Full Title

Deducing genotypes for loci of interest from SNP array data via haplotype sharing, demonstrated for apple and cherry

Short title

Deducing locus genotypes from SNP array data

Authors

Alexander Schaller[1,2], Stijn Vanderzande[1], Cameron Peace[1]*

[1] Department of Horticulture, Washington State University, Pullman, WA, United States of America
[2] Present address: Department of Environmental Horticulture, University of Florida, Gainesville, FL, United States of America

* Corresponding author
Email: cpeace@wsu.edu (CP)

Author Contributions: Conceptualization, C.P.; Data Curation, S.V.; Funding Acquisition, C.P.; Investigation, A.S.; Methodology, A.S., S.V. and C.P.; Project Administration, C.P.; Supervision, S.V., C.P.; Validation, A.S.; Visualization, A.S., C.P.; Writing – Original Draft Preparation, A.S.; Writing – Review & Editing, A.S., S.V. and C.P.

Abstract

Breeders, collection curators, and other germplasm users require genetic information, both genome-wide and locus-specific, to effectively manage their genetically diverse plant material. SNP arrays have become the preferred platform to provide genome-wide genetic profiles for elite germplasm and could also provide locus-specific genotypic information. However, genotypic information for loci of interest such as those within PCR-based DNA fingerprinting panels and trait-predictive DNA tests is not readily extracted from SNP array data, thus creating a disconnect between historic and new data sets. This study aimed to establish a method for deducing genotypes at loci of interest from their associated SNP haplotypes, demonstrated for two fruit crops and three locus types: quantitative trait loci *Ma* and *Ma3* for acidity in apple, apple fingerprinting microsatellite marker GD12, and Mendelian trait locus R_f for sweet cherry fruit color. Using phased data from an apple 8K SNP array and sweet cherry 6K SNP array, unique haplotypes spanning each target locus were associated with alleles of important breeding parents. These haplotypes were compared via identity-by-descent (IBD) or identity-by-state (IBS) to haplotypes present in germplasm important to U.S. apple and cherry breeding programs to deduce target locus alleles in this germplasm. While IBD segments were confidently tracked through pedigrees, confidence in allele identity among IBS segments used a shared length threshold. At least one allele per locus was deduced for 64–93% of the 181 individuals. Successful validation compared deduced R_f and GD12 genotypes with reported and newly obtained genotypes. Our approach can efficiently merge and expand genotypic data sets, deducing missing data and identifying errors, and is appropriate for any crop with SNP array data and historic genotypic data sets, especially where linkage disequilibrium is high. Locus-specific genotypic information extracted from genome-wide SNP data is expected to enhance confidence in management of genetic resources.

Introduction

Accurate genotypic information on identity, parentage, ancestry, breeding value, and performance potential informs effective germplasm management and use [1]. Historically, fruit breeders and collection curators have relied on meticulous passport and crossing records to be confident about identity, parentage, and ancestry and relied on phenotypic data to estimate genetic potential. Increasingly, locus-specific DNA tests for key traits, often based on simple PCR markers, have been used to determine the genotypes (i.e., allelic combinations) at trait loci of interest for cultivars and selections (e.g., [2–5]). In addition, small panels of neutral genetic markers have routinely been employed by germplasm managers to identify duplicates, infer pedigree relationships among germplasm individuals (mostly parent-child relationships), and to calculate overall relatedness among germplasm individuals. (e.g., [6–10]).

Single nucleotide polymorphisms (SNPs) have rapidly become the genetic marker of choice and are replacing previously developed marker types for a given organism. SNP arrays characterizing thousands of loci across the genome have been developed for fruit crops to provide desired genotypic information genome-wide [1, 11–22]. SNP arrays have been used to determine general relatedness among individuals as well as identify specific pedigree relationships [23–27]. SNP arrays have also been used to make genome-wide predictions for apple, cherry, and peach, in which breeding value and performance potential were based on cumulative information from tiny-effect alleles across the genome and a few large-effect alleles of quantitative trait loci (QTLs) [28–31]. In the RosBREED project [22, 32, 33], SNP arrays were developed and used in apple, cherry, and peach on large breeding germplasm sets that were pedigree-connected and included numerous important breeding parents and their ancestors ([34])

to identify and dissect loci influencing fruit quality and disease resistance traits and identify favorable and unfavorable alleles and their associated SNPs [35–45]. The data obtained from these SNP arrays were curated, which included combining SNPs into haploblocks delimited by historic recombination events and establishing the set of observed multi-SNP haplotypes at each haploblock for all genotyped germplasm individuals [46].

Despite the utility of SNP arrays, routine use in germplasm management and breeding of fruit crops is still limited. SNP arrays are most suited to named and clonally replicated individuals, such as cultivars, selections, parents, and germplasm collection accessions (rather than transient breeding seedlings, for example) because of the genotyping cost for each individual – tens of dollars (rather than cents to a few dollars) – and the data management burden. Thus, breeders and germplasm managers still rely on diagnostic information provided by locus-specific assays involving DNA markers such as SSRs, SCARs, or single SNPs targeting loci they are familiar with. In most cases, however, allelic information for those targeted loci is not directly provided by SNP arrays. Instead, genotypes for QTLs, Mendelian trait loci (MTLs), or any loci of interest such as multi-locus SSRs are not the immediate output of SNP arrays and are hidden within a sea of data points. Without dedicated methods to extract such information, previously obtained genotypic data sets relying on locus-specific markers risk becoming incompatible with newly generated SNP array data sets and germplasm users risk losing previous investments to characterize their material. Additionally, extraction of this information from newly scanned germplasm would enable comparisons to previous genotypic data sets without the need to invest in both genome-wide and locus-specific markers.

Therefore, if germplasm users could readily determine for any SNP array-genotyped individual its relatedness-revealing or functional (trait-influencing) alleles at loci of interest, they would be able to utilize their germplasm with increased confidence as well as merge informative data sets that are incompatible currently. Consequently, the objective of this study was to develop and validate a method to readily deduce alleles for any locus using genome-wide SNP array data and demonstrate it in apple and sweet cherry.

Materials and Methods

Data set

This study involved 121 apple and 60 cherry cultivars and their previously obtained genome-wide SNP data. A wide assortment of apple germplasm forming the RosBREED apple Crop Reference Set was previously assembled [34] and genotyped using the 8K SNP array [11]. In sweet cherry, a Crop Reference Set was also previously assembled [34], and the Breeding Pedigree Set of additional germplasm to specifically represent the Pacific Northwest Sweet Cherry Breeding Program [47], was also included. This cherry germplasm was genotyped using the 6K SNP array [12]. For both crops, the SNP data was quality-checked, phased, and haploblocked to result in two parental haplotypes for each individual in discrete units across each chromosome [46]. Only data for the chromosomes containing the target loci were used in this study. For apple, 247, 129, and 226 SNPs in 59, 53, and 55 haploblocks (HBs) covering chromosomes 3, 8, and 16 were included, respectively. For sweet cherry, 191 SNPs in 26 haploblocks covering chromosome 3 were included (S1 Tables).

Loci targeted and haploblock positions

Three types of loci were targeted. The first type was QTL, represented by the *Ma* and *Ma3* in apple. Both QTLs influence fruit acidity; the two loci together were reported to explain 66% of phenotypic variance among breeding germplasm derived from nine important apple breeding parents [40]. The second type of locus was MTL, represented by R_f in sweet cherry. This locus was reported to be associated with fruit color in sweet cherry, with two functional alleles that determine the major market classes of “mahogany” and “blush” [4]. The third type of locus was multi-allelic microsatellite, represented by GD12 in apple. This SSR [6, 48] is a component of a multi-SSR fingerprinting panel recommended for apple by the European Cooperative Programme for Plant Genetic Resources *Malus*/*Pyrus* working group [49], commonly used for studies of apple germplasm relatedness.

Genomic positions of the QTLs, MTL, and SSR were determined in relation to haploblocks by identifying the physical position of each locus in the appropriate reference genome and comparing this physical position to those of SNPs in the 8K apple [11, 46] and 6K cherry array [34] and the SNPs’ associated haploblocks. For *Ma*, the reported genomic position of the marker [50, 51] was used, and its physical position was determined on the GDDH13 v1.1 apple whole genome sequence [54] accessed via the Genome Database for Rosaceae [52] using a BLAST search [53]. For *Ma3*, the physical location of the informative SNP identified in [43] on the GDDH13 v1.1 apple whole genome sequence [54] was used as the location of the locus. The physical position in the sweet cherry genome [55] of Pav-R_f -SSR [4] was used for the R_f locus. The physical position of GD12 was determined by a BLAST search [53] of the SSR primer sequences against the GDDH13 v1.1 apple whole genome sequence [54] accessed via the Genome Database for Rosaceae [52].

Allele assignment and haplotype sharing via IBD or IBS

Reported genotypes of cultivars and their ancestors were assembled for the QTLs, MTL, and SSR. For *Ma* and *Ma3*, both functional alleles of nine important breeding parents and their ancestral origins were obtained from their previous allocations in [40] and [43]. For R_f, both functional alleles assigned to 17 cultivars and traced to their ancestral sources were obtained from [4]. While unique SNP haplotypes had previously been determined for R_f alleles, the interval spanned did not correspond with haploblocks subsequently delimited by [46]. Therefore, haplotype designations from [4] were used as historically recorded alleles to be deduced here using haplotype data from [46]. For GD12, both alleles for 20 ancestors included in the germplasm set of [34] were obtained from GRIN-Global (www.ars-grin.gov). Reported alleles were then associated with the haplotypes of the haploblocks they were located within (R_f) or with the combined haplotypes of the two haploblocks they were located between (*Ma*, *Ma3*, and GD12). “Haplotype pattern” hereafter refers to such single or combined haplotypes associated with specific locus alleles. In cases of multiple alleles of a target locus being associated with a single flanking haplotype pattern, haplotypes of additional upstream and downstream haploblocks were added one haploblock at a time until multi-haploblock haplotype patterns were uniquely associated with each functional allele. For adding such haploblocks, the locus was kept at the center and haploblocks were added to the shortest distance first, then additional haploblocks added progressively on each side so that flanking haploblocks downstream and upstream of the target locus were of near-equal genetic length.

Once all alleles for each locus had been associated with one unique haplotype pattern in cultivars with historically recorded genotypes (i.e., a single haplotype pattern was not associated with more than one locus allele), these haplotype patterns were traced back through the pedigree to the earliest known ancestor containing that haplotype pattern and this ancestor was then considered the ancestral source for that allele. Next, the haplotype patterns present in all cultivars and selections with unknown genotypes in each Crop Reference Set were compared to the haplotype patterns for the ancestral sources to identify all cases of haplotype sharing (Figure 1). Where a cultivar with an unknown genotype shared its haplotype pattern with an ancestral source, the locus allele of the source that was embedded within the haplotype pattern was assigned to the cultivar. In cases where a locus was located within a haploblock, the haplotypes of that haploblock were compared between each cultivar with an unknown genotype and the ancestral sources. Where the inheritance could be traced via known pedigree connections to a shared ancestor, the allele was noted to be deduced via identity-by-descent (IBD). For cultivars whose newly assigned alleles could not be traced through known pedigree connections to ancestral sources, alleles were noted as deduced via identity-by-state (IBS).

The total length of extended haplotype sharing with ancestral sources across adjacent haploblocks to the trait locus was also recorded. In cases where the same flanking haplotypes were identical to more than one ancestral source for the same allele, alleles were assigned according to IBD if possible or else according to the allele of the ancestral source with the longest extended shared haplotype. While IBD segments could be tracked through the pedigree for high confidence in identity, there was less certainty about IBS segments being truly identical between individuals, especially for short segments. Therefore, alleles assigned via IBS were only listed and considered successfully deduced if they had a longer extended shared haplotypes than the shortest extended shared haplotype observed for IBD segments in the data set, which was 9.4 cM.

Figure 1: Allele deduction via IBD (identity-by-descent) or IBS (identity-by-state) to tracking shared haplotypes in which an allele of interest is embedded, exemplified by the *G-ma3* allele of the father of 'Delicious'. Shown in green are extended haplotypes in coupling-phase linkage with *G-ma3* that are shared with the father of 'Delicious' without disruption by recombination; all other haplotypes are in gray. (**A**) The position of the *Ma3* locus is shown relative to haploblocks of chromosome 8 (only the immediately flanking haploblocks are shown). The exact position of *Ma3* is indicated using the physical position of the informative SNP identified in [43] and the flanking haploblocks are encompassed in the QTL regions identified in [40] and [43]. The *G-ma3* allele, shown as a black dot, is flanked by haplotypes 7 and 4 of HB-8-18 and HB-8-19, respectively. (**B**) Extended haplotypes in which the *G-ma3* allele is embedded that are shared with a particular ancestor (father of 'Delicious') via IBD or IBS are shown for various cultivars. The entire length of chromosome 8, horizontal bars, is displayed for all individuals.

Validation of deduced alleles

Locus genotypes of an additional 28 and 43 individuals, not previously used to associate haplotype alleles with locus alleles, were extracted from [4] for R_f and GRIN-Global for GD12, respectively. In addition, 49 individuals were independently genotyped for GD12 as follows: DNA for each individual was extracted according to [56], the GD12 SSR was amplified using primers and PCR conditions described in [6], resulting amplicons were separated and detected with an Applied Biosystems® 3730 DNA Analyzer, and observed amplicons were scored using GeneMarker® software. The proportion of new genotypic data for GD12 that matched allele deductions of each individual was calculated as the accuracy of deduction. Mismatches were examined carefully to determine whether they were due to incorrect genotype calls in the GRIN-Global data set or real mismatches between deductions from SNP data and observations from marker genotyping. It was not possible to validate the results of *Ma* and *Ma3* because allele designations were based on QTL analyses and such analyses, or the data required to conduct such analyses, were not available for any of the individuals with deduced genotypes.

Results

Locus genomic positions

The physical position of the apple *Ma* locus was determined to be between HB-16-5 and HB-16-6. *Ma3* was determined to be between HB-8-18 and HB-8-19, which was situated at the end of the consensus QTL positions determined in [40] and [43]. The sweet cherry R_f locus was determined to be within HB-3-17. The physical position of the apple GD12 locus was determined to be between HB-3-26 and HB-3-26a.

Allele deduction and validation for QTLs

Successful deduction of alleles via both IBD and IBS of extended shared haplotypes with ancestral sources of the nine important breeding parents was achieved for a high proportion of individuals of both apple QTLs. In total, at least one allele was deduced for 64% and 73% of the Crop Reference Set cultivars and selections for *Ma* and *Ma3,* respectively (S2 Tables). Complete genotypes (two alleles) for both the *Ma* and *Ma3* loci were deduced for 16 cultivars (14% of the 113 cultivars, excluding the nine important breeding parents), and at least one allele of both loci was deduced for a further 49 cultivars (43%). For the *Ma* locus alone, complete genotypes were deduced for 23 cultivars (20%) and one allele for 49 cultivars (42%). At the *Ma* locus, 70 homologs matched via IBD and 25 homologs via IBS, for which the IBS threshold was established as ≥9.4 cM (Table 1). For the *Ma3* locus, complete genotypes were deduced for 38 cultivars (34%) and one allele for 44 cultivars (39%). At the *Ma3* locus, 91 homologs matched via IBD and 29 homologs via IBS (Table 1). No alleles for *Ma* and *Ma3* could be deduced for 26 individuals that had no pedigree connection and also did not share extended haplotypes with the nine important breeding parents. Among the individuals with missing allele information for *Ma*, 24 unique haplotypes patterns were observed, with just six of these accounting for 76% of the undeduced allele cases. *Ma3* had 21 unique haplotypes not assigned to a known functional allele, with just three of these representing 58% of undeduced allele cases.

Table 1: Alleles deduced at the acidity trait loci *Ma* and *Ma3* for apple cultivars and selections by haplotype sharing. Individuals in **bold** shared the haplotypes via IBD, while the rest shared haplotypes via IBS. Underlined individuals are the nine important breeding parents of [40]. *Italicized* individuals are ancestors of the important breeding parents for which allele assignment was also determined in [40]. Individuals annotated with an asterisk (*) are ancestral sources of alleles. Immediately flanking haplotypes of the trait loci were identical for the following sets of alleles: $B_{F2/J}$-*Ma*, B_{Ws}-*Ma*, and *G-Ma*; *H-ma* and *I-ma*; C_{F2}-*Ma3* and C_{ES}-*Ma3*.

Allele (ancestral source)	Cultivars with shared haplotypes at the locus (length of extended shared haplotypes in cM)
A-Ma (Duchess of Oldenburg)	**Duchess of Oldenburg***, **Honeycrisp** (68.2)
$B_{F2/J}$-*Ma* (UP_Jonathan)	***Jonathan*** (68.2), **Enterprise** (39.3), $F_2$26829-2-2 (32.9) ***Idared***, **Arlet**, **Fiesta** (28.4), **Empress**, **Jonamac** (27.5), **Fireside**, **Minnewashta** (25), PRI 14-126 (22.1), Co-op 15, **Topaz** (17.9), PRI 1661-2 (14.9) **Akane**, **Delorgue**, **Sansa** (13.4)
B_{Ws}-*Ma* (Winesap)	**Winesap* (68.2), Aurora Golden Gala**, **Delicious**, **Splendour**, **BC 8S-27-43**, **Braeburn**, **Nicola**, **Scired**, **Sonya**, **WA 2** (39.3), **Fuji** (23.9), Montgomery (20.8), **NY 88** (17.9), Cox's Orange Pippin, Elstar, Ingrid Marie, James Grieve, Kidd's Orange Red, Lord Lambourne (12.0), **Empire** (9.4)
C-Ma (UP_Lady Williams)	***Lady Williams*** (68.2), **Cripps Pink** (17.9)

D-Ma (Frostbite)	**Frostbite***, ***Keepsake***, **Sweet 16** (68.2), **Honeycrisp** (22.1), M. Floribunda 821 (11.2)
E-Ma (NJ 27)	**WA 5**, **Co-op 15** (68.2)
F-Ma (NJ 136055)	***NJ 90*** (68.2), **WA 1** (9.4)
G-Ma (Grimes Golden)	**Grimes Golden***, ***Goldrush***, ***Golden Delicious***, **Ambrosia**, **Blushing Golden**, **Gala**, **Ginger Gold**, **Cripps Red**, **NY 752** (68.2), **Scifresh** (65.7), **PRI 14-126** (59.2), **WA 1** (37.5), **Autumn Crisp** (20.8), **Chinook** (14.9)
H-ma (UP_Golden Delicious)	***Golden Delicious***, **Delblush**, **Honeygold**, **Pinova** (68.2), **Arlet** (59.2), **Sunrise** (41.9), **Prima** (31.7), Wealthy, Fireside (24.9), **Splendour**, **WA 5**, **Chinook, Sciros** (24.0), **Cripps Pink** (22.1), **Jonafree**, **Tsugaru**, Esopus Spitzenburg, Jonathan, Akane, Monroe, NY 88, NY 752, Northern Spy, Keepsake, Sweet 16, Worcester Pearmain, Fortune (14.9), **Elstar** (13.4)
I-ma (UP_Delicious)	**Delicious**, ***Gala***, ***Kidd’s Orange Red***, **BC 8S-27-43**, **NY 543**, **Sansa**, **Scired**, **Sonya**, **Spartan**, **WA 2** (68.2), **Nicola** (54.1), **Ambrosia** (20), **Aurora Golden Gala** (14.9)
J-ma (McIntosh)	**McIntosh***, **Macoun**, **Regent** (68.2), **Cortland** (60.6), ***PRI 1661-2*** (47.6), **Liberty** (34.8), **Jonamac** (31.7), **Fantazja** (24.9), **Enterprise** (14.4)
*A-Ma3 (*Granny Smith)	**Granny Smith*** (62.0), ***Lady Williams***, **Cripps Red** (47.5), **Cripps Pink** (43.4), Frostbite, Keepsake, Sweet 16 (11.9)
B_{Mc}-Ma3 (McIntosh)	**McIntosh*, Regent** (62.0), Goodland (56.9), **Cortland** (47.5), **Enterprise**, **PRI 1661-2** (43.7)
C_{ES}-Ma3 (Esopus Spitzenburg)	**Esopus Spitzenburg*** (62.0), ***Idared***, ***Jonathan***, **Arlet**, **Autumn Crisp**, **Burgundy**, **Jonafree**, **Monroe**, **NY 752** (61.3), **Sawa** (34.3)
C_{F2}-Ma3 (F$_2$26829-2-2)	**F$_2$26829-2-2*** (62.0), Ben Davis, Cortland (54.4), Early Cortland (50.5), **Dayton** (48.5), **Prima** (40.8), ***PRI 14-126*** (39.2), ***Co-op 15***, **WA 5**, **PRI 1661-2** (26.9), ***Liberty*** (25.7), **NY 65707-19** (22.0)
D-Ma3 (McIntosh)	**McIntosh***, **Empire**, **Jonamac**, **Macoun** (62.0), **Sunrise** (55.7), ***NJ 90*** (47.5), ***Spartan***, **WA 1**, **NY 65707-19** (35.2), **Fantazja** (29.2)
E-Ma3 (Grimes Golden)	**Grimes Golden***, ***Golden Delicious***, **Arlet** (62.0), **Cameo** (48.6), **PRI 14-126** (41.8), **BC 8S-27-43**, **Gala**, **Nicola, Scired**, **Sciros**, **Sonya** (36.8), Delorgue (34.8), Granny Smith, Macoun, Liberty (13.8)
B_{GG+DO}-ma3 (Grimes Golden / Duchess of Oldenburg)	**Honeycrisp** (14.2)
F-ma3 (UP_Golden Delicious)	***Golden Delicious***, **Aurora Golden Gala, Splendour**, **WA 5**, **Ambrosia**, **Delblush**, **Ginger Gold**, **Cripps Red**, **Tsugaru** (62.0)**, Blushing Golden** (59.3), **WA 2** (57.8), **NY 752** (55.7), **Autumn Crisp** (54.5), **Cripps Pink**, **Honeygold**, **Silken** (48.6), ***Goldrush*** (40.7), **Pinova** (40.3), **WA 1** (37.6), **Elstar** (35), **Sunrise** (30.2)
G-ma3 (UP_Delicious)	**Delicious**, **Ambrosia**, **NJ 90**, **Spartan** (62.0), **Empire** (60.8), **Aurora Golden Gala** (57.3), ***Gala***, ***Kidd’s Orange Red***, **Scifresh** (55.8), **Fuji** (51.6), **Cameo** (48.6), **Splendour**, **Braeburn**, **Chinook**, **Nicola**, **Scired**, **Sciros**, **Sonya**, **WA 2** (45.2), **NY 543** (40.3), **Chinook** (38.4), Wagener, Idared, Fiesta (36.1), Cox’s Orange Pippin, Clivia (21.3), NY 543

	(18.3), Lodi, Ginger Gold, Montgomery (17.1), Fortune (15.2), Fireside, Minnewashta (14.2)
H-ma3 (Winesap)	**Winesap*** (62.0), **Delicious, Melrose** (60.1), Co-op 17 (38.5), Fortune (22.8)
I-ma3 (UP_Jonathan)	***Jonathan***, **Akane**, **Sansa**, **Tsugaru**, **Wealthy**, **Fireside** (62.0), **Empress** (51.0), **Jonamac** (49.7), **Oriole** (44.5), **Enterprise** (25.7)
J-ma3 (Northern Spy)	**Northern Spy*** (60.2), ***Keepsake*** (60.8), **Jonafree** (34.8), **Honeycrisp** (31.7), James Grieve (17.5), Cox's Orange Pippin, Kidd's Orange Red, Fiesta, Ingrid Marie (15.6)

Allele deduction and validation for an MTL

Both alleles of R_f were deduced for 45 (75%) of the 60 sweet cherry cultivars and selections (including the 17 ancestral sources) and one allele for an additional seven individuals (12%), resulting in at least one deduced allele for 86% selections of the sweet cherry Crop Reference Set (S2 Tables). Via IBD, 86 homologs matched, while 11 homologs matched via IBS (S3 Table). No alleles could be deduced for eight individuals due to missing haplotype data (five) or unique haplotype patterns (three). For the homologs that could not be deduced, 10 unique haplotype patterns were detected, however none were common. All deduced alleles matched genotypes reported by [4], resulting in a 100% deduction accuracy for this locus.

Allele deduction and validation for an SSR

For GD12, both alleles were deduced for 81 (67%) of the 121 cultivars and one allele for 28 cultivars (24%; S2 Tables). Among deduced alleles, 167 were deduced via IBD and 23 via IBS (S4 Table). It was not possible to deduce any alleles for 12 individuals of the Crop Reference Set because they were not pedigree-connected to others and their haplotypes did not match via IBS to the ancestral sources. For the undeduced alleles of GD12, 31 unique haplotype patterns were observed with seven of those patterns being present in more than one of the undeduced individuals. A total of 93 deduced alleles were validated using newly obtained SSR data for 49 individuals (95% of alleles present). Of the remaining five alleles, four could not be validated due to poor DNA quality of two individuals and resulting lack of PCR amplicons, while the last allele was associated with a unique haplotype pattern. Thus, all allele deductions that could be validated via independent and de novo genotyping were correct.

For validation of 77 allele deductions with GRIN-Global data, three deduced alleles (4%) did not match the reported alleles, each occurring in a separate cultivar (Arlet, Early Cortland, and Worcester Pearmain). Further comparison of the reference alleles, deduced alleles, and extended haplotypes of these three cultivars with those of their parents, siblings, and offspring indicated that alleles were likely deduced correctly but that the GRIN-Global data contained errors (S5 Tables). 'Arlet' was deduced as "155":"195" but reported as "155":"155". Its parents were reported as "155":"195" ('Golden Delicious') and "155":"155" ('Idared'), thus making both genotypes possible. However, one homolog of 'Arlet' matched the 'Golden Delicious' "195"-containing homolog across the entirety of chromosome 3. Thus, it was deduced that 'Arlet' should be "155":"195". 'Early Cortland' was deduced as "155":"187" but reported as "155":"155". Its parents were reported as "155":"187" ('Cortland') and "155":"187" ('Lodi'). However, one homolog of 'Early Cortland' matched the "187"-containing homolog of 'Cortland', inherited in turn from 'McIntosh', for the entirety of the chromosome and shared 48.5 cM across the GD12 locus with 'McIntosh'. 'Early Cortland' also shared 38-48.5 cM across this locus with other individuals that had inherited the 187-containing homolog from 'McIntosh'. Thus, it was deduced that

Early Cortland should be “155”:“187”. ‘Worcester Pearmain’ was deduced as “155”:”155” but reported as “155”:“187”. Although no parental information was available for this cultivar, its offspring had validated alleles of “155”:“155” (‘Discovery’) and “155”:“155” (‘Lord Lambourne’) and both individuals were determined to have inherited two different ‘Worcester Pearmain’ homologs. Thus, it was deduced that Worcester Pearmain should be “155”:“155”.

Identity-by-state among ancestral sources

Both QTLs had cases of ancestral IBS for the same functional allele (S6 Tables). For the *Ma* locus, cases of haplotype-sharing between two sources with different functional alleles occurred with the *A-Ma/J-ma* alleles. These ancestral sources only shared the immediate flanking haplotypes, so it was possible to differentiate between functional alleles by considering one additional haploblock on either side. All other functional alleles could be differentiated with just the two flanking haplotypes. In order to differentiate between ancestral sources sharing the same functional allele, it was necessary to include up to four flanking haploblocks on each side for a 7.7 cM haplotype pattern across the locus. For the *Ma3* locus, there were two cases where it was necessary to distinguish between different functional alleles with the same flanking haplotypes. The first case was between B_{Mc}*-Ma3, E-Ma3*, B_{GG+DO}*-ma3*, and *J-ma3* and the second case was between *D-Ma3, F-ma3*, and *H-ma3*. For the first case, up to five flanking haploblocks were needed on either side for a 12.7 cM haplotype pattern across the locus. The second case needed up to two flanking haplotypes on either side for 5.5 cM across the locus. All other functional alleles could be distinguished by just the flanking haplotypes. To differentiate between all ancestral sources sharing the same functional allele and the same haplotypes immediately around the locus (although they might represent an IBD allele just beyond the known pedigree), up to eight adjacent haploblocks on either side were needed, totaling up to 16.4 cM across the locus.

For the MTL, there was one case (haplotype 6) where it was necessary to include flanking haploblocks to distinguish between different functional alleles. Inclusion of both flanking haploblocks on each side provided 14.4 cM extended haplotypes that fully distinguished between the R_f and r_f alleles (S6 Tables). In all other cases for the MTL, it was only necessary to include the haploblock in which the locus was embedded. To effectively differentiate among ancestral sources with the same functional allele, up to eight haploblocks on each side of the locus were needed, for up to 39.7 cM in total across the locus (S6 Tables). The first case involved six individuals (‘Ambrunes’, ‘Bertiolle’, ‘Emperor Francis’, ‘Empress Eugenie’, ‘Napoleon’, and ‘Schmidt’) that all had haplotype 2, associated with the recessive r_f allele. All shared 1–7 haplotypes on either side of the locus, totaling 14.2—36.3 cM across the R_f locus. In the second case, both individuals (MIM 17 and MIM 23) had haplotype 18, associated with r_f and shared eight flanking haplotypes on either side of the locus totaling 39.7 cM across the R_f locus. In the third case, ‘Summit’ and ‘Schmidt’ shared seven flanking haplotypes on either side of the locus (36.3 cM), with haplotype 23 associated with the dominant R_f allele. The fourth case was between ‘Blackheart’ and PMR-1, which had haplotype 5 associated with R_f and shared seven flanking haplotypes on either side of the locus (36.3 cM). The fifth case was between ‘Ambrunes’, ‘Bertiolle’, and ‘Cristobalina’, all of which had haplotype 8 associated with R_f and shared one or two flanking haplotypes on either side of the locus (14.2–16.1 cM).

For the GD12 SSR locus, to differentiate among all functional alleles from different ancestral sources ,up to three flanking haploblocks on either side of the locus (4.5 cM) were needed (S6 Tables). There were four cases of IBS among multiple ancestral sources (S6 Tables). The first case involved the “155” allele that was shared by ‘Beauty of Bath’, ‘Cox’s Orange Pippin’, ‘Esopus Spitzenburg’, ‘Granny Smith’, ‘Malinda’, ‘McIntosh’, both homologs of ‘Northern Spy’, ‘Wagener’, both homologs of ‘Worcester Pearmain’, and the unknown parent of ‘Golden Delicious’. These nine ancestral sources shared 2–24 haploblocks (2.6–54.5 cM) across the GD12 locus. This “155” allele was the most common in the germplasm, with 59 additional cultivars having the allele and six of them matching via IBS. The second case of a shared haplotype was for the “157” allele of ‘Beauty of Bath’, ‘Esopus Spitzenburg’, ‘Montgomery’, ‘Rome Beauty’, and ‘Russian Seedling’. These ancestral sources shared up to 9.8 cM across the locus, so it was necessary to expand four haploblocks on both sides of the locus to differentiate among them. The third case was the “159” allele of ‘Ben Davis’, ‘Cox’s Orange Pippin’, ‘Granny Smith’, ‘Malinda’, ‘Winesap’, and the father of Delicious’ (UP_Delicious). These individuals shared the same flanking haplotype pattern at the locus, so it was necessary to include 9–24 haploblocks on both sides of the locus (15.29–54.5 cM) to differentiate them all. The fourth case was the “187” allele of ‘McIntosh and ‘Montgomery’ for which 2.9 cM was shared across the locus, so it was necessary to include three additional haploblocks on both sides to differentiate these ancestral sources.

Discussion

We successfully developed, demonstrated, and validated a method to deduce alleles from SNP array data for various types of loci that extrapolates known allele information for a few individuals to a larger germplasm set. While the method was demonstrated in apple and sweet cherry, it could be expanded to other types of loci and other crops that have SNP arrays or other genome-wide data available, especially where linkage disequilibrium is high. This approach enables germplasm users to extract information from previously characterized loci and extend this information to further individuals genotyped with the SNP array. In all cases where alleles could be deduced via IBD (i.e., for which inheritance of haplotypes could be traced from a shared ancestor), allele assignments were made with higher confidence than via IBS.

The developed method can be used to confirm reported genotypes. SSR genotyping is not always accurate as was identified here and has been reported in other studies [57–59]. Confirmation of reported results is important to ensure accuracy of published allele information for individuals. In all three cases where the GD12 allele did not match the GRIN-Global data, there were other validated individuals that had extended haplotypes matching the individual across the locus. While it is possible that a double recombination occurred at the location of the GD12 locus, these are rare events and highly unlikely to occur in the same genomic position in all three individuals. Alternatively, while parent-child and parent-parent-child errors (also called Mendelian-inconsistent errors) can be detected relatively easily, Mendelian-consistent errors (genotyping errors that do not infringe on Mendel’s inheritance laws) are harder to detect and require the phasing of linked loci [46]. Although no Mendelian-inconsistent errors were observed in the GRIN-Global data set, it is unlikely that any Mendelian-consistent errors were detected and resolved, especially because no or few flanking markers were available to conduct such error removal. Thus, it is more likely that the GRIN-Global data was incorrect as there were no possibilities for correction of Mendelian-consistent errors. Application of the method here easily identified the genotypic errors and could be systematically performed for listed genotypes of other loci in GRIN-Global datasets or reported elsewhere.

Cases of IBS among ancestral sources were detected for all loci investigated in this study. For IBS among ancestral sources with different functional alleles, the identical segments were often very short, with the longest being 14.5 cM (certain individuals with haplotype 6 of the R_f locus of cherry). However, to differentiate among ancestral sources with the same functional allele, it was often necessary to examine extended haplotypes on one or both sides of the target loci. Recent studies have reported that many historic apple cultivars are closely related with unknown recent shared ancestors [27, 61]. Thus, while these extended haplotype patterns with identical functional alleles were treated as originating different ancestral sources in this study, it is likely that in many of these cases a shared recent ancestral is the source of the allele. Therefore, both the IBD and IBS deductions capitalized on a high degree of linkage disequilibrium among the cultivated germplasm.

The many haplotypes observed in both apple and cherry that were not able to be associated with known alleles via IBD or IBS present opportunities for further research. While most of these allele-unassigned haplotypes were from individuals not pedigree-connected with other germplasm or poorly represented in the germplasm, there were also cases of haplotypes present in common ancestors but not represented in previous QTL studies. For example, the second *Ma* allele of the ancestor ‘McIntosh´ had extended haplotype-sharing via IBD or IBS with eight other cultivars but was not functionally characterized in the multi-parent study [40]. Therefore, its association with high or low acidity is unclear. To ascertain allele effects, an efficient approach would be to conduct DNA testing, or ideally QTL analyses, for sets of individuals representing the most common undetermined haplotypes (highlighted in S5 Tables). The method established in this study could then be applied to quickly deduce allele identities for all individuals sharing those haplotypes, efficiently expanding the number and proportion of germplasm individuals with genotypic information for loci of interest. Thus, the availability of a reference data set covering all or most of the observed haplotypes in relevant germplasm would be of much value for confident germplasm usage.

Opportunities for improvement of this method include determining extended haplotypes that are unambiguously associated with each allele as well as extending the method to unphased SNP data. Alleles deduced via IBD in this study were deduced unambiguously because pedigrees of these individuals were known, an approach originally outlined in [60], enabling establishment of IBD relationships for chromosomal regions among individuals. The ability to unambiguously assign alleles for loci where pedigree connections are unknown would greatly expand the allele information available. To do so, additional diagnostic SNPs could be developed and specifically used to genotype key individuals, or these additional SNPs could be included in future genome-wide assays. However, for immediate use of genotypic data sets in which ambiguity persists, some efficient shortcuts are available. Establishing thresholds of shared haplotype lengths by empirically determining the lengths at which matching of a known allele is unambiguous would enable rapid and confident allele assignment in IBS cases. Here, ≥9.4 cM was used as the threshold, taken from the minimum shared length observed via IBD with an ancestral source (which allowed for recombination to shorten shared haplotypes) among all the examined loci. Other methods for establishing confidence of deductions could be used, relying on empirical observations or theoretical calculations. For individuals with shared haplotypes that are not above the thresholds of unambiguity, alleles could be assigned according to their longest match, with the degree of confidence assigned according to the previously described empirical observations. Additionally, expanding the approach to unphased data could enable rapid extraction of valuable information from genome-wide SNP assays (such as SNP arrays or genotyping-by-sequencing), bypassing the time and effort for the data curation step of phasing, although at the expected cost of some loss of accuracy. Ultimately, the automation of such a method could enable genome-wide SNP data to be rapidly interpreted into allele information simultaneously for any and many loci, instead of obtaining information from one DNA test or genetic marker at a time. A streamlined process would further increase the ability for germplasm users to quickly gain allelic information about loci of interest for their germplasm while providing increased confidence in the utilization of genetic resources.

Acknowledgments

Jack Klipfel's assistance with conducting the SSR genotyping is gratefully acknowledged. This work was funded by USDA's National Institute of Food and Agriculture-Specialty Crop Research Initiative Project "RosBREED: Combining disease resistance and horticultural quality in new rosaceous cultivars" (2014-51181-22378) and the USDA National Institute of Food and Agriculture Hatch project 1014919, Crop Improvement and Sustainable Production Systems (WSU reference 00011).

References

1. Peace C. DNA-informed breeding of rosaceous crops: promises, progress and prospects. Hortic. Res. 2017; 4:17006.
2. Zhu Y, Barritt B. Md-ACS1 and Md-ACO1 genotyping of apple (*Malus* × *domestica* Borkh.) breeding parents and suitability for marker-assisted selection. Tree Genet. Genomes. 2008; 4:555-562.
3. Longhi S, Hamblin MT, Trainotti L, Peace CP, Velasco R, Costa F. A candidate gene based approach validates Md-PG1 as the main responsible for a QTL impacting fruit texture in apple (*Malus* x *domestica* Borkh). BMC Plant Biol. 2013; 13:37.
4. Sandefur P, Oraguzie N, Peace C. A DNA test for routine prediction in breeding of sweet cherry fruit color, Pav-R_f-SSR. Mol. Breeding. 2016; 36.
5. Vanderzande S, Piaskowski J, Luo F, Edge-Garza D, Klipfel J, Schaller A, et al. Crossing the finish line: How to develop diagnostic DNA tests as breeding tools after QTL discovery. J. Hortic. 2018; 5:228.
6. Hokanson S, Szewc-McFadden A, Lamboy W, McFerson J, Hokanson SC, Szewc-Mcfadden AK, Lamboy WF, McFerson JR. Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus* x *domestica* Borkh core subset collection. Theor. Appl. Genet. 1998; 97: 671-683.
7. Rosyara UR, Sebolt AM, Peace C, Iezzoni AF. Identification of the paternal parent of 'Bing' sweet cherry and confirmation of descendants using SNP markers. J. Amer. Soc. Hort. Sci. 2014; 139:148-156.
8. Lassois L, Denancé C, Ravon E, Guyader A, Guisnel R, Hibrand-Saint-Oyant L, et al. Genetic diversity, population structure, parentage analysis, and construction of core collections in the French apple germplasm based on SSR markers. Plant Mol. Biol. Rep. 2016; 34:827-844.
9. Urrestarazu J, Miranda C, Santesteban L, Royo J. Genetic diversity and structure of local apple cultivars from Northeastern Spain assessed by microsatellite markers. Tree Genet. Genomes. 2012; 8:1163-1180.
10. Urrestarazu J, Denancé C, Ravon E, Guyader A, Guisnel R, Feugey L, et al. 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. 2016; 16:130.
11. Chagné D, Crowhurst RN, Troggio M, Davey MW, Gilmore B, Lawley C, et al. Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. PLoS One. 2012; 7:e31745.
12. Peace C, Bassil N, Main D, Ficklin S, Rosyara UR, Stegmeir T, et al. Development and evaluation of a genome-wide 6K SNP array for diploid sweet cherry and tetraploid sour cherry. PLoS One. 2012; 7:e48305.
13. Verde I, Bassil N, Scalabrin S, Gilmore B, Lawley CT, Gasic K, et al. Development and evaluation of a 9K SNP array for peach by internationally coordinated SNP detection and validation in breeding germplasm. PLoS One. 2012; 7(4):e35668.
14. Le Paslier M-C, Choisne N, Scalabrin S, Bacilieri R, Berard AA, Bounon R, et al. The GrapeReSeq 18k Vitis genotyping chip. 9th International Symposium Grapevine Physiology and Biotechnology. La Serena, Chile; 2013. P.123

15. Bianco L, Cestaro A, Sargent DJ, Banchi E, Derdak S, Di Guardo M, et al. Development and validation of a 20K single nucleotide polymorphism (SNP) whole genome genotyping array for apple (*Malus* × *domestica* Borkh). PLoS One. 2014; 9(10):e110377.
16. Bianco L, Cestaro A, Linsmith G, Muranty H, Denancé C, Théron A, et al. Development and validation of the Axiom(®) Apple480K SNP genotyping array. Plant J. 2016; 86(1):62-74.
17. Bassil NV, Davis TM, Zhang H, Ficklin S, Mittmann M, Webster T, et al. Development and preliminary evaluation of a 90 K Axiom® SNP array for the allo-octoploid cultivated strawberry *Fragaria* × *ananassa*. BMC Genomics. 2015; 16:155.
18. Faivre-Rampant P, Zaina G, Jorge V, Giacomello S, Segura V, Scalabrin S, et al. New resources for genetic studies in *Populus nigra*: genome-wide SNP discovery and development of a 12k Infinium array. Mol. Ecol. Resour. 2016; 16:1023-1036.
19. Peace C, Bianco L, Troggio M, van de Weg E, Howard NP, Cornille A, et al. Apple whole genome sequences: recent advances and new prospects. Hortic. Res. 2019; 6:59.
20. Aranzana MJ, Decroocq V, Dirlewanger E, Eduardo I, Gao ZS, Gasic K, et al. *Prunus* genetics and applications after de novo genome sequencing: achievements and prospects. Hortic. Res. 2019; 6:58.
21. Vanderzande S, Zheng P, Cai L, Barac G, Gasic K, Main D, et al. The cherry 6+9K SNP array: a cost-effective improvement to the cherry 6K SNP array for genetic studies. Sci. Rep. 2020; 10:7613.
22. Iezzoni A, McFerson J, Luby JJ, Gasic K, Whitaker V, Bassil N, et al. RosBREED: bridging the chasm between discovery and application to enable DNA-informed breeding in rosaceous crops. Hortic. Res. 2020; 7:177.
23. Howard NP, van de Weg E, Bedford DS, Peace CP, Vanderzande S, Clark MD, et al. Elucidation of the 'Honeycrisp' pedigree through haplotype analysis with a multi-family integrated SNP linkage map and a large apple Hortic. Res. 2017; 4:17003.
24. Larsen B, Toldam-Andersen TB, Pedersen C, Ørgaard M. Unravelling genetic diversity and cultivar parentage in the Danish apple gene bank collection. Tree Genet. Genomes. 2017; 13:14.
25. Vanderzande S, Micheletti D, Troggio M, Davey MW, Keulemans J. Genetic diversity, population structure, and linkage disequilibrium of elite and local apple accessions from Belgium using the IRSC array. Tree Genet. Genomes. 2017; 13:125.
26. van de Weg E, Di Guardo M, Jänsch M, Socquet-Juglard D, Costa F, Baumgartner IO, et al. Epistatic fire blight resistance QTL alleles in the apple cultivar 'Enterprise' and selection X-6398 discovered and characterized through pedigree-informed analysis. Mol. Breeding. 2017; 38:5.
27. Howard NP, Peace C, Silverstein KAT, Poets A, Luby JJ, Vanderzande S, et al. The use of shared haplotype length information for pedigree reconstruction in asexually propagated outbreeding crops, demonstrated for apple and sweet cherry. Hortic. Res. 2021; 8:202.
28. Kumar S, Chagné D, Bink MC, Volz RK, Whitworth C, Carlisle C. Genomic selection for fruit quality traits in apple (*Malus* × *domestica* Borkh.). PLoS One. 2012; 7(5):e36674.
29. Piaskowski J, Hardner C, Cai L, Zhao Y, Iezzoni A, Peace C. Genomic heritability estimates in sweet cherry reveal non-additive genetic variance is relevant for industry-prioritized traits. BMC Genet. 2018; 19:23.
30. Hardner CM, Hayes BJ, Kumar S, Vanderzande S, Cai L, Piaskowski J, et al. Prediction of genetic value for sweet cherry fruit maturity among environments using a 6K SNP array. Hortic Res. 2019; 6:6.
31. Hardner C, Kumar S, Main D, Peace C. Global genomic prediction in horticultural crops: Promises, progress, challenges and outlook. Front. Agr. Sci. Eng. 2021; 8:353-355.
32. Iezzoni A, Weebadde C, Luby JJ, Yue C, van de Weg E, Fazio G, et al. RosBREED: Enabling marker-assisted breeding in Rosaceae. Acta Hortic. 2009; 859: 389-394.
33. Iezzoni A, Weebadde C, Peace C, Main D, Bassil NV, Coe M, et al., editors. Where are we now as we merge genomics into plant breeding and what are our limitations? Acta Hortic. 2016; 1117:1-5.

34. Peace C, Luby JJ, van de Weg WE, Bink MCAM, Iezzoni AF. A strategy for developing representative germplasm sets for systematic QTL validation, demonstrated for apple, peach, and sweet cherry. Tree Genet. Genomes. 2014; 10:1679-1694.
35. Guan Y, Peace C, Rudell D, Verma S, Evans K. QTLs detected for individual sugars and soluble solids content in apple. Mol. Breeding. 2015; 35:135.
36. Fresnedo-Ramírez J, Bink MCAM, van de Weg E, Famula TR, Crisosto CH, Frett TJ, et al. QTL mapping of pomological traits in peach and related species breeding germplasm. Mol. Breeding. 2015; 35:166.
37. Cai L, Voorrips RE, van de Weg E, Peace C, Iezzoni A. Genetic structure of a QTL hotspot on chromosome 2 in sweet cherry indicates positive selection for favorable haplotypes. Mol. Breeding. 2017; 37:85.
38. Howard NP, van de Weg E, Tillman J, Tong CBS, Silverstein KAT, Luby JJ. Two QTL characterized for soft scald and soggy breakdown in apple (*Malus* × *domestica*) through pedigree-based analysis of a large population of interconnected families. Tree Genet. Genomes. 2018; 14:2.
39. Chagné D, Vanderzande S, Kirk C, Profitt N, Weskett R, Gardiner SE, et al. Validation of SNP markers for fruit quality and disease resistance loci in apple. Hortic. Res. 2019; 6:30.
40. Verma S, Evans K, Guan Y, Luby JJ, Rosyara UR, Howard NP, et al. Two large-effect QTLs, *Ma* and *Ma3*, determine genetic potential for acidity in apple fruit: breeding insights from a multi-family study. Tree Genet. Genomes. 2019; 15:18.
41. Luo F, Norelli JL, Howard NP, Wisniewski M, Flachowsky H, Hanke M-V, et al. Introgressing blue mold resistance into elite apple germplasm by rapid cycle breeding and foreground and background DNA-informed selection. Tree Genet. Genomes 2020; 16:28.
42. Rawandoozi ZJ, Hartmann TP, Carpenedo S, Gasic K, da Silva Linge C, Cai L, et al. Identification and characterization of QTLs for fruit quality traits in peach through a multi-family approach. BMC Genomics. 2020; 21:522.
43. Rymenants M, Weg E, Auwerkerken A, Wit I, Czech A, Nijland B, et al. Detection of QTL for apple fruit acidity and sweetness using sensorial evaluation in multiple pedigreed full-sib families. Tree Genet. Genomes. 2020; 16:71.
44. Crump WW, Peace C, Zhang Z, McCord P. Detection of breeding-relevant fruit cracking and fruit firmness QTLs in sweet cherry via pedigree-based and genome-wide association approaches. Front. Plant Sci. 2021; 13:823250.
45. Kostick SA, Teh SL, Norelli JL, Vanderzande S, Peace C, Evans KM. Fire blight QTL analysis in a multi-family apple population identifies a reduced-susceptibility allele in 'Honeycrisp'. Hortic. Res. 2021; 8:28.
46. Vanderzande S, Howard NP, Cai L, Da Silva Linge C, Antanaviciute L, Bink MCAM, et al. High-quality, genome-wide SNP genotypic data for pedigreed germplasm of the diploid outbreeding species apple, peach, and sweet cherry through a common workflow. PLoS One. 2019; 14(6):e0210928.
47. Oraguzie NC, Watkins CS, Chavoshi MS, Peace C. Emergence of the Pacific Northwest sweet cherry breeding program. Acta Hortic. 2017; 1161:73-78.
48. Hemmat M, Weeden N, Brown S. Mapping and Evaluation of *Malus* × *domestica* microsatellites in apple and pear. J. Amer. Soc. Hortic. Sci. 2003; 128.
49. Evans KM, Fernández- Fernández F, Laurens F, Feugey L, van de Weg WE (2007) Harmonizing fingerprinting protocols to allow comparisons between germplasm collections. Eucarpia. XII Fruit Selection Symposium. Zaragoza, Spain, pp.57–58.
50. Bai Y, Dougherty L, Li M, Fazio G, Cheng L, Xu K. A natural mutation-led truncation in one of the two aluminum-activated malate transporter-like genes at the *Ma* locus is associated with low fruit acidity in apple. Mol. Genet. Genomics. 2012; 287:663-678.

51. Ma B, Liao L, Zheng H, Chen J, Wu B, Ogutu C, et al. Genes encoding aluminum-activated malate transporter II and their association with fruit acidity in apple. Plant Genome. 2015; 8(3):1-14.
52. Jung S, Lee T, Cheng CH, Buble K, Zheng P, Yu J, et al. 15 years of GDR: New data and functionality in the Genome Database for Rosaceae. Nucleic Acids Res. 2019; 47(D1):D1137-D1145.
53. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J. Mol. Biol. 1990; 215:403-410.
54. Daccord N, Celton JM, Linsmith G, Becker C, Choisne N, Schijlen E, et al. High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. Nat. Genet. 2017; 49:1099-1106.
55. Shirasawa K, Isuzugawa K, Ikenaga M, Saito y, Yamamot T, Hirakawa H, et al. The genome sequence of sweet cherry (*Prunus avium*) for use in genomics-assisted breeding. DNA Res. 2017; 24: 499-508.
56. Edge-Garza DA, Rowland TV, Haendiges S, Peace C. A high-throughput and cost-efficient DNA extraction protocol for the tree fruit crops of apple, sweet cherry, and peach relying on silica beads during tissue sampling. Mol. Breeding. 2014; 34:2225-2228.
57. This P, Jung A, Boccacci P, Borreng J, Botta R, Costantini L, et al. Development of a standard set of microsatellite reference alleles for identification of grape cultivars. Theor. Appl. Genet. 2004; 109: 1448-1458.
58. Cabe PR, Baumgarten A, Onan K, Luby JJ, Bedford D. Using microsatellite analysis to verify breeding records: A study of ‘Honeycrisp’ and other cold-hardy apple cultivars. J. Amer. Soc. Hortic. Sci. 2005; 40:15-17.
59. Ordidge M, Litthauer S, Venison E, Blouin-Delma M, Fernandez-Fernandez F, Höfer, M, et al. Towards a joint international database: Alignment of SSR marker data for European collections of cherry germplasm. Plants. 2021; 10:1243.
60. van de Weg E, Voorrips R, Finkers HJ, Kodde LP, Jansen J, Bink M. Pedigree genotyping: A new pedigree-based approach of QTL identification and allele mining. Acta Hortic. 2004; 663.
61. Muranty H. Denancé C, Feugey L, Crépin JL, Barbier Y, Tartarini S, et al. Using whole-genome SNP data to reconstruct a large multi-generation pedigree in apple germplasm. BMC Plant Biol. 2020; 20:2.

Supplementary information captions

S1 Tables: **SNPs included in the study.** Details are provided on each SNP's name, NCBI dbSNP accession identifier, linkage group and genetic position, haploblock, and chromosome and physical position. For apple, details were extracted from Vanderzande et al. (2019)*. For sweet cherry, SNP name and identifier and physical chromosome and position were extracted from Vanderzande et al. (2020); genetic position and haploblock were extracted from Vanderzande et al. (2019)*. *Dataset available at https://www.rosaceae.org/publication_datasets, accession number tfGDR1038.

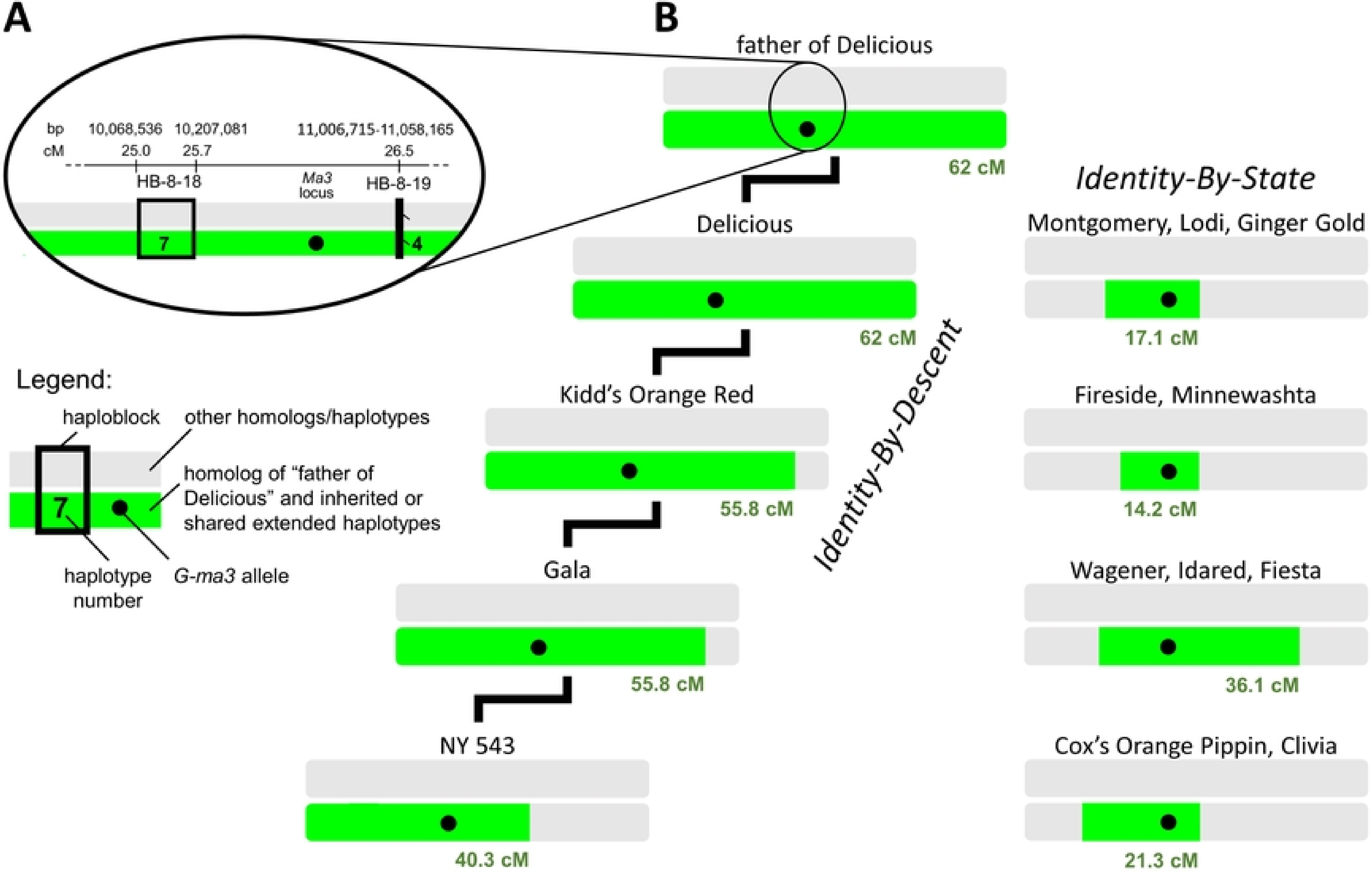
S2 Tables: All alleles deduced for four loci utilized in this study for apple and cherry.

S3 Table: Alleles deduced for the R_f locus for sweet cherry cultivars and selections by haplotype sharing. Sharing via IBD is shown in **bold**, otherwise sharing was via IBS. Individuals annotated with an asterisk (*) are ancestral sources of alleles. For H13, 'Windsor' and 'Venus' shared the same extended haplotypes with 'Blackheart' and 'PMR-1', so they are listed under H13 for both ancestral sources.

S4 Table: Alleles deduced for the GD12 locus for various apple cultivars and selections by haplotype sharing. Sharing via IBD is shown in **bold**, otherwise sharing was via IBS. Individuals annotated with an asterisk (*) are ancestral sources of alleles.

S5 Tables: Haplotype comparisons for three cultivars with alleles deduced for the SSR GD12 that did not match reported genotypes on GRIN-Global. As evidence of correct deduction, extended haplotype patterns are shown for the cultivars and their parents, some siblings, and some offspring. Extended haplotype patterns are color-coded by ancestral source.

S6 Tables: Display of extended haplotypes of ancestral sources needed to differentiate among all functional alleles. Flanking haplotypes with the same background shades have the same pattern. Extended haplotypes in black font were necessary for differentiation among functional alleles, those in blue font were necessary for differentiation among ancestral sources, those in green font represent the locus and its immediately flanking haplotypes, and the rest are in gray font.


A
bp
10,068,536
10,207,081
11,006,715-11,058,165
cM
25.0
25.7
26.5
HB-8-18
Ma3 locus
HB-8-19
7
4
Legend:
haploblock
other homologs/haplotypes
homolog of "father of Delicious" and inherited or shared extended haplotypes
7
haplotype number
G-ma3 allele
B
father of Delicious
62 cM
Delicious
62 cM
Kidd's Orange Red
55.8 cM
Gala
55.8 cM
NY 543
40.3 cM
Identity-By-Descent
Identity-By-State
Montgomery, Lodi, Ginger Gold
17.1 cM
Fireside, Minnewashta
14.2 cM
Wagener, Idared, Fiesta
36.1 cM
Cox's Orange Pippin, Clivia
21.3 cM

Figure