et al. Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. Nat Biotechnol. 2014;32:656–62.
10. Xu Q, Chen L-L, Ruan X et al. The draft genome of sweet orange (Citrus sinensis). Nat Genet. 2013;45:59–66.
11. Chaturvedi H, Mitra G. Clonal propagation of Citrus from somatic callus cultures. HortScience. 1974;9:118–20.
12. Wang QM, Wang L. An evolutionary view of plant tissue culture: somaclonal variation and selection. Plant Cell Rep. 2012;31:1535–47.
13. Rodríguez Lopez CM, Wetten AC, Wilkinson MJ. Progressive erosion of genetic and epigenetic variation in callus-derived cocoa (Theobroma cacao) plants. New Phytol. 2010;186:856–68.
14. Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. Nat Rev Genet. 2011;12:363–76.
15. Eid J, Fehr A, Gray J et al. Real-time DNA sequencing from single polymerase molecules. Science. 2009;323:133–8.
16. Sedlazeck FJ, Rescheneder P, Smolka M et al. Accurate detection of complex structural variations using single-molecule sequencing. Nat Methods. 2018;15:461–8.
17. Gaur U, Li K, Mei S et al. Research progress in allele-specific expression and its regulatory mechanisms. J Appl Genet. 2013;54:271–83.
18. Skelly DA, Johansson M, Madeoy J et al. A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. Genome Res. 2011;21:1728–37.
19. Edger PP, Smith R, MR MK et al. Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. Plant Cell. 2017;29:2150–67.
20. Zhao B, Cao J-F, Hu G-J et al. Core cis-element variation conforms subgenome-biased expression of a transcription factor that functions in cotton fiber elongation. New Phytol. 2018;218:1061–75.
21. Shao L, Xing F, Xu C et al. Patterns of genome-wide allele-specific expression in hybrid rice and the implications on the genetic basis of heterosis. Proc Natl Acad Sci U S A. 2019;116:5653–8.
22. Shen R, Wang L, Liu X et al. Genomic structural variation-mediated allelic suppression causes hybrid male sterility in rice. Nat Commun. 2017;8:1310–9.
23. Pastholdt A, Larson NB, Marcon C et al. Nonsyntenic genes drive highly dynamic complementation of gene expression in maize hybrids. Plant Cell. 2014;26:3939–48.
24. Waters AJ, Makarevitch I, Noshay J et al. Natural variation for gene expression responses to abiotic stress in maize. Plant J. 2017;89:706–17.
25. Zhang X, Borevitz JO. Global analysis of allele-specific expression in Arabidopsis thaliana. Genetics. 2009;182:943–54.
26. Peschke VM, Phillips RL, Gengenbach BG. Discovery of transposable element activity among progeny of tissue culture-derived maize plants. Science. 1987;238:804–7.
27. Hirochika H, Sugimoto K, Otsuki Y et al. Retrotransposons of rice involved in mutations induced by tissue culture. Proc Natl Acad Sci U S A. 1996;93:7783–8.
28. Lin C, Lin X, Hu L et al. Dramatic genotypic difference in, and effect of genetic crossing on, tissue culture-induced mobility of retrotransposon Tos17 in rice. Plant Cell Rep. 2012;31:2057–63.
29. Kazuhiro K, Kazuki T, Masamitsu W et al. The plant mTE mPing is mobilized in anther culture. Nature. 2003;421:167–70.
30. Komatsu M, Shimamoto K, Kyoizuka J. Two-step regulation and continuous retrotransposition of the rice LINE-type retrotransposon karma. Plant Cell. 2003;15:1934–44.
31. Kaeppler SM, Kaeppler HF, Rhee Y. Epigenetic aspects of somaclonal variation in plants. Plant Mol Biol. 2000;43:179–88.
32. Miguel C, Marum L. An epigenetic view of plant cells cultured in vitro: somaclonal variation and beyond. J Exp Bot. 2011;62:3713–25.
33. Escher PW, Simon SA et al. Plants regenerated from tissue culture contain stable epigenome changes in rice. Elife. 2013;2:e00354.
34. Ong-Abdullah M, Ordway JM, Jian N et al. Loss of Karma transposon methylation underlies the mantled somaclonal variant of oil palm. Nature. 2019;525:533–7.
35. Gent J, Ellis NA, Guo L et al. CHH islands: de novo DNA methylation in near-gene chromatin regulation in maize. Genome Res. 2013;23:628–37.
36. Cassells AC, Curry RF. Oxidative stress and physiological, epigenetic and genetic variability in plant tissue culture: implications for micropropagators and genetic engineers. Plant Cell Tissue Organ Cult. 2003;64:145–57.
37. Etienne H, Bertrand B. Somaclonal variation in Coffea arabica: effects of genotype and embryogenic cell suspension age on frequency and phenotype of variants. Tree Physiol. 2003;23:419–26.
38. Belles-Boix E, Hamant O, Witiak SM et al. KNAT6 an Arabidopsis homeobox gene involved in meristem activity and organ separation. Plant Cell. 2006;18:1900–7.
39. Hung FY, Chen F-F, Li C et al. The Arabidopsis DLD1/2-HDA6 histone modification complex is functionally associated with CCA1/LHY in regulation of circadian clock genes. Nucleic Acids Res. 2018;46:10669–81.
40. Luo M, Yu C-W, Chen F-F et al. Histone deacetylase HDA6 is functionally associated with AS1 in repression of KNOX genes in Arabidopsis. PLoS Genet. 2012;8:e1003114.
41. Alkan C, Coe BP, Eichler EE. Genome structural variation underlying a major fitness trade-off in Arabidopsis thaliana. BMC Genomics. 2008;9:118–39.
42. Rodriguez-Leal D, Xu C, Kwon C-T et al. Evolution of buffering in a genetic circuit controlling plant stem cell proliferation. Nat Genet. 2019;51:786–92.
43. Poetsch AR, Boulton SJ, Luscombe NM. Genomic landscape of oxidative DNA damage and repair reveals regioselective protection from mutagenesis. Genome Biol. 2018;19:215–37.
44. Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. Nat Rev Genet. 2017;18:71–86.
45. Naito K, Zhang F, Tsukiyama T et al. Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. Nature. 2009;461:1130–4.
46. Ke L, Zhou Z, Xu X-W et al. Evolutionary dynamics of lin-cRNA transcription in nine citrus species. Plant J. 2019;98:912–27.
47. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.
48. Li H, Handsaker B, Wysoker A et al. The sequence alignment/map and SAMtools. Bioinformatics. 2009;25:2078–9.
49. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19:1655–64.
50. Boeve V, Popova T, Bleakley K et al. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics. 2012;28:423–5.
51. Harris RS. Improved pairwise alignment of genomic DNA. Ph.D. Thesis. The Pennsylvania State University; 2007.
52. Blanchette M, Kent WJ, Riemer C et al. Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res. 2004;14:708–15.
54. Hubisz MJ, Pollard KS, Siepel A. PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief Bioinformatics*. 2010;12:41–51.

55. Cingolani P, Platts A, Wang Le L et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012;6:80–92.

56. Cingolani P, Platts A, Wang Le L et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012;6:80–92.

57. Hackl T, Hedrich R, Schultz J et al. proovread: large-scale high-accuracy PacBio correction through iterative short read consensus. *Bioinformatics*. 2014;30:3004–11.

58. Robinson JT, Thorvaldsdóttir H, Winckler W et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29:24–6.

59. Altschul SF, Gish W, Miller W et al. Basic local alignment search tool. *J Mol Biol*. 1990;215:403–10.

60. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.

61. Kim D, Pertea G, Trapnell C et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. * Genome Biol*. 2013;14:295–311.

62. Trapnell C, Williams BA, Pertea G et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010;28:511–5.

63. Du Z, Zhou X, Ling Y et al. agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res*. 2010;38:64–70.

64. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9:559–71.

65. Su G, Morris JH, Dernchak B et al. Biological network exploration with Cytoscape 3. *Curr Protoc Bioinformatics*. 2014;47:11–24.

66. Romanel A, Lago S, Prandi D et al. ASEQ: fast allele-specific studies from next-generation sequencing data. *BMC Medical Genom*. 2015;8:9–20.

67. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114–20.

68. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*. 2011;27:1571–2.

69. Krzywinski M, Schein J, Birol I et al. Circos: an information aesthetic for comparative genomics. *Genome Res*. 2009;19:1639–45.

70. Akalin A, Kormaksson M, Li S et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol*. 2012;13:87–95.