

# Fruit Quality Traits Have Played Critical Roles in Domestication of the Apple

M. Awais Khan, Kenneth M. Olsen, Valpuri Sovero, Mosbah M. Kushad, and Schuyler S. Korban\*

## Abstract

With its long history of cultivation, the domesticated apple, *Malus × domestica* Borkh., is an excellent model for studying domestication in long-lived perennial plants. As apples have been transported from their center of origin in the Tian Shan region in central Asia and moved along the famous Silk Road trading path, they have undergone selection for distinct phenotypic traits. In this study, a high-throughput single nucleotide polymorphism (SNP) genotyping assay is used to investigate relationships of *Malus* species, draw inferences on the domestication history, identify traits critical for domestication, and assess potential genetic loci under selection. A total of 160 *Malus* accessions, including wild species, old and new apple cultivars, and advanced selections, were genotyped. Of 1536 SNPs from GoldenGate oligonucleotide pool assays (OPAs), 901 SNPs fulfilled filtering criteria. A principal component analysis (PCA) revealed that most *M. × domestica* accessions grouped together. A total of 67 loci, including 13 genomic SNPs and 54 genic SNPs, have been identified as potential targets for selection during evolution of the apple genome from its wild progenitor *M. sieversii*. Genes introgressed from local wild species into *M. × domestica* are associated with adaptation to local environments, while genes for fruit quality traits are derived from *M. sieversii*.

**S**INCE ANCIENT TIMES, plants have been undergoing selection for desirable traits through the process of domestication. During domestication, wild species of plants and animals undergo intensive artificial selection based on desirable morphological characters that eventually fixes beneficial alleles derived from wild species or from new mutations (Paran and Van der Knaap, 2007), collectively referred to as the “domestication syndrome” (Hammer, 1984). Over many centuries, selection for traits of interest has gone on both intentionally, with targeted desired traits, and unintentionally, by propagating favorable phenotypes of each generation (Gepts, 2004). Domestication of species provides not only insights into cultures and lifestyles of human civilizations, but also serves as a well-designed natural experiment to gain new knowledge about evolution, and allows us to pursue studies on the influence of selection on modifying functions of genes controlling desirable traits (Purugganan and Fuller, 2009; Olsen and Wendel, 2013). To date, several studies have focused on the genetics and genomics of the

M.A. Khan, and S.S. Korban, Dep. of Natural Resources & Environmental Sciences, and V. Sovero and M. M. Kushad, Dep. of Crop Sciences, Univ. of Illinois, Urbana, IL 61801; K.M. Olsen, Dep. of Biology, Washington Univ., St. Louis, MO 63130-4899; S.S. Korban, Dep. of Biology, Univ. of Massachusetts-Boston, Boston 02125. M.A. Khan, current address, International Potato Center (CIP), La Molina, Lima, Peru. Received 25 Apr. 2014.

\*Corresponding author (korkban@illinois.edu).

Published in The Plant Genome 7  
doi: 10.3835/plantgenome2014.04.0018  
© Crop Science Society of America  
5585 Guilford Rd., Madison, WI 53711 USA  
An open-access publication

All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permission for printing and for reprinting the material contained herein has been obtained by the publisher.

**Abbreviations:** EST, expressed sequence tag;  $F_{st}$ , subpopulation relative to the total variance; GWAS, Genome-Wide Association Study in Genome Association and Prediction Integrated Tool, an R package; IBS, identity-by-state matrix; K, number of subpopulations; LD, linkage disequilibrium; LG, linkage group; NADP, nicotinamide adenine dinucleotide phosphate; NCBI, National Center for Biotechnology Information; OPA, oligonucleotide pool assay; PC, principal component; PCA, principal component analysis; QTL, quantitative trait locus; SNP, single nucleotide polymorphism.

evolution of animals (e.g., dogs (Pollinger et al., 2010), horses (Outram et al., 2009), and cattle (Götherström et al., 2005), and plants, such as maize (*Zea mays* L.; Wang et al., 1999; Clark et al., 2004; Hufford et al., 2012) and rice (*Oryza sativa* L.; Konishi et al., 2006; Fuller et al., 2009; Huang et al., 2012). Evaluation and analysis of genome-wide genetic diversity and linkage disequilibrium (LD) in crops and their wild progenitors can be used in “selective sweep” mapping and in other population genetic studies to identify genomic regions targeted for selection during domestication (Tian et al., 2009). Two of the best-known examples of extended LD are the maize *tbt1* gene that controls plant architecture (Wang et al., 1999; Clark et al., 2004), and the *Y1* gene that controls kernel color (Palaisa et al., 2004).

Unlike self-compatible annual crop species, wherein lifecycles are short and several generations can fix favorable genetic combinations over a relatively short time, many fruit trees have long lifecycles, and are self-incompatible (Hummer and Janick, 2009). Therefore, it will be very useful to investigate whether or not selection during domestication of fruit trees has undergone a different mechanism to rapidly shape their genomes in relatively few generations during cultivation (Miller and Gross, 2011). Thus far, there are few studies on domestication of fruit crops. In one recent study (Myles et al., 2011), archeological findings have confirmed that wine grape (*Vitis vinifera* subsp. *vinifera*), a vegetatively-propagated vine crop, originated in the Near East. However, following domestication and as grape cultivation moved westward towards Europe, wild *V. vinifera* subsp. *sylvestris* from Western Europe contributed to *V. vinifera* subsp. *vinifera* cultivars, resulting in significant introgression into modern Western European cultivars.

The cultivated apple, *M. × domestica* Borkh., can serve as an excellent model to study domestication in long-lived perennial fruit trees. The *Malus* genus includes about 25 to 33 species. Progenitors of *M. × domestica* have been proposed to have arisen ~8 to 12 million years ago in forests close to the borders of Kazakhstan and China (Kellerhals, 2009). Incidence of high genetic diversity in both wild and cultivated apples of the Tian Shan region in central Asia, availability of chloroplast and nuclear DNA phylogenies, as well as common fruit and tree morphologies strongly suggest that the domesticated apple is more similar to *M. sieversii* (abundant in Tian Shan forests) rather than to other wild apples of Western Europe and North America (Juniper and Mabblerley, 2006). Neolithic agricultural techniques were pioneered in the Fergana Valley in Uzbekistan, initiating early efforts of human selection and propagation of desirable traits, and sweet apples, similar to extant apples, first appeared there. Subsequently, apples have been transported from their center of origin, along the Silk Road trading path, to Europe and then to the Americas while undergoing further selection for distinct phenotypic traits based on new environmental conditions and different organoleptic preferences. During this path of

domestication, secondary introgression into the genome of domesticated apples must have occurred from additional wild *Malus* species (Harris et al., 2002).

Large fruit size and low-fruit acidity (sweetness) have likely played important roles in selection during domestication of the apple (Juniper and Mabblerley, 2006). Among *Malus* species and hybrids, there are large differences in fruit size, from ~4 cm or less in diameter for the wild progenitor *M. sieversii* (Harris et al., 2002) up to 10 cm for the modern cultivar Spokane Beauty, but the molecular basis is not conclusively known. Differences in fruit size among *Malus* species have been attributed to differences in cell number and/or cell size (Harada et al., 2005). It has also been reported that ‘Golden Delicious’, a large-sized apple, differs from small-sized apple cultivars both in early cell production (proliferation) as well as in the duration of cell proliferation, suggesting that both factors influence final cell number and fruit size of apple (Malladi and Hirst, 2010).

Although it is widely accepted that the origin of the apple is from *M. sieversii* in the Tian Shan region and its domestication occurred along the commercial Silk Road path, the extent of contribution, as well as the number of other *Malus* species and traits that might have been critical in apple domestication remain in question. In this study, a high-throughput SNP genotyping assay (Khan et al., 2012a) is used to investigate relationships of wild *Malus* species with the cultivated apple, *M. × domestica*, draw inferences on the domestication history of apples, identify traits critical for domestication, and assess potential genetic loci under selection.

## Materials and Methods

### Plant Material

In this study, 160 accessions from 30 worldwide *Malus* species, including accessions from wild species, old and new cultivars, as well as advanced selections from a germplasm collection located at the University of Illinois at Urbana-Champaign, were used (Table 1, Supplementary Table S1). Out of 30 *Malus* species and a single hybrid group (advanced selections from breeding programs and cultivars originating from crosses among different *Malus* species), 20 species had only one accession and five species, including *M. × domestica*, *M. orientalis*, *M. prunifolia*, *M. sieversii*, and *M. sylvestris* had 52, 11, 5, 21, and 5 accessions, respectively. The hybrid group consisted of 31 accessions, and each of *M. baccata* and *M. micromalus* had four accessions. Species with fewer than four accessions were excluded from some analysis, such as analysis of relationships of species, unless stated otherwise.

All accessions have been propagated (by grafting) onto Bud-9 apple rootstocks, and were 6 yr in age. These were grown and maintained in an experimental orchard (in a completely randomized block with at least four replications per accession) at the University of Illinois at Urbana-Champaign. The origin of this germplasm collection has been described previously by Potts et al. (2012). Briefly, these accessions are a subset of the entire

**Table 1. Accessions from 30 *Malus* species, including accessions from wild *Malus* species, *M. × domestica*, and advanced selections from the germplasm collection at University of Illinois, Urbana-Champaign, used in the study.**

Species	No.	Species	No.	Species	No.	Species	No.
<i>M. angustifolia</i>	1	<i>M. halliana</i>	1	<i>M. prattii</i>	1	<i>M. sublobata</i>	1
<i>M. arnoldiana</i>	1	<i>M. hartwigii</i>	2	<i>M. prunifolia</i>	5	<i>M. sylvestris</i>	5
<i>M. asiatica</i>	2	<i>M. honanensis</i>	1	<i>M. robusta</i>	2	<i>M. toringo</i>	2
<i>M. baccata</i>	4	<i>Malus</i> hybrid spp.	31	<i>M. rockii</i>	1	<i>M. transitoria</i>	1
<i>M. bhutanica</i>	1	<i>M. kansuensis</i>	1	<i>M. sieversii</i>	21	<i>M. yunnanensis</i>	1
<i>M. coronaria</i>	1	<i>M. kirghisorum</i>	1	<i>M. sikkimensis</i>	1	<i>M. zhaojiaoensis</i>	1
<i>M. dawsoniana</i>	1	<i>M. microcarpa</i>	4	<i>M. soulardii</i>	1	<i>M. zumi</i>	1
<i>M. × domestica</i>	52	<i>M. orientalis</i>	11	<i>M. spectabilis</i>	1	Total	160

*Malus* germplasm maintained at the clonal repository of the Plant Genetic Resources Unit in Geneva, NY. This collection has been selected by the US Apple Crop Germplasm Committee (CGC) for its wide diversity for various horticultural characters, such as fruit size, fruit ripening, cold hardiness, and resistance to diseases and insects, among others, and botanical traits, such as leaf shape and leaf color (Korban and Tartarini, 2009).

### DNA Extraction and Genotyping

Young leaves were collected from trees of each of 160 *Malus* accessions (Table 1), growing in the orchard, and were quickly transported to the lab on ice. Leaves were freeze-dried in liquid N, and crushed into fine powder for DNA extraction.

Genomic DNA extraction was done using the Cetyltrimethyl Ammonium Bromide (CTAB) extraction method, but with slight modifications to remove polysaccharides, polyphenols, and RNA, followed by a phenol-chloroform extraction method as previously described (Khan et al., 2012a). DNA was quantified using NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE), and normalized to a total of 250 ng per sample. DNA from each sample was genotyped using the Illumina 1536 GoldenGate assay on a BeadStation system (Illumina Inc., San Diego, CA) at the W.M. Keck Center for Functional Genomics (University of Illinois at Urbana-Champaign), as previously described (Khan et al., 2012a). The GoldenGate OPAs used for genotyping consisted of 1536 SNPs, consisting of 1411 genic SNPs along with an additional 125 genomic SNPs (Khan et al., 2012a).

Scoring of SNPs after removal of outliers, and scaling of raw hybridization intensity data of SNPs were performed using the genotyping function in the BeadStudio package (Illumina, San Diego, CA) as recommended by Illumina and as previously described (Khan et al., 2012a). If needed, normalized intensity values were manually inspected and corrected, and SNPs with a GenCall (GC) score  $\geq 0.35$ , based on an average GC scores for genotypes, and showing errors in segregation were removed. Clean data were used for further analysis.

### Population Genetic Structure

PLINK (Purcell et al., 2007) was used to check for missing values at the level of individual SNPs as well as at the

accession level, and for minor allele frequency for genotyping. Accessions with  $\geq 15\%$  missing data and SNPs with  $\geq 10\%$  missing data were excluded, while SNPs with allele frequencies  $\geq 0.05$  were kept for further analysis. Filtered SNP data were used for assessing population groupings and relationships among *Malus* accessions.

A PCA was performed on genotypic data from 901 SNPs of 160 *Malus* accessions from 30 species and one hybrid group (includes advanced selections and cultivars resulting from crosses of two species) in R. Relationships among accessions (Table 1) were inferred based on PCA1 and PCA2. All data were graphically displayed with XLSTAT. Population structure was also assessed for these same accessions using 901 SNPs in STRUCTURE (Pritchard et al., 2000) using the admixture model. The parameters used in STRUCTURE were 1,000,000 burnin period followed by 10,000,000 MCMC runs, along with five replications each for K (number of sub-populations) values of 1 to 15.

A nearest neighbor approach was also employed to detect outliers, as implemented in PLINK (Purcell et al., 2007). This consisted of closely related *Malus* accessions from the first nearest neighbor to the fifth nearest neighbor, and these were subjected to network analysis and graphical display in Pajek (Batagelj and Mrvar, 2004). A pairwise identity-by-state (IBS) similarity matrix was also calculated using PLINK (Purcell et al., 2007). This was also used to obtain a distance matrix (1 – IBS) which was then used in SplitsTree 4 (Huson and Bryant, 2006) to calculate and display the NeighborNet network.

### Linkage Disequilibrium Decay

The chromosomal locations of SNPs were determined using BLAST (Basic Local Alignment Search Tool) searches against the apple draft genome sequence (Velasco et al., 2010) posted on the Genome Database for Rosaceae (<http://www.rosaceae.org/>, verified 24 July 2014). The physical positions of 882 SNPs on all 17 linkage groups (LGs) of the apple was used to calculate genome-wide LD patterns in all 160 *Malus* accessions from 31 species and *M. × domestica* accessions using PLINK (Purcell et al., 2007). Pairwise  $r^2$  values were visually displayed by plotting physical distance against  $r^2$  values using XLSTAT. The LD decay threshold was inferred by drawing a trend line based on nonlinear logarithmic regression curve of  $r^2$  as a function of physical distance.

## Loci under Selection

Two approaches were used to scan SNPs for signatures of selection: a subpopulation relative to the total variance ( $F_{st}$ )-based outlier detection method implemented in LOSITAN (Antao et al., 2008) and a Bayesian genome scan approach as implemented in BayeScan (Foll and Gaggiotti, 2008). Loci showing signatures of selection detected by each method were compared, and overlapping loci were identified and considered candidate SNPs under selection. Allele frequencies for the candidate SNPs were calculated across the different species, and the identities of genes containing SNPs were assessed through blastn BLAST searches against the NCBI (National Center for Biotechnology Information) database for *M. × domestica* expressed sequence tags (ESTs) and unigenes. Corresponding apple EST sequences with highest similarities were recorded and BLAST searched using blastx against nonredundant protein sequences.

## Trait Evaluation

Accessions were evaluated for various phenotypic traits over 2 yr (2010 and 2011) as follows:

**Leaf traits.** Three mature (fully-expanded) leaves from each genotype were collected, and scanned using a scanner (Epson GT-15000). Scanned images were imported into LAMINA software (Bylesjö et al., 2008) to quantify several leaf parameters including leaf area, leaf perimeter, leaf indentation, and leaf circularity. LAMINA was calibrated using 30 × 30 mm bar, and default settings were used for the analysis as manual threshold value (1 to 255) = 100, minimum object size (% of total image size) = 0.05, minimum object density (% of total image size) = 10, serration detection pixel threshold = 10, and boundary coordinates = 50.

**Flowering.** Twice a week, and for a period of 4 wk during the months of April to May in 2010 and 2011, data were recorded on flowering dates for each accession. Date was recorded for each genotype when most of the flowers of an accession were in full bloom.

**Fruit quality traits.** A minimum of three mature fruits were collected from each of 160 *Malus* accessions. Fruits were kept in cold storage and evaluated within 2 wk of collection for 10 fruit quality traits as described below. For many wild accessions, fruit size was small, and 20 to 50 fruits per genotype were collected instead.

**Fruit dimension parameters.** Measurements of fruit circumference, length, and diameter at midpoint were recorded in centimeters using calipers, while fresh weight was recorded in grams of individual fruits. Subsequently, data collected from each tree of an accession were averaged over all trees to obtain a single measurement per genotype.

**Peel (skin) and flesh coloration.** Fruit skin pigmentation was measured using a Minolta CR-200 (Minolta Co. Ltd., Japan) reflectance colorimeter for each of three fruits per accession from three different locations along the surface of a fruit. The Minolta CR-200 records colors in CIE  $L^*a^*b^*$  color space coordinates, wherein  $L^*$  corresponds to lightness, and  $a^*$  and  $b^*$  correspond to

chromaticity indices. For each genotype, the average of measurements per genotype was used to calculate chroma ( $C^*$ ) and hue angle (H) as follows:  $C^* = (a^{*2} + b^{*2})^{1/2}$  and  $H = \tan^{-1}(b^*/a^*)$  (Lipka et al., 2012). Fruits were then cut in half, and fruit flesh color was also measured using the colorimeter at two different places on the flesh;  $C^*$  and H were calculated as described above.

**Sugar content and acidity:** Peeled fruit flesh from each genotype was crushed, and extracted juice was used to determine sugar content and acidity. For wild *Malus* species with small fruit sizes, at least 20 fruits were pooled and crushed to obtain sufficient juice for acidity and sugar content analyses. A drop of apple juice from each accession was placed on a hand-held refractometer (Leica Inc., Buffalo, NY). Brix values were recorded as total soluble solids and used as equivalent to the percentage of sugar. Total titratable acid was measured by titrating 1.0 mL of apple juice of each accession to an endpoint pH of 8.2 using 0.01 M sodium hydroxide. The concentration of malic acid (MA) was calculated according to the Organization for Economic Cooperation and Development (OECD) guidelines as follows:

$$\%MA = \text{titer} \times AMF \times 100/10 \text{ (mL juice);}$$

an acid multiplication factor (AMF) of 0.0067 was used.

## Distribution and Correlation of Phenotypic Traits across *Malus* Species

Best linear unbiased predictions (BLUPs) were extracted for four traits related to fruit dimension parameters, four leaf attributes, skin and flesh color, sugar and malic acid content, and flowering date using the linear mixed model in R-package lme4. Pairwise Pearson correlation coefficients between traits were calculated using the data for all accessions, and PCA was performed based on all traits, for all species together and for different combinations of species, using R packages gplots and FactoMineR.

## Genome-Wide Association Study

Phenotypic data for 14 traits and genotypic data for 901 SNPs collected from 158 accessions were used for Genome-Wide Association Study (GWAS). This was conducted in GAPIT (Genome Association and Prediction Integrated Tool), an R package available at <http://www.maizegenetics.net/gapit> (verified 25 July 2014). GAPIT uses the Efficient Mixed Model Association (EMMA) method (Lipka et al., 2012). Sequences containing SNPs with significant associations were BLAST searched against the NCBI database for *M. × domestica* ESTs using blastn to retrieve full length sequences of ESTs. Then, retrieved EST sequences were searched against the unigene dataset. Sequences of apple ESTs with highest similarities to sequences of significant SNPs were determined through BLAST searches using blastx against nonredundant protein sequences to confirm sequence identities.

## Results

### Relationship of the Domesticated Apple with Wild *Malus* Species

A total of 160 *Malus* accessions, part of a core collection of 200 accessions maintained at five different sites around the United States, including Urbana, IL, and Geneva, NY, previously collected from different parts of the world by the U.S. National Plant Germplasm System, including wild species, old and new apple cultivars, and advanced selections, were genotyped (Table 1). Of 1536 SNPs from the GoldenGate OPAs (Khan et al., 2012a), 901 SNPs fulfilled the filtering criteria of  $\leq 15\%$  missing values for accessions,  $\leq 10\%$  missing values for SNPs, and with allele frequencies of  $\geq 0.05$ . A PCA based on these 901 SNPs revealed that most *M. × domestica* (domesticated apple) accessions grouped together (Fig. 1A). Accessions of hybrid genotypes also showed high similarities with *M. × domestica*, as most hybrids were crosses between *M. × domestica* and wild species or with progenies of such crosses. Overall, most accessions of domesticated apples and hybrids clustered together. However, a few accessions of *M. sieversii*, two accessions of *M. orientalis*, one accession of *M. prunifolia*, and two accessions, PI 588757 and PI 589421, identified as *M. × hartwigii* Koehne (*M. baccata* × *M. halliana*) and *Malus* spp. (catch-all record for unidentified species) by the USDA GRIN database, from the “others” group of accessions, also showed stronger similarities with *M. × domestica*. Furthermore, *M. sieversii* and the majority of *M. orientalis* accessions were grouped together, while four out of five accessions of *M. sylvestris* used in this study clustered together. Similarly, most *M. micromalus* accessions were also grouped together.

Bayesian analysis of genetic structure using STRUCTURE software (Pritchard et al., 2000) has indicated that there are five clusters ( $K = 5$ ) in the population used in the study (Fig. 1B). Delta- $K$  values (Evanno et al., 2005) and log-likelihood ( $\ln p[D]$ ) values estimated for all 160 *Malus* accessions as well as for 102 accessions of seven *Malus* species confirmed that  $K = 5$  is the optimal number of clusters. Of seven *Malus* species, only *M. sylvestris* is represented by a single cluster, while four are in two clusters and another two show admixtures from all five clusters. Although the orange color component is clearly attributed to the genome of *M. sylvestris*, three accessions of *M. sylvestris* show that some portion of their genomes belong to other clusters, mainly of green, red, and blue color components (Fig. 1). The clusters represented by red and green color components can be equally attributed to genomes of either *M. micromalus* and *M. baccata* or *M. sieversii* and *M. orientalis*, respectively. Thus, these latter genomes cannot be readily assigned to a single cluster. Moreover, a single accession of *M. micromalus* is an admixture of the genome represented by green color, while a single accession of *M. baccata* has a completely different genome composition represented by green and blue components of the genome. Two additional accessions of *M. baccata* have predominantly red components

along with some contributions from orange and green clusters. For genomes of two *M. orientalis* accessions, it is revealed that they predominantly belong to the blue cluster. The genome of *M. prunifolia* shows an admixture of green, blue, and red components, wherein the proportions of blue and green components are larger than that of the red component. As to be expected, the genome of *M. × domestica* has contributions from genomes represented by all five clusters. However, neither blue nor yellow components of the *M. × domestica* genome are substantially present in other species.

Relationship assessments were conducted for 102 *Malus* accessions based on identity-by-state (IBS) distance matrix (1 – IBS) using 901 SNPs, and these results were similar to those obtained using PCA and Bayesian-based clustering. Most accessions of *M. × domestica* were closely related, and some accessions showed strong genetic similarities with multiple accessions (Fig. 1C). For example, apple cultivars Rosemary Russet, Smith Jonathan, Jonathan, and Golden Delicious exhibited strong genetic relationships with multiple accessions or cultivars in the collection used in this study. Interestingly, five accessions of *M. sylvestris* did not have direct relationships with *M. × domestica* accessions; whereas, *M. sieversii*, *M. orientalis*, *M. prunifolia*, *M. baccata*, and *M. micromalus* all exhibited multiple network connections to cultivated apples (Fig. 1C).

The NeighbourNet network has further clarified relationships between accessions and their species (Fig. 1D). Similar to above findings, most accessions of *M. × domestica* have grouped together, and *M. sieversii* and *M. orientalis* have also predominantly grouped together. Furthermore, ‘Nagano’ (*M. prunifolia*) along with two accessions (99TU-20-01 [PI 4539] and RUS 98 05-05 [PI 633822]) of *M. orientalis* have strong genetic similarities with accessions of the domesticated apple, similar to findings obtained by PCA (Fig. 1A). The accession ‘Nagano’ may have been mislabeled, while the two *M. orientalis* accessions are open-pollinated, according to the description in GRIN, suggesting an uncertain pedigree. Four *M. sieversii* accessions, including KAZ 96 07-07 (PI 613958), KAZ 95 08-06 (PI 613976), KAZ 96 07-06 (PI 613994), and KAZ 96 08-16 (PI 613998) show genetic similarities to two groups of old apple cultivars, represented by ‘Petrel’ and ‘Irish Peach’, and are grouped similar to our earlier observations (Fig. 1A). Another important finding of the NeighbourNet network is the observed genetic similarities of domesticated apples to both *M. sieversii* and *M. orientalis*, as well as to other wild *Malus* species including *M. sylvestris*. Among popular domesticated cultivars close to *M. sieversii* and *M. orientalis* are Cox’s Orange Pippin, Anna, Empire, Fuji, Nova Easygro, and McIntosh. Other popular domesticated apple cultivars that have genetic similarities with other wild *Malus* species (*M. sylvestris*, *M. baccata*, *M. prunifolia*, and *M. micromalus*) include Irish Peach, Granny Smith, Gala, Golden Delicious, and Honey Crisp. Three apple cultivars, Crimson Beauty (PI 589024),

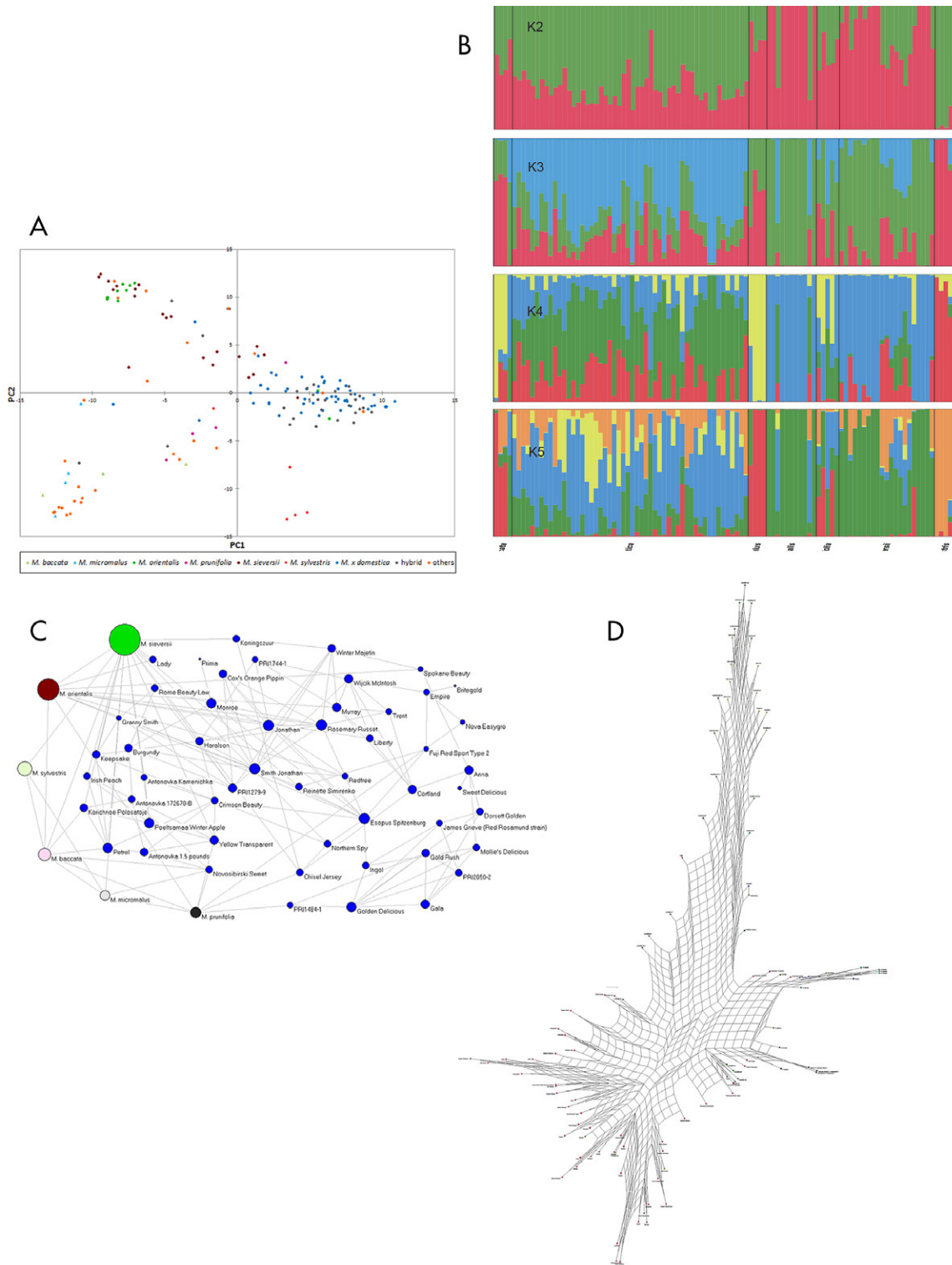


Figure 1. (A) Principal component (PC) analysis based on genotypic data from 901 single nucleotide polymorphisms (SNPs) showing population genetic structure in 160 *Malus* accessions from 30 species and one hybrid group (advanced selections from breeding programs and cultivars originating from crosses of different *Malus* species). Relationship of accessions is shown based on PC1 and PC2. Accessions from seven species, as well as the hybrid group, are colored. Note: Group named “others” includes 58 accessions from rest of the species (Table 1). (B) Population genetic structure ( $K = 2$  to 5) in seven *Malus* species (*M. x domestica*, *M. orientalis*, *M. sylvestris*, *M. sieversii*, *M. micromalus*, *M. prunifolia*, and *M. baccata*). Genetic structure was estimated using a Markov chain Monte Carlo (MCMC) algorithm implemented in STRUCTURE 2.2 using a total of 901 SNPs. (C) Network relationships for 102 *Malus* accessions from seven species inferred from an identity-by-state distance matrix. Species other than *M. x domestica* are collapsed into one circle each to emphasize the relationship of domesticated cultivars with different wild species. The relationship is based on nearest neighbor approach implemented in PLINK to detect outliers using 901 SNPs in 102 *Malus* accessions from the first nearest neighbor to the fifth nearest neighbor. (D) Network built with Neighbor-Net option in SplitsTree 4 using 901 SNPs based on identity-by-state (IBS) distance matrix (1 - IBS).

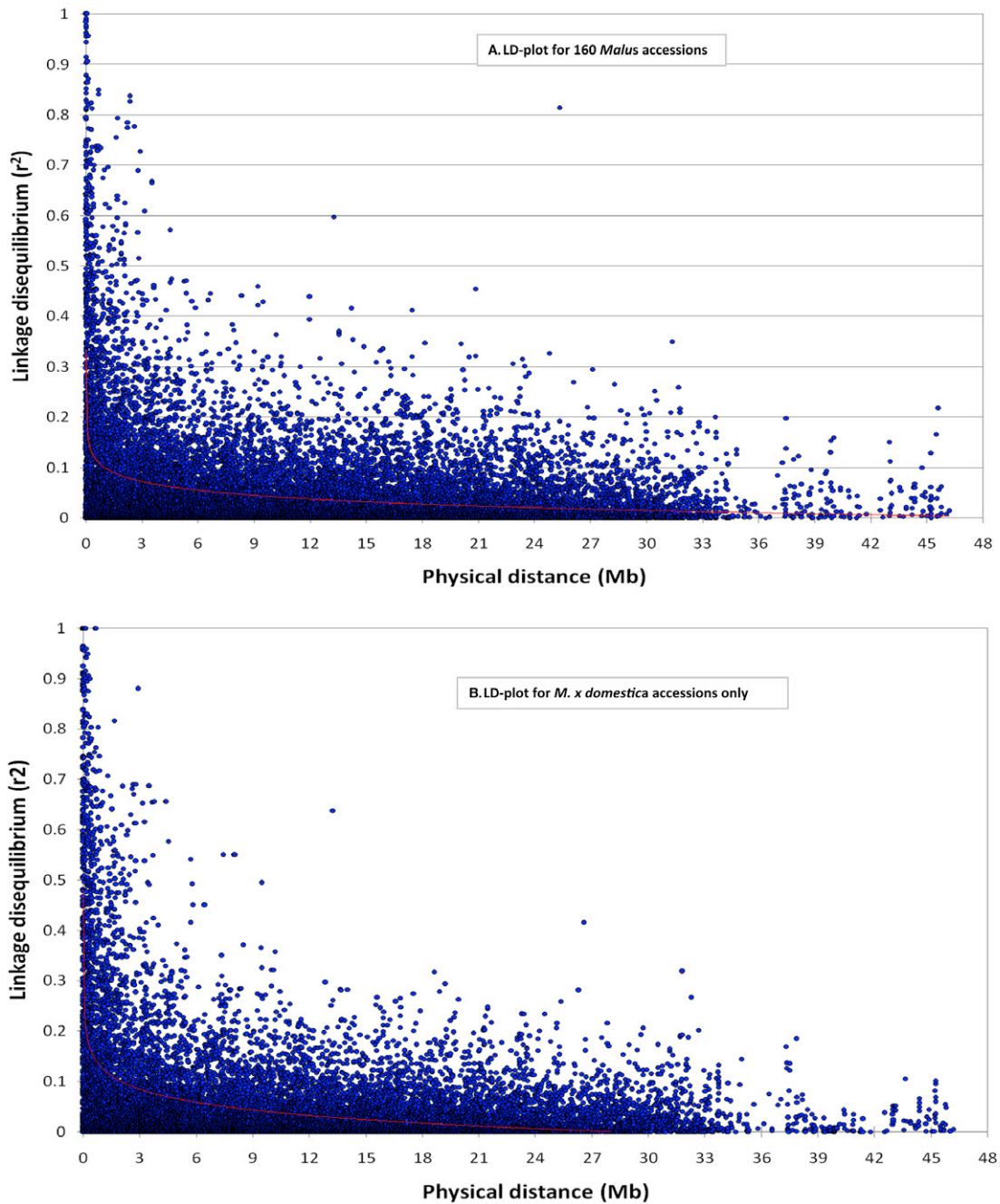


Figure 2. Genome-wide linkage disequilibrium (LD) decay patterns for (A) 160 *Malus* accessions and (B) *M. x domestica* accessions only. The LD decay was estimated using 882 single nucleotide polymorphisms (SNPs) along 17 chromosomes in PLINK and plotted in XLSTAT. Plots illustrate physical distances between pairs of SNPs with their corresponding  $r^2$  values; the red line indicates the trend line based on a nonlinear logarithmic regression curve of  $r^2$  on physical distance.

Antonovka 172670-B (PI 589956), and Antonovka 1.5 Pounds, show strong similarities to the *M. sieversii* and *M. orientalis* group as well as to *M. sylvestris*.

### Linkage Disequilibrium Decay

A BLAST search against the apple draft genome sequence (Velasco et al., 2010) allowed for assigning physical positions of 882 SNPs out of 901 SNPs. Based on the trend line, an  $r^2$ -value of 0.1 was determined to be the threshold for significant LD decay. Pairwise LD patterns indicated that for the population used in this study, all 160 *Malus*

accessions, LD decayed rapidly to an average  $r^2$ -value of 0.2 or less within 2.8 Mb of physical distance for most SNPs (Fig. 2A); while LD decay for *M. x domestica* accessions was more rapid, 0.2  $r^2$  for about 2.5 Mb (Fig. 2B).

### Supporting Evidence for Selection

The  $F_{st}$ -based outlier detection method implemented in LOSITAN (Antao et al., 2008) identified 51 SNPs as potentially showing signatures of selection. A total of 16 candidates were identified following the Bayesian genome scan in BayeScan (Foll and Gaggiotti, 2008), and

among these, 14 were shared with LOSITAN candidates. Over both detection methods, a total of 67 detected candidate SNPs have been identified. Of these, 13 were genomic SNPs, identified using the genome sequence of apple (Velasco et al., 2010), 54 were genic SNPs (Khan et al., 2012a), using EST sequences, and a single genic SNP (MdSNPui04662) whose chromosomal position could not be established (Table 2). The remaining loci were distributed across all 17 *Malus* chromosomes. The chromosome with the highest number of loci under selection

was Chromosome 7, with nine candidate loci, and this was followed by Chromosomes 12, 5, and 17 with 8, 7, and 6 loci, respectively (Table 2).

Allele frequencies differ significantly among species for SNPs detected as potential candidates for selection. BLAST searches of sequences containing candidate SNPs against the NCBI EST database for *M. × domestica* using blastn identified those corresponding *Malus* ESTs. This was followed by blastx searches of EST sequences against nonredundant protein sequences, and this provided a

**Table 2. Inferred functions and gene ontology classifications of loci showing signatures of selection between wild and domesticated apples. Loci under selection were identified using Fst-based outlier detection method implemented in LOSITAN and a Bayesian method for genome scan implemented in BayeScan.**

Single nucleotide polymorphism	Chromosome	Position (bp)	Protein ID	Unigene/blastx†	Id (%)	Len (aa)	<i>M. sieversii</i>	<i>M. × domestica</i>	<i>M. hybrid</i>
MdSNPui04662 <sup>B</sup>	0	34966610	XP_003537933.1	predicted: salt tolerance protein-like isoform 1, COL domain class transcription factor	84.5	237	1	1	1
GDsnp01916 <sup>L</sup>	1	13185230					0.9	0.47	0.43
MdSNPui07528 <sup>L</sup>	1	29037653	XP_002277103.1	asparagine synthetase 1	94.3	581	1	0.9	0.88
MdSNPui08653 <sup>L</sup>	2	19130120	XP_003531078.1	predicted: 26S proteasome non-ATPase regulatory subunit RPN12A-like	92.1	266	1	0.8	0.83
MdSNPui09323 <sup>L</sup>	2	29798888	XP_002263156.1	predicted: dnaJ protein homolog, heat shock protein	96.4	415	1	0.86	0.88
GDsnp00588 <sup>L</sup>	3	2107141	NP_181757.2	transcription factor bHLH130	77	122	0.83	0.31	0.53
MdSNPui08596 <sup>L</sup>	3	4382784					0.93	0.54	0.76
MdSNPui11939 <sup>L</sup>	4	12449449					0.88	0.4	0.48
MdSNPui01585 <sup>L</sup>	4	19100858	XP_002318956.1	acidity, vacuolar H <sup>+</sup> -translocating inorganic pyrophosphatase	94.8	756	1	0.9	0.93
MdSNPui06825 <sup>L</sup>	4	22399444	XP_002336919.1	putative CC-NBS-LRR protein	53.1	245	0.8	0.29	0.24
MdSNPui06330 <sup>L</sup>	5	291398	XP_003610595.1	Proline-rich protein	77	260	1	0.78	0.79
MdSNPui08785 <sup>L</sup>	5	680454	XP_003607876.1	F-box protein PP2-B10, Phloem protein 2 (pp2-1 gene), cultivar Florina,	46.5	202	1	0.8	0.9
MdSNPui08770 <sup>L</sup>	5	5292467	XP_002266308.2	predicted: oryzain alpha chain-like, cysteine protease, putative	68.3	221	1	0.8	0.66
MdSNPui04326 <sup>L</sup>	5	6362466	XP_002316737.1	chloroplast LHClI type I chlorophyll a-b binding precursor protein	98.9	263	1	0.85	0.67
MdSNPui07414 <sup>L</sup>	5	17952130	XP_002266481.2	sugar, predicted: dCTP pyrophosphatase 1-like	89.5	114	1	0.89	0.9
MdSNPui03960 <sup>L</sup>	5	18984890	NP_178171.1	acidity, 3-isopropylmalate dehydrogenase 2	85.4	402	1	0.86	0.74
MdSNPui08927 <sup>L</sup>	5	22399567	NP_188284.1	predicted: translocase of chloroplast 132, chloroplastic-like	82.8	612	0.73	0.18	0.22
MdSNPui07915 <sup>L</sup>	6	20851825	XP_002275117.1	predicted: putative pectinesterase/pectinesterase inhibitor 26-like	52.1	241	1	0.81	0.79
MdSNPui11553 <sup>L</sup>	7	20876	transcribed	heme binding protein			1	0.78	0.9
MdSNPui09322 <sup>BL</sup>	7	5479896	XP_002263156.1	dnaJ protein homolog, heat shock protein	96.4	415	1	0.73	0.72
MdSNPui01415 <sup>L</sup>	7	11215023	NP_188258.1	GDSL esterase/lipase APG, zinc finger protein	78.9	350	0.93	0.53	0.48
GDsnp00699 <sup>L</sup>	7	20880172	XP_003550092.1	transmembrane protein 70 homolog, mitochondrial-like	82.2	240	1	0.93	0.91
GDsnp02291 <sup>L</sup>	7	21051042					0.93	0.49	0.52
MdSNPui09399 <sup>L</sup>	7	21683503	XP_002280495.1	predicted: putative pterin-4-alpha-carbinolamine dehydratase 1-like	79.5	209	0.98	0.56	0.69
MdSNPui09396 <sup>BL</sup>	7	21698261	XP_002280495.1	predicted: putative pterin-4-alpha-carbinolamine dehydratase	79.5	209	1	0.7	0.79
MdSNPui06769 <sup>BL</sup>	7	22658462	XP_001699012.1	peptidyl-prolyl cis-trans isomerase, FKBP-type, FK506-binding protein, putative	66	103	1	0.61	0.76
MdSNPui03340 <sup>L</sup>	7	24570799	NP_194094.1	reticulon-like protein B2-like isoform 1	67.3	268	0.85	0.3	0.41

(cont'd)



**Table 2. Continued.**

Single nucleotide polymorphism	Chromosome	Position (bp)	Protein ID	Unigene/blastx <sup>†</sup>	Id (%)	Len (aa)	<i>M. sieversii</i>	<i>M. × domestica</i>	<i>M. hybrid</i>
MdSNPui03173 <sup>L</sup>	8	3443435	NP_173114.1	Alpha-helical ferredoxin, NADH-quinone oxidoreductase subunit I	95.5	221	0.9	0.45	0.52
MdSNPui08601 <sup>BL</sup>	8	12201364	XP_003545336.1	predicted: LOB domain-containing protein 38	81.4	70	0.98	0.54	0.43
MdSNPui06951 <sup>L</sup>	8	21839764	XP_002264360.2	predicted: extensin-3-like	53.6	166	1	0.84	0.72
MdSNPui00311 <sup>L</sup>	9	15310382	XP_002284964.1	elongation factor 1 alpha	98.2	166	0.93	0.53	0.45
MdSNPui11107 <sup>L</sup>	9	31173087	XP_002277882.1	predicted: probable WRKY transcription factor 23, WRKY1	78	195	1	0.87	0.93
MdSNPui09716 <sup>L</sup>	10	7720319	XP_002277937.1	fruit size, putative axial regulator YABBY 2, YABBY domain class transcription factor	79.8	182	1	0.87	0.86
MdSNPui00408 <sup>L</sup>	10	8218445	XP_003563110.1	beta-ureidopropionase-like	89.7	397	1	0.81	0.9
GDsnp00258 <sup>L</sup>	10	30458919	NP_001154325.1	adenine nucleotide alpha hydrolases-like protein	73	241	1	0.79	0.78
MdSNPui11654 <sup>L</sup>	11	3990834	NP_196380.4	lipid transporter	62.7	118	0.65	0.15	0.22
MdSNPui04695 <sup>L</sup>	11	11357665	XP_002313510.1	aquaporin, MIP family, PIP subfamily, aquaporin PIP2	91.5	283	0.9	0.46	0.59
MdSNPui04696 <sup>L</sup>	11	11357905	XP_002313510.1	aquaporin, MIP family, PIP subfamily, aquaporin PIP2	91.5	283	0.88	0.35	0.5
GDsnp01640 <sup>L</sup>	11	15273189	XP_002279648.2	protein FAF-like, chloroplastic-like	56.2	104	0.98	0.59	0.72
MdSNPui05178 <sup>L</sup>	11	32270624	XP_003590616.1	putative senescence-associated protein SAG102	58.3	229	1	0.87	0.97
GDsnp00173 <sup>L</sup>	12	306576	XP_002277003.1	Glucan endo-1,3-beta-glucosidase precursor, putative	83.5	478	0.93	0.49	0.45
MdSNPui01058 <sup>L</sup>	12	1336099	XP_003520929.1	NADH-Cytochrome b5 reductase, putative	92.1	277	1	0.72	0.74
MdSNPui10357 <sup>L</sup>	12	1496415	XP_003553837.1	membrane steroid-binding protein 2-like	70.4	186	0.85	0.38	0.41
MdSNPui01007 <sup>BL</sup>	12	26392506	XP_002284203.1	ubiquitin-conjugating protein-like	100	147	0.85	0.34	0.36
GDsnp01874 <sup>L</sup>	12	30306267					1	0.86	0.88
GDsnp01647 <sup>BL</sup>	12	30409507					1	0.7	0.79
GDsnp01523 <sup>L</sup>	12	31227148					0.95	0.57	0.66
GDsnp00296 <sup>BL</sup>	12	31342412					0.98	0.49	0.62
MdSNPui10109 <sup>L</sup>	13	25537095	XP_002284638.1	predicted: 1-aminocyclopropane-1-carboxylate oxidase homolog 1	58.1	362	1	0.95	0.9
MdSNPui05417 <sup>BL</sup>	13	32069849	NP_194351.1	ATP synthase subunit G protein, hydrogen-transporting ATP synthase, rotational mechanism, putative	86.8	121	1	0.59	0.74
GDsnp01888 <sup>L</sup>	14	232566					0.68	0.14	0.16
MdSNPui09720 <sup>L</sup>	14	13545797	NP_850199.1	AP2-associated kinase <i>A. thaliana</i> , serine/threonine protein kinase, putative	62.5	517	1	0.87	0.83
MdSNPui08089 <sup>L</sup>	14	23936419	NP_565969.1	rhodanese-like domain-containing protein	78.5	232	0.88	0.44	0.48
MdSNPui00937 <sup>L</sup>	15	15668840					1	0.81	0.84
MdSNPui06548 <sup>BL</sup>	15	22681289	XP_002310744.1	disease resistance protein	58.3	156	0.98	0.42	0.41
MdSNPui08800 <sup>BL</sup>	15	46732617	XP_002281376.1	40S ribosomal protein S30	98.4	61	1	0.67	0.67
GDsnp01179 <sup>BL</sup>	16	4301368					1	0.69	0.67
MdSNPui08640 <sup>BL</sup>	16	4483457	NP_176247.2	predicted: pantothenate kinase 2-like, partial	79.2	173	0.95	0.44	0.36
MdSNPui07532 <sup>BL</sup>	16	4736021	XP_002270485.1	alpha-glucan water dikinase, chloroplastic	87.9	495	1	0.66	0.57
MdSNPui00129 <sup>B</sup>	16	5097313	XP_002273994.1	alanine aminotransferase 2 isoform 1	93.3	480	1	0.75	0.64
MdSNPui07063 <sup>L</sup>	16	8964268	XP_003594789.1	Ras-related small GTP-binding protein	97.4	192	1	0.81	0.83
MdSNPui11018 <sup>L</sup>	17	458698					0.6	0.1	0.12
MdSNPui04666 <sup>BL</sup>	17	5786868	XP_003537933.1	predicted: salt tolerance protein-like isoform 1, COL domain class transcription factor	81.9	237	1	0.73	0.72
MdSNPui00867 <sup>L</sup>	17	19195304	NP_172119.1	putative medium chain acyl-CoA oxidase	79.1	569	1	0.85	0.88
MdSNPui02268 <sup>L</sup>	17	19660061	XP_003551741.1	acidity, predicted: NADP-dependent malic enzyme-like	84.5	128	0.93	0.46	0.52
MdSNPui04667 <sup>L</sup>	17	20057880	XP_003537933.1	predicted: salt tolerance protein-like isoform 1, COL domain class transcription factor	81.9	237	1	0.79	0.81
MdSNPui11052 <sup>L</sup>	17	20831486	NP_568328.1	Plastidic glucose transporter 4, putative sugar transporter	85.5	543	1	0.96	0.88

<sup>†</sup> ATP, adenosine triphosphate; CC, coiled-coil domain; dCTP, deoxycytidine triphosphate; dnaJ, deoxyribonucleic acid chaperon protein; FAF, *Finogaldia magna* adhesion factor; FKBP, FK-binding protein; GDSL, Gly-Asp-Ser-(Leu) [GDS(L)] motif protein; LOB, lateral organ boundaries; LRR, leucine-rich repeat; NADH, nicotinamide adenine dinucleotide hydrogenated; NADP, nicotinamide adenine dinucleotide phosphate; NBS, nucleotide-binding site; PIP, prolactin-inducible protein; WRKY, conserved WRKYGQK peptide sequence and a zinc finger motif (CX4-7CX22-23HXH/C; YABBY, Adaxial-abaxial polarity transcription factor.

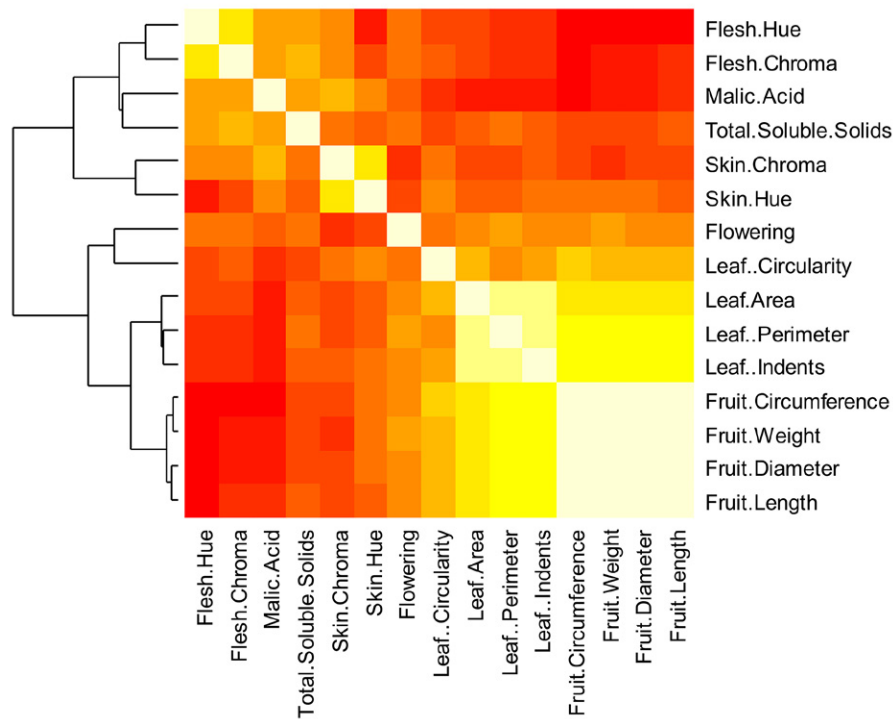


Figure 3. Pearson correlation analysis for all fruit quality and leaf traits for 131 apple accessions. Traits are grouped based on correlations among them, lighter colors correspond to stronger correlations.

listing of predicted possible functional associations for 55 SNP loci. A total of 12 SNPs did not show similarities with any database entry, including four genic SNPs (Khan et al., 2012a) and eight genomic SNPs (Velasco et al., 2010).

Among those 55 loci with functional associations, five loci (MdSNPui01585, MdSNPui07414, MdSNPui03960, MdSNPui09716, and MdSNPui02268) showed sequence similarities (80–95%) with transcripts predicted to be associated with sugar metabolism, acidity, and fruit size. Moreover, a total of seven loci on Chromosome 5 exhibited selection signatures. Of these, the SNP sequence of MdSNPui07414 showed similarity with dCTP pyrophosphatase 1-like, a sugar related protein. A second SNP of Chromosome 5, MdSNPui03960, showed similarity to an acidity related protein, 3-isopropylmalate dehydrogenase 2. In addition, two other SNPs of particular interest, MdSNPui09716 and MdSNPui02268, were located on Chromosomes 10 and 17, respectively. An EST sequence from NCBI (CV631928.1) harboring SNP MdSNPui09716 showed high sequence similarity (E-value =  $2e^{-111}$ ) to YABBY-2, a transcription factor previously reported to be associated with tomato fruit development (Bartley and Ishida, 2003) and more recently identified in apple (Costa et al., 2010), exhibiting the highest levels of change in mRNA abundance during fruit development. The sequence of the second SNP of interest, MdSNPui02268, was similar to that of an NADP (nicotinamide adenine dinucleotide phosphate)-dependent malic enzyme-like element. Overall, allele frequencies of these five SNPs were similar in both *M. × domestica* and the hybrid group, but differed significantly from allele frequencies in *M. sieversii*

and (Table 2), which was consistent for selection of alleles during domestication.

### Distribution, Correlation, and Clustering of Phenotypic Traits

Phenotypic data for 131 accessions were used for assessing trait distributions, correlations among traits, and percentage contribution of traits to phenotypic diversity using PCA. Overall, phenotypic traits were normally distributed, with the exception of fruit size parameters, as well as sugar content and acidity. As expected, fruits of wild *Malus* species were small in size, with high acidity and low sugar content. Pairwise Pearson correlation coefficients between these traits revealed that fruit size, including fruit weight, diameter, length, and circumference, yielded high positive correlations (Fig. 3). Moreover, leaf character parameters showed strong positive correlations amongst themselves as well as with fruit size parameters. Both fruit size and leaf parameters had strong negative correlations with sugar content, acidity, as well as with both peel (or skin) and flesh coloration (Fig. 3). Overall, fruit size, leaf attributes, and flowering time were clustered together. Similarly, sugar content, acidity, peel and flesh coloration were clustered together (Fig. 3). Malic acid content and peel coloration covaried in the same direction, while flesh coloration and sugar content also covaried in the same direction (Fig. 4). The first principal component (PC1), predominantly of fruit size parameters and leaf attributes, explained ~43% of the observed variation in *Malus* accessions used in this study. The PC2, predominantly of peel, flesh

## Variables factor map (PCA)

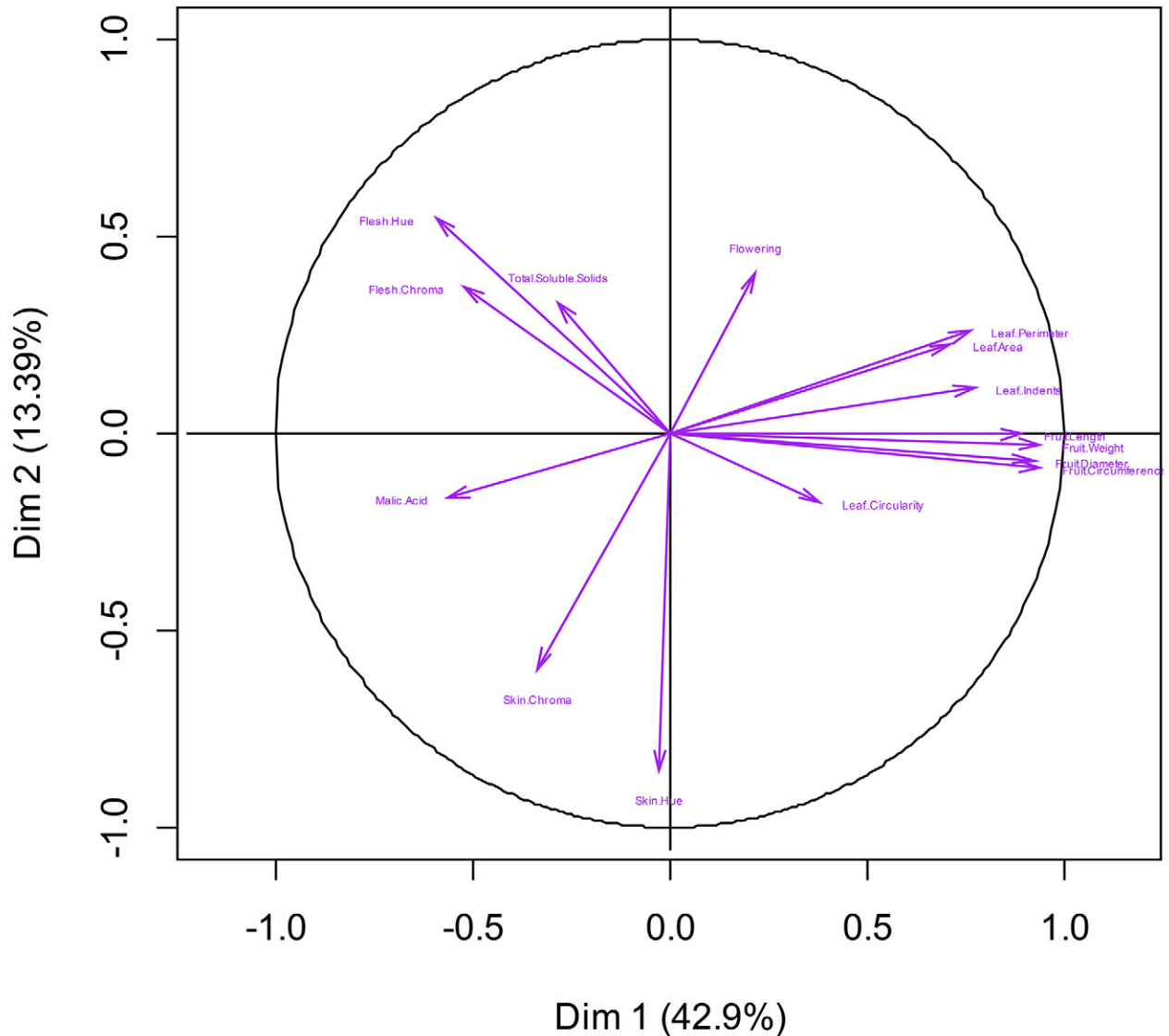


Figure 4. Principal component analysis (PCA) showing Dimension Parameters (Dim) 1 and 2 for all fruit quality traits and leaf attributes, skin, and flesh color in 131 *Malus* accessions, demonstrating grouping and direction of the traits.

coloration, sugar content, and acidity, explained ~13% of the observed variation; whereas, PC3, primarily of flesh coloration and sugar content, explained ~9% of the observed variation (Fig. 5).

The PCA for 131 *Malus* accessions, *M. × domestica* vs. *M. sieversii*, and *M. × domestica* vs. *M. sylvestris* did not reveal any large differences between the PC and variation explained. For those major contributors to PC2, including malic acid content and leaf traits, they were generally responsible for higher variation between *M. × domestica* vs. *M. sieversii* than for either 131 *Malus* accessions or *M. × domestica* vs. *M. sylvestris* (Fig. 5). Moreover, the amount of variation explained by fruit flesh color and sugar content, key contributors to PC3, between *M. × domestica* and *M. sieversii*, was

significantly higher compared with their contributions toward variation among 131 *Malus* accessions as well as between *M. × domestica* and *M. sylvestris* (Fig. 5). Overall, this may point to the important role of acidity and color traits in the divergence of *M. × domestica* from *M. sieversii* as both parameters were negatively correlated with fruit traits, which contributed minimally to the observed variation between the two species.

### Genome-Wide Association Study

Following Bonferroni's correction for multiple comparisons, no significant SNPs associated with observed phenotypic variations were detected. However, on visual inspection, 36 SNPs with the highest *p*-values were deemed to be likely candidates associated with one or

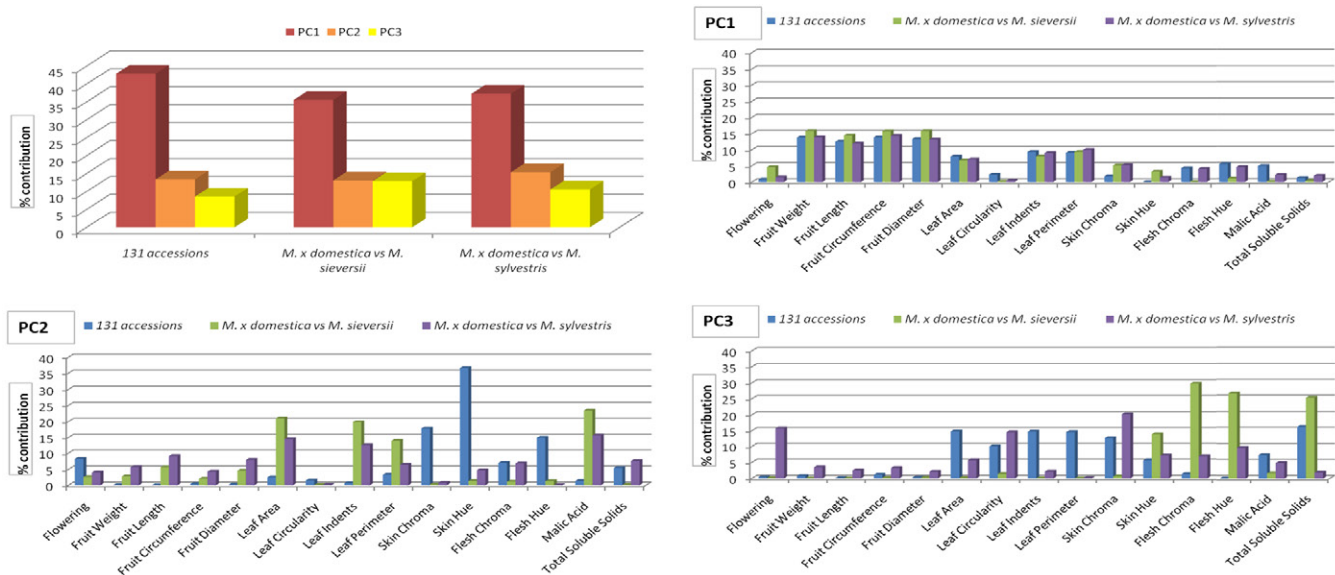


Figure 5. Principal component (PC) analysis showing overall contribution and importance of fruit quality attributes, leaf traits, and skin and flesh color for the first three PCs (i.e., PC1, PC2, and PC3) for 131 accessions, between *M. x domestica* and *M. sieversii* and between *M. x domestica* and *M. sylvestris*.

more traits from all 14 phenotypic traits evaluated (Table 3). Among these SNPs, three were from the draft apple genome sequence (Velasco et al., 2010), while the remainder (33) were EST-derived SNPs (Khan et al., 2012a). All these SNPs were distributed across 15 chromosomes with no associations identified on Chromosomes 10 and 13. Each of five SNPs showed potential associations on Chromosomes 3, 4, and 15, while four potential associations were found on Chromosome 12, and three associations were found on each of Chromosomes 5, 6, and 16. Allele frequencies for these SNPs ranged from 0.1 to 0.5.

The EST sequences for all genic SNPs with putative associations were obtained using blastn searches against the apple (*M. x domestica*) EST NCBI database. Unigenes of corresponding apple EST sequences with highest similarities and blastx searches against nonredundant protein sequences aided in assigning functions to genes containing these SNPs. It was found that two SNPs (MdSNPui10632, MdSNPui10633) from Chromosome 15 were deemed good candidates as they revealed associations with malic acid content and with fruit size parameters. In fact, the sequence of the MdSNPui10632 SNP, associated with malic acid content, was predicted to be similar to an NADP-dependent malic enzyme-like (Table 3, Fig. 6); while the sequence of the MdSNPui10633 SNP, associated with fruit size parameters, was similar to malate dehydrogenase (oxaloacetate-decarboxylating; NADP+). In addition, these two SNPs were very closely located to each other. Yet another SNP of potential interest, MdSNPui03155, located on Chromosome 11, was associated with peel color intensity, and was present in a gene with sequence similarity to a sucrose synthase 2-like protein (Table 3).

## Discussion

### Population Genetic Characteristics of Domesticated Apples

Many of the apple cultivars (*M. x domestica*) used in this study are predominantly derived from and/or linked to a few founder cultivars, thus affirming their close relationships to one another and suggesting a shared common ancestry. For example, cultivars Rosemary Russet, Smith Jonathan, Jonathan, Golden Delicious, and McIntosh have strong genetic relationships with many cultivars as demonstrated by their links to multiple accessions or cultivars. These findings are consistent with an earlier report (Noiton and Alspach, 1996), wherein Cox's Orange Pippin, Golden Delicious, Red Delicious, Jonathan, and McIntosh have been deemed as the most frequent founder cultivars. However, Red Delicious was not included in our study sample. In addition, Cox's Orange Pippin was not detected as a founder and this might be due to the composition of the different sets of cultivars used in this study. This may also account for the detection of Rosemary Russet as a key founder in this study, but not in a previous report (Noiton and Alspach, 1996).

Overall, a high level of diversity is observed within the *M. x domestica* collection used in this study. Considering the global genetic diversity in domesticated apples, it is reasonable to assume that a certain level of genetic diversity for various economic horticultural traits is present within *M. x domestica*. It is known that most current apple cultivars have been identified either as chance seedlings or bud sports of a single cultivar or developed from breeding programs (Noiton and Alspach, 1996). According to Evans et al. (2011), the majority of apple breeding programs around the world, and particularly European apple breeding programs, rely on improved

**Table 3. Single nucleotide polymorphisms (SNPs) significantly associated with 14 phenotypic traits identified using Genome-Wide Association Study (GWAS) in GAPIT (Genome Association and Prediction Integrated Tool), an R package. GAPIT uses Efficient Mixed Model Association (EMMA) method55. Sequences of SNPs showing significant associations were BLAST searched against the National Center for Biotechnology Information database for *Malus × domestica* ESTs using blastn and the unigenes for the corresponding apple EST sequence with highest similarity were recorded. The EST sequence was also BLAST searched using blastx against nonredundant protein sequences for homology. Chromosome position (bp), minor allele frequency (MAF), protein id from sequence blast results for top hit against nonredundant protein sequences, and GWAS significance level (p-value) for the trait for each SNP is also provided.**

SNPs	Chrom	Position (bp)	Protein ID	Unigene/blastx <sup>†</sup>	MAF	Flowering	Fruit circumference
MdSNPui04946	1	9582771	XP_003633964.1	reticulon-like protein B1 isoform 2 <i>V. vinifera</i> 79.0 214	0.3		
MdSNPui07247	2	8794039	XP_002276608.1	MLO9 protein <i>V. vinifera</i> 79.2 557	0.1		
MdSNPui11355	3	1020516	NP_001031195.4	SC35-like splicing factor 33 <i>A. thaliana</i> 75.2 214	0.4		
MdSNPui01364	3	7039674	NP_189621.1	general transcription factor 2-related zinc finger protein <i>A. thaliana</i> 47.0 230	0.2	0.003	
MdSNPui00312	3	9737978	XP_003547694.1	elongation factor 1-alpha-like isoform 1 <i>G. max</i> 98.7 446	0.3		
MdSNPui07830	3	16053768	XP_003634629.1	probable LRR receptor-like serine/threonine-protein kinase At4g08850-like <i>V. vinifera</i> 75.7 470	0.4		
MdSNPui05042	3	32301103	XP_003520230.1	glutathione S-transferase-like <i>G. max</i> 77.4 212	0.3		
MdSNPui11046	4	4403759	XP_003525411.1	PR-5 protein <i>G. max</i> 85.6 221	0.2		
MdSNPui02526	4	4548762	NP_176500.1	rhomboid-like 2 <i>A. thaliana</i> 78.2 243	0.4		
MdSNPui07106	4	4692027	NP_567380.2	Ara4-interacting protein <i>A. thaliana</i> 72.3 158	0.3		
MdSNPui11858	4	6878878	NP_196891.1	xyloglucan:xyloglucosyl transferase <i>A. thaliana</i> 83.3 286	0.3	0.004	
MdSNPui11898	4	11733909	NP_190930.1	pyrophosphorylase 4 <i>A. thaliana</i> 89.3 215	0.3		
MdSNPui07242	5	12313108	NP_565719.1	actin depolymerizing factor 6 <i>A. thaliana</i> 83.6 145	0.3		0.002
MdSNPui02859	5	17835394	NP_193544.1	ribosomal protein L32-1 <i>A. thaliana</i> 88.7 132	0.3		
GDsnp02729	5	27341501			0.1		
MdSNPui07718	6	4990256	NP_568645.1	duplicated SANT DNA-binding domain-containing protein <i>A. thaliana</i> 45.5 121	0.3		
MdSNPui07719	6	4990499	NP_568645.1	duplicated SANT DNA-binding domain-containing protein <i>A. thaliana</i> 45.5 121	0.3		
MdSNPui01624	6	5961708	NP_567522.5	peptidyl-prolyl cis-trans isomerase <i>A. thaliana</i> 71.2 333	0.5		
MdSNPui02362	7	21049993	XP_003541082.1	PREDICTED: 4-hydroxy-3-methylbut-2-enyl diphosphate reductase-like <i>G. max</i> 91.8 379	0.5		
MdSNPui11906	8	13931723	XP_002274966.1	profilin-1 isoform 1 <i>V. vinifera</i> 87.0 130	0.4		
MdSNPui04262	9	1793397	XP_003553833.1	carbonic anhydrase, chloroplastic-like isoform 1 <i>G. max</i> 80.3 329	0.4		
MdSNPui03155	11	34496174	XP_003521575.1	sucrose synthase 2-like <i>G. max</i> 87.0 59	0.2		
MdSNPui07111	12	21458032	XP_003537829.1	V-type proton ATPase 16 kDa proteolipid subunit-like isoform 1 <i>G. max</i> 100.0 163	0.1		
MdSNPui07110	12	21458804	XP_003537829.1	V-type proton ATPase 16 kDa proteolipid subunit-like isoform 1 <i>G. max</i> 100.0 163	0.3		
MdSNPui06752	12	30130898	NP_001236007.1	putative aldo/keto reductase <i>G. max</i> 83.4 319	0.2		
GDsnp02228	12	30184029			0.1		
MdSNPui10600	14	17709952	XP_002269986.1	putative Holliday junction resolvase <i>V. vinifera</i> 83.8 167	0.4		
MdSNPui10632	15	3202683	XP_003522719.1	NADP-dependent malic enzyme-like <i>G. max</i> 89.2 582	0.3		
MdSNPui10633	15	3203185	NP_196728.1	malate dehydrogenase (oxaloacetate-decarboxylating)(NADP+) <i>A. thaliana</i> 86.7 578	0.3		
MdSNPui05276	15	12489150	XP_003524915.1	70 kDa peptidyl-prolyl isomerase-like isoform 1 <i>G. max</i> 85.4 547	0.3		
MdSNPui02219	15	14789358	XP_002285072.1	DEAD-box ATP-dependent RNA helicase 56-like <i>V. vinifera</i> 94.2 427	0.2		0.001
MdSNPui04592	15	25096687	XP_003540182.1	probable serine/threonine-protein kinase DDB_G0279405-like <i>G. max</i> 80.5 169	0.3		
MdSNPui04087	16	2650965	NP_001189504.1	uvrB/uvrC motif-containing protein <i>A. thaliana</i> 78.5 219	0.4		
GDsnp01179	16	4301368			0.2		
MdSNPui08271	16	15864142	NP_188935.1	ADP-ribosylation factor C1 <i>A. thaliana</i> 92.8 181	0.2	0.001	
MdSNPui05184	17	11061114	XP_003580532.1	36.4 kDa proline-rich protein-like <i>B. distachyon</i> 60.4 105	0.4		

<sup>†</sup> ADP, adenosine triphosphate; NADP, nicotinamide adenine dinucleotide.

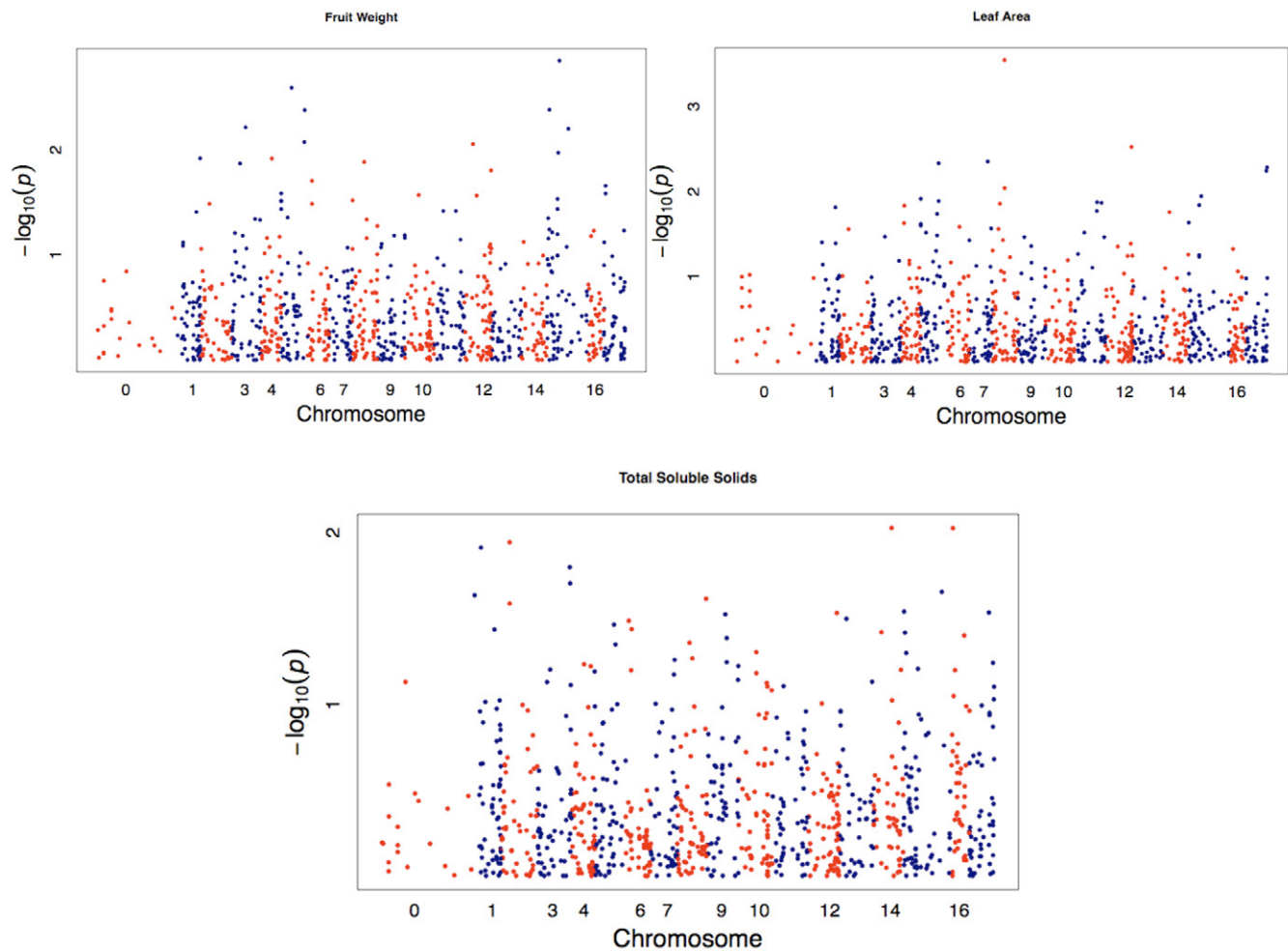


Figure 6. Genome-wide association study (GWAS) for leaf area, total soluble solids, and fruit weight for 158 apple accessions showing single nucleotide polymorphisms (SNPs) with significant associations compared with background. Colors represent SNPs located on alternating chromosomes.

traditional cultivars that have been either bred or discovered as chance seedlings several decades or centuries ago. Thus, only a few cultivars have served as the germplasm pool of most modern cultivars. Concerns over the use of a narrow pool of cultivars of domesticated apples in breeding programs have also been raised previously (Noiton and Alspach 1996), as this would highly influence future sustainability of apple production by decreasing horticultural and fruit quality attributes, rendering them prone to diseases and abiotic stresses. Similar to Noiton and Alspach (1996), a thoughtful approach for increasing the genetic basis of our apple improvement programs is needed. Patzak et al. (2012) used 10 SSRs in 273 *M. × domestica* accessions, including 130 reference world cultivars and 143 old and local genotypes. Their results suggest that many recent breeding programs are exploiting wider diversity in the *Malus* germplasm, thus contributing to enhanced sustainability of the domesticated apple.

### Relationships of Wild *Malus* Species with *M. × domestica*

The observed high numbers of relationships between *M. sieversii* and domesticated apples in the network analysis (Fig. 1C) suggest that *M. sieversii* is the primary progenitor of *M. × domestica*, as previously reported in earlier studies (Velasco et al., 2010; Harrison and Harrison, 2011; Cornille et al., 2012). The elevated LD in the pooled dataset compared to *M. × domestica* likely corresponds to genome-wide genetic differentiation among the different *Malus* species. There are also secondary introgressions from multiple *Malus* species into the *M. × domestica* genome following its origination from *M. sieversii*. Bayesian-based population genetic structure analysis (Fig. 1B) identified five clusters ( $K = 5$ ). In addition to *M. sylvestris*, a total of seven species could not be separated into distinct clusters. *M. micromalus* and *M. baccata* showed similar compositions of their genomes while *M. sieversii* and *M. orientalis* showed similar patterns in their genomes. This might be due to the fact that markers used for assessing these relationships were not sufficiently informative to distinguish beyond two

clusters for these four species. However, patterns shared by these species indicated that *M. micromalus* and *M. baccata* were as close to each other as were *M. sieversii* and *M. orientalis*, and these relationships were further confirmed using additional analyses (Fig. 1A, C, D). As for accessions showing admixture, this could be due to likely coancestry or secondary introgression from cocultivation or original location of the plant material.

In this study, the genome composition of *M. × domestica* shows admixture for all five clusters, but it has a major contribution from the blue component and some contribution from the yellow component (Fig. 1B). It is notable that both blue and yellow population components are hardly represented in the wild species used in the study. Based on these findings, it is proposed that the closest wild progenitor populations are either unsampled in this study or they are extinct. However, as presented in previous studies (Velasco et al., 2010; Cornille et al., 2012), substantial contributions from *M. sieversii*, the presumed progenitor, and *M. sylvestris*, the major secondary contributor to the domesticated apple genome, are observed.

As mentioned above, these findings are in agreement with previous studies (Velasco et al., 2010; Harrison and Harrison, 2011; Cornille et al., 2012), reporting that *M. × domestica* must have originated from *M. sieversii*. However, these results do not offer sufficient confidence in the claim that *M. sylvestris*, the European wild apple, has made only minor contributions to *M. × domestica* when compared with other wild species of *Malus*. Whether or not *M. sylvestris* harbors traits different from all other wild apple species that will render it uniquely capable of introgressing into *M. × domestica* is yet to be determined. Moreover, it is important to assess the extent to which domesticated apples from specific regions carry hallmarks of wild apples from their neighborhoods. Depending on the set of accessions and their geographic regions of origin, these factors can lead to different determinations of such relationships. In this study, several accessions of both *M. orientalis* and *M. prunifolia* have been included, and findings have suggested that both species have made considerable contributions to the domesticated apple along the Silk Road. For example, it has been previously reported that ‘Mandshurica 2330’ (PI 322713) is a separate species (‘Mandshurica’; Velasco et al., 2010). However, this is in contrast to previous USDA reports as well as to findings obtained in this study with twice as many *M. baccata* included, which demonstrate that the accession PI 322713 is indeed similar to other *M. baccata* accessions. Thus, a wider sampling of accessions along the Silk Road will provide a more comprehensive overview of the relationships among various *Malus* species, as well as their relationships and contributions to the domesticated apple, *M. × domestica*, genome.

In a recent study (Cornille et al., 2012), a comprehensive set of 839 accessions consisting of 368, 168, 215, 40, and 48 accessions of *M. × domestica*, *M. sieversii*, *M. orientalis*, *M. sylvestris*, and *M. baccata*, is used to assess genetic relationships among *Malus* using 26 SSR markers.

Based on their findings, *M. sieversii* has been confirmed as the progenitor of the domesticated apple. Moreover, it has been suggested there is a bidirectional gene flow between domesticated apples and their wild species, and *M. sylvestris* is deemed as the main secondary contributor to the *M. × domestica* genome (Cornille et al., 2012). Nevertheless, these findings remain inconclusive as to the extent of contribution as well as the number of introgressed wild species into the *M. × domestica* genome.

### Signatures of Selection across the *Malus* Genome and Corresponding Distribution of Phenotypic Traits in *M. × domestica* vs. Wild Species

During genome evolution of domesticates, genetic regions as well as genes undergoing selection are predicted to show significant reductions in genetic diversity when compared with neutral genes, as well as to higher levels of differentiation between the progenitor and the domesticated population (Tanksley and McCouch, 1997; Yamasaki et al., 2005). In this study, a total of 67 loci, including 13 genomic SNPs (Velasco et al., 2010) and 54 genic SNPs (Khan et al., 2012a) have been identified as potential targets for selection during the evolution of the *M. × domestica* genome from its wild progenitor *M. sieversii*. It is likely that there are some false positives among these detected loci. Nevertheless, loci identified by both  $F_{st}$  and Bayesian-based outlier detection methods, serving as potential targets for selection, as well as exhibiting significant differences in allele frequencies among *M. × domestica* and *M. sieversii*, are likely to be strong candidates. Regardless of the possible role of these loci in *Malus* domestication, these are proposed as potential candidates for selection, and deserve further investigation.

Except for a single SNP (MdSNPui04662), chromosomal positions of 66 SNPs were established along all 17 LGs of the *Malus* genome. It is presumed that genes associated with key traits for apple domestication are present in the regions with most loci under selection (Table 2). Besides clusters of disease resistance, quantitative trait loci (QTLs) for fire blight disease (Calenge et al., 2005; Durel et al., 2009; Khan et al., 2006; Khan et al., 2012b), and QTLs for plant vigor and fruit quality traits of apple have also been reported to be present along these LGs (King et al., 2001; Liebhard et al., 2003; Longhi et al., 2012). Quantitative trait loci with moderate effects for variation in fruit texture (sensory traits, crispiness, and firmness, among others) on LGs 7, 5, and 12 have been reported (King et al., 2001). In addition, a strong QTL (LOD = 7) for flowering time on LG 17 has also been identified (Liebhard et al., 2003). Furthermore, QTLs for each of fruit flesh firmness and fruit weight, located on LG 12, along with two QTLs for flowering time with moderate effects, located on LG 7, as well as two QTLs for increased plant length, located on LG 17, have also been reported (Liebhard et al., 2003).

In this study, marked differences in allele frequencies for loci between domesticated and wild apples have been observed, suggesting that selection against one of the alleles might have occurred. It is interesting to note that allele frequencies in *M. × domestica* are in agreement with those of the hybrid group used in this study, thus suggesting that similar forces of selection are active in genomes of both groups. Based on BLAST searches of sequences containing candidate SNPs, five loci (MdSNPui01585, MdSNPui07414, MdSNPui03960, MdSNPui09716, and MdSNPui02268) show >80% sequence similarities to genes known to control sugar content, acidity, and fruit size. These are key traits during early domestication events of apple genome as reported previously by Juniper and Mabberley (2006).

On LG 5, a single sequence containing SNP MdSNPui07414 shows similarity to dCTP pyrophosphatase 1-like a sugar related protein, while the sequence of another SNP, MdSNPui03960, from LG 5 shows similarity to an acidity related protein, 3-isopropylmalate dehydrogenase 2. Two additional SNPs of major interest are MdSNPui09716 and MdSNPui02268, located on LGs 10 and 17, respectively. The sequence of SNP MdSNPui09716 shows similarity to a previously predicted fruit quality-related locus (Costa et al., 2010), while the sequence of SNP MdSNPui02268 is similar to NADP-dependent malic enzyme-like. Altogether, these five loci are potentially the best candidates for future studies.

In this study, distribution and correlation of traits in PCA analyses also point towards the potential role of specific phenotypic traits in apple domestication. In general, fruits of wild *Malus* species are characterized by small size, high acidity, and low sugar content. This is also accompanied with strong positive correlations among fruit dimension parameters. Furthermore, these parameters have strong negative correlations with sugar content, acidity (malic acid content), as well as with both peel and flesh coloration (Fig. 3). These phenotypic covariance findings suggest that when larger fruit size is favored during domestication, selection for low acidity and high sugar content is observed. The positive correlations of malic acid content and peel color, as well as of flesh color and sugar content, suggest that skin coloration serves as an indicator of fruit ripening and exposure to sun would contribute to accumulation of high sugar content. The PC1 of phenotypic traits, which predominantly involves fruit dimension parameters and leaf attributes, explains ~43% of the variation, but does not seem to account for observed variation between *M. × domestica* and *M. sieversii*. The malic acid content, as well as flesh color and sugar content, main contributors to variations explained by PC2 and PC3, respectively, accounted for more of the observed variation between *M. × domestica* and *M. sieversii* than among all 131 accessions or between *M. × domestica* and *M. sylvestris* (Fig. 5), and are likely to be key contributing traits from *M. sieversii* during domestication of apples. In contrast to previous studies (Juniper and Mabberley, 2006), fruit dimension parameters are significantly

different between *M. × domestica* and *M. sieversii*; however, they do not seem to be the main traits under selection between these two species. Instead, our findings indicate that other sensory attributes, such as color and taste, in the form of low acidity, are more important during the domestication process.

## Genome-Wide Association Study

Following Bonferroni's correction of multiple comparisons in significance tests, none of the SNPs above the significance threshold could be identified for trait associations in the GWAS. This could be due to the following two reasons: (i) rapid LD decay, and (ii) the small size of the association mapping population, which can both influence the power of association mapping (Khan and Korban, 2012). It is expected that, with a larger size population and a more dense SNP coverage across the genome, signals of SNPs and genomic regions would be above the threshold. To assess whether trends toward significant associations were present in our dataset, every LG was visually inspected to identify trends for elevated associations. A total of 36 SNPs with the highest *p*-values were identified using visual inspection (Table 3), and as valuable candidates for further research, are reported here. The distribution of SNPs on different chromosomes provided further support of the level of genetic complexity of the traits. Although these were not significant, elevated associations with traits and sequence similarities with transcripts relevant to key traits emphasized the potential roles of the SNPs in trait expression. For example, SNPs MdSNPui10632 and MdSNPui10633 on LG 15 showing associations with malic acid and fruit dimension parameters along with sequence similarities to malic acid metabolism related enzymes, as well as SNP MdSNPui03155 on LG 11 with association to skin color pigmentation (or hue) and sequence similarity with sucrose synthase 2-like protein, are all good candidates for further studies (Table 3).

## Supplemental Information Available

Supplemental information is included with this article.

## Acknowledgments

This work was funded by a grant received from USDA-NIFA-SCRI grant AG 2009-51181-06023. Partial funding was also provided by the University of Illinois Office of Research Project 875-325, and University of Illinois Office of Research Project 875-922.

## References

- Antao, T., A. Lopes, R. Lopes, A. Beja-Pereira, and G. Luikart. 2008. LOSITAN: A workbench to detect molecular adaptation based on a  $F_{st}$ -outlier method. *BMC Bioinform.* 9:323. doi:10.1186/1471-2105-9-323
- Bartley, G.E., and B.K. Ishida. 2003. Developmental gene regulation during tomato fruit ripening and in-vitro sepal morphogenesis. *BMC Plant Biol.* 3:4. doi:10.1186/1471-2229-3-4
- Batagelj, V., and A. Mrvar. 2004. Pajek—Analysis and visualization of large networks. Springer, Berlin. Available at [http://link.springer.com/content/pdf/10.1007/978-3-642-18638-7\\_4](http://link.springer.com/content/pdf/10.1007/978-3-642-18638-7_4) (verified 18 July 2014).
- Bylesjö, M., V. Segura, R.Y. Soolanayakanahally, A.M. Rae, J. Trygg, P. Gustafsson, S. Jansson, and N.R. Street. 2008. LAMINA: A tool for rapid quantification of leaf size and shape parameters. *BMC Plant Biol.* 8:82. doi:10.1186/1471-2229-8-82



- Calenge, F., D. Drouet, C. Denancé, W.E. Van de Weg, M.-N. Brisset, J.-P. Paulin, and C.-E. Durel. 2005. Identification of a major QTL together with several minor additive or epistatic QTLs for resistance to fire blight in apple in two related progenies. *Theor. Appl. Genet.* 111:128–135. doi:10.1007/s00122-005-2002-z
- Clark, R.M., E. Linton, J. Messing, and J.F. Doebley. 2004. Pattern of diversity in the genomic region near the maize domestication gene *tb1*. *Proc. Natl. Acad. Sci. USA* 101:700–707. doi:10.1073/pnas.2237049100
- Cornille, A., P. Gladieux, M.J. Smulders, I. Roldán-Ruiz, F. Laurens, B. Le Cam, A. Nersesyan, J. Clavel, M. Olonova, L. Feugey, I. Gabrielyan, X.G. Zhang, M.I. Tenaillon, and T. Giraud. 2012. New insight into the history of domesticated apple: Secondary contribution of the European wild apple to the genome of cultivated varieties. *PLoS Genet.* 8:e1002703. doi:10.1371/journal.pgen.1002703
- Costa, F., C.P. Peace, S. Stella, S. Serra, S. Musacchi, M. Bazzani, S. Sansavini, and W.E. Van de Weg. 2010. QTL dynamics for fruit firmness and softening around an ethylene-dependent polygalacturonase gene in apple (*Malus × domestica* Borkh.). *J. Exp. Bot.* 61:3029–3039. doi:10.1093/jxb/erq130
- Durel, C.-E., C. Denancé, and M.-N. Brisset. 2009. Two distinct major QTL for resistance to fire blight co-localize on linkage group 12 in apple genotypes Evereste and *Malus floribunda* clone 821. *Genome* 52:139–147. doi:10.1139/G08-111
- Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* 14:2611–2620. doi:10.1111/j.1365-294X.2005.02555.x
- Evans, K.M., A. Patocchi, F. Rezzonico, F. Mathis, C.E. Durel, F. Fernández-Fernández, A. Boudichevskaia, F. Dunemann, M. Stankiewicz-Kosyl, L. Gianfranceschi, M. Komjanc, M. Lateur, M. Madduri, Y. Noordijk, and W.E. van de Weg. 2011. Genotyping of pedigreed apple breeding material with a genome-covering set of SSRs: Trueness-to-type of cultivars and their parentages. *Mol. Breed.* 28:535–547. doi:10.1007/s11032-010-9502-5
- Foll, M., and O. Gaggiotti. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics* 180:977–993. doi:10.1534/genetics.108.092221
- Fuller, D.Q., L. Qin, Y. Zheng, Z. Zhao, X. Chen, L.A. Hosoya, and G.-P. Sun. 2009. The domestication process and domestication rate in rice: Spikelet bases from the Lower Yangtze. *Science* 323:1607–1610. doi:10.1126/science.1166605
- Gepts, P. 2004. Crop domestication as a long-term selection experiment. *Plant Breed. Rev.* 24:1–44.
- Götherström, A., C. Anderung, L. Hellborg, R. Elburg, C. Smith, D.G. Bradley, and H. Ellegren. 2005. Cattle domestication in the Near East was followed by hybridization with aurochs bulls in Europe. *Proc. Royal Soc. B Biol. Sci.* 272:2345–2351.
- Hammer, K. 1984. Das domestikationssyndrom. *Kulturpflanze* 32:11–34. doi:10.1007/BF02098682
- Harada, T., W. Kurahashi, M. Yanai, Y. Wakasa, and T. Satoh. 2005. Involvement of cell proliferation and cell enlargement in increasing the fruit size of *Malus* species. *Sci. Hortic. (Amsterdam)* 105:447–456. doi:10.1016/j.scienta.2005.02.006
- Harris, S.A., J.P. Robinson, and B.E. Juniper. 2002. Genetic clues to the origin of the apple. *Trends Genet.* 18:426–430. doi:10.1016/S0168-9525(02)02689-6
- Harrison, N., and R.J. Harrison. 2011. On the evolutionary history of the domesticated apple. *Nat. Genet.* 43:1043–1044. doi:10.1038/ng.935
- Huang, X., N. Kurata, X. Wei, Z.X. Wang, A. Wang, Q. Zhao, Y. Zhao, K. Liu, H. Lu, W. Li, Y. Guo, Y. Lu, C. Zhou, D. Fan, Q. Weng, C. Zhu, T. Huang, L. Zhang, Y. Wang, L. Feng, H. Furuumi, T. Kubo, T. Miyabayashi, X. Yuan, Q. Xu, G. Dong, Q. Zhan, C. Li, A. Fujiyama, A. Toyoda, T. Lu, Q. Feng, Q. Qian, J. Li, and B. Han. 2012. A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490:497–503. doi:10.1038/nature11532
- Hufford, M.B., X. Xu, J. van Heerwaarden, T. Pyhäjärvi, J.M. Chia, R.A. Cartwright, R.J. Elshire, J.C. Glaubitz, K.E. Guill, S.M. Kaeppler, J. Lai, P.L. Morrell, L.M. Shannon, C. Song, N.M. Springer, R.A. Swanson-Wagner, P. Tiffin, J. Wang, G. Zhang, J. Doebley, M.D. McMullen, D. Ware, E.S. Buckler, S. Yang, and J. Ross-Ibarra. 2012. Comparative population genomics of maize domestication and improvement. *Nat. Genet.* 44:808–813. doi:10.1038/ng.2309
- Hummer, K.E., and J. Janick. 2009. Rosaceae: Taxonomy, economic importance, genomics. In: K.M. Folta and S.E. Gardiner, editors, *Genetics and genomics of Rosaceae*. Springer, New York. p. 1–17.
- Huson, D.H., and D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23:254–267. doi:10.1093/molbev/msj030
- Juniper, B.E., and D.J. Mabblerley. 2006. *The story of the apple*. Timber Press, Portland, OR.
- Kellerhals, M. 2009. Introduction to apple (*Malus × domestica*). In: K.M. Folta and S.E. Gardiner, editors, *Genetics and genomics of Rosaceae*. Springer, New York. p. 73–84.
- Khan, M.A., B. Duffy, C. Gessler, and A. Patocchi. 2006. QTL mapping of fire blight resistance in apple. *Mol. Breed.* 17:299–306. doi:10.1007/s11032-006-9000-y
- Khan, M.A., Y. Han, Y.F. Zhao, and S.S. Korban. 2012a. A high-throughput apple SNP genotyping platform using the GoldenGate™ assay. *Gene* 494:196–201. doi:10.1016/j.gene.2011.12.001
- Khan, M.A., and S.S. Korban. 2012. Association mapping in forest trees and fruit crops. *J. Exp. Bot.* 63:4045–4060. doi:10.1093/jxb/ers105
- Khan, M.A., Y. Zhao, and S.S. Korban. 2012b. Molecular mechanisms of pathogenesis and resistance to the bacterial pathogen *Erwinia amylovora*, causal agent of fire blight disease in Rosaceae. *Plant Mol. Biol. Rep.* 30:247–260. doi:10.1007/s11105-011-0334-1
- King, G.J., J.R. Lynn, C.J. Dover, K.M. Evans, and G.B. Seymour. 2001. Resolution of quantitative trait loci for mechanical measures accounting for genetic variation in fruit texture of apple (*Malus pumila* Mill.). *Theor. Appl. Genet.* 102:1227–1235. doi:10.1007/s001220000530
- Konishi, S., T. Izawa, S.Y. Lin, K. Ebana, Y. Fukuta, T. Sasaki, and M. Yano. 2006. An SNP caused loss of seed shattering during rice domestication. *Science* 312:1392–1396. doi:10.1126/science.1126410
- Korban, S.S., and S. Tartarini. 2009. Apple structural genomics. In: K.M. Folta and S.E. Gardiner, editors, *Genetics and genomics of Rosaceae*. Springer, New York. p. 85–119.
- Liebard, R., M. Kellerhals, W. Pfammatter, M. Jertmini, and C. Gessler. 2003. Mapping quantitative physiological traits in apple (*Malus × domestica* Borkh.). *Plant Mol. Biol.* 52:511–526. doi:10.1023/A:1024886500979
- Lipka, A.E., F. Tian, Q. Wang, J. Peiffer, M. Li, P.J. Bradbury, M.A. Gore, E.S. Buckler, and Z. Zhang. 2012. GAPIT: Genome association and prediction integrated tool. *Bioinformatics* 28:2397–2399. doi:10.1093/bioinformatics/bts444
- Longhi, S., M. Moretto, R. Viola, R. Velasco, and F. Costa. 2012. Comprehensive QTL mapping survey dissects the complex fruit texture physiology in apple (*Malus × domestica* Borkh.). *J. Exp. Bot.* 63:1107–1121. doi:10.1093/jxb/err326
- Malladi, A., and P.M. Hirst. 2010. Increase in fruit size of a spontaneous mutant of ‘Gala’ apple (*Malus × domestica* Borkh.) is facilitated by altered cell production and enhanced cell size. *J. Exp. Bot.* 61:3003–3013. doi:10.1093/jxb/erq134
- Miller, A.J., and B.L. Gross. 2011. From forest to field: Perennial fruit crop domestication. *Am. J. Bot.* 98:1389–1414. doi:10.3732/ajb.1000522
- Myles, S., A.R. Boyko, C.L. Owens, P.J. Brown, F. Grassi, M.K. Aradhya, B. Prins, A. Reynolds, J.M. Chia, D. Ware, C.D. Bustamante, and E.S. Buckler. 2011. Genetic structure and domestication history of the grape. *Proc. Natl. Acad. Sci. USA* 108:3530–3535. doi:10.1073/pnas.1009363108
- Noiton, D.A., and P.A. Alspach. 1996. Founding clones, inbreeding, coancestry, and status number of modern apple cultivars. *J. Am. Soc. Hortic. Sci.* 121:773–782.
- Olsen, K.M., and J.F. Wendel. 2013. A Bountiful Harvest: Genomic insights into crop domestication phenotypes. *Annu. Rev. Plant Biol.* 64:47–70. doi:10.1146/annurev-arplant-050312-120048
- Outram, A.K., N.A. Stear, R. Bendrey, S. Olsen, A. Kasparov, V. Zaibert, N. Thorpe, and R.P. Evershed. 2009. The earliest horse harnessing and milking. *Science* 323:1332–1335. doi:10.1126/science.1168594
- Palaisa, K., M. Morgante, S. Tingey, and A. Rafalski. 2004. Long-range patterns of diversity and linkage disequilibrium surrounding the maize Y1 gene are indicative of an asymmetric selective sweep. *Proc. Natl. Acad. Sci. USA* 101:9885–9890. doi:10.1073/pnas.0307839101

- Paran, I., and E. Van der Knaap. 2007. Genetic and molecular regulation of fruit and plant domestication traits in tomato and pepper. *J. Exp. Bot.* 58:3841–3852. doi:10.1093/jxb/erm257
- Patzak, J., F. Paprstein, A. Henychová, J. Sedláč, and D. Somers. 2012. Comparison of genetic diversity structure analyses of SSR molecular marker data within apple (*Malus × domestica*) genetic resources. *Genome* 55:647–665. doi:10.1139/g2012-054
- Pollinger, J.P., K.E. Lohmueller, E. Han, H.G. Parker, P. Quignon, J.D. Degenhardt, A.R. Boyko, D.A. Earl, A. Auton, A. Reynolds, K. Bryc, A. Brisbin, J.C. Knowles, D.S. Mosher, T.C. Spady, A. Elkahloun, E. Geffen, M. Pilot, W. Jedrzejewski, C. Greco, E. Randi, D. Bannasch, A. Wilton, J. Shearman, M. Musiani, M. Cargill, P.G. Jones, Z. Qian, W. Huang, Z.L. Ding, Y.P. Zhang, C.D. Bustamante, E.A. Ostrander, J. Novembre, and R.K. Wayne. 2010. Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* 464:898–902. doi:10.1038/nature08837
- Potts, S.M., Y. Han, M.A. Khan, M.M. Kushad, A.L. Rayburn, and S.S. Korban. 2012. Genetic diversity and characterization of a core collection of *Malus* germplasm using simple sequence repeats (SSRs). *Plant Mol. Biol. Rep.* 30:827–837. doi:10.1007/s11105-011-0399-x
- Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
- Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A. Ferreira, D. Bender, J. Maller, P. Sklar, P.I. de Bakker, M.J. Daly, and P.C. Sham. 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
- Purugganan, M.D., and D.Q. Fuller. 2009. The nature of selection during plant domestication. *Nature* 457:843–848. doi:10.1038/nature07895
- Tanksley, S.D., and S.R. McCouch. 1997. Seed banks and molecular maps: Unlocking genetic potential from the wild. *Science* 277:1063–1066. doi:10.1126/science.277.5329.1063
- Tian, F., N.M. Stevens, and E.S. Buckler. 2009. Tracking footprints of maize domestication and evidence for a massive selective sweep on chromosome 10. *Proc. Natl. Acad. Sci. USA* 106:9979–9986. doi:10.1073/pnas.0901122106
- Velasco, R., A. Zharkikh, J. Affourtit, A. Dzingra, A. Cestaro, A. Kalyanaraman, P. Fontana, S.K. Bhatnagar, M. Troggo, D. Pruss, S. Salvi, M. Pindo, P. Baldi, S. Castelletti, M. Cavaiuolo, G. Coppola, F. Costa, V. Cova, A. Dal Ri, V. Goremykin, M. Komjanc, S. Longhi, P. Magnago, G. Malacarne, M. Malnoy, D. Micheletti, M. Moretto, M. Perazzolli, A. Si-Ammour, S. Vezzulli, E. Zini, G. Eldredge, L.M. Fitzgerald, N. Gutin, J. Lanchbury, T. Macalma, J.T. Mitchell, J. Reid, B. Wardell, C. Kodira, Z. Chen, B. Desany, F. Niazi, M. Palmer, T. Koepke, D. Jiwan, S. Schaeffer, V. Krishnan, C. Wu, V.T. Chu, S.T. King, J. Vick, Q. Tao, A. Mraz, A. Stormo, K. Stormo, R. Bogden, D. Ederle, A. Stella, A. Vecchiatti, M.M. Kater, S. Masiero, P. Lasserre, Y. Lespinasse, A.C. Allan, V. Bus, D. Chagné, R.N. Crowhurst, A.P. Gleave, E. Lavezzo, J.A. Fawcett, S. Proost, P. Rouzé, L. Sterck, S. Toppo, B. Lazzari, R.P. Hellens, C.E. Durel, A. Gutin, R.E. Bumgarner, S.E. Gardiner, M. Skolnick, M. Egholm, Y. Van de Peer, F. Salamini, and R. Viola. 2010. The genome of the domesticated apple (*Malus × domestica* Borkh.). *Nat. Genet.* 42:833–839. doi:10.1038/ng.654
- Wang, R.-L., A. Stec, J. Hey, L. Lukens, and J. Doebley. 1999. The limits of selection during maize domestication. *Nature* 398:236–239. doi:10.1038/18435
- Yamasaki, M., M.I. Tenaillon, I.V. Bi, S.G. Schroeder, H. Sanchez-Villeda, J.F. Doebley, B.S. Gaut, and M.D. McMullen. 2005. A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. *Plant Cell* 17:2859–2872. doi:10.1105/tpc.105.037242