Article

Genome sequencing of 'Fuji' apple clonal varieties reveals genetic mechanism of the spur-type morphology

Yudong Cai [1,2,10], Xiuhua Gao[1,10], Jiangping Mao [1,10] ✉, Yu Liu[1,10], Lu Tong[1,10], Xilong Chen [1,10], Yandong Liu[1], Wenyan Kou[2], Chuanjun Chang[1], Toshi Foster [3], Jialong Yao [3], Amandine Cornille[4], Muhammad Mobeen Tahir[1], Zhi Liu[5], Zhongye Yan[5], Siyi Lin[1], Fengwang Ma [1], Juanjuan Ma[1], Libo Xing[1], Na An[1], Xiya Zuo[1], Yanrong Lv[1], Zhengyang Zhao[1], Wenqiang Li[6], Qianjin Li[7], Caiping Zhao[1], Yanan Hu[1], Hangkong Liu[1], Chao Wang[1], Xueyan Shi[1], Doudou Ma[1], Zhangjun Fei [8,9] ✉, Yu Jiang [2] ✉ & Dong Zhang [1] ✉

Somatic variations can give rise to bud sports with advantageous traits, serving as the foundation for bud sport breeding in perennial plants. Here, we report a fully phased genome assembly of 'Fuji' apple, enabling comprehensive identification of somatic variants across 74 clonally propagated 'Fuji' varieties. Phylogenetic analysis indicates that spur-type and early-maturation traits in 'Fuji' sport varieties arise from multiple independent events. Several putative functional somatic variants have been identified, including a spur-type-specific deletion in the promoter of the TCP transcription factor gene *MdTCP11*. DNA methylation level of the deletion-associated miniature inverted-repeat transposable element is lower in spur-type varieties compared to standard-type varieties, while the expression of *MdTCP11* is significantly higher. Overexpression of *MdTCP11* in apple decreases plant height, highlighting its important role in the development of spur-type apple varieties. This study sheds light on the cloning history of 'Fuji' and provides valuable resources for apple breeding.

Somatic mutations occur in all living organisms, but studies on these mutations have largely focused on mosaic organs in the context of human aging, cancer, and neurodegeneration. In plants, somatic variations refer to genetic changes that arise during the mitotic cell cycle. They are widespread in perennials, such as fruit tree crops, and contribute to genetic diversity, giving rise to new traits known as bud sports. Bud sports are a main source of new cultivars widely used in apples, citruses, pears, and other fruit trees[1]. Vegetative propagation of bud sports enables the selection and breeding of clonal varieties, preserving somatic variations over long periods of time. The ability to

---

[1]College of Horticulture, Yangling Sub-Center of National Center for Apple Improvement, State Key Laboratory for Crop Stress Resistance and High-Efficiency Production, Northwest A&F University, Yangling 712100 Shaanxi, P. R. China. [2]College of Animal Science and Technology, Northwest A&F University, Yangling 712100 Shaanxi, P. R. China. [3]The New Zealand Institute for Plant and Food Research Ltd, Private Bag 92169, Auckland 1142, New Zealand. [4]Université Paris Saclay, INRAE, CNRS, AgroParisTech, GQE-Le Moulon, Gif-sur-Yvette, France. [5]Liaoning Institute of Polomogy, Yingkou 115009 Liaoning, P. R. China. [6]China Apple Research System Xian Guoyou Association, Qianxian 713300 Shaanxi, P. R. China. [7]Apple Industry R&D Center of Luochuan County, Luochuan 727400 Shaanxi, P. R. China. [8]Boyce Thompson Institute, Cornell University, Ithaca, NY 14853, USA. [9]US Department of Agriculture-Agricultural Research Service, Robert W. Holley Center for Agriculture and Health, Ithaca, NY 14853, USA. [10]These authors contributed equally: Yudong Cai, Xiuhua Gao, Jiangping Mao, Yu Liu, Lu Tong, Xilong Chen. ✉e-mail: mjp588@nwafu.edu.cn; zf25@cornell.edu; yu.jiang@nwafu.edu.cn; afant@nwafu.edu.cn

clonally propagate mosaic or homogeneous somatic mutations is not only valuable for expanding genetic diversity but also allows us to identify mutations responsible for novel traits. However, the high heterozygosity of the apple genome poses challenges in accurately identifying somatic mutations, as many heterozygous germline variants could be misclassified as somatic variants. Furthermore, structural variations or rearrangements between the two haploid genomes can further complicate somatic mutation detection.

In China, 33.9% of apple cultivars are selected and bred through bud sport breeding[2]. 'Fuji' is one of the most prominent cultivars in apple bud sport breeding, with numerous clonally propagated varieties selected from successive generations of 'Fuji' bud sports. Approximately 73.6% of apple cultivars planted in China are 'Fuji' bud sport clones, displaying a wide range of new traits such as spur-type growth habits, red peel, and early maturation. The primary apple-producing regions in China face challenges such as poor soil quality, frequent drought conditions, and a scarcity of suitable dwarf rootstocks. The widespread adoption of spur-type varieties in high-density cultivation, driven by these agricultural conditions, has significantly enhanced apple productivity and orchard establishment. Spur-type apple trees exhibit more concentrated flower bud differentiation and spur formation, leading to increased fruit yield and reduced pruning costs. They are characterized by a compact architecture and reduced size compared to the open-canopy form of standard-type varieties[3]. With shorter, thicker internodes, spur-type trees produce shorter branches, making them more suitable for pruning and fruit harvesting[4-6]. Previous studies have shown that somatic mutations often manifest as bud sports in apples, pears, citruses, and grapes[7-11]. In addition, some structural variants (SVs) associated with transposable element (TE) activities have also been linked to the development of bud sports[12-14]. While multiple studies have explored the formation mechanisms of spur-type varieties from the perspectives of endogenous hormones, transcriptomes, and epigenetics[15-17], the genomic basis of spur-type bud sports still remains unclear.

In this study, we report the assembly of a fully phased chromosome-level genome of 'Fuji'. Using this haplotype-resolved genome as a reference, we detect high-confidence somatic variants from 74 clonally propagated 'Fuji' varieties and uncover major candidate variants and regulatory mechanisms underlying the spur-type morphology in apples. This study lays the foundation for understanding the mechanisms underlying bud-sport emergence and the causative variants of bud sports, and provides valuable resources and insights for apple breeding.

Results

Phased diploid genome assembly of 'Fuji'

'Fuji' was bred in 1939 through a cross between 'Delicious' and 'Ralls Janet'. To assemble a fully phased 'Fuji' genome, we generated 101.36 Gb of PacBio HiFi sequencing data for four and 404.65 Gb of ultra-long Oxford Nanopore Technology (ONT) reads for 13 clonally propagated 'Fuji' varieties (Supplementary Data 1). To distinguish different haplotypes inherited from the two parents, we also generated 239.77 Gb of short-read data for 'Ralls Janet' and 'Delicious' (Supplementary Data 1). Using the trio-binning approach, we first generated four haplotype-resolved contig assemblies for four varieties ('Nagafu No. 2', 'Fengfu2021', 'Red general', and 'Red general spur') with HiFi reads. The ultra-long ONT reads longer than 50 kb (48.45 Gb in total) were integrated into this process to improve the continuity of the assemblies. Using the assembly with the best continuity ('Nagafu No. 2'), we first employed a reference-guided method for scaffolding, and then generated 78.43 Gb Hi-C reads for 'Nagafu No. 2' to guide the manual correction of potential misassemblies (Supplementary Fig. 1). Finally, we used haplotype-resolved contigs from all four varieties to patch gaps in the two haploid assemblies, resulting in a near-complete, fully phased diploid genome of 'Fuji' (Fig. 1a).

The phased assembly of 'Fuji' contained two haploid genomes (haplomes), one inherited from 'Ralls Janet' (hereafter Fuji_Ral) and the other from 'Delicious' (hereafter Fuji_Del). The total lengths of the 17 nuclear chromosomes of Fuji_Del and Fuji_Ral were 654.62 Mb and 653.37 Mb, respectively, which are comparable to or larger than those of recently published genome assemblies of domesticated apples[14,17-24]. Fourteen chromosomes had paired telomeres in at least one haplome, while the remaining three chromosomes had only single telomeres (Supplementary Data 2). The two haplomes showed a high level of similarity to each other (Supplementary Fig. 2) and exhibited a high degree of collinearity with the GDDH13 genome[17] (Supplementary Fig. 3).

We comprehensively evaluated the quality of Fuji_Ral and Fuji_Del. The contig N50 values for the two haploid genomes were 37.30 Mb and
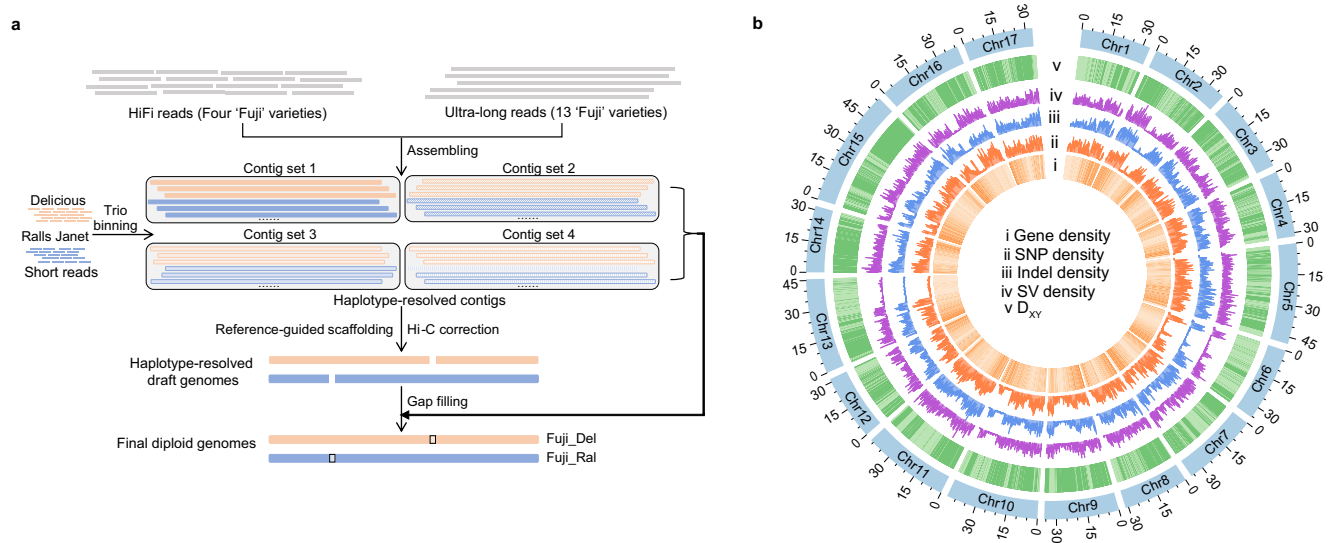

Fig. 1 | Genomic features of 'Fuji'. a Assembly strategy of the fully phased 'Fuji' genome. HiFi reads of four 'Fuji' clonally propagated varieties were assembled into contigs, integrated with ultra-long reads of 13 additional 'Fuji' clonally propagated varieties. Short reads from 'Fuji' parents, 'Delicious' and 'Ralls Janet', were used for phasing. **b** Genomic features of the 'Fuji' genome assembly (window size of 0.1 Mb). D_{XY}, absolute genetic divergence between 'Ralls Janet' and 'Delicious'. Source data are provided as a Source Data file.

36.89 Mb, respectively, which were higher than or comparable to those of previously published apple genomes (Supplementary Data 2). The LTR assembly index (LAI) was 20.73 for Fuji_Ral and 21.34 for Fuji_Del, meeting the 'gold standard'[25]. The BUSCO[26] completeness rates were 98.3% for of Fuji_Ral and 98.8% for Fuji_Del. Both haploid genomes achieved high-quality consensus base calls, with quality values (QV) of 59.06 for Fuji_Ral and 58.70 for Fuji_Del. Furthermore, k-mer-based evaluations also confirmed the phasing accuracy of the haplotype-resolved 'Fuji' genome assembly (Supplementary Note 1 and Supplementary Figs. 4–8).

We predicted a total of 50,646 and 50,678 genes in Fuji_Ral and Fuji_Del, respectively, and most homologous genes between 'Fuji' and GDDH13 showed high similarity, with a median of 99.60% (Supplementary Fig. 9). We annotated 458.75 Mb and 456.80 Mb of repetitive sequences, accounting for 69.34% and 69.35% of Fuji_Ral and Fuji_Del, respectively. The repetitive sequence composition of the two haploid genomes was similar (Supplementary Fig. 10). We also detected insertions and deletions between the two haploid genomes of 'Fuji', uncovering haplotype-specific sequences amounting to 73.87 Mb. This observation was further validated by genomic data from different 'Fuji' varieties, indicating that these sequences likely represent hemizygous regions within the 'Fuji' genome (Supplementary Data 3).

To accurately identify genetic differences between the two haplomes in the 'Fuji' cultivar population, we further sequenced 74 clonally propagated varieties of 'Fuji', generating a total of 2,215.10 Gb of Illumina sequences (Supplementary Data 4). Using both genome alignment and read-based methods at the population level, we detected 4,478,495 single-nucleotide polymorphisms (SNPs), 480,098 small insertions/deletions (indels), and 16,734 SVs between the two 'Fuji' haplomes (Supplementary Data 5 and 6). These variants represented heterozygous alleles shared across different 'Fuji' varieties, inherited from the common ancestor (referred to as germline variants). Interestingly, we observed an unequal distribution of these heterozygous regions across the genome (Fig. 1b). By scanning the nucleotide divergence between 'Ralls Janet' and 'Delicious', the two parents of 'Fuji', along the 'Fuji' genome, we found that regions with low divergence coincided with those with low heterozygosity in the 'Fuji' genome (Fig. 1b), in line with the formation history of 'Fuji'.

Somatic variations in 'Fuji' clonally propagated varieties

During the clonal propagation of 'Fuji', different varieties have been selected based on traits arising from somatic variations. However, given the high heterozygosity of the 'Fuji' genome, there are far more heterozygous germline variants between the two haploid genomes than somatic variants. This can result in numerous germline variants being misidentified as somatic variants. To minimize the misidentification of germline variants as somatic variants, we used the phased diploid 'Fuji' genome as the reference, assigning each read to the appropriate haplome (Fig. 2a). In total, we detected 68,965 somatic SNPs across the 74 individuals, including 532 deleterious mutations (SIFT score ≤0.05). Furthermore, we also identified 27,757 somatic indels and 1848 somatic SVs (Supplementary Data 7–9).

We calculated the density of somatic SNPs in different genomic regions and found that somatic variants showed a similar pattern to germline variants (Fig. 2b and Supplementary Fig. 11). The highest density was found in the upstream and downstream regions of genes, followed by 5' UTRs, 3' UTRs, and introns. In contrast, the lowest density was observed in exons, consistent with the conserved nature of exon regions. Among the somatic SNPs, the proportion of the two transition types was higher than that of the four transversion types (Fig. 2c, d). C->T mutations were the most frequent among transition types, and C->A mutations were the most frequent among transversion types (Fig. 2c). This bias in mutation types is commonly observed in other plant species and aligns with the expected signatures of somatic variants[27–29].

Population structure of 'Fuji' varieties

We performed principal component analysis (PCA) using all germline and somatic SNPs detected based on one haploid genome (Fuji_Ral) to evaluate the genetic relationships among different cultivars. The first and second principal components clearly distinguished the three groups: 'Fuji', 'Ralls Janet', and 'Delicious' (Fig. 3a). To further confirmed these genetic relationships, we conducted a kinship analysis, which revealed that the 74 'Fuji' varieties were clonally related and were the offspring of 'Ralls Janet' and 'Delicious', consistent with the known breeding history of 'Fuji' (Fig. 3b).

To further analyze the genetic relationships among the clonally propagated 'Fuji' varieties, we performed PCA on the 74 varieties using somatic SNPs. The results did not reveal a clear genetic stratification (Fig. 3c). Furthermore, we performed phylogenetic analysis to track the evolutionary relationships among these varieties. The results revealed that spur-type and early-maturing varieties were distributed across multiple clades (Fig. 3d), indicating that both traits evolved independently in varieties from different clades. Notably, one clade consisted solely of spur-type varieties (referred to as the spur-type clade), suggesting that these ten varieties might descend from a common spur-type ancestor. Similarly, we identified a clade enriched with early-maturing varieties (referred to as the early-maturing clade).

Genetic basis of spur-type and early-maturation traits

Spur-type growth habit is a crucial economic trait in apple production. The trunk height and crown width of spur-type clonal varieties are shorter than those of standard-type 'Fuji' varieties. Additionally, spur-type varieties exhibit a higher spur rate compared to standard-type 'Fuji' varieties (Fig. 4a). To further investigate the differences in characteristics between spur-type and standard-type varieties, we used the spur-type 'Liquan spur' and the standard-type 'Yanfu No. 8' as representative varieties. The average shoot length and internode length of 'Liquan spur' were significantly shorter than those of 'Yanfu No. 8' (Fig. 4b and Supplementary Fig. 12). Furthermore, the contents of gibberellin A_3 (GA_3) and GA_4 were lower in 'Liquan spur' compared to 'Yanfu No. 8' (Fig. 4c). Cell differentiation plays a crucial role in internode elongation[30]. Therefore, we examined the morphological characteristics of cells in both spur-type and standard-type varieties. In the paraffin sections, the cell length was significantly shorter in 'Liquan spur' compared to 'Yanfu No. 8', while the cell number in the same area was significantly higher in 'Liquan spur' (Fig. 4d–f). These results strongly suggest that cell length and density are critical factors in determining the internode length of apple trees, consistent with findings from a previous study[30].

Phylogenetic analysis revealed two clades enriched with spur-type and early-maturing varieties, respectively (Fig. 3d). To identify candidate somatic variants associated with spur-type and early-maturation traits, we detected clade-specific variants based on these phylogenetic relationships (Supplementary Data 10). In the spur-type clade we identified 89 SNPs and eight indels specific to spur-type varieties, while in the early-maturing clade, we detected 44 SNPs and two indels specific to the early-maturing varieties (Supplementary Data 11 and 12). We randomly selected 12 spur-type-specific SNPs and verified them in four spur-type and four standard-type varieties using Sanger sequencing (Supplementary Data 13). These clade-specific variants did not appear in other spur-type or early-maturing varieties, further supporting that spur-type and early-maturing varieties that clustered into single clades, respectively, arose from a common ancestor. In contrast, spur-type and early-maturing varieties in other clades likely arose independently.

A TCP-like gene contributes to the spur-type morphology

We further examined whether somatic SVs contributed to the spur-type and early-maturation phenotypes. By comparing the frequencies of each somatic SV across different clades, we did not identify any SVs specific to early-maturing varieties in the early-maturing clade. However, we discovered a 167-bp deletion specific to varieties in the spur-
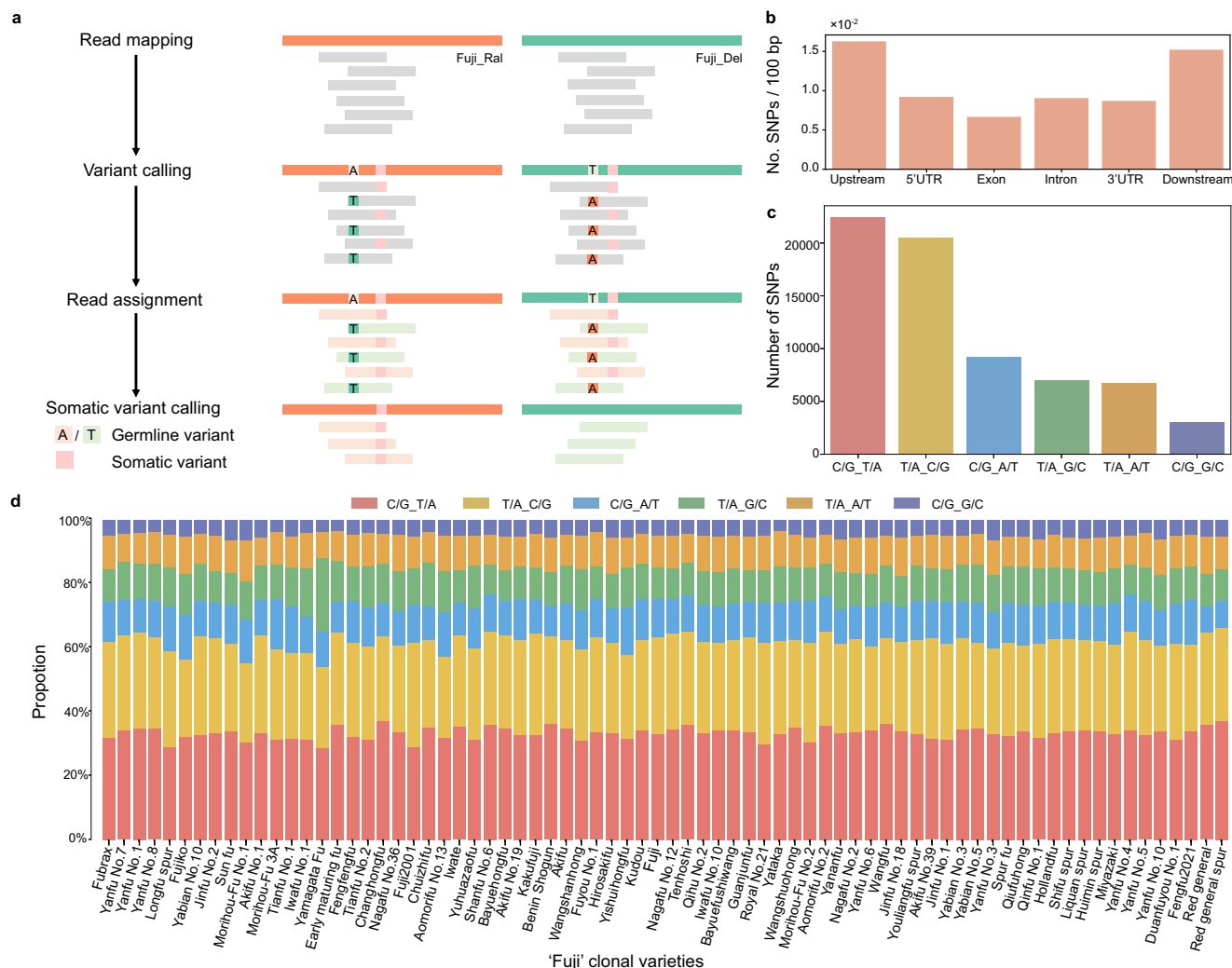
Fig. 2 | Somatic variants in clonally propagated 'Fuji' varieties. a Pipeline for somatic variant detection. Each read is aligned to Fuji_Ral and Fuji_Del, respectively, and then assigned to the optimal haploid genome based on the alignment quality. b Densities of somatic SNPs across different gene features. c Number of different types of somatic SNPs. d Distribution of different types of somatic SNPs in each variety. Source data are provided as a Source Data file.

type clade (Fig. 5a). This variant was present in all ten spur-type varieties within the clade but absent in spur-type varieties in other clades and in standard-type varieties. Notably, this 167-bp deletion overlapped with a 205-bp miniature inverted-repeat transposable element (MITE), encompassing 115 bp of the MITE sequence (Fig. 5b). The MITE existed as homozygous in other apple cultivars and their wild ancestors (Supplementary Fig. 13), indicating that the transposable element was inserted at an early stage. We selected five spur-type (from the spur-type clade) and five standard-type varieties for PCR validation of this deletion, achieving a verification rate of 100% (Fig. 6a). The SV was located in the promoter region of a TCP-like gene and contained a cis-acting element involved in GA response (Fig. 5b). It has been previously reported that genes from the TCP transcription factor (TF) family are involved in regulating internode length and branching[31,32]. Phylogenetic analysis indicated that this TCP-like gene was highly homologous to Arabidopsis TCP11 (Fig. 5c), so we named it MdTCP11. Subcellular localization analysis revealed that MdTCP11 was localized in the nucleus of N. benthamiana leaves (Supplementary Fig. 14).

Using time-series expression data from the spur-type variety 'Liquan Spur' and the standard-type variety 'Yanfu No. 8', we found that MdTCP11 was differentially expressed between these two varieties at 103, 112, and 134 days after germination (Fig. 5d), consistent with the observed difference in plant height over time, indicating a potential

role of MdTCP11 in the development of spur-type morphology (Fig. 5e). RNA-Seq data further confirmed the differential expression of MdTCP11 between the spur-type and the standard-type varieties (Fig. 5f). Interestingly, other spur-type varieties not in the spur-type clade also showed higher expression of MdTCP11 compared to standard-type varieties (Fig. 5f), despite the absence of the 167-bp deletion in the promoter of MdTCP11 in these spur-type varieties. The differential expression of MdTCP11 was verified by qRT-PCR (Supplementary Fig. 15). We speculate that the upregulation of MdTCP11 in spur-type varieties may contribute to the formation of spur-type morphology.

To investigate the methylation status of the MITE, we conducted a genome-wide analysis of cytosine methylation in shoot tips of 'Liquan spur' and 'Yanfu No. 8' and site-specific bisulfite sequencing (BS-seq) of leaves. We found that the average DNA methylation level of the MITE in both shoot tips and leaves of the standard-type varieties was higher than in the spur-type varieties (Fig. 6b and Supplementary Fig. 16). To further determine the correlation between the MITE methylation and the expression level of MdTCP11 with gibberellin, we treated GL-3 plants with the DNA methylation inhibitor 5-AZA under control and exogenous gibberellin conditions (Fig. 6c). After treatment with 5-AZA, the expression of MdTCP11 was significantly increased, confirming that methylation of the MITE represses MdTCP11 expression. Additionally,
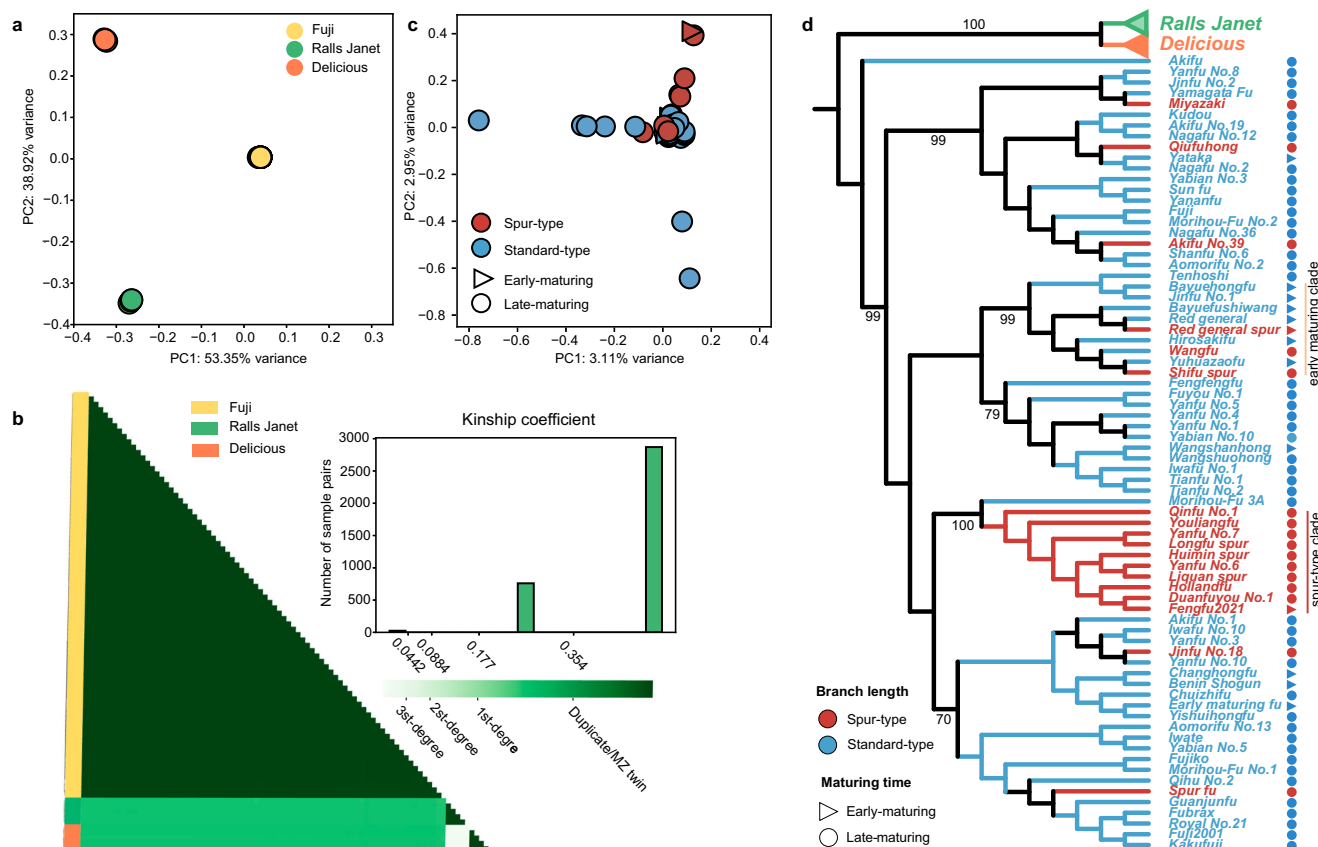
Fig. 3 | **Population structure of 'Fuji' varieties. a** PCA of 74 'Fuji' clonal varieties and five individuals from each of the two parents, 'Ralls Janet' and 'Delicious', using all identified SNPs. **b** Kinship analysis of 'Fuji' clonal varieties along with 'Ralls Janet' and 'Delicious' individuals. The histogram in the upper right corner shows the frequency distribution of the coefficient of relatedness among all individuals. **c** PCA of 'Fuji' varieties based on somatic SNPs. **d** Phylogenetic relationships among 'Ralls Janet', 'Delicious', and the 74 'Fuji' varieties based on somatic SNPs. Source data are provided as a Source Data file.

under exogenous gibberellin treatment, the expression of *MdTCP11* was increased by 5-AZA. These results support the notion that MITE methylation is negatively correlated with *MdTCP11* expression under gibberellin treatment (Fig. 6c).

To further verify that the 167-bp deletion responds to GA treatment, we analyzed the promoter activity. The constructed vectors were transformed into *Agrobacterium* EHA105 and immediately transformed into apple calli. The infected apple calli were incubated on the GA₃-containing medium, after which LUC (luciferase) activities and GUS (β-glucuronidase) staining were assessed (Fig. 6d and Supplementary Fig. 17). The constitutive 35S promoter, *MdTCP11-ΔMITE-pro*, and *MdTCP11-pro* exhibited vigorous GUS activities. Following GA₃ treatment, GUS and LUC activities in *MdTCP11-ΔMITE-pro* and *MdTCP11-pro* were significantly inhibited. Furthermore, GUS and LUC activities in *MdTCP11-ΔMITE-pro* were higher than those in *MdTCP11-pro*. These results suggest that deletion of the MITE sequence may enhance the *MdTCP11* promoter activity (Fig. 6d and Supplementary Fig. 17), aligning with the heterozygous MITE deletion and increased *MdTCP11* expression observed in spur-type varieties (Figs. 6a, 5d). Additionally, GA₃ was found to inhibit the expression of *MdTCP11* (Fig. 6c, d and Supplementary Fig. 17), consistent with the lower GA₃ concentration and higher *MdTCP11* expression level in spur-type varieties (Figs. 4c, 5d). To characterize the function of *MdTCP11*, we overexpressed *MdTCP11* in both apple and *Nicotiana tabacum* (Supplementary Fig. 18). Overexpression of *MdTCP11* resulted in reduced growth in both apple and *N. tabacum*, and *MdTCP11* overexpressing transgenic apple and *N. tabacum* lines displayed shorter plant height and average internode length compared to their corresponding wild types (Fig. 6e–j).

Discussion

Somatic variations are prevalent in organisms and manifest across all stages of growth and development. Their generation and accumulation have long been recognized as important factors in aging and the occurrence of tumors in humans. Therefore, studying the occurrence and development of somatic variations is of great significance. In plants, somatic variations are widely used in bud selection for fruit tree breeding[2]; however, the mechanisms underlying somatic variations remain largely unexplored.

Reference genomes serve as the foundation for detecting genetic variants and are important resources for studying the genetic mechanisms underlying trait formation. Currently, common variant detection pipelines are based on haploid genome assemblies, which introduce bias towards the reference genomes during read mapping and interfere with subsequent analyses[33]. This makes it challenging to accurately identify causal variants, particularly in species with high heterozygosity such as apple. With the aid of long-read and Hi-C data, several high-quality phased apple genomes have been assembled recently, providing important resources for apple genetics[19-24]. In this study, we generated HiFi and ultra-long ONT reads, as well as Hi-C data for multiple 'Fuji' clonal varieties. Combined with short read data from the two parents of 'Fuji', we assembled a high-quality fully phased 'Fuji' genome using a trio-based approach, with individual haplotypes successfully assigned to the two parents. Trio-based approach combines the high accuracy of trio-based phasing with the capability of long reads to bridge complex genomic regions, thereby offering unparalleled efficacy in segregating a genome into its paternal and maternal haplotypes[34]. This contrasts with the recently published 'Fuji' genomes, in which the parental origins of individual
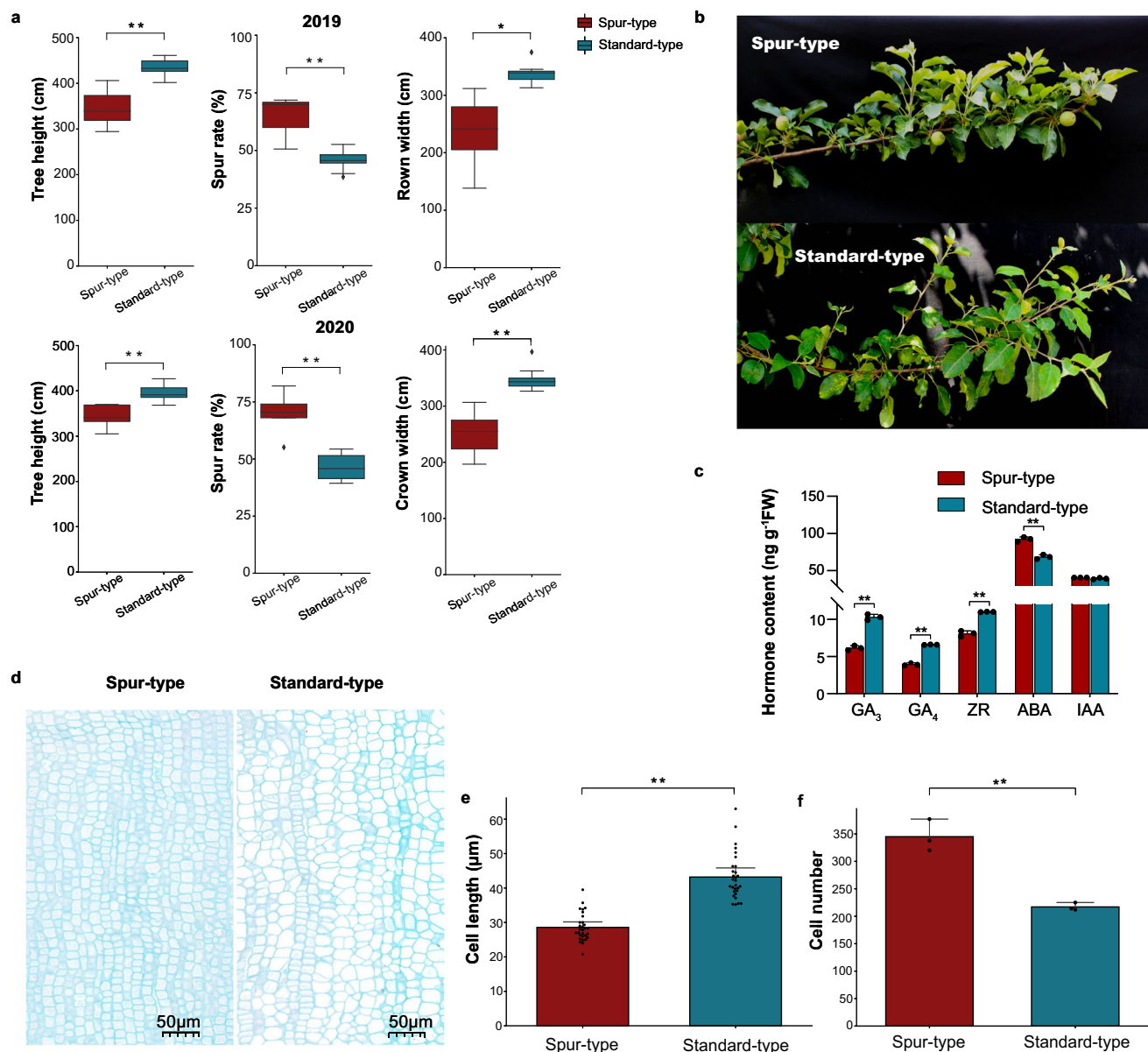
Fig. 4 | Physiological characteristics of standard-type and spur-type 'Fuji' varieties. a Spur rate, crown width, and trunk height of standard-type varieties ($n=10$) and spur-type 'Fuji' varieties (n = 5) listed in Supplementary Data 14. In each boxplot, the center line represents the median, the lower and upper bounds indicate the first and third quartiles, respectively, whiskers extend to 1.5 times the interquartile range, and points indicate outliers. **b** Branches of standard-type 'Yanfu No. 8' (Y8) and spur-type 'Liquan spur' (LQD). **c** Contents of GA$_3$, GA$_4$, trans-zeatin-riboside (ZR), abscisic acid (ABA), and indole-3-acetic acid (IAA) in Y8 and LQD. Data are means ± SE (n = 3 biological replicates). **d** Anatomical observations of a longitudinal section of the LQD and Y8 stems. The experiment was repeated three times independently with similar results. **e** Cell length in the stem longitudinal sections of Y8 and LQD. Values represent mean ± SE (n = 30 individual cells). **f** Cell number in the stem longitudinal sections of Y8 and LQD. Data are means ± SE (n = 3 measurements). * and ** indicate significant differences at $P < 0.05$ and 0.01, respectively (two-sided Student's t test). Source data are provided as a Source Data file.

haplotypes remain unresolved[22,23] (Supplementary Note 1 and Supplementary Figs. 5–7).

In plants, somatic variations can be passed on to the next generation. It has been shown that the proportion of somatic mutations passed on to the next generation in perennials (approximately 50%) is much higher than in annual plants (approximately 10%). Therefore, somatic variations may have a greater contribution to the genetic diversity of perennials[28]. Due to the high heterozygosity of the 'Fuji' genome, the number of inherent germline variants present between the two haplotypes is much higher than the number of somatic variants, which increases the likelihood of germline variants being misidentified as somatic variants. In addition, SVs between the two haplotypes can complicate short-read alignments, leading to the

identification of false genetic variants. To systematically identify somatic variants in the 'Fuji' bud-sport varieties and study their functional impacts, we developed a pipeline based on the phased diploid genome, inspired by the reference flow approach[35]. The pipeline aligns reads to the two haploid genomes and assigns them to the appropriate haplotypes based on alignment quality. Compared to directly aligning reads to a single or diploid genome, this method yields fewer mismatches, higher alignment scores, and overall higher alignment quality (Supplementary Figs. 19–21). This approach mitigates most false positives caused by heterozygous germline variants inherited from the common ancestor of 'Fuji' varieties, allowing for the detection of somatic variants with greater confidence. Furthermore, we jointly used long-read data from 13 varieties and short-read data from 74 varieties
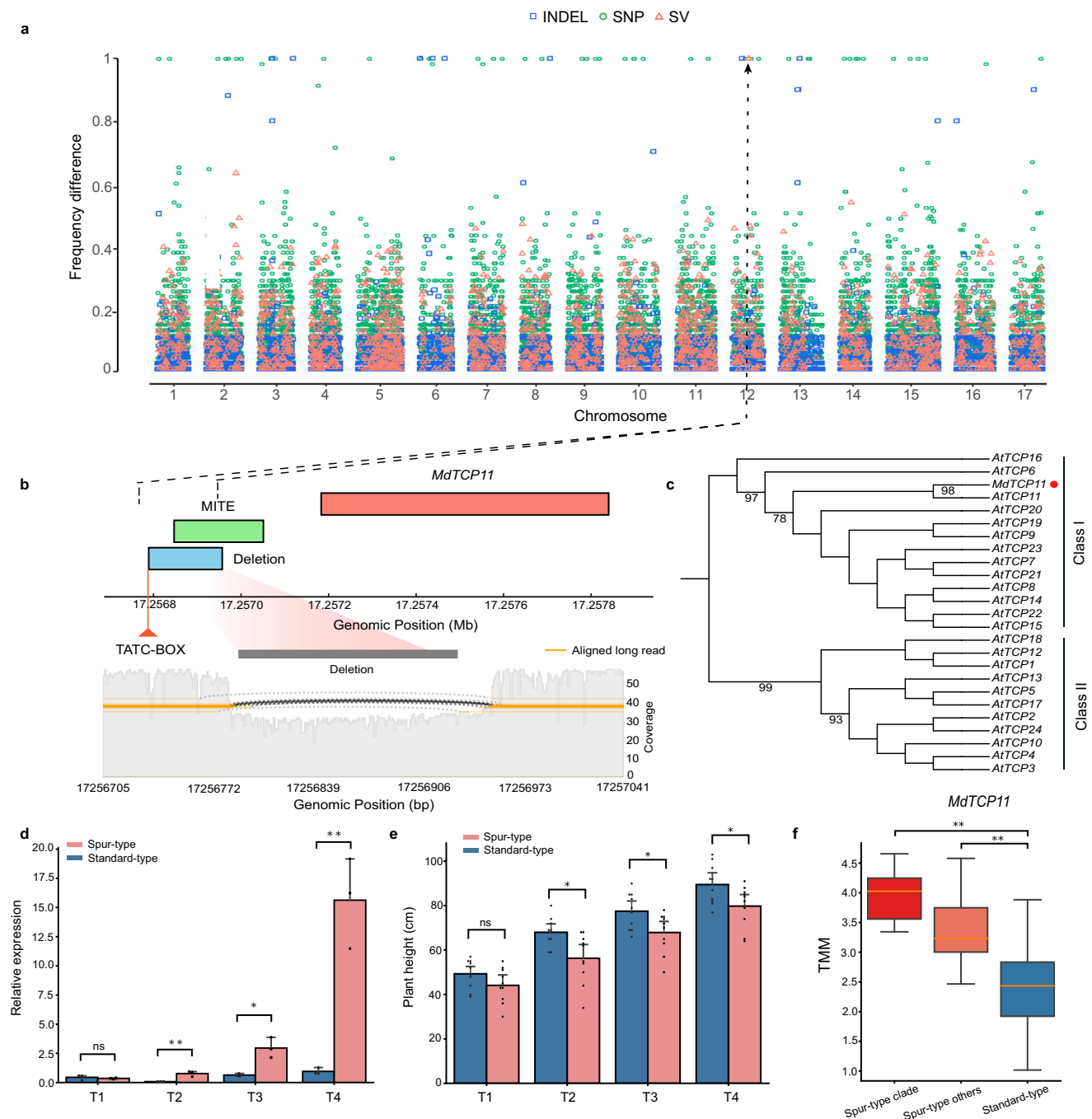
Fig. 5 | A 167-bp deletion and the expression of *MdTCP11* associated with spur-type growth habit. a Frequency difference of somatic variants between spur-type varieties in the spur-type clade and standard-type varieties. The y-axis represents the difference in frequency of samples with somatic mutations between the two groups. **b** Genomic positions of the 167-bp spur-type clade-specific deletion, MITE, and *MdTCP11*. The 167-bp deletion was located in the promoter of *MdTCP11* and contained a TATC-BOX. The read depth and sequence alignments around this deletion are shown below. **c** Phylogenetic analysis of MdTCP11 and its homologs in Arabidopsis. Protein sequences of the TCP domain were used for the phylogenetic analysis with the maximum likelihood method. **d** Expression of *MdTCP11* in shoot tips at different developmental stages of spur-type 'Liquan spur' and standard-type 'Yanfu No. 8'. T1, T2, T3, and T4 correspond to 90, 103, 112, and 134 days after germination, respectively. Data are means ± SE (n = 3 biological replicates). **e** Plant height of 'Liquan spur' and 'Yanfu No. 8' at different stages. Data are mean ± SE (n = 10 individual plants). **f** Expression of *MdTCP11* in shoot tips of mature spur-type varieties in the spur-type clade (n = 9 individual plants), other spur-type varieties (n = 8 individual plants), and standard-type 'Fuji' varieties (n = 59 individual plants) based on RNA-Seq data. * and ** indicate significant differences at *P* < 0.05 and 0.01, respectively (two-sided Student's t test). ns, not significant. Source data are provided as a Source Data file.

to construct an SV dataset. This accurate and comprehensive collection of somatic variants provides an important foundation for identifying causal variants underlying interesting traits.

The bud sports of plants are usually caused by genetic and epigenetic variations. Unlike epigenetic variations, genetic variations are

stable and independent of environmental influences. In this study, we reconstructed the evolutionary relationships of 'Fuji' bud-sport varieties using high-confidence somatic SNPs. The results indicated that two important agronomic traits, spur-type growth habit and early maturation, originated from multiple independent events in different
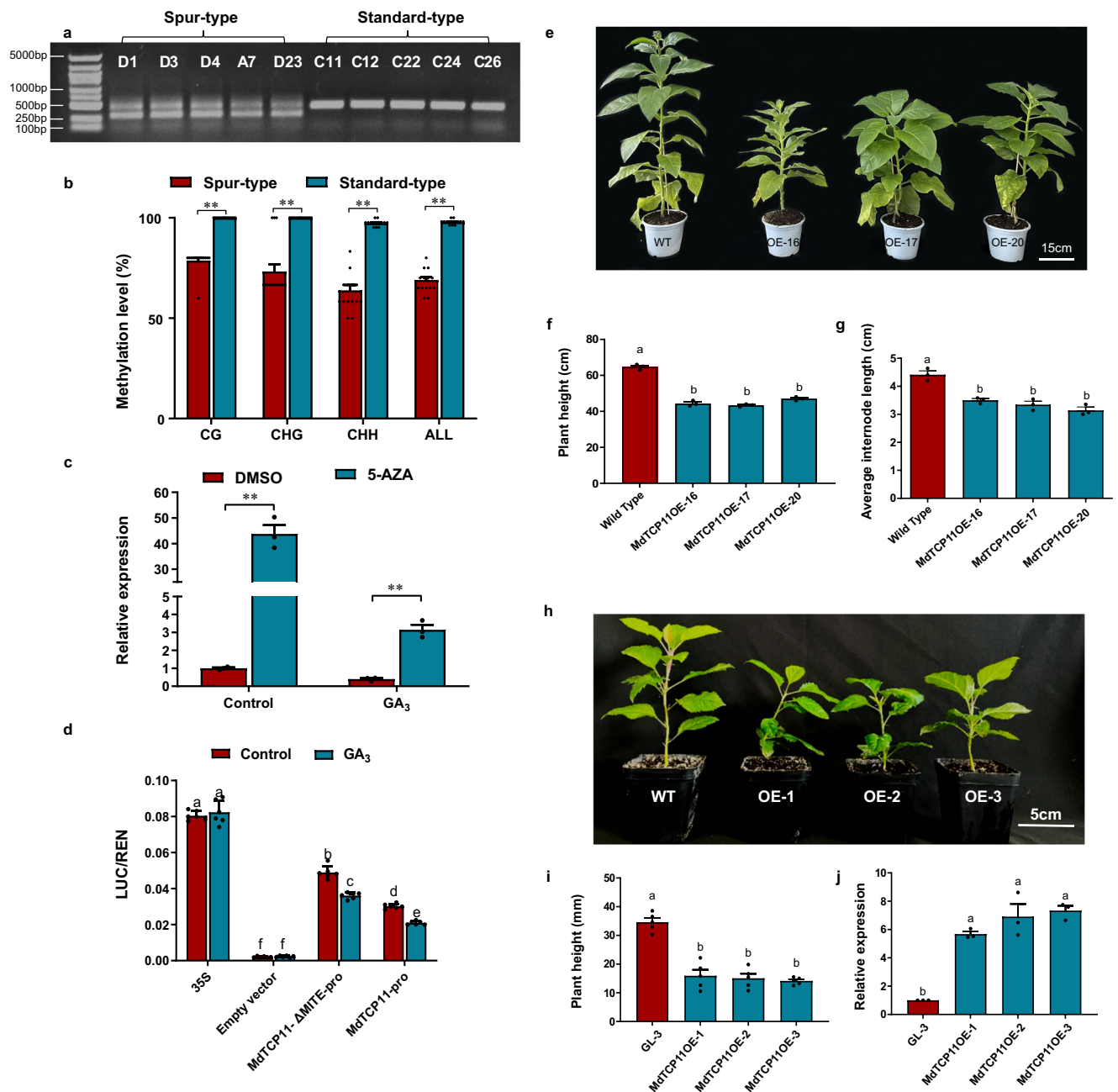
Fig. 6 | Promoter activity analysis and functional characterization of *MdTCP11*.
a PCR analysis of the 167-bp deletion in the *MdTCP11* promoter in different 'Fuji' clonal varieties. Detailed information about the varieties is provided in Supplementary Data 4. PCR amplification and electrophoresis were repeated three times independently with similar results. **b** Methylation levels of the MITE in standard-type and spur-type apple varieties. Data are presented as mean ± SE (n = 15 biological replicates). **c** Expression analysis of *MdTCP11* treated with 5-AZA under control and exogenous gibberellin conditions. DMSO, the solvent for 5-AZA, was used as the control. Data are presented as mean ± SE (n = 3 biological replicates). * and ** indicate significant differences at $P < 0.05$ and 0.01, respectively (two-sided Student's *t*-tests). **d** LUC activity of the *MdTCP11* promoter treated with GA₃. LUC-TCP11-ΔMITE-pro represents the promoter of the spur-type variety, while LUC-TCP11-pro represents the promoter of the standard-type variety. Data are presented as mean ± SE (n = 6 biological replicates). **e** Morphology of transgenic *N. tabacum* lines overexpressing *MdTCP11* and the wild type (WT). **f, g** Plant height (**f**) and average internode length (**g**) of transgenic *N. tabacum* lines overexpressing *MdTCP11*. Values are means ± SE (n = 3 biological replicates). **h** Morphology of transgenic apple lines overexpressing *MdTCP11* and the wild-type 'GL-3'. **i** Plant height of transgenic apple lines overexpressing *MdTCP11* and the wild-type 'GL-3'. Data are presented as mean ± SE (n = 5 biological replicates). **j** Relative expression of transgenic apple lines overexpressing *MdTCP11* and the wild-type 'GL-3'. Data are presented as mean ± SE (n = 3 biological replicates). Different letters in (**d, f, g, i, j**) indicate significant differences at $P < 0.01$ (one-way ANOVA with Tukey's post-hoc comparisons). Source data are provided as a Source Data file.

'Fuji' bud-sport varieties. Interestingly, each trait emerged from one major event that affected multiple varieties. This finding contrasts with the result of cluster analysis using epigenetic variants, where spur-type and early-maturing varieties were dispersed[36]. This suggests two different strategies employed in breeding spur-type and early-maturing bud-sport varieties: (1) further selection from existing spur-type or early-maturing varieties to leverage existing genetic variations, and (2) reliance on newly generated somatic variations from standard-type varieties to acquire spur-type or early-maturing traits.

Spur-type apples represent a valuable germplasm resource characterized by dwarfing, compactness, ease of management, early fruiting, and high yield. They are important for dwarf and dense apple

cultivation. In apple, the 'double dwarf' cultivation method, which combines spur-type apple varieties with dwarf rootstocks, has emerged as a key strategy for intensive dwarf and dense cultivation. Spur-type bud sports achieve a dwarf tree architecture by shortening the duration of shoot vegetative growth. This is different from dwarfing rootstocks, which reduce the proportion of lateral buds that develop into long shoots during the early stages of tree development[37,38]. Compared to standard-type varieties, spur-type varieties exhibited reduced annual extension growth, a greater tendency to produce spurs rather than shoots, and a smaller trunk cross-sectional area[39,40]. Consistent with these features, we observed significant differences in spur rate and crown width between standard-type and spur-type bud-sport varieties (Supplementary Fig. 12). Furthermore, the contents of GA_3 and GA_4 were lower in spur-type bud-sport varieties compared to standard-type varieties (Fig. 4c). It has been reported that exogenous gibberellin application promotes shoot growth and increases internode length, while gibberellin biosynthesis inhibitors can inhibit internode growth[15], indicating a positive correlation between gibberellin and internode length of apple branches.

The formation of the spur-type trait is complicated, and spur-type varieties with different genetic origins are likely caused by different variants. In this study, we detected a number of genetic variants specific to spur-type varieties that clustered together on the phylogenetic tree (Supplementary Data 10). These somatic variants may have contributed to the development of this important trait. Among these somatic variants, we noticed one SV located in the promoter of a TCP TF family gene, *MdTCP11*. The TCP TF family is divided into two classes: class I and class II. Class I TCP TFs generally promote cell growth and proliferation, while class II TCP TFs typically repress these processes[32,41–45]. *MdTCP11* belongs to class I. Interestingly, our research demonstrated that *MdTCP11* inhibits apple growth, behaving similarly to a class II TCP TF[46]. However, this finding aligns with a previous report on *GrTCP11* from cotton, which, when expressed in *Arabidopsis thaliana*, inhibits root hair elongation[47]. Hormones play crucial roles in plant growth. We found that GA_3 inhibited the promoter activity of *MdTCP11* (Fig. 6d and Supplementary Fig. 17a). In contrast, overexpression of *MdTCP11* reduced GA_3 content in both tobacco and apple (Supplementary Fig. 18). These findings suggest the existence of a potential negative feedback mechanism involving GA_3 in the regulation of apple internode development. Additionally, ABA is known to inhibit bud growth[48,49]. In our study, spur-type varieties exhibited higher levels of both GA_3 and ABA compared to standard-type varieties (Fig. 4c). We also observed that ABA- and GA-related genes were differentially expressed between spur-type and standard-type varieties (Supplementary Fig. 22). Moreover, we found that ABA treatment significantly enhanced the promoter activity of *MdTCP11* (Supplementary Fig. 17b). Together, these results suggest potential crosstalk between ABA and GA in regulating *MdTCP11* expression and the spur-type phenotype.

In this study, our transcriptome profiling data revealed that *MdTCP11* was expressed at a higher level in spur-type varieties compared to standard-type varieties. Analysis of promoter activity and methylation levels indicated that the expression of *MdTCP11* is negatively regulated by DNA methylation of the MITE in its promoter, and that the 167-bp deletion involving the MITE in the promoter could results in an increased expression level of *MdTCP11*. Notably, although the deletion was specific to the 10 spur-type varieties in the spur-type clade, we also observed that the expression of *MdTCP11* was higher in other spur-type varieties, dispersed across different clades in the phylogenetic tree, compared to standard-type varieties (Fig. 5f). This suggests that altered *MdTCP11* expression might also be involved in the emergence of other spur-type varieties, indicating the presence of different causal variants regulating *MdTCP11* expression in these spur-type varieties. The precise molecular mechanisms underlying the spur-type phenotype in these varieties will require further investigation.

Methods
Plant materials and genome sequencing
A total of 84 apple samples, including 74 'Fuji' clonal varieties, five 'Ralls Janet' varieties, and five 'Delicious' varieties (Supplementary Data 4), were collected at the Apple Demonstration Nursery of Yangling Modern Agriculture Technology Park, Northwest Agriculture & Forestry University, Shaanxi, China (34°52′N, 108°7′E). All the 'Fuji' varieties were grafted onto the M.26 rootstock in 2016. Genomic DNA was extracted from young leaves using the phenol-chloroform method. DNA libraries were constructed and sequenced on the Illumina NovoSeq platform.

PacBio SMRT libraries were prepared following the standard protocol provided by Pacific Biosciences (CA, USA) and sequenced on the PacBio Sequel II platform. Ultra-long Nanopore libraries were constructed with the SQK-LSK109 Ligation Sequencing 1D kit (Oxford Nanopore Technologies, UK), followed by size-selection (>50 kb) using the SageHLS HMW system (Sage Science, USA), as per the manufacturer's guidelines. The resulting libraries were sequenced on the PromethION platform. Genomic DNA of 'Nagafu No. 2' was used to construct Hi-C libraries, which were sequenced using the MGISEQ T7 platform.

Genome assembly and annotation
The 'Fuji' genome was assembled using PacBio HiFi reads, with the aid of ONT ultra-long reads, parental short reads, and Hi-C data. First, hifiasm[50] (v0.19.5) was applied independently to HiFi reads from four 'Fuji' clonal varieties. During the assembly process, the parental short reads were integrated into the trio binning stage to generate haplotype-resolved contigs. ONT ultra-long reads were then integrated to improve the continuity of the assembly, producing a telomere-to-telomere level assembly. Next, Hi-C contact maps of the two haploid assemblies were generated using the 3D-DNA pipeline and Juicer tools[51] with default parameters, which were used to facilitate manual checking and correction of potential misassemblies. Collinearity analysis between the 'Fuji' genome and the GDDH13 genome was performed using SyRI[52] with default parameters and visualized using plotsr[53]. The 20-mers were extracted from Illumina and HiFi reads of 'Fuji' and its two parents ('Delicious' and 'Ralls Janet') using Meryl (https://github.com/marbl/meryl). Subsequently, QV scores and hamming error rates were calculated using Merqury[54].

EDTA[55] and RepeatModeler (http://www.repeatmasker.org/RepeatModeler/) were used to annotate repeat sequences in the assembly. First, EDTA was employed to identify LTR retrotransposons and DNA elements, and RepeatModeler was then used to predict de novo TE sequences from the assembly. Finally, the two libraries were merged and imported into RepeatMasker (http://www.repeatmasker.org/RepeatMasker/) to scan the 'Fuji' genome assembly for repetitive elements.

Protein-coding genes were predicted from the 'Fuji' genome using the BRAKER3 pipeline[56], which integrates RNA-Seq and protein homology information. RNA-Seq reads were aligned to the soft-masked 'Fuji' genome using HISAT2[57] (v2.2.1). Protein sequences from *Malus domestica, Malus sieversii, Malus sylvestris, Pyrus communis*, and *Prunus persica* were downloaded from GDR (https://www.rosaceae.org/) and aligned to the 'Fuji' genome as protein homology evidence. As a complement to de novo gene predictions, we also mapped the gene annotations of GDDH13 onto the 'Fuji' genome using Liftoff[58] (v1.6.3).

SNP and small indel calling
Illumina paired-end reads were filtered and trimmed using fastp[59] v0.22.0. To detect germline variants, the cleaned reads were mapped to one haploid genome of 'Fuji' (Fuji_Ral) using BWA-MEM[60] with default parameters. Variants (SNPs and small indels) were called using the HaplotypeCaller module of GATK[61]. The resulting raw variants were

filtered using GATK with parameters 'QD > 2.0||MQ > 40.0||FS > 60.0 || SOR > 3.0 || MQRankSum > -12.5 || ReadPosRankSum > -8.0'. Heterozygous alleles shared by more than 80% of 'Fuji' varieties were retained for the germline variant set.

To detect somatic variants, the cleaned reads were aligned to the two haploid 'Fuji' genomes (Fuji_Ral and Fuji_Del), respectively, using BWA-MEM, resulting in a pair of BAM files for each sample. Next, alignments were filtered by comparing the alignment quality of the same pair of reads against the two haploid genomes. Specifically, if a pair of reads exhibited a higher alignment score and a lower edit distance to Fuji_Ral compared to Fuji_Del, this pair of reads were excluded from the BAM file for Fuji_Del, and vice versa. This process generated a pair of cleaned BAM files for each sample, in which most reads were appropriately assigned to one haploid genome. Alignments with same qualities against both haploid genomes remained in both BAM files. We then used GATK HaplotypeCaller to detect variants from the cleaned BAM files with default parameters. Raw variants were filtered using GATK, employing the same parameters as for filtering raw germline variants, and variants shared by more than 80% of 'Fuji' varieties were excluded to further reduce false positives. Since a subset of reads were retained in both BAM files, some genetic variants were reported in both haploid genomes. We further employed CrossMap[62] to merge and eliminate redundancies in genetic variants detected in both haplomes, resulting in the final somatic variant collection. All variants were annotated using SnpEff[63], and deleterious variants were identified using the SIFT4G protocol[64].

Structural variant calling

To identify high-confidence SVs, we employed a strategy combining long-read and short-read sequencing data. First, HiFi and Nanopore reads were aligned to the 'Fuji' genome using minimap2[65]. SVs were then detected and accurate SV boundaries were generated using Sniffles[66] (v2.2). Next, we employed GraphTyper[67] (v2.7.5) to genotype these SVs in the 74 'Fuji' clonal varieties using Illumina short-read data. To identify potential hemizygous regions in the genome, we first compared Fuji_Ral and Fuji_Del to detect insertions and deletions between these two haploid genomes using svim-asm[68] (v1.0.3) and Minigraph[69] (v0.19). These SVs between the two haploid genomes were then verified in the 74 clonal varieties using the results from GraphTyper. The reliable sequences, identified in more than 80% of samples, were preserved as the final diploid genome hemizygous sequence collection.

Population genetic analyses

To infer the genetic relationships among the collected samples, we calculated the pairwise robust kinship estimator using the KING software[70] with all SNPs. Principal component analysis (PCA) was performed using EIGENSOFT[71]. A phylogenetic tree was constructed using iqtree[72] (v2.0.3) with the 'MFP + ASC' model[73] and 1,000 bootstraps, and then visualized using iTOL[74].

Identification of clade-specific variants

The 'Fuji' population was grouped based on their phenotypes and phylogenetic relationships (Fig. 3d). The frequency difference of each somatic variant between standard-type and spur-type varieties within the spur-type clade, as well as between late-maturing and early-maturing varieties within the early-maturing clade, was calculated. Variants specific to spur-type or early-maturing varieties were identified and then annotated using SnpEff[63] (v5.0).

Anatomical observations and morphological measurements

Stems of 'Yanfu No. 8' and 'Liquan spur' were fixed in formaldehyde−acetic acid−alcohol solution. The samples were dehydrated and embedded in paraffin. Sections were cut with a microtome (Leica DM2000, Germany) and stained with 0.1% toluidine blue (TB). Cell length and cell number were measured in stem longitudinal sections, with three biological replicates from each variety. All spur-type varieties were grafted onto the M.26 rootstock, and morphological measurements were carried out at the end of the growing season. The height and crown width of three trees from each variety were measured with a meter ruler and caliper. Additionally, the number of 1-year-old spurs (<5 cm) graving on 2-year-old branches was counted, and the spur rate was calculated as the number of spur shoots divided by the total number of shoots. 'Fuji' varieties used for measuring spur rate, crown width, and trunk height are provided in Supplementary Data 14. Three randomly selected main branches from each tree were used to count the spur rate. 'Liquan spur' and 'Yanfu No. 8' were used to represent spur-type and standard-type varieties, respectively. Three trees from each of these two varieties were used to measure the length and internode length of 1-year-old shoots on three randomly selected main branches of each tree. The average internode length was calculated by dividing the total shoot length by the number of internodes.

Hormone extraction and measurement

Shoot tips (5–8 mm below the apical region, including the shoot meristem without any leaves or petioles) from 'Yanfu No. 8' and 'Liquan spur' were collected at 60 days after flowering. Approximately 0.2 g of frozen shoot tip samples were ground into powder and used to measure GA_3, GA_4, ZR, ABA, and IAA contents. Each tube of sample was filled with 1 mL of extraction buffer (ethyl acetate), shaken at 4 °C for 30 min, and then centrifuged at $12,000 \times g$ for 5 min. The supernatant was transferred to a new centrifuge tube and dried with nitrogen gas. The dried samples were dissolved in 150 μL of methanol and filtered through a 0.22-μm membrane[75,76]. Three biological replicates were used for each sample. Detection and analysis of the hormones were performed with high-performance liquid chromatography (Waters 2498/UV; Visible Detector, Shaanxi, China). External standards for GA_3 (Sigma, 48880, USA), GA_4 (Sigma, G7276, USA), ZR (Sigma, Z0375, USA), ABA (Sigma, 5.30339, USA), and IAA (Sigma, I3750, USA) were used for quantitative analyses.

Subcellular localization of MdTCP11

The CDS of *MdTCP11* was inserted into the pC2300-GFP vector, which was then introduced into *Agrobacterium* strain GV3101 (WEIDI, AC1001, China), and subsequently transformed into *N. benthamiana*. After 3 days of culture in the dark, GFP fluorescence was detected using a confocal laser-scanning microscope with excitation at 488 nm (Zeiss LSM 510 Meta, Jena, Germany).

GUS staining

The promoter of *MdTCP11* (*MdTCP11-pro*) and the *MdTCP1-ΔMITE-pro* sequence were inserted into the *pCAMBIA1381-GUS* vector, and the 35S:GUS and the empty *pCAMBIA1381-GUS* vector were used as positive and negative controls, respectively. The resulting constructs were transferred into *Agrobacterium tumefaciens* strain EHA105 (WEIDI, AC1010, China) using *Agrobacterium*-mediated transformation to obtain transient transgenic apple calli. GUS staining was performed after 3 days of dark culture using 5-bromo-4-chloro-3-indolyl-β-glucuronide (X-gluc) as a substrate. Apple calli (*Malus domestica* cv. 'Orin') used for injection were grown on MS medium (M519-100L, Phyto-Technology Laboratories, USA) containing 0.4 mg/L 6-BA (A600743-0025, Sangon Biotech, China) and 1.5 mg/L 2,4-D (A425360-0001, Sangon Biotech, China) at 25 °C under dark conditions and were subcultured every 2 weeks.

Dual-luciferase assays

The *MdTCP11-pro* (2000 bp) and *MdTCP11-ΔMITE-pro* sequences were inserted into the pGreen II 0800-LUC vector, with the 35S promoter serving as a positive control and the empty pGreen II 0800-LUC vector

as a negative control. The constructs were transformed into *Agrobacterium* strain EHA105 and then introduced into apple calli. After 3 days of culture in the dark, LUC and REN activities were quantified using a dual-luciferase reporter assay system (Promega, E1910) with Double-Luciferase Reporter Assay Kit (*TransDetect*, FR201-01-V2, China). At least ten biological replicates were performed for each transformation, and the ratios of LUC to REN were calculated for treatments and controls to assess the promoter activity of *MdTCP11*.

Transcriptome and comparative genomic analysis of *MdTCP11*
Shoot tips from mature clonally propagated 'Fuji' varieties were collected from for RNA-Seq analysis. Total RNA was isolated using the Tiangen Total RNA Extraction Kit (DP441, Beijing, China), followed by mRNA enrichment using magnetic beads conjugated with oligo (dT). RNA-Seq libraries were then constructed using the NEBNext Ultra RNA Library Prep Kit and sequenced on the Illumina HiSeq 2500 platform. Raw RNA-Seq reads were filtered using fastp[59] (v0.22.0) with default parameters. The resulting cleaned reads were mapped to the Fuji_Ral genome using STAR[77] (v2.7.4). Following mapping, raw counts for each gene were derived using featureCounts[78] (v2.0.3) and normalized to TMM (Trimmed Mean of M-values) using the R package EdgeR[79] (v3.36.0).

qRT-PCR analysis was performed using the CFX Connect Real-Time PCR Detection System (USA) and 2× SYBR Green Pro Taq HS Premix II (Accurate Biotechnology, Hunan, China), with three biological and three technical replicates. The *ACTIN* gene was used as the reference gene. Primers used for qRT-PCR are listed in Supplementary Data 15. Relative transcript levels were calculated using the $2^{-\Delta\Delta Ct}$ method[80].

To investigate the presence or absence of the MITE in the promoter of *MdTCP11* across different apple species, genome assemblies and annotation files in gff3 format were retrieved for 13 *Malus* accessions from NCBI (accession numbers PRJNA869488, PRJNA927238, and PRJNA591623). RepeatMasker (v4.1.0) was used to annotate repeat sequences in these 13 apple genomes. Orthologous gene pairs of *MdTCP11* were identified using MCScan[81] with default parameters. Promoter sequences were defined as the 2-kb region upstream of the start codon. Finally, gene structure annotation and repeat annotation were integrated to determine whether the MITE was present in the promoter of *MdTCP11* across these 13 apple genomes.

Bisulfite sequencing library construction and data analysis
Genomic DNA extracted from the shoot tips of 'Liquan spur' and 'Yanfu No. 8' was used to construct bisulfite sequencing (BS-Seq) libraries using a Super Plant Genomic DNA Kit (TIANGEN, DP360, China) following the manufacturer's instructions. The libraries were sequenced on the Illumina HiSeq 2500 platform. Adapters and low-quality reads were removed from the raw data using Trimmomatic[82] (v0.36) with default parameters, and the cleaned reads were then mapped to the Fuji_Ral genome using Bismark[83] (v0.22.3) with default parameters. The methylation level of each cytosine was then calculated based on the alignments.

DNA methylation assay
For locus-specific BS-seq, DNA was extracted from the leaves of 'Nagafu No.2' (standard-type) and 'Yanfu No.6' (spur-type) using a Super Plant Genomic DNA Kit (TIANGEN, DP360, China). Approximately 300 ng of DNA was treated with bisulfite using an EZ DNA Methylation-Gold Kit (Zymo Research, D5005, USA) and amplified by PCR using Takara Premix Ex Taq™ Hot Start Version (Takara Research, RR030A, Japan). The amplified products were recovered from an agarose gel using a Universal DNA Purification Kit (TIANGEN, DP214, China). The purified product was cloned into the pBM16A vector using a pBM16A Topsmart Cloning Kit (BIOMED, China). At least 15 clones were sequenced per genotype using the Sanger sequencing

technology, and sequencing results were analyzed using the Kismeth online software[84]. Treatment with 5-AZA (Sigma, A3656, USA) was conducted following a previously described protocol with minor modifications[85]. Briefly, subcultured 'GL-3' plants were transferred to Murashige & Skoog medium (MS; M519-100L, PhytoTechnology Laboratories, USA) supplemented with either 7 µg/L of 5-AZA (Sigma, A3656, USA) or DMSO (Sigma, D2650, USA) and incubated for 3 weeks under long-day conditions (14 h light, 10 h dark) at 25 °C. Subsequently, the plants were treated with 2 mg/L of GA3 for 24 h.

Genetic transformation
Tissue-cultured 'GL-3' (a progeny of the 'Royal Gala' apple) was used for apple transformation[86]. 'GL-3' and transgenic plants were subcultured every 4 weeks on MS medium (M519-100L, PhytoTechnology Laboratories, USA) containing 0.2 mg/L 6-BA (A600743-0025, Sangon Biotech, China), 0.2 mg/L IBA (A600725-0025, Sangon Biotech, China) under long-day conditions (14 h light/10 h dark) at 25 °C.

The *pC2300-MdTCP11-GFP* vector was transformed into *Agrobacterium* strain EHA105 (WEIDI, AC1010, China), which was then transformed into the leaves of apple 'GL-3' and *Nicotiana tabacum L. cv*. NC89. Transgenic apple and *N. tabacum* microshoots were maintained in vitro, with subcultures performed every 4 weeks. Both apple and *N. tabacum* were grown in a growth room under long-day conditions (16 h light/8 h dark, 25 °C). Plant height and internode length were measured in transgenic *N. tabacum* lines overexpressing *MdTCP11* and wild-type (WT) plants, and plant height was measured in transgenic apple lines overexpressing *MdTCP11* and WT plants. Five biological replicates from each line were conducted.

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability
Raw sequencing data have been deposited in China National Center for Bioinformation (CNCB) under the BioProject accession PRJCA024962. The genome assembly and annotation files are available at Figshare [https://doi.org/10.6084/m9.figshare.27317043]. Source data are provided with this paper.

References
1. Shamel, A. D. & Pomeroy, C. S. Bud mutations in horticultural crops. *J. Hered.* **27**, 487–494 (1936).
2. Deng, X. et al. Retrospection and prospect of fruit breeding for last four decades in China (in Chinese). *J. Fruit Sci.* **36**, 514–520 (2019).
3. Costes, E., Lauri, P. É. & Regnard, J. L. Analyzing fruit tree architecture: Implications for tree management and fruit production. *Hortic. Rev.* **32**, 1–61 (2006).
4. Lapins, K. O. Spur-type growth habit in 60 apple progenies. *J. Am. Soc. Hortic. Sci.* **99**, 568–572 (1974).
5. Quinlan, J. D. & Tobutt, K. R. Manipulating fruit tree structure chemically and genetically for improved performance. *HortScience* **25**, 60–64 (1990).
6. Warrington, I. J., Ferree, D. C., Schupp, J. R., Dennis, F. G. & Baugher, T. A. Strain and rootstock effects on spur characteristics and yield of 'Delicious' apple strains. *J. Amer. Soc. Hort. Sci.* **115**, 348–356 (1990).
7. Wang, L. et al. Somatic variations led to the selection of acidic and acidless orange cultivars. *Nat. Plants* **7**, 954–965 (2021).
8. Ou, C. et al. A 14 nucleotide deletion mutation in the coding region of the *PpBBX24* gene is associated with the red skin of 'Zaosu Red' pear (*Pyrus pyrifolia* White Pear Group): a deletion in the *PpBBX24* gene is associated with the red skin of pear. *Hortic. Res.* **7**, 39 (2020).

9. Islam et al. Transcriptome analysis of an apple (*Malus × domestica*) yellow fruit somatic mutation identifies a gene network module highly associated with anthocyanin and epigenetic regulation. *J. Exp. Bot.* **66**, 7359–7376 (2015).

10. Han, M. et al. Insertion of a solo LTR retrotransposon associates with spur mutations in 'Red Delicious' apple (*Malus × domestica*). *Plant Cell Rep.* **36**, 1375–1385 (2017).

11. Wu, G. A. et al. Genomics of the origin and evolution of *Citrus*. *Nature* **554**, 311–316 (2018).

12. Tian, Y. et al. Transposon insertions regulate genome-wide allele-specific expression and underpin flower colour variations in apple (*Malus* spp.). *Plant Biotechnol. J.* **20**, 1285–1297 (2022).

13. Cai, X. et al. Transposable element insertion: a hidden major source of domesticated phenotypic variation in *Brassica rapa*. *Plant Biotechnol. J.* **20**, 1298–1310 (2022).

14. Wang, T. et al. Pan-genome analysis of 13 *Malus* accessions reveals structural and sequence variations associated with fruit traits. *Nat. Commun.* **14**, 7377 (2023).

15. Song, C. et al. miRNA and degradome sequencing reveal miRNA and their target genes that may mediate shoot growth in spur type mutant "Yanfu 6". *Front Plant Sci.* **8**, 441 (2017).

16. Jiang, S. et al. Methylation of MdMYB1 locus mediated by RdDM pathway regulates anthocyanin biosynthesis in apple. *Plant Biotechnol. J.* **18**, 1736–1748 (2020).

17. Daccord, N. et al. High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nat. Genet.* **49**, 1099 (2017).

18. Zhang, L. et al. A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. *Nat. Commun.* **10**, 1494 (2019).

19. Sun, X. et al. Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. *Nat. Genet.* **52**, 1423–1432 (2020).

20. Khan, A. et al. A phased, chromosome-scale genome of 'Honeycrisp' apple (*Malus domestica*). *Gigabyte* **2022**, 1–15 (2022).

21. Su, Y. et al. Phased telomere-to-telomere reference genome and pangenome reveal an expansion of resistance genes during apple domestication. *Plant Physiol.* **195**, 2799–2814 (2024).

22. Peng, H. et al. A haplotype-resolved genome assembly of *Malus domestica* 'Red Fuji. *Sci. Data* **11**, 592 (2024).

23. Li, W. et al. Near-gapless and haplotype-resolved apple genomes provide insights into the genetic basis of rootstock-induced dwarfing. *Nat. Genet.* **56**, 505–516 (2024).

24. Zhang, H. et al. A Haplotype-resolved, chromosome-scale genome for *Malus domestica* Borkh. 'WA 38'. *G3 Genes Genomes Genet.* jkae222 (2024).

25. Ou, S., Chen, J. & Jiang, N. Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Res.* **46**, e126 (2018).

26. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).

27. Ren, Y. et al. Somatic mutation analysis in *Salix suchowensis* reveals early-segregated cell lineages. *Mol. Biol. Evol.* **38**, 5292–5308 (2021).

28. Wang, L. et al. The architecture of intra-organism mutation rate variation in plants. *PLoS Biol.* **17**, e3000191 (2019).

29. Watson, J. M. et al. Germline replications and somatic mutation accumulation are independent of vegetative life span in *Arabidopsis*. *Proc. Natl. Acad. Sci.* **113**, 12226–12231 (2016).

30. Ripetti, V., Escoute, J., Verdeil, J. L. & Costes, E. Shaping the shoot: the relative contribution of cell number and cell shape to variations in internode length between parent and hybrid apple trees. *J. Exp. Bot.* **59**, 1399–1407 (2008).

31. Koyama, T., Furutani, M., Tasaka, M. & Ohme-Takagi, M. TCP transcription factors control the morphology of shoot lateral organs via negative regulation of the expression of boundary-specific genes in *Arabidopsis*. *Plant Cell* **19**, 473–484 (2007).

32. Kieffer, M., Master, V., Waites, R. & Davies, B. TCP14 and TCP15 affect internode length and leaf shape in Arabidopsis. *Plant J.* **68**, 147–158 (2011).

33. Valiente-Mullor, C. et al. One is not enough: On the effects of reference genome for the mapping and subsequent analyses of short-reads. *PLoS Comput. Biol.* **17**, e1008678 (2021).

34. Koren, S. et al. De novo assembly of haplotype-resolved genomes with trio binning. *Nat. Biotechnol.* **36**, 1174–1182 (2018).

35. Chen, N.C., Solomon, B., Mun, T., Iyer, S. & Langmead, B. Reference flow: reducing reference bias using multiple population genomes. *Genome Biol.* **22**, 8 (2021).

36. Du, X. et al. The assessment of epigenetic diversity, differentiation, and structure in the 'Fuji' mutation line implicates roles of epigenetic modification in the occurrence of different mutant groups as well as spontaneous mutants. *PLoS ONE* **15**, e0235073 (2020).

37. Lauri, P., Maguylo, K. & Trottier, C. Architecture and size relations: an essay on the apple (*Malus × domestica*, Rosaceae) tree. *Am. J. Bot.* **93**, 357–368 (2006).

38. Seleznyova, A. N., Tustin, D. S. & Thorp, T. G. Apple dwarfing rootstocks and interstocks affect the type of growth units produced during the annual growth cycle: precocious transition to flowering affects the composition and vigour of annual shoots. *Ann. Bot.* **101**, 679–687 (2008).

39. Heinicke, D. R. High-density apple orchards: planning, training, and pruning. *US Department of Agriculture Handbook* 458 (1975).

40. Lane, W., MacDonald, R. & Ketchie, D. Evaluation of quality traits and yield of 28 strains of Delicious apple. *Fruit Var. J. USA* **50**, 167–174 (1996).

41. Li, C., Potuschak, T., Colón-Carmona, A., Gutiérrez, R. A. & Doerner, P. *Arabidopsis* TCP20 links regulation of growth and cell division control pathways. *Proc. Natl. Acad. Sci.* **102**, 12978–12983 (2005).

42. Viola, I. L., Uberti Manassero, N. G., Ripoll, R. & Gonzalez, D. H. The *Arabidopsis* class I TCP transcription factor AtTCP11 is a developmental regulator with distinct DNA-binding properties due to the presence of a threonine residue at position 15 of the TCP domain. *Biochem. J.* **435**, 143–155 (2011).

43. Davière, J.-M. et al. Class I TCP-DELLA interactions in inflorescence shoot apex determine plant height. *Curr. Biol.* **24**, 1923–1928 (2014).

44. Mao, J. et al. Cytokinin-responsive MdTCP17 interacts with MdWOX11 to repress adventitious root primordium formation in apple rootstocks. *Plant Cell* **35**, 1202–1221 (2023).

45. Li, G. et al. Molecular mechanism of MdWUS2–MdTCP12 interaction in mediating cytokinin signaling to control axillary bud outgrowth. *J. Exp. Bot.* **72**, 4822–4838 (2021).

46. Martín-Trillo, M. & Cubas, P. TCP genes: a family snapshot ten years later. *Trends Plant Sci.* **15**, 31–39 (2010).

47. Hao, J. et al. GrTCP11, a cotton TCP transcription factor, inhibits root hair elongation by down-regulating jasmonic ccid pathway in *Arabidopsis thaliana*. *Front. Plant Sci.* **12**, 769675 (2021).

48. González-Grandío, E. et al. Abscisic acid signaling is controlled by a BRANCHED1/HD-ZIP I cascade in *Arabidopsis* axillary buds. *Proc. Natl. Acad. Sci.* **114**, E245–E254 (2017).

49. Yao, C. & Finlayson, S. A. Abscisic acid is a general negative regulator of arabidopsis axillary bud growth. *Plant Physiol.* **169**, 611–626 (2015).

50. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175 (2021).

51. Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).

52. Goel, M., Sun, H., Jiao, W.-B. & Schneeberger, K. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol.* **20**, 277 (2019).

53. Goel, M. & Schneeberger, K. plotsr: visualizing structural similarities and rearrangements between multiple genomes. *Bioinformatics* **38**, 2922–2926 (2022).

54. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 245 (2020).

55. Ou, S. et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* **20**, 275 (2019).

56. Gabriel, L. et al. BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS and TSEBRA. *Genome Res.* **34**, 769–777 (2024).

57. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).

58. Shumate, A. & Salzberg, S. L. Liftoff: accurate mapping of gene annotations. *Bioinformatics* **37**, 1639–1643 (2021).

59. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

60. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* https://doi.org/10.48550/arXiv.1303.3997 (2013).

61. McKenna, A. et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

62. Zhao, H. et al. CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* **30**, 1006–1007 (2014).

63. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **6**, 80–92 (2012).

64. Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M. & Ng, P. C. SIFT missense predictions for genomes. *Nat. Protoc.* **11**, 1–9 (2016).

65. Li, H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* **37**, 4572–4574 (2021).

66. Sedlazeck, F. J. et al. Accurate detection of complex structural variations using single-molecule sequencing. *Nat. Methods* **15**, 461–468 (2018).

67. Eggertsson, H. P. et al. GraphTyper2 enables population-scale genotyping of structural variation using pangenome graphs. *Nat. Commun.* **10**, 5402 (2019).

68. Heller, D. & Vingron, M. SVIM-asm: structural variant detection from haploid and diploid genome assemblies. *Bioinformatics* **36**, 5519–5521 (2021).

69. Li, H., Feng, X. & Chu, C. The design and construction of reference pangenome graphs with minigraph. *Genome Biol.* **21**, 265 (2020).

70. Manichaikul, A. et al. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).

71. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).

72. Minh, B. Q. et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).

73. Schrempf, D., Minh, B. Q., Von Haeseler, A. & Kosiol, C. Polymorphism-aware species trees with advanced mutation models, bootstrap, and rate heterogeneity. *Mol. Biol. Evol.* **36**, 1294–1301 (2019).

74. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).

75. Ivanov Dobrev, P. & Kamínek, M. Fast and efficient separation of cytokinins from auxin and abscisic acid and their purification using mixed-mode solid-phase extraction. *J. Chromatogr. A* **950**, 21–29 (2002).

76. Djilianov, D. L. et al. Dynamics of endogenous phytohormones during desiccation and recovery of the resurrection plant species haberlea rhodopensis. *J. Plant Growth Regul.* **32**, 564–574 (2013).

77. Dobin, A. et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

78. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

79. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

80. Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative CT method. *Nat. Protoc.* **3**, 1101–1108 (2008).

81. Wang, Y. et al. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).

82. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

83. Krueger, F. & Andrews, S. R. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).

84. Gruntman, E. et al. Kismeth: Analyzer of plant methylation states through bisulfite sequencing. *BMC Bioinformatics* **9**, 371 (2008).

85. Li, S. et al. SUVH1, a Su(var)3–9 family member, promotes the expression of genes targeted by DNA methylation. *Nucleic Acids Res.* **44**, 608–620 (2016).

86. Dai, H. et al. Development of a seedling clone with high regeneration capacity and susceptibility to *Agrobacterium* in apple. *Sci. Hortic.* **164**, 202–208 (2013).

Acknowledgements

Author contributions

D.Z., Y.J., Z.F. and J.Mao designed and coordinated the study. Y.C., Yu L., L.T. and W.K. performed genomic and transcriptomic analyses. Y.C. and J.Mao performed the data analysis. X.G., Yu L., L.T., C.C., Yandong L., X.Z., C.C., S.L., X.S. and D.M. performed the experiments and data analysis. Z.Y., Z.L., F.M., Z.Z., W.L. and Q.L. provided samples. Y.C., J.Mao, D.Z., Z.F. and Yu L. wrote the paper with contributions from T.F., J.Y., A.C., X.C., M.M.T., J.Ma, L.X., N.A., Yanrong L., C.Z., C.W., Y.H. and H. L.

Competing interests

The authors declare no competing interests.

Additional information