

Identification of favorable SNP alleles related to fruit traits in diverse apple germplasm

A Reum Han

Kyungpook National University College of Agriculture and Life Sciences

Seon-Ae Kim

Apple Research Institute, National Institute of Horticultural and Herbal Science, RDA

Gi Hoon Kim

Kyungpook National University College of Agriculture and Life Sciences

Cheol Choi (✉ cc31@knu.ac.kr)

Kyungpook National University College of Agriculture and Life Sciences <https://orcid.org/0000-0002-7102-0151>

Research article

Keywords: Apple, Fruit quality, Genotyping by sequencing (GBS), Genome-wide association study (GWAS), High resolution melting (HRM), Marker-assisted selection (MAS)

Posted Date: December 20th, 2019

DOI: <https://doi.org/10.21203/rs.2.19413/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: The Apple (*Malus × domestica* Borkh.) is a valuable fruit crop worldwide, and receives considerable attention as one of the model plants of the Rosaceae family. The primary purpose of apple breeding programs is to generate novel apple cultivars with fruit quality traits that have high commercial value. To advance fruit-tree breeding systems, it is necessary to investigate the association between genomes and targeted traits. Genome-wide association studies (GWAS) are promising to analyze the associations between the genome and traits in fruit tree crops.

Results: In this study, we evaluated 10 major fruit quality traits including titratable acidity (TA), soluble solids content (SSC), and skin color from 301 apple germplasms for four years (2015-2018). GWAS analysis was performed using SNP data generated via genotyping by sequencing (GBS) and identified SNPs significantly related to the fruit quality traits. For TA, significant association loci were detected on chromosome 16 and genes in the candidate regions related to malate transporter. The GWAS loci of SSC was found on chromosome 15, where genes related to sucrose synthase and transporter are located. Significant SNPs associated with fruit skin color were identified on genetic region near the *MYB10* gene on chromosome 9, which regulates anthocyanin biosynthesis. SNPs identified by the GWAS were further confirmed with high resolution melting (HRM) analysis to discover specific polymorphisms in the melting curve.

Conclusion: Overall, these results could identify several candidate genes and SNP markers associated with the fruit quality traits, thus validation of these SNPs enabled marker-assisted selection (MAS). The candidate SNPs and genes observed in this study will contribute to a better understanding of genetic basis for the important fruit quality traits and provide tools for generation of novel cultivars with the quality traits for advancement of apple industry.

Background

Apple (*Malus domestica* Borkh.) is an economically important temperate fruit-tree crop [1], and its industry is globally relatively large. For more advancement of apple industry, many apple breeders are keen interested in development of novel cultivars with high fruit quality traits such as richness of titratable acidity (TA), soluble solids content (SSC), skin color, flesh firmness, and weight, which are major factors attracting to consumers' preferences [2–7]. However, production of new apple cultivars from conventional breeding is relatively slow and unpredictable owing to its self-incompatibility, high heterozygosity, complex traits controlled by multi loci, and relatively long juvenile period that makes it time-consuming to confirm the traits of the fruit compared with annual crops [3, 8, 9].

Apple breeders have attempted to utilize their extensive genetic and phenotypic diversity in order to develop high quality apples, and improvements in quantitative fruit quality traits had also been accomplished by the development and application of molecular markers [10]. Application of marker-assisted selection (MAS) in apple breeding program promises to dramatically improve fruit quality traits

[11, 12]. Genetic molecular marker information can be obtained using the approaches such as genotyping by sequencing (GBS) using next-generation sequencing (NGS) [13] and DNA microarrays [14, 15]. The primary advantage of GBS is that GBS is not restricted by an established set of genotypes compared to DNA microarrays [16].

Quantitative trait locus (QTL) mapping was primarily approach to determine markers related to targeted traits. However, the mapping method has the problem of limited crossover of markers in different populations, and the requirement for large populations linked to the traits [17, 18]. An alternative to this mapping is genome-wide association studies (GWAS), which do not use a designed mapping population in contrast to past mapping techniques [19, 20]. In plant breeding system, the understanding of genotype-to-phenotype relationships is an important perspective. GWAS can explore genotype-phenotype associations in diverse populations of unrelated cultivars, and investigation of several important quantitative fruit quality traits (titratable acidity, soluble solids content, skin color, flesh firmness, and weight) had been performed using GWAS [2–6, 21].

In this study, we used GWAS to confirm single nucleotide polymorphisms (SNPs) relevant to apple fruit quality characteristics. To enable marker-assisted selection (MAS), the genome-wide SNP markers generated by GBS were applied in GWAS studies and their abilities were validated to predict fruit quality traits in diverse apple cultivars. We present the results of a GWAS investigating 10 fruit quality traits by exploiting phenotypic characterization in 301 samples of apple germplasm, and using genotyping-by-sequencing technology.

Results

Distribution of phenotypic variations

Phenotyping was conducted for a total of 301 apple genotypes over four years (2015-2018) in order to identify the distributions of and correlation among different traits (Table S1, Fig.1). A normal distribution was observed for TA, SSC, skin chroma, skin hue, flesh firmness, fruit length, fruit diameter, and fruit weight (Fig. 1), whereas TA showed a continuous variation with a range from 0.1 to 3.0 % over four years. Generally, most cultivars had a distribution ranging from 0.2 to 1.0 % (Fig. 1). SSC ranged from 10.1 to 23.8°brix and reached 11.0-16.0°brix in apple genotypes of more than 80 %, while the skin chroma ranged from 250–350 and skin hue fell between -1 and 2 for about 80 % of the genotypes. In case of flesh firmness, it fell mainly in the 50-80N range, and fruit length and diameter from most of the fruits were 60-70 mm and 70-80 mm, respectively. In term of weight, except crab apple group belonged under 50 g which primarily produces small fruits, most apple phenotypes have 150-250 g, and the distribution of fruit weight resembled a normal distribution.

To prove significant correlations among the traits, Pearson's correlation coefficient was applied in this study (Fig. 2). According to the correlation analysis, all phenotypic traits did not show large differences over the four years. However, fruit length, diameter, and weight were positively correlated with each other ($r > 0.90^{**}$), explaining their mutual interdependence, while a highly significant negative correlation was

indicated fruit length, diameter, and weight ($r = -0.72^{**}$, -0.70^{**} , and -0.61^{**}) regarding flesh firmness. In addition, TA also exhibited a highly significant negative correlation with fruit length, diameter, and weight, but it was significantly positively correlated with flesh firmness ($r = 0.48^{**}$).

Genotyping by sequencing

The GBS data set for the 301 apple genotypes was generated from an average of 3,840,814 raw Illumina sequencing reads from the GBS library. The mapped reads covered an average of 63,058 regions and had a depth of 15.57 depth, with approximately 2 % coverage of the reference genome (Table 1). A total of 13,146,988 SNP loci were called from the GBS sequencing data, including homozygous (8,260,098), heterozygous (2,385,549), and other SNPs (2,501,483) (Table 2). The SNPs were filtered according to specified criteria (Supplementary Table 2). The SNP data were first filtered with a minimum 5 % MAF, which retained 261,969 SNPs. Then, the data set was further filtered to less than 20 % and 10 % missing data per locus, and 51,434 and 12,608 SNPs were obtained, respectively. The SNPs were distributed evenly on chromosomes 1 to 17.

Population genetic analysis

To verify the genetic distances among the 301 genotypes, phylogenetic trees containing 12,608 SNP genotypes were constructed using neighbor joining (NJ) (Fig. 4). According to the phylogenetic analysis, the genotypes clustered into four different major groups. In addition, to better comprehend the genetic structure of the genotypes, the SNP data was analyzed with STRUCTURE software using the MCMC model. Assessment of population structure was performed using genome-wide SNPs, whereas K values ranged from 3 to 10. When delta K was between 4 and 6, a change was significantly appeared about number of clusters (Fig. 3a). This classification of the population was correspond with measured delta K values in four runs with different conditions. This observation was also in accordance with results of the phylogenetic analysis (Fig. 4), thus, the optimum population mean was supposed to be 4. In addition, occurrence of the different genetic variances was also indicated by K value (Fig. 3b).

Genome-wide association analysis for quantitative traits

For association analysis of 10 fruit quality traits among the genotypes, genome-wide association studies were conducted, and the results are presented in Figs. 5 and 6, and Table 3. Four noteworthy association observed from the association analysis are highlighted in Fig. 6. According to the analysis, TA, SSC and skin chroma were identified as being significantly associated with specific SNPs, whereas SNPs with lower significance were identified for the other fruit quality traits (Table 3, Fig. 5). The valid associations of SNPs with all traits were examined on quantile–quantile (Q-Q) plots in advance of confirmation with manhattan plots, and were confirmed to deviate from a uniform diagonal distribution, meaning that the association with significant SNPs was probably valid (Fig. 5 (k-t)). In the Manhattan plot, the red horizontal line indicates a significant line denoting the $-\log_{10}(P)$ of the value significance threshold, and the solid blue line indicates a suggestive line ($-\log_{10}(P) = 5$).

Over the four years investigated, we could identify the SNPs that are significantly associated with titratable acidity (TA) on chromosome 16 (position: 1466019) (Table 3, Fig. 6a), whereas only one SNP that was significantly associated with SSC was found on chromosome 15 (Fig. 6b). In addition, a significant SNPs associated with SSC/TA were also identified near the front of chromosome 16 (Fig. 6c). Moreover, SNPs significantly associated with chroma of skin color were also appeared on chromosome 9 (position: 29585162), and those related to hue of skin color identified on chromosomes 12, 2, and 3, respectively (Table 3, Fig. 5e). Despite they did not surpass the significance threshold, peaks above the suggestive line were observed on chromosomes 17, 8, 3, 9, and 12, associated with flesh firmness, fruit length, fruit diameter, fruit length/diameter ratio, and fruit weight, respectively (Fig. 5(f-j)).

To identify of candidate genes highly related to the fruit traits, a 300 kilobase pair (kb) range centered on each meaningful loci was analyzed. From the analysis, meaningful candidate genes associated with TA, SSC, SSC/TA, and skin chroma could be identified. *Ma1* (MDP0000252114) and *Ma2* (MDP0000244249) genes, which code for malic acid transporters, associated with TA and SSC/TA could be identified to be located within 4 kb – 20 kb from the SNP (Table 3, Fig. 6 a, c). In addition, *SPS3* (MDP0000414968), *SPS4* (MDP0000783676), *HK1* (MDP0000309677), *SUSY1* (MDP0000250070), and *SUT1* (MDP0000275743) were identified to be a candidate gene for SSC in the regions surrounding the SNP (chromosome 15: 11673596, 42605470) (Table 3, Fig. 6b). Moreover, a MYB transcription factor (MDP0000259614) associated with fruit skin chroma was identified to be located approximately 200 kb from the three SNPs on chromosome 9 (chromosome 9: 29585162, 29298755, 29692945) (Table 3, Fig. 6d).

Development of SNP markers and HRM validation

From GWAS analysis and gene annotation, the SNP alleles favorable for fruit quality traits were selected (Table 3), and several SNP markers related to TA and SSC (listed in Table 3) were converted to HRM markers. Based on the initial HRM analysis using all primer pairs, one primer pair for each of the three SNPs with the most definite melt curve patterns was chosen for further analysis. These markers had noticeable differences in HRM curves and were correlated with acidity and sugar content to a certain extent in apple germplasm (Fig. 7).

Related to TA, two-primer set was tested in the core collection and the reference set, including 'Maypole', 'Purple wave', 'Ginger gold', 'Crimson king', 'Virginia gold' and 'Hongro'. The C16S39 (chromosome 16: 1545939) and C16S63 (chromosome 16: 1332163) primers were distinguishable by their 1.3% and 1.0% acidity level, respectively (Fig. 7 a, b). In melting curve of C16S39, the SNP alleles of apple cultivars showing more than 1.4 acidity level tend to indicate GG, and cultivars with less than 1.3 % acidity level were AA or AG (Fig. 7a). Similarly in C16S63, the SNP alleles of apple cultivars with more than 1.1 % acidity level tend to indicate AA or AG and cultivars of less than 1.0 acidity level indicated GG (Fig. 7b). In relation to SSC, C15S96 of SSC-HRM marker (chromosome 15: 11673596) was tested in the core collection and reference set including 'Golden delicious', 'Granny smith', 'Gala', 'Crimson king', 'Spur earliblaze', and 'Hongro'. The C15S96 primer was distinguishable from the 14°brix SSC level (Fig. 7c). In

Fig. 7c, blue represents apple cultivars of 14brix or less and red indicates cultivars of more than 14°brix, respectively.

Discussion

Development of apple new cultivars with novel traits is a prerequisite for advancement of global apple industry. However, development of such cultivars by conventional breeding is relatively long and difficult owing to its self-incompatibility, high heterozygosity, and relatively long juvenile period that makes it time-consuming to confirm the desired traits [8]. Hence, apple breeders tried alternative ways that will help in generation of novel cultivars. Since then, genomic technologies that elucidate the genetic foundation of commercially essential traits had been developed and increasingly used to advance the breeding program [22, 23]. GWAS have been performed to discover the genetic architecture of the apple, in addition to determine genetic loci associated with desirable traits [24]. Understanding genomic region and the number of loci controlling a trait is important for determining genetic structure, and is also useful for constructing effective breeding strategies.

Generally, as the factors, such as the number of cultivars analyzed and the degree of phenotypic variations, can influence reliability of GWAS analysis, sufficient collection of germplasms for this analysis is important [24]. Recently, McClure et al. (2018) conducted the GWAS of apples using 172 genotypes for two years and developed some SNPs associated with fruit traits. In our study, we performed GWAS of apples using 301 germplasms with a broad range of fruit quality characteristics for four years. As the germplasm could be categorized into four clusters, the population used in this study was found to be suitable for the GWAS analysis. Firstly, analysis of fruit quality traits (TA, SSC, skin chroma, skin hue, flesh firmness, fruit length, fruit diameter, and fruit weight) for 301 apple genotypes was performed for four years (2015–2018), and we observed a normal distribution for the traits. Pearson's pairwise correlation analysis for the traits across years indicated that TA, SSC, and flesh firmness were positively correlated, but a negative correlation was found between TA and fruit length, fruit diameter, and fruit weight.

Despite utilization of NGS technologies in various plants, the GBS approach used here for genotyping is more reliable, efficient, cheaper, rapid and has considerably high throughput [25]. Particularly, GBS has been valuable in the sequencing of plants with complicated genomes and large size [26]. In addition, GBS can achieve the sequencing of many samples at the same time and incurs lower costs [13]. Moreover, the efficiency of GWAS analysis using GBS has been demonstrated in soybean and maize [13]. In several genome-wide association studies based on GBS, the principal contribution about power of GWAS analysis was revealed to be trustworthy phenotyping and high-throughput genotyping. In addition, there are some recent researches reporting about GBS-GWAS of apples, however, the population used in those studies are relatively less in comparison with the present study [7, 27]. In our study, application of GBS could identify abundant SNPs, which are useful for GWAS analysis and covered approximately 2% of the reference genome, similar to previous studies.

The analyses were performed using the Efficient Mixed Model Association (EMMA) implementation and Mixed linear model (MLM), which was designed as a model optimized for quantitative traits [28, 29]. Although the shortage of significant SNPs-phenotype associations for some traits, a number of notable associations were detected for most traits. The SNPs strongly associated with TA were located on chromosome 16 in data from four years (Fig. 5a). This peak was identified in the genomic region of the Ma1 (MDP0000252114) and Ma2 (MDP0000244249) genes, which function as malic transporters, and are known to regulate apple acidity [30]. It has also been reported that the genetic loci involved in acidity of the apple are located on chromosomes 8 and 16 [31]. In our study, GWAS results for TA / SSC indicated main peaks and sub peaks at chromosome 16 and 8, and their aspect was similar to the TA results (Fig. 5a, c). Since SSC is the most influential trait in the environment, genetic loci which show a statistically significant association with SSC showed different patterns from year to year. However, significant SNPs were identified most frequently on chromosome 15 over the four years. Chromosome 15 contains genes involved in sucrose transport and sucrose synthesis, and SNPs with high peaks are located close to these genes (SPS3 (MDP0000414968), SPS4 (MDP0000783676), HK1 (MDP0000309677), SUSY1 (MDP0000250070), and SUT1 (MDP0000275743) (Fig. 5b). A convincing peak related to apple fruit skin color was detected on chromosome 9 over the four years (Fig. 5d). This peaked point emphasizes a genetic region including the R2R3 MYB transcription factor (MdMYB1; MDP0000259614), which is known to regulate skin color of apple fruit [3–5, 7, 32]. The degree of the association found in this research and the findings in previous studies suggest that these locus plays an important role in determining fruit skin color within the various apple germplasms. In addition, fruit firmness that influences fruit texture is also an instrumental feature for consumers. Therefore, there has been a concentrate on the firmness in apple breeding in order to maintain post-harvest fruit quality during long-term storage. According to other studies of QTL for flesh firmness, genetic loci related to flesh firmness were appeared on chromosome 10 including genomic regions for candidate genes such as 1-aminocyclopropane-1-carboxylate oxidase (ACO) and polygalacturonase (PG) located up to these QTL regions [2, 5–7, 18]. In this study, meaningful GWAS peaks for flesh firmness were found on chromosomes 17, 1, 3, 4, 5, 16, and 10, and SNPs (chromosome 10: 12900899, 28459403) most significantly associated with firmness were located 5 Mb upstream of PG1 (MDP0000326734) and 4 Mb downstream of ACO1 (MDP0000195885). Genetic loci which are associated with fruit length and fruit diameter have a tendency to appear on chromosome 8 and 9. SNPs with significant associations on chromosome 8 were located close to auxin and proteasome-related genes that regulate cell division and growth (Table 3). Consequentially, several traits such as firmness, fruit length, fruit diameter, L/D ratio, and fruit weight were identified with SNPs below the expected significance threshold.

Recent research suggests that more than millions of SNPs may be needed to a highly influential GWAS in populations with various apple genotypes, owing to relatively rapid LD decay compared to other crops [4]. In addition, the number of fruit samples collected was restricted owing to needed time and labor scarcities, and it is therefore possible that replication would have better observed characteristic variation. Recently, McClure et al. (2018) conducted the GWAS of apples using 172 genotypes for two years. In our study, we conducted GWAS of apples using 301 germplasms for four years and developed the number of

SNP markers (~ 51,434 SNPs), thus, the GWAS results obtained from our work would be more reliable than previous studies. However, a further GWAS in apples with more genetic diversity and larger population sizes would give more precise information to clarify the influence of these factors in identifying associations of genotype and phenotype.

HRM analysis was performed to apply the locus to the population in relation to the acidity and soluble solids content of the GWAS analysis. The two markers capable of distinguishing TA were identified by their acidity of 1.0% and 1.3%, respectively (Fig. 7a, b). One marker that could distinguish the soluble solids content was able to classify based on 14°brix (Fig. 7c). However, the main concern of marker development research is reproducibility in other populations. The exact mechanism by which the four SNP molecular markers identified in this study can distinguish between apple cultivars with high acidity and apple cultivars with high soluble solid content will need to be studied in the future. This research approach also needs to be applied to more genetic resources and crossbreeding populations to see if they are possible candidates for molecular markers. The development of SNP molecular markers by HRM analysis is required to ensure the reproducibility of test results, and the higher the number of species to be identified, the higher the accuracy. Overall, we present a foundation to develop selection markers for marker assisted breeding, genomic selection, and MAS in fruit quality traits, which will facilitate breeders to cost-effectively and more quickly generate new apple cultivars for advancement of global apple industry.

Conclusions

Notable associations as well as interesting candidate genes were found for several fruit quality traits including TA, SSC, SSC/TA, and skin chroma. The SNPs strongly associated with TA exist in the genomic regions of the Ma1 (Chromosome 16: MDP0000252114) and Ma2 (MDP0000244249) genes, while those associated with SSC are located close to the regions of SPS3 (MDP0000414968), SPS4 (MDP0000783676), HK1 (MDP0000309677), SUSY1 (MDP0000250070), and SUT1 (MDP0000275743) on chromosome 15. However, those strongly associated with TA/SSC were located on the chromosomes 8 and 16. In addition, the SNP significantly associated with fruit skin color was detected at a genomic region that harbors an R2R3 MYB transcription factor (MdMYB1; MDP0000259614) on chromosome 9. Moreover, SNPs (chromosome 10: 12900899, 28459403) most significantly associated with firmness were located 5 Mb upstream of PG1 (MDP0000326734) and 4 Mb downstream of ACO1 (MDP0000195885), while genetic loci which are associated with fruit length and fruit diameter have a tendency to appear on chromosome 8 and 9. Development of the SNP markers and candidate genes related to the fruit quality traits might benefit for future apple breeding program by decreasing the time and financial cost required for developing and selecting desirable genotypes. In addition, once function of the candidate genes has been confirmed, they can be specifically edited to improve new cultivars with desired traits for advancement of global apple industry.

Methods

Plant materials

The apple germplasm used in this study consisted of 301 cultivars, including 114 cultivars from the core collection and 44 cultivars belong to reference set for breeding based on pedigree selection (Table S1). For DNA extraction and phenotyping, young leaves and fruits were collected at the Apple Research Institute (Gunwi, Korea).

Fruit assessment and phenotypic data analysis

A total of 301 apple germplasm samples were assessed for ten agronomic traits: Titratable Acidity (TA); Soluble Solids Content (SSC); TA/SSC; Skin Color (Skin Chroma and Skin Hue); Flesh Firmness; Fruit Length; Fruit Diameter; Fruit Length/Diameter Ratio (L/D ratio); and Fruit Weight, in Korea, at the Apple Research Institute in Gunwi, over four years: 2015, 2016, 2017, and 2018.

The titratable acidity (TA) was measured by titrating 5.0mL of apple juice of each cultivar to an endpoint pH of 8.1 with 0.1N sodium hydroxide using an automatic titrator (DL 15, Mettler Toledo, USA) and calculated as malic acid (MA) content. Soluble solids content (SSC) was measured using a digital refractometer (PR-32α, ATAGO, Japan). The fruit skin color was determined by CIELAB (L^* , a^* , b^*) for each cultivar from three different locations along the surface of a fruit using a reflectance colorimeter (CR-400/410, KONICA MINOLTA, Japan). Skin chroma and skin hue angle were calculated as follows:

$SC = (a^{*2} + b^{*2})^{1/2}$, and

$SH = \tan^{-1} (b^*/a^*)$

[33, 34], where L^* indicates lightness, and a^* and b^* indicate chromaticity indices. Flesh firmness (FF) was estimated using a fruit hardness tester (FT327, EFFEGI, Italy) with an 11mm-diameter plunger. Fruit length, fruit diameter, and L/D ratio were measured in millimeters using calipers (ABSOLUTE Digimatic Calliper Series 500, Mitutoyo, Japan) and fruit weight was recorded in grams with an electronic scale.

Phenotypic data were analyzed using SPSS software (IBM SPSS Statistics 25, USA) [35]. Pearson's pairwise correlation coefficient was calculated for fruit quality traits. The correlation analysis was performed with all traits over four years. The asterisks (* or **) indicate statistically significant correlation at p -values of <0.05 or <0.01, respectively.

Genotyping by sequencing and genotype imputation

Extraction of DNA was conducted using DNeasy Plant Mini Kits (QIAGEN, Germany) according to the manufacturer's protocol. The concentration ($\mu\text{g } \mu\text{l}^{-1}$) and quality (A260/A230 and A260/A280 ratio) of genomic DNA were evaluated using a PLUS Spectrophotometer (Bio-Rad, USA) and a NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific, USA). Visual evaluation was performed using electrophoresis on a 0.8% agarose gel.

Single nucleotide polymorphism (SNP) genotyping was obtained using genotyping by sequencing (GBS). For GBS library preparation and sequencing, DNA samples were sent to SEEDERS for GBS library preparation using the restriction enzyme *ApeK1*, and sequenced on a Illumina HiSeq 2000 instrument. The raw sequencing data and SNP calling were performed using the SolexaQA v.1.13 package [36], BWA (Burrows-Wheeler Aligner) program and SAM tools. Short reads were aligned to the apple reference genome, “Golden Delicious” [1]. To ensure the accuracy of the results, biallelic SNPs with a missing rate of less than 0.1-0.3 and minor allele frequency (MAF) greater than 0.05 were used in the analysis (Table 2).

Population genetic analysis

Population genetic analysis was performed using 12,608 SNPs. Population structure analysis was performed using a Bayesian Markov Chain Monte Carlo model (MCMC) implemented using the STRUCTURE software [37]. Two iterations were used for each K value (population number) from three to 10. The length of the burn-in period and number of MCMC replication were set for four run conditions; Burnin/MCMC of 10000/10000; 10000/50000; 10000/75000; and 20000/100000. The most favorable *K* value was identified by the log probability of the data (LnP(D)) and delta K using STRUCTURE HARVEST [38]. To construct a phylogenetic tree, neighbor-joining (NJ) clustering was carried out using the Geneious 10.2.4 software.

Genome-wide association study

GWAS was conducted on 301 apple cultivars from phenotype data from 2015-2018. Among the SNPs identified as significant, 51,434 SNP loci with less than 20% missing data and more than 5% MAF were used for GWAS (Table 1). All analyses were performed using the Efficient Mixed Model Association (EMMA) implementation and Mixed linear model (MLM) of the Genomic Association and Prediction Integrated Tool (GAPIT) in R [33, 39]. The genome-wide significance thresholds for each association study were assigned using Bonferroni correction according to the effective number of SNPs [33]. Significance thresholds for a 5% false discovery rate (FDR) were calculated using the Benjamini-Hochberg FDR correction [40]. Manhattan plots and Q-Q plots for association mapping were visualized using the qqman package in R [39, 41]. The red horizontal line indicates the $-\log_{10}(P)$ value significance threshold and the blue solid line indicates a suggestive line ($-\log_{10}(P) = 5$).

Gene annotation

The genomic regions upstream and downstream of the significant SNPs associated with each fruit trait were searched to discover the annotated genes by scanning the genome using ~500 Kb windows. SNPs at promising loci were retrieved to predict genes for the corresponding regions from the NCBI database (<https://www.ncbi.nlm.nih.gov/genome/>) and the Genome Database for Rosaceae (GDR) (<https://www.rosaceae.org>).

Development of SNP markers and HRM validation

The primers for high-resolution melting (HRM) were designed based on the SNPs identified by genotyping-by sequencing of genetic loci associated with quantitative traits, according to GWAS results using PRIMER 3 (NCBI/Primer-BLAST: <http://www.ncbi.nlm.nih.gov/tools/primer-blast/>). PCR for HRM analysis was performed using the Rotor-Gene 6000™ (Corbett Research, Australia) using Type-it HRM PCR Kits (QIAGEN, Germany). The reaction mixture in a 20 µl final volume contained 2x HRM PCR Master Mix, 10 µM forward primer, 10 µM reverse primer, genomic DNA and RNase-free water. The cycling and melting conditions were as follows: initial denaturation at 95°C for 10 minutes, followed by 50 cycles at 95°C for 15 seconds, 60°C for 15 seconds and 72°C for 30 seconds, and then 90°C for one minute, 40°C for one minute and a melt from 65°C to 90°C, rising by 0.1°C per second. The primers were tested using DNA previously isolated and diluted to 50 ng/µl. All reactions were conducted in duplicate. Melting curve data were obtained and analysed using the Rotor-Gene Q Software (v2.3.1). Analysis of HRM variants was conducted according to differences in the pattern of the T_m values and melting curves.

Declarations

Acknowledgments

Not applicable.

Funding

This work was carried out with the support of "Cooperative Research Program for Agriculture Science and Technology Development (Project No. PJ01311503)" Rural Development Administration, Republic of Korea.

Availability of data and materials

All data generated or analysed during this study are included in this published article (and its Additional file 1 and 2).

Author's contributions

ARH, GHK and CC conceived and designed the research. ARH performed the experiments and analyzed data. Phenotyping was investigated by SAK. ARH and CC drafted the initial manuscript and discussed to all authors about results. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Abbreviations

GWAS: Genome-wide association studies; TA: Titratable acidity; SSC: Soluble solids content; GBS: Genotyping by sequencing; HRM: High resolution melting; MAS: Marker-assisted selection; NGS: Next-generation sequencing; QTL: Quantitative trait locus; SNP: Single nucleotide polymorphism; L/D: Length/diameter; SPS: Sucrose phosphate synthase; SUSY: Sucrose synthase; SUT: Sucrose transporter; HK: Hexokinase; ACO: 1-aminocyclopropane-1-carboxylate oxidase; PG: Polygalacturonase; EMMA: Efficient mixed model association; MLM: Mixed linear model; NJ: Neighbor-joining

References

1. Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Fontana P, Bhatnagar SK, Troggio M, Pruss D, et al. The genome of the domesticated apple (*Malus × domestica* Borkh.). *Nat Genet.* 2010;42(10):833-839.
2. Kumar S, Garrick DJ, Bink MC, Whitworth C, Chagne D, Volz RK. Novel genomic approaches unravel genetic architecture of complex traits in apple. *BMC Genom.* 2013;14(1):393-405.
3. Kumar S, Raulier P, Chagne D, Whitworth C. Molecular-level and trait-level differentiation between the cultivated apple (*Malus × domestica* Borkh.) and its main progenitor *Malus sieversii*. *Plant Genet Resour.* 2014;12(3):330-340.
4. Migicovsky Z, Gardner KM, Money D, Sawler J, Bloom JS, Moffett P, Chao CT, Schwaninger H, Fazio G, Zhong GY, et al. Genome to Phenome Mapping in Apple Using Historical Data. *Plant Genome.* 2016; doi:10.3835/plantgenome2015.11.0113.
5. Amyotte B, Bowen AJ, Banks T, Rajcan I, Somers DJ. Mapping the sensory perception of apple using descriptive sensory evaluation in a genome wide association study. *PLoS One.* 2017;12(2):e0171710.
6. Moriya S, Kunihiya M, Okada K, Iwanami H, Iwata H, Minamikawa M, Katayose Y, Matsumoto T, Mori S, Sasaki H, et al. Identification of QTLs for Flesh Mealiness in Apple (*Malus × domestica* Borkh.). *Hortic J.* 2017; doi:10.2503/hortj.MI-156.
7. McClure KA, Gardner KM, Douglas GM, Song J, Forney CF, DeLong J, Fan L, Du L, Toivonen P, Somers DJ. A genome-wide association study of apple quality and scab resistance. *Plant Genome.* 2018; doi:10.3835/plantgenome2017.08.0075.
8. Sassa H, Mase N, Hirano H, Ikehashi H. Identification of self-incompatibility-related glycoproteins in styles of apple (*Malus × domestica*). *Theor Appl Genet.* 1994;89(2):201-205.
9. Brown SK, Maloney KE. Genetic improvement of apple: breeding, markers, mapping and biotechnology. *Apple: Botany, Production and Uses.* 2003; doi:10.1079/9780851995922.0031.

10. Hancock JF. Temperate fruit crop breeding: germplasm to genomics. Springer Netherlands. 2008; doi:10.1007/978-1-4020-6907-9.
11. Dirlewanger E, Graziano E, Joobeur T, Garriga-Caldere F, Cosson P, Howad W, Arus P. Comparative mapping and marker-assisted selection in Rosaceae fruit crops. *Proc Natl Acad Sci U S A*. 2004;101(26):9891-9896.
12. Liebhard R, Koller B, Gianfranceschi L, Gessler C. Creating a saturated reference map for the apple (*Malus × domestica* Borkh.) genome. *Theor Appl Genet*. 2003;106(8):1497-1508.
13. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS One*. 2011;6(5):e19379.
14. Bianco L, Cestaro A, Sargent DJ, Banchi E, Derdak S, Di Guardo M, Salvi S, Jansen J, Viola R, Gut I. Development and validation of a 20K single nucleotide polymorphism (SNP) whole genome genotyping array for apple (*Malus× domestica* Borkh). *PLoS One*. 2014;9(10):e110377.
15. Bianco L, Cestaro A, Linsmith G, Muranty H, Denance C, Theron A, Poncet C, Micheletti D, Kerschbamer E, Di Pierro EA, et al. Development and validation of the Axiom®Apple480K SNP genotyping array. *Plant J*. 2016;86(1):62-74.
16. Chagne D. Whole Genome Sequencing of Fruit Tree Species. *Adv Bot Res*. 2015; doi: 10.1016/bs.abr.2015.04.004.
17. Soto-Cerda BJ, Cloutier S. Association mapping in plant genomes. *Genetic Diversity in Plants*. 2012; doi:10.5772/33005.
18. Costa F. MetaQTL analysis provides a compendium of genomic loci controlling fruit quality traits in apple. *Tree Genet Genomes*. 2015;11(1):819-829.
19. Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES. Association mapping: critical considerations shift from genotyping to experimental design. *Plant Cell*. 2009;21(8):2194-2202.
20. Khan MA, Korban SS. Association mapping in forest trees and fruit crops. *J Exp Bot*. 2012;63(11):4045-4060.
21. McClure KA, Gardner KM, Douglas GM, Song J, Forney CF, DeLong J, Fan L, Du L, Toivonen PMA, Somers DJ, et al. A Genome-Wide Association Study of Apple Quality and Scab Resistance. *Plant Genome*. 2018; doi:10.3835/plantgenome2017.08.0075.
22. Isik F, Kumar S, Martinez-Garcia PJ, Iwata H, Yamamoto T. Acceleration of Forest and Fruit Tree Domestication by Genomic Selection. *Adv Bot Res*. 2015; doi:10.1016/bs.abr.2015.05.002.
23. Edge-Garza DA, Luby JJ, Peace C. Decision support for cost-efficient and logistically feasible marker-assisted seedling selection in fruit breeding. *Mol Breed*. 2015;35(12):223.
24. Korte A, Farlow A. The advantages and limitations of trait analysis with GWAS: a review. *Plant methods*. 2013;9:29-29.
25. Sonah H, Bastien M, Iquira E, Tardivel A, Legare G, Boyle B, Normandeau E, Laroche J, Larose S, Jean M, et al. An Improved Genotyping by Sequencing (GBS) Approach Offering Increased Versatility and

- Efficiency of SNP Discovery and Genotyping. *PLoS One*. 2013;8(1):e54603.
26. Deschamps S, Llaca V, May GD. Genotyping-by-Sequencing in Plants. *Biol*. 2012;1(3):460-483.
 27. Larsen B, Migicovsky Z, Jeppesen AA, Gardner KM, Toldam-Andersen TB, Myles S, Orgaard M, Petersen MA, Pedersen C. Genome-Wide Association Studies in Apple Reveal Loci for Aroma Volatiles, Sugar Composition, and Harvest Date. *Plant Genome*. 2019; doi: 10.3835/plantgenome2018.12.0104.
 28. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat genet*. 2012;44(7):821-824.
 29. Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM. Mixed linear model approach adapted for genome-wide association studies. *Nat genet*. 2010;42(4):355-360.
 30. Khan SA, Beekwilder J, Schaart JG, Mumm R, Soriano JM, Jacobsen E, Schouten HJ. Differences in acidity of apples are probably mainly caused by a malic acid transporter gene on LG16. *Tree Genet Genomes*. 2013;9(2):475-487.
 31. Jia D, Shen F, Wang Y, Wu T, Xu X, Zhang X, Han Z. Apple fruit acidity is genetically diversified by natural variations in three hierarchical epistatic genes: MdSAUR37, MdPP2CH and MdALMTII. *Plant J*. 2018;95(3):427-443.
 32. Ban Y, Honda C, Hatsuyama Y, Igarashi M, Bessho H, Moriguchi T. Isolation and functional analysis of a MYB transcription factor gene that is a key regulator for the development of red coloration in apple skin. *Plant Cell Physiol*. 2007;48(7):958-970.
 33. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z. GAPIT: genome association and prediction integrated tool. *Bioinformatics*. 2012;28(18):2397-2399.
 34. Khan MA, Olsen KM, Sovero V, Kushad MM, Korban SS. Fruit quality traits have played critical roles in domestication of the apple. *Plant Genome*. 2014; doi:10.3835/plantgenome2014.04.0018.
 35. Coakes SJ, Steed L. SPSS: Analysis Without Anguish Using Spss Version 14.0 for Windows. John Wiley & Sons, Inc. 2009.
 36. Cox MP, Peterson DA, Biggs PJ. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics*. 2010;11(1):485-490.
 37. Pritchard JK, Stephens M, Donnelly P. Inference of Population Structure Using Multilocus Genotype Data. *Genet*. 2000;155(2):945-959.
 38. Earl DA, vonHoldt BM. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet*. 2012;4(2):359-361.
 39. Team RC. A language and environment for statistical computing. R Foundation for Statistical Computing. 2013; <http://www.R-project.org>.
 40. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B*. 1995;57(1):289-300.

41. Turner SD. qqman: an R package for visualizing GWAS results using QQ and manhattan plots. Biorxiv. 2014; doi:10.1101/005165.

Tables

Table 1. Regions of alignment mapped to the reference genome cv. Golden Delicious in the 301 apple germplasm, based on GBS results

	Mapped regions					Reference genome coverage (%)
	Total No.	Average depth	Median depth	Total length (bp)	Average length (bp)	
SUM	19,169,677	-	-	3,160,895,874	-	-
AVG	63,058	15.57	8.84	10,397,683	162.50	1.98

Table 2. Single nucleotide polymorphism (SNP) information in the 301 apple germplasm, based on GBS results

	Total SNPs	No. of SNPs (Hom ^a)	No. of SNPs (Het ^b)	No. of SNPs (etc.)
SUM	13,146,988	8,260,098	2,385,549	2,501,341
AVG	43,678	27,442	7,925	8,310

^aHom: homozygous SNPs, read depth ≥ 90%
^bHet: heterozygous SNPs, 40% ≤ read depth ≤ 60%

Table 3. Significant genetic loci associated with apple fruit quality traits as determined by GWAS and annotations of the candidate genes.

Trait	Marker	SNP Alleles	P value	Gene	Annotation	Region annotation ^a
TA	chr16: 1466019	C/T	8.62×10 ⁻⁹	LOC103402725	methionine aminopeptidase 1B	LOC103402712 aluminum-activated malate transporter 4-like
	chr16: 1466001	A/G	9.56×10 ⁻⁹			LOC103402726 AP2/ERF and B3 domain-containing transcription factor RAV1
	chr16: 1350941	G/T	2.36×10 ⁻⁸	LOC103402717	unknown	
	chr16: 1378405	C/T	1.72×10 ⁻⁷	LOC103402719	CTL-like protein DDB_G0274487	
	chr16: 1545939	A/G	9.37×10 ⁻⁷	LOC103402728	unknown	
	chr16: 1348804	A/G	1.31×10 ⁻⁶	LOC103402717	unknown	
	chr16: 1450841	A/C	2.23×10 ⁻⁶	-		
	chr16: 1316991	A/C	2.59×10 ⁻⁶	LOC103402710	ACD11 homolog protein	
	chr16: 1313364	C/T	4.59×10 ⁻⁶	LOC103402711	dihydrolipoyllysine-residue acetyltransferase component 4 of pyruvate dehydrogenase complex	
	chr16: 1332163	A/G	6.00×10 ⁻⁶	LOC103402712	aluminum-activated malate transporter 4-like	
SSC	chr9: 7620153	C/T	1.97×10 ⁻¹⁰	LOC103442928	ribonuclease 3-like protein 3	LOC103442640 endoglucanase 11-like LOC103442646 glucan endo-1,3-beta-glucosidase 12 LOC103442637 pyrophosphate—fructose 6-phosphate 1-phosphotransferase subunit beta-like
	chr9: 7620179	C/G	1.97×10 ⁻¹⁰	LOC103442928	ribonuclease 3-like protein 3	
	chr15: 40638748	A/G	1.02×10 ⁻⁹	LOC103402139	unknown	LOC103402127 ethylene-responsive transcription factor ERF022-like LOC103402221 GDSL esterase/lipase At5g55050-like
	chr4: 18771078	C/T	1.55×10 ⁻⁹	LOC103433852	chloroplast stem-loop binding protein of 41 kDa b, chloroplastic	LOC103434062 glucose-6-phosphate 1-dehydrogenase 4, chloroplastic
	chr15: 11673596	A/T	2.37×10 ⁻⁹	LOC103400498	tRNA(Ile)-lysidine synthase	LOC103400550 probable plastidic glucose transporter 2
	chr6: 944785	C/G	4.47×10 ⁻⁹	LOC103436726	U-box domain-containing protein 44-like	LOC103436720 glucan endo-1,3-beta-D-glucosidase -like LOC103436736 glycosyltransferase family 92 protein RCOM_0530710
	chr15: 11673448	G/T	4.75×10 ⁻⁹	LOC103400498	tRNA(Ile)-lysidine synthase	LOC103400550 probable plastidic glucose transporter 2
	chr17: 2360290	A/G	5.22×10 ⁻⁹	LOC103404338	methyl-CpG-binding domain-containing protein 9	LOC103404306 probable galactinol-sucrose galactosyltransferase 5

Table 3. (Continued)

SSC/TA	chr16: 1316991	A/C	6.95×10 ⁻¹¹	LOC103402710	ACD11 homolog protein	LOC103402712 aluminum-activated malate transporter 4-like
	chr16: 1316988	C/T	1.60×10 ⁻¹⁰			LOC103402726 AP2/ERF and B3 domain-containing transcription factor RAV1
	chr16: 1450841	A/C	1.30×10 ⁻⁹	-		
	chr16: 1466019	C/T	1.43×10 ⁻⁹	LOC103402725	methionine aminopeptidase 1B	
	chr16: 1350941	G/T	3.60×10 ⁻⁹	LOC103402717	unknown	
	chr16: 1466001	A/G	3.63×10 ⁻⁹	LOC103402725	methionine aminopeptidase 1B	
	chr16: 1469777	A/C	6.73×10 ⁻⁹	LOC103402725	methionine aminopeptidase 1B	
Skin chroma	chr9: 29585162	A/G	3.28×10 ⁻¹²	LOC103444207	unknown	LOC103444202 transcription factor MYB75-like
	chr9: 30309556	A/T	3.54×10 ⁻¹¹	LOC103444238	universal stress protein PHOS32	
	chr9: 29298755	A/G	7.71×10 ⁻¹¹	LOC103444196	unknown	
	chr9: 29692945	A/C	9.20×10 ⁻¹¹	LOC103444361	threonylcarbamoyladenosine dehydratase (tRNA)	
Skin hue	chr12: 18919840	A/G	7.06×10 ⁻¹²	LOC103450207	unknown	
Flesh firmness	chr17: 10332533	C/T	4.02×10 ⁻⁷	LOC103405095	vacuolar protein sorting-associated protein 53 A-like	LOC103405256 transmembrane 9 superfamily member 12-like LOC103405265 7-deoxyloganetin glucosyltransferase-like LOC103405248 cytochrome P450 724B1-like
Fruit length	chr8: 15846862	C/T	1.66×10 ⁻⁷	LOC103441332	golgin subfamily B member 1-like	LOC103441212 proteasome subunit beta type-6 LOC103441218 auxin-responsive protein SAUR72-like
Fruit diameter	chr3: 19969159	C/T	2.07×10 ⁻⁷			
Fruit L/D ratio	chr9: 22755626	A/G	5.71×10 ⁻⁷	LOC103443904	ATP synthase subunit delta	LOC103443903 somatic embryogenesis receptor kinase 2 LOC103443896 GDSL esterase/lipase At1g29670-like
Fruit weight	chr2: 2551153	A/G	3.88×10 ⁻⁶	LOC103447817	N-alpha-acetyltransferase 35, NatC auxiliary subunit-like	LOC103447799 probable sucrose-phosphate synthase 1

^aThe genes located in the regions were within approximately 500Kb upstream and downstream of the SNPs of interest

TA- titratable acidity, SSC- soluble solids content, L/D- length/diameter

Figures



Frequency distribution for phenotypic data of fruit quality traits at harvest time. (a) Titratable acidity; (b) Soluble solids content; (c) Skin chroma; (d) Skin hue; (e) Flesh firmness; (f) Fruit length; (g) Fruit diameter; (h) Fruit weight.

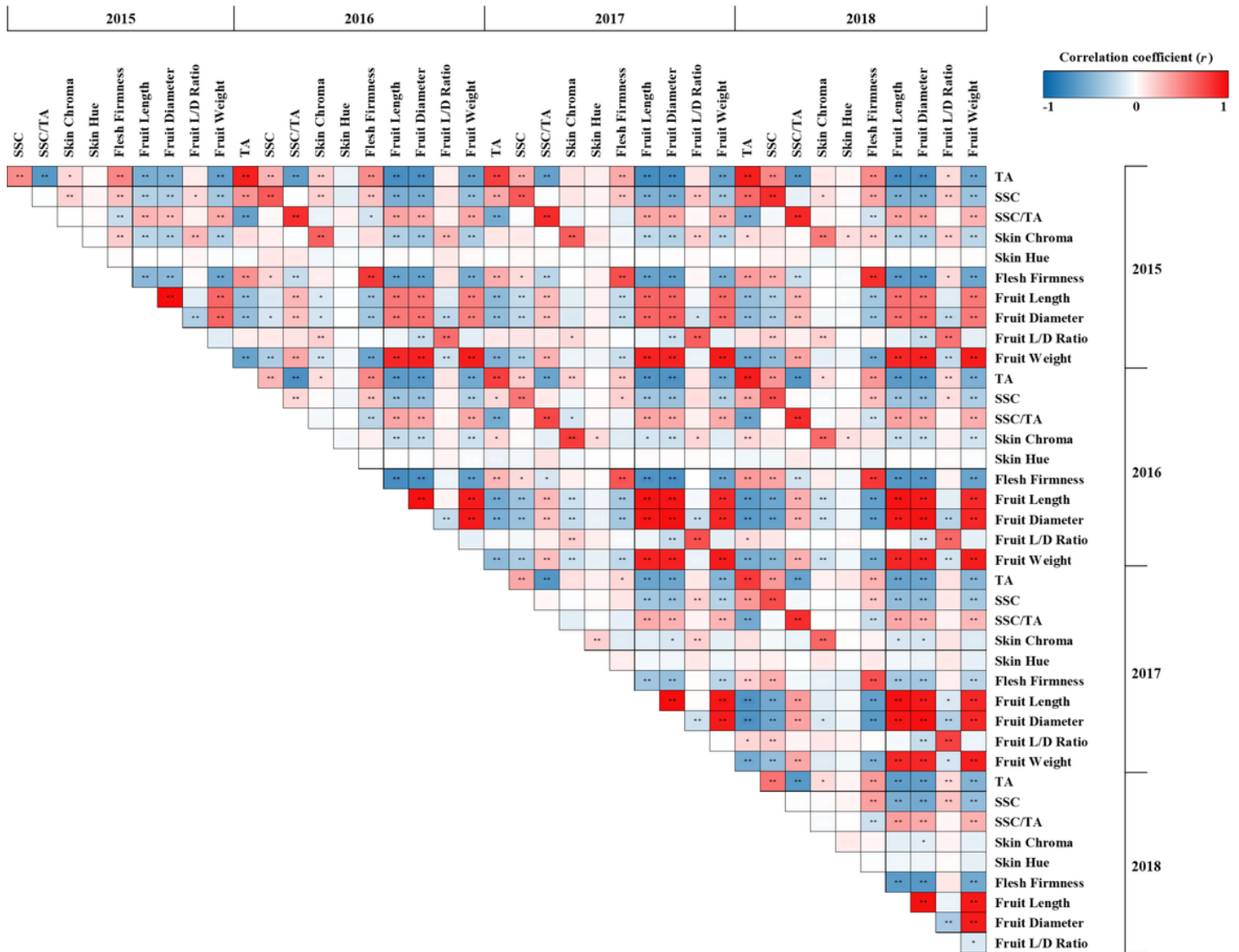


Figure 2

Pearson correlation coefficient (r) for the phenotype data of fruit quality traits at harvest over four years (2015, 2016, 2017, and 2018). Correlation values range from -1 to 1, and colors ranging from dark blue to dark red, representing positive and negative correlations, respectively. **. Correlation is significant at the 0.01 level. *. Correlation is significant at the 0.05 level. TA- Titratable Acidity. SSC- Soluble Solids Content. L/D- Length/Diameter.



cluster.



Figure 5

The GWAS results of fruit quality traits in 301 apple germplasms. Manhattan plots (a-j) of chromosomal positions and Q-Q plots (k-t) of SNP markers for titratable acidity (TA), soluble solids content (SSC), SSC/TA, skin chroma, skin hue, flesh firmness, fruit length, fruit diameter, fruit L/D ratio, and fruit weight using EMMA manual of GAPIT. In all plots, the red horizontal line indicates a significant line denoting the $-\log_{10}(P)$ value significance threshold and the blue solid line indicates a suggestive line ($-\log_{10}(P) = 5$).



Figure 6

Manhattan plots for (a) Titratable acidity (TA), (b) Soluble solids content (SSC), (c) SSC/TA, and (d) Skin Chroma for 301 apple germplasm in 2015, 2016, 2017, and 2018. The bottom row shows chromosome size of Manhattan plots containing the most significant peak SNPs, with the position of genes related to trait, gene name, and corresponding MDP ID from GDR. In all plots the red horizontal line indicates the significant line denoted the $-\log_{10}(P)$ value significance threshold and the solid blue line indicates a suggestive line ($-\log_{10}(P) = 5$).



Figure 7

Melting curves of an HRM analysis for markers developed from SNPs associated with titratable acidity (TA) and soluble solids content (SSC). The left side is a melting curve analysis (normalized graph), and the right side is a difference graph. The HRM markers are as follows: (a) TA_C16S39 and (b) TA_C16S63 (c) SSC_C15S96.

Supplementary Files

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

- [Additionalfile1.docx](#)
- [Additionalfile2.docx](#)