

# *Genome-wide association mapping of flowering and ripening periods in apple*

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open access

Urrestarazu, J., Muranty, H., Denancé, C., Leforestier, D., Ravon, E., Guyader, A., Guisnel, R., Feugey, L., Aubourg, S., Celton, J.-M., Daccord, N., Dondini, L., Gregori, R., Lateur, M., Houben, P., Ordidge, M. ORCID: <https://orcid.org/0000-0003-0115-5218>, Paprstein, F., Sedlak, J., Nybom, H., Garkava-Gustavsson, L., Troggio, M., Bianco, L., Velasco, R., Poncet, C., Théron, A., Moriya, S., Bink, M. C. A. M., Laurens, F., Tartarini, S. and Durel, C.-E. (2017) Genome-wide association mapping of flowering and ripening periods in apple. *Frontiers in Plant Science*, 8. 01923. ISSN 1664-462X doi: 10.3389/fpls.2017.01923 Available at <https://centaur.reading.ac.uk/73385/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.3389/fpls.2017.01923>

Publisher: Frontiers

copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

## **CentAUR**

Central Archive at the University of Reading

Reading's research outputs online



# Genome-Wide Association Mapping of Flowering and Ripening Periods in Apple

## OPEN ACCESS

### Edited by:

Marion S. Röder,  
Leibniz-Institut für Pflanzengenetik und  
Kulturpflanzenforschung (IPK),  
Germany

### Reviewed by:

Ahmad M. Alqudah,  
Leibniz-Institut für Pflanzengenetik und  
Kulturpflanzenforschung (IPK),  
Germany  
Robert Henry,  
The University of Queensland,  
Australia  
Raquel Sánchez-Pérez,  
University of Copenhagen, Denmark

### \*Correspondence:

Jorge Urrestarazu  
jorge.urrestarazu@unavarra.es  
Charles-Eric Durel  
charles-eric.durel@inra.fr

†These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Plant Genetics and Genomics,  
a section of the journal  
Frontiers in Plant Science

**Received:** 31 July 2017

**Accepted:** 24 October 2017

**Published:** 10 November 2017

### Citation:

Urrestarazu J, Muranty H, Denancé C,  
Leforestier D, Ravon E, Guyader A,  
Guisnel R, Feugey L, Aubourg S,  
Celton J-M, Daccord N, Dondini L,  
Gregori R, Lateur M, Houben P,  
Ordridge M, Paprstein F, Sedlak J,  
Nybohm H, Garkava-Gustavsson L,  
Troggio M, Bianco L, Velasco R,  
Poncet C, Theron A, Moriya S and  
Bink MCAM, Laurens F, Tartarini S and  
Durel C-E (2017) Genome-Wide  
Association Mapping of Flowering and  
Ripening Periods in Apple.  
Front. Plant Sci. 8:1923.  
doi: 10.3389/fpls.2017.01923

Jorge Urrestarazu<sup>1,2,3\*</sup>, Hélène Muranty<sup>1†</sup>, Caroline Denancé<sup>1</sup>, Diane Leforestier<sup>1</sup>,  
Elisa Ravon<sup>1</sup>, Arnaud Guyader<sup>1</sup>, Rémi Guisnel<sup>1</sup>, Laurence Feugey<sup>1</sup>, Sébastien Aubourg<sup>1</sup>,  
Jean-Marc Celton<sup>1</sup>, Nicolas Daccord<sup>1</sup>, Luca Dondini<sup>2</sup>, Roberto Gregori<sup>2</sup>, Marc Lateur<sup>4</sup>,  
Patrick Houben<sup>4</sup>, Matthew Ordridge<sup>5</sup>, Frantisek Paprstein<sup>6</sup>, Jiri Sedlak<sup>6</sup>, Hilde Nybohm<sup>7</sup>,  
Larisa Garkava-Gustavsson<sup>8</sup>, Michela Troggio<sup>9</sup>, Luca Bianco<sup>9</sup>, Riccardo Velasco<sup>9</sup>,  
Charles Poncet<sup>10</sup>, Anthony Theron<sup>10</sup>, Shigeki Moriya<sup>1,11</sup>, Marco C. A. M. Bink<sup>12,13</sup>,  
François Laurens<sup>1</sup>, Stefano Tartarini<sup>2</sup> and Charles-Eric Durel<sup>1\*</sup>

<sup>1</sup> Institut de Recherche en Horticulture et Semences UMR 1345, INRA, SFR 4207 QUASAV, Beaucouzé, France,

<sup>2</sup> Department of Agricultural Sciences, University of Bologna, Bologna, Italy, <sup>3</sup> Department of Agricultural Sciences, Public

University of Navarre, Pamplona, Spain, <sup>4</sup> Plant Breeding and Biodiversity, Centre Wallon de Recherches Agronomiques,

Gembloux, Belgium, <sup>5</sup> School of Agriculture, Policy and Development, University of Reading, Reading, United Kingdom,

<sup>6</sup> Research and Breeding Institute of Pomology Holovousy Ltd., Horice, Czechia, <sup>7</sup> Department of Plant Breeding, Swedish

University of Agricultural Sciences, Kristianstad, Sweden, <sup>8</sup> Department of Plant Breeding, Swedish University of Agricultural

Sciences, Alnarp, Sweden, <sup>9</sup> Fondazione Edmund Mach, San Michele all'Adige, Italy, <sup>10</sup> Plateforme Gentyane, INRA, UMR

1095 Genetics, Diversity and Ecophysiology of Cereals, Clermont-Ferrand, France, <sup>11</sup> Apple Research Station, Institute of

Fruit Tree and Tea Science, National Agriculture and Food Research Organization (NARO), Morioka, Japan, <sup>12</sup> Wageningen

UR, Biometris, Wageningen, Netherlands, <sup>13</sup> Hendrix Genetics, Boxmeer, Netherlands

Deciphering the genetic control of flowering and ripening periods in apple is essential for breeding cultivars adapted to their growing environments. We implemented a large Genome-Wide Association Study (GWAS) at the European level using an association panel of 1,168 different apple genotypes distributed over six locations and phenotyped for these phenological traits. The panel was genotyped at a high-density of SNPs using the Axiom® Apple 480 K SNP array. We ran GWAS with a multi-locus mixed model (MLMM), which handles the putatively confounding effect of significant SNPs elsewhere on the genome. Genomic regions were further investigated to reveal candidate genes responsible for the phenotypic variation. At the whole population level, GWAS retained two SNPs as cofactors on chromosome 9 for flowering period, and six for ripening period (four on chromosome 3, one on chromosome 10 and one on chromosome 16) which, together accounted for 8.9 and 17.2% of the phenotypic variance, respectively. For both traits, SNPs in weak linkage disequilibrium were detected nearby, thus suggesting the existence of allelic heterogeneity. The geographic origins and relationships of apple cultivars accounted for large parts of the phenotypic variation. Variation in genotypic frequency of the SNPs associated with the two traits was connected to the geographic origin of the genotypes (grouped as North+East, West and South Europe), and indicated differential selection in different growing environments. Genes encoding transcription factors containing either NAC or MADS domains were identified as major candidates within the small confidence intervals computed for the associated genomic regions. A

strong microsynteny between apple and peach was revealed in all the four confidence interval regions. This study shows how association genetics can unravel the genetic control of important horticultural traits in apple, as well as reduce the confidence intervals of the associated regions identified by linkage mapping approaches. Our findings can be used for the improvement of apple through marker-assisted breeding strategies that take advantage of the accumulating additive effects of the identified SNPs.

**Keywords:** adaptive traits, association genetics, germplasm collection, GWAS, *Malus* × *domestica* Borkh., microsynteny, quantitative trait loci, SNP

## INTRODUCTION

Flowering time in temperate plants is influenced by multiple environmental factors related to temperature and day length at different periods of the year (Wilczek et al., 2009; Cook et al., 2012; Abbott et al., 2015). For crop cultivation, floral timing is of utmost importance, because it is a major yield determinant (Jung et al., 2017). Temperate fruit trees use bud dormancy for adaption to seasonality (Campoy et al., 2011; Sánchez-Pérez et al., 2014; Ionescu et al., 2017); flowering occurs uniformly when the chilling and heating requirements associated with winter and spring have been fulfilled. In the context of global climate change, increasing temperatures tend to result in an acceleration of springtime phenological events (Hänninen and Tanino, 2011; Cook et al., 2012), with implications for both the risk of frost damage (Cannell and Smith, 1986; Vitasse et al., 2014) and the photosynthetic capacity of the trees (Ensminger et al., 2008). Moreover, this advance is responsible for several morphological disorders/abnormalities, including bud burst delay, low burst rate, irregular floral or leaf budbreak and poor fruit set (Erez, 2000; Celton et al., 2011; Dirlwanger et al., 2012; Abbott et al., 2015). Disruptions in synchronization of flowering may disturb pollination for self-incompatible cultivars, while modifications of fruit harvesting periods can cause problems with orchard management and fruit marketing (Dirlwanger et al., 2012). Breeding programs mainly focus on improvement of yield and fruit quality, but additional objectives like climate change adaptation receive increased attention. Genetic control of flowering and ripening periods plays a crucial role, since adaptation to different growing environments affects fruit quality (Chagné et al., 2014; Jung et al., 2017).

Flowering time is regulated by an intricate signaling network of multiple genes that integrates both endogenous and exogenous stimuli to induce flowering under the most favorable conditions

(Boss et al., 2004; Amasino, 2005). Fruit ripening control involves coordinated regulation of many metabolic pathways (Johnston et al., 2009; Pirona et al., 2013; Chagné et al., 2014), resulting in the conversion of starch to sugars, reduced acidity, reduced flesh firmness, changes in color and an increase in aroma/flavor volatile compounds. Both traits are quantitatively inherited in most fruit tree species (Celton et al., 2011; Pirona et al., 2013; Castède et al., 2014; Chagné et al., 2014).

Association mapping exploits the linkage disequilibrium (LD) present among individuals from natural populations or germplasm collections to dissect the genetic basis of complex trait variation (Neale and Savolainen, 2004; Aranzana et al., 2005; Balding, 2006; Myles et al., 2009). Germplasm collections generally contain more genetic diversity than segregating progenies and, since association mapping exploits all the recombination events that have occurred in the evolutionary history of the association panel, a much higher mapping resolution is expected (Zhu et al., 2008; Myles et al., 2009; Ingvarsson and Street, 2011). In addition, the number of QTLs that can be mapped for a given phenotype is not limited to the segregation products in a specific cross, but rather by the number of QTLs underlying the trait and the degree to which the studied population captures the genetic species-wide diversity (Zhu et al., 2008; Yano et al., 2016). Association mapping has recently been applied to fruit tree species such as peach (Micheletti et al., 2015), apricot (Mariette et al., 2016) and apple (Leforestier et al., 2015; Migicovsky et al., 2016; Di Guardo et al., 2017; Farneti et al., 2017), especially after the release of high-density SNP arrays with uniform coverage of the whole genome (Chagné et al., 2012; Verde et al., 2012; Bianco et al., 2014) or Genotyping-by-Sequencing (Gardner et al., 2014).

The high density Axiom® Apple 480 K SNP array (Bianco et al., 2016) developed within the EU-FruitBreedomics project (Laurens et al., 2012; <http://www.fruitbreedomics.com>) was used for the first time in the present study to perform GWAS. Here, we focused on the analysis of the genetic control of flowering and ripening periods in a panel of almost 1,200 different genotypes distributed over six apple collections managed by six European institutes. We identified one genomic region associated with flowering period and three with ripening period. Co-variation between the genotypic frequencies at the significant SNPs and three major geographic groupings of genotypes was explored, and candidate genes were identified in the detected genomic

**Abbreviations:** CRA-W, Centre Wallon de Recherche Agronomique [Gembloux (Belgium)]; EBIC, Extended Bayesian Information Criterion; GDDH13 genome, version (v1.1) released for the apple genome based on the doubled haploid GDDH13; GWAS, Genome-Wide Association Study; INRA, Institut National de la Recherche Agronomique [Angers (France)]; LD, linkage disequilibrium; MLM, multi-locus mixed model; NFC, University of Reading [Brogdale (United Kingdom)]; PCA, Principal Component Analysis; QTL, Quantitative Trait Loci; RBIPH, Research and Breeding Institute of Pomology Holovousy [Holovousy (Czech Republic)]; SLU, Swedish University of Agricultural Sciences [Alnarp (Sweden)]; SNP, Single Nucleotide Polymorphism; UNIBO, University of Bologna [Bologna (Italy)].

regions. To our knowledge, this is the largest association study ever performed in a fruit tree species considering both population size and SNP marker density.

## MATERIALS AND METHODS

### Plant Material

The association panel consisted of 1,168 different diploid apple genotypes corresponding to accessions preserved in six European germplasm collections (Table S1). The uniqueness of these genotypes was confirmed with SSR markers in a previous analysis (Urrestarazu et al., 2016). Some accessions corresponding to genotypes present at multiple locations were maintained in order to adjust phenotypic data between collections. Especially, ten standard genotypes (“Alkmene,” “Ananas Reinette,” “Discovery,” “Golden Delicious,” “Ingrid Marie,” “James Grieve,” “Jonathan,” “Reine des Reinettes” (= “King of the Pippins”), “Reinette de Champagne” and “Winter Banana”) were included from almost all collections. The association panel comprised mainly genotypes corresponding to old local/national cultivars, and the majority could be classified into three geographic groups according to their area of origin in Europe [North+East (141 different genotypes), South (148) and West (775)]; the remaining 104 corresponded to recent cultivars, germplasm originating from other worldwide regions, or were of unknown origin (Urrestarazu et al., 2016).

### Phenotypic Data Analysis

Phenotypic data for flowering and ripening periods were scored on an ordinal scale from 1 to 9, and consisted of both historical data in germplasm databases and of new data acquired in recent years (2012–2014) using the same scoring scales. Flowering period was assessed by recording dates of Fleckinger phenological flower stages F or F2 (Fleckinger, 1964), and then assigning a score on the ordinal scale by comparison to reference cultivars. Assessments for flowering period were performed over a period of 3–19 years except for NFC where only a single average value (assessed over 10 years) was available (Table 1). Ripening period was determined by observing pre-ripening drop of healthy fruits, ground- and over-color of fruits, taste of fruits and/or iodine starch index. It was recorded over 3–13 years (Table 1).

Genotypic means obtained for each genotype by adjusting for year and site effects were used as phenotypes for association analysis. When analyzing individual collections, the genotypic means were estimated using a linear model taking into account the year effect (Equation 1), while we considered the combined effect of site and year (Equation 2) for the whole analysis, i.e., all the collections were combined into a single analysis:

$$P_{ik} = \mu + Y_i + g_k + e_{ik} \quad (1)$$

$$P_{ijk} = \mu + (Y_i \times S_j) + g_k + e_{ijk} \quad (2)$$

where for (Equation 1),  $P_{ik}$  is the phenotypic value of the  $k$ th genotype in the  $i$ th year;  $\mu$  is the mean value of the trait;  $Y_i$  is the fixed effect of the  $i$ th year on the trait;  $g_k$  is the random genotypic effect of genotype  $k$ ; and  $e_{ik}$  is the residual term of the

model. For (Equation 2),  $\mu$ ,  $Y_i$ , and  $g_k$  have the same meanings as in (Equation 1);  $P_{ijk}$  refers to the phenotypic value of the  $k$ th genotype in the  $i$ th year in the  $j$ th site;  $S_j$  is the fixed effect of the  $j$ th site; and  $e_{ijk}$  is the residual term of the model. Heritability of genotypic means ( $h^2$ , here called broad-sense heritability) was estimated for each individual collection as:

$$h^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_\varepsilon^2}{n}} \quad (3)$$

where  $\sigma_G^2$  is the variance of genotype effect,  $\sigma_\varepsilon^2$  is the variance of the residual term, and  $n$  is the mean number of observations per genotype. These analyses were performed using “R” software (R Core Team, 2014), in particular the packages effects (Fox, 2003), lme4 (Bates and Sarkar, 2007) and FactoMineR (Lê et al., 2008).

### SNP Genotyping

The 1,168 apple genotypes were genotyped with the Axiom® Apple 480 K array containing 487,249 SNPs evenly distributed over the 17 apple chromosomes (Bianco et al., 2016). Bianco et al. (2016) applied stringent filters that resulted in a set of 275,223 robust SNPs for GWAS. Further details on the development of the SNP array, genotyping process, and the filtering pipeline procedure can be found in Bianco et al. (2016). All presented results use the SNP positions on the latest version (v1.1) released for the apple genome based on the doubled haploid GDDH13 (hereafter, GDDH13 genome; Daccord et al., 2017; see also <https://iris.angers.inra.fr/gddh13/> for the genome browser).

### Kinship, Population Structure and Linkage Disequilibrium Estimates

GEMMA software (Zhou and Stephens, 2012) was used to estimate the standardized relatedness matrix (**K**) between the genotypes at the whole population level and within each collection. A Principal Component Analysis (PCA) of the SNP data was performed using PLINK (Purcell et al., 2007) and the ten largest Eigenvalues were used to control for population structure (**Q**). Matrix **Q** was constructed for the whole population as well as for each collection separately.

LD was studied between sets of SNPs spanning regions of 10 kb randomly sampled along the genome. These sets were obtained by a random choice of 50 contigs larger than 10 kb on each chromosome, followed by the selection of SNPs spanning a random region of 10 kb in each of these contigs. LD was estimated as squared allele frequency correlations ( $r^2$ ) and as  $r^2$  corrected for population structure and relatedness ( $r_{vs}^2$ ) using the R-package LDcorSV (Mangin et al., 2012). In addition, local LD ( $r^2$ ) was assessed for chromosomal regions of 1 Mb surrounding the SNPs retained as cofactors in the GWAS (see next section for details) and displayed in LD maps and network plots using “LDheatmap” and “network” R-packages, respectively.

### Genome-Wide Association Study (GWAS)

The GWAS method was applied both at the whole population level and for each collection independently. GWAS were conducted with correction for population structure (**Q**) and



**TABLE 1** | Averages and ranges for the genotypic means for flowering and ripening periods.

Population	Phenotypic assessments			Genotypic adjusted means		
	No. years (range)	data/cvr.	bs_h <sup>2</sup>	Mean	SD	Range
<b>FLOWERING PERIOD</b>						
Whole population	29 (1985–2014)	5.45	0.82	4.79	1.15	1.73–9.24
INRA	4 (2009–2012)	3.00	0.88	5.58	1.45	2.56–9.52
UNIBO	6 (1987–1992)	7.60	0.84	5.26	0.99	2.57–7.81
CRA-W	19 (1985–2007)	4.90	0.88	3.97	1.03	1.25–7.25
RBIPH	13 (1995–2010)	5.00	0.85	4.42	0.83	3.03–8.61
NFC	1 <sup>a</sup>	–	–	4.91	0.92	2.00–9.00
SLU	3 (2012–2014)	3.00	0.81	3.66	0.85	1.99–6.56
<b>RIPENING PERIOD</b>						
Whole population	22 (1987–2014)	5.37	0.95	5.43	2.05	0.54–9.95
INRA	10 (2002–2014)	4.86	0.95	6.89	1.77	1.62–9.26
UNIBO	13 (1987–2014)	7.83	0.96	6.51	1.87	0.98–9.19
CRA-W	10 (1987–2008)	4.37	0.87	4.90	1.15	1.12–8.38
RBIPH	5 (2006–2010)	5.07	0.92	5.00	1.56	1.00–7.60
NFC	3 (1999–2013)	2.93	0.87	5.94	1.77	2.00–8.33
SLU	3 (2012–2014)	2.91	0.98	3.75	1.44	1.00–7.00

<sup>a</sup>A single average value was available at NFC, assessed over 10 years (different years according to the cultivars).

modeling phenotypic covariance with the kinship matrix (**K**) implemented in a modified version of the multi-locus mixed model (MLMM) proposed by Segura et al. (2012). The Extended Bayesian Information Criterion (EBIC, Chen and Chen, 2008) was used to select the model that best fitted our data. A genome-wide significance threshold was determined using a Bonferroni correction at 5%. MLMM uses a stepwise mixed-model regression with forward inclusion and backward elimination of SNPs re-estimating the variance components of the model at each step. MLMM divides the phenotypic variance into genetic variance (explained by structure, by kinship, and by SNPs included as cofactors in the model), and unexplained variance (residual variance), suggesting a natural stopping criterion (genetic variance = 0) for including cofactors, and allowing to estimate the explained and unexplained heritable variance for each trait. The causal-variant heritability tagged by all possible genotyped SNPs, was quantified for each trait at the step 0 of MLMM, i.e., when the structure, kinship and residual variances were estimated with no SNP included as cofactors in the model. The part of variance explained (PVE) by the significant SNP(s) as well as the part of variance due to population structure and due to kinship, were estimated at the optimal step of the MLMM (i.e., stopping criterion).

To establish 95% confidence intervals for the significant SNPs retained as cofactors in the whole population, we conducted a re-sampling approach as proposed by Hayes (2013). The full set of individuals with phenotypic data was randomly split into two subsets with equal size; this procedure was repeated 50 times for each trait, and then a GWAS was run on each subset as explained above. The standard error ( $se(x)$ ) of the position of an underlying association was estimated as the median absolute deviation of the positions of the SNPs retained as cofactors on

each chromosome over all subsets. Then, the 95% confidence interval was calculated as the position of the most significant SNPs retained as cofactors in the analysis of the whole population  $\pm 1.96 se(x)$ .

## Effects of the SNPs Identified as Cofactors

The SNPs identified as cofactors were analyzed toward mode and size of allelic effects. The mode of gene action at each SNP was estimated for the whole population and (when possible) for each of the three geographic groups using the ratio of dominance ( $d$ ) to additive ( $a$ ) effects calculated from the mean of the genotypic means for each genotypic class. The dominance effect was calculated as the difference between the mean observed within the heterozygous class and the mean across both homozygous classes ( $d = G_{AB} - 0.5 (G_{AA} + G_{BB})$ , where  $G_{ij}$  is the trait mean in the  $ij$ th genotypic class). To classify the mode, we used the following ranges, similar to Węgrzyn et al. (2010). No dominance was defined for small absolute values, i.e.,  $|d/a| \leq 0.50$ ; partial or complete dominance was defined as values in the range  $0.50 < |d/a| < 1.25$ ; and over- or under-dominance pertained to values of  $|d/a| > 1.25$ .

To assess the joint effect of the different allelic combinations (i.e., genetic variants) defined by the SNPs identified for each trait, mean and statistical significance among the most frequent genetic variants were calculated by the Tukey-Kramer test ( $\alpha = 0.05$ ).

Dominance and epistatic effects among the identified SNPs were tested in a model including their additive effects with correction for population structure (**Q**) and modeling phenotypic covariance with the kinship matrix (**K**). Percentages of variance explained by additive plus dominance effects, and by additive

plus dominance and epistatic effects were estimated in a hierarchical sequence, using a cumulative  $R^2$  metric.

### In Silico Candidate Gene Research

Chromosomal regions corresponding to approximate 95% confidence intervals for the position of SNPs retained as cofactors in the whole population for flowering and ripening periods were investigated for *in silico* candidate gene identification using GDDH13 genome (Daccord et al., 2017). The annotations of protein-coding and non-protein-coding genes of the regions of interest were identified using GDDH13 genome v1.1 browser (<https://iris.angers.inra.fr/gddh13/>). Annotations regarding the biochemical function of genes (mainly provided by InterproScan) were enriched by the biological functions inferred from the putative orthologs identified in *Arabidopsis thaliana*, *Solanum lycopersicum*, and *Prunus persica* genomes. Furthermore, structures of predicted genes and intergenic regions were systematically investigated to detect eventual mis/not-annotated genes and pseudogenes (stop codons and/or frameshifts in their CDS) in the regions of interest.

## RESULTS

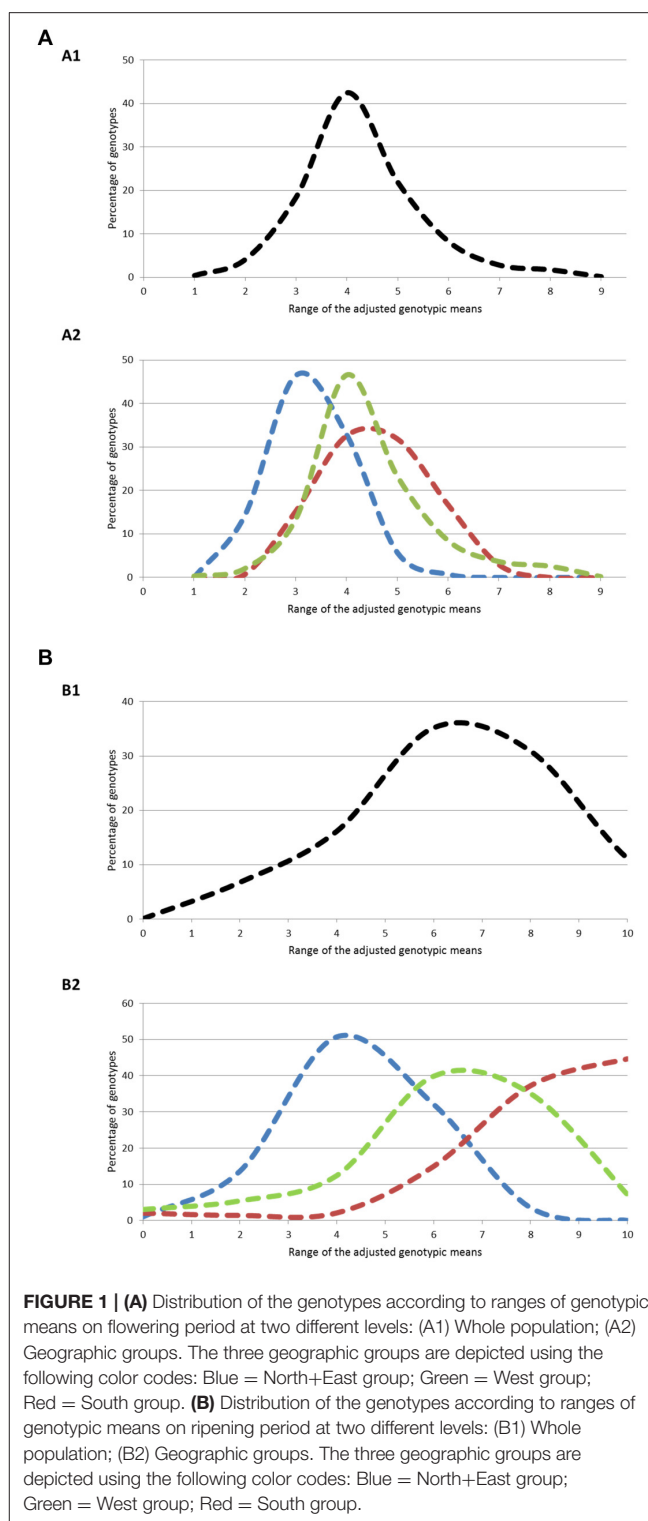
### Phenotypic Variation

Large phenotypic variation was observed at both whole population and collection level (Table 1). In the individual collections, average flowering period varied in genotypic means from 3.66 to 5.58 (on the 1–9 scale), while average ripening period varied from 3.75 to 6.89. Heritability was consistently high ( $>0.80$ ). The two traits were significantly correlated when calculated across all genotypes in the whole population ( $r = 0.44$ ;  $p\text{-value} = 2.2e^{-16}$ ), see Figure S1.

The three geographical groups differed considerably for both traits (Figure 1). For flowering period, 94% of the genotypes had genotypic means between 2 and 5 (mean value = 3.77) in the North+East group, while 96 and 83% of the genotypes varied between 3 and 7 and between 3 and 6 for the South and West groups respectively, with almost identical mean values (South: 5.03; West: 4.99). For ripening period, phenotypic variation was even higher: 83% of the genotypes in the North+East group had a genotypic mean below 5 (mean value = 3.41), while 91% in the South group had values above 5 (mean value = 7.49). The West group showed an intermediate distribution (mean value = 5.48).

### Population Structure and Linkage Disequilibrium

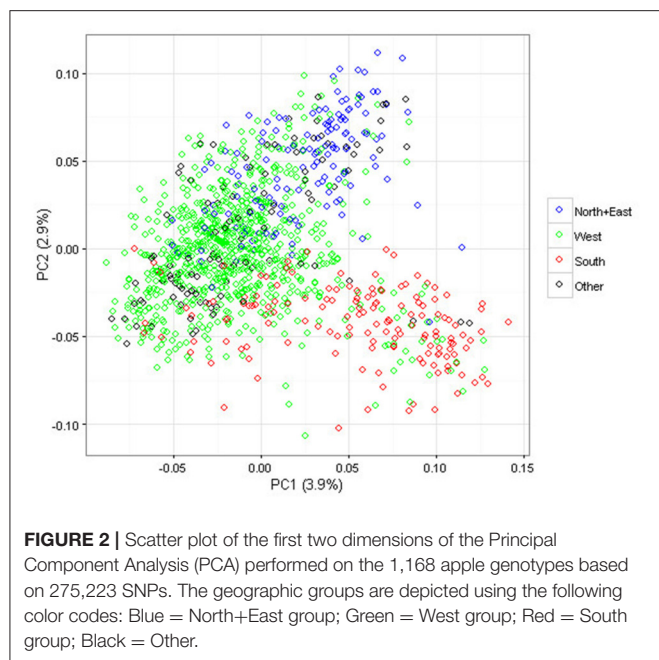
PCA was applied to summarize global genetic marker variation in the association panel: the ten largest Eigenvalues used to describe the whole population structure explained 17% of the overall variation. In Figure 2, the first two components of the PCA are represented. Genetic discrimination between the genotypes classified according to their geographic group of origin is visible in the bi-dimensional plot; genotypes from the North+East group are located in the upper part along the Y axis, while those from the West and South groups mostly occur on the left and right side along the X axis, respectively. The three groups North+East, West and South, explained 30 and 37% of the



**FIGURE 1 | (A)** Distribution of the genotypes according to ranges of genotypic means on flowering period at two different levels: (A1) Whole population; (A2) Geographic groups. The three geographic groups are depicted using the following color codes: Blue = North+East group; Green = West group; Red = South group. **(B)** Distribution of the genotypes according to ranges of genotypic means on ripening period at two different levels: (B1) Whole population; (B2) Geographic groups. The three geographic groups are depicted using the following color codes: Blue = North+East group; Green = West group; Red = South group.

variation for the first two dimensions of the PCA but less than 6.5% for the next eight dimensions.

LD ( $r^2$ ) was very variable in the SNP sets of 10 kb randomly sampled along the genome, spanning the entire range from absence to complete LD (Figure S2). The distribution of LD



was highly asymmetric, with half of the marker pairs showing a  $r^2$  value below 0.1 ( $r^2_{vs}$  value below 0.07 when corrected for relatedness and population structure). LD decay curves were very flat (**Figure 3**); mean  $r^2$  values were 0.24, 0.21, and 0.19 at 100 bp, 1 kb, and 5 kb, respectively, while mean  $r^2_{vs}$  values were 0.20, 0.17, and 0.13, respectively. Half of the adjacent marker pairs occurred within 587 bp, while 90% occurred within 4,975 bp. To estimate LD between a causal variant in the middle of a marker interval and its flanking markers, mean  $r^2$  values for marker pairs at half these distances (i.e., 293.5 and 2,487.5 bp) were computed: in the whole population, mean  $r^2$  values were 0.23 and 0.19 without correction, and 0.19 and 0.14 with correction.

## Genome-Wide Association Study (GWAS) Without Cofactor Inclusion

### Flowering Period

Using a single-locus mixed model with control for population structure and relatedness, 50 SNPs were significantly associated with flowering period for the whole population (Table S2). In a quantile-quantile (Q-Q) plot (not shown), close adherence was found between the observed and expected  $-\log_{10}(p)$  values till around 3, indicating that the significant SNPs are unlikely to be biased by population structure and relatedness. A strong association signal was found on chromosome 9 (49 SNPs) with a Bonferroni correction threshold of 5% ( $-\log_{10}(p) > 6.74$ ). The remaining SNP was located on the fictive chromosome 0 containing all unassigned scaffolds. The SNPs located on chromosome 9 spanned a distance of 3.24 Mb (265,164–3,509,888 bp).

Analyses of the individual collections revealed significant associations for flowering period only at INRA (29 SNPs), NFC (2 SNPs) and RBIPH (1 SNP) (Table S2). Twenty-one of these SNPs, all on chromosome 9, were also significant

for the whole population. Of the 11 SNPs identified only in individual collections, nine were located on chromosome 9, one on chromosome 4 (RBIPH), and one on chromosome 11 (NFC).

### Ripening Period

For ripening period, 82 SNPs exhibited a significant association for the whole population with adjustment for population structure and relatedness (Table S3). The Q-Q plot (not shown) was similar to the previous one. Most SNPs (70) were located on chromosome 3, spanning a distance of 2.05 Mb (29,196,200–31,243,065 bp). Nine SNPs were located on chromosome 16 and spanned a distance of 274.3 kb (9,032,064–9,306,332 bp), while three SNPs could not be mapped.

When GWAS was performed for each collection separately, numbers of significant SNPs were 38, 12, and 8 for NFC, INRA and SLU, respectively, two for both RBIPH and UNIBO, and only one for CRA-W (Table S3). Thirty-one of the 43 SNPs identified in the analyses of individual collections, showed a significant association also in the whole population. When analyses were carried out at collection-scale, all the identified SNPs were located on chromosome 3, except for three that were unmapped; none of the SNPs located on chromosome 16 with a significant association in the whole population, were identified in the GWAS of the individual collections.

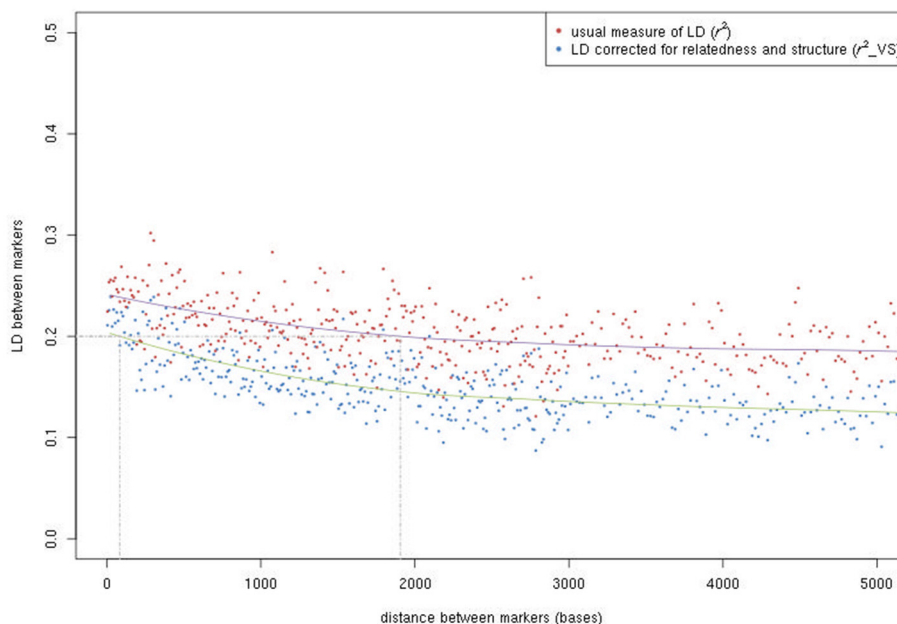
## Genome-Wide Association Study (GWAS) Using SNPs as Cofactors

To further dissect the signal from the chromosomal regions containing the sets of significant SNPs for each trait, we performed a GWAS with MLM using SNPs as cofactors (Segura et al., 2012). MLM handles the putatively confounding effect of significant SNPs elsewhere on the genome, which considerably outperforms the existing single-locus mixed models by reducing the number of significantly associated SNPs rather than the number of peaks (Sauvage et al., 2014). The part of variance explained by structure and kinship estimated at step 0 of MLM (i.e., when no SNP were included as cofactors in the model) was 0.78 for flowering period and 0.84 for ripening period (Table 2).

### Flowering Period

The optimal MLM according to the EBIC criterion for the whole population retained two SNPs for flowering (FB\_AFFY\_0496090 “SNP.9-1” and FB\_AFFY\_0495650 “SNP.9-2”), both located on chromosome 9, only 27 kb apart (Table 3). These two SNPs were significantly associated with flowering period also in the initial analysis based on a single-locus mixed model (Table S2). With this optimal model, 8.9% of the whole phenotypic variance was explained by the pair of SNPs retained, 27.3% corresponded to the underlying population structure of the association panel, and 38.6% was associated with kinship (**Figure 4A**). In the re-sampling analysis conducted for estimating an approximate 95% confidence interval, the number of cofactors retained in the model was one in 76 subsets, two in 23 subsets and three in one subset (Table S4). SNP.9-1 and SNP.9-2 were selected as cofactors in 35 and 38 subsets, respectively, while other SNPs from the same chromosome were selected as cofactors in 45 subsets, among which FB\_AFFY\_4941692





**FIGURE 3 |** LD decay according to the physical distance between SNPs. Both the usual  $r^2$  and the  $r^2$  after correcting for relatedness and population structure ( $r^2_{VS}$ ) are given.

**TABLE 2 |** Summary of trait associations at the optimal models according to the EBIC criterion.

Population	No. cultivars	Model 0 PVE by (structure + kinship)	Optimum model			
			PVE by (structure + cofactors + kinship)	No. associations without cofactors	No. significant cofactors	PVE by cofactors
FLOWERING PERIOD						
Whole population	1,126	0.78	0.75	50	2	0.09
INRA	251	0.93	0.90	29	1	0.13
UNIBO	166	0.74	0.74	0	0	0.00
CRA-W	221	0.72	0.72	0	0	0.00
RBIPH	177	0.79	0.58	1	2	0.27
NFC	288	0.78	0.77	2	4	0.33
SLU	159	0.80	0.80	0	0	0.00
RIPENING PERIOD						
Whole population	1,149	0.84	0.85	82	6	0.17
INRA	260	0.84	0.84	12	1	0.13
UNIBO	178	0.88	0.86	2	1	0.16
CRA-W	217	0.70	0.65	1	1	0.12
RBIPH	176	0.80	0.78	2	2	0.18
NFC	293	0.97	0.92	38	1	0.22
SLU	160	0.94	0.89	8	4	0.28

Part of Variance Explained (PVE) by population structure, cofactors and kinship, the number of associations without cofactors, the number of significant cofactors, the PVE by cofactors, and the ratio PVE by kinship/PVE by cofactors and kinship are showed. Data obtained for individual collections and the whole populations are provided.

(“SNP.9-5”) was selected in 25 subsets (Table S4). For this region, the length of a 95% confidence interval was estimated at 157 kb.

GWAS of each single collection retained one, two and four SNPs (Tables 2, 3) for the collections of INRA, RBIPH, and NFC,

respectively, but none for the collections of CRA-W, SLU, and UNIBO. The part of variance explained by the markers selected in the optimal models for each collection was 13% (INRA), 27% (RBIPH), and 33% (NFC) (Table 2; Figure 4A). One of the two SNPs identified in the MLM analysis of the whole population,

**TABLE 3 |** Summary of associations identified by Multi-Locus Mixed Model (MLMM) at the optimal models according to the EBIC criterion for flowering and ripening periods in the whole population and in the six individual collections.

Population	SNP code	SNP short name	No. Cofactor <sup>a</sup>	Location of the SNPs		Alleles	p-value	MAF
				Chromosome	Position			
FLOWERING PERIOD								
Whole population	FB_AFFY_0496090	SNP.9-1	1	9	530,386	G/T <sup>b</sup>	1.33E-08	0.11
Whole population	FB_AFFY_0495650	SNP.9-2	2	9	557,419	A/G	6.81E-08	0.13
INRA	FB_AFFY_0495650	SNP.9-2	1	9	557,419	A/G	1.06E-12	0.18
RBIPH	FB_AFFY_6830175	SNP.4-1	1	4	968,334	C/T	3.16E-09	0.01
RBIPH	FB_AFFY_1629518	SNP.9-3	2	9	925,476	A/G	8.01E-08	0.14
NFC	FB_AFFY_6873601	SNP.4-2	4	4	7,719,622	A/G	3.94E-07	0.24
NFC	FB_AFFY_7355751	SNP.9-4	2	9	1,938,744	C/T	3.15E-08	0.11
NFC	FB_AFFY_2782466	SNP.11-1	1	11	12,422,656	A/C	8.30E-09	0.02
NFC	FB_AFFY_9818101	SNP.12-1	3	12	14,536,815	C/T	5.35E-08	0.15
RIPENING PERIOD								
Whole population	FB_AFFY_6730867	SNP.3-3	4	3	30,430,113	A/G <sup>c</sup>	6.76E-15	0.10
Whole population	FB_AFFY_7541229	SNP.3-4	5	3	30,465,002	C/T	8.51E-10	0.09
Whole population	FB_AFFY_4981462	SNP.3-6	2	3	30,700,183	C/T	4.39E-19	0.18
Whole population	FB_AFFY_1209620	SNP.3-7	1	3	30,726,252	A/G	1.28E-13	0.41
Whole population	FB_AFFY_3795860	SNP.10-1	6	10	38,390,484	A/G	1.76E-08	0.23
Whole population	FB_AFFY_6370928	SNP.16-1	3	16	9,146,297	C/T	5.16E-12	0.14
INRA	FB_AFFY_1253936	SNP.3-5	1	3	30,590,166	A/C	3.03E-14	0.08
UNIBO	FB_AFFY_1253936	SNP.3-5	1	3	30,590,166	A/C	6.75E-09	0.06
CRA-W	FB_AFFY_4741632	SNP.3-2	1	3	30,318,639	A/G	6.71E-08	0.11
RBIPH	FB_AFFY_4981462	SNP.3-6	1	3	30,700,183	C/T	3.60E-10	0.20
RBIPH	FB_AFFY_4836781	SNP.15-1	2	15	10,377,731	C/T	3.18E-08	0.37
NFC	FB_AFFY_4981462	SNP.3-6	1	3	30,700,183	C/T	1.37E-18	0.14
SLU	FB_AFFY_0899559	SNP.3-1	4	3	24,220,838	A/G	7.21E-07	0.10
SLU	FB_AFFY_1209620	SNP.3-7	1	3	30,726,252	A/G	7.51E-15	0.33 <sup>d</sup>
SLU	FB_AFFY_6239519	SNP.13-1	3	13	1,889,560	G/T	1.41E-07	0.24
SLU	FB_AFFY_3879540	SNP.16-2	2	16	10,298,660	G/T	2.34E-07	0.27

<sup>a</sup> Order of inclusion of the SNPs at the optimal model in MLMM according to the EBIC criterion.

<sup>b</sup> The allele associated with an early flowering period is highlighted in bold. The alternative allele is thus associated with a late flowering period.

<sup>c</sup> The allele associated with an early ripening period is highlighted in bold. The alternative allele is thus associated with a late ripening period.

<sup>d</sup> The allele found in SLU with the lowest frequency was the opposite to the one that appeared in the lowest frequency in the whole population and the other five individual collections.

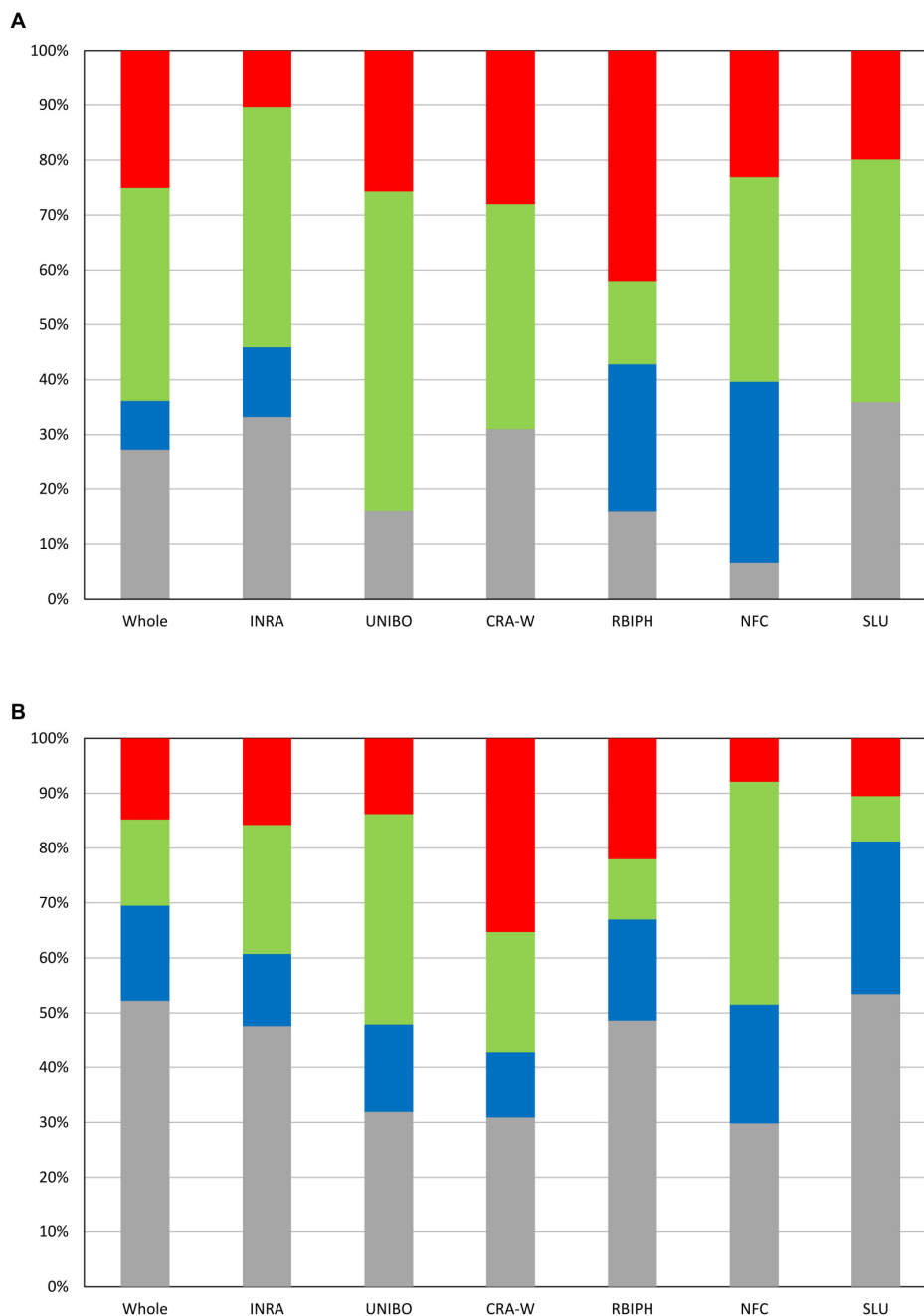
SNP.9-2, was also found for the INRA collection. Neither the two SNPs selected for the RBIPH collection (chromosomes 4 and 9) nor the four identified in the NFC collection (chromosomes 4, 9, 11, and 12) were retained in the analysis of the whole population.

### Ripening Period

The optimal model according to the EBIC criterion retained six SNPs for the whole population, four at the bottom of chromosome 3 (FB\_AFFY\_6730867, FB\_AFFY\_7541229, FB\_AFFY\_4981462, and FB\_AFFY\_1209620, denoted as “SNP.3-3,” “SNP.3-4,” “SNP.3-6,” and “SNP.3-7”), one at the bottom of chromosome 10 (FB\_AFFY\_3795860 “SNP.10-1”), and another on the top of chromosome 16 (FB\_AFFY\_6370928 “SNP.16-1”) (Table 3). The four SNPs identified on chromosome 3 were clustered two by two with a distance of only 35 and 26 kb within each cluster and a distance of about 296 kb between clusters. The SNP retained on chromosome 10 (SNP.10-1) and one of the four retained on chromosome 3 (SNP.3-3) did not show a significant

association with ripening period in the analysis based on a single-locus mixed model (Table S3). Altogether, the six SNPs explained 17.2% of the phenotypic variation, whereas population structure and kinship explained 52.2 and 15.7%, respectively (Figure 4B). When estimating the approximate 95% confidence interval with a re-sampling analysis, the number of cofactors retained varied between one and five, and between two and four in 94 of the 100 subsets (Table S5). SNP.3-6 and SNP.3-7, selected as the two first cofactors in the whole collection, were selected in 55 and 85 subsets, respectively. Other SNPs from chromosome 3 were selected in 59 subsets (Table S5). The length of a 95% confidence interval was estimated at only 152 kb for chromosome 3, but 1.39 Mb for chromosome 10, and 426 kb for chromosome 16.

The optimal model for the analysis of each individual collection retained at least one SNP per collection (Tables 2, 3): four in the SLU collection, two in the RBIPH collection, and one each in the remaining collections. Part of variance explained by the markers identified for each collection, ranged



**FIGURE 4 |** Partition of variance at the optimal models according to EBIC for the whole population and the six individual collections for flowering period **(A)** and ripening period **(B)**. Gray: part of variance explained by structure; Blue: part of variance explained by SNPs retained as cofactors; Green: part of variance explained by kinship; Red: residual variance.

from 12% (CRA-W) to 28% (SLU) (**Table 2; Figure 4B**), with an average of 18%. Two out of the six SNPs retained in the whole population were identified also in some of the individual collections, i.e., SNP.3-6 in NFC and RBIPH, and SNP.3-7 in SLU, both of them belonging to the same lower cluster previously defined on chromosome 3. For INRA and UNIBO the same single SNP (FB\_AFFY\_1253936 “SNP.3-5”) on chromosome 3 was selected by MLM and was located in between the two

previously identified SNP clusters. In brief, the analyses of the individual collections identified six SNPs additional to the six ones identified in the whole population, three on chromosome 3, and one on each of chromosomes 13, 15, and 16.

For each trait, Manhattan plots obtained with the single-locus mixed model and the multi-locus mixed model for the whole population and for the individual collections, are shown in Figures S3, S4.

## Linkage Disequilibrium among SNPs Identified as Cofactors

Pairwise LD was assessed to test the independence of SNPs identified as cofactors for each trait in the whole population as well as in each geographic group. For flowering period, low LD ( $r^2 = 0.27$ ;  $r^2_{vs} = 0.12$ ) was detected in the whole population between the two SNPs associated with the trait despite being located only 27 kb apart (Table 4A). Analysis of the results at the geographic-group level found almost complete equilibrium between these two SNPs in the North+East and South groups ( $r^2 = 4E^{-04}$  and 0.06, respectively), while LD was much higher ( $r^2 = 0.41$ ;  $r^2_{vs} = 0.22$ ) in the West group. For ripening period, variable LD values were found among the four SNPs identified as cofactors located on chromosome 3 (Table 4B). Intermediate to high  $r^2$  values were found between SNP.3-3, SNP.3-4, and SNP.3-6, whereas low values were observed between these three SNPs and SNP.3-7. In the North+East and West groups,  $r^2$

values between these four SNPs were quite similar to those found in the whole population. By contrast, very low  $r^2$  were found for the South group, except for the pair SNP.3-3/SNP.3-6 ( $r^2 = 0.43$ ).

LD interconnections between the eight SNPs retained as cofactors and other SNPs residing within their surrounding regions of 1 Mb (Figure S6) showed that SNP.16-1 exhibited the highest number of connections with other SNPs at  $r^2 > 0.70$ , i.e., 53 SNPs delineating a region of 763 kb (Table S6). None or only a few (maximum 16) SNPs in the neighborhood of the remaining seven retained SNPs were linked with them at  $r^2 > 0.70$ . Accordingly, none of the triangular LD heat maps for the above mentioned regions showed a LD spatial pattern suggesting that they are organized in blocks of moderate/high LD (Figure S5). Conversely, networks of moderate LD ( $r^2 \sim 0.5$ – $0.6$ ) were more frequently observed for the above mentioned regions (i.e., five regions with more than 40 connected SNPs).

**TABLE 4 | (A)** Pairwise LD between the two SNPs associated with flowering period in the whole population (A1) and in the three geographic groups: North+East (A2), West (A3), and South (A4). **(B)** Pairwise LD between the four SNPs associated with ripening period on chromosome 3 in the whole population (B1) and the three geographic groups: North+East (B2), West (B3), and South (B4).

A				B					
(A1) Whole population				(B1) Whole population					
SNPs as cofactors	SNP.9-1	SNP.9-2	MAF	SNPs as cofactors	SNP.3-3	SNP.3-4	SNP.3-6	SNP.3-7	MAF
SNP.9-1	1.00	0.12	0.11	SNP.3-3	1.00	0.55	0.31	0.02	0.10
SNP.9-2	0.27	1.00	0.13	SNP.3-4	0.71	1.00	0.27	0.06	0.09
				SNP.3-6	0.56	0.54	1.00	0.22	0.18
				SNP.3-7	0.11	0.06	0.22	1.00	0.41
(A2) North+East group				(B2) North+East group					
SNPs as cofactors	SNP.9-1	SNP.9-2	MAF	SNPs as cofactors	SNP.3-3	SNP.3-4	SNP.3-6	SNP.3-7 <sup>a</sup>	MAF
SNP.9-1	1.00	2.0E-03	0.09	SNP.3-3	1.00	0.79	0.36	0.09	0.28
SNP.9-2	4.2E-04	1.00	0.11	SNP.3-4	0.83	1.00	0.32	0.11	0.27
				SNP.3-6	0.48	0.45	1.00	0.33	0.45
				SNP.3-7 <sup>a</sup>	0.16	0.21	0.28	1.00	0.33 (0.67) <sup>a</sup>
(A3) West group				(B3) West group					
SNPs as cofactors	SNP.9-1	SNP.9-2	MAF	SNPs as cofactors	SNP.3-3	SNP.3-4	SNP.3-6	SNP.3-7	MAF
SNP.9-1	1.00	0.22	0.13	SNP.3-3	1.00	0.69	0.34	0.04	0.06
SNP.9-2	0.41	1.00	0.14	SNP.3-4	0.77	1.00	0.37	0.07	0.06
				SNP.3-6	0.49	0.55	1.00	0.21	0.12
				SNP.3-7	0.09	0.12	0.21	1.00	0.39
(A4) South group				(B4) South group					
SNPs as cofactors	SNP.9-1	SNP.9-2	MAF	SNPs as cofactors	SNP.3-3	SNP.3-4	SNP.3-6	SNP.3-7	MAF
SNP.9-1	1.00	0.07	0.09	SNP.3-3	1.00	0.02	0.41	0.02	0.11
SNP.9-2	0.06	1.00	0.05	SNP.3-4	0.01	1.00	0.04	4.7E-03	0.01
				SNP.3-6	0.43	0.07	1.00	0.03	0.13
				SNP.3-7	0.03	0.02	0.04	1.00	0.17

Values below the diagonal line refer to the usual  $r^2$  and above the diagonal line refer to  $r^2_{vs}$  (i.e., with correction for relatedness and population structure) MAF of the SNPs are given for the whole population and the three geographic groups.

<sup>a</sup> The allele found in the North-East group with the lowest frequency at the SNP.3-7 was the opposite to the one that appeared in the lowest frequency in the whole population and the other two geographic groups.

Pairwise LD between SNPs are highlighted using the following color scale: red ( $r^2 = 1$ ) and yellow ( $r^2 = 0$ ).



## Allele Frequencies, Effects and Genetic Variants for the SNPs Identified as Cofactors

For each of the eight SNPs identified as cofactors in the analysis of the whole population, the minor (i.e., less frequent) allele remained the same across the individual collections, the three geographic groups and the whole population, except for two cases: SLU collection and North+East group for SNP.3-7 associated with ripening period (Tables 3, 4A,B). In the two latter cases, the allele of SNP.3-7 with the lowest frequency was the alternate one to that in the other five collections and two geographic groups thus exhibiting a strong shift in the frequency of the G allele associated to early ripening period. Large differences between geographic groups were also observed for the minor allele frequencies (MAF) of the four SNPs associated with ripening period located on chromosome 3 (Table 4B), again indicating a North-South gradient.

The phenotypic effects of the two SNPs identified for flowering period indicated a strong mean difference ( $>1.9$ ) between genotypic means for genotypes homozygous for the alternative alleles (Table S7; Figure S7). Dominance effects and epistatic interaction effects were significant, despite explaining a very small part of the variance (Table 5; Table S7). For ripening period, an even higher mean difference (frequently  $>2.5$ ) was observed between the genotypic means of alternative homozygous genotypes for the four SNPs on chromosome 3 (Table S8; Figure S8), while less variation was found for SNP.10-1 and SNP.16-1. Variation in genotypic frequencies at each SNP was again very pronounced between geographic groups (Figure S8). Globally, dominance effects and epistatic interaction effects between the six SNPs were not significant, except some partial dominance occasionally observed for SNP.3-3, SNP.3-4, and SNP.10-1 (Table 5; Table S8).

The joint effect associated with the two SNPs identified for flowering period was assessed by comparing the average values for genotypes with different combinations of alleles in the whole population. Among the five genetic variants with a frequency above 1% (Table 6), variants 1 and 5 combining, respectively, the two alleles associated with early (GG/AA)

and late (TT/GG) flowering at a homozygous stage differed on average by 3.73 corresponding to  $3.24 \sigma$  (in standard-deviation units). The double heterozygous variant 3 (GT/GA) exhibited an intermediate value. For the four SNPs identified on chromosome 3 for ripening period, only 26 combinations out of the 81 potential variants were observed, 10 of which accounted for  $\sim 95\%$  of the association panel (Table 6). The genetic variants accumulating homozygous alleles associated to early ripening period (variant 10: AA/TT/TT/GG) or late ripening period (variant 1: GG/CC/CC/AA) differed by 4.63 on average corresponding to  $2.25 \sigma$ . Out of the 397 genotypes belonging to variant 1, only 4.3% belonged to the North+East group, while 69.7 and 19.7% belonged to the West and South groups, respectively, representing 12, 35.7, and 52.7% of the total genotypes from North+East, West, and South groups, respectively. By contrast, the infrequent variant 10 ( $\sim 2\%$ ) was common in the North+East group (52.2%) but more scarce (17.4%) and totally absent in the West and South groups, respectively. Multiple comparisons indicated no significant differences between variants 1 and 8, between variants 4, 5, 6, and 7, or between variants 9 and 10 (Table 6).

## Candidate Gene Identification

For flowering period, we considered the interval 451,830–635,974 bp (i.e., 184 kb) on chromosome 9, corresponding to the fusion of the 95% confidence intervals of the two SNPs selected as cofactors. In this interval, we found 28 gene models (Table S9) including putative transcription factors containing e.g., a NAM/NAC (MD09G1006400), a WRKY (MD09G1008800), a SBP (MD09G1008900) domain, and a putative glutaredoxin (MD09G1007400). In a second run, we also considered the 95% confidence interval covering SNP.9-5 which was selected in 25 subsets of the re-sampling analysis (Table S4) despite not being detected in the initial analysis. The corresponding interval 654,780–811,891 bp (i.e., 157 kb) was almost contiguous to the previous one, thus defining a wider region of  $\sim 360$  kb (451,830–811,891 bp). Thirty-eight additional gene models were found in this enlarged interval (Table S9) including a putative SRF transcription factor containing a MADS- and a K-box (MD09G1009100), another putative SRF transcription factor (not detected by automatic annotation pipeline), and a gene model containing a SWIB/MDM2 domain (MD09G1011600).

For ripening period, we considered two intervals on chromosome 3, one corresponding to the fusion of the 95% confidence intervals of SNP.3-6 and SNP.3-7 which overlapped (30,624,429–30,802,006 bp, i.e., 178 kb), and the second for the confidence interval of SNP.3-3 and SNP.3-4 (30,354,359–30,540,756 bp, i.e., 186 kb). Only eleven gene models were found in the first interval and 24 in the second, with 6 additional gene models in between (Table S10). Two successive genes encoding a putative transcription factor containing a NAM/NAC domain (MD03G1222600 and MD03G1222700) were found in the very close vicinity of SNP.3-6 and SNP.3-7, both SNPs located in between the two genes. An Ultrapetala transcription factor (MD03G1220200) was found close to SNP.3-3, and a protein tyrosine kinase

**TABLE 5 |** Test of dominance and epistatic effects among the SNPs selected as cofactors in the GWAS of the whole population for flowering and ripening periods.

Trait	Effects	d.f.	F-test	p-value	PVE (%)
Flowering period	Additive	2	79.8	4.3E-33	9.1
	Dominance	2	5.9	2.7E-03	0.7
	Epistatic	4	9.0	3.7E-07	2.0
Ripening period	Additive	6	106.6	3.4E-106	17.4
	Dominance	6	1.1	3.6E-01	0.2
	Dominance of SNP.3-4	1	4.8	2.8E-02	0.1
	Dominance of SNP.3-6	1	4.5	3.3E-02	0.1
	Epistatic	41 <sup>a</sup>	1.2	2.3E-01	1.3

<sup>a</sup>Some combinations of SNP genotypes did not exist in the whole population, which reduced the df for all interactions between 6 SNPs.

**TABLE 6 |** Joint effect of the two SNPs associated with flowering period in the whole population and of the ten most frequent genetic variants defined by the four SNPs on chromosome 3 associated with ripening period in the whole population.

Genetic variant	Genotypes at SNPs <sup>a,b</sup>	N° cultivars	Frequency	Mean	Median	SD	Min	Max	Tukey groups
<b>FLOWERING PERIOD</b>									
Variant 1	<b>GG/AA</b>	760	0.67	4.65	4.73	0.95	1.73	8.87	a
Variant 2	<b>GT/GA</b>	126	0.11	5.83	5.73	1.32	2.73	8.88	b
Variant 3	<b>GG/GA</b>	121	0.11	4.43	4.47	1.13	1.73	8.24	a
Variant 4	<b>GT/AA</b>	89	0.08	4.49	4.67	1.00	2.34	7.37	a
Variant 5	TT/GG	11	0.01	8.38	7.82	1.08	5.73	9.24	c
<b>RIPENING PERIOD</b>									
Variant 1	GG/CC/CC/AA	397	0.35	6.75	6.80	1.43	2.21	9.84	a
Variant 2	GG/CC/CC/AG	336	0.29	5.69	5.63	1.35	0.88	9.50	b
Variant 3	GG/CC/CC/GG	73	0.06	4.72	4.81	1.61	1.21	8.48	c
Variant 4	GG/CC/CT/AG	61	0.05	3.75	3.74	1.73	0.88	8.77	d
Variant 5	AG/CT/CT/AG	59	0.05	3.75	3.85	1.35	1.21	7.15	d
Variant 6	AG/CT/CT/GG	44	0.04	2.89	2.54	1.32	0.54	6.85	d
Variant 7	GG/CC/CT/GG	39	0.03	3.49	3.21	1.61	0.55	6.66	d
Variant 8	AG/CC/CT/AA	29	0.03	7.44	7.67	1.51	4.82	9.80	a
Variant 9	AG/CT/TT/GG	28	0.02	2.01	2.09	0.62	1.16	4.14	e
Variant 10	AA/TT/TT/GG	23	0.02	2.11	2.21	1.10	0.54	4.42	e

<sup>a</sup> The allele associated with an early flowering period is highlighted in bold; order of SNPs is as follows: SNP:9-1/SNP:9-2.

<sup>b</sup> The allele associated with an early ripening period is highlighted in bold; order of SNPs is as follows: SNP:3-3/SNP:3-4/SNP:3-6/SNP:3-7.

(MD03G1221300) close to SNP.3-4. On chromosome 10, we considered the 95% confidence interval 37,695,471–39,085,497 bp for SNP.10-1 and found 153 gene models (Table S11). Among them were four putative transcription factors, two of which contained a NAM/NAC domain (MD10G1288300 and MD10G1299900) while another two contained an Apetala-2 domain (MD10G1290400 and MD10G1290900). A carbohydrate phosphorylase putatively involved in starch metabolism was also identified (MD10G1289300). On chromosome 16, we considered a 95% confidence interval 8,933,453–9,359,141 bp for SNP.16-1 and found 38 gene models (Table S12). Together with two putative transcription factors encoding either a NAM/NAC domain or a TIFY domain (MD16G1125800 and MD16G1127400, respectively), we especially identified a gene model encoding an auxin responsive protein (MD16G1124300) and another gene model encoding a sugar bidirectional transporter (MD16G1125300).

A nearly perfect microsynteny (with some minor rearrangement) was revealed between apple and peach in all the four confidence interval genomic regions estimated in our study, as shown by the numerous conserved homologs between the two species in those regions (Tables S9–S12).

## DISCUSSION

### Genomic Regions Controlling Variation in Phenological Traits

The SNPs retained as cofactors in the GWAS on the whole population defined one genomic region controlling flowering period and three controlling ripening period. Additional regions were identified when conducting GWAS for individual

collections. The associations found accounted for varying levels of trait variation (0–33% for flowering period; 12–28% for ripening period) across the whole population and individual collections. We applied a conservative approach in identifying SNPs as cofactors for *p*-values below a defined threshold of Bonferroni correction at 5%. Implementation of those stringent parameters was essential to eliminate false positives, but have probably sacrificed some true associations with small effects.

The top of chromosome 9 was recently indicated as being involved in the genetic control of flowering or bud burst period (Celton et al., 2011; Allard et al., 2016). The regions pointed out in these contributions overlap with the confidence interval found in our study although the region indicated by Allard et al. (2016) is shifted slightly downstream since the very top of the chromosome was not mapped in their experiment. The regions indicated in these studies were, however, much larger than the confidence interval we report: Celton et al. (2011) examined a region of almost 16 cM corresponding to 4.04 Mb and comprising 983 gene models in the apple genome v1.0 of the Genome Database for *Rosaceae* (GDR, <https://www.rosaceae.org/>), whereas Allard et al. (2016) indicated a region of 10 cM corresponding to 1.8 Mb and comprising 622 gene models. The numerous recombination events accumulated in our association panel reduced the associated region to 184 kb with only 28 gene models in the GDDH13 genome. An extended interval of ~360 kb was nevertheless proposed to take into account the results of the re-sampling analysis, thus generating a final set of 66 candidate gene models. Interestingly, Trainin et al. (2016) identified a common haplotype on the top of chromosome 9 shared by a small subset of mostly Israeli apple cultivars adapted to low-chill conditions such as the well-known “Anna.” They

defined an interval of about 1.7 Mb but suggested that the genetic factor/s responsible for early bud-break could be located in a region of about only 190 kb (between SNP-A6-2 and SNP-A4). Mapping these SNPs on the GDDH13 genome, we found the corresponding interval to be 730,978–923,844 bp, which overlaps the extended interval accounting for SNP.9-5 (451,830–811,891 bp). This co-localization raises the question of the allelic control of flowering period in that particular genomic region as described below.

Chromosomes 3, 10, and 16 have shown associations with ripening period in previous linkage mapping studies (Liebhard et al., 2003; Kenis et al., 2008; Chagné et al., 2014; Kunihiya et al., 2014). None of these studies attempted to define a confidence interval for the physical position of the reported QTLs, thus preventing an accurate comparison of the precision in QTL location between studies. Recently, Migicovsky et al. (2016) did not find any associations with ripening period on chromosomes 10 and 16 in a GWAS based on single-locus tests, but identified associations with two SNPs on chromosome 3 located within the coding region of NAC18.1 (GenBank ID: NM\_001294055.1) which corresponds to a gene model (MD03G1222600) at position ~30,697,000 bp of GDDH13 genome. Interestingly, this position fits perfectly within the 95% confidence interval of SNP.3-6/SNP.3-7 (30,624,429–30,802,006 bp). Since this genomic region has been identified in various environments and genetic backgrounds, it therefore appears to potentially be a major factor in the genetic control of ripening period.

## GWAS on Phenological Traits Suggests Presence of Allelic Heterogeneity

For each trait, MLM analysis for the whole population retained SNPs in weak LD despite being in close vicinity. Two SNPs retained as cofactors for flowering period on chromosome 9 were only 27 kb apart. Four SNPs retained for ripening period on chromosome 3 spanned a region of 296 kb, with two sub-regions spanning only 35 and 26 kb, respectively. Identification of multiple significant SNPs within or near a single gene may suggest either allelic heterogeneity or the presence of an untyped causal variant that requires multiple SNPs to be adequately tagged, or both (Atwell et al., 2010; Dickson et al., 2010; Segura et al., 2012). Allelic heterogeneity refers to the presence of more than two functional alleles of a given gene affecting a phenotypic trait (Wood et al., 2011). Indeed, the biallelic nature of SNPs reduces their ability to tag multiple alleles and explains the need for several SNPs to tag them. Also, maximizing the genetic variance in the association panel by including geographically distant accessions with both different and complex evolutionary histories is expected to improve resolution, but has the potential to introduce genetic heterogeneity (i.e., multiple causal variants with various dates of appearance and frequencies) which can generate false “synthetic” associations when only single-locus tests are used (Korte and Farlow, 2013). Fortunately, the MLM approach is able to disentangle the contribution of genetic heterogeneity by including “competing” variants as cofactors within the mixed model setting and thus helps to discard false

“synthetic” associations (Segura et al., 2012; Korte and Farlow, 2013). For flowering period, the two detected SNPs can either fit with allelic heterogeneity or untyped causal variant requesting more than one SNP. But more interestingly, the co-localization of our confidence interval with the small genomic region identified by Trainin et al. (2016) for the extreme phenotype of low-chilling requirement, opens the question of the local genomic architecture of this trait. Since bud-break and consequently flowering period of Israeli cultivars occur much earlier than in traditional European cultivars (Trainin et al., 2016), either two different polymorphic genes or a single gene with at least three alleles may be responsible for the co-location of detectable genotypic variation for flowering period and early bud-break. In the latter case, at least two alleles would control the genotypic difference we observed here for flowering period, and another more “extreme” allele would confer the early bud-break of Israeli cultivars. Alternatively, this extreme allele could be proposed as an epi-allele when considering epigenetic control (Ríos et al., 2014). For ripening period, a model including the two nearby genomic regions detected on chromosome 3 can also be proposed with the presence of two closely positioned genes (~300 kb apart), each with possible allelic heterogeneity. Such a complex pattern of association has never been highlighted before for flowering and ripening periods in apple. Nevertheless, additional genetic studies would be required to be certain about the multi-allelic and multi-genic architecture of the detected regions by using e.g., local haplotype sharing methods (Xu and Guan, 2014) provided that sufficient SNPs are available.

## Unexplained Genetic Variation May Be Accounted for by Multiple Factors

The limited number of detected genomic regions associated with the traits and the low/moderate amount of phenotypic variance accounted for by the retained SNPs suggests that several, if not many additional genomic regions are involved in the genetic control of these traits. Here, as with other GWAS, we were challenged by the so-called “missing heritability” syndrome (i.e., traits exhibiting both high heritability and tiny effect variants; Maher, 2008; Manolio et al., 2009; Visscher et al., 2010; Yang et al., 2010; Zuk et al., 2012). In our experiment, a significant proportion of the phenotypic variance not captured by the SNP cofactors could be explained by relatedness accounting for polygenic effects (15–58% for flowering period, 8–41% for ripening period) and population structure mostly accounting for genetic differentiation over geographic groups (7–36% for flowering period, 30–53% for ripening period). The large proportion of phenotypic variance under genetic control clearly indicates that additional genomic regions are still to be discovered. Interestingly, at the whole population level, the part of variance explained by relatedness for flowering period (39%) was more than twice the estimate for ripening period (16%), while the inverse was observed for the part of variance explained by structure (27% for flowering period, 52% for ripening period), thus indicating differential impact of relatedness and geographic structure on these two phenological traits.



Several factors may have hampered the detection of additional genomic regions. Genetic architecture consisting of many common variants with small effects and/or rare variants with large effects can reduce the statistical power of GWAS (Brachi et al., 2011; Gibson, 2011; Stranger et al., 2011; Korte and Farlow, 2013). The wide diversity in our association panel may have favored the inclusion of several rare variants with strong effects that could not be detected in the present study. The rapid LD decay and the LD pattern between causal variants and genotyped SNPs are two other limiting factors (Manolio et al., 2009; Visscher et al., 2010; Stranger et al., 2011). Despite the use of a high-density SNP array, it is possible that some genomic regions with causal variants were insufficiently covered by SNPs (i.e., null or incomplete LD), thus preventing detection of the corresponding variance. Denser genotyping may be required to find new associations given that both their effect and frequency are large enough to be detected by GWAS. Also, other factors may account for the unexplained genetic variation: (i) quality and precision of the phenotypic (historical) data (Myles et al., 2009; Migicovsky et al., 2016), (ii) genotype  $\times$  environment (G $\times$ E) interactions, (iii) epistatic effects that were not systematically investigated in our experiment, or even, (iv) epigenetic variation.

## Population Structure and Geographic Adaptation

Our association panel consisted mostly of local and/or old dessert apple cultivars selected as representative subsets by each institute. The phenotypic differences observed in the geographic-scale analyses (North+East, South and West groups) are probably explained by adaptive selection to different environments. Adaptive traits are frequently filtered by environmental gradients that coincide with patterns of population structure due to the differential fixation of alleles among groups of cultivars, following diversifying selection and/or genetic drift (Atwell et al., 2010; Brachi et al., 2011; Lasky et al., 2015; Nicolas et al., 2016). Despite genetic structure being weak at the whole population scale in our study (only 17% of the genotypic variation was explained by the ten largest Eigenvalues of the PCA), this structure explained a moderate (flowering period: 27%) or even high (ripening period: 52%) proportion of the phenotypic variance in GWAS. These results are in line with the phenotypic differences observed at a geographic scale, since the first two principal components were highly associated with geographic grouping (30 and 37%). A similar observation was made by Migicovsky et al. (2016). Differential selection together with genetic drift where the germplasm originated (North+East, West and South) may have favored or selected specific alleles or combinations of alleles in different geographic regions/environments. A good example is given by SNP.3-7, which was associated to ripening period with a frequency of 67% for its G allele in the North+East group but only 17% in the South group. Similarly, when considering the genetic variants combining the four SNPs retained on chromosome 3, variant 10 combining all earliness-associated SNP alleles at a homozygous state, was very common in accessions of the North+East group while totally absent in the South. In apple, harvest period is

probably the trait with the strongest impact of geographical adaptation, since local weather conditions define the length of harvesting season.

## Putative Functions of Genes Controlling Phenotypic Variation in Apple Flowering Period

Gene models of particular interest were identified in the interval defined on chromosome 9, including a putative NAC gene (MD09G1006400). NAC-domain proteins are transcription factors involved in the genetic control of flowering time in *Arabidopsis* (Yoo et al., 2007), where two NAC proteins in association with a JM14 gene (a histone demethylase) apparently take part in flowering time regulation (Ning et al., 2015). In addition, a putative WRKY transcription factor was identified. This gene model (MD09G1008800, corresponding to MDP0000154734 in GDR) was cited by Trainin et al. (2016) as a putative candidate for early bud-break of Israeli apple cultivars. The WRKY gene family was recently proposed to play a role in dormancy regulation in peach (Chen et al., 2016). Based on RNAseq, MD09G1008800 transcription was detected mainly in apple roots and only slightly in fruits, thus limiting its potential role in flowering.

Three other candidate gene models are of special interest for the genetic control of flowering period. MD09G1009100 and another non-predicted gene model are similar to SRF transcription factors containing a MADS domain, putatively homologous to the *FLC* (*FLOWERING LOCUS C*) gene involved with the *FRIGIDA* gene in vernalization response of *Arabidopsis* (reviewed by Amasino and Michaels, 2010). MADS-box genes, such as the DAM (dormancy associated MADS-box) family members, were previously shown to be the master regulators of dormancy establishment and maintenance in *Prunus* and *Pyrus* species (Bielenberg et al., 2008; Ubi et al., 2010; Yamane et al., 2011; Saito et al., 2013; Sánchez-Pérez et al., 2014; Zhebentyayeva et al., 2014). Related DAM-like genes with dormancy-dependent expression have been identified in other perennial species such as leafy spurge (Horvath et al., 2008, 2010), raspberry (Mazzitelli et al., 2007), blackcurrant (Hedley et al., 2010), and kiwifruit (Wu et al., 2012). Also, MD09G0010600 is predicted as a SWIB/MDM2-domain containing gene, a member of a family of chromatin remodeling complexes that modify DNA accessibility by restructuring nucleosomes (Jerzmanowski, 2007). These three genes (MDP0000167381/MDP0000126259, MDP0000296123, and MDP0000315892/MDP0000317368 in GDR v1.0, respectively) were also mentioned by Trainin et al. (2016). Conversely, the other candidate genes highlighted by these authors were located outside of our largest confidence interval, as were all the candidate genes cited by Celton et al. (2011). Finally, special attention should be given to the MADS-domain containing gene (MD09G1009100 = MDP0000167381 = MDP0000126259 in its shorter version) since it was upregulated in several differential expression situations when comparing the low chilling requirement sport “Castel Gala” with “Royal Gala” (Porto et al., 2015). By contrast, the other two genes (MADS-box: MDP0000207984, and *PRE1*-like:



MDP0000320691) highlighted by Porto et al. (2015), were located either on another chromosome or considerably downstream on chromosome 9.

Whilst the candidate genes we identified did not encompass all of those that have previously been proposed to have a role in flowering time, it is clear that a number of them have putative roles in the regulation of flowering time in apple or other plants.

## Putative Functions of Genes Controlling Phenotypic Variation in Apple Ripening Period

Concerning ripening period, three main genomic regions were identified (on chromosomes 3, 10, and 16) with candidate genes belonging to the NAC family, surrounded by other genes putatively involved in apple ripening. In the genomic region of chromosome 3, two NAC transcription factors (MD03G1222600 and MD03G1222700) are strongly indicated as candidate genes for the control of this trait. A member of this gene family (i.e., ppa008301m, according to the *P. persica* genome version v1.0) was identified in a major locus on chromosome 4 controlling maturity date in peach, and a 9 bp DNA insertion in its last exon was described as a variant putatively linked to early ripening (Pirone et al., 2013). Most interestingly, one of the two NAC genes of apple (i.e., MD03G1222700) cited above showed to be the best homolog of this particular peach NAC gene which was renamed Prupe.4G186800 in the *P. persica* genome version v2.1 (Verde et al., 2017). The second apple NAC gene was also showed to be homolog to the second peach NAC gene cited by Pirone et al. (2013) (i.e., ppa007577m.v1.0 equivalent to Prupe.4G187100.v2.1), and a strong microsynteny was observed between *Malus* and *Prunus* all along the analyzed confidence interval (Table S10). The importance of NAC transcription factors in controlling fruit ripening traits has been described also in tomato (Zhu et al., 2014) and kiwifruit (Nieuwenhuizen et al., 2015). Very recently, two NAC members (called SINAC4/9) were indicated as regulators of ethylene biosynthesis and ethylene-related genes in tomato (Kou et al., 2016).

The genes identified on chromosome 10 appeared to be involved in the same metabolic pathways: two NAM/NAC (MD10G1288300 and MD10G1299900) and two Apetala2 (MD10G1290400 and MD10G1290900) transcription factors. Members of the plant-specific APETALA2/ethylene response factor (AP2/ERF) superfamily of transcription factors act downstream of the ethylene signaling pathway and are strongly conserved throughout the plant kingdom (Xie et al., 2016). They are apparently associated with several plant developmental and growth processes, including fruit ripening (Licausi et al., 2010; Karlova et al., 2014; Xie et al., 2016).

On chromosome 16, two additional putative transcription factors encoding either a NAM/NAC domain (MD16G1125800) or a TIFY domain (MD16G1127400) were identified as well as one gene encoding an auxin responsive protein (MD16G1124300) and one gene for a sugar bidirectional

transporter (MD16G1125300) with high homology with a senescence associated protein (SAG 29) of *Arabidopsis*. TIFY transcription factors comprise a plant-specific family involved in the regulation of various developmental processes and responses to phytohormones. Among the 30 members of this family characterized in apple (Li et al., 2015), are the jasmonate zim-domain (JAZ) proteins, known to be repressors of JA signaling and, consequently, actors of the cross-talk among multiple hormone signaling pathways including ethylene and gibberellins (An et al., 2016). The expression patterns of genes in the JA biosynthesis pathway was found to be correlated with genes in the ethylene biosynthesis pathway, emphasizing the role of JA biosynthesis and its signaling on apple fruit maturation (Lv et al., 2015).

Altogether, candidate genes identified after GWAS highlight the probable role of transcription factors, controlling the ethylene biosynthesis or regulatory pathway, for ripening in apple. Other candidates such as the gene encoding for an auxin responsive protein may also be considered since an ethylene–auxin interplay at a late ripening stage has been proposed in apple (Tadiello et al., 2016).

## FUTURE PERSPECTIVES

GWAS mapping is a powerful tool for the identification of genomic regions associated with important traits, but results can be restricted by too much genetic heterogeneity, insufficient marker density and an overly strong impact of population structure. In the present study, narrow genomic regions controlling two phenological traits and a rather low number of candidate genes were identified, while other regions remained unidentified because of the relationship between traits and geographic structure. Enlarging the diversity panel with more genotypes, especially from Southern and Northern+Eastern groups, might improve detection of loci associated with those traits in each geographic group. Also, a combination of linkage and association analyses may achieve higher statistical power and resolution (Jansen et al., 2003; Flint-Garcia et al., 2005; Pascual et al., 2016) and reduce the confidence interval of the detected genomic regions which would call for validating the function of certain candidate genes by genetic transformation, especially gene editing (Busov et al., 2005; Nishitani et al., 2016). Still, the current set of phenotypic and genotypic data is already useful to establish genome-wide predictions (Meuwissen et al., 2001; Muranty et al., 2015) of the breeding values of the studied genotypes for these two traits.

## AUTHOR CONTRIBUTIONS

JU, HM, and DL carried out the statistical analyses under the supervision of C-ED and ST. CD and ER managed the leaf sample collection and part of DNA extraction. AT performed the remaining DNA extraction and the whole runs of the Axiom® plates on the GeneTitan® under the supervision of CP. CD performed the whole SNP genotyping analysis and validation

process with the help of HM and C-ED. Selection of germplasm and acquiring phenotypic data were performed by C-ED, AG, RG, LF, ML, PH, MO, FP, JS, HN, RGr, LD, and ST. MB and SM brought expertise about GWAS methodology and results interpretation. SA performed the candidate gene analysis thanks to the new apple genome sequence developed by ND and J-MC. C-ED conceived and coordinated the study. FL coordinated the EU FruitBreedomics project including this study. JU, HM, and C-ED wrote the manuscript with decisive contributions of SA and LD. HN, MO, MT, LB, RV, MB, LG-G, and SM critically reviewed the manuscript. All authors read and approved the final manuscript.

## FUNDING

This work has been funded under the EU seventh Framework Programme by the FruitBreedomics project No. 265582: Integrated Approach for increasing breeding efficiency in fruit tree crops (<http://www.fruitbreedomics.com/>).

## REFERENCES

- Abbott, A. G., Zhebentyayeva, T., Barakat, A., and Liu, Z. (2015). "The genetic control of bud-break in trees," in *Advances in Botanical Research*, Vol. 74, eds P. Christophe and A. B. Anne-Françoise (San Diego, CA: Academic Press), 201–228.
- Allard, A., Legave, J. M., Martinez, S., Kelner, J. J., Bink, M. C. A. M., Di Guardo, M., et al. (2016). Detecting QTLs and putative candidate genes involved in budbreak and flowering time in an apple multiparental population. *J. Exp. Bot.* 67, 2875–2888. doi: 10.1093/jxb/erw130
- Amasino, R. M. (2005). Vernalization and flowering time. *Curr. Opin. Biotechnol.* 16, 154–158. doi: 10.1016/j.copbio.2005.02.004
- Amasino, R. M., and Michaels, S. D. (2010). The timing of flowering. *Plant Physiol.* 154, 516–520. doi: 10.1104/pp.110.161653
- An, X. H., Hao, Y. J., Li, E. M., Xu, K., and Cheng, C. G. (2016). Functional identification of apple MdJAZ2 in *Arabidopsis* with reduced JA-sensitivity and increased stress tolerance. *Plant Cell Rep.* 36, 255–265. doi: 10.1007/s00299-016-2077-9
- Aranzana, M. J., Kim, S., Zhao, K., Bakker, E., Horton, M., Jakob, K., et al. (2005). Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. *PLoS Genet.* 1:e60. doi: 10.1371/journal.pgen.0010060
- Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., et al. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465, 627–631. doi: 10.1038/nature08800
- Balding, D. J. (2006). A tutorial on statistical methods for population association studies. *Nat. Rev. Genet.* 7, 781–791. doi: 10.1038/nrg1916
- Bates, D. M., and Sarkar, D. (2007). *lme4: Linear Mixed-effects Models using Eigen and Classes*. R package version 0.99875–99876.
- Bianco, L., Cestaro, A., Linsmith, G., Muranty, H., Denancé, C., Théron, A., et al. (2016). Development and validation of the Axiom® Apple 480 K SNP genotyping array. *Plant J.* 86, 62–74. doi: 10.1111/tpj.13145
- Bianco, L., Cestaro, A., Sargent, D. J., Banchi, E., Derdak, S., Di Guardo, M., et al. (2014). Development and validation of a 20K single nucleotide polymorphism (SNP) whole genome genotyping array for apple (*Malus × domestica* Borkh.). *PLoS ONE* 9:e110377. doi: 10.1371/journal.pone.0110377
- Bielenberg, D. G., Wang, Y., Li, Z., Zhebentyayeva, T., Fan, S., Reighard, G. L., et al. (2008). Sequencing and annotation of the evergrowing locus in peach (*Prunus Persica* [L.] Batsch) reveals a cluster of six MADS-box transcription factors as

## ACKNOWLEDGMENTS

The authors thank Vincent Segura for his help in defining the process for testing dominance and epistatic effects with MLM, the Migale (<http://migale.jouy.inra.fr/>) and GenoToul (<http://bioinfo.genotoul.fr/>) bioinformatics platforms for giving access to computing facilities, the past and present curatorial staff and field technicians for maintaining the apple collections at INRA Experimental Unit (UE Horti), CRA-W, RBIPH, AUB-UNIBO, NFC-Brogdale (UK), and SLU, and acknowledge Defra (UK) for supporting the characterization of the collections. JU has been partially supported by an Early Stage Research Fellowship of the Institute of Advanced Studies (University of Bologna).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2017.01923/full#supplementary-material>

- candidate genes for regulation of terminal bud formation. *Tree Genet. Genomes* 4, 495–507. doi: 10.1007/s11295-007-0126-9
- Boss, P. K., Bastow, R. M., Mylne, J. S., and Dean, C. (2004). Multiple pathways in the decision to flower: enabling, promoting, and resetting. *Plant Cell* 16, S18–S31. doi: 10.1105/tpc.015958
- Brachi, B., Morris, G. P., and Borevitz, J. O. (2011). Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biol.* 12:232. doi: 10.1186/gb-2011-12-10-232
- Busov, V. B., Brunner, A. M., Meilan, R., Filichkin, S., Ganio, L., Gandhi, S., et al. (2005). Genetic transformation: a powerful tool for dissection of adaptive traits in trees. *New Phytol.* 167, 9–74. doi: 10.1111/j.1469-8137.2005.01412.x
- Campoy, J. A., Ruiz, D., and Egea, J. (2011). Dormancy in temperate fruit trees in a global warming context: a review. *Sci. Hortic.* 130, 357–372. doi: 10.1016/j.scienta.2011.07.011
- Cannell, M. G. R., and Smith, R. I. (1986). Climate warming, spring budburst and frost damage on trees. *J. Appl. Ecol.* 23, 177–191. doi: 10.2307/2403090
- Castède, S., Campoy, J. A., Quero-García, J., Le Dantec, L., Lafargue, M., Barreneche, T., et al. (2014). Genetic determinism of phenological traits highly affected by climate change in *Prunus avium*: flowering date dissected into chilling and heat requirements. *New Phytol.* 202, 703–715. doi: 10.1111/nph.12658
- Celton, J. M., Martinez, S., Jammes, M. J., Bechti, A., Salvi, S., Legave, J. M., et al. (2011). Deciphering the genetic determinism of bud phenology in apple progenies: a new insight into chilling and heat requirement effects on flowering dates and positional candidate genes. *New Phytol.* 192, 378–392. doi: 10.1111/j.1469-8137.2011.03823.x
- Chagné, D., Crowhurst, R. N., Troggio, M., Davey, M. W., Gilmore, B., Lawley, C., et al. (2012). Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. *PLoS ONE* 7:e31745. doi: 10.1371/journal.pone.0031745
- Chagné, D., Dayatilake, D., Diack, R., Oliver, M., Ireland, H., Watson, A., et al. (2014). Genetic and environmental control of fruit maturation, dry matter and firmness in apple (*Malus × domestica* Borkh.). *Hortic. Res.* 1:14046. doi: 10.1038/hortres.2014.46
- Chen, J., and Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model space. *Biometrika* 95, 759–771. doi: 10.1093/biomet/asn034
- Chen, M., Tan, Q., Sun, M., Li, D., Fu, X., Chen, X., et al. (2016). Genome-wide identification of WRKY family genes in peach and analysis of WRKY expression during bud dormancy. *Mol. Genet. Genomics* 291, 1319–1332. doi: 10.1007/s00438-016-1171-6

- Cook, B. I., Wolkovich, E. M., and Parmesan, C. (2012). Divergent responses to spring and winter warming drive community level flowering trends. *Proc. Natl. Acad. Sci. U.S.A.* 109, 9000–9005. doi: 10.1073/pnas.1118364109
- Daccord, N., Celton, J. M., Linsmith, G., Becker, C., Choisine, N., Schijlen, E., et al. (2017). High-quality *de novo* assembly of the apple genome and methylome dynamics of early fruit development. *Nat. Genet.* 49, 1099–1106. doi: 10.1038/ng.3886
- Dickson, S. P., Wang, K., Krantz, I., Hakonarson, H., and Goldstein, D. B. (2010). Rare variants create synthetic genome-wide associations. *PLoS Biol.* 8:e1000294. doi: 10.1371/journal.pbio.1000294
- Di Guardo, M., Bink, M. C. A. M., Guerra, W., Letschka, T., Lozano, L., Busatto, N., et al. (2017). Deciphering the genetic control of fruit texture in apple by multiple family-based analysis and genome-wide association. *J. Exp. Bot.* 68, 1451–1466. doi: 10.1093/jxb/erx017
- Dirlwanger, E., Quero-Garcia, J., Le Dantec, L., Lambert, P., Ruiz, D., Dondini, L., et al. (2012). Comparison of the genetic determinism of two key phenological traits, flowering and maturity dates, in three *Prunus* species: peach, apricot and sweet cherry. *Heredity* 109, 280–292. doi: 10.1038/hdy.2012.38
- Ensminger, I., Schmidt, L., and Lloyd, J. (2008). Soil temperature and intermittent frost modulate the rate of recovery of photosynthesis in Scots pine under simulated spring conditions. *New Phytol.* 177, 428–442. doi: 10.1111/j.1469-8137.2007.02273.x
- Erez, A. (2000). “Bud dormancy: phenomenon, problems and solutions in the tropics and subtropics,” in *Temperate Fruit Crops in Warm Climates*, ed A. Erez (Dordrecht: Kluwer Academic Publishers), 17–48. doi: 10.1007/978-94-017-3215-4\_2
- Farneti, B., Di Guardo, M., Khomenko, I., Cappellin, L., Biasioli, F., Velasco, R., et al. (2017). Genome-wide association study unravels the genetic control of the apple volatilome and its interplay with fruit texture. *J. Exp. Bot.* 68, 1467–1478. doi: 10.1093/jxb/erx018
- Fleckinger, J. (1964). “Phénologie et arboriculture fruitière,” in *Le bon jardinier*, (Tome I, 2ème partie), eds P. Grisvard and V. C. Chaudun (La Maison Rustique), 362–372.
- Flint-Garcia, S. A., Thuillet, A., Yu, J., Pressoir, G., Romero, S. M., Mitchell, S. E., et al. (2005). Maize association population: a high-resolution platform for quantitative trait locus dissection. *Plant J.* 44, 1054–1064. doi: 10.1111/j.1365-313X.2005.02591.x
- Fox, J. (2003). Effect displays in R for generalised linear models. *J. Stat. Soft.* 8, 1–27. doi: 10.18637/jss.v008.i15
- Gardner, K. M., Brown, P., Cooke, T. F., Cann, S., Costa, F., Bustamante, C., et al. (2014). Fast and cost-effective genetic mapping in apple using next-generation sequencing. *G3* 4, 1681–1687. doi: 10.1534/g3.114.011023
- Gibson, G. (2011). Rare and common variants: twenty arguments. *Nat. Rev. Genet.* 13, 135–145. doi: 10.1038/nrg3118
- Hänninen, H., and Tanino, K. (2011). Tree seasonality in a warming climate. *Trends Plant Sci.* 16, 412–416. doi: 10.1016/j.tplants.2011.05.001
- Hayes, B. (2013). Overview of statistical methods for Genome-Wide Association Studies (GWAS). *Methods Mol. Biol.* 1019, 149–169. doi: 10.1007/978-1-62703-447-0\_6
- Hedley, P. E., Russell, J. R., Jorgensen, L., Gordon, S., Morris, J. A., Hackett, C. A., et al. (2010). Candidate genes associated with bud dormancy release in blackcurrant (*Ribes nigrum* L.). *BMC Plant Biol.* 10:202. doi: 10.1186/1471-2229-10-202
- Horvath, D. P., Chao, W. S., Suttle, J. C., Thimmapuram, J., and Anderson, J. V. (2008). Transcriptome analysis identifies novel responses and potential regulatory genes involved in seasonal dormancy transitions of leafy spurge (*Euphorbia esula* L.). *BMC Genomics* 9:536. doi: 10.1186/1471-2164-9-536
- Horvath, D. P., Sung, S., Kim, D., Chao, W., and Anderson, J. (2010). Characterization, expression and function of DORMANCY ASSOCIATED MADS-BOX genes from leafy spurge. *Plant Mol. Biol.* 73, 169–179. doi: 10.1007/s11103-009-9596-5
- Ingvarsson, P. K., and Street, N. R. (2011). Association genetics of complex traits in plants. *New Phytol.* 189, 909–922. doi: 10.1111/j.1469-8137.2010.03593.x
- Ionescu, I. A., Moller, B. L., and Sánchez-Pérez, R. (2017). Chemical control of flowering time. *J. Exp. Bot.* 68, 369–382. doi: 10.1093/jxb/erw427
- Jansen, R. C., Jannink, J. L., and Beavis, W. D. (2003). Mapping quantitative trait loci in plant breeding populations: use of parental haplotype sharing. *Crop Sci.* 43, 829–834. doi: 10.2135/cropsci2003.8290
- Jerzmanowski, A. (2007). SWI/SNF chromatin remodeling and linker histones in plants. *Biochim. Biophys. Acta* 1769, 330–345. doi: 10.1016/j.bbaexp.2006.12.003
- Johnston, J. W., Gunaseelan, K., Pidakala, P., Wang, M., and Schaffer, R. J. (2009). Co-ordination of early and late ripening events in apples is regulated through differential sensitivities to ethylene. *J. Exp. Bot.* 60, 2689–2699. doi: 10.1093/jxb/erp122
- Jung, C., Pillen, K., Staiger, D., Coupland, G., and von Korff, M. (2017). Editorial: recent advances in flowering time control. *Front. Plant Sci.* 7:2011. doi: 10.3389/fpls.2016.02011
- Karlova, R., Chapman, N., David, K., Angenent, G. C., Seymour, G. B., and de Maagd, R. A. (2014). Transcriptional control of fleshy fruit development and ripening. *J. Exp. Bot.* 65, 4527–4541. doi: 10.1093/jxb/eru316
- Kenis, K., Keulemans, J., and Davey, M. W. (2008). Identification and stability of QTLs for fruit quality traits in apple. *Tree Genet. Genomes* 4, 647–661. doi: 10.1007/s11295-008-0140-6
- Korte, A., and Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* 9:29. doi: 10.1186/1746-4811-9-29
- Kou, X., Liu, C., Han, L., Wang, S., and Xue, Z. (2016). NAC transcription factors play an important role in ethylene biosynthesis, reception and signaling of tomato fruit ripening. *Mol. Genet. Genomics* 291, 1205–1217. doi: 10.1007/s00438-016-1177-0
- Kunihisa, M., Moriya, S., Abe, K., Okada, K., Haji, T., Hayashi, T., et al. (2014). Identification of QTLs for fruit quality traits in Japanese apples: QTLs for early ripening are tightly related to preharvest fruit drop. *Breed. Sci.* 64, 240–251. doi: 10.1270/jsbbs.64.240
- Lasky, J. R., Upadhyaya, H. D., Ramu, P., Deshpande, S., Hash, C. T., Bonnette, J., et al. (2015). Genome-environment associations in sorghum landraces predict adaptive traits. *Sci. Adv.* 1, 1–13. doi: 10.1126/sciadv.1400218
- Laurens, F., Aranzana, M. J., Arús, P., Bassi, D., Bonany, J., Corelli, L., et al. (2012). Review of fruit genetics and breeding programmes and a new European initiative to increase fruit breeding efficiency. *Acta Hortic.* 929, 95–102. doi: 10.17660/ActaHortic.2012.929.12
- Leforestier, D., Ravon, E., Muranty, H., Cornille, A., Lemaire, C., Giraud, T., et al. (2015). Genomic basis of the differences between cider and dessert apple varieties. *Evol. Appl.* 8, 650–661. doi: 10.1111/eva.12270
- Lê, S., Josse, J., and Husson, F. (2008). FactoMineR: an R package for multivariate analysis. *J. Stat. Soft.* 25, 1–18. doi: 10.18637/jss.v025.i01
- Li, X., Yin, X., Wang, H., Li, J., Guo, C., Gao, H., et al. (2015). Genome-wide identification and analysis of the apple (*Malus × domestica* Borkh.) TIFY gene family. *Tree Genet. Genomes* 11, 1–13. doi: 10.1007/s11295-014-0808-z
- Licausi, F., Giorgi, F. M., Zenoni, S., Ost, F., Pezzotti, M., and Perata, P. (2010). Genomic and transcriptomic analysis of the AP2/ERF superfamily in *Vitis vinifera*. *BMC Genomics* 11:719. doi: 10.1186/1471-2164-11-719
- Liebhart, R., Kellerhals, M., Pfammatter, W., Jertmini, M., and Gessler, C. (2003). Mapping quantitative physiological traits in apple (*Malus × domestica* Borkh.). *Plant Mol. Biol.* 52, 511–526. doi: 10.1023/A:1024886500979
- Lv, J., Rao, J., Johnson, F., Shin, S., and Zhu, Y. (2015). Genome-wide identification of jasmonate biosynthetic genes and characterization of their expression profiles during apple (*Malus × domestica*) fruit maturation. *Plant Growth Reg.* 75, 355–364. doi: 10.1007/s10725-014-9958-0
- Maher, B. (2008). Personal genomes: the case of the missing heritability. *Nature* 456, 18–21. doi: 10.1038/456018a
- Mangin, B., Siberchicot, A., Nicolas, S., Doligez, A., This, P., and Cierco-Ayrolles, C. (2012). Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Heredity* 108, 285–291. doi: 10.1038/hdy.2011.73
- Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorf, L. A., Hunter, D. J., et al. (2009). Finding the missing heritability of complex diseases. *Nature* 461, 747–753. doi: 10.1038/nature08494
- Mariette, S., Tai, F. W. J., Roch, G., Barre, A., Chague, A., Decroocq, S., et al. (2016). Genome-wide association links candidate genes to resistance to *Plum pox virus* in apricot (*Prunus armeniaca*). *New Phytol.* 209, 773–784. doi: 10.1111/nph.13627
- Mazzitelli, L., Hancock, R. D., Haupt, S., Walker, P. G., Pont, S. D., McNicol, J., et al. (2007). Co-ordinated gene expression during phases of dormancy release in raspberry (*Rubus idaeus* L.) buds. *J. Exp. Bot.* 58, 1035–1045. doi: 10.1093/jxb/erl266



- Meuwissen, T., Hayes, B., and Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.
- Micheletti, D., Dettori, M. T., Micali, S., Aramini, V., Pacheco, I., Linge, C. D. S., et al. (2015). Whole-genome analysis of diversity and SNP-major gene association in peach germplasm. *PLoS ONE* 10:e0136803. doi: 10.1371/journal.pone.0136803
- Migicovsky, Z., Gardner, K. M., Sawler, J., Money, D., Bloom, J. S., Zhong, G. Y., et al. (2016). Genome to phenotype mapping in apple using historical data. *Plant Genome* 9, 1–15. doi: 10.3835/plantgenome2015.11.0113
- Muranty, H., Troggio, M., Sadok, I. B., Rifai, M. A., Auwerkerken, A., Banchi, E., et al. (2015). Accuracy and responses of genomic selection on key traits in apple breeding. *Hortic. Res.* 2:15060. doi: 10.1038/hortres.2015.60
- Myles, S., Peiffer, J., Brown, P. J., Ersoz, E. S., Zhang, Z., Costich, D. E., et al. (2009). Association mapping: critical considerations shift from genotyping to experimental design. *Plant Cell* 21, 2194–2202. doi: 10.1105/tpc.109.068437
- Neale, D. B., and Savolainen, O. (2004). Association genetics of complex traits in conifers. *Trends Plant Sci.* 9, 325–330. doi: 10.1016/j.tplants.2004.05.006
- Nicolas, S. D., Péros, J. P., Lacombe, T., Launay, A., Le Paslier, M. C., Bérard, A., et al. (2016). Genetic diversity, linkage disequilibrium and power of a large grapevine (*Vitis vinifera* L.) diversity panel newly designed for association studies. *BMC Plant Biol.* 16:74. doi: 10.1186/s12870-016-0754-z
- Nieuwenhuizen, N. J., Chen, X., Wang, M. Y., Matich, A. J., Perez, R. L., Allan, A. C., et al. (2015). Natural variation in monoterpene synthesis in kiwifruit: transcriptional regulation of terpene synthases by NAC and ETHYLENE-SENSITIVE3-like transcription factors. *Plant Physiol.* 167, 1243–1258. doi: 10.1104/pp.114.254367
- Ning, Y. Q., Ma, Z. Y., Huang, H. W., Mo, H., Zhao, T. T., Li, L., et al. (2015). Two novel NAC transcription factors regulate gene expression and flowering time by associating with the histone demethylase JM14. *Nucleic Acids Res.* 43, 1469–1484. doi: 10.1093/nar/gku1382
- Nishitani, C., Hirai, N., Komori, S., Wada, M., Okada, K., Osakabe, K., et al. (2016). Efficient genome editing in apple using a CRISPR/Cas9 system. *Sci. Rep.* 6:13481. doi: 10.1038/srep31481
- Pascual, L., Albert, E., Sauvage, C., Duangjit, J., Bouchet, J. P., Bitton, F., et al. (2016). Dissecting quantitative trait variation in the resequencing era: complementarity of bi-parental, multi-parental and association panels. *Plant Sci.* 242, 120–130. doi: 10.1016/j.plantsci.2015.06.017
- Pirone, R., Eduardo, I., Pacheco, I., Da Silva, L. C., Miculan, M., Verde, I., et al. (2013). Fine mapping and identification of a candidate gene for a major locus controlling maturity date in peach. *BMC Plant Biol.* 13:166. doi: 10.1186/1471-2229-13-166
- Porto, D. D., Bruneau, M., Perini, P., Anzanello, R., Renou, J. P., Pessoa dos Santos, H., et al. (2015). Transcription profiling of the chilling requirement for bud break in apples: a putative role for *FLC*-like genes. *J. Exp. Bot.* 66, 2659–2672. doi: 10.1093/jxb/erv061
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- R Core Team, A. (2014). *R: A Language and Environment for Statistical Computing*. [Internet]. Vienna: R Foundation for Statistical Computing. Available online at: <http://www.Rproject.org/>
- Ríos, G., Leida, C., Conejero, A., and Badenes, M. L. (2014). Epigenetic regulation of bud dormancy events in perennial plants. *Front. Plant Sci.* 5:247. doi: 10.3389/fpls.2014.00247
- Saito, T., Bai, S., Ito, A., Sakamoto, D., Saito, T., Ubi, B. E., et al. (2013). Expression and genomic structure of the dormancy-associated MADS box genes MADS13 in Japanese pears (*Pyrus pyrifolia* Nakai) that differ in their chilling requirement for endodormancy release. *Tree Physiol.* 33, 654–667. doi: 10.1093/treephys/tpt037
- Sánchez-Pérez, R., Del Cuetto, J., Dicenta, F., and Martínez-Gómez, P. (2014). Recent advancements to study flowering time in almond and other *Prunus* species. *Front. Plant Sci.* 5:334. doi: 10.3389/fpls.2014.00334
- Sauvage, C., Segura, V., Bauchet, G., Stevens, R., Do, P. T., Nikoloski, Z., et al. (2014). Genome-Wide Association in tomato reveals 44 candidate loci for fruit metabolic traits. *Plant Physiol.* 165, 1120–1132. doi: 10.1104/pp.114.241521
- Segura, V., Vilhjálmsson, B. J., Platt, A., Korte, A., Seren, Ü., Long, Q., et al. (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat. Genet.* 44, 825–830. doi: 10.1038/ng.2314
- Stranger, B. E., Stahl, E. A., and Raj, T. (2011). Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics* 187, 367–383. doi: 10.1534/genetics.110.120907
- Tadiello, A., Longhi, S., Moretto, M., Ferrarini, A., Tononi, P., Farneti, B., et al. (2016). Interference with ethylene perception at receptor level sheds light on auxin and transcriptional circuits associated with the climacteric ripening of apple fruit (*Malus x domestica* Borkh.). *Plant J.* 88, 963–975. doi: 10.1111/tpj.13306
- Trainin, T., Zohar, M., Shimoni-Shor, E., Doron-Faigenboim, A., Bar-Ya'akov, I., Hatib, K., et al. (2016). A unique haplotype found in apple accessions exhibiting early bud-break could serve as a marker for breeding apples with low chilling requirements. *Mol. Breed.* 36:158. doi: 10.1007/s11032-016-0575-7
- Ubi, B. E., Sakamoto, D., Ban, Y., Shimada, T., Ito, A., Nakajima, I., et al. (2010). Molecular cloning of dormancy associated MADS-box gene homologs and their characterization during seasonal endodormancy transitional phases of Japanese pear. *J. Am. Soc. Hort. Sci.* 135, 174–182.
- Urrestarazu, J., Denancé, C., Ravon, E., Guyader, A., Guisnel, R., Feugey, L., et al. (2016). Analysis of the genetic diversity and structure across a wide range of germplasm reveals prominent gene flow in apple at the European level. *BMC Plant Biol.* 16:130. doi: 10.1186/s12870-016-0818-0
- Verde, I., Bassil, N., Scalabrin, S., Gilmore, B., Lawley, C. T., Gasic, K., et al. (2012). Development and evaluation of a 9K SNP array for peach by internationally coordinated SNP detection and validation in breeding germplasm. *PLoS ONE* 7:e35668. doi: 10.1371/journal.pone.0035668
- Verde, I., Jenkins, J., Dondini, L., Micali, S., Pagliarini, G., Vendramin, E., et al. (2017). The Peach v2.0 release: high-resolution linkage mapping and deep resequencing improve chromosome-scale assembly and contiguity. *BMC Genomics* 18:225. doi: 10.1186/s12864-017-3606-9
- Visscher, P. M., Yang, J., and Goddard, M. E. (2010). A commentary on ‘common SNPs explain a large proportion of the heritability for human height’ by Yang et al. (2010). *Twin Res. Hum. Genet.* 13, 517–524. doi: 10.1375/twin.13.6.517
- Vitassey, Y., Lenz, A., and Körner, C. (2014). The interaction between freezing tolerance and phenology in temperate deciduous trees. *Front. Plant Sci.* 5:541. doi: 10.3389/fpls.2014.00541
- Wegrzyn, J. L., Eckert, A. J., Choi, M., Lee, J. M., Stanton, B. J., Sykes, R., et al. (2010). Association genetics of traits controlling lignin and cellulose biosynthesis in black cottonwood (*Populus trichocarpa*, *Salicaceae*) secondary xylem. *New Phytol.* 188, 15–32. doi: 10.1111/j.1469-8137.2010.03415.x
- Wilczek, A. M., Roe, J. L., Knapp, M. C., Cooper, M. D., Lopez-Gallego, C., Martin, L. J., et al. (2009). Effects of genetic perturbation on seasonal life history plasticity. *Science* 323, 930–934. doi: 10.1126/science.1165826
- Wood, A. R., Hernandez, D. G., Nalls, M. A., Yaghootkar, H., Gibbs, J. R., Harries, L. W., et al. (2011). Allelic heterogeneity and more detailed analyses of known loci explain additional phenotypic variation and reveal complex patterns of association. *Hum. Mol. Genet.* 20, 4082–4092. doi: 10.1093/hmg/ddr328
- Wu, R. M., Walton, E. F., Richardson, A. C., Wood, M., Hellens, R. P., and Varkonyi-Gasic, E. (2012). Conservation and divergence of four kiwifruit SVP-like MADSbox genes suggest distinct roles in kiwifruit bud dormancy and flowering. *J. Exp. Bot.* 63, 797–807. doi: 10.1093/jxb/err304
- Xie, X. I., Yin, X. R., and Chen, K. S. (2016). Roles of APETALA2/Ethylene-Response factors in regulation of fruit quality. *Crit. Rev. Plant Sci.* 35, 120–130. doi: 10.1080/07352689.2016.1213119
- Xu, H., and Guan, Y. (2014). Detecting local haplotype sharing and haplotype association. *Genetics* 197, 823–838. doi: 10.1534/genetics.114.164814
- Yamane, H., Ooka, T., Jotsu, H., Hosaka, Y., Sasaki, R., and Tao, R. (2011). Expressional regulation of PpDAM5 and PpDAM6, peach (*Prunus persica*) dormancy-associated MADS-box genes, by low temperature and dormancy breaking reagent treatment. *J. Exp. Bot.* 62, 3481–3488. doi: 10.1093/jxb/err028
- Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42, 565–571. doi: 10.1038/ng.608
- Yano, K., Yamamoto, E., Aya, K., Takeuchi, H., Lo, P. C., Hu, L., et al. (2016). Genome-wide association study using whole-genome sequencing rapidly



- identifies new genes influencing agronomic traits in rice. *Nat. Genet.* 48, 927–934. doi: 10.1038/ng.3596
- Yoo, S. Y., Kim, Y., Kim, S. Y., Lee, J. S., and Ahn, J. H. (2007). Control of flowering time and cold response by a NAC-domain protein in *Arabidopsis*. *PLoS ONE* 2:e642. doi: 10.1371/journal.pone.0000642
- Zhebentyayeva, T. N., Fan, S., Chandra, A., Bielenberg, D. G., Reighard, G. L., Okie, W. R., et al. (2014). Dissection of chilling requirement and bloom date QTLs in peach using a whole genome sequencing of sibling trees from an F2 mapping population. *Tree Genet. Genomes* 10, 35–51. doi: 10.1007/s11295-013-0660-6
- Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* 44, 821–824. doi: 10.1038/ng.2310
- Zhu, C. S., Gore, M., Buckler, E. S., and Yu, J. M. (2008). Status and prospects of association mapping in plants. *Plant Genome* 2, 121–133. doi: 10.3835/plantgenome2008.02.0089
- Zhu, M., Chen, G., Zhou, S., Tu, Y., Wang, Y., Dong, T., et al. (2014). A new tomato NAC (NAM/ATAF1/2/CUC2) transcription factor, SINAC4, functions as a positive regulator of fruit ripening and carotenoid accumulation. *Plant Cell Physiol.* 55, 119–135. doi: 10.1093/pcp/pct162
- Zuk, O., Hechter, E., Sunyaev, S. R., and Lander, E. S. (2012). The mystery of missing heritability: genetic interactions create phantom heritability. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1193–1198. doi: 10.1073/pnas.1119675109

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer AMA and handling editor declared their shared affiliation.

Copyright © 2017 Urrestarazu, Muranty, Denancé, Leforestier, Ravon, Guyader, Guisnel, Feugey, Aubourg, Celton, Daccord, Dondini, Gregori, Lateur, Houben, Ordidge, Paprstein, Sedlak, Nybom, Garkava-Gustavsson, Troglio, Bianco, Velasco, Poncet, Théron, Moriya, Bink, Laurens, Tartarini and Durel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.