Journal of Experimental Botany, Vol. 63, No. 8, pp. 2895–2908, 2012
doi:10.1093/jxb/err464 Advance Access publication 13 February, 2012
This paper is available online free of all access charges (see http://jxb.oxfordjournals.org/open_access.html for further details)

Journal of Experimental Botany
www.jxb.oxfordjournals.org

RESEARCH PAPER

Genetic analysis of metabolites in apple fruits indicates an mQTL hotspot for phenolic compounds on linkage group 16

Sabaz Ali Khan[1,*], Pierre-Yves Chibon[1,*], Ric C.H. de Vos[2,4,5], Bert A. Schipper[2,4,5], Evert Walraven[3], Jules Beekwilder[2], Thijs van Dijk[1], Richard Finkers[2], Richard G.F. Visser[1], Eric W. van de Weg[2], Arnaud Bovy[1,5], Alessandro Cestaro[6], Riccardo Velasco[6], Evert Jacobsen[1] and Henk J. Schouten[2,†]

[1] Wageningen UR Plant Breeding, Wageningen University and Research Centre, 6708 PB Wageningen, The Netherlands
[2] Wageningen University and Research Centre, PO Box 16, 6700 AA, Wageningen, The Netherlands
[3] Wageningen University and Research Centre, Lingewal 1, 6668 LA Randwijk, The Netherlands
[4] Netherlands Metabolomics Centre, Einsteinweg 55, 2333 CC Leiden, The Netherlands
[5] Centre for BioSystems Genomics, PO Box 98, 6700 AB, Wageningen, The Netherlands
[6] Istituto Agrario San Michele all'Adige Research and Innovation Centre, Edmund Mach Foundation, Trento, Italy

* These authors contributed equally to this work
† To whom correspondence should be addressed. E-mail: henk.schouten@wur.nl

Received 23 June 2011; Revised 25 November 2011; Accepted 22 December 2011

Abstract

Apple (*Malus*×*domestica* Borkh) is among the main sources of phenolic compounds in the human diet. The genetic basis of the quantitative variations of these potentially beneficial phenolic compounds was investigated. A segregating F_1 population was used to map metabolite quantitative trait loci (mQTLs). Untargeted metabolic profiling of peel and flesh tissues of ripe fruits was performed using liquid chromatography–mass spectrometry (LC-MS), resulting in the detection of 418 metabolites in peel and 254 in flesh. In mQTL mapping using MetaNetwork, 669 significant mQTLs were detected: 488 in the peel and 181 in the flesh. Four linkage groups (LGs), LG1, LG8, LG13, and LG16, were found to contain mQTL hotspots, mainly regulating metabolites that belong to the phenylpropanoid pathway. The genetics of annotated metabolites was studied in more detail using MapQTL®. A number of quercetin conjugates had mQTLs on LG1 or LG13. The most important mQTL hotspot with the largest number of metabolites was detected on LG16: mQTLs for 33 peel-related and 17 flesh-related phenolic compounds. Structural genes involved in the phenylpropanoid biosynthetic pathway were located, using the apple genome sequence. The structural gene *leucoanthocyanidin reductase* (*LAR1*) was in the mQTL hotspot on LG16, as were seven transcription factor genes. The authors believe that this is the first time that a QTL analysis was performed on such a high number of metabolites in an outbreeding plant species.

Key words: *Malus*×*domestica* Borkh, genetical metabolomics, LC-MS, MapQTL, MetaNetwork, untargeted and targeted mQTL mapping.

Introduction

The fruit of apple (*Malus*×*domestica* Borkh) is a rich source of phytochemicals including phenolic compounds (Gerhauser, 2008). There is increasing evidence that apples are an important source of various compounds that are beneficial for human health. For example, their consumption has been associated with a risk reduction of many human diseases, such as asthma, type-2 diabetes, thrombotic stroke, ischaemic heart disease, and various cancers (Eberhardt *et al.*, 2000; Mcghie *et al.*, 2005). Some of the major phenolic compounds isolated and identified from apple are procyanidins, anthocyanins, chlorogenic acid, hydroxycinnamic acid, flavan-3-ols, such as (–)-epicatechin, (+)-catechin, and gallaocatechin, and phloridzin and quercetin glycosides (Mazza and Velioglu, 1992; Lu and Foo, 1997; Awad *et al.*, 2000; Treutter, 2001).

The aim of the current study is the elucidation of the genetic basis of metabolic variability in apple fruits. This study was initiated without any *a priori* knowledge with regard to the specific metabolite groups. For that reason, large-scale liquid chromatography–mass spectrometry (LC-MS)-based metabolic profiling was chosen.

Metabolomics is defined as the large-scale analysis of metabolites in an organism, and it concerns the simultaneous measurement of these metabolites in a given biological system (Dixon and Strack, 2003). Metabolomics is developing as an important functional genomics tool in crop plants, including fruit trees (Carrari and Fernie, 2006; Moco *et al.*, 2006). Although QTLs have been mapped in cultivated apples for different traits such as disease resistance (Calenge *et al.*, 2004; Calenge and Durel, 2006; Khan *et al.*, 2006), fruit quality (King *et al.*, 2001; Liebhard *et al.*, 2003; Davey *et al.*, 2006), and tree morphology (Kenis and Keulemans, 2007), there is only one report on the genetic mapping of a large number of metabolites in apple fruits, and, in that case, on volatiles (Dunemann *et al.*, 2009).

The LC-MS metabolomics showed numerous metabolic compounds in the segregating F_1 population, both in peel and in flesh of the fruits, allowing mQTL (metabolomic quantitative locus) mapping. Standard QTL mapping software is designed to map individual traits, one by one, and is not suited to map hundreds of metabolites simultaneously. Therefore, it was decided to use the software MetaNetwork (Fu *et al.*, 2007). MetaNetwork enables simultaneous genome-wide screening of numerous traits. Keurentjes *et al.* (2006) used MetaNetwork to find mQTLs for secondary metabolites in an *Arabidopsis thaliana* recombinant inbred line (RIL) population.

In the current research, metabolites from untargeted metabolic profiling were mapped. The majority of the mapped metabolites belong to the phenylpropanoid pathway. A major mQTL hotspot was found on linkage group (LG) 16. Only one structural gene, *leucoanthocyanidin reductase* (*LAR1*), was detected in the mQTL hotspot, as were seven transcription factor genes. The authors believe that this is the first time that such an extensive mQTL mapping was performed in a highly heterozygous and cross-pollinating crop species such as apple.

Materials and methods

Plant materials

For mQTL mapping, a segregating F_1 population from the cross 'Prima'×'Fiesta' was used. This population was also used for the first international reference linkage map of apple covering all chromosomes (Maliepaard *et al.*, 1998). In this study a subset of 113 progeny and both parents was used. For the F_1 population, two trees per genotype were present.

Harvesting and storage of the apples

Mature fruits of all genotypes were harvested in September and October 2008 in a trial orchard in Elst, The Netherlands. The maturity of the fruits was assessed by checking the colour of the peel, the taste, and the browning of the seeds. For each progeny, >10 fruits from each of the two trees were harvested separately, while for the two parents 'Prima' and 'Fiesta' fruits from five trees (five replicates) were harvested. The fruits were harvested randomly from different sides of each individual tree to balance out possible differences due to environmental factors such as light. Fruits were oversampled in the field to forestall the possible damage or decay during transit. After harvesting, fruits were immediately stored at 0 °C in a cold storage room to minimize enzymatic activities. Once fruits for all of the genotypes were harvested, these were moved to a storage room at 20 °C for 7 d. This was done to mimic the storage conditions in a consumer's household.

Selection and grinding of apples

Samples of eight apples per genotype were selected. For the progeny genotypes, four apples from each of the two trees of one genotype were combined as one sample, giving one replicate per genotype. For each sample, the individual fruit was cut transversely to obtain a 1 cm thick round slice, and the round slice was peeled. The peel (1.4 mm thick) was chopped into small pieces and snap-frozen in a separate beaker with liquid nitrogen. The core was removed from the flesh and a slice (3.2 mm thick) of the flesh was also chopped into small pieces and snap-frozen. This was repeated for all of the eight apples of one genotype, and the samples from a tissue were pooled per genotype. The samples were then ground using an IKA coffee grinder (model A11 basic). The powder for the flesh and peel was collected separately in 50 ml falcon tubes and stored at –80 °C. For the parents 'Prima' and 'Fiesta', the samples were treated separately in five replicates each and subhjected to the same procedures as described for the progeny.

Extract preparation

The aqueous methanol extracts were prepared as described by De Vos *et al.* (2007), with minor modification (Keurentjes *et al.*, 2006). Ice-cold 99.9% methanol (1.5 ml) acidified with 0.133% (v/v) formic acid was added to each plant sample (final methanol concentration of 75%, assuming 90% water in the 500±5 mg tissues). The ensuing steps, from sonication to the injection of the samples and separation using the Alliance 2795 HT system, were performed as described by De Vos *et al.* (2007). The separation was performed at 40 °C, by applying a 45 min gradient of 5–35% acetonitrile in water (acidified with 0.1% formic acid) at a flow rate of 0.19 ml min^{-1}. The compounds eluting from the column were detected online, first by a Waters 996 photodiode array detector at 200–700 nm and then by a Q-TOF Ultima MS (Waters) with an electron spray ionization (ESI) source. Ions were detected in negative mode in the range of *m*/*z* 80–1500 at a resolution of 10 000, using a scan time of 900 ms and an interscan delay of 100 ms. The desolvation temperature was 250 °C, with a nitrogen gas flow of 500 l h^{-1}, the capillary spray was 2.75 kV, the source temperature was 120 °C, the cone voltage was 35 V with 50 l h^{-1} nitrogen gas flow, and the collision energy was 10 eV.

The mass spectrometer was calibrated as described by De Vos *et al.* (2007). MassLynx software version 4.0 (Waters) was used to control all instrumentation and for the calculation of accurate masses.

Pre-processing the data set

Unbiased mass peak picking and baseline correction of the raw LC-MS data was performed using Metalign software (De Vos *et al.*, 2007; Lommen, 2009; www.metalign.nl) with a signal to noise ratio of ⩾3. Thus, a total of 18 582 mass signals were extracted from the peel samples and 11 817 signals from the flesh samples. Both peak lists were subsequently filtered for signals present in >10 samples, resulting in 4830 peel signals and 2826 flesh signals. A single metabolite may produce a number of mass

peaks, due to natural isotopes, unavoidable fragmentation, and adduct formation in the ESI source, resulting in data redundancy. Therefore, mass signals originating from the same metabolite were grouped, based on their corresponding retention time and intensity pattern over samples, using MSClust software (Tikunov *et al.*, 2011) that can be freely downloaded from the Metalign website (www.metalign.nl). From the clustered mass signals (i.e. reconstructed metabolites), the most representative signal per mass peak cluster was taken for further data analyses. The metabolite signals were ^{10}log transformed to normalize the variances among the metabolites.

Genetic linkage maps

Genetic linkage maps were available for both 'Prima' and 'Fiesta', representing the 17 linkage groups of apple. The maternal map consists of 562 markers and the paternal map consists of 452 markers, including diversity array technology (DArT), amplified fragment length polymorphism (AFLP), restriction fragment length polymorphism (RFLP), nucleotide-binding site leucine-rich repeat (NBS-LRR), simple sequence repeat (SSR), and random amplified polymorphic DNA (RAPD) markers, and some isozymes (Schouten *et al.*, 2011).

In the untargeted mQTL mapping of metabolites in apple, the individual maps of 'Prima' and 'Fiesta' were used, as the MetaNetwork could not incorporate the integrated map of cross-pollinating crops such as apple. In the targeted mQTL mapping using MapQTL® 6.0, an integrated map of both parents was constructed and used for the analysis of the annotated metabolites. The integrated map contained 801 markers, spanning 1348 cM.

Untargeted mQTL mapping of metabolites using MetaNetwork

MetaNetwork was designed for the mQTL analysis of homozygous RIL populations of inbreeding plants such as *A. thaliana* (Fu *et al.*, 2007), and was therefore not applicable for the analysis of a segregating F_1 population of an outcrossing species. Therefore, the data were transformed, using single parental maps, giving $2\times17=34$ linkage groups. Hereby each parent was considered to be derived from a cross between two inbred lines, and the F_1 progeny was considered to be the result of a backcross. The linkage phase information from the linkage map was used to assign F_1 marker alleles to the respective parental inbred lines, thus giving the dichotomous marker scores as required by MetaNetwork. Missing marker data were imputed using information of flanking markers if they were within a 20 cM distance and in a non-recombinant segment.

MetaNetwork implements a two-part parametric model per trait, combining a non-parametric approach (Wilcoxon–Mann–Whitney test; Brem *et al.*, 2002; Yvert *et al.*, 2003) with a parametric test [analysis of variance (ANOVA)]. The non-parametric test uses a user-defined spike value to distinguish qualitative segregation from quantitative differences. The value chosen as the spike was 37, because this value was the noise level in the LC-MS analysis. MetaNetwork also allows setting a threshold for the significance of mQTL by performing permutation tests on samples. A bootstrap procedure was performed with a type I error of 5% (default value of MetaNetwork) for finding an mQTL considering all genetic markers. This procedure rendered a $-^{10}\log(P)$ threshold of 3.8 for individual marker–trait combinations. This threshold was used for all analyses.

Annotation of metabolites

MetaNetwork revealed clusters of mQTLs on the apple genome. The underlying LC-MS mass peaks of mQTL clusters were annotated by comparing their accurate mass and retention time with standards, with the metabolite databases Moto (Moco *et al.*, 2006), KNapSack, Metabolome Japan, and the Dictionary of Natural Products. The results are shown in Supplementary Tables S1 and S2 available at *JXB* online.

Targeted mQTL mapping of annotated metabolites using MapQTL® 6.0

The aim of the study was to look at the genetics of these annotated metabolites in more detail, such as revealing the allelic contributions to traits, and performing analysis of cofactors to filter out the effect of strong mQTLs. MetaNetwork appeared to be less suitable for these deeper analyses of individual metabolites. Therefore, it was decided to use MapQTL® (Van Ooijn, 2009*a*) for this. Interval mapping was applied, followed by rMQM (restricted multiple QTL mapping) with regression algorithm, Haldane's mapping function, with a mapping step size of 1, and independent LOD (logarithm of odds) test statistics. The threshold for mQTL significance was determined using a genome-wide permutation test with 1000 iterations, which gave $\alpha=0.005$ for the 17 chromosomes of apple, to obtain a 95% confidence interval. Two LOD support intervals were used to estimate the range in centiMorgans where the mQTLs reside. Markers near mQTL peaks or at mQTL peaks were used as cofactors for rMQM. This was followed by another round of cofactor selection by using markers from the newly found minor mQTLs from the rMQM. The results from this second round were recorded as the final result. An mQTL was named as a minor QTL if its LOD score was close to or just at the threshold level.

Mapping of the metabolites that segregated as a monogenetic trait

The metabolites procyanidin dimer I, procyanidin dimer II, procyanidin trimer I, procyanidin trimer II, (+)-catechin, and (–)-epicatechin had only one mQTL, segregating in a clear 3:1 ratio. This single locus explained a major part of the variation in the metabolite content (up to 81%; Supplementary Table S3 at *JXB* online). These metabolites were treated as monogenic traits and were integrated into the genetic linkage map by JoinMap 4.0 (Van Ooijn, 2009*b*). This was performed with the aim of locating the positions of the underlying genes more precisely.

Testing additional SSR loci for LG16

To map the monogenetically segregating metabolites more precisely, 17 additional SSR loci at the upper part of LG16 were tested for the 42 progeny that showed recombination in this genetic area. These SSRs along with their primers have been previously published in other apple molecular marker linkage maps (Liebhard *et al.*, 2002; Kenis and Keulemans, 2005; Silfverberg-Dilworth *et al.*, 2006; Celton *et al.*, 2009). The 17 SSR loci along with their primer sequences are listed in Supplementary Table S7 at *JXB* online.

Locating structural genes of the phenylpropanoid pathway in the apple genome

To find the position of the orthologous genes on the 17 chromosomes of apple, the DNA sequences of the structural genes of *A. thaliana* (table 2 in Lillo *et al.*, 2008) were aligned to the entire genome sequence of the apple cv. 'Golden Delicious' (Velasco *et al.*, 2010).

Results

Untargeted mQTL mapping of metabolites showed 669 mQTLs in peel and flesh

From the cross of 'Prima'×'Fiesta', 113 progeny individuals were analysed by accurate LC-MS. A total of 18 582 and 11

817 mass signals were detected in peel and flesh tissues, respectively. Clustering of the mass signals based on their corresponding retention time and abundance profile across samples resulted in 672 centrotypes: 418 and 254 for peel and flesh, respectively. In the following sections, these centrotypes are named metabolites.

In view of the genetic mapping, the distributions of these metabolites were studied. The ^{10}log transformation appeared to provide Gaussian distributions in both peel and flesh for the majority of the metabolites (data not shown). In the untargeted mQTL mapping using MetaNetwork, a total of 669 mQTLs were detected (Table 1), spread over all of the 17 linkage groups of the apple genome (Fig. 1). Not all of the metabolites showed mQTLs; 50% of the metabolites in peel and 44% in flesh exhibited statistically significant mQTLs. Figure 2 shows that several mQTLs had very high –log(*P*) values.

LG16 has a strong hotspot of mQTLs in both parental genotypes and in both peel and flesh for phenolic compounds

The striking thing in Figs 1 and 2 is the strong hotspot of mQTLs on LG16. On other linkage groups such as LG1 and LG13, many mQTLs were also detected, but these were not as strongly clustered as on LG16. The mQTLs on LG16 clustered mainly around a single locus. Notable also is that the hotspot of mQTLs on LG16 was present in both parents and in both tissues, in contrast to the mQTL hotspot on LG8, which was explicitly present in 'Prima' but absent in 'Fiesta' (Fig. 1).

For peel 69 and for flesh 30 metabolites were annotated (Supplementary Tables S1, S2 at *JXB* online). Of the annotated metabolites, 81 out of 99 were phenolic compounds belonging to the two groups of phenylpropanoids and polyphenols (Supplementary Tables S1, S2).

The mQTL hotspot on LG16 is not caused by the co-localizing major locus for pH

Maliepaard *et al.* (1998) previously mapped the pH of apple fruits on LG16 in the same segregating population. They observed monogenic inheritance for low versus high acidity and mapped the corresponding gene. The current mQTL hotspot on LG16 was close to that gene (Fig. 3). Remarkably, both parents had one dominant allele for low pH at the LG16 mQTL hotspot. Maliepaard *et al.*, 1998 denoted the locus as *Ma* for malic acid, being the major acid in apple, although they measured pH rather than malic acid itself. As the pH might influence different enzymatic processes and biochemical reactions in plant cells, differences in the pH may possibly have caused the mQTL hotspot. This hypothesis was evaluated. Both at low and high pH, high levels of the metabolites were found (Fig. 4). Apparently, the dominant allele for high acidity was in repulsion to the allele for a high level of metabolite content in both parents. Consequently, the occurrence of the hotspot was not a side effect of major differences in the pH.

Targeted mQTL mapping

The genetics of these annotated metabolites were studied in more detail using MapQTL® 6.0. Similar to the MetaNetwork analysis, the hotspot with the highest number of mQTLs was detected on LG16. Among the 69 annotated metabolites in the peel, 33 had an mQTL on LG16 (Fig. 5A; Supplementary Table S3 at *JXB* online). The majority of these metabolites represented procyanidins of various chain lengths, including the monomeric building blocks of procyanidins, the flavan-3-ols (+)-catechin and (–)-epicatechin. The same region of LG16 showed mQTLs for quinic acid, phenolic esters, coumaroyl hexoside, kaempferol glycosides, and phloridzin. Interestingly, all these metabolites originate from the phenylpropanoid pathway (Fig. 5A, B). Most of the metabolites had only a single mQTL; however, a few metabolites were found to have some additional minor mQTLs on other LGs (Fig. 5A, B). An rMQM analysis revealed in addition several minor mQTLs (Supplementary Tables S3, S4).

For the 69 annotated metabolites in the peel tissue, mQTLs were located on five different linkage groups (Fig. 5A; Supplementary Table S3). LG1 contained specific mQTLs for quercetin glycosides. There is also an mQTL for kaempferol glycosides on LG1. Glycosides of isorhamnetin had an mQTL on LG13. Chlorogenic acid showed an mQTL on LG17. A distinguished group of mQTLs, mapped on LG8, was formed by alcohol glycosides such as octane-di-ol hexoside and phenylethanol glycoside (Supplementary Table S3).

The genetic loci controlling metabolite content in the peel also appear to control that in the flesh, although less significantly

Like those in peel, mQTLs in flesh were detected on the same five linkage groups. Most of the metabolites showing mQTLs in the peel also showed mQTLs in the flesh; however, the number of mQTLs in the flesh was lower than that in the peel (Figs 1, 2). Like in peel, in flesh kaempferol glycosides also had an mQTL on LG1 and LG16. In

Table 1. MetaNetwork results of mQTL mapping in peel and flesh of apple from the F_1 mapping population from 'Prima'×'Fiesta'

	'Prima' peel	'Prima' flesh	'Fiesta' peel	'Fiesta' flesh	Total peel	Total flesh	Total
No. of metabolites with at least one mQTL	184	77	169	67			
No. of mQTLs	288	101	200	80	488	181	669
No. of markers with at least one mQTL	133 (28%)	62 (13%)	81 (24%)	50 (13%)	214	112	326

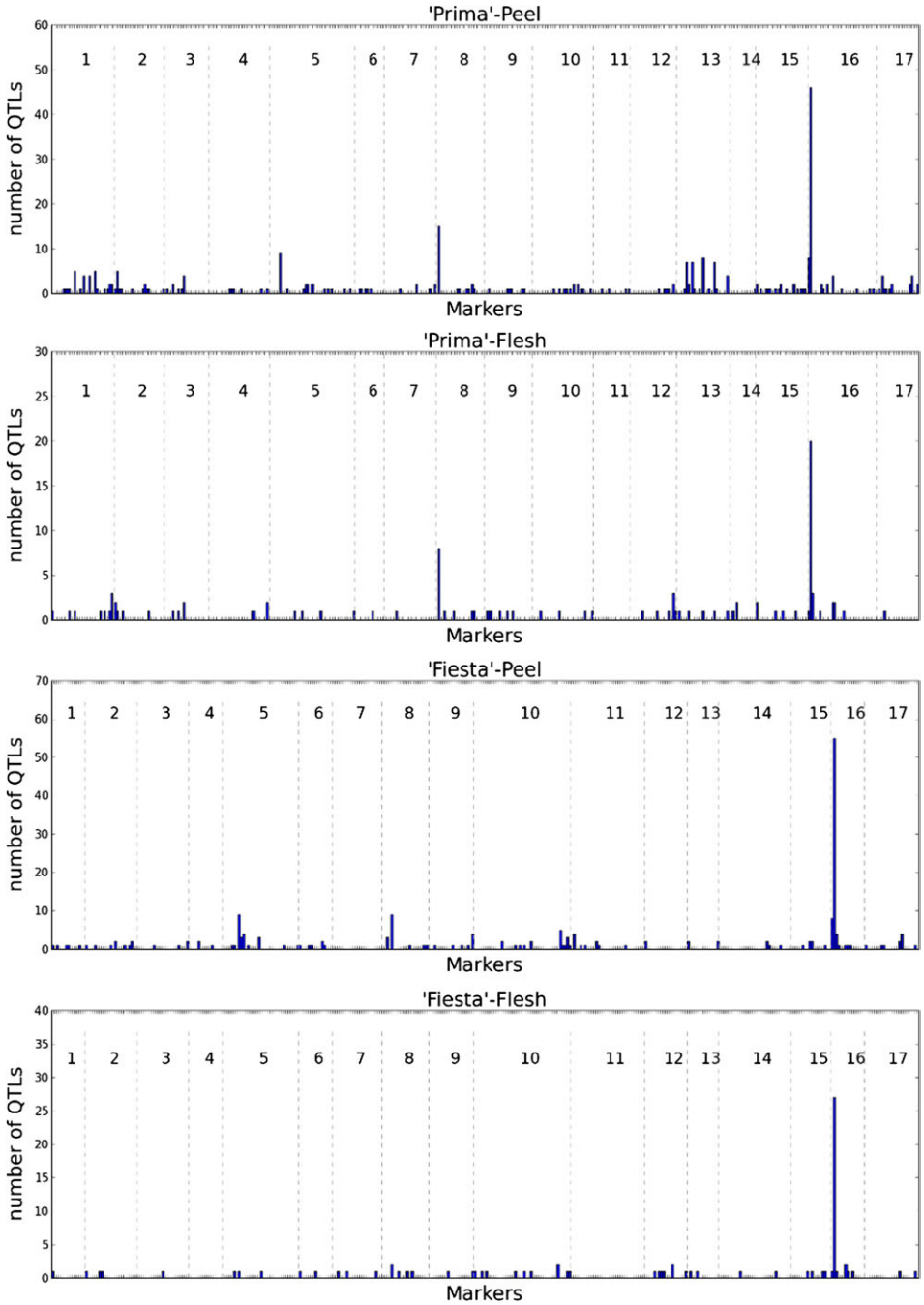

Fig. 1. Number of mQTLs over the apple genome. The linkage groups are separated by vertical dotted lines. In this figure, markers are ordered and positioned equidistantly, thus ignoring their genetic distances.

contrast to peel, quercetin rhamnoside is the only quercetin glycoside which had a clear mQTL on LG8. As in the peel, several octane-di-ol glycosides also had mQTLs on LG8 (Supplementary Table S4 at *JXB* online). LG16 contained mQTLs for procyanidins at the same genetic region as in the peel. In this genetic area on LG16, mQTLs for phenolic esters, (+)-catechin, and (–)-epicatechin were also found (Fig. 5B). Chlorogenic acid, which had an mQTL in the peel on LG17, had a minor mQTL in the same genetic region in the flesh (Supplementary Table S4). Glucuronic acid which is not part of the phenylpropanoid pathway also has an mQTL on LG16 (Supplementary Table S4).

The levels of metabolites in the LG16 mQTL hotspot were controlled by a single, dominant locus present in both parents

The different metabolites were found to have a clear 3:1 segregation and therefore could be mapped. Surprisingly these metabolites mapped on one locus on LG16. Supplementary Tables S3 and S4 at *JXB* online show the effects of different parental allele pairs at the LG16 hotspot on metabolite levels in the progeny, allowing the detection of dominant, recessive, or additive genetic effects. For the hotspot of mQTLs on LG16, both 'Prima' and 'Fiesta' had

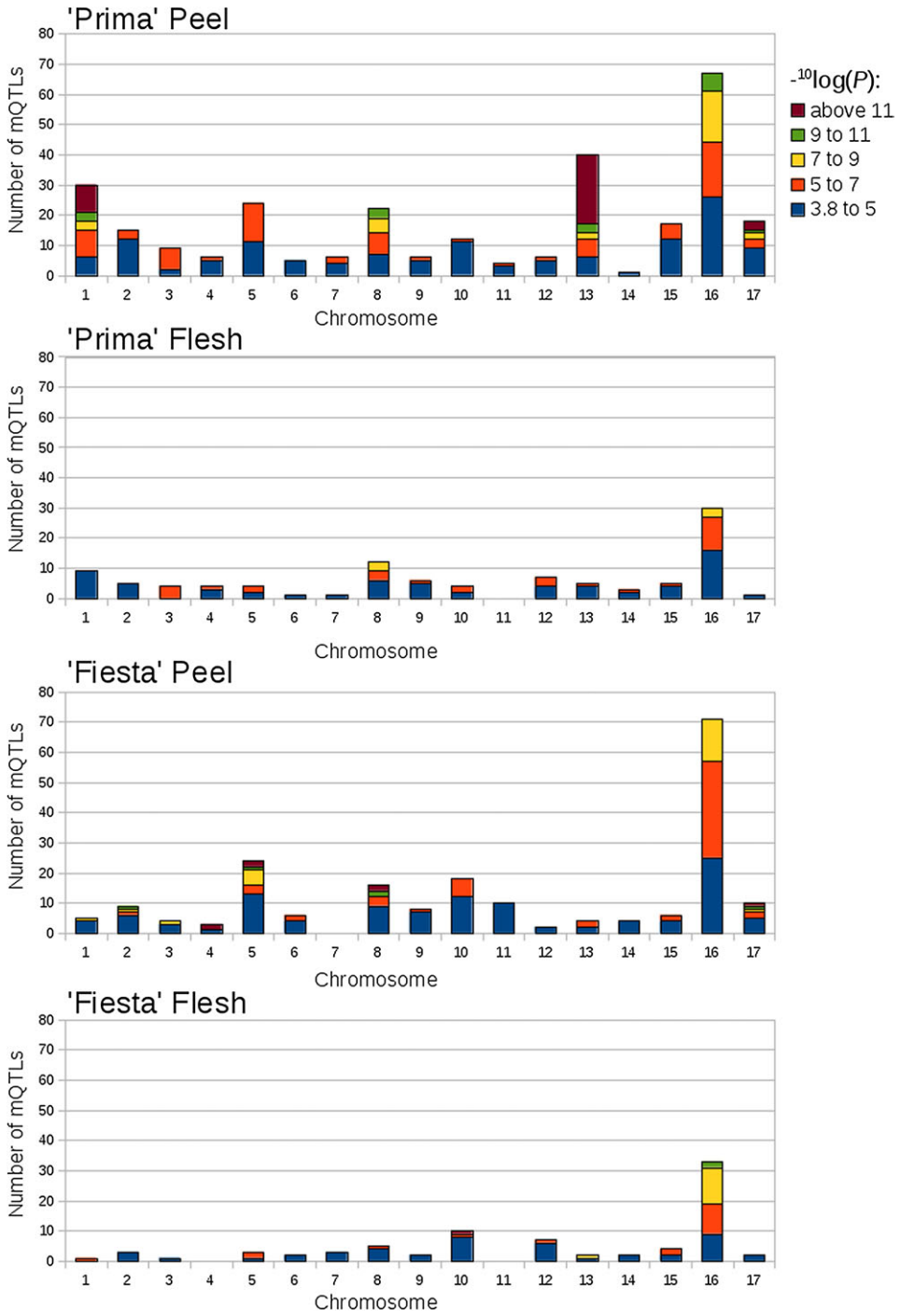

Fig. 2. Significant mQTLs with range of log *P*-values over the apple genome. An mQTL was considered as significant if its -log *(P)*-value was >3.8.

one dominant and one recessive allele each (*Mm*). The combination of the two dominant alleles in the progeny (*MM*) occasionally showed a further increase in the metabolite level, indicating an additive effect or incomplete dominance in these cases.

As LG16 showed a cluster of many mQTLs, this particular hotspot was analysed in more detail. The metabolites (+)-catechin, (–)-epicatechin, two of the procyanidin dimers, and two procyanidin trimers gave one major mQTL per parent, showing two contrasting groups representing monogenic segregation. Figure 4 shows a typical example based on the procyanidin dimer II, indicating a Mendelian 3:1 segregation of *Mm*×*Mm* ($\chi_{3:1}$=0.87; $P > 0.05$), whereby the amount of this metabolite was apparently predominantly controlled by a single locus. Both parents were heterozygous and the *Mm* and *MM* offspring genotypes showed a similar average content for this metabolite, indicating full dominance in both peel and flesh (Fig. 4). The effect of a single dominant allele was on average an increase of

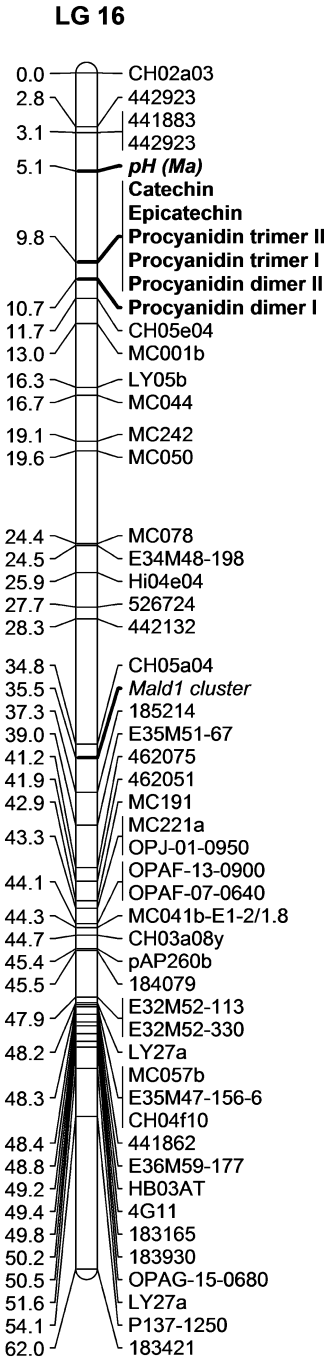

Fig. 3. Mapping of (+)-catechin, (–)-epicatechin, several procyanidins, and pH (*Ma*) on linkage group 16.

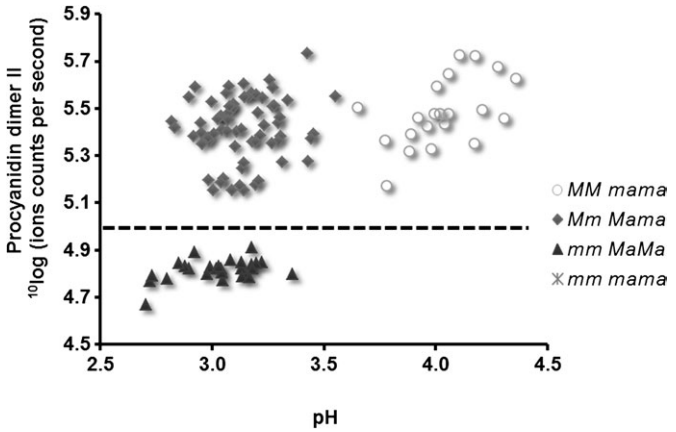

Fig. 4. A scatter plot showing the distribution of F_1 progeny of 'Prima'×'Fiesta' over the four genotype classes for low/high pH and low/high procyanidin dimer II content, whereby procyanidin dimer II represents the metabolites that share the strong mQTL on LG16. The trait pH co-localizes to this hotspot. The dominant allele for a high metabolite level is denoted as *M*, and for a low pH (high acidity; presumably high level of malic acid) as *Ma*. As the dominant alleles *M* and *Ma* are in repulsion phase in both parents, giving as alleles in the gametes *Mma* and *mMa*, the progeny segregate into three genotypes, lacking the genotype *mm mama*. The horizontal dashed line represents the 3:1 clear segregation for the procyanidin dimer II that demonstrates that the two classes (i.e. *mm* and *Mm+MM*) show full dominance.

0.62 compared with the recessive allele on the ^{10}log scale (Fig. 4). This resembles a 4.2-fold increase on a linear scale.

Procyanidins, phenolic esters, (+)-catechin, (–)-epicatechin, and kaempferol hexose rhamnose showed similar segregation patterns, apparently being controlled by the same dominant and recessive alleles of LG16 from both parents. Coumaroyl hexoside and quinic acid appeared to be controlled by the same locus, but, in contrast to phenolic esters and other phenolic compounds, their level was negatively correlated to the other phenolic compounds that mapped at this hotspot (Supplementary Tables S5, S6 at *JXB* online).

Graphical genotyping of the mQTL hotspot on LG16

The metabolites that segregated according to a 3:1 ratio and had only one mQTL behaved as monogenic traits, and could be mapped as genetic markers, which was true for (+)-catechin, (–)-epicatechin, two procyanidin dimers, and two procyanidin trimers. In cases where the relative metabolite level of a progeny was high, it was not clear whether that genotype had inherited the dominant allele from the mother, the father, or from both. Only in the case of a low metabolite level was it evident that both the mother and the father provided the recessive allele. Therefore, as for all dominant markers that segregate in a 3:1 fashion, the marker information could only be used for 25% of the progeny. In spite of this limitation, it was still worthwhile and helpful to locate the genetic window of the locus (Fig. 6).

For more detailed mapping of this locus on LG16, additional SSR markers in the LG16 mQTL hotspot were mapped. Several of the new SSR loci mapped in the genetic window (Fig. 6). The names and allelic sizes of these markers are given in Supplementary Table S7 at *JXB* online. Graphical genotyping of the metabolite 'procyanidin dimer II' revealed that the gene causing the mQTL hotspot on LG16 is located between the locus NH26a and the locus Ch05e04 in 'Prima', and between

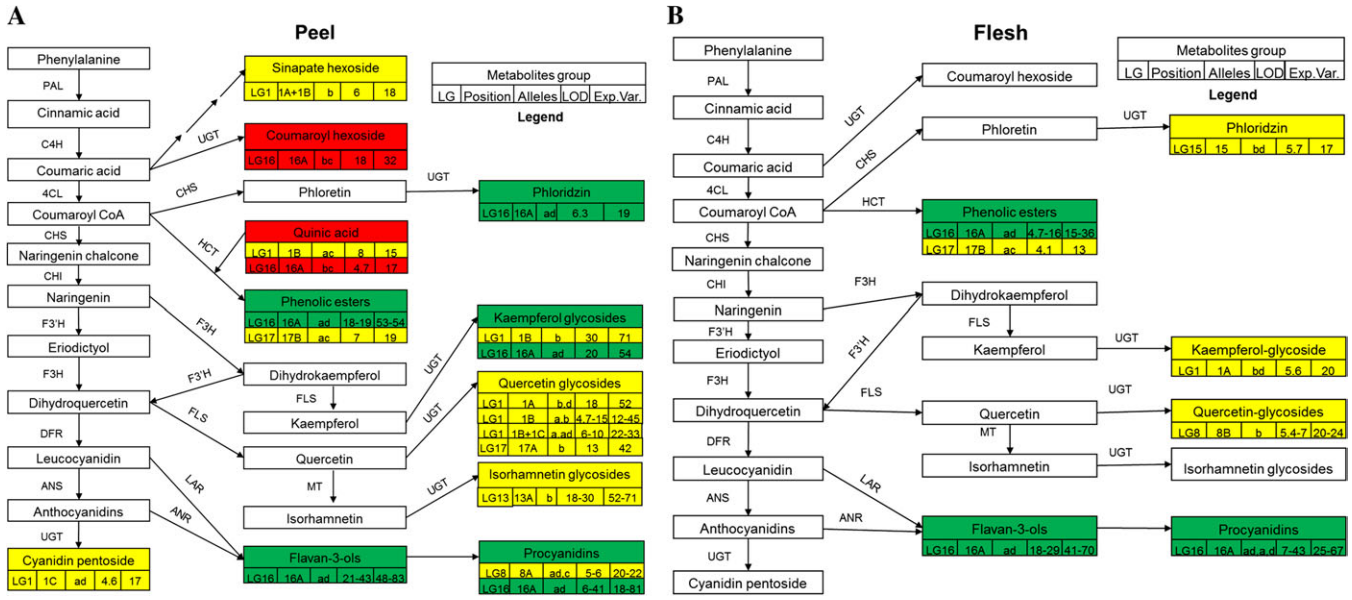

Fig. 5. The phenylpropanoid pathway of phenolic compounds in two apple fruit tissues, peel (A) and flesh (B). The metabolites for which mQTLs were found are presented in coloured boxes. Colourless boxes show the metabolites that were not detected in the present analysis or have no mQTL. Boxes with green colour indicate mQTLs of which the + alleles are in the coupling phase. Boxes with yellow colour show mQTLs for metabolites other than on LG16A. The metabolites in the red box show a negative correlation with the metabolites in the green boxes, having an mQTL on LG16A. The linkage group (LG) where an mQTL was located is given. If different mQTLs were present on different regions of an LG, these regions are distinguished with the letters A, B, C, etc. The alleles 'a' and 'b' originate from the parent 'Prima', and the alleles 'c' and 'd' originate from the parent 'Fiesta', thus following JoinMap codes for outcrossers. As many metabolites in the phenylpropanoid pathway were mapped, for the purpose of simplicity, metabolites that belong to a similar group of compounds are shown as a group (e.g. phenolic esters is a group of several metabolites). Gene names are abbreviated as: *phenylalanine ammonia-lyase* (*PAL*), *cinnamate-4-hydroxylase* (*C4H*), *4-coumaroyl:CoA-ligase* (*4CL*), *chalcone isomerase* (*CHI*), *chalcone synthase* (*CHS*), *flavonone 3' hydroxylase* (*F3'H*), *dihydroflavonol 4-reductase* (*DFR*), *hydroxycinnamoyl-CoA quinate/shikimate hydroxycinnamoyl transferase* (*HCT*), *leucoanthocyanidin 4-reductase* (*LAR*), *UDP-glycosyltransferase* (*UGT*), *flavonol synthase* (*FLS*), and *anthocyanidin synthase* (*ANS*).

the locus Ch02a03 and the locus Hi15a13 in 'Fiesta' (Fig. 6).

The structural gene LAR *and seven transcription factors are at the mQTL hotspot*

The results of the alignments of structural genes of *A. thaliana* against apple are shown in Table 2 and also in Figs 5 and 6. Using the apple genome sequence (http://genomics.research.iasma.it/gb2/gbrowse/apple/), only the structural gene *LAR* was found in the mQTL hotspot on LG16 among the 15 different structural genes of the phenylpropanoid pathway (Figs 5, 6). A closer look revealed that the published 'Golden Delicious' genome sequence had at least five *LAR*-like sequences in overlapping contigs at this locus. However, a sequence homology search using the EMBOSS software package revealed that these sequences were identical or highly homologous, and actually only two different genomic sequences were found. These two sequences probably represent the two alleles for the *LAR* gene. Apart from *LAR*, seven putative transcription factor genes were also identified in the genetic window of the mQTL hotspot on LG16 (Table 3). Two of these belong to the *MYB* class, three to the basic *helix–loop–helix* (*bHLH*) class, one to the *bZIP* class, and one to the *AP2* class of transcription factor genes.

Discussion

Locating genes that are responsible for variation in metabolite levels in the progeny

Many genes involved in the biosynthesis of phenylpropanoids and flavonoids in different plant species have been identified and appear to occur throughout the plant kingdom (Dixon and Steele, 1999; Winkel-Shirley, 2001). In apple, most of the structural genes and several regulatory genes have been isolated (Kim *et al.*, 2003; Takos *et al.*, 2006; Szankowski *et al.*, 2009). However, it is not yet clear which of these genes are critical for the variation in metabolite levels between tissues and genotypes. Genetic mapping of metabolites is a useful step to elucidate which genes are critical for this variation.

By combining the metabolomic data with genetic linkage maps, 488 mQTLs were detected in peel and 254 mQTLs in flesh, using the software MetaNetwork. To gain more insight into the biochemical pathways regulated by the detected mQTLs, the centrotypes that showed highly

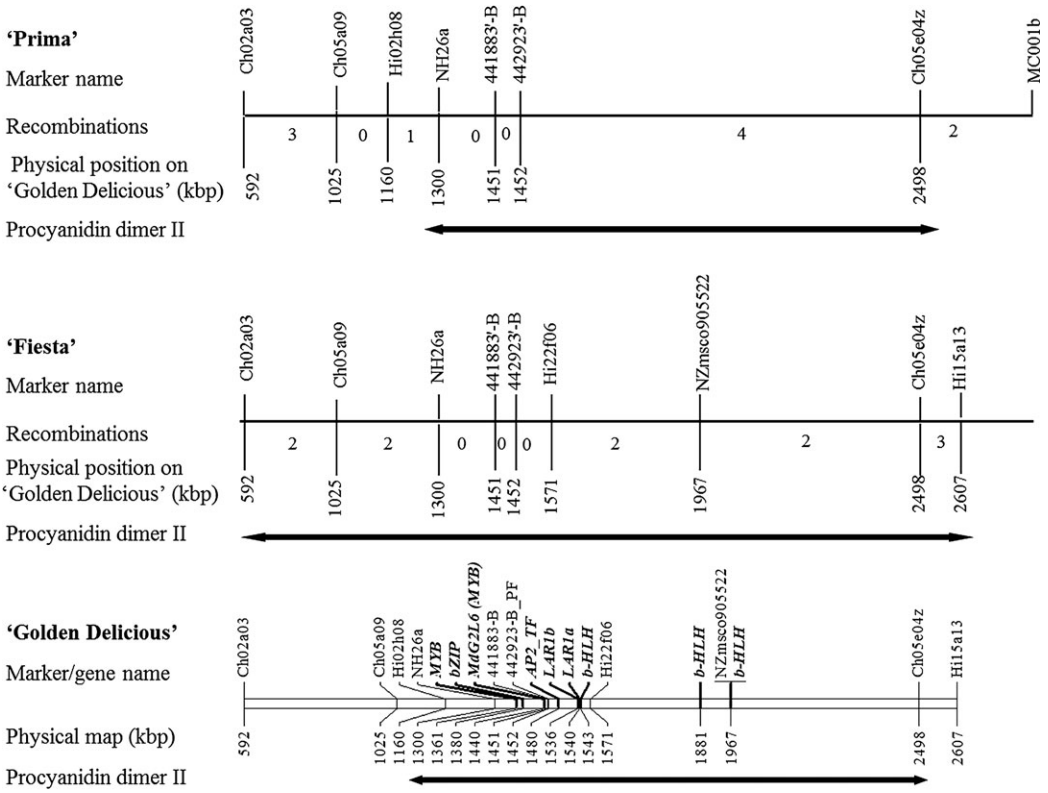

Fig. 6. Genetic linkage maps of 'Prima' and 'Fiesta' and a physical map of the apple cv. 'Golden Delicious' for the mQTL hotspot region on LG16. Procyanidin dimer II was used as a representative for the metabolites that mapped to the LG16 mQTL hotspot. The mQTL regions were genetically mapped as monogenic traits by means of graphical genotyping in both parents and are indicated as horizontal green arrows. The structural gene *leucoanthocyanidin reductase* (*LAR*) of the phenylpropanoid pathway appeared to be present in this region, according to the putative genes in the whole genome sequence of cv. 'Golden Delicious'. Seven putative transcription factor genes including *MYB* and *bHLH* were also detected in this region. The structural gene *LAR* and transcription factor genes are presented in bold text.

probable mQTLs were manually annotated using accurate mass, in-source MS/MS, and UV/Vis absorbance information. The annotation revealed several groups of metabolites in both peel and flesh, including phenylpropanoid esters and glycosides (such as coumaroyl-hexoside, chlorogenic acid, and coumaroyl-quinic acid) and flavonoids [such as (+)-catechin, (–)-epicatechin, procyanidins, quercetin, and kaempferol glycosides] (Supplementary Tables S1, S2 at *JXB* online). Except for the alcohol glycosides and glucuronic acid, all of these metabolites originate from the phenylpropanoid and flavonoid biosynthetic pathway. This bias towards the phenylpropanoid pathway was at least partially caused by the sample preparation (aqueous methanol extracts) and the LC-MS system used [C18-reversed phase LC and ESI ionization in negative mode (Moco *et al.*, 2006)]. Other apple fruit metabolites, such as apolar steroids and volatiles, could not be detected with these parameter settings and thus did not show up in the approach used here.

No mQTL was detected for many metabolites

For about half the number of the metabolites, no mQTL was detected. A reason may be the complex genetic regulation of metabolites by several loci. This would hamper detection of these mQTLs. Another reason can be strong environmental effects compared with genetic effects for some metabolites. A third reason can be that levels of metabolites can be just above noise level for some progeny and within the noise for other progeny. That would also hamper detection of mQTLs. A metabolite is regarded as being present in the case where ⩾10 progeny genotypes show that metabolite above the noise level.

A flavonoid mQTL hotspot is located on LG16

From MetaNetwork analysis, an mQTL hotspot on LG16 was found for both the parents and both the peel and flesh tissues (Figs 1, 2). In the genetic window of the mQTL hotspot, 271 putative genes were detected using the published genome of apple (Velasco *et al.*, 2010; http://www.rosaceae.org/projects/apple_genome). To find the putative underlying gene(s) causing this mQTL hotspot, the structural genes of the phenylpropanoid pathway were positioned on the chromosomes of apple, as shown in Fig. 5 and Table 2. In this case, the position of the structural genes was not determined by means of genetic mapping in a segregating population, but via the alignment

Table 2. Structural genes of the phenylpropanoid pathway in *Arabidopsis* and apple

Gene	Full name	*Arabidopsis* locus	Known apple sequence	Homologues in *Arabidopsis*	Genetic positions in apple	At mQTL hotspot on LG16
4CL	*4-Coumarate-CoA ligase*	At1g65060	GO565912, GR882782, GO577298, GO568847, etc	12	3 (LG1,3,7)	No
ANR	*Anthocyanin reductase*	At1g61720	AY830130	1	3 (LG5)	No
ANS	*Anthocyanin synthase*	At4g22870	AF117269	5	4 (LG6)	No
C3H	*p-Coumarate 3-hydroxylase*	At2g40890	TC28151 (http://compbio.dfci.harvard.edu/)	1	6 (LG8,15)	No
C4H	*Trans-cinnamate 4-monooxygenase*	At2g30490	GO549874	1	2 (LG3,11)	No
CHI	*Chalcone isomerase*	At3g55120	X68978	3	9 (LG1,11,14)	No
CHS	*Chalcone synthase*	At4g34850	X68977	4	3 (LG2,5,7)	No
DFR	*Dihydroflavonol 4-reductase*	At5g42800	AF117268	2	2 (LG8,12)	No
F3'H	*Flavonoid 3'-monooxygenase*	At5g07990	Apple_0223.261.C2.Contig645, Apple_0223.261.C1.Contig644 (http://titan.biotec.uiuc.edu/cgi-bin/ESTWebsite/estima_start?seqSet=apple)	1	2 (LG6,14)	No
F3H	*Flavanone 3-hydroxylase*	At3g51240	AF117270	2	2 (LG2,5)	No
FLS	*Flavonol synthase*	At5g08640	AF119095	2	2 (LG0)	No
HCT	*Shikimate O-hydroxycinnamoyltransferase*	At5g48930	Apple_0223.2950.C2.Contig4990, Apple_0223.850.C1.Contig1757	1	3 (LG9,17)	No
LAR	*Leucoanthocyanidin reductase*	AJ550154[a]	AY830131, AY830132	0	5 (LG13,16)	Yes
PAL	*Phenyalanine ammonium lyase*	At2g37040	Apple_0223.263.C1.Contig648, Apple_0223.215.C2.Contig537, Apple_0223.263.C2.Contig649	4	6 (LG1,4,8,12)	No
UGT	*UDP-dependent glycosyltransferase*	At5g17050	AF117267	3	6 (LG0,1,7,9)	No

[a]Not found in *Arabidopsis* but in *Desmodium uncinatum*.

Table 3. Transcription factor genes at the mQTL hotspot on LG16

Putative gene	Apple sequence	Position on chromosome (bp)	Size (kbp)
MYB	MDP0000375685	1 361 220–1 362 093	1.3
MYB	MDP0000703817	1 440 436–1 442 198	1.7
bHLH	MDP0000319726	1 543 934–1 555 640	11.7
bHLH	MDP0000154272	1 881 558–1 884 164	2.6
bHLH	MDP0000261293	1 967 365–1 970 040	2.6
AP2	MDP0000939633	1 475 660–1 476 865	1.2
bZIP	MDP0000250967	1 376 596–1 386 527	9.9

of the known sequences of these structural genes in *A. thaliana* with the first draft whole-genome sequence of the apple cv. 'Golden Delicious' (Velasco *et al.*, 2010). Only the structural gene *LAR* was found in the mQTL hotspot on LG16. *LAR* catalyses the conversion of leucocyanidins into the flavan-3-ols (+)-catechin and (–)-epicatechin, which are the building blocks of procyanidins (Fig. 5). Both the flavan-3-ols and the procyanidins showed an mQTL in this region. The *LAR* gene may explain the mQTL hotspot on LG16, as 23 procyanidins in peel and 13 procyanidins in flesh were mapped to this locus, as well as the two flavan-3-ols (+)-catechin and (–)-epicatechin in both tissues. This was also observed with the alleles involved: the level of flavan-3-ols was increased by the 'a' allele from 'Prima' and/or the 'd' allele from 'Fiesta' (Fig. 5). These same alleles also increased the level of procyanidins (Fig. 5; Supplementary Tables S3, S4 at *JXB* online).

A contra-indication for *LAR* being the gene responsible for the hotspot is the presence of mQTLs at this locus for chlorogenic acid and coumaroyl quinic acid (phenolic esters), phloridzin, and kaempferol glycosides. These metabolites are upstream of the substrate for *LAR* (Fig. 5).

Furthermore, their levels were simultaneously increased with flavan-3-ols and procyanidins (Fig. 5). A possible explanation for this is the presence of a transcription factor gene that regulates the structural genes for these metabolites. In view of this, a search was made for transcription factor genes at the mQTL hotspot and seven transcription factor genes were detected here. Some of these transcription factor genes belong to the *MYB* and *bHLH* types of transcriptional regulators. One of these transcription factor genes may be responsible for the mQTL hotspot for the phenolic esters and kaempferol glycosides and possibly also for the mQTL hotspot for flavan-3-ols, procyanidins, and coumaroyl hexoside.

Coumaroyl hexoside was negatively correlated with procyanidins, indicating a key role for 4CL

The level of coumaroyl hexoside (Fig. 5A) was negatively correlated with the level of flavan-3-ols and procyanidins (Supplementary Table S5, S6 at *JXB* online) in the progeny, as indicated by a red colour in Fig. 5A. This was also indicated by the alleles that increase the levels of coumaroyl hexoside: whereas the levels of flavan-3-ols and procyanidins were elevated by the presence of the marker alleles 'a' and 'd', the coumaroyl hexoside was elevated when these alleles were absent (Fig. 5A; Supplementary Tables S3, S4). This may be explained by a strong sink effect for the production of flavan-3-ols and procyanidins, thus competing with coumaroyl hexoside for the substrate coumaric acid. This indicates that the enzyme 4-coumarate:CoA ligase (4CL) is the critical factor. If this enzyme has a low activity, coumaroyl acid may accumulate, giving a stronger flow to the side branch that leads to coumaroyl hexoside. However, if 4CL were to be more active, the downstream metabolites would be enhanced, at the cost of coumaric acid and coumaroyl hexoside. Strikingly, the gene *4CL* is not at the mQTL hotspot but on other chromosomes (Table 2). This may indicate a feedback mechanism from downstream genes or downstream metabolites to 4CL.

Figure 5A also shows that quinic acid was negatively correlated with flavan-3-ols, procyanidins, and phenolic esters. This may again be a result of a sink effect. High activity of 4CL may lead to high levels of coumaroyl-CoA, and therewith to high levels of phenolic esters, but also to depletion of the other precursor of phenolic esters, namely quinic acid.

Structural genes of the phenylpropanoid pathway in other studies

Several structural genes for flavonoid biosynthesis have been described in apple. Takos *et al.* (2006) described two *LAR* genes and an *Anthocyanidin reductase* (*ANR*) gene, detected in cDNA from the peel of the red apple cv. 'Cripps Red'. They named the two *LAR* genes *MdLAR1* and *MdLAR2*. BLAST results of the sequences of these genes from apple revealed that only *MdLAR1* is present in the LG16 mQTL hotspot. The *MdLAR2* is located on LG13. A major part of LG16 contains homoeologous sequences of LG13 due to whole genome duplication (Maliepaard *et al.*, 1998; Velasco *et al.*, 2010). Park *et al.* (2006) also identified *LAR* and *ANR* in apple fruits by statistically analysing the expressed sequence tags (ESTs). *ANR* utilizes anthocyanidin and *LAR* uses leucocyanidin as substrate. Both *ANR* and *LAR* participate in synthesis of flavan-3-ol monomers, whereas these monomers are the building blocks of procyanidin polymers (Xie *et al.*, 2003). In grape, *ANR* and *LAR* genes strongly influence procyanidin accumulation and composition during berry development (Bogs *et al.*, 2005).

In pear (*Pyrus communis* L.), a species closely related to apple, Fischer *et al.* (2007) isolated cDNAs for the prominent genes in flavonoid biosynthesis mentioned in Fig. 5 via homology with the apple sequences. They found high homology to apple in the DNA and cDNA. Substrate specificities of the recombinant enzymes expressed in yeast were determined for physiological and non-physiological substrates and found to be in general agreement with the characteristic pear flavonoid metabolite pattern (Fischer *et al.*, 2007). In strawberry, another member of the rosaceous family, genes in the flavonoid pathway could be clearly classified into two groups according to their expression pattern; one having two transcription peaks at early and late stages (i.e. *FaANR*, *FaANS*, *FaCHI*, *FaFHT*, and *FaLAR*), and the other showing an up-regulation trend with a single peak at the turning and/or ripening stage [i.e. *FaDFR*, *FaFGT*, *FaFLS*, and *FaMYB* (Almeida *et al.*, 2007)]. This shows that the expression pattern for flavonoid genes can be different and the fruit stage for the expression of certain flavonoid genes can be very critical.

Transcription factor genes of the phenylpropanoid pathway in other studies

Bogs *et al.* (2007) characterized a grapevine *MYB* transcription factor gene, *VvMYBPA1*. This regulatory gene was shown to be able to activate the *LAR* and *ANR* genes, and several other flavonoid pathway genes in grapevines (Bogs *et al.*, 2007). Two other *MYB* genes, *VvMYBPA2* (Terrier *et al.*, 2009) and *VvMYB5b* (Deluc *et al.*, 2008), were also found to promote procyanidin biosynthesis in grapes. *VvMYBPA1*, *VvMYBPA2*, and *VvMYB5b* were compared with the *MYB* gene in the hotspot on LG16, but no convincing homology was found.

Three transcription factor genes from *Arabidopsis* that regulate procyanidin accumulation have been identified: the *MYB* transcription factor *TRANSPARENT TESTA 2* (Nesi *et al.*, 2001), the *bHLH* transcription factor *TRANSPARENT TESTA* 8 (Nesi *et al.*, 2000), and *TRANSPARENT TESTA GLABRA* (*AtTTG1*; Nesi *et al.*, 2000). For the latter gene, Brueggemann *et al.* (2010) found a functional homologue in apple. Li *et al.* (2007) reported the up-regulatation of mRNA for several structural enzymes of the flavonoid pathway in apple by overexpressing the maize leaf colour (*Lc*) transcription factor gene. A BLAST analysis was performed to identify putative homologues of

these transcription factor genes on the apple genome, particularly in the mQTL hotspot on LG16. None of the above transcription factor genes showed close homologues at this locus, and thus these cannot explain the mQTL hotspot on LG16.

Phenotypic buffering

Fu *et al.* (2009) analysed a segregating *A. thaliana* population for variation in transcript, protein, and metabolite abundance. They mapped QTLs for 40 580 molecular and 139 phenotypic traits, and found six QTL hotspots with major, system-wide effects. For the vast majority of the 500 000 single nucleotide polymorphisms (SNPs) between the two parental lines, no or a minor impact on the phenotype was detected. The authors interpreted this lack of dramatic changes by genetic variation as robustness of the system. The six hotspots are exceptions. These hotspots seem to correspond to a few molecular fragilities of an otherwise robust regulatory system (Fu *et al.*, 2009). In another study, Keurentjes *et al.* (2006) have described these hotspots in more detail for metabolites. Their results show striking similarities to the present results in apple, although the *Arabidopsis* population contained only up to two alleles per gene because of the homozygous parents, whereas in the apple population up to four alleles were present per gene. Although a series of genes are involved in the pathway of phenolic compounds, a large extent of quantitative variation in these compounds is explained by one locus only, namely the hotspot on LG16.

Quercetin glycosides are not controlled by the flavonoid mQTL hotspot on LG16

Quercetin glycosides are commonly found in apple fruits (Van der Sluis et al., 2001). Although quercetin glycosides are part of the phenylpropanoid and flavonoid biosynthetic pathways, they did not exhibit any significant mQTL on LG16 (Fig. 5). Instead, these compounds had mQTLs on LG1 in peel. On LG1, the structural gene *UGT* (*UDP-dependent glycosyltransferase*) is located, which is responsible for the glycosidation of quercetins. This *UGT* gene may be responsible for the mQTL on LG1 (Fig. 5). Takos *et al.* (2006) identified and characterized an *UGT* gene in apple, using a functional genomics approach. As *UGT* genes consist of a large gene family, further studies would be needed to verify which *UGT* gene would be responsible for the glycosidation of quercetin. In flesh, quercetin glycosides showed an mQTL on LG8 (Fig. 5).

Another quercetin derivative, isorhamnetin (a methoxylated form of quercetin), had a strong mQTL on LG13. Possibly, a gene for methoxylation is located on LG13 (Fig. 5). It was observed that both quercetin metabolites (quercetin glycosides and isorhamnetin glycosides) were not dependent on LG16, and no free, unmodified quercetin was detected in apple. Together, these observations suggest that the rate-limiting step in the formation of quercetin derivatives in apple is determined by the modifying enzymes (UGT and *O*-methyltransferase), and that the flux of phenylpropanoid towards quercetin is adapted to the availability of modification opportunities.

Consequences of tight genetic linkage of the dominant alleles for high levels of metabolites to the recessive alleles for pH

Figures 3 and 4 illustrate that in both parents the dominant alleles for high levels of metabolites are genetically tightly linked to the recessive alleles for high pH. This has consequences for apple breeding. In Northern Europe, apples with a low pH are usually preferred to those with a high pH, and therefore should contain the dominant allele for low pH. As the dominant allele for low pH is in repulsion phase to the dominant allele for high metabolites, at least in the genotypes investigated here, the selection for the dominant low pH allele implies the selection for the recessive allele for low levels of the phenolic compounds. Therefore, progeny that are more acidic have higher chances of having lower levels of procyanidins and other phenolic compounds. This can be solved by the selection of progeny that have one dominant allele for low pH from one parent and one dominant allele for high metabolite levels from the other parent. This is the *MmMama* group in Fig. 4. In the southern countries of Europe and in Asia, consumers usually prefer a higher pH. In that case, the desired absence of the dominant allele for pH is automatically combined with the presence of the dominant allele for high levels of metabolites. This is the *MMmama* group in Fig. 4. For apple, it usually takes 6–8 years after sowing to obtain fruits that can be evaluated for pH and metabolites. Selection of the desired progeny for these fruit traits is feasible already at a very young stage, using DNA from leaves of seedlings and DNA markers (Fig. 3).

Follow-up studies

In subsequent studies, the expression profiles of the *MdLAR1* candidate gene, other structural genes of the phenylpropanoid pathway, and the seven transcription factor genes found in the mQTL hotspot will be studied in progeny that have either low or high procyanidin levels. Final proof of their involvement needs complementation studies. Next, to increase the level of these beneficial metabolite(s), the most promising alleles may be inserted into existing, highly popular apple cultivars with low procyanidin levels, for example by means of a cisgenesis approach. Cisgenesis is defined as 'the genetic modification of a recipient plant with natural gene(s) from a sexually compatible plant' (Schouten *et al.*, 2006*a*, *b*). Whatever the outcome of these follow-up studies, the knowledge obtained from the current study of the mQTL hotspot genes is already of use for the breeding of new cultivars with increased levels of these putatively beneficial metabolites through application in marker-assisted breeding.

Supplementary data

Table S1. Annotation of metabolites found in apple peel.

Table S2. Annotation of metabolites found in apple flesh.

Table S3. QTL mapping of metabolites in apple peel.

Table S4. QTL mapping of metabolites in apple flesh.

Table S5. Correlation of metabolites found in peel of apple fruits.

Table S6. Correlation of metabolites found in flesh of apple fruits.

Table S7. Additional simple sequence repeat (SSR) loci tested for LG16.

Acknowledgements

We are grateful to Remmelt Groenwold for his assistance in the orchard. This project was financially supported by PiDON, Inova Fruit B.V., and the Higher Education Commission (HEC) of Pakistan. RdV acknowledges the Centre for Biosystems Genomics and The Netherlands Metabolomics Centre, both initiatives under the auspices of the Netherlands Genomics Initiative, for additional financing.

References

Almeida JRM, D'Amico E, Preuss A, *et al*. 2007. Characterization of major enzymes and genes involved in flavonoid and proanthocyanidin biosynthesis during fruit development in strawberry (*Fragaria×ananassa*). *Archives of Biochemistry and Biophysics* **465,** 61–71.

Awad MA, De Jager A, Van Westing LM. 2000. Flavonoid and chlorogenic acid levels in apple fruit: characterisation of variation. *Scientia Horticulturae* **83,** 249–263.

Bogs J, Downey MO, Harvey JS, Ashton AR, Tanner GJ, Robinson SP. 2005. Proanthocyanidin synthesis and expression of genes encoding leucoanthocyanidin reductase and anthocyanidin reductase in developing grape berries and grapevine leaves. *Plant Physiology* **139,** 652–663.

Bogs J, Jaffé FW, Takos AM, Walker AR, Robinson SP. 2007. The grapevine transcription factor *VvMYBPA1* regulates proanthocyanidin synthesis during fruit development. *Plant Physiology* **143,** 1347–1361.

Brem RB, Yvert G, Clinton R, Kruglyak L. 2002. Genetic dissection of transcriptional regulation in budding yeast. *Science* **296,** 752–755.

Brueggemann J, Weisshaar B, Sagasser M. 2010. A WD40-repeat gene from *Malus* × *domestica* is a functional homologue of Arabidopsis thaliana TRANSPARENT TESTA GLABRA1. *Plant Cell Reports* **29,** 285–294.

Calenge F, Durel CE. 2006. Both stable and unstable QTLs for resistance to powdery mildew are detected in apple after four years of field assessments. *Molecular Breeding* **17,** 329–339.

Calenge F, Faure A, Goerre M, Gebhardt C, Van De Weg WE, Parisi L, Durel CE. 2004. Quantitative trait loci (QTL) analysis reveals both broad-spectrum and isolate-specific QTL for scab resistance in an apple progeny challenged with eight isolates of Venturia inaequalis. *Phytopathology* **94,** 370–379.

Carrari F, Fernie AR. 2006. Metabolic regulation underlying tomato fruit development. *Journal of Experimental Botany* **57,** 1883–1897.

Celton JM, Tustin DS, Chagne D, Gardiner SE. 2009. Construction of a dense genetic linkage map for apple rootstocks using SSRs developed from *Malus* ESTs and Pyrus genomic sequences. *Tree Genetics and Genomes* **5,** 93–107.

Davey MW, Kenis K, Keulemans J. 2006. Genetic control of fruit vitamin C contents. *Plant Physiology* **142,** 343–351.

De Vos RCH, Moco S, Lommen A, Keurentjes JJB, Bino RJ, Hall RD. 2007. Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nature Protocols* **2,** 778–791.

Deluc L, Bogs J, Walker AR, Ferrier T, Decendit A, Merillon JM, Robinson SP, Barrieu F. 2008. The transcription factor *VvMYB5b* contributes to the regulation of anthocyanin and proanthocyanidin biosynthesis in developing grape berries. *Plant Physiology* **147,** 2041–2053.

Dixon RA, Steele CL. 1999. Flavonoids and isoflavonoids—a gold mine for metabolic engineering. *Trends in Plant Science* **4,** 394–400.

Dixon RA, Strack D. 2003. Phytochemistry meets genome analysis, and beyond. *Phytochemistry* **62,** 815–816.

Dunemann F, Ulrich D, Boudichevskaia A, Grafe C, Weber WE. 2009. QTL mapping of aroma compounds analysed by headspace solid-phase microextraction gas chromatography in the apple progeny 'Discovery'×'Prima'. *Molecular Breeding* **23,** 501–521.

Eberhardt MV, Lee CY, Liu RH. 2000. Antioxidant activity of fresh apples. *Nature* **405,** 903–904.

Fischer TC, Gosch C, Pfeiffer J, Halbwirth H, Halle C, Stich K, Forkmann G. 2007. Flavonoid genes of pear (*Pyrus communis*). *Trees – Structure and Function* **21,** 521–529.

Fu JY, Keurentjes JJB, Bouwmeester H, *et al*. 2009. System-wide molecular evidence for phenotypic buffering in. *Arabidopsis. Nature Genetics* **41,** 166–167.

Fu J, Swertz MA, Keurentjes JJB, Jansen RC. 2007. MetaNetwork: a computational protocol for the genetic study of metabolic networks. *Nature Protocols* **2,** 685–694.

Gerhauser C. 2008. Cancer chemopreventive potential of apples, apple juice, and apple components. *Planta Medica* **74,** 1608–1624.

Kenis K, Keulemans J. 2005. Genetic linkage maps of two apple cultivars (*Malus×domestica* Borkh.) based on AFLP and microsatellite markers. *Molecular Breeding* **15,** 205–219.

Kenis K, Keulemans J. 2007. Study of tree architecture of apple (*Malus×domestica* Borkh.) by QTL analysis of growth traits. *Molecular Breeding* **19,** 193–208.

Keurentjes JJB, Fu JY, de Vos CHR, Lommen A, Hall RD, Bino RJ, van der Plas LHW, Jansen RC, Vreugdenhil D, Koornneef M. 2006. The genetics of plant metabolism. *Nature Genetics* **38,** 842–849.

Khan MA, Duffy B, Gessler C, Patocchi A. 2006. QTL mapping of fire blight resistance in apple. *Molecular Breeding* **17,** 299–306.

Kim SH, Lee JR, Hong ST, Yoo YK, An G, Kim SR. 2003. Molecular cloning and analysis of anthocyanin biosynthesis genes preferentially expressed in apple skin. *Plant Science* **165,** 403–413.

King GJ, Lynn JR, Dover CJ, Evans KM, Seymour GB. 2001. Resolution of quantitative trait loci for mechanical measures accounting for genetic variation in fruit texture of apple (*Malus pumila* Mill.). *Theoretical and Applied Genetics* **102,** 1227–1235.

Li H, Flachowsky H, Fischer TC, Hanke MV, Forkmann G, Treutter D, Schwab W, Hoffmann T, Szankowski I. 2007. Maize *Lc* transcription factor enhances biosynthesis of anthocyanins, distinct proanthocyanidins and phenylpropanoids in apple (*Malus domestica* Borkh. *Planta* **226,** 1243–1254.

Liebhard R, Gianfranceschi L, Koller B, Ryder CD, Tarchini R, Van De Weg E, Gessler C. 2002. Development and characterisation of 140 new microsatellites in apple (Malus×domestica Borkh.). *Molecular Breeding* **10,** 217–241.

Liebhard R, Kellerhals M, Pfammatter W, Jertmini M, Gessler C. 2003. Mapping quantitative physiological traits in apple (*Malus×domestica* Borkh.). *Plant Molecular Biology* **52,** 511–526.

Lillo C, Lea US, Ruoff P. 2008. Nutrient depletion as a key factor for manipulating gene expression and product formation in different branches of the flavonoid pathway. *Plant, Cell and Environment* **31,** 587–601.

Lommen A. 2009. Metalign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. *Analytical Chemistry* **81,** 3079–3086.

Lu Y, Foo LY. 1997. Identification and quantification of major polyphenols in apple pomace. *Food Chemistry* **59,** 187–194.

Maliepaard C, Alston FH, Van Arkel G, *et al*. 1998. Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. *Theoretical and Applied Genetics* **97,** 60–73.

Mazza G, Velioglu YS. 1992. Anthocyanins and other phenolic compounds in fruits of red-flesh apples. *Food Chemistry* **43,** 113–117.

Mcghie TK, Hunt M, Barnett LE. 2005. Cultivar and growing region determine the antioxidant polyphenolic concentration and composition of apples grown in New Zealand. *Journal of Agricultural and Food Chemistry* **53,** 3065–3070.

Moco S, Bino RJ, Vorst O, Verhoeven HA, De Groot J, Van Beek TA, Vervoort J, Ric De Vos CH. 2006. A liquid chromatography–mass spectrometry-based metabolome database for tomato. *Plant Physiology* **141,** 1205–1218.

Nesi N, Debeaujon I, Jond C, Pelletier G, Caboche M, Lepiniec L. 2000. The TT8 gene encodes a basic helix–loop–helix domain protein required for expression of DFR and BAN genes in Arabidopsis siliques. *The Plant Cell* **12,** 1863–1878.

Nesi N, Jond C, Debeaujon I, Caboche M, Lepiniec L. 2001. The Arabidopsis TT2 gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. *The Plant Cell* **13,** 2099–2114.

Park S, Sugimoto N, Larson MD, Beaudry R, Van Nocker S. 2006. Identification of genes with potential roles in apple fruit development and biochemistry through large-scale statistical analysis of expressed sequence tags. *Plant Physiology* **141,** 811–824.

Schouten HJ, Krens FA, Jacobsen E. 2006*a*. Cisgenic plants are similar to traditionally bred plants: international regulations for genetically modified organisms should be altered to exempt cisgenesis. *EMBO Reports* **7,** 750–753.

Schouten HJ, Krens FA, Jacobsen E. 2006*b*. Do cisgenic plants warrant less stringent oversight? [6]. *Nature Biotechnology* **24,** 753.

Schouten HJ, van de Weg WE, Carling J, *et al*. 2011. Diversity arrays technology (DArT) markers in apple for genetic linkage maps. *Molecular Breeding,* 1–16.

Silfverberg-Dilworth E, Matasci CL, Van De Weg WE, *et al*. 2006. Microsatellite markers spanning the apple (*Malus×domestica* Borkh.) genome. *Tree Genetics and Genomes* **2,** 202–224.

Szankowski I, Li H, Flachowsky H, Höfer M, Hanke MV, Fischer T, Forkmann G, Treutter D, Schwab W, Hoffmann T. 2009. Metabolic engineering of flavonoid biosynthesis in apple (*Malus×domestica* Borkh.). *Acta Horticulturae* **814,** 511–516.

Takos AM, Ubi BE, Robinson SP, Walker AR. 2006. Condensed tannin biosynthesis genes are regulated separately from other flavonoid biosynthesis genes in apple fruit skin. *Plant Science* **170,** 487–499.

Terrier N, Torregrosa L, Ageorges A, Vialet S, Verriés C, Cheynier V, Romieu C. 2009. Ectopic expression of *VvMybPA2* promotes proanthocyanidin biosynthesis in grapevine and suggests additional targets in the pathway. *Plant Physiology* **149,** 1028–1041.

Tikunov YM, Laptenok S, Hall RD, Bovy A, de Vos RCH. 2011. MSClust: a tool for unsupervised mass spectra extraction of chromatography–mass spectrometry ion-wise aligned data. *Metabolomics,* 1–5.

Treutter D. 2001. Biosynthesis of phenolic compounds and its regulation in apple. *Plant Growth Regulation* **34,** 71–89.

Van der Sluis AA, Dekker M, De Jager A, Jongen WMF. 2001. Activity and concentration of polyphenolic antioxidants in apple: effect of cultivar, harvest year, and storage conditions. *Journal of Agricultural and Food Chemistry* **49,** 3606–3613.

Van Ooijn JW. 2009*a*. MapQTL® 6.0. *Sofware for the mapping of quantitative trait loci in experimental populations of diploid species*. Wageningen, The Netherlands: Kyazma BV.

Van Ooijn J. 2009*b*. JoinMap® 4. *Software for the calculation of genetic linkage maps in experimental populations*. Wageningen, The Netherlands: Kyazma BV.

Velasco R, Zharkikh A, Affourtit J, *et al*. 2010. The genome of the domesticated apple (*Malus* × *domestica* Borkh.). *Nature Genetics* **42,** 833–839.

Winkel-Shirley B. 2001. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiology* **126,** 485–493.

Xie DY, Sharma SB, Paiva NL, Ferreira D, Dixon RA. 2003. Role of anthocyanidin reductase, encoded by *BANYULS* in plant flavonoid biosynthesis. *Science* **299,** 396–399.

Yvert G, Brem RB, Whittle J, Akey JM, Foss E, Smith EN, Mackelprang R, Kruglyak L. 2003. Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nature Genetics* **35,** 57–64.