


RESEARCH

Open Access



# A metabolic perspective of selection for fruit quality related to apple domestication and improvement

Qiong Lin<sup>1,2,3†</sup>, Jing Chen<sup>1†</sup>, Xuan Liu<sup>1†</sup>, Bin Wang<sup>4†</sup>, Yaoyao Zhao<sup>1</sup>, Liao Liao<sup>5</sup>, Andrew C. Allan<sup>6</sup>, Chongde Sun<sup>3</sup>, Yuquan Duan<sup>1</sup>, Xuan Li<sup>1</sup>, Donald Grierson<sup>3,7</sup>, Julian C. Verdonk<sup>2</sup>, Kunsong Chen<sup>3</sup>, Yuepeng Han<sup>5</sup> and Jinfeng Bi<sup>1\*</sup> 

<sup>†</sup>Qiong Lin, Jing Chen, Xuan Liu, and Bin Wang contributed equally to this work.

\*Correspondence: bjfcaas@126.com

<sup>1</sup> Key Laboratory of Agro-Products Quality and Safety Control in Storage and Transport Process, Ministry of Agriculture and Rural Affairs/Institute of Food Science and Technology, Chinese Academy of Agricultural Sciences, Beijing 100193, China

<sup>2</sup> Horticulture and Product Physiology, Department of Plant Sciences, Wageningen University, Wageningen 6708 PD, The Netherlands

<sup>3</sup> College of Agriculture and Biotechnology, Zhejiang University, Zijingang Campus, Hangzhou 310058, China

<sup>4</sup> Wuhan Metware Biotechnology Co., Ltd., Wuhan 430070, China

<sup>5</sup> Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China

<sup>6</sup> The New Zealand Institute for Plant and Food Research Limited, Auckland Mail Centre, Auckland 1142, New Zealand

<sup>7</sup> Plant and Science Crop Division, School of Biosciences, University of Nottingham, Sutton Bonington Campus, Loughborough LE12 5RD, UK

## Abstract

**Background:** Apple is an economically important fruit crop. Changes in metabolism accompanying human-guided evolution can be revealed using a multiomics approach. We perform genome-wide metabolic analysis of apple fruits collected from 292 wild and cultivated accessions representing various consumption types.

**Results:** We find decreased amounts of certain metabolites, including tannins, organic acids, phenolic acids, and flavonoids as the wild accessions transition to cultivated apples, while lysolipids increase in the “Golden Delicious” to “Ralls Janet” pedigree, suggesting better storage. We identify a total of 222,877 significant single-nucleotide polymorphisms that are associated with 2205 apple metabolites. Investigation of a region from 2.84 to 5.01 Mb on chromosome 16 containing co-mapping regions for tannins, organic acids, phenolic acids, and flavonoids indicates the importance of these metabolites for fruit quality and nutrition during breeding. The tannin and acidity-related genes *Myb9-like* and *PH4* are mapped closely to fruit weight locus *fw1* from 3.41 to 3.76 Mb on chromosome 15, a region under selection during domestication. Lysophosphatidylethanolamine (LPE) 18:1, which is suppressed by *fatty acid desaturase-2* (*FAD2*), is positively correlated to fruit firmness. We find the fruit weight is negatively correlated with salicylic acid and abscisic acid levels. Further functional assays demonstrate regulation of these hormone levels by *NAC-like activated by Apetala3/Pistillata* (*NAP*) and *ATP binding cassette G25* (*ABCG25*), respectively.

**Conclusions:** This study provides a metabolic perspective for selection on fruit quality during domestication and improvement, which is a valuable resource for investigating mechanisms controlling apple metabolite content and quality.

**Keywords:** Apple, Metabolome, mGWAS, Taste, Storage, Quality, Fruit weight



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Apple is one of the most economically important fruit crops all over the world. Due to their high nutritional value and strong ecological adaptability, apples are preferred by both growers and consumers [1, 2]. Following efforts of domestication and cultivar breeding, the global production of apples in 2020 has exceeded 86 million tons [3]. Moreover, modern growers are keen to produce cultivars tailored to the apple industry that focus on and cater to discerning consumers, with changing preferences and dietary habits, who are highly aware of the nutritional and health benefits of fruits [4].

Apple belongs to the genus *Malus* within the family Rosaceae. The cultivated apples belong to *M. × domestica* Borkh., while other *Malus* species are commonly known as wild apples or crab apples. Genetic and morphological evidence has suggested that cultivated apples are more closely related to *M. sieversii*, a wild species native to the Tianshan Mountains in central Asia, than other wild species [5–7]. As apples spread along the Silk Road between Asia and Europe, their genetic makeup changed due to hybridization with local wild apples. Following the primary progenitor *M. sieversii*, cultivated apples obtained intensive introgressions from *M. sylvestris* [6–8]. Fruit taste and size are two of the most important characteristics in apple breeding. The content and composition of organic acids, soluble tannins, and soluble sugars contribute to the overall taste of apples and other fruits [9]. Wild apple fruits are astringent and sour due to higher contents of tannins and organic acids than cultivated apples, which are not astringent and less acidic [10]. Also, wild apples are smaller than cultivated apples, suggesting an extensive selection for fruit size during apple domestication [7, 11]. While the domestication of maize and tomato was initiated from ancestral species with small fruits, the domestication of apple benefits from *M. sieversii* bearing large fruits [12].

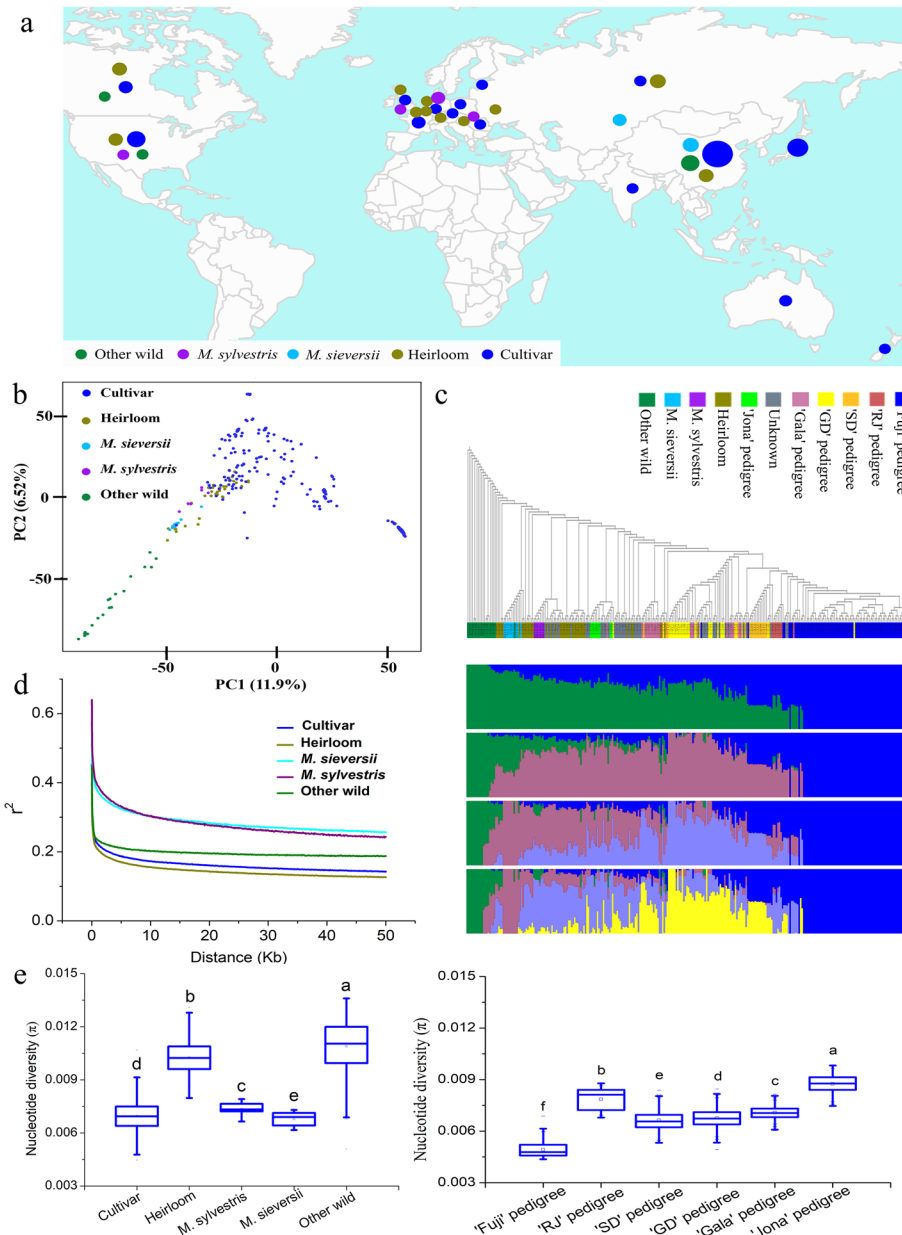
Thousands of apple cultivars have been reported worldwide due to diverse hybridization and introgression processes [13]. Besides the fruit taste and size, other characteristics such as texture [14], storage life [15], and nutrition [16] vary in apples with different pedigrees. Recent studies have demonstrated that metabolism changes within a species are greater than expected [17, 18], and this genetic diversity determines the naturally occurring variation in metabolites. Integrating metabolic analysis with other omics has been widely used in identifying genes and elucidating plant pathways [19, 20]. In particular, metabolome-based genome-wide association studies (mGWAS) have been proven efficient in clarifying natural variation in metabolism and its genetic control in tomato [21], peach [22], rice [23], jujube [24], and other crops. GWAS has been recently used to map the genetic loci for important phenotypic traits in apples [25–27]. The diversity of apples has rendered them an “ideal model” to interpret the genetic basis underlying the accumulation of metabolites and, in turn, the regulation of the metabolome.

In the present study, we collected 292 accessions of wild and cultivated apples to compare the genomic and metabolic differences among apples of different pedigrees. A comprehensive population genomic analysis and metabolomic variation related to apple domestication and improvement were conducted. Subsequently, mGWAS was carried out to identify the key genes associated with specific modifications. The results provide insights into genetic variants associated with apple domestication and improvement, and the data are valuable for clarification of the mechanisms controlling quality traits.

## Results

### Genome variation of wild and cultivated apple accessions

We collected and re-sequenced 292 apple accessions worldwide (Fig. 1a), including 11 *M. sieversii*, seven *M. sylvestris*, 19 other wild species, 38 heirlooms, and 217



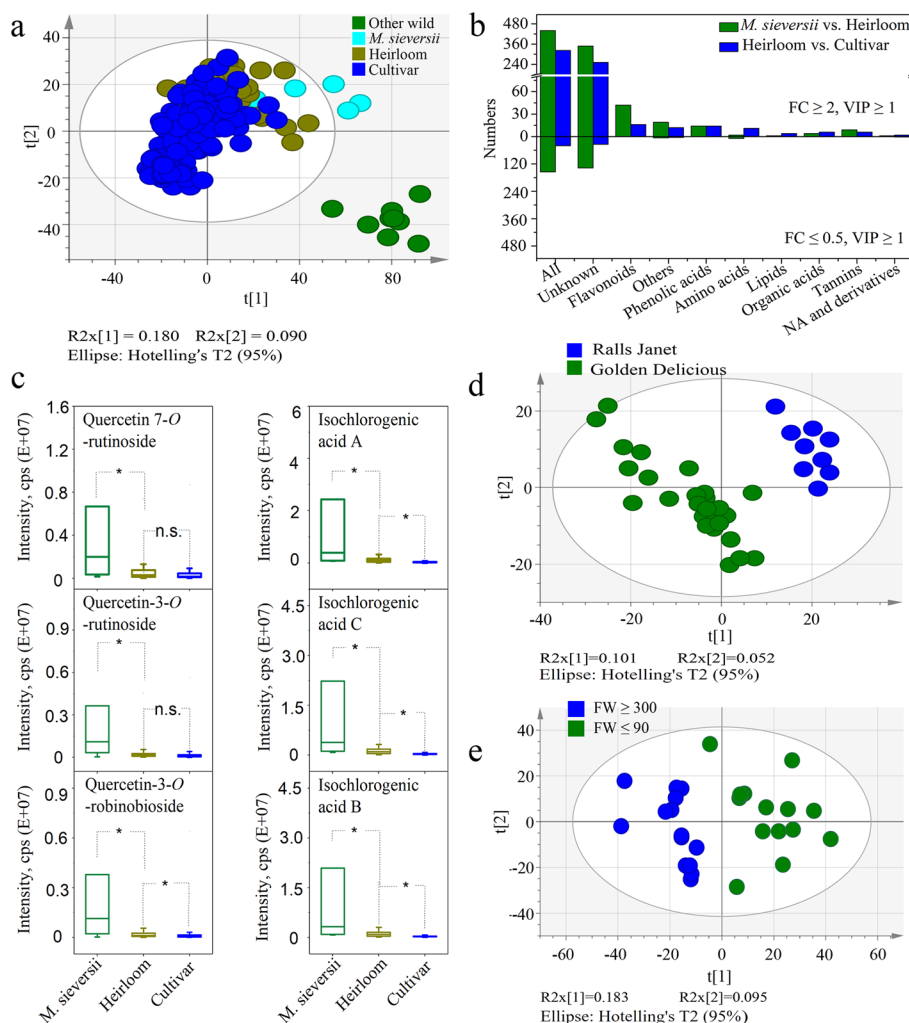
**Fig. 1** Genome-wide relationship in cultivated apple and wild species. **a** Geographic distribution of apple accessions. The circle represents the sample size. **b** Principal component analysis (PCA) of apple accessions. **c** The phylogenetic tree and population structure of apple accessions. "RJ" represents "Ralls Janet," "SD" represents "Starking Delicious," "GD" represents "Golden Delicious" and "Jona" represents "Jonathan." The accession orders and positions on the X-axis are the same. **d** Linkage disequilibrium (LD) decay patterns in *M. sieversii*, *M. sylvestris*, other wild, heirloom, and cultivar groups. **e** Nucleotide diversity ( $\pi$ ) between different groups. Lowercase letters indicate the significant differences based on Fisher's least significant difference (LSD) test at  $p \leq 0.01$

cultivars representing different pedigrees and consumption types, which yielded 2.67 Terabyte (Tb) high-quality cleaned sequences, with a median depth of  $11.42 \times$  and coverage of 94.06% of the assembled genome (Additional file 2: Tables S1 and S2). A total of 2,956,396 high-quality single-nucleotide polymorphisms (SNPs) were selected for genetic variation analyses (Additional file 2: Table S3). Principal component analysis (PCA) results indicated that the cultivars were separated from wild accessions and were much closer to *M. sieversii* and *M. sylvestris*, while heirlooms were distributed from wild to cultivars (Fig. 1b). The phylogenetic tree supported the clustering of the wild and cultivated clades, where the *M. sylvestris* and all apple cultivars situated in a large monophyletic cluster. Population structure analysis demonstrated that the domesticated apples along with *M. sieversii* and *M. sylvestris* were separated from other wild species when  $K$  ranged from three to five (Fig. 1c). In general, the distribution of heirlooms was partially intermingled with wild and apple cultivars, indicating a transition of heirlooms from wild to cultivars. Likewise, the “Fuji,” “Ralls Janet,” and “Starking Delicious” pedigrees formed three sub-clusters in a monophyletic clade, which suggests the “Starking Delicious” and “Ralls Janet” as the pollen parents of “Fuji” [28]. Rapid linkage disequilibrium (LD) decays were observed in cultivars and heirlooms, and the process was found slow in *M. sieversii* and *M. sylvestris* (Fig. 1d). Nucleotide diversities were different between groups, with other wild species exhibiting the highest degree of genetic variation while the lowest was observed in the “Fuji” pedigree due to erosion caused by breeding (Fig. 1e).

#### Variation of metabolites related to apple domestication and quality improvement

Apple flesh was collected from 270 accessions comprising nine other wild accessions, six *M. sieversii*, 38 heirlooms, and 217 cultivars, and we constructed an MS2 spectral tag (MS2T) library using a widely targeted metabolic profiling method [29]. A data matrix of 2575 metabolites, including 486 annotated and 2089 unannotated metabolites, was generated (Additional file 2: Table S4). Among the annotated metabolites, 275 were confirmed by standards, and 211 were putatively annotated as described previously, including flavonoids, phenolic acids, organic acids, lipids, alkaloids, tannins, amino acids, nucleotides, and others (Additional file 1: Fig. S1a). Over 65% of the metabolites exhibited a coefficient of variation (CV) greater than 50%, indicating significant variation between apple accessions (Additional file 1: Fig. S1b). Of these, flavonoids revealed the highest CV with an average of up to 349%, followed by phenolic acids and lipids with an average of 202 and 201%, respectively (Additional file 2: Table S4). PCA analysis of all metabolites divided the apple accessions into four groups, namely other wild, *M. sieversii*, heirlooms, and cultivars, where the other wild group was separated from the other three groups (Fig. 2a). The cultivars were further grouped into six clades, namely “Jonathan,” “Gala,” “Golden Delicious,” “Starking Delicious,” “Ralls Janet,” and “Fuji” pedigrees, respectively (Additional file 1: Fig. S2). The results obtained from metabolite differentiation were generally consistent with the phylogenetic tree relationships, indicating a strong relationship between changes in metabolite and genomic variation of apples.

A total of 595 and 363 different metabolites were identified in the comparison between *M. sieversii* and heirloom and between heirloom and cultivar (Additional file 2: Tables S5 and S6), respectively, revealing the metabolome variation in apples during domestication



**Fig. 2** Profiling of metabolites among different apple accessions. **a** PCA analysis of all metabolites in other wild accessions, *M. sieversii*, heirlooms, and cultivars. **b** Differentiation of metabolites between *M. sieversii*/ heirloom and heirloom/ cultivar comparisons. **c** Box plots for the contents of flavonoids (quercetin 7-O-rutinoside, quercetin-3-O-rutinoside, and quercetin-3-O-robinobioside) and phenolic acids (isochlorogenic acid A, isochlorogenic acid B, and isochlorogenic acid C) with large fold changes between *M. sieversii* and heirloom groups. **d** PCA analysis between the “Ralls Janet” and “Golden Delicious” pedigrees, and **e** between high-weight ( $\geq 300$ ) and low-weight ( $\leq 90$ ) domesticated apples. t[1] and t[2] represent the first and second principal component.  $R^2 \times [1]$  and  $R^2 \times [2]$  indicate the proportion of variance explained by t[1] and t[2], respectively. The colored dots indicate different apple accessions and the ellipses indicate the 95% confidence interval for each data set

and improvement [26]. Among these, the levels of most flavonoids, phenolic acids, organic acids, and tannins decreased when the *M. sieversii* transitioned into heirloom and from heirloom to cultivar (Fig. 2b). However, most flavonoids such as quercetin-3-O-robinobioside, quercetin-3-O-rutinoside (rutin), and quercetin 7-O-rutinoside and phenolic acids such as isochlorogenic acid A, isochlorogenic acid B, and isochlorogenic acid C exhibited large fold change (FC) of up to 20 and 10, respectively, between *M. sieversii* and heirloom. However, the value was lower in comparison between heirloom and cultivar, with an average of 3 (Fig. 2c). It is well known that tannins and organic acids contribute to an astringent property and acid taste in fruits, while flavonoids and phenolic acids

have been associated with the astringent property and bitter taste in fruits [30–32]. Consequently, it can be deduced that acidity, astringency, and bitter tastes are the primary characteristics selected for removal during domestication.

Improved storage life is another selection criterion for fruit quality by the breeders. Fruits of the “Ralls Janet” pedigree are characterized by durable storability and firm flesh [33], which can be passed on to offspring, while fruits of the “Golden Delicious” pedigree soften during storage [34]. The comparison of all metabolites in the “Ralls Janet” and “Golden Delicious” pedigree divided them into two distinct groups (Fig. 2d). The two groups differed in 361 metabolites, of which 64 were annotated. Interestingly, seven lysolipids were accumulated to significantly higher levels in the “Ralls Janet” pedigree than in the “Golden Delicious” pedigree (Additional file 2: Table S7). Of these, lysophosphatidylethanolamine (LPE) 18:1 and LPE 18:1 (2n: isomer) were highly accumulated, with FC of up to 4.39 and 3.68, respectively. Lysolipids have been reported to improve cell membrane stability [35, 36] and mediate hormonal regulation of growth and senescence [37, 38]. Thus, the accumulation of lysolipids, particularly LPE 18:1, could potentially enhance the apple’s storage life.

Fruit weight is an important agronomical trait for many domesticated crops. Accordingly, the potential metabolites that could affect domesticated apple fruit weight were investigated. PCA analysis of metabolites divided the domesticated apples into high- and low-weight groups (Fig. 2e). The two groups differed in 805 metabolites, of which 181 were annotated (Additional file 2: Table S8). Interestingly, the phytohormones abscisic acid (ABA) and salicylic acid (SA) accumulated significantly higher levels in low-weight apples, with FC of 3.69 and 1.90, respectively. Moreover, 14 soluble tannins and eight organic acids were significantly accumulated in low-weight fruits, with an average FC of up to 4.54 and 3.75, respectively (Additional file 2: Table S8). The total tannin and titratable acidity decreased with an increase in weight (Additional file 1: Fig. S3), suggesting that tannin and organic acid content were inversely correlated with the evolution of enhanced fruit weight.

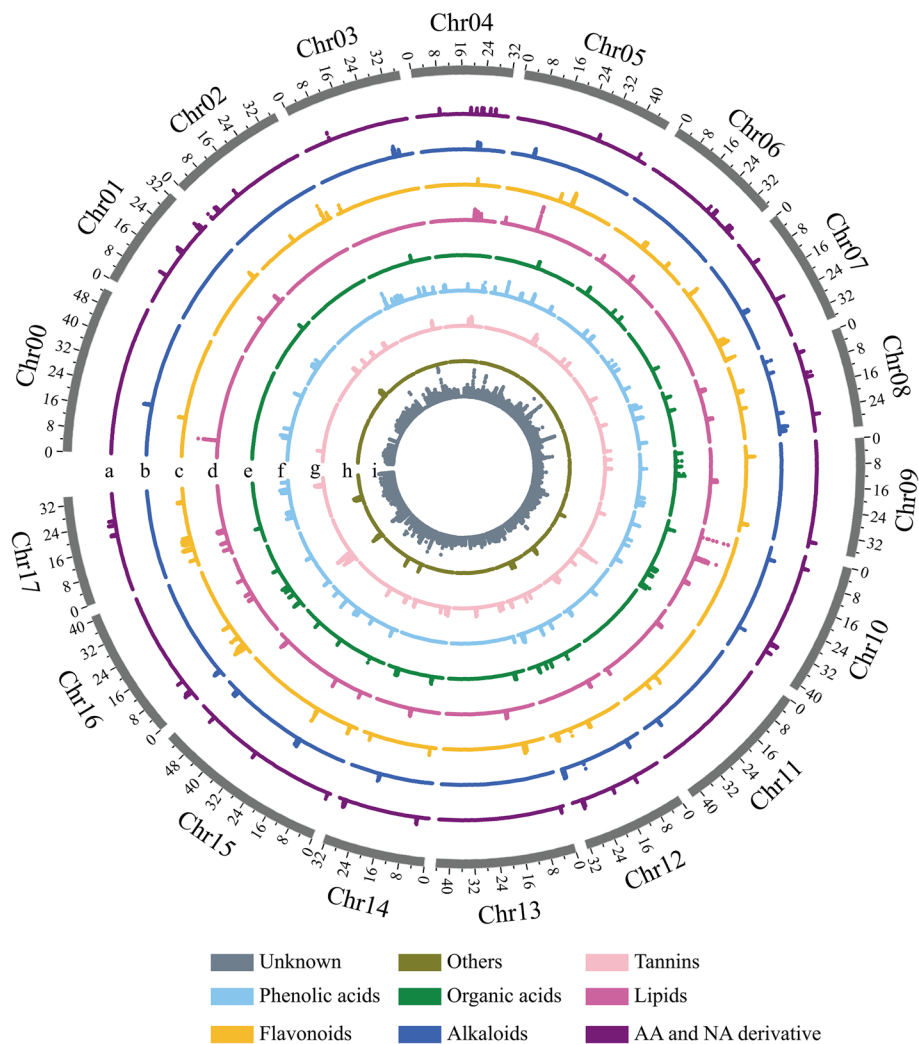
#### Genetic basis of metabolite content in different apple accessions

We then performed mGWAS for 270 apple accessions using a linear mixed model (LMM) approach implemented in GEMMA to reduce false positives by considering genome-wide patterns of genetic correlation. A total of 222,877 significant SNPs associated with 2205 apple metabolites were detected (Additional file 2: Table S9). Manhattan plots of significant SNPs of metabolites, including tannins, phenolic acids, organic acids, lipids, flavonoids, alkaloids, amino acids, nucleic acid derivatives, others, and unknown metabolites are illustrated (Fig. 3). Furthermore, an overlap of the co-mapped region of flavonoids, organic acids, tannins, and phenolic acids from 2.84 to 5.01 Mb on chromosome (Chr) 16 was investigated, suggesting this region is important for improving taste, quality, and nutrition in apples.

#### Domestication based on fruit taste and its underlying biosynthetic network

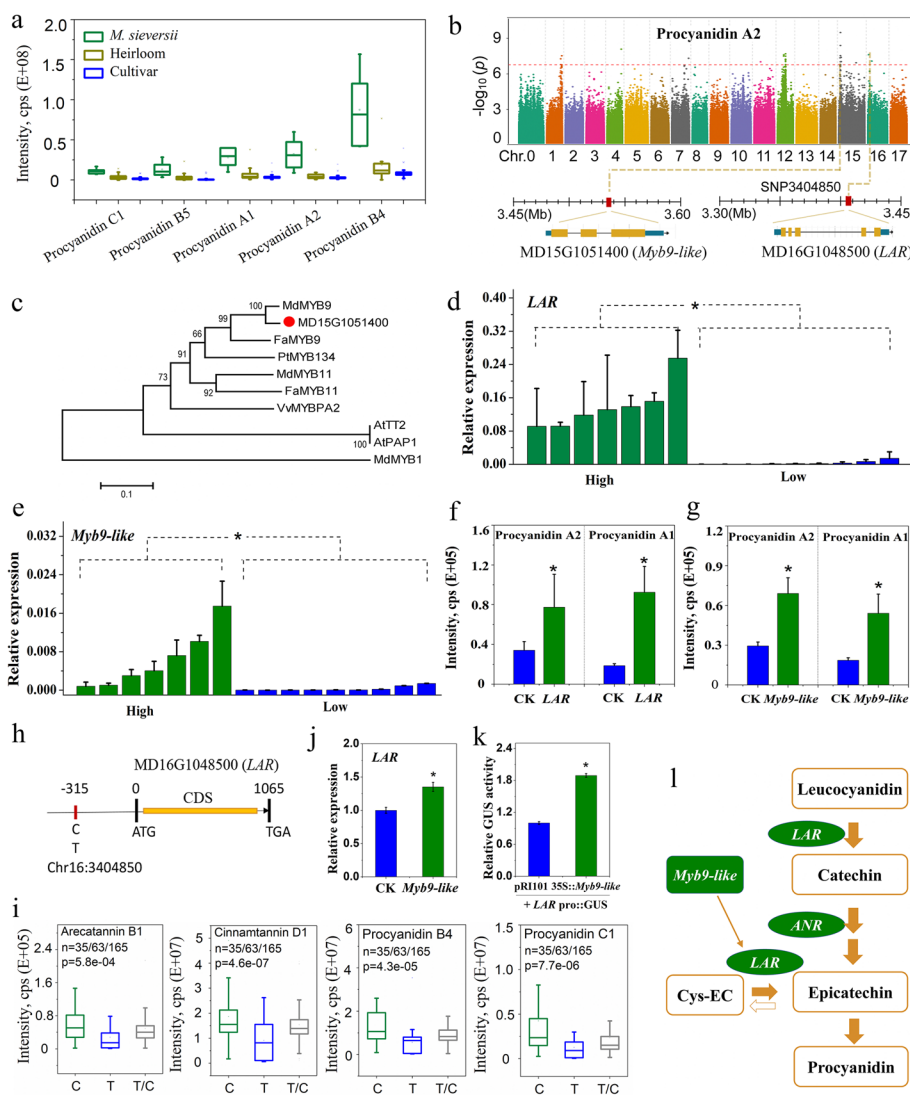
Tannins, including procyanidin A1, procyanidin A2, procyanidin B4, procyanidin B5, and procyanidin C1, were accumulated to higher levels in the *M. sieversii* group than in heirlooms and cultivars (Fig. 4a). From GWAS results, a strong peak on Chr





**Fig. 3** Genome distribution of significant SNPs related to all metabolites. The circles from outside to center represent amino acids and nucleic acids (a), alkaloids (b), flavonoids (c), lipids (d), organic acids (e), others (f), phenolic acids (g), soluble tannins (h), and unknown metabolites (i). The unit for chromosome length is Mb. The *P*-value of all significant SNPs associated with metabolites are listed in Additional file 2: Table S9

16 co-mapped with procyanidin A2, procyanidin C1, procyanidin B4, cinnamtannin D1, and arecatannin B1 was identified (Additional file 1: Fig. S4), estimating a candidate region from 1.55 to 5.57 Mb harboring one putative gene (Additional file 2: Table S10). MD16G1048500 contains the consensus sequence defining *leucoanthocyanidin reductase (LAR)* [39] and is 32 kb from the lead SNP (Chr16:3,372,301) (Fig. 4b). In addition, a strong peak on Chr 15 was identified for procyanidin A2, indicating a candidate region from 1.82 to 4.81 Mb harboring two putative genes (Additional file 2: Table S10). One of these, *Myb9-like* (MD15G1051400), has the potential to regulate anthocyanin and procyanidin metabolisms (Fig. 4b). A phylogenetic tree of reported R2R3 MYBs revealed that MD15G1051400 was closely related to *MdMYB9* and *FaMYB9* (Fig. 4c; Additional file 1: Fig. S5), which are involved in procyanidin synthesis by regulating *LAR* and *anthocyanidin reductase (ANR)* [40, 41],



**Fig. 4** Identifying metabolites and genes potentially involved in apple domestication. **a** The contents of tannins with significant differential accumulation in *M. sieversii*/ heirloom and heirloom/ cultivar comparisons. **b** Manhattan plots of procyanidin A2 by GWAS. **c** A phylogenetic tree for MD15G1051400 and other R2R3 MYB members. **d,e** qRT-PCR analysis of *Myb9-like* and *LAR* (*leucoanthocyanidin reductase*) in high- and low-procyanidin content apples. **f,g** Results of transient overexpression of *Myb9-like* and *LAR* in apple fruits than empty vector (pGreenII 62-SK) injections. **h** Nucleotide polymorphism identified in the promoter of *LAR*. **i** Box plot for procyanidins contents correlated with C and T in Chr16:3404850. **j** qRT-PCR analysis of *LAR* in *Myb9-like* overexpressing apple fruit. **k** GUS transactivation assay in apples. The GUS reporter vector contains the *LAR* promoter, while the effector vector contains *Myb9-like*. **l** A simplified metabolic biosynthetic pathway of procyanidins in fruit. ANR, anthocyanidin reductase; Cys-EC, 4β-(S-cysteiny)-epicatechin. Data are expressed as mean ± sd. \*\* indicate  $p \leq 0.05$  by *t*-test

suggesting that *Myb9-like* may regulate procyanidin synthesis in apple. Fruit weight quantitative trait locus (QTL), *fw1*, was co-located with vacuolar acidification-related *PH4* (MD15G1051000) [42] as well as *Myb9-like* from 3.41 to 3.76 Mb on Chr 15, this region was found in the selective sweeps of *M. sieversii*/ heirloom and *M. sylvestris*/ heirloom comparisons but not in heirloom/ cultivar comparison (Additional file 1: Fig. S6), the results were consistent with a previous report [26].

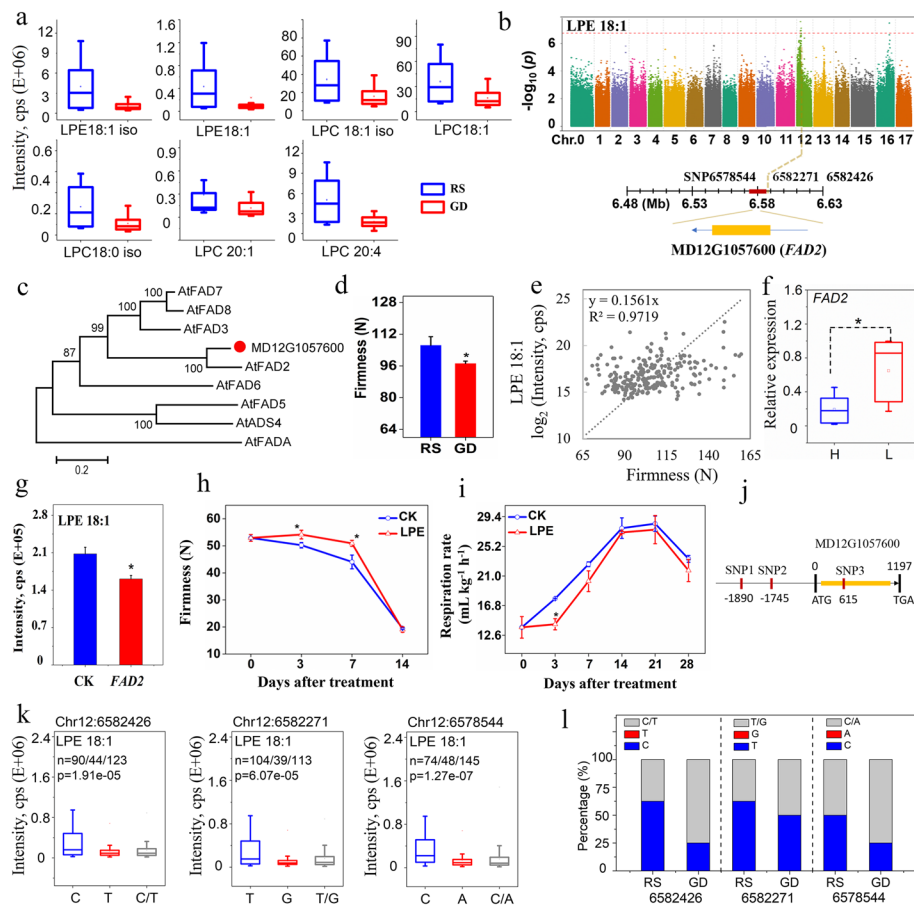


The expression levels of *Myb9-like* and *LAR* were higher in fruits with high procyanidin than those low in procyanidin (Fig. 4d, e). Transiently overexpressing *Myb9-like* and *LAR* exhibited a significantly higher level of procyanidin A2 and procyanidin A1 in apple fruits than the controls (Fig. 4f, g). A significant SNP (Chr16:3,404,850) was located at –315 bp upstream of the *LAR* gene (Fig. 4h), where allelic mutations of C and T were significantly correlated with the contents of tannins such as procyanidin B4, procyanidin C1, cinnamtannin D1, and arecatannin B1 (Fig. 4i). The allelic mutation in the promoter could potentially lead to differential expression of *LAR* in apples with a high and low level of procyanidin. Furthermore, transiently overexpressing *Myb9-like* increased *LAR* expression in apple flesh (Fig. 4j), and the *Myb9-like* specifically activated GUS transcription as driven by the *LAR* promoter (Fig. 4k). *LAR* enzyme catalyzes leucocyanidin to catechin and converts 4 $\beta$ -(S-cysteinyl)-epicatechin (Cys-EC) to epicatechin [43, 44]. Thus, it is inferred that *Myb9-like* might participate in procyanidin synthesis by promoting *LAR* expression (Fig. 4l).

#### Improvement of metabolites related to the firmness of fruits

Seven lysolipids, in particular LPE 18:1 implicated in firmness and storage life of fruits [45], accumulated to higher levels in the “Ralls Janet” pedigree than in the “Golden Delicious” pedigree (Fig. 5a). Based on GWAS results, a strong peak on Chr 12 co-mapped for LPE 18:1 and LPE 18:1 (2n isomer) was identified, with estimated candidate region from 5.93 to 10.01 Mb, harboring three putative genes (Additional file 1: Fig. S7; Additional file 2: Table S10). One of these, MD12G1057600, which encodes *fatty acid desaturase 2* (*FAD2*), is 0.72 kb from the lead SNP (Chr12:6,578,653) (Fig. 5b). A phylogenetic tree of the well-characterized *Arabidopsis* FAD family demonstrated that MD12G1057600 was closely related to *AtFAD2* (Fig. 5c), which converts LPE precursor PE 18:1 to PE 18:2 [46], suggesting that the activity of *FAD2* may reduce LPE 18:1 accumulation.

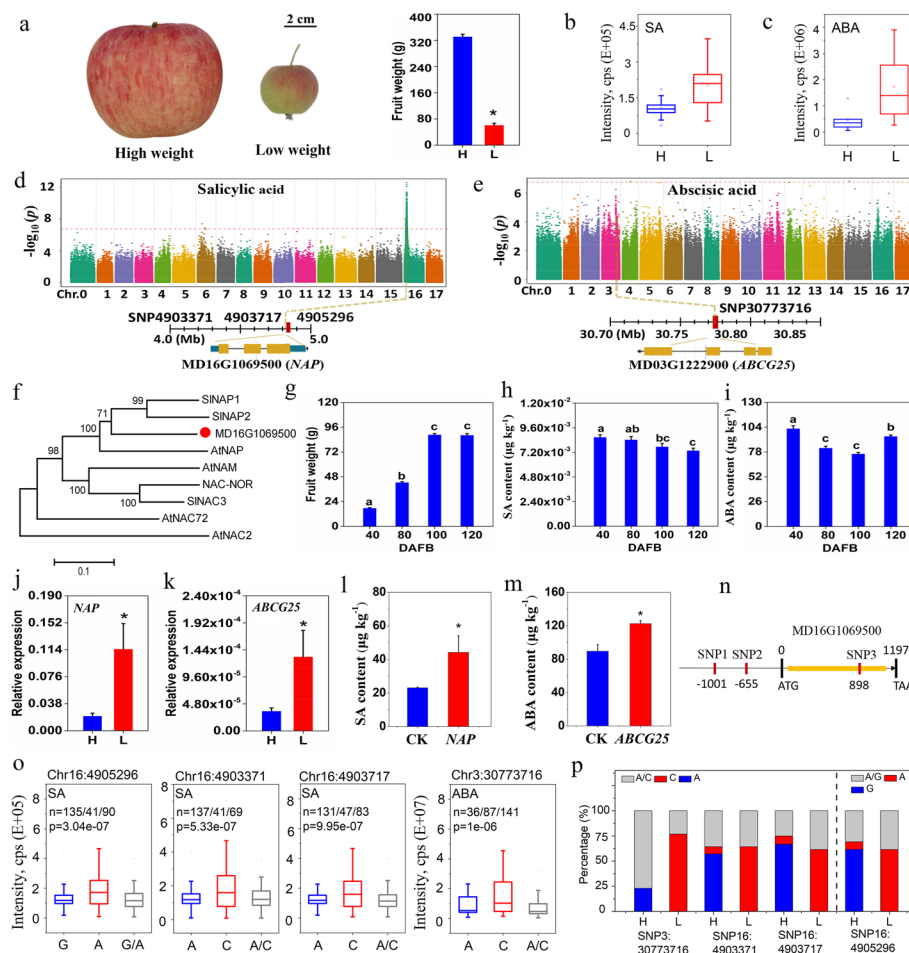
Consistent with their enhanced storage life, apples from “Ralls Janet” pedigree exhibited higher firmness than the “Golden Delicious” pedigree (Fig. 5d). A close correlation between the content of LPE 18:1 and the firmness of cultivated apple accessions was investigated (Fig. 5e). Furthermore, a significant difference was observed in transcript levels of *FAD2* in apple accessions with high and low LPE 18:1 content (Fig. 5f). Transient overexpression of *FAD2* in apple fruit led to a decreased level of LPE 18:1 than that of the control (Fig. 5g). To test the effects of LPE 18:1 on fruit firmness, apples were sprayed with LPE 18:1 before storage. Compared with the control, LPE 18:1 maintained apple firmness at 3 and 7 days of storage (Fig. 5h). Also, the respiration rate in LPE 18:1-treated fruit was reduced during storage (Fig. 5i). Furthermore, three SNPs were significantly correlated with LPE 18:1 content. Chr12:6582426 and Chr12:6582271 located at –1890 and –1745 bp upstream of *FAD2*, respectively, and Chr12:6578544 located at 615 bp in the coding sequence might induce a non-synonymous amino acid change in *FAD2* (Fig. 5j). Allelic mutations that correlated with high and low LPE 18:1 content were mainly found in the “Ralls Janet” and “Golden Delicious” groups, respectively (Fig. 5k, l). Thus, *FAD2* is a strong candidate for suppressing the accumulation of LPE 18:1, which was potentially maintaining the firmness and prolonging the storage life of apples.



**Fig. 5** Functional interpretation of GWAS for lysophosphatidylethanolamine (LPE) 18:1 content in apple. **a** Lysolipid contents of apples in the “Ralls Janet” (RS) and “Golden Delicious” (GD) pedigrees. **b** Manhattan plots of LPE 18:1 content by GWAS. **c** An unrooted phylogenetic tree for MD12G1057600 and *Arabidopsis* FAD2 members. **d** The firmness of apples in “RS” and “GD” pedigrees. **e** Correlation analysis ( $R^2$ ) between firmness and LPE 18:1 in domesticated apples. **f** qRT-PCR analysis of *FAD2* transcripts in domesticated fruits with high and low LPE 18:1. **g** Transient overexpression of *FAD2* in apple fruits. **h** The firmness and respiration rate of apples treated with LPE 18:1 during storage. **i** The firmness and respiration rate of apples treated with LPE 18:1 during storage. **j** Nucleotide polymorphisms identified in the promoter and coding sequence of *FAD2*. **k** Box plots of LPE 18:1 correlated with identified nucleotide polymorphisms. **l** Nucleotide polymorphisms in the “RS” and “GD” pedigrees. The data are expressed as mean  $\pm$  sd. “\*\*” indicates  $p \leq 0.05$  by *t*-test. The correlation was evaluated based on the Pearson correlation coefficient

### Metabolome alterations related to fruit weight targeted selection

The phytohormones ABA and SA, involved in cell division and elongation of plants [47, 48], were accumulated to high levels in low-weight fruits (Fig. 6a–c). Based on GWAS results, a strong peak associated with SA contents was identified on Chr 16 (Fig. 6d), with an estimated region from 2.70 to 4.99 Mb harboring three putative genes (Additional file 2: Table S10). MD16G1069500 encoding a transcription factor, *NAC-like activated by Apetala3/Pistillata* (*NAP*), was identified in this region. Based on a phylogenetic tree of reported *NACs*, MD16G1069500 was closely related to tomato *SINAP1* and *SINAP2* (Fig. 6f), which were reported to regulate SA biosynthesis and fruit size in tomatoes [49, 50]. In addition, a strong peak on Chr 3 was identified, which was associated with ABA content, with an estimated candidate region from 30.57 to 32.58 Mb, harboring two putative genes (Additional file 2: Table S10).



**Fig. 6** Functional analysis of GWAS for SA and ABA content related to apple fruit weight. Fruit weight (a), SA (b), and ABA (c) contents in high- (H) and low- (L) weight domesticated fruits. **d,e** Manhattan plots of SA and ABA contents by GWAS. **f** An unrooted phylogenetic tree was constructed for MD15G1051400 and other *NAC* members. Fruit weight (g), SA (h), and ABA (i) contents during fruit development. **j,k** qRT-PCR analysis of *NAP* (MD16G1069500) and *ABCG25* (MD03G1222900) in H and L apple groups. **l,m** Results of transient overexpression of *NAP* and *ABCG25* in apple fruits. **n** Nucleotide polymorphisms identified in the promoter and coding sequence of *NAP*. **o** Box plot of SA and ABA correlated with identified nucleotide polymorphism. **p** Nucleotide polymorphisms in H and L apple groups. DAFB, days after full bloom. Data are expressed as mean  $\pm$  sd. **\*\*** indicate  $p \leq 0.05$  by *t*-test

MD03G1222900 contains the consensus sequence of an ABA transporter, *ATP binding cassette G25* (*ABCG25*), which controls the ABA level in *Arabidopsis* [51] (Fig. 6e).

Consistent with its role in restricting the increase in fruit weight, SA content decreased during fruit development, while ABA content decreased from 40 to 100 days after full bloom (DAFB) and increased to 120 DAFB, possibly due to the role of ABA in fruit ripening (Fig. 6g–i) [52]. The transcript level of *NAP* was significantly higher in low-weight apples than in high-weight apples, which correlated with the SA contents in high- and low-weight groups (Fig. 6j). Transiently overexpressing *NAP* in apple fruits indicated that SA content was significantly increased, which further supports the role of *NAP* in SA synthesis (Fig. 6l). The transcript level of *ABCG25* was significantly higher in low-weight apples, which correlated with the ABA content

in high- and low-weight apples (Fig. 6k). Also, apple fruits transiently overexpressing *ABCG25* contained a significantly higher ABA content than the controls (Fig. 6m). Two significant SNPs (Chr16:4903371 and Chr16:4903717) located upstream, and Chr16:4905296 located at 898 bp in *NAP* may lead to a non-synonymous amino acid change within this gene (Fig. 6n). Additionally, the Chr3:30773716 located at 2865 bp in *ABCG25* may lead to a non-synonymous amino acid change in *ABCG25*. Allelic mutations that led to low and high SA and ABA contents were mainly found in the high- and low-weight groups, respectively (Fig. 6o, p). Thus, it is concluded that accumulation of SA and ABA, which appear to be affected by the genes, *NAP* and *ABCG25*, respectively, may restrict fruit enlargement in apples.

## Discussion

The metabolome defines what we ingest from a plant product and determines its attractiveness and nutritional value. Plant metabolites play important roles in growth, cellular replenishment, whole-plant resource allocation, and the adaptation of plants to the environment [53, 54]. Consequently, the metabolome is a direct reflection of the physiological status and a connection between the genotype and the phenotype of a plant [55]. Given the importance of plant metabolism to plant development and for human health, many genetic studies have been conducted to identify metabolic pathways in apples [56], tomatoes [21], peach [22], rice [23], jujube [24], and other crops. With increasing genome resources, GWAS has been conducted on apples to clarify important agronomical traits that contribute to yield and quality [7, 26]. In the current study, mGWAS data analysis aimed at understanding the effects of domestication and improvement on the metabolic composition of apples was provided. A total of 222,877 significant SNPs were identified that are associated with 2205 apple metabolites. This information provides a valuable resource for further studies on biosynthetic pathways involved in determining the contents of metabolites in apple fruit and provides a reference for metabolic pathway elucidation to other crops. An interesting overlap of the co-mapped region of flavonoids, organic acids, soluble tannins, and phenolic acids from 2.84 to 5.01 Mb on Chr 16 was investigated. Based on previous reports, several important genes or QTLs related to apple fruit quality have been identified in this region [26, 57–59]. In our study, a NAC transcription factor *NAP* potentially related to an increment in fruit weight was identified at a 4.90~4.91 Mb region on Chr 16. Thus, the mGWAS data identified a region on Chr 16 as having great value for apple quality and nutrition breeding.

Apple accessions were divided into two major groups based on genomic and metabolomic criteria that distinguished between the wild and cultivated apples. Although many omics studies have been conducted on crop species and their progenitors [60, 61], very few researches have been performed at the metabolic level that allows evaluation of the impact of domestication on metabolites diversity and their contents [62]. In recent studies, the impact of domestication on the fruit metabolome in tomatoes [21], the combined nature and human selections on peach fruit metabolome [22], and the variation in primary metabolism in a diverse lettuce population were investigated [63]. A previous study suggested that cultivated apples were domesticated from *M. sieversii* in Central Asia [12]. In this study, the genomes and metabolomes of *M. sieversii* and heirloom (domestication), and between heirloom and cultivar (improvement) were evaluated [26],

and the question was addressed of how domestication and improvement have altered fruit chemical composition. This study provided a direct evaluation of the chemical compositional consequences of apple domestication and improvement at such a scale.

Apple domestication resulted in the reduction of certain compounds that produce adverse effects on taste, including tannins [30], organic acids [10, 64], flavonoids, and phenolic acids [32] but not sugars, which are consistent with previous reports [26]. Also, organic acid content has been identified as the key criterion for the evolution of taste during domestication in citrus [65], lettuce [63], jujube [66], and pear [67], which showed both distinct and overlapping aspects of genetic control of metabolism between species. Flavonoids have been associated with astringency and bitter taste in fruit [31, 32], which characteristics are to be removed during domestication. Other research has also suggested that the flavonoids is excluded from selection during the domestication of citrus, tomato, rice, wheat, soybean, and barley [68, 69]. Interestingly, the tannins and organic acids were significantly accumulated at different levels based on the comparisons between the heirloom and *M. sieversii* groups (Fig. 2a, b) and between high- and low-weight domesticated apples (Fig. 2e), which were also found to be negatively correlated with fruit weight (Additional file 1: Fig. S3). Furthermore, the locus for the fruit weight, *fw1*, was located in a similar region with acidity and tannin-related genes, e.g., *PH4* and *Myb9-like*, from 3.41 to 3.76 Mb on Chr 15. It was also shown in tomato that selection for alleles of genes associated with larger fruits altered metabolite profiles as a consequence of linkage with nearby genes [21]. However, fruit acidity and astringency are genetically controlled by other major genes such as *Ma1* and *LAR* on Chr 16, whereas fruit weight is regulated by multiple genes [26]. Thus, fruit enlargement may partially affect the acidity and astringency in apples.

The control of firmness has been a major goal for improving fruit storage and shelf life [70, 71]. The loss of turgor pressure in fruits leads to a decrease in firmness [72], in addition to changes in the cell wall and the action of antioxidants [73]. The integrity of the fruit cell membranes can be damaged during storage, particularly at low temperatures, causing the loss of turgor pressure. We demonstrated that lysolipid LPE 18:1 is potentially maintaining the firmness of apple, with the “Ralls Janet” pedigree identified as a variety with firm flesh, prolonged storage life, and relatively high LPE 18:1 content (Fig. 5a, d). Previous reports have depicted that unsaturation of fatty acids in lipids, such as LPE and lysophosphatidylcholine (LPC), could improve cell membrane stability [35–37, 74] and mediate hormonal regulation of growth and senescence [37, 38], while saturated fatty acids produce the opposite effect [75, 76]. The ratio of unsaturated to total fatty acids has been reported greater in “Ralls Janet” than in “Starking Delicious,” which increases as the storage temperature decreases [77]. Other research has also found that pre- and postharvest LPE treatments promote fruit color and enhance shelf life [45]. Particularly, the LPE-treated banana had higher pulp firmness, lower ion leakage from peel tissue, a thicker peel, and better shelf life as compared to the control [78]. Thus, it can be concluded that fruits with a higher content of lysolipids, particularly LPE 18:1, have better storage and shelf life.

A comparison of the metabolites in low- and high-weight domesticated groups demonstrated that the increment of SA and ABA is associated with fruit enlargement. The growth phase of fruit involves an active period of cell division, followed by a cell



expansion phase, and a combination of these determines the maximum final fruit size [79]. Recent studies have reinforced the importance of SA and ABA in regulating plant cell division or organ enlargement. For example, the dwarf phenotype was identified in *Arabidopsis* mutants with high SA content, such as *constitutive pathogen response* (*Scpr5*) [80], *accelerated cell death* (*acd6-1*) [81], and *aberrant growth and death* (*agd2*) [82]. ABA has been reported to be crucial in inhibiting cell division in plants and tissues [48, 83, 84]. Our results demonstrate that smaller fruits contain higher SA and ABA contents than bigger fruits. ABA was also found to promote ripening in some fruits [85], suggesting the possible dual roles of ABA and SA in regulating fruit size, development, and ripening. We also found evidence that the NAP transcription factor is associated with fruit size in apples based on the accumulation of SA. Previous research reported that tomato fruits that overexpressed *SINAP2* were smaller than the wild-type despite the increase in the fruit yield [49]. Analysis of different hormones revealed higher SA and ABA contents in the *SINAP1* overexpressing tomato plants. Moreover, *SINAP1* was observed to directly activate the expression of genes in both SA and ABA biosynthetic pathways [50], implying that *NAP* may affect both SA and ABA synthesis in apples. However, the mechanisms involved in interactions between SA and ABA and their role in fruit enlargement requires further research.

In addition to providing information on metabolic divergences in well-characterized pathways, the data can also provide insight into unknown pathways. Metabolomic or sequencing data alone can provide valuable information on plant metabolism, but the combination of these two data can create a synergistic effect, which is beneficial for the discovery of metabolic pathways [86]. In particular, these approaches can reveal novel metabolites and genes that are involved in plant metabolism. These data have also laid the foundation for cross-species comparative approaches that can provide insights into the evolution of metabolism in other crops [87]. However, we should also understand that although the multiomics method provides large data, most of them are difficult to utilize currently. For example, of the 2575 distinct metabolites identified, only 486 (18.87%) could be annotated (Additional file 2: Table S4). Therefore, the characterization of these chemicals through biochemical approaches is essential for the identification of the underlying candidate genes, which will be possible to construct novel key biological pathways of apples in the future.

## Conclusion

Metabolite changes in apple accessions have accompanied human-guided evolutionary selection processes. We have associated significant SNPs with apple metabolites. Important flavor compounds, namely organic acids, soluble tannins, flavonoids, and phenolic acids, were co-mapped to a 2.84~5.01 Mb region on Chr 16. In addition, the fruit weight locus was co-mapped with organic acid and tannin-related genes in a 3.41~3.76 Mb region on Chr 15. LPE 18:1, which contents were reduced by the activity of *FAD2* and found higher in the “Ralls Janet” pedigree than the “Golden delicious” pedigree, was also found beneficial for maintaining firmness and storability. The levels of two plant hormones, SA and ABA, with higher levels detected in smaller fruits (Fig. 6a–c), and their accumulation was affected by *NAP* and *ABCG25*, respectively. In summary, our results provide a metabolic perspective of selection for fruit quality related to domestication



and improvement in apples. The information offered in this article can be used to investigate the mechanisms underlying apple metabolite content and quality attributes.

## Materials and methods

### Fruit material and treatment

A total of 292 accessions with a wide range of geographic origins (Additional file 2: Table S1) were used in this study, including 270 collected from the apple germplasm repository in Liaoning, China, and 22 wild accessions reported previously [7, 26]. In total, 217 cultivars, 38 heirlooms, 11 *M. sieversii*, seven *M. sylvestris*, and 19 accessions of other wild species, including three *M. baccata*, three *M. sieboldii*, seven *M. spectabilis*, and six *M. toringoides* (Additional file 2: Table S2), were investigated in this study. The fruit flesh was frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ .

For developmental metabolites analysis and postharvest treatment, “Hongyu” and “Molisi” apples were collected from an orchard in Beijing, China. The “Hongyu” apples were collected at 40, 80, 100, and 120 DAFB. For LPE 18:1 treatment, “Molisi” apples at the commercial ripening stage were randomly divided into two groups and were separately dipped in water (as a control) and  $100\text{ mg L}^{-1}$  of LPE 18:1 for 10 min. The physical data for “Molisi” were measured at 0, 3, 7, 14, 21, and 28 days of storage. All the fruits were stored at  $20 \pm 1^{\circ}\text{C}$  with 90% relative humidity. Each treatment involved 120 fruit samples with three replicates.

### Physiological index measurements

Firmness was evaluated by a texture analyzer (TA.HD plus, Stable Micro Systems, UK) with a probe of 7.5 mm diameter, and results were expressed in Newton’s (N) force. The measurement speed and distance were set at  $2\text{ mm s}^{-1}$  and 10 mm, respectively. Two apples were sealed in a 1.5-L chamber for 2 h at  $20^{\circ}\text{C}$ ; then, the  $\text{CO}_2$  level was monitored by a gas analyzer (F950, FELIXD, USA) to calculate the respiration rate presented as  $\text{mL kg}^{-1}\text{ h}^{-1}$ . To measure titratable acid, 1 g of flesh powder was mixed with 10 mL of water and centrifuged for 10 min at  $12,000 \times g$  at  $4^{\circ}\text{C}$ . The supernatant was titrated to pH 8.2 with  $0.01\text{ mol L}^{-1}$  of sodium hydroxide. ABA and SA contents were measured by enzyme-linked immunosorbent assay (ELISA) kits (Welab, China). Each measurement contained three biological replicates.

### Metabolic profiling

Apple flesh was freeze-dried and ground for 1 min at 30 Hz using a mixer mill (MM 400, Retsch, Germany) with zirconia beads. A total of 80 mg of flesh powder was homogenized with 1 mL of 70% aqueous methanol at  $4^{\circ}\text{C}$  overnight. After centrifuging for 10 min at 12,000 rpm, supernatants were filtered using a filter with a pore size of  $0.22\text{ }\mu\text{m}$  (SCAA-104, ANPEL, China). The samples were measured by a high-performance liquid chromatography (HPLC)-electrospray ionization (ESI)-mass spectrometer (MS)/MS system (HPLC, Shim-pack UFLC Shimadzu CBM30A system, China; MS, Applied Biosystems 6500 plus Q TRAP, China). For HPLC system, Waters ACQUITY UPLC HSS T3 C18 ( $1.8\text{ }\mu\text{m}$ ,  $2.1\text{ mm} \times 100\text{ mm}$ ) column was used. Water (0.04% acetic acid) and acetonitrile (0.04% acetic acid) were used as a solvent system. The program was set as follows: 95:5 v/v at 0 min, 5:95 v/v at 11.0 min, 5:95 v/v at 12.0 min, 95:5 v/v at 12.1 min,

95:5 v/v at 15.0 min at a flow rate of 0.35 mL min<sup>-1</sup>, the temperature of 40 °C, and injection volume of 2 µL. The effluent was connected to an ESI-triple quadrupole-linear ion trap (Q TRAP) MS.

Linear ion trap (LIT) and triple quadrupole (QQQ) scans were collected on a Q TRAP MS using an API 6500 Q TRAP LC/MS/MS System equipped with an ESI Turbo Ion-Spray interface, operated in a positive ion mode and controlled by Analyst 1.6 software (AB Sciex, Germany). Operation parameters of the ESI source were as follows: turbo ion spray; source temperature of 550 °C; ion spray voltage (IS) of (+) 5500 V and (–) 4500 V; ion source gas I (GSI), gas II (GSI), and curtain gas (CUR) were set at 55, 60, and 25 psi, respectively; and collision gas (CAD) set at medium. Instrument tuning and mass calibration were performed using 10 and 100 µmol L<sup>-1</sup> polypropylene glycol solutions in QQQ and LIT modes, respectively. QQQ scans were obtained from multiple reaction monitoring (MRM) experiments with the collision gas (nitrogen) set to 5 psi. By further decluttering potential (DP) and collision energy (CE) optimization, the DP and CE for individual MRM transitions were performed. A specific set of MRM transitions was monitored for each period based on the metabolites eluted during this period.

### Genome sequencing

Genomic DNA was obtained using a Plant DNA Mini Kit (Aidlab Biotech Ltd., China). After constructing the DNA library, an Illumina HiSeq XTen/NovaSeq/BGI platform was used for sequencing, with 150 bp read lengths, conducted by a credible service provider (Biomarker Technologies, China). High-quality sequences were obtained by removing reads containing the sequencing adapters or that of low quality based on the following criteria: pair-end reads with > 10% “N” bases or reads with more than 50% of the bases have a Phred-like score < 20.

### SNP calling

All clean reads of 292 apple accessions were mapped to the reference genome (GDDH13 v.1.1). The mapping results were sorted, and duplications were removed by rmdup in SAMTools (v.1.9) [88]. The HaplotypeCaller module in GATK v.3.8 was used for calling of SNPs within the whole accessions, which were filtered according to the parameters: QD < 2.0, MQ < 40.0, FS > 60.0, QUAL < 30.0, MQRankSum < -12.5, ReadPos-RankSum < -8.0, -clusterSize 2 -clusterWindowSize 5. The high-quality SNPs were further selected with the following criteria: minor allele frequency (MAF) ≥ 5% and missing rate per site ≤ 10%. Based on the reference genome, SNP annotation was conducted by snpEff software [89], and SNPs were classified as intergenic, upstream or downstream regions, and exons or introns. SNPs identified in coding exons were further categorized into synonymous or non-synonymous SNPs.

### Population genetics analysis and selective sweeps identification

The phylogenetic tree was constructed from the distance matrix in PHYLIP software v.3.69 with 1000 bootstrap replicates and was visualized by iTOL [90]. PCA of whole-genome SNPs was performed with EIGENSOFT 6.0 software smartpca program, and the first three eigenvectors were plotted in two/ three dimensions [91]. Population structure was modeled by ADMIXTURE v.1.3.0 [92], and TASSEL v.3.0 [93] was used

to calculate kinship value. LD coefficients based on  $r^2$  values between pair-wised high-quality SNPs were calculated by PopLDdecay software with default parameters [94]. The potential selective signatures were estimated by the genetic differentiation index ( $F_{ST}$ ) for *M. sieversii*/ heirloom, *M. sylvestris*/ heirloom, and heirloom/ cultivar comparisons. The region of sweeps associated with domestication and improvement were selected within the top 5% of windows of  $F_{ST}$ .

#### Metabolites-based GWAS analysis

A total of 2,956,396 SNPs were used to carry out GWAS. mGWAS was conducted by the linear mixed model implemented in GEMMA. The genome-wide significance thresholds were determined using the Bonferroni test threshold ( $p = 1.71E^{-7}$ ).

#### RNA extraction and quantitative real-time PCR (qRT -PCR)

Total RNA was extracted from the apple flesh using the cetyltrimethylammonium bromide (CTAB) method [95]. cDNA was synthesized, and qRT-PCR was performed on a Fast Real-Time PCR System (Applied Biosystems 7500, USA). The primers are listed in Additional file 2: Table S11.

#### Transient overexpression

The coding sequences were cloned into NotI/ SmaI sites of a pGreenII 0029 62-SK vector, which were subsequently transformed into *Agrobacterium tumefaciens* strain GV3101. *Agrobacterium* cells containing the overexpressed constructs were separately infiltrated into the apples. The opposite sides of each apple were infiltrated with cells containing the empty vector as controls. Tissues around the infiltration sites were collected after 3 days in the chamber.

#### GUS transactivation assay

The coding sequence of *Myb9-like* was cloned into the binary pRI101 vector to generate an effector construct. The reporter construct was cloned, according to the *LAR* promoter sequence, upstream of the GUS reporter gene in the binary vector pBI101. The reporter vector and the effector vector were transformed into *A. tumefaciens* GV3101. The opposite side of each apple was infiltrated with cells containing the pRI101 vector and reporter construct as control. Tissues around the infiltration sites were collected after 3 days in the chamber.

#### Statistical analyses

Coefficients of variations (CV) of each metabolite were independently calculated as follows:  $s/m$ , where “s” and “m” represent the standard deviation and mean of each metabolite in all accessions, respectively. Differential metabolites between different groups were determined by partial least squares to discriminate analysis (PLS-DA) with the variable importance for projection (VIP) values  $\geq 1$ , followed by both ANOVA ( $p \leq 0.05$ ), using SIMCA software package (v.14.1). Significant differences were determined by Fisher's least significant difference (LSD) test at the level of  $p \leq 0.01$ .

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-023-02945-6>.

**Additional file 1: Fig. S1.** Classification and coefficient of variation values for all metabolites detected in apple flesh. **Fig. S2.** PCA analysis of all metabolite contents detected in other wild, *M. sieversii*, heirloom, and six apple cultivar groups. **Fig. S3.** Contents of total tannins and titratable acid in apple accessions among *M. sieversii* group, low-weight, and high-weight domesticated groups. **Fig. S4.** Co-mapped regions of soluble tannins identified by GWAS on Chr 16. **Fig. S5.** Differences between amino acid sequences of MdMyb9 and MD15G1051400. **Fig. S6.** Whole-genome screen for selective sweeps during apple domestication and improvement. The horizontal gray dotted lines indicate genome-wide thresholds that were estimated based on the top 5% of *Fst* values. **Fig. S7.** Co-mapped regions of LPE identified by GWAS on Chr 12.

**Additional file 2: Table S1.** The list of collected 292 varieties. **Table S2.** Classification of 292 apple accessions. **Table S3.** Statistical results of heredity and variation. **Table S4.** Scheduled MRM transitions for widely targeted metabolite analysis in apple flesh. **Table S5.** Different metabolites identified between *M. sieversii* and heirloom groups. **Table S6.** Different metabolites identified between heirloom and cultivar groups. **Table S7.** Different metabolites identified among "Ralls Janet" and "Golden Delicious" series. **Table S8.** Different metabolites identified between low- and high-weight groups. **Table S9.** All significant SNPs for all apple metabolites. **Table S10.** Candidate genes for tannins, lipids, SA, and ABA identified by GWAS. **Table S11.** Primers used in this study.

**Additional file 3.** Review history.

### Acknowledgements

We would like to thank Prof. Kun Wang, Dr. Yuan Gao, and Dr. Dajiang Wang from the National Repository of Apple and Pear Germplasm Resources for providing the research materials. We thank Ang Li, Yifeng Chai, Congcong Song, and Ting Fang from the Institute of Food Science and Technology, Chinese Academy of Agricultural Sciences for their efforts in sample collection. We would like to thank Dr. Chao Feng from South China Botanical Garden, Chinese Academy of Sciences for his helpful comments during the preparation of this article. We acknowledge the direct and indirect support of researchers and non-scientific staff.

### Peer review information

Wenjing She was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

### Review history

The review history is available as Additional file 3.

### Authors' contributions

Q.L., X.Liu, B.W., and J.B. designed and managed the project. Q.L., J.C., X.Liu, B.W., Y.Z., L.L., Y.D., X.Li, Y.H., and J.B. collected samples and performed experiments. Q.L., J.C., X.Liu, B.W., Y.Z., L.L., A.A., C.S., J.V., Y.H., and J.B. performed data analyses. Q.L., J.C., X.Liu, B.W., L.L., A.A., D.G., Y.H., and J.B. wrote the manuscript. Q.L., B.W., L.L., A.A., C.S., D.G., K.C., Y.H., and J.B. revised the manuscript. All authors read and approved the manuscript.

### Funding

We acknowledge the financial support from the National Key Research and Development Program of China (2022YFD2100100), the earmarked fund for China Agriculture Research System (CARS-27), the Agricultural Science and Technology Innovation Program of the Institute of Food Science and Technology, Chinese Academy of Agricultural Sciences (CAAS-ASTIP-G2022-IFST-02), and the China Scholarship Council (201903250033).

### Availability of data and materials

The datasets generated during the current study are available in the Genome Sequence Archive (GSA) database in the BIG Data Center under accession numbers CRA009278 [96] and CRA009404 [97]. Raw genomes of 11 samples were obtained from the GSA database in the BIG Data Center under accession number CRA003964 [98]. Raw genomes of 12 samples were obtained from the NCBI Sequence Read Archive under accession SRP075497 [99]. The metabolic profiling data in this paper has been deposited in the OMIX database, China National Center for Bioinformatics/ Beijing Institute of Genomics, Chinese Academy of Sciences under accession number OMIX003609 [100]. Details about the samples can also be found in Additional file 2: Table S1. All data generated or analyzed during this study are included in this published article and its supplementary information files. We confirm that no other scripts and software were used other than those mentioned in the "Materials and methods" section.

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

Received: 2 September 2022 Accepted: 18 April 2023

Published online: 26 April 2023

## References

1. Eberhardt MV, Lee CY, Liu RH. Antioxidant activity of fresh apples. *Nature*. 2000;405:903–4.
2. Bondonno NP, Bondonno CP, Ward NC, Hodgson JM, Croft KD. The cardiovascular health benefits of apples: whole fruit vs. isolated compounds. *Trends Food Sci Tech*. 2017;69:243–56.
3. Food and Agriculture Organization Corporate Statistical Database. 2022. <http://www.fao.org/faostat/en/#data/QCL>. Accessed 5 July 2022.
4. O'Rourke D. Economic importance of the world apple industry. In: Korban SS, editor. *The Apple Genome*. Springer; 2021. p. 1–18.
5. Daccord N, Celton JM, Linsmith G, Becker C, Choisine N, Schijlen E, et al. High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. *Nat Genet*. 2017;49:1099–106.
6. Harris SA, Robinson JP, Juniper BE. Genetic clues to the origin of the apple. *Trends Genet*. 2002;18:426–30.
7. Duan N, Bai Y, Sun H, Wang N, Ma Y, Li M, et al. Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. *Nat Commun*. 2017;8:1–11.
8. Sun X, Jiao C, Schwaninger H, Chao CT, Ma Y, Duan N, et al. Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. *Nat Genet*. 2020;52:1423–32.
9. Kader AA. Flavor quality of fruits and vegetables. *J Sci Food Agr*. 2008;88:1863–8.
10. Ma B, Chen J, Zheng H, Fang T, Ogutu C, Li S, et al. Comparative assessment of sugar and malic acid composition in cultivated and wild apples. *Food Chem*. 2015;172:86–91.
11. Yao JL, Xu J, Cornille A, Tomes S, Karunairatnam S, Luo Z, et al. A micro RNA allele that emerged prior to apple domestication may underlie fruit size evolution. *Plant J*. 2015;84:417–27.
12. Cornille A, Giraud T, Smulders MJ, Roldán-Ruiz I, Gladieux P. The domestication and evolutionary ecology of apples. *Trends Genet*. 2014;30:57–65.
13. Morgan J, Richards A, Selborne TEO, Dowle E. *The new book of apples: the definitive guide to over 2000 varieties*. Random House; 2013.
14. Di Guardo M, Bink MC, Guerra W, Letschka T, Lozano L, Busatto N, et al. Deciphering the genetic control of fruit texture in apple by multiple family-based analysis and genome-wide association. *J Exp Bot*. 2017;68:1451–66.
15. Costa F, Cappellin L, Fontanari M, Longhi S, Guerra W, Magnago P, et al. Texture dynamics during postharvest cold storage ripening in apple (*Malus × domestica* Borkh.). *Postharvest Biol Tec*. 2012;69:54–63.
16. Hyson DA. A comprehensive review of apples and apple components and their relationship to human health. *Adv Nutr*. 2011;2:408–20.
17. Colantonio V, Ferrão LFV, Tieman DM, Bliznyuk N, Sims C, Klee HJ, et al. Metabolomic selection for enhanced fruit flavor. *P Natl Acad Sci USA*. 2022;119:e2115865119.
18. Fang C, Luo J. Metabolic GWAS-based dissection of genetic bases underlying the diversity of plant metabolism. *Plant J*. 2019;97:91–100.
19. Tohge T, Fernie AR. Combining genetic diversity, informatics and metabolomics to facilitate annotation of plant gene function. *Nat Protoc*. 2010;5:1210–27.
20. Wei G, Tian P, Zhang F, Qin H, Miao H, Chen Q, et al. Integrative analyses of nontargeted volatile profiling and transcriptome data provide molecular insight into VOC diversity in cucumber plants (*Cucumis sativus*). *Plant Physiol*. 2016;172:603–18.
21. Zhu G, Wang S, Huang Z, Zhang S, Liao Q, Zhang C, et al. Rewiring of the fruit metabolome in tomato breeding. *Cell*. 2018;172:249–61.
22. Cao K, Wang B, Fang W, Zhu G, Chen C, Wang X, et al. Combined nature and human selections reshaped peach fruit metabolome. *Genome Biol*. 2022;23:1–25.
23. Fang H, Shen S, Wang D, Zhang F, Zhang C, Wang Z, et al. A monocot-specific hydroxycinnamoylputrescine gene cluster contributes to immunity and cell death in rice. *Sci Bull*. 2021;66:2381–93.
24. Zhang Z, Shi Q, Wang B, Ma A, Wang Y, Xue Q, et al. Jujube metabolome selection determined the edible properties acquired during domestication. *Plant J*. 2021;109:1116–33.
25. Farneti B, Di Guardo M, Khomenko I, Cappellin L, Biasioli F, Velasco R, et al. Genome-wide association study unravels the genetic control of the apple volatilome and its interplay with fruit texture. *J Exp Bot*. 2017;68:1467–78.
26. Liao L, Zhang W, Zhang B, Fang T, Wang X-F, Cai Y, et al. Unraveling a genetic roadmap for improved taste in the domesticated apple. *Mol Plant*. 2021;14:1454–71.
27. Chen P, Li Z, Zhang D, Shen W, Xie Y, Zhang J, et al. Insights into the effect of human civilization on *Malus* evolution and domestication. *Plant Biotechnol J*. 2021;19:2206–20.
28. Minamikawa MF, Kunihiisa M, Noshita K, Moriya S, Abe K, Hayashi T, et al. Tracing founder haplotypes of Japanese apple varieties: application in genomic prediction and genome-wide association study. *Hortic Res-England*. 2021;8:49.
29. Chen W, Gong L, Guo Z, Wang W, Zhang H, Liu X, et al. A novel integrated method for large-scale detection, identification, and quantification of widely targeted metabolites: application in the study of rice metabolomics. *Mol Plant*. 2013;6:1769–80.
30. Wu W, Zhu QG, Wang WQ, Grierson D, Yin XR. Molecular basis of the formation and removal of fruit astringency. *Food Chem*. 2021;372:131234.
31. Huang R, Xu C. An overview of the perception and mitigation of astringency associated with phenolic compounds. *Compr Rev Food Sci F*. 2021;20:1036–74.
32. Drewnowski A. The science and complexity of bitter taste. *Nutr Rev*. 2001;59:163–9.

33. Sakurai K, Brown SK, Weeden N. Self-incompatibility alleles of apple cultivars and advanced selections. *HortScience*. 2000;35:116–9.
34. Lau O, Liu Y, Yang S. Effects of fruit detachment on ethylene biosynthesis and loss of flesh firmness, skin color, and starch in ripening Golden Delicious apples. *J Am Soc Hortic Sci*. 1986;111:731–4.
35. Cowan AK, Leung C, Santori C. Lyso-phosphatidylethanolamine (LPE) as a plant bioregulator. *Acta Hortic*. 2006;727:527–36.
36. Ryu SB, Karlsson BH, Özgen M, Palta JP. Inhibition of phospholipase D by lysophosphatidylethanolamine, a lipid-derived senescence retardant. *P Natl Acad Sci USA*. 1997;94:12717–21.
37. Rietz S, Dermendjiev G, Oppermann E, Tafesse FG, Effendi Y, Holk A, et al. Roles of Arabidopsis patatin-related phospholipases a in root development are related to auxin responses and phosphate deficiency. *Mol Plant*. 2010;3:524–38.
38. Watanabe M, Balazadeh S, Tohge T, Erban A, Giavalisco P, Kopka J, et al. Comprehensive dissection of spatiotemporal metabolic shifts in primary, secondary, and lipid metabolism during developmental senescence in Arabidopsis. *Plant Physiol*. 2013;162:1290–310.
39. Khan SA, Schaart JG, Beekwilder J, Andrew CA, Yury MT, Evert J, et al. The mQTL hotspot on linkage group 16 for phenolic compounds in apple fruits is probably the result of a *leucoanthocyanidin reductase* gene at that locus. *BMC Res Notes*. 2012;5:618.
40. Schaart JG, Dubos C, Romero De La Fuente I, van Houwelingen AM, de Vos RC, Jonker HH, et al. Identification and characterization of MYB-bHLH-WD40 regulatory complexes controlling proanthocyanidin biosynthesis in strawberry (*Fragaria × ananassa*) fruits. *New Phytol*. 2013;197:454–67.
41. An XH, Tian Y, Chen KQ, Liu XJ, Liu DD, Xie XB, et al. *MdMYB9* and *MdMYB11* are involved in the regulation of the JA-induced biosynthesis of anthocyanin and proanthocyanidin in apples. *Plant Cell Physiol*. 2015;56:650–62.
42. Quattrocchio F, Verweij W, Kroon A, Spelt C, Mol J, Koes R. PH4 of Petunia is an R2R3 MYB protein that activates vacuolar acidification through interactions with basic-helix-loop-helix transcription factors of the anthocyanin pathway. *Plant Cell*. 2006;18:1274–91.
43. Pang Y, Abeysinghe ISB, He J, He X, Huhman D, Mewan KM, et al. Functional characterization of proanthocyanidin pathway enzymes from tea and their application for metabolic engineering. *Plant Physiol*. 2013;161:1103–16.
44. Yu K, Jun JH, Duan C, Dixon RA. *VvLAR1* and *VvLAR2* are bifunctional enzymes for proanthocyanidin biosynthesis in grapevine. *Plant Physiol*. 2019;180:1362–74.
45. Amaro AL, Almeida DP. Lysophosphatidylethanolamine effects on horticultural commodities: A review. *Postharvest Biol Tec*. 2013;78:92–102.
46. Nakamura Y. Plant phospholipid diversity: emerging functions in metabolism and protein-lipid interactions. *Trends Plant Sci*. 2017;22:1027–40.
47. Rivas-San Vicente M, Plasencia J. Salicylic acid beyond defence: its role in plant growth and development. *J Exp Bot*. 2011;62:3321–38.
48. Myers PN, Setter TL, Madison JT, Thompson JF. Abscisic acid inhibition of endosperm cell division in cultured maize kernels. *Plant Physiol*. 1990;94:1330–6.
49. Ma X, Zhang Y, Turečková V, Xue GP, Fernie AR, Mueller-Roeber B, et al. The NAC transcription factor *SINAP2* regulates leaf senescence and fruit yield in tomato. *Plant Physiol*. 2018;177:1286–302.
50. Wang J, Zheng C, Shao X, Hu Z, Li J, Wang P, et al. Transcriptomic and genetic approaches reveal an essential role of the NAC transcription factor *SINAP1* in the growth and defense response of tomato. *Hortic Res-England*. 2020;7:209.
51. Park Y, Xu ZY, Kim SY, Lee J, Choi B, Lee J, et al. Spatial regulation of ABCG25, an ABA exporter, is an important component of the mechanism controlling cellular ABA levels. *Plant Cell*. 2016;28:2528–44.
52. Jia HF, Chai YM, Li CL, Lu D, Luo JJ, Qin L, et al. Abscisic acid plays an important role in the regulation of strawberry fruit ripening. *Plant Physiol*. 2011;157:188–99.
53. Obata T, Fernie AR. The use of metabolomics to dissect plant responses to abiotic stresses. *Cell Mol Life Sci*. 2012;69:3225–43.
54. Sulpice R, Mckeown PC. Moving toward a comprehensive map of central plant metabolism. *Annu Rev Plant Biol*. 2015;66:187–210.
55. Diego HS, Henning R, Ute K, Michael KU, Joachim K. Metabolome-ionome-biomass interactions: what can we learn about salt stress by multiparallel phenotyping? *Plant Signal Behav*. 2008;3:598–600.
56. Zhang L, Hu J, Han X, Li J, Gao Y, Richards CM, et al. A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. *Nat Commun*. 2019;10:1–13.
57. Bai Y, Dougherty L, Li M, Fazio G, Cheng L, Xu K. A natural mutation-led truncation in one of the two aluminum-activated malate transporter-like genes at the Ma locus is associated with low fruit acidity in apple. *Mol Genet Genomics*. 2012;287:663–78.
58. Chagné D, Krieger C, Rassam M, Sullivan M, Fraser J, André C, et al. QTL and candidate gene mapping for polyphenolic composition in apple fruit. *BMC Plant Biol*. 2012;12:1–16.
59. Bilbrey EA, Williamson K, Hatzakis E, Miller DD, Fresnedo-Ramírez J, Cooperstone JL. Integrating genomics and multiplatform metabolomics enables metabolite quantitative trait loci detection in breeding-relevant apple germplasm. *New Phytol*. 2021;232:1944–58.
60. Bellucci E, Bitocchi E, Ferrarini A, Benazzo A, Biagetti E, Klie S, et al. Decreased nucleotide and expression diversity and modified coexpression patterns characterize domestication in the common bean. *Plant Cell*. 2014;26:1901–12.
61. Koenig D, Jiménez-Gómez JM, Kimura S, Fulop D, Chitwood DH, Headland LR, et al. Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *P Natl Acad Sci USA*. 2013;110:E2655–62.
62. Beleggia R, Rau D, Laidò G, Platani C, Nigro F, Fragasso M, et al. Evolutionary metabolomics reveals domestication-associated changes in tetraploid wheat kernels. *Mol Biol Evol*. 2016;33:1740–53.
63. Zhang W, Alseekh S, Zhu X, Zhang Q, Fernie AR, Kuang H, et al. Dissection of the domestication-shaped genetic architecture of lettuce primary metabolism. *Plant J*. 2020;104:613–30.



64. Zhen Y, Zheng LJ, Nie JY, Li ZX, Cheng Y. Evaluation indices of sour flavor for apple fruit and grading standards. *J Integr Agr*. 2018;17:994–1002.
65. Wang L, He F, Huang Y, He J, Yang S, Zeng J, et al. Genome of wild mandarin and domestication history of mandarin. *Mol Plant*. 2018;11:1024–37.
66. Huang J, Zhang C, Zhao X, Fei Z, Wan K, Zhang Z, et al. The jujube genome provides insights into genome evolution and the domestication of sweetness/acidity taste in fruit trees. *Plos Genet*. 2016;12:e1006433.
67. Wu J, Wang Y, Xu J, Korban SS, Fei Z, Tao S, et al. Diversification and independent domestication of Asian and European pears. *Genome Biol*. 2018;19:1–16.
68. Misha P, Ronald K, Francesca MQ. Alteration of flavonoid pigmentation patterns during domestication of food crops. *J Exp Bot*. 2019;70:3719–35.
69. Frydman A, Liberman R, Huhman DV, Carmeli-Weissberg M, Sapir-Mir M, Ophir R, et al. The molecular and enzymatic basis of bitter/non-bitter flavor of citrus fruit: evolution of branch-forming rhamnosyltransferases under domestication. *Plant J*. 2013;73:166–78.
70. Chang H, Xu J, Macqueen LA, Aytac Z, Peters MM, Zimmerman JF, et al. High-throughput coating with biodegradable antimicrobial pullulan fibres extends shelf life and reduces weight loss in an avocado model. *Nat Food*. 2022;3:428–36.
71. Konopacka D, Plocharski W. Effect of storage conditions on the relationship between apple firmness and texture acceptability. *Postharvest Biol Tec*. 2004;32:205–11.
72. Khin MM, Zhou W, Yeo SY. Mass transfer in the osmotic dehydration of coated apple cubes by using maltodextrin as the coating material and their textural properties. *J Food Eng*. 2007;81:514–22.
73. Shi Y, Li BJ, Su G, Zhang M, Grierson D, Chen KS. Transcriptional regulation of fleshy fruit texture. *J Integr Plant Biol*. 2022;0:1–24.
74. Holk RUPA, Scherer GF. Fatty acids and lysophospholipids as potential second messengers in auxin action. Rapid activation of phospholipase A2 activity by auxin in suspension-cultured parsley and soybean cells. *Plant J*. 1998;16:601–11.
75. He Y, Li Z, Tan F, Liu H, Zhu M, Yang H, et al. Fatty acid metabolic flux and lipid peroxidation homeostasis maintain the biomembrane stability to improve citrus fruit storage performance. *Food Chem*. 2019;292:314–24.
76. Pourcel L, Routaboul JM, Kerhoas L, Caboche M, Lepiniec L, Debeaujon I. TRANSPARENT TESTA10 encodes a laccase-like enzyme involved in oxidative polymerization of flavonoids in Arabidopsis seed coat. *Plant Cell*. 2005;17:2966–80.
77. Kimura S, Kanno M, Yamada Y, Takahashi K, Murashige H, Okamoto T. The contents of conjugated lipids and fatty acids and the cold tolerance in the mitochondria of Starking Delicious and Ralls Janet apples. *Agr Bio Chem*. 1982;46:2895–902.
78. Ahmed Z, Palta JP. Postharvest dip treatment with a natural lysophospholipid plus soy lecithin extended the shelf life of banana fruit. *Postharvest Biol Tec*. 2016;113:58–65.
79. Gillaspay G, Ben-David H, Gruitsem W. Fruits: a developmental perspective. *Plant Cell*. 1993;5:1439–51.
80. Bowling SA, Clarke JD, Liu Y, Klessig DF, Dong X. The *cpr5* mutant of Arabidopsis expresses both NPR1-dependent and NPR1-independent resistance. *Plant Cell*. 1997;9:1573–84.
81. Rate DN, Cuenca JV, Bowman GR, Guttman DS, Greenberg JT. The gain-of-function Arabidopsis *acd6* mutant reveals novel regulation and function of the salicylic acid signaling pathway in controlling cell death, defenses, and cell growth. *Plant Cell*. 1999;11:1695–708.
82. Rate DN, Greenberg JT. The Arabidopsis aberrant growth and *death2* mutant shows resistance to *Pseudomonas syringae* and reveals a role for NPR1 in suppressing hypersensitive cell death. *Plant J*. 2001;27:203–11.
83. Zeevaert JA, Creelman RA. Metabolism and physiology of abscisic acid. *Annu Rev Plant Phys*. 1988;39:439–73.
84. Finkelstein RR, Gampala SS, Rock CD. Abscisic acid signaling in seeds and seedlings. *Plant Cell*. 2002;14:S15–45.
85. Li S, Chen K, Grierson D. Molecular and hormonal mechanisms regulating fleshy fruit ripening. *Cells*. 2021;10:1136.
86. Bernonville T, Papon N, Clastre M, O'Connor SE, Courdavault V. Identifying missing biosynthesis enzymes of plant natural products. *Trends Pharmacol Sci*. 2020;41:142–6.
87. Giovannoni J. Tomato multiomics reveals consequences of crop domestication and improvement. *Cell*. 2018;172:6–8.
88. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAM-tools. *Bioinformatics*. 2009;25:2078–9.
89. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang 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*. 2012;6:80–92.
90. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res*. 2019;47:W256–9.
91. Herrando-Pérez S, Tobler R, Huber CD. Smartsnp, an R package for fast multivariate analyses of big genomic data. *Methods Ecol Evol*. 2021;12:2084–93.
92. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19:1655–64.
93. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*. 2007;23:2633–5.
94. Zhang C, Dong SS, Xu JY, He WM, Yang TL. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics*. 2019;35:1786–8.
95. Li A, Chen J, Lin Q, Zhao Y, Duan Y, Wai SC, et al. Transcription factor *MdWRKY32* participates in starch-sugar metabolism by binding to the *MdBam5* promoter in apples during postharvest storage. *J Agr Food Chem*. 2021;69:14906–14.
96. Lin Q, Chen J, Liu X, Wang B, Zhao YY, Liao L, et al. A metabolic perspective of selection for fruit quality related to apple domestication and improvement. Datasets. NGDC Genome Sequence Archive. 2023; CRA0092784. <https://ngdc.cncb.ac.cn/gsa/browse/CRA009278>.

97. Lin Q, Chen J, Liu X, Wang B, Zhao YY, Liao L, et al. A metabolic perspective of selection for fruit quality related to apple domestication and improvement. Datasets. NGDC Genome Sequence Archive. 2023; CRA009404. <https://ngdc.cncb.ac.cn/gsa/browse/CRA009404>.
98. Liao L, Zhang W, Zhang B, Fang T, Wang XF, Cai Y, et al. Unraveling a genetic roadmap for improved taste in the domesticated apple. Datasets. NGDC Genome Sequence Archive. 2021; CRA003964. <https://ngdc.cncb.ac.cn/gsa/browse/CRA003964>.
99. Duan N, Bai Y, Sun H, Wang N, Ma Y, Li M, et al. Genome re-sequencing reveals the history of apple and supports a two-stage model for fruit enlargement. Datasets. NCBI Sequence Read Archive. 2017; PRJNA322175. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA322175>.
100. Lin Q, Chen J, Liu X, Wang B, Zhao YY, Liao L, et al. A metabolic perspective of selection for fruit quality related to apple domestication and improvement. Datasets. NGDC OMIX. 2023; OMIX003609. <https://ngdc.cncb.ac.cn/omix/release/OMIX003609>.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

