Identification and Characterization of Important Quantitative Trait Loci for Soluble Solids and Titratable Acidity for Germplasm in the University of Minnesota Apple Breeding Program

A Thesis
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Baylee Ann Miller

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

James J. Luby

December 2020

Acknowledgements

I would like to acknowledge the large group of individuals that provided the many years of data collection, moral support, and teaching throughout the duration of this project. Much of the data that was utilized in this project was collected years before I arrived, and for that I am grateful as it provided an opportunity to carry out a research project unlike those previously for the UMN apple breeding program. The consistent help, determination, and willingness of all involved is immensely appreciated.

I greatly appreciate the ongoing efforts to neatly maintain the apple orchard performed by the apple crew at the University of Minnesota Horticultural Research Center, including Winford McIntosh, and many of whom I have had the pleasure of working with since 2016 and into the future.

My colleagues and peers at the University of Minnesota, specifically fellow graduate students in the Applied Plant Sciences graduate program, have been a constant source of help, support, and laughter. My lab mates in the fruit breeding group, namely Elizabeth Blissett and Seth Wannemuehler, have provided countless hours of assistance and motivation, especially during the busy harvest seasons. My coworker and supervisor, John Tillman, has provided an immense amount of technical teaching, support, and encouragement throughout my undergraduate and graduate education as well as into the beginning of my career. The genetic analysis portion of this project would not have been possible without the help, teaching, and assistance that Nicholas Howard has provided to others and myself. There have been dozens of members in the fruit group prior to my arrival, and without their efforts, my project would not have been possible. This group has consistently encouraged and supported me in my academic and scientific growth and knowledge, and for that, I am thankful.

This project was partially funded by the USDA's National Institute of Food and Agriculture – Specialty Crop Research Initiative project, “RosBREED: Enabling marker-assisted breeding in Rosaceae” (2009-51181-05808). Additional funding was provided by the royalty funds from the patented releases of cultivars. I have received additional financial support from the Daniels Fellowship in Horticulture and travel grants to attend the RosBREED annual meeting in 2018 and the Plant and Animal Genomics conference in 2019.

My education experience at the University of Minnesota has been one of great opportunities and never-ending learning, and would not have been possible without my advisor, Dr James Luby. I had the pleasure of working in the fruit breeding group as an undergraduate student, which embraced me and inspired me to continue my education and ultimately attain a full-time researcher position with the same group. My time with this group has been nothing but wonderful, and for that I am incredibly thankful.

The members of my committee, Dr Matthew Clark and Dr James Bradeen, have been greatly supportive of me and this research project, and have made the experience seamless.

Dedication

This thesis is dedicated to my mother, Julee Miller, who has inspired, encouraged, and supported me in my academic endeavors. Without her, I would not have discovered this path and taken this journey that has filled my life with curiosity, wonderful friends, and a lifelong passion for science.

Abstract

Apple fruit acidity and sweetness are two of the major trait components involved in apple seedling sensory evaluation. Published studies have alluded to some of the genetic components of apple fruit acidity and sweetness, but few have included an array of germplasm relevant to the University of Minnesota apple breeding program. With the release and subsequent frequent use of ‘Honeycrisp’ and ‘Minneiska’ apple cultivars as parents at the University of Minnesota, the germplasm set deviates from other breeding programs. In order to increase breeding efficiency and increase overall quality of apple seedlings, this study describes the genetic components of apple fruit acidity and sweetness and provides breeding insights to negate the creation of undesirable apple seedlings. This study uses data from 2010 to 2018 to characterize a wide but relevant array of germplasm, using six major families, three of which have ‘Honeycrisp’ as a parent, and three of which have ‘Minneiska’ as a parent. Three major loci associated with variation in titratable acidity content on linkage groups 1, 8, and 16, and two loci associated with variation in soluble solids content on linkage groups 1 and 13 were identified, and haplotypes were characterized for each locus. The conclusions from this study provide insights for designing crosses that create seedlings with desirable ranges of acidity and sweetness characteristics.

Table of Contents

List of Figures

List of Tables

CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW

Introduction

The domesticated apple (*Malus x domestica*) is a commonly consumed fruit product in the United States, with domestic fresh fruit utilization reaching nearly seven billion pounds in 2018 (USDA 2019). The United States Department of Agriculture (USDA) released a global apple production report in 2018 (USDA 2020) that indicates that China produces nearly half of the world's supply of apples while the next largest producer, the European Union, produced roughly twenty percent of the world's supply. The total apple production globally in 2018 was 68.6 million metric tons (USDA 2020). U.S. apple production in 2010 encompassed more than twenty unique cultivars (USDA 2012), indicating high diversity in the marketplace and the continued need for specialization for distinct commercial uses. Cultivars have desirable characteristics for fresh-eating, cooking, cider production, or processing, with each cultivar filling a role within its respective market. Many cultivars already fill many needs across most markets but breeders continue to develop novel cultivars. Apples are one of a few produce items that specifically brand and advertise different cultivars, encouraging apple breeders to capture consumers by creating cultivars with unique characteristics that satisfy consumer needs.

Apple breeding is a long-term investment, as trees do not produce fruit until sexual maturity, which often takes three to ten years when trees are grown on their own roots. However, sexual maturity can be reached much quicker with cultural practices such as grafting, pruning, and bark ringing (Janick and Moore 1996). Once a tree reaches reproductive maturity, the fruit may be evaluated to determine if the individual can

provide an enhanced product or a product that carries novel traits. The timeline from an initial cross to the release of a new cultivar commonly takes 15-30 years.

Current Public U.S. Apple Breeding Programs

Three university-based apple breeding programs are active in the United States as of 2020. These include Cornell University (CU), the University of Minnesota (UMN), and Washington State University (WSU). The newest program is led by Dr. Kate Evans at WSU. It began in 1994 and has since released two cultivars, both exclusive to Washington state growers. 'WA2' was released in 2009 and its fruit are sold under the trademark 'Sunrise Magic'. More recently, 'WA38' was released in 2017 and its fruit are sold under the trademark 'Cosmic Crisp' (Wilhite 2014). The program at CU started in the late 1890's and is currently led by Dr. Susan Brown. This program has released 66 cultivars with the two most recent being 'NY1' and 'NY2'. Their fruit are sold under the trademarks 'SnapDragon' and 'RubyFrost', respectively (Garris 2013). The oldest apple breeding program is at the UMN, dating back to 1878. Almost 30 cultivars have been released by the UMN apple breeding program, but the most notable release is 'Honeycrisp'. The release of 'Honeycrisp' led to an increased consumer interest in apple fruit texture, juiciness, and flavor, as many cultivars before it lacked these qualities. Many apple cultivars have been developed by the UMN program and programs at WSU and CU using 'Honeycrisp' as a parent, including 'WA38' ('Enterprise' x 'Honeycrisp'), 'NY1' (unnamed New York selection x 'Honeycrisp'), 'Minneiska' ('Honeycrisp' x 'Minnewashta'), and 'MN55' ('Honeycrisp' x 'AA44').

Until recently, most breeding efforts have focused on appearance and storage ability of fruit, which are important qualities for marketability. In recent decades, consumers have become more interested in produce variety and diversity, and consumers are willing to pay more for a better tasting apple when provided with information about the cultivar (Gallardo et al. 2018). Many current breeding targets include combining several high quality traits into one cultivar in order to improve the fresh-eating experience of apple consumption. These traits include, but are not limited to, astringency, flavor, acidity, sweetness, and storage ability. Breeders continue to hold these traits to high standards to ensure a pleasurable eating experience. In addition, new cultivars must have desirable horticultural characteristics including disease resistance.

Importance of Sugar and Acid for Consumers

Two important traits for eating quality of apples are the perception of sweetness and acidity. Harker et al. (2002) found that sugar and acid both contribute to sensory sweetness and overall consumer acceptance, and that the perception of acidity is highly correlated with measured titratable acidity (TA) content. Sweetness is related to the composition and quantity of sugars in fruit. Sugar content is commonly estimated using a refractometer to estimate the percent of soluble solids in a juice sample. Most soluble solids in apple are sugars, so soluble solids content (SSC), measured as °Brix, approximates the percent sugar content. Individual components of sugar in a solution are often determined using High Performance Liquid Chromatography (Guan et al. 2015). However, studies have suggested that sugar content and quantity does not always directly relate to sweetness that a consumer perceives (Harker et al. 2002). Acidity contributes to

perception of both sourness and sweetness. Acidity is commonly estimated indirectly by measuring pH (probe or paper) or directly by titration. The main acid contributing to perceived apple fruit acidity is malic acid (Harker et al. 2002). Malic acid, which is diprotic, constitutes 90% of total acids in apple (Yamaki 1984) making TA content a more informative measurement than pH alone for estimating fruit acidity (Harker et al. 2002).

Benchmark measurements for sugar and acid content have been estimated in order for a cultivar to be considered "edible". Visser and Verhaegh (1978) suggest that apples with a pH between 3.1 and 3.3 and with a sugar content of 15 °Brix or greater possess characteristics of a desirable fresh eating apple.

In order to gauge consumer acceptance of fruit cultivars, a study by Jayasena et al. (2008) attempted to determine an optimal ratio of sugar to acid. They evaluated consumer acceptance of 'Crimson Seedless' table grapes at different °Brix/percent acid ratios (different ratios were obtained by using fruit at different ripening stages) and found significant differences between different ratios and their rating of acceptance. The findings from this study suggest that both sugar and acid are crucial to consumer acceptance of grapes and other fruit products, opposed to solely sugar or acid.

Some apple cultivars are known to taste especially sour (tart), which is a result of a high acid and/or low sugar content. 'Granny Smith' is a commonly available example of a tart cultivar. Conversely, some apple cultivars such as 'Delicious', 'Fuji', and 'McIntosh' are characterized as sweet, due to their low acid and/or high sugar content. People commonly have a favorite cultivar, as some people find certain characteristics interesting and appealing while others may dislike those same features, which suggests

that consumer acceptance of apples is a function of multiple traits. If breeders can understand the genetic contributions to variation for these different traits, they can predict crosses that may produce apples that are favored by consumers.

Biological Development of Acid Content

Perceived acidity in apple fruit is primarily due to the presence or absence of malate within the vacuole of fruit cells (Ulrich 1970). A study performed by Beruter (2004) found evidence that malate content in apple fruit is determined by the accumulation of malate in the vacuole, and is not heavily dependent on the rate at which malate is synthesized. Beruter (2014) found that content of sorbitol, a precursor for malate synthesis, has a very similar content in high acid and low acid fruit, further suggesting that malate content and fruit acidity is not determined solely by malate production and instead by malate accumulation in the vacuole.

A candidate gene has been identified for a malic acid transporter, and activity of the corresponding gene product has been found to associate with variation in acidity content. This gene on chromosome 16 was identified by Khan et al. (2013) and is responsible for transporting malate from the cytosol into the vacuole.

Jia et al. (2018) investigated previously unidentified genes associated with variation in acidity content in a different region of the apple genome. A ‘Jonathan’ x ‘Golden Delicious’ family and ‘Zisai Pearl’ x ‘Red Fuji’ family were used to identify candidate genes for acidity. They identified 61, 157, and 130 genes within three genetically defined regions on chromosome 8 that control variation in this trait. Variation in a promoter region controlling a vacuolar H^+ ATPase (which is responsible for malate

accumulation in the vacuole) was identified and correlated with high and low acid phenotypes. Jia et al. (2018) identified a candidate gene that was consistent with previous research suggesting a malate transporter gene as being responsible for variation in apple fruit acidity content. Jia et al. (2018) provides additional biological explanations for variation in malate content (acidity) and identified several single nucleotide polymorphism (SNP) markers that may be used in breeding practices.

Biological Development of Sugar Content

The accumulation of sugar in developing apple fruits has been well studied. Candidate genes have been identified for mechanisms involved in sugar transportation, as well as enzymes that are involved in the metabolism of sugar (Li et al. 2012).

Beruter (1985) noted that sucrose was undetectable during very early development of apple fruit. They found that concentrations of several forms of sugar, including sorbitol, glucose, and fructose, begin at high levels but decrease in early fruit development up until “June drop”, the time when apple trees shed excess fruit and subsequently put more resources into the apples that are retained after dropping.

Sugar accumulation in maturing fruit is due to the conversion of starch into sugar. Starch is accumulated in mature fruit from carbohydrates that are produced by photosynthesis in plant leaves and then transported mainly as sorbitol to the fruit during development. Sorbitol content is highest in early stages of fruit development, decreases during development, and rises again when nearing fruit maturity (Sturm and Stampar 1999).

Fructose is the most abundant sugar in mature apple fruit, as it has concentrations twice as high as sucrose, sorbitol and glucose (Zhen et al. 2018).

Interaction of Sugar and Acid during Fruit Development and Maturation

Sugar and acid content of apple fruit are dynamic during the development and maturation process. During fruit development on the tree, acid content increases until fruit maturity and sugar content increases during fruit development and continues through maturation. After fruit maturation and harvest, acid content decreases and sugar content increases. The decrease in starch content and simultaneous increase of sugar content during development and maturation indicates fruit maturity. This can be visualized by spraying a potassium-iodine solution on a fruit that has been cut in half, exposing the flesh (Blanpied and Silsby, 1992). Starch reacts with the potassium-iodine solution resulting in black flesh color. This assay is quantitative and as fruit develop and mature, they show less black staining, indicating the conversion of starch to sugar. Etienne et al. (2013) proposed that high sugar content in mature apple fruit might increase the conversion of malate to citrate via glycolysis. However, contradictory findings have been reported by Coombe et al. (1976), which suggest that sugar may not be available for respiration due to sugar being stored in vacuoles, therefore causing the fuel for respiration to shift towards acids, resulting in lowering acidity levels as fruit mature during storage.

Apple Genetics and Genomics

The development of new apple cultivars has been enhanced by modern technology and new breeding methods that have been developed over the last few

decades. The use of DNA sequence information for apple breeding became a reality when the genome of 'Golden Delicious' (*Malus x domestica*) was sequenced and published by Velasco et al. (2010). The 'Golden Delicious' genome was sequenced using a whole-genome shotgun approach. Seventeen linkage groups (LGs) were assembled, corresponding to the seventeen chromosomes in apple, from the utilization of 1,643 DNA markers. The genome contains over 57,000 putative genes, over 11,000 of which were not present in plant genomes published at that time. The data also suggest a recent genome wide duplication event that occurred in an ancestor with nine chromosomes, followed by many chromosomal translocations, deletions, and the loss of one entire chromosome, resulting in the seventeen chromosomes observed in most modern diploid cultivars. The 'Golden Delicious' genome has been sequenced further since 2010, and the *Malus* x *domestica* GDDH13 v1.1 genome ('Golden Delicious' doubled haploid) is the most recently reported apple whole-genome sequence by Daccord et al. (2017). In addition to resequencing an apple genome, they created a high-density linkage map, haplotype map, and a map of DNA methylation. These genome sequences provide a framework from which genetic markers can be mapped and referenced across different cultivars. With the use of DNA markers, quantitative trait loci (QTLs) can be identified. QTLs are regions in the genome of a specific set of germplasm that are associated with variation for an observed multi-genic trait. After the identification of QTLs, causal gene(s) may be investigated to unveil the biological nature of the trait of interest.

Apple is a highly heterozygous, outcrossing species. Due to these properties, progeny produced from crosses are often highly diverse. In apple breeding programs, individuals are seldom replicated for evaluation until the later stages of testing, due to the

large plant size, long period to fruit evaluation, and labor-intensive nature of apple growing. Controlling the outcomes of crosses using selection based on DNA markers can help enrich the breeding pool with desirable individuals and produce phenotypes that are more predictable and desirable.

RosBREED Projects

The RosBREED Initiative (hereafter referred to as the "RosBREED1 project") was started in September 2009 with funding from the USDA Specialty Crops Research Initiative, in anticipation of the sequencing of several rosaceous fruit crop genomes, including peach in 2013 (Verde et al. 2013), apple in 2010 (Velasco et al. 2010), and strawberry in 2011 (Shulaev et al. 2011). The goal was to establish collaboration between rosaceous crop breeding programs in the United States to utilize newly available genetic information for DNA-informed breeding (Iezzoni et al. 2010). For apple, the RosBREED1 project led to the creation of the International RosBREED SNP Consortium 8 K Illumina Infinium array v1 (hereafter referred to as the "8K SNP array"), a single nucleotide polymorphism (SNP) marker array with approximately 8,000 markers from apple (Chagne et al. 2012a) and the creation of other Simple Sequence Repeat (SSR) markers and Diversity Array Technology (DART) markers (Iezzoni et al. 2010). The 8K SNP array was created by sequencing 27 apple cultivars that have been frequently used in breeding programs and aligning their sequences to the previously sequenced 'Golden Delicious' genome (Chagne et al. 2012a). Over two million SNPs were identified across the genomes of the selected cultivars and the SNPs were thinned to a set of 7,867 after stringent filtering criteria, with 5,554 of those markers being polymorphic across the

different cultivars that were used to develop the array. The creation of the 8K SNP array resulted in a publicly available array that has been used by many breeding programs across a wide variety of apple germplasm. The practical use of this array serves as a framework for QTL identification, QTL analysis, genomic selection, and creating marker-assisted selection protocols for mapped traits. The RosBREED1 project also defined protocols for harvesting apple fruit and phenotyping several traits in apple including soluble solids and acid content (Evans et al. 2012).

The RosBREED1 project was followed by initiation of a new five year grant titled "RosBREED: Combining disease resistance and horticultural quality in new rosaceous cultivars", hereafter referred to as the "RosBREED2 project" (Iezzoni et al. 2017). Apple continued to be part of this project, as integration of novel disease resistance and improvement of fruit quality traits were determined to be important and worthy of continued collaboration.

The funding of the RosBREED2 project supported the creation of several genetic markers for important traits in apple including resistance to apple scab (*Venturia inaequalis*) and fire blight (*Erwinia amylovora*), fruit acidity, and fruit skin color (Evans and Peace 2017). These traits are important breeding targets for the UMN apple breeding program. The types of genetic markers commonly used include SNPs, SSRs, and insertion-deletion (Indel) markers. Many of these markers have been developed by U.S. breeding programs, including WSU and CU (Evans and Peace 2017).

Mapping QTLs in Apple

Genetic Linkage Maps

In order to map QTLs to a genome, a linkage map of markers must be constructed. Early linkage maps for apple varied in the number of markers, types of markers, and number of LGs. Conner et al. (1997) created three linkage maps from two families and an integration of one of the shared parents of the two families. These linkage maps were created using 110 to 238 randomly amplified polymorphic DNA (RAPD) markers mapped to 16 to 19 LGs.

Liebhard et al. (2003a) constructed a dense linkage map for apple using a family of seedlings from a 'Fiesta' x 'Discovery' cross. The markers included SSRs, Sequence Characterized Amplified Regions (SCARs), RAPDs, and amplified fragment length polymorphism (AFLP) markers, with 840 in total. These markers were used to create a map for both parents, which each contained 17 LGs. Four maps (Hemmat et al. 1994, Conner et al. 1997, Gianfranceschi et al. 1998, Maliepaard et al. 1998) were published prior to the map of Liebhard et al. (2003a), but only two of them contained the expected 17 LGs. The map developed by Liebhard et al. (2003a) had greatly increased marker density and genome coverage, allowing easy application of the map to other populations. Additional linkage maps have been created for apple, and maps are constantly being updated with the knowledge from newly genotyped individuals and from new SNP arrays. Additionally, research groups with unique germplasm often generate linkage maps that are tailored specifically to their breeding program. One such high quality genetic map that has been used for germplasm at the UMN is described by Howard et al. (2017) and was used to determine the previously unknown parents of 'Honeycrisp'.

FlexQTL™ Mapping Software

With the creation of linkage maps containing plentiful markers on all 17 chromosomes, QTLs for many traits have been mapped to the apple genome using a variety of mapping software. An effective method of QTL mapping for apple has been developed for utilization of pedigree-connected populations of related plants using FlexQTL™ software (Bink et al. 2002, Bink et al. 2012).

FlexQTL™ is specifically suited to mapping QTL in apple by using related families that exist as a part of a breeding program (Bink et al. 2014). Like other mapping software, FlexQTL™ uses mapped genetic markers and phenotype information to accurately map traits to the target genome (Chagne et al. 2012b). One advantage of utilizing FlexQTL™ for mapping traits to the apple genome is that the pool of individuals used for mapping can include multiple, related families. Peace et al. (2014) suggest that one can add more individuals from different families (including those from other genetic backgrounds and geographic locations) to the analysis to improve QTL validation power. The benefits of using multiple populations have been exhibited by Mora et al. (2017). They used 1467 individuals from 18 families across five peach breeding programs. This collective set of individuals improved the analysis by creating a large set of individuals with more variability than one family could provide by itself, and by increasing the power of the analysis for detecting QTL. They determined that up to 47% of the QTL discovered were from using families derived from breeding germplasm rather than families derived from commercial cultivars, indicating that additional QTL exist in the breeding germplasm that are not present in current cultivars. This finding supports the idea that breeding programs benefit from the integration of novel alleles in the creation of new cultivars, as novel alleles may be able to complement the alleles present in current

released cultivars. Several studies (Bink et al. 2014, Mora et al. 2017, Mangandi et al. 2017) demonstrate the utility of FlexQTL^{TM} for obligate outcrossing species as it is capable of performing analyses with multiple families. The software benefits from diverse germplasm, especially when family sizes may too be small to provide power for a single bi-parental analysis. Additionally, with perennial plants such as apples, families can be evaluated for QTL in multiple years, which may increase confidence in findings and help shed light on environmental effects on traits.

QTL Identified in Apple for Acid Content

The inheritance of sugar and acid content has been extensively researched in apple. Many QTL have been reported for sugar and acid content (Table 1.1) and several have been integrated into breeding practices. Nybom (1959) was the first to postulate and provide evidence for the *Ma* (malic acid) locus, which characterizes a major effect on fruit acidity via recessive inheritance of low fruit acidity. Nearly all of the cultivars Nybom (1959) observed appeared to be heterozygous at this locus, suggesting that the heterozygous state may confer a selective advantage. Nybom hypothesized that many genes may be involved in apple fruit acidity and these genes may lie in other regions of the genome.

Maliepaard et al. (1998) used a family of 152 individuals from a 'Prima' x 'Fiesta' cross and several types of genetic markers to construct a genetic linkage map with 17 LGs for each parent in order to map the *Ma* locus that Nybom (1959) identified. Acidity was measured using pH paper and bromocresol to determine the relative pH of fruit from each individual. The individuals were grouped based on acidity, with a pH of

3.8 being the threshold separating the high and low acidity groups. Variation in acidity was mapped to LG16 and was named the *Ma* locus, assuming that the predominant acid in apple fruit is malic acid and that it was likely the locus hypothesized by Nybom (1959). They determined that the recessive homozygous state of this locus results in pH higher than 3.8 and that both parents in this study are heterozygous, supporting the findings from Nybom (1959).

Liebhard et al. (2003b) identified the same *Ma* locus that Maliepaard et al. (1998) reported, and also identified a QTL on LG8 for fruit acidity using individuals from a 'Fiesta' x 'Discovery' cross. These two loci accounted for 42% and 46% of the trait variance, respectively, and appeared to act additively. Tightly linked markers for each locus were identified. Several other traits such as large fruit size, earlier ripening, and higher sugar content appeared to be positively correlated with higher acidity in this study.

The *Ma* locus was further investigated by Xu et al. (2012) with two half-sib families and based on variation for pH and TA among 438 individuals. They found that the *Ma* locus exhibits an additive effect, with the *Ma* allele conferring high acidity and the *ma* allele conferring low acidity. This locus was fine-mapped and determined to fall between two SSR markers, a region that spans approximately 150 kbp with reference to the 'Golden Delicious' genome. Xu et al. (2012) also identified two minor QTL related to apple fruit acidity, naming the locus on LG6 *M2* and the locus on LG1 *M*3. This study indicates difficulty with determining the segregating classes of acidity in individuals due to the appearance of a dominance effect of the LG16 locus when using pH or TA to characterize the *Ma* locus. However, segregating classes in this study enabled estimation of the additive effect of one copy of the *Ma* allele at the LG16 to be -0.27 pH units and

+1.94 g/L for TA. Additionally, this study utilized data on consumer acceptance related to pH level and suggested that fruit from individuals of the *MaMa* genotype are likely too acidic for dessert apple consumption.

A study evaluating the effects of the *Ma* locus was also performed by Iwanami et al. (2012). They suggested that the *Ma* locus has a dominant allele, with a very small difference in TA between individuals that carried two copies of the dominant allele and heterozygous individuals. However, their analysis suggests a polygene model, which contradicts the findings and conclusions from Liebhard et al. (2003b) and Xu et al. (2012).

Validation for the LG16 QTL has been performed and a candidate gene for this locus has been proposed by Khan et al. (2013). They suggest that a gene controlling malic acid transport from the cytosol into the vacuole by regulating production of malate transporter proteins is the basis for the *Ma* locus. They observed a segregating population of individuals from a 'Prima' x 'Fiesta' cross and characterized the segregating phenotypes based upon the individuals' genotypes at the *Ma* locus. The pH of fruit was measured at three developmental stages and confirmed that acidity in apple fruit decreases during the maturation process. They also observed that the sourness (due to high acidity) of fruit appeared to be caused by an overall increased content malic acid content rather than a slower breakdown of malic acid during maturation. They identified 27 potential malic acid transporters, indicating quantitative control over acid content, and confirmed the hypothesis proposed by Maliepaard et al. (1998) that malic acid is the predominant acid affecting the pH of apple fruit. The findings of this study are consistent with those of Xu et al. (2012) in different populations. Khan et al. (2013) identified

additional potential causal genes that may influence pH, located on LG6, LG13 and LG14.

Other loci associated with fruit TA have been associated with SSR markers on LG1, LG6, and LG7 by Liu et al. (2016) in a single family. Four loci were detected on LG1 spanning from 56.2 cM to 80.0 cM. A limitation of this study is sparse marker density with only 64 polymorphic SSR markers across the 17 LGs.

A study performed by Jia et al. (2018) identified four major QTLs for apple fruit acidity. Three QTLs on LG8 were detected in addition to the *Ma* locus on LG16. The three QTLs on LG8 were investigated for SNPs that correlated with variation for TA. Individuals were characterized based on their phenotype and SNP states. The three QTLs mapped to LG8 are very close to each other (12.4 Mbp to 12.8 Mbp, 13.4 Mbp to 13.8 Mbp, and 15.1 Mbp to 16.3 Mbp) and should be further investigated in appropriate crosses to confirm that there are indeed three separate QTLs. Zhang et al. (2012) also mapped fruit acidity to LG8 in a segregating progeny of a 'Jonathan' x 'Golden Delicious'.

The collective QTLs identified in the apple genome for malic acid or fruit acidity lie on LGs 1, 2, 6, 7, 8, 10, 13, 15, 16 and 17, with LG8 and LG16 being the predominant LGs identified in most studies (Table 1.1). The numerous reported QTL support the hypothesis that acidity in apple fruit is a quantitative trait, one that is complex and not completely explained by one or few QTL.

QTL Identified in Apple for Sugar Content

Several QTLs have been mapped for SSC on 12 of the 17 apple LGs based on associations with SNP, SSR and sequence-related amplified polymorphism (SRAP) markers (Kennis et al. 2008, Liu et al. 2016, Guan et al. 2015, Liebhard et al. 2003b).

Several QTLs have been mapped for individual sugars. QTLs for sucrose content have been found on LG1, 3, 4, 9, and 12 (Guan et al. 2015, Sun et al. 2015). QTLs for fructose content have been found on LG1, LG3 and LG15 (Guan et al. 2015, Sun et al. 2015). The QTL on LG1 for fructose content identified by Guan et al. (2015) has been translated into a single marker characterization, which is currently being utilized by the UMN apple breeding program for parent selection.

Zhen et al. (2018) performed a study where *Sugar Will Eventually be Exported Transporters* (SWEET) genes were hypothesized and discovered in apple. These genes encode proteins that are associated with the movement of sugars from the leaves to the fruit during fruit development and maturation. Twenty five genes were identified and mapped to the apple genome, falling on chromosomes 3, 4, 5, 6, 10, 11, 12, 13, 14, 16, and 17 of the 'Golden Delicious' genome. SSR primers were developed for nine of these genes, and three (one on chromosome 4, one on 10, and one on 16 near the *Ma* locus) highly correlated with variation in sugar accumulation in fruit. However, these genes alone explained only a small portion of phenotypic variance in their germplasm (ranging from 0.7% for sucrose to 8.4% for total sugar content).

A summary of QTLs detected for sugar or sugar related components of apple can be found in Table 1.2.

Haplotyping and Characterizing QTL

Once a QTL is mapped, haplotypes spanning QTL can be traced through pedigrees to understand the inheritance of traits and are useful in associating markers that can be used in MAS. This has been performed for evaluation of QTL mapping statistical models (Bink et al. 2002), simulations of pairwise relatedness (Bink et al. 2008), and soft scald and soggy breakdown disorder (Howard et al. 2018). An Indel marker and a SNP marker correlated with TA variation at the *Ma* locus (Chagne et al. 2019, Verma et al. 2019). These markers are currently utilized in parent selection in the UMN apple breeding program. Another locus for fruit acidity has been mapped to LG8 and an SSR marker and a SNP marker have been identified that correlate with variation in TA content conditioned by this locus. However, the TA predictions provided by these markers at the LG8 and LG16 loci are inconsistent across parents in the UMN apple breeding program. For instance, the Indel marker at the LG16 locus for 'Minneiska' predicts two alleles associated with low TA, while the SNP marker at the LG16 locus for 'Minneiska' predicts only one allele associated with low TA. This same discrepancy is observed at the LG8 locus SNP marker and LG8 locus SSR marker. The LG8 locus SNP marker predicts two alleles associated with low TA for 'Minneiska', while the LG8 locus SSR marker predicts one allele associated with low TA. These different predictions for the same trait locus in an individual indicates the importance of validating QTL in unique germplasm sets.

Verma et al. (2019) performed haplotype effect analyses for fruit acidity on the breeding material at WSU. The major QTLs on LG8 (*Ma3*) and LG16 (*Ma*) were identified, and they determined that an additive allele dosage model explains 66% of the observed variation in TA content. Haplotypes from the selected material were identified

and characterized as high TA or low TA. The addition of a high TA haplotype had an average +1.8 g/L effect at harvest. This study provides a prediction for apple fruit acidity based on additive alleles at LG8 and LG16, and such information can be easily translated and used for breeding purposes when using germplasm of a similar background to that used in the study.

The importance of fine mapping QTL and performing QTL validation are shown with the discrepancies seen between different predictive markers for the same trait locus. Individual markers have been validated in independent studies and marker validation is important to consider for each individual program with a unique set of germplasm. Genetic background can have additional effects on traits even when major loci have been identified, as QTLs of smaller effect may be present in some germplasm sets and not in others. In several instances in UMN apple breeding germplasm, two markers for the same trait locus produce conflicting trait value predictions, creating an uncertain phenotype prediction. This may be due to unique haplotypes existing in UMN germplasm, or due to the markers being physically distant from the causal genetic regions. Such discrepancies complicate the breeding process and can lead a breeder to potentially make crosses that produce undesirable progeny due to a lack of marker validation in their set of germplasm. Therefore, many genetic studies may be of limited use for extrapolation to the relevant breeding programs without further validation. While major loci have been identified for components of acidity and sugar, minor loci can also have an impact on trait expression when combined with certain major loci, and are not detected in all sets of germplasm.

Genotype by Environment Effects

When evaluating phenotype data to perform QTL mapping, it is important to consider not only the unique individuals being evaluated, but also the effect that environmental factors have due to location (e.g. climate, topography), year (e.g. severe weather events, accumulation of growing degree days), or management practices on the expression of traits. Biennial bearing nature of apple can contribute to year-to-year variation, as not all seedlings may be evaluated every year.

The effect of year-to-year variation on sugar content was studied by Guan et al. (2015). A sample of 20 fruits were collected from a set of 265 apple seedlings at the WSU apple breeding program for two years. QTLs were detected for various sugar components and they determined that some loci appeared in the same regions for both years, while other loci shifted up to 20 cM on the same linkage group between years. Some loci were unstable between years. This study exemplifies the complexity of quantitative traits and limitations when utilizing a single year of phenotypic data.

Utilization of Genetic Markers

Genetic markers have been developed for various traits and are being integrated into breeding decisions in several ways. Crossing schemes can benefit from genetic markers linked to desired traits as breeders can predict progeny phenotypes based upon the genotypes of the parents using marker-assisted parent selection (MAPS) (Collard and Mackill, 2008). For validated loci, this method only requires the genotyping of important individuals in a program that are used as parents, and they need to be genotyped only once in perennial crops since the individual can be used as a parent for multiple years. MAPS allows breeders to eliminate crossing choices that result in incompatibility or that

are predicted to produce progeny that may have undesirable phenotypes. MAPS can save time and money that would have been spent raising inferior plants that would be subsequently eliminated from the breeding pool.

Another use of genetic markers is for marker-assisted seedling selection (MASS) (Collard and Mackill, 2008). This is the screening of seedlings from a desired population for a trait(s) of interest using linked markers to facilitate selection and culling. Seedlings can be analyzed for genetic markers as early as the first true leaf. This can help reduce time and money spent on raising seedlings by culling undesirable seedlings before they are planted into a field. In contrast with MAPS, MASS can be substantially more expensive, as hundreds of plants may be screened from a single cross, compared to screening the two parents.

While QTLs and genetic markers have been identified and investigated for many traits in apple, some traits are more desirable targets than others for marker-assisted selection. Traits that can be observed only in mature plants, such as fruit quality, are especially good candidates for marker-assisted selection. Apple trees bear fruit after three to eight years of growing in an orchard when grafted onto rootstock (Janick and Moore 1996). Plants vary for the rate of maturation depending on genetics and the environment. Fruit acidity and fruit sugar content can only be observed once a tree bears fruit. By targeting these traits at the parental selection stage or seedling selection stage, a program can reduce the amount of time and money spent raising seedlings that would ultimately be discarded upon fruiting and sensory evaluation. This concept was verified by Wannemuehler et al. (2019) by examining cost efficiency of utilizing MASS markers for simulated apple, peach, and strawberry breeding programs. Additionally, by utilizing

parent selection to produce progeny that lack undesirable traits, the progeny pool will be enriched with desirable traits, aiding the future generations of breeding.

Thesis Objectives

A major hurdle in the UMN apple breeding program involves the elimination of trees that produce fruit with undesirable acid and sugar content. Fruit evaluation takes a considerable amount of time and the evaluator's palate becomes fatigued especially quickly when tasting fruit with high acid levels (Harker et al. 2002). Many loci associated with sugar and acid content have been reported from a wide variety of germplasm, as summarized in Table 1.1 and Table 1.2. A subset of these loci are likely associated with variation in UMN germplasm. WSU researchers have performed QTL identification, analysis, and validation for acid content and sugar content in material that is specific to their breeding program, but those efforts may not accurately predict phenotypes for UMN germplasm (Guan et al. 2015, Verma et al. 2019). Similar research can benefit the UMN apple breeding program and help bridge the gap between reported literature and application to the breeding process. The goal of this study is to improve breeding practices for acid and sugar content in apple fruit by pursuing the following objectives:

1. Identify QTLs contributing to variation for SSC and TA content using individuals in the UMN apple breeding program to confirm previous literature and identify new loci specific and/or important to UMN germplasm.
2. Identify segregating haplotypes at QTLs for SSC and TA in individuals used for QTL analysis.

3. Characterize haplotype effects for SSC and TA for important UMN breeding germplasm and create recommendations for breeding application.

Table 1.1. Mapped QTLs in Apple for Traits Related to Acid Content.

Trait	Genetic Marker Type	Linkage Group	Publication
Citric Acid	SNP	8	Sun et al. 2015
Citric Acid	SNP	15	Sun et al. 2015
Fruit Acidity	"Molecular Markers"	8	Liebhard et al. 2003
Fruit Acidity	"Molecular Markers"	16	Liebhard et al. 2003
Fruit Total Acid	SNP	8	Sun et al. 2015
Malate	SNP	8	Jia et al. 2018
Malate	SNP	8	Jia et al. 2018
Malate	SNP	16	Jia et al. 2018
Malic Acid	SNP	8	Sun et al. 2015
pH	SSR	1	Xu et al. 2012
pH	Dart/SSR	6	Khan et al. 2013
pH	SSR	6	Xu et al. 2012
pH	Dart/SSR	13	Khan et al. 2013
pH	Dart/SSR	14	Khan et al. 2013
pH	Several	16	Maliepaard et al. 1998
Sensory Acidity	SNP	1	Rymenants et al. 2020
Sensory Acidity	SNP	8	Rymenants et al. 2020
Sensory Acidity	SNP	16	Rymenants et al. 2020
Titratable Acid	SSR/SRAP	1	Liu et al. 2016
Titratable Acid	SSR/SRAP	1	Liu et al. 2016
Titratable Acid	SSR/SRAP	1	Liu et al. 2016
Titratable Acid	SSR/SRAP	1	Liu et al. 2016
Titratable Acid	SSR	1	Xu et al. 2012
Titratable Acid	SSR	2	Kenis et al. 2008
Titratable Acid	SSR/SRAP	6	Liu et al. 2016
Titratable Acid	SSR	6	Xu et al. 2012
Titratable Acid	SSR/SRAP	7	Liu et al. 2016
Titratable Acid	SNP	8	Verma et al. 2019
Titratable Acid	SSR	10	Kenis et al. 2008
Titratable Acid	SSR	13	Kenis et al. 2008
Titratable Acid	SSR	15	Kenis et al. 2008
Titratable Acid	SSR	16	Kenis et al. 2008
Titratable Acid	SNP	16	Verma et al. 2019
Titratable Acid	SSR	17	Kenis et al. 2008

Table 1.2. Mapped QTLs in Apple for Traits Related to Sugar Content.

Trait	Genetic Marker Type	Linkage Group	Publication
Fructose	SNP	1	Guan et al. 2015
Fructose	SNP	1	Sun et al. 2015
Fructose	SNP	3	Guan et al. 2015
Fructose	SNP	15	Guan et al. 2015
Glucose	SNP	1	Guan et al. 2015
Glucose	SNP	2	Guan et al. 2015
Glucose	SNP	3	Guan et al. 2015
Glucose	SNP	15	Guan et al. 2015
Glucose	SNP	16	Guan et al. 2015
Soluble Solids	SSR/SRAP	1	Liu et al. 2016
Soluble Solids	SNP	2	Guan et al. 2015
Soluble Solids	SSR	2	Kenis et al. 2008
Soluble Solids	SNP	3	Guan et al. 2015
Soluble Solids	"Molecular Markers"	3	Liebhard et al. 2003
Soluble Solids	"Molecular Markers"	6	Liebhard et al. 2003
Soluble Solids	SSR/SRAP	7	Liu et al. 2016
Soluble Solids	"Molecular Markers"	8	Liebhard et al. 2003
Soluble Solids	"Molecular Markers"	9	Liebhard et al. 2003
Soluble Solids	SSR	10	Kenis et al. 2008
Soluble Solids	SNP	12	Guan et al. 2015
Soluble Solids	SNP	13	Guan et al. 2015
Soluble Solids	"Molecular Markers"	14	Liebhard et al. 2003
Soluble Solids	SNP	15	Guan et al. 2015
Sorbitol	SNP	1	Guan et al. 2015
Sorbitol	SNP	3	Guan et al. 2015
Sorbitol	SNP	5	Guan et al. 2015
Sorbitol	SNP	9	Guan et al. 2015
Sorbitol	SNP	11	Guan et al. 2015
Sorbitol	SNP	13	Guan et al. 2015
Sorbitol	SNP	15	Guan et al. 2015
Sucrose	SNP	1	Sun et al. 2015
Sucrose	SNP	3	Guan et al. 2015
Sucrose	SNP	4	Guan et al. 2015
Sucrose	SNP	9	Guan et al. 2015
Sucrose	SNP	12	Guan et al. 2015
Total Sugar	SSR/SRAP	2	Liu et al. 2016

CHAPTER 2: GENETIC COMPONENTS OF SOLUBLE SOLIDS CONTENT IN IMPORTANT UNIVERSITY OF MINNESOTA APPLE GERMPLASM

Introduction

When an apple breeder is determining if an apple is suitable for the consumer market, many sensory traits are considered. Texture, sweetness, sourness, and flavor are a few of the major components that determine if an apple cultivar will be successful (Harker et al. 2008). Within the University of Minnesota (UMN) apple breeding germplasm, high-quality texture has been achieved and successfully passed on to subsequent generations from the cultivar Honeycrisp. However, not all 'Honeycrisp' descendants have a desirable level of sweetness. For a trait such as sweetness, many genetic components are likely responsible for trait expression and the genetic components have not been well understood.

Sensory sweetness is determined by sugar and acid content. Soluble solids content (SSC) is used to estimate sugar content and is, therefore, an important trait considered during apple cultivar development. The perception of sweetness is important for at least reasons; sweetness is a characteristic desired by consumers (Harker et al. 2002) and is required for the sensory perception of more complex flavor compounds (Chitarrini et al. 2020). Fruit with low sugar content may taste too sour (due to proportionally high acid content) or insipid (due to proportionally low acid content). Fruit with both low sugar content and low acid content will lack flavor complexity, whereas fruit with high levels of both sugar content and acid content will taste overpowering (Rodbotten et al. 2009). Selecting a proper balance of both sugar and acid becomes very important for a breeder,

and even more so when considering how fruit will taste after a period of cold storage, as many apple cultivars are stored for several weeks or months prior to being sold to the final consumer.

Many factors determine the sugar content in apple fruit. The environment provides the sunlight, water, and air necessary for sugar production via photosynthesis. While variations in the environment can result in variation in sugar content (Kondo 1992), the genetic makeup of an apple tree will also influence sugar accumulation. Studies have attempted to elucidate the genetic components contributing to variation for SSC in fruit (Guan et al. 2015, Liu et al. 2016, Kenis et al. 2008, Liebhard et al. 2003a), and have suggested complex, multigenic inheritance. Additionally, the genetic background of the germplasm will have an effect on which (if any) genetic loci are expressed.

Guan et al. (2015) identified five quantitative trait loci (QTLs) responsible for variation in SSC suggesting that that no single gene is the major determinant of SSC in Washington State University (WSU) apple breeding germplasm. This study was perhaps lacking a large sample size, a diverse germplasm set, and replicated data that may have been able to strengthen the QTL mapping results and subsequent analyses. Additionally, due to small sizes of segregating families, haplotype analysis was not performed, limiting the study's utility in enabling meaningful breeding decisions regarding SSC QTLs.

Lastly, the sugar content in fruit changes during a period of cold storage. In the 2019/2020 market year, the United States produced 11 billion pounds of apples (USDA, 2020), most of which underwent a period of cold storage prior to the fruit being sold to the final consumer. Apple fruit is composed of living tissue that continues to respire after

harvest, is climacteric, and therefore the chemical components of apple fruit change over time. While producers hold apples at low temperatures to minimize respiration (Kader et al. 1989), respiration still occurs and often results in loss of mass over long periods of time (Ghafir et al. 2009). Additionally, the acid content in apples often decreases in storage while the sugar content increases (Gorny and Kader 1996). The rates of change for this inverse relationship vary depending on the apple cultivar (Suni et al. 2000). The genetic basis of this relationship has not been well studied and understanding this relationship can help a breeder determine if an individual genotype will have desirable fruit after being stored for an extended period of time, prior to subjecting fruit to an extended period of cold storage.

This study seeks to utilize important germplasm in the UMN apple breeding program to uncover genetic components of SSC in apple fruit. QTL discovery analyses used a wide range of germplasm from 2010 through 2018. SSC was examined in fruit at harvest, 10-weeks post storage, and, in some years, 20-weeks post storage to investigate the inheritance of SSC at harvest, after storage, and the difference between harvest and 10-weeks post storage. Six families comprised of the most relevant contemporary germplasm were used for QTL mapping and haplotype analysis. Two major QTLs were identified and characterized for SSC variation at harvest, 10-weeks post storage, and the difference between harvest and 10-weeks post storage for important germplasm in the UMN apple breeding program.

Materials and Methods

Plant Material

Over the course of eight years of fruit evaluation (2010-2018, excluding 2013), 1095 unique apple cultivars, advanced selections, or seedlings were observed for at least one year at the Horticultural Research Center in Chanhassen, Minnesota (Appendix Table 1A). All trees had fruit examined at the time of harvest, 10-weeks post storage, and, in some cases, 20-weeks post storage, when sufficient fruit were available as described below in the phenotyping protocol. The plant material utilized for this study encompasses many families, including six of major focus, that are part of the active breeding efforts at the UMN apple breeding program. With the exception of one family, all trees were evaluated from unselected, non-replicated seedlings that were budded onto 'Budagovsky 9' or 'Geneva 11' rootstock. One family, 'Honeycrisp' x 'Minnewashta', was exclusively harvested from a randomized, replicated planting budded onto 'Budagovsky 9' rootstock.

The germplasm evaluated over this nine-year period can be broken into several different subsets based on the initial intent and purpose for including the families (Appendix Table 1A). The individuals evaluated during the 2010-2012 period were selected specifically for the RosBREED1 project, which aimed to characterize a set of diverse germplasm within three apple breeding programs in the United States: Cornell University (CU), WSU, and UMN. Germplasm evaluated at the UMN was not evaluated for its major contributions to current breeding efforts, but rather to represent the diversity found within the breeding germplasm, including important ancestors. The individuals evaluated during the 2014-2016 period were selected specifically for the RosBREED2 project, which aimed to identify the genetic components of important breeding trait targets, such as disease resistance, storage disorders, and basic fruit chemistry. This set of

germplasm targeted contemporary breeding material that was more relevant to specific breeding goals. The germplasm set evaluated in 2017-2018 included specific families that have large breeding relevance due to their high positive selection rates and relevant ancestral origin in relation to contemporary parental selection for crosses. The entire set of pedigree-connected germplasm was included in QTL analyses to attempt to detect major QTLs for SSC (Peace et al. 2014) while six families evaluated from 2014-2018 were analyzed separately for haplotype analysis as they had larger family sizes to provide greater power for estimation of haplotype effects at identified QTLs.

The six families of focus for this study are 'Honeycrisp' x 'Minnewashta', 'Honeycrisp' x MN1836, 'Honeycrisp' x MN1915, 'Minneiska' x 'MN55', 'Minneiska' x 'Wildung', and 'Minneiska' x MN1965. These families captured a substantial proportion of genetic diversity being utilized for breeding efforts, and also focused on important breeding parents, namely, 'Honeycrisp' and 'Minneiska'. Within the ancestral background of these families are many released cultivars that served as parents or grandparents: 'Honeycrisp', 'Minneiska' (trademarked as 'SweeTango'), 'Minnewashta' (trademarked as 'Zestar!'), 'Wildung' (trademarked as 'SnowSweet'), 'MN55' (trademarked as 'First Kiss' and 'Rave'), 'Keepsake', 'State Fair', 'Sweet Sixteen', 'Ginger Gold', 'Sharon', and 'Fireside'. These six families represent the direction of contemporary breeding efforts in the UMN apple breeding program and had minimal quantitative SSC data recorded for them prior to this study.

Phenotyping Protocol

With the RosBREED1 project, a standardized phenotyping protocol was developed for numerous apple fruit quality traits (Evans et al. 2012). This protocol was developed for consistency of data collection across three institutions (UMN, WSU, and CU). This standard protocol included measuring several simple chemistry components: pH, titratable acidity (TA, in g/L in reference to malic acid), and SSC (°Brix).

Fruit was harvested from each tree when the starch pattern index measured between four and six (Blanpied and Silsby 1992). Five to fifteen fruit were harvested from each tree; five fruit were immediately juiced and the remainder were immediately placed into a 4°C cooler for ten weeks or twenty weeks. Fruit removed from storage for evaluation sat at room temperature for one week before processing. Fruit were processed by juicing the combined top halves of at least three fruit, creating one juice sample per individual. Juice samples were immediately placed into a -80°C freezer. Juice samples were removed and allowed to reach room temperature prior to chemistry data collection. The juice collected at harvest will be hereafter referred to as "at harvest", the juice collected after ten weeks of refrigerated storage and one week at room temperature will be referred to as "10-weeks post storage", and the juice collected after twenty weeks of refrigerated storage and one week at room temperature will be referred to as "20-weeks post storage".

Apple juice samples were evaluated using a digital refractometer (Atago PAL-1 3810) to measure two technical replicates of SSC for each sample. The technical replicates were averaged together to create one SSC value for each individual at each harvest/storage evaluation. For genotypes in the 'Honeycrisp' x 'Minnewashta' family that were planted and harvested in replicate, SSC measurements were taken for each

replicated tree, and their observed values were averaged together to create one mean value for each genotype.

Genetic Markers

Each individual was genotyped using a SNP array. The methods for DNA extraction described by Clark et al. (2014) were used for this study.

The genetic marker set used for QTL detection was derived from two SNP arrays developed specifically for apple: the IRSC apple 8K SNP array v1 (Chagne et al. 2012) (hereafter referred to as the "8K SNP array") and the 20K Infinium SNP array (Bianco et al. 2014) (hereafter referred to as the "20K SNP array"). At the beginning of data collection in 2010, trees were genotyped utilizing the 8K array, as it was economical and provided sufficient genomic coverage to perform QTL analyses. Trees that were utilized after 2016 or that were re-genotyped to confirm identity were genotyped with the 20K SNP array, which superseded the 8K SNP array and remains currently available. The 20K SNP array includes over 2,000 SNPs from the previous array and provides additional SNPs including some that provide higher marker saturation on the ends of chromosomes and within regions that have been identified to play key roles in several major fruit traits (Bianco et al. 2014).

A subset of 2,213 markers that are common to both arrays was identified in order to perform QTL analyses that included individuals genotyped with either array (Appendix Table 1B). A detailed description of the reproducible SNP curation and the genetic map that was used for this study is described in Howard et al. (2017) and

Vanderzande et al. (2019). All genotyped individuals had their genetic marker sets trimmed to include only the markers in this common marker subset, further referred to as the "combined marker set". The combined marker set was used for all QTL and haplotype analyses.

Best Linear Unbiased Predictions and QTL Mapping

Three Best Linear Unbiased Prediction (BLUP) models were utilized to adjust observed values for a year effect based on the overall average value of all individuals for each year of a specified time period (2010-2012, 2014-2018, and 2010-2018), and then all observed values across all specified years were averaged to create an overall average value for SSC for each individual. The BLUP models enabled utilization of the replicated data over several years, since simple averaging of all measurements for each individual cannot accurately capture variation due to different environmental factors each year. FlexQTL^{TM} software (Bink et al. 2002, Bink et al. 2012) was utilized to perform QTL mapping for each year and storage evaluation using observed values, and for the three adjusted multi-year datasets.

Haplotyping Protocol

After genomic regions of interest were identified from QTL mapping, each region was individually evaluated using the six key families in this study. Data sets were constructed that included SNP markers in the QTL regions for all offspring individuals in the six key families in order to estimate haplotype effects. Phased haplotypes from each

region were obtained from the "mhaplotypes.csv" output file from FlexQTL^{TM} for each individual in the six key families. SNP markers that were non-informative (i.e. were fixed for all individuals in the six key families) were removed from haplotype characterization. Haplotype names were assigned for each unique haplotype based upon the most distant genotyped ancestor possessing that haplotype, derived identity by descent. If a recombination was observed, the novel recombinant haplotype was named based upon the individual that the received the recombinant haplotype and the parent that it was inherited from. Offspring in the six key families that inherited recombinant haplotypes from one or both parents were removed from further haplotype analysis as the minimal representation of the recombinant provided insufficient power to estimate their effects.

Statistical Analyses of Diplotype and Haplotype Effects and Segregation Distortion

The genotype of each individual for a QTL region was characterized in two ways: the first method utilizes the diplotype of each individual (an identity based on its combined haplotypes, its "diplotype name"), and the second method creates two identities for each individual, one for each haplotype it carries, its "haplotype names". Evaluation of diplotypes and haplotypes was performed in order to evaluate groups of individuals that share the same diplotype at a given locus, and to evaluate groups of individuals that share at least one copy of a haplotype at a given locus. Due to the highly related nature of the families observed in this study, haplotype effects for a QTL could be estimated across families as in many instances, identical haplotypes by descent are found in a least two separate families.

Two methods were used to examine diplotype and haplotype effects: (1) analysis of each of the six key families separately using their BLUP 2014-2018 SSC values, and (2) analysis of individuals using their BLUP 2014-2018 SSC values in the six key families but ignoring their family identity. Each family was investigated separately due to more than one locus being identified as influencing the trait value, therefore accounting for the confounding effects of multiple loci segregating in an individual family. All families were evaluated collectively to estimate overall diplotype and haplotype effects at each locus.

A chi-squared test was performed for each family for each locus to identify segregation distortion from Mendelian expectation. Progeny were classified based upon their diplotype state and were further included in diplotype and haplotype analyses if they have SSC data in the BLUP 2014-2018 SSC dataset.

Analysis of Variance (ANOVA) tests were performed on each family for each QTL to evaluate if the variation for SSC among diplotypes was significant. ANOVA tests were also performed for the combined six families for diplotype effects and haplotype effects at each locus, and a Tukey's honestly significant difference (Tukey's HSD) test ($p= 0.05$) was performed when the ANVOA was significant ($p<0.10$) or nearly significant. A Tukey's HSD test was performed in order to identify which, if any, groups of diplotypes or haplotypes have statistically significant variation among each other.

Results

Variation for Soluble Solids Content

SSC data was collected over eight harvest seasons between 2010 and 2018. Figure 2.1 shows the distribution of SSC at harvest values for all genotypes obtained in each year. 1,044 genotypes were evaluated at least once during these eight years. All years appear to have a normal or slightly skewed distribution of SSC values, but yearly medians appear to slightly increase over the years.

The distributions of adjusted harvest values for all individuals for each of the multi-year BLUP models (2010-2018, 2010-2012, and 2014-2018) are in Figure 2.2. The 2010-2018 and 2014-2018 models appear to be similar, but 2014-2018 values are shifted higher. All distributions appear to be normal or slightly skewed.

The distributions of the 2014-2018 BLUP SSC values at harvest for all individuals in each of the six families of focus for this study are in Figure 2.3. Most distributions are normal, with the exception of 'Minneiska' x 'Wildung', which appears to be skewed towards higher SSC values. The distributions of the 2014-2018 BLUP SSC values at 10-weeks post storage for all individuals in each of the six families of focus for this study can be found on Figure 2.4. Most distributions are normal, with the exception of 'Minneiska' x 'MN55', which appears to be bimodal and 'Minneiska' x 'Wildung' which appears to be skewered towards higher SSC values.

The 2014-2018 BLUP SSC values for each individual within the six families at harvest and at 10-weeks post storage were plotted. An overall coefficient of determination $R^2 = 0.64$ ($p = 2.0 \times 10^{-39}$) is obtained (Figure 2.5). The slope for the linear line of best fit is 1.1 and the y intercept is 0.30.

QTL Mapping

LGs containing a QTL for SSC that were detected and mapped using FlexQTL™ in each analysis are summarized in Table 2.1 including sample size and transformed Bayes Factor for each evaluation period-year-model. For each linkage group with a QTL, the transformed Bayes Factor (BF) defined as [2*ln][Bayes Factor] is presented. Bink et al. (2012) considered 5<BF<10 and BF>10 as "strong" and "decisive" evidence, respectively, for QTL in genomic regions indicated in FlexQTL™ analyses.

The regions (in centimorgans) on significant LGs (BF>5) are defined in Table 2.2. The regions on LGs that were consistent (multiple different years/models and/or evaluation periods) and significant and were used in further analyses are described below.

A QTL on LG1 was detected for variation in SSC in six analyses (Table 2.1). Two of these QTL had a 5<BF<10 (strong evidence, analyses: 2010-2012 BLUP at 20-weeks post storage and 2014-2018 BLUP at harvest). Two of these QTL had a BF>10 (decisive evidence, analyses: 2010-2018 BLUP at harvest and 2010-2018 BLUP at 20-weeks post storage). LG1 was only detected in the analyses using BLUP SSC values. The QTL region consistently identified spans 34 cM to 63 cM (Table 2.2).

A QTL on LG13 was detected for variation in SSC in five analyses (Table 2.1). Two of these QTL had a 5<BF<10 (strong evidence, analyses: 2016 at harvest, and 2014-2018 BLUP at 10-weeks post storage). Three of these QTL had a BF>10 (decisive evidence, analyses: 2018 at harvest, 2014-2018 BLUP at harvest, 2010-2018 BLUP at harvest). LG13 was identified only in plant material that was evaluated in the 2014-2018 period. The QTL region consistently identified spans 17 cM to 58 cM (Table 2.2).

Haplotyped Loci

For the QTLs identified on LG1 and LG13, the minimum number of informative SNPs needed to define the segregating haplotypes at each locus was determined.

The LG1 locus was thinned to six SNP markers from 51.28 cM to 56.57 cM, which spanned 29,522,664 bp to 30,943,508 bp in reference to the 'Golden Delicious' doubled-haploid genome (GDDH13) (Table 2.3). One of the flanking markers that Guan et al. (2015) identified on LG1 as significant for fructose content was retained after thinning. Several markers used by Guan et al. (2015) are redundant in this set of germplasm, providing no additional information, and are therefore not used in further haplotype analysis. Six unique parental haplotypes were found at the LG1 locus.

The LG13 locus was thinned to eighteen SNP markers from 53.38 cM to 65.94 cM, which spanned 15,869,389 bp to 23,862,860 bp in reference to GDDH13 (Table 2.4). One of the flanking markers that Guan et al. (2015) identified on LG13 as significant for SSC was retained after thinning. Ten unique parental haplotypes were found at the LG13 locus.

The parental haplotypes and associated names for all SSC loci can be found in Table 2.5. The basis for naming and name selection can be found in Table 2.6.

Segregation Distortion

Chi-squared tests were performed for each family at each locus to determine if expected segregation ratios of diplotypes were observed (Table 2.7). Segregation distortion was detected for the region on LG1 for 'Honeycrisp' x MN1915 (p=0.048) and 'Minneiska' x MN1965 (p=0.026). For the QTL region for SSC on LG13, the 'Honeycrisp' x 'Minnewashta' family exhibited segregation distortion (p=0.002). Hypotheses for causes of distortion are described later.

Diplotype Associations for LG1 and LG13 with Family as a Factor

Families were investigated individually at each locus to determine if they segregated for significant diplotype effects for SSC at harvest, at 10-weeks post storage, and the difference between harvest and 10-weeks post storage. These analyses extend past the analyses performed by FlexQTL™ and included additional statistical testing. The ANOVAs for each family at harvest can be found in Table 2.8, ANOVAs for 10-weeks post storage can be found in Table 2.9, and ANOVAs for the difference between harvest and 10-weeks post storage can be found in Table 2.10.

ANOVA indicated one family ('Honeycrisp' x 'Minnewashta') had significant variation among LG13 diplotypes at harvest (p= 0.042, Table 2.8). One family ('Honeycrisp' x MN1915) had significant variation among LG1 diplotypes at 10-weeks post storage (p= 0.054, Table 2.9). Two families ('Honeycrisp' x MN1915 and 'Minneiska' x 'MN55') had significant variation among LG13 diplotypes for the difference in SSC between harvest and 10-weeks post storage (p= 0.089 and p= 0.091, respectively, Table 2.10).

Diplotype Associations for LG1 at Harvest and at 10-Weeks Post Storage

Diplotypes at the LG1 locus were evaluated (Table 2.11). When utilizing all individuals from all six families, ANOVA detected significant variation among diplotypes for SSC at harvest ($p= 5.7x10^{-6}$). Two statistically significant diplotype groupings were determined using a Tukey's HSD test (Table 2.11). The SSC diplotype mean values at harvest ranged from 14.2 °Brix (GrimesRecomGrimesRecom) to 12.5 °Brix (Frostbite1Malinda2). There appears to be an even gradient of diplotype means for SSC at harvest, except the Frostbite1Malinda2 diplotype, which was 0.6 ° Brix lower than the next lowest diplotype and is the only diplotype placed exclusively in group B.

When diplotypes from all six families were evaluated for SSC at 10-weeks post storage, ANOVA detected significant variation among LG1 diplotypes ($p= 1.1x10^{-4}$). Two statistically significant diplotype groupings were determined using a Tukey's HSD test (Table 2.11). The SSC diplotype mean values at 10-weeks post storage ranged from 15.7 °Brix (Aspa2NWGreening1) to 13.0 °Brix (Frostbite1Malinda2). There appears to be an even gradient of diplotype means for SSC at harvest.

Haplotype Associations for LG1 at Harvest and at 10-Weeks Post Storage

Haplotypes at the LG1 locus were evaluated (Table 2.12). When utilizing all individuals from all six families, ANOVA detected significant variation among haplotypes for SSC at harvest ($p= 5.6x10^{-4}$). Two statistically significant haplotype

groupings were determined using a Tukey's HSD test (Table 2.12). The SSC haplotype mean values at harvest ranged from 14.1 °Brix (Aspa2) to 13.2 °Brix (Frostbite1).

When haplotypes from all six families were evaluated for SSC at 10-weeks post storage, ANOVA detected significant variation among LG1 haplotypes ($p= 0.0882$). A Tukey's HSD test did not identify multiple significant groups at the $p= 0.05$ level. The SSC haplotype mean values at 10-weeks post storage ranged from 15.2 °Brix (Aspa2) to 13.7 °Brix (Frostbite1).

Diplotype Associations for LG13 at Harvest and at 10-Weeks Post Storage

Diplotypes at the LG13 locus were evaluated (Table 2.13). When utilizing all individuals from all six families, ANOVA detected significant variation among diplotypes for SSC at harvest ($p= 7.4x10^{-10}$). Two statistically significant diplotype groupings were determined using a Tukey's HSD test (Table 2.13). The SSC diplotype mean values at harvest ranged from 14.2 °Brix (GoldenDel2Goodland2) to 12.1 °Brix (Keepsake2MinSFRecom). There appears to be an even gradient of mean diplotype SSC values at harvest, with the exception of diplotypes that contain the MinSFRecom haplotype, which exhibited exceptionally low SSC compared to the other diplotypes.

When individuals from all six families were evaluated for SSC at 10-weeks post storage, ANOVA detected significant variation among LG13 diplotypes ($p= 1.7x10^{-5}$). Three statistically significant diplotype groupings were determined using a Tukey's HSD test (Table 2.13). The SSC diplotype mean values at 10-weeks post storage ranged from 15.7 °Brix (GoldenDel2Sharon2) to 12.5 °Brix (Keepsake2MinSFRecom). There appears

to be an even gradient in diplotype SSC mean values at 10-weeks post storage, with the exception of diplotypes that contain the MinSFRecom haplotype, which exhibited exceptionally low SSC compared to the other diplotypes.

Haplotype Associations for LG13 at Harvest and at 10-Weeks Post Storage

Haplotypes at the LG13 locus were evaluated (Table 2.14). When utilizing all individuals from all six families, ANOVA detected significant variation among haplotypes for SSC at harvest ($p= 1.4x10^{-13}$). Three statistically significant haplotype groupings were determined using a Tukey's HSD test (Table 2.14). The SSC haplotype mean values at harvest ranged from 14.1 °Brix (Sharon2) to 12.1 °Brix (MinSFRecom). The MinSFRecom haplotype was the only haplotype placed in the group C, and is 1.0 °Brix lower than the next lowest haplotype (Duchess2, mean= 13.1 °Brix).

When haplotypes from all six families were evaluated for SSC at 10-weeks post storage, ANOVA detected significant variation among LG13 haplotypes ($p= 1.41x10^{-6}$). Two statistically significant haplotype groupings were determined using a Tukey's HSD test (Table 2.14). The SSC haplotype mean values at 10-weeks post storage ranged from 15.6 °Brix (GoldenDel2) to 12.6 °Brix (MinSFRecom). The lowest haplotype (MinSFRecom) was the only haplotype placed into group B, and is 1.2 °Brix lower than the next lowest haplotype (Duchess2, mean 13.8 °Brix).

Haplotype Associations for LG1 and LG13 for SSC Difference between Harvest and 10-Weeks Post Storage

Haplotypes at the LG1 locus were evaluated for the difference in SSC between harvest and 10-weeks post storage (Table 2.15). ANOVA did not detect significant variation among haplotypes for SSC difference between harvest and 10-weeks post storage ($p= 0.113$). The 2014-2018 BLUP SSC mean haplotype values at harvest and at 10-weeks post storage for each of the six families were plotted for LG1 (Figure 2.6). An overall coefficient of determination $R^2= 0.83$ ($p= 0.012$) was obtained. The slope for the linear line of best fit is 1.7 and the y intercept is -9.0.

Haplotypes at the LG13 locus were evaluated for the difference in SSC between harvest and 10-weeks post storage (Table 2.16). ANOVA detected significant variation among haplotypes for SSC difference between harvest and 10-weeks post storage ($p= 0.019$). However, a Tukey's HSD test did not determine any statistically significant haplotype groupings at $p=0.05$. The 2014-2018 BLUP SSC mean haplotype values at harvest and at 10-weeks post storage for each of the six families were plotted for LG13 (Figure 2.7). An overall coefficient of determination $R^2= 0.81$ ($p= 8.1x10^{-4}$) was obtained. The slope for the linear line of best fit is 1.3 and the y intercept is -3.2. One haplotype, MinSFRecom, is separated from the other haplotypes as it has an unusually low mean SSC at harvest and at 10-weeks post storage.

Discussion

Genomic Regions Contributing to SSC Variation Determined Using QTL Mapping

The LG1 locus consistently detected by QTL analyses spans the 51.28 cM to 56.57 cM region (29,522,664 bp to 30,943,508 bp in reference to the 'Golden Delicious'

doubled-haploid genome) (Table 2.2). This region coincides with the region previously identified by Guan et al. (2015) for QTLs associated with fructose, glucose, sucrose, and sorbitol content. One of the SNP markers used by Guan et al. (2015) to describe the region identified for fructose was the SNP marker located at 51.28 cM for this study (Table 2.3). A QTL was also identified by Liu et al. (2016) on the distal end of LG1, which is also consistent with the QTL region identified in this study.

The LG13 locus consistently detected by QTL analyses spans the 53.38 cM to 65.94 cM region (15,869,389 bp to 23,862,860 bp in reference to the 'Golden Delicious' doubled-haploid genome) (Table 2.2). This region is consistent with the QTLs identified by Guan et al. (2015) on the distal end of LG13 for sorbitol content at 20-weeks post storage (68.40 cM in reference to the combined marker dataset), and for SSC at 20-weeks post storage (65.94 cM in reference to the combined marker dataset) (Table 2.4).

Segregation Distortion

Two families exhibited segregation distortion for diplotypes at the LG1 locus. LG1 exhibited segregation distortion in families observed by Zhang et al. (2012). They noted that the markers that exhibited distortion are closely linked to the *Vf* (*Rvi6*) gene, a gene associated with apple scab resistance that resides on LG1 (Maliepaard et al. 1998). Self-incompatibility or chromosome loss was proposed by Zhang et al. (2012) as a possible explanation for segregation distortion on LG1. The Rvi6_Vf6_M799 marker, associated with *Rvi6* apple scab resistance (Bus et al. 2011) originating from *Malus floribunda*, is located at 27.9 Mbp on chromosome one of GDDH13 (Daccord et al.

2017). This marker is in close proximity to the LG1 locus identified in this study, which begins at 29.5 Mbp, rendering the *Rvi6* marker within about 2.6 Mbp. Selection for this marker may result in segregation distortion at the LG1 locus.

The 'Honeycrisp' x 'Minnewashta' family exhibited significant segregation distortion (p= 0.003) at LG13. Clark et al. (2014) described distortion on LG13 in a study of families with 'Honeycrisp' as one of the parents. They did not advance a hypothesis to explain the distortion observed, but due to high relatedness between the family in Clark et al. (2014) and the family observed to have distortion in this study, it is possible the distortion was caused by the same factor.

Utilization of BLUP Models and the Effects on QTL Mapping

The QTL analyses that utilized BLUP adjusted data were compared to the QTL analyses that utilized only the observed data (Table 2.1). Generally, the BLUP analyses consistently identified similar QTL that were commonly detected using observed values, and detected QTL that were not identified in any single year analyses.

For further haplotype analysis, the 2014-2018 BLUP dataset for SSC data (at harvest and at 10-weeks post storage) for the six families of interest was utilized. This dataset included larger family sizes based on the germplasm of highest interest and in analyses it exhibited clear, significant QTLs on LG1 and LG13.

Haplotype Naming

While many haplotypes could be traced back to the most distant genotyped ancestor, some haplotypes could only be traced to an intermediate ancestor that was homozygous in the selected region (for example, the US2Ark haplotype at the LG13 locus). While the haplotypes were traced back as far as possible using current genetic data, some ancestors have missing genetic data (such as SNPs that could not be characterized for some individuals). These unique parental haplotypes may in fact be more related than could be determined in this study. Additional genetic data may further complete the pedigrees and indicate higher degrees of relatedness.

Diversity and Origin of LG1 Haplotypes

In the six families evaluated in this study, six unique parental haplotypes were identified at the LG1 QTL (Table 2.6). These haplotypes originate from: the unknown parent (designated as paternal) of 'Aspa', the unknown parent (designated as maternal) of 'Frostbite', a recombinant haplotype from 'Grimes Golden' which was inherited by 'Golden Delicious', the unknown parent (designated as paternal) of 'Malinda', the unknown parent (designated as maternal) of 'Northern Spy', and the unknown parent (designated as maternal) of 'Northwest Greening'. One of these cultivars, 'Frostbite' was released by the UMN. It was extensively used in breeding by the UMN in the mid-20th for its cold-hardiness and is a grandparent of 'Honeycrisp'. 'Malinda' is a parent or grandparent of at least nine released cultivars from the UMN and is a grandparent of 'Honeycrisp'. The low diversity at this LG1 locus indicates high selective advantage and low rates of crossover events during inheritance into offspring.

Diversity and Origin of LG13 Haplotypes

In the six families evaluated in this study, ten unique parental haplotypes were identified at the LG13 QTL (Table 2.6). These haplotypes originate from: a recombinant haplotype from 'State Fair' which was inherited by MN1836, the unknown parent (designated as paternal) of 'Duchess', the unknown parent (designated as paternal) of 'Golden Delicious', the unknown parent (designated as paternal) of 'Goodland', the unknown parent (designated as paternal) of 'Keepsake', a recombinant haplotype from 'State Fair' inherited by 'Minnewashta', the unknown parent (designated as paternal) of 'Northern Spy', the unknown parent (designated as maternal) of 'Northwest Greening', the unknown parent (designated as paternal) of 'Sharon', and one of the unknown parents of US2Ark. The highly diverse region for the LG13 locus contains only three haplotypes that are found in more than one parent, the Duchess2, Goodland2, and Keepsake2 haplotypes, which are present in 'Honeycrisp', 'Minneiska' (a cross between 'Honeycrisp' and 'Minnewashta'), and 'Minnewashta'. These three parents have been used extensively in the UMN apple breeding program, but the presence of additional haplotypes indicates that this region has not been fixed across other unrelated UMN selections and cultivars.

Linkage Group 1 Locus Diplotype Analysis

Several studies (Liu et al. 2016, Sun et al. 2015, Guan et al. 2015) have identified the distal end of LG1 as contributing to variation in fructose, soluble solids, or sucrose.

The regions in previous studies coincide with the region detected through QTL analyses for SSC in this study as well as a region contributing to variation for TA (discussed in Chapter 3).

When each family was evaluated for LG1 diplotype effects on SSC at harvest, none of the ANOVA p-values were significant at the 90% confidence level (Table 2.8). Lack of variation due to the effect of the LG1 locus may be due to several factors: 1) germplasm at the UMN is mostly fixed for diplotype effects at the LG1 locus, resulting in no significant differences between diplotypes within families; and/or 2) family size and SSC measurement precision are insufficient to capture the true variation among diplotypes; and/or 3) SSC is heavily influenced by the environment, such that variation due to non-genetic factors overwhelmed the genetic effects contributing to SSC.

In contrast, when all individuals from all six families were included in ANOVA (Table 2.11), variation among diplotypes was significant at harvest ($p= 5.7x10^{-6}$). However, the differences in haplotype means for SSC were small in the context of sensory perception. For example, a 1.7 °Brix difference between two apples may not result in a noticeable difference in sensory sweetness when one also considers TA content (Harker et al. 2002).

Linkage Group 1 Locus Haplotype Analysis

Haplotype groups at the LG1 locus for all six families for SSC were evaluated (Table 2.12). ANOVA utilizing all individuals from all six families indicated significant variation among haplotypes at harvest ($p= 5.6x10^{-4}$) and at 10-weeks post storage (p=

0.088). Haplotypes with the highest averages at harvest had the highest averages at 10-weeks post storage (Table 2.12).

However, the range from the highest group of haplotype means to the lowest group of haplotype means at harvest was 0.9 °Brix. From a sensory perspective, 0.9 °Brix is a small difference in sweetness levels (Harker et al. 2002). Due to the low variation between haplotype means at this locus, breeding decisions for sensory sweetness should not be solely based upon the LG1 locus.

Linkage Group 13 Locus Diplotype Analysis

One study has previously reported a QTL on LG13 influencing variation in SSC and sorbitol content (Guan et al. 2015), consistent with the QTL detected in this study. The 18 polymorphic SNP markers used to define the haplotypes at this locus range from 53.38 cM to the end of LG13 at 68.40 cM (Table 2.6).

When each family was evaluated for diplotype effects on SSC at harvest by ANOVA, only ‘Honeycrisp’ x ‘Minnewashta’ had significant variation among diplotypes ($p= 0.041$, Table 2.8).

Several diplotypes can be compared across families due to the inter-relatedness of the families used in this study. Of the 18 unique offspring diplotypes across the six families evaluated, three of the diplotypes were identified in multiple families. Diplotypes that occur in several families do not appear to have consistent mean values, suggesting that other loci or genetic background influence SSC values.

While three haplotypes (Duchess2, Goodland2, and Keepsake2) were frequent in these six families, several other unique haplotypes were contributed from sources other than ‘Honeycrisp’, ‘Minneiska’, and ‘Minnewashta’. This suggests that this locus has not been fixed within UMN breeding germplasm and other loci are likely also contributing to SSC variation.

Diplotype groups at the LG13 locus for all six families were evaluated (Table 2.13). ANOVA utilizing individuals from all six families indicated significant variation among diplotypes at harvest ($p= 7.4x10^{-10}$) and at 10-weeks post storage ($p= 1.7x10^{-5}$). Two statistically significant diplotype groups were identified for SSC at harvest using Tukey’s HSD test (Table 2.13). The two lowest diplotypes, which both carry the MinSFRecom haplotype, had the same mean value of 12.1 °Brix, which was 1.1 °Brix lower than the next lowest diplotype. The MinSFRecom haplotype may contribute to a lower production of SSC. When the diplotype groups that contain the MinSFRecom haplotype were removed from the LG13 diplotype analysis, variation among diplotype groups at this locus was not significant.

The remaining diplotypes at the LG13 locus differ by only 1.0 °Brix between the highest and lowest means. Due to a low sample size in many of the diplotype groups, the variation observed within each diplotype is high, causing difficulty in statistical separation of the diplotype means.

The diverse background at the LG13 locus is indicated by a large number of diplotypes. Due to the resulting small sample size for many of the diplotypes, their means for 10-weeks post storage should be observed with caution. However, diplotypes that contain the MinSFRecom haplotype continued to exhibit low SSC, consistent with the

findings at harvest. This haplotype may be undesirable to a breeder and selection against this haplotype may be warranted.

Linkage Group 13 Locus Haplotype Analysis

Haplotype groups at the LG13 locus varied significantly at harvest ($p= 1.36x10^{-13}$) and at 10-weeks post storage ($p= 1.4x10^{-6}$) based on ANOVA utilizing all individuals from all six families (Table 2.14). However, ranges of values within and between groups of haplotypes at this locus are very small.

The MinSFRecom haplotype was 1.0 °Brix lower than the next lowest haplotype at harvest and had the lowest mean at 10-weeks post storage. This difference in SSC values is minimal and will likely go undetected by a fruit evaluator when tasting.

Due to a diverse background at the LG13 locus, many haplotypes are observed, resulting in small sample sizes. From a breeding application standpoint, this locus is of minor interest. This LG13 locus should be considered along with other SSC loci if SSC is being targeted for selection.

Relationship of SSC at Harvest and at 10-Weeks Post Storage - Linkage Group 1 Locus Haplotypes

The LG1 locus haplotype SSC means at harvest were highly predictive of haplotype means at 10-weeks post storage (Figure 2.6). The large coefficient of determination ($R^2= 0.83$, $p= 0.012$) suggests that haplotypes at this locus do not have a

substantial impact on the SSC difference between harvest and 10-weeks post storage. Likewise, ANOVA does not indicate significant variation among haplotypes for all six families ($p= 0.1129$), as haplotypes exhibited similar means for SSC difference between harvest and 10-weeks post storage (Table 2.15). Comparing the linear regression obtained for all individuals in all six families without factoring any loci (Figure 2.5) to the LG1 linear regression indicates that only a portion of variation of SSC at harvest and at 10-weeks post storage is explained by the LG1 locus.

The relative SSC at harvest for LG1 haplotype means are highly predictive of most of the relative LG1 haplotype means at 10-weeks post storage. The NorthSpy1 haplotype is a noteworthy exception with its large deviation from the linear regression, as its SSC at 10-weeks post storage is at least 1.5 °Brix higher than the regression predicts. While a Tukey's HSD test did not indicate any statistically significant haplotype comparisons at the $p=0.05$ level (Table 2.15), the NorthSpy1 haplotype may be of interest for a breeder when considering SSC at 10-weeks post storage due to its high mean value.

Relationship of SSC at Harvest and at 10-Weeks Post Storage - Linkage Group 13 Locus Haplotypes

The LG13 locus haplotype SSC means at harvest were highly predictive of LG13 haplotype means at 10-weeks post storage (Figure 2.7). The large coefficient of determination ($R^2= 0.81$, $p= 8.1x10^{-4}$) suggests that haplotypes at this locus do not have a substantial impact on the mean SSC difference between harvest and 10-weeks post storage. Although ANOVA does indicate significant variation among haplotypes for all

six families (p= 0.0185), the Tukey's HSD test does not indicate any statistically significant haplotype groupings (Table 2.16). This may be due to the high number of unique haplotypes found at the LG13 locus, resulting in haplotype groups of relatively small sizes with high variation within them. Comparing the linear regression obtained for all individuals in all six families without factoring any loci (Figure 2.5) to the LG13 linear regression indicates that only a portion of variation of SSC at harvest and at 10-weeks post storage is explained by the LG13 locus.

The relative harvest SSC LG13 haplotype means are predictive of most of the relative haplotype means at 10-weeks post storage. The MinSFRecom haplotype is noteworthy, as its SSC at harvest and at 10-weeks post storage was at least 1.0 °Brix lower than the other haplotypes, yet its mean difference is similar to the other haplotypes. While a Tukey's HSD test did not indicate any statistically significant haplotype comparisons at the p= 0.05 level (Table 2.16), a breeder may wish to select against this haplotype when considering SSC at harvest and at 10-weeks post storage due to its low mean value.

Linkage Group 16 Locus

The region identified on LG16 contributing to variation for SSC resides in the same region associated with variation for acidity content and previously identified as the *Ma* locus (Verma et al. 2019). While the LG16 locus was significant in three QTL analyses for SSC in this study (BF>5, Table 2.2), this locus does not appear consistently, and was therefore not investigated. Further investigation on additional or larger families

may identify this locus as an additional factor contributing to variation for SSC in certain germplasm.

Confounding Trait Selection

Sweetness is only one of many traits that breeders must consider in fruit marketed after extended cold storage. Selection for other fruit traits such as texture, soft scald storage disorder, bitter pit, blue mold, and other storage disorders may influence variation for sweetness if genetic linkage is a factor. For example, a QTL for firmness following storage has been identified on LG1 (King et al. 2000) and may be physically linked to the LG1 locus identified for SSC in this study. Due to these additional layers of selection being performed for other traits, unintended selection for haplotypes that influence sweetness may occur.

The families utilized in this study were selected specifically to examine certain traits, however, they may have undergone unintentional or intentional human selection or natural selection prior to data collection. Segregation distortion in a genomic region may occur regardless of experimental design or maintenance. An example is the case of an allele for a gene on LG9 conditioning zonal leaf chlorosis when heterozygous and lethality when homozygous (Howard et al. 2019). While zonal leaf chlorosis could be clearly identified, not all cases of segregation distortion can be easily explained. For traits such as SSC that are controlled by multiple loci, the consequences of segregation distortion may be even more difficult to discern. When multiple loci are confounded with distorted ratios, identification and characterization of loci becomes increasingly

challenging. A study with larger sample sizes than this study would be needed in order to address this topic for SSC in apple.

Conclusions

QTLs for variation in SSC were identified on LG1 and LG13 and explain a small proportion of variation in SSC from a practical breeding perspective. Genetic effects contributed to few statistically significant differences in SSC among diplotype or haplotype groups, limiting the application of these findings to a wider range of families and individuals in the UMN apple breeding program.

One haplotype of particular interest is the MinSFRecom haplotype inherited to 'Minnewashta' from 'State Fair' at the LG13 locus. This haplotype appears to contribute to significantly lower SSC at harvest and at 10-weeks post storage compared to the other haplotypes investigated in this study at the LG13 locus. This haplotype is found in many recent UMN apple selections (MN2015, MN2016, MN2044, MN2050, MN2067, MN2070, MN2094, MN2105, MN2133), indicating this haplotype has been selected for in offspring.

An additional haplotype of interest is the Aspa2 haplotype at the LG1 locus, which is found in MN1965. At harvest, this haplotype was associated with high SSC and low TA (described further in Chapter 3). This haplotype is unique in that all other haplotypes with higher SSC are also associated with higher TA, except Aspa2. Therefore, a breeder should be aware that the Aspa2 haplotype may contribute to individuals having high SSC and low TA content, which could lead to flat tasting fruit. A breeder should

consider avoiding using parents that carry this haplotype or crossing them to parents with haplotypes contributing to higher TA, or using marker assisted seedling selection to eliminate seedlings that inherit this haplotype. This haplotype is found in many recent UMN apple selections (MN2011, MN2062, MN2073, MN2076, MN2117, MN2121, MN2124, MN2124, MN2140, MN2151, MN2155), indicating this haplotype has been passed on to offspring and may be linked to a desirable trait. Additionally, this haplotype is also found in the cultivar 'Ginger Gold', which may have been used extensively in other breeding programs and widens the implications from this study.

Soluble solids content appears to be expressed as a quantitative trait (Blissett 2019) that is likely controlled by many loci having small effects that could not be detected in this study. While previous studies have reported significant QTLs for SSC, analyses of the loci identified for the six families in this study indicate that these loci alone account for only a small portion of the variation observed in these families. QTL mapping with larger family sizes or a genome-wide modeling approach may be able to identify new or additional loci of smaller effects and have greater predictive ability. Alternatively, breeding efforts may be more productively focused on other traits related to sensory perception and consumer acceptance.

Figure 2.1. Distributions of SSC at Harvest across Years. A boxplot laid over a violin plot for the observed SSC values at harvest for each year in this study from 2010 to 2018. No data were collected in 2013. Each bolded point represents outliers. The box plots indicate the 25th, 50th, and 75th percentiles for each dataset.

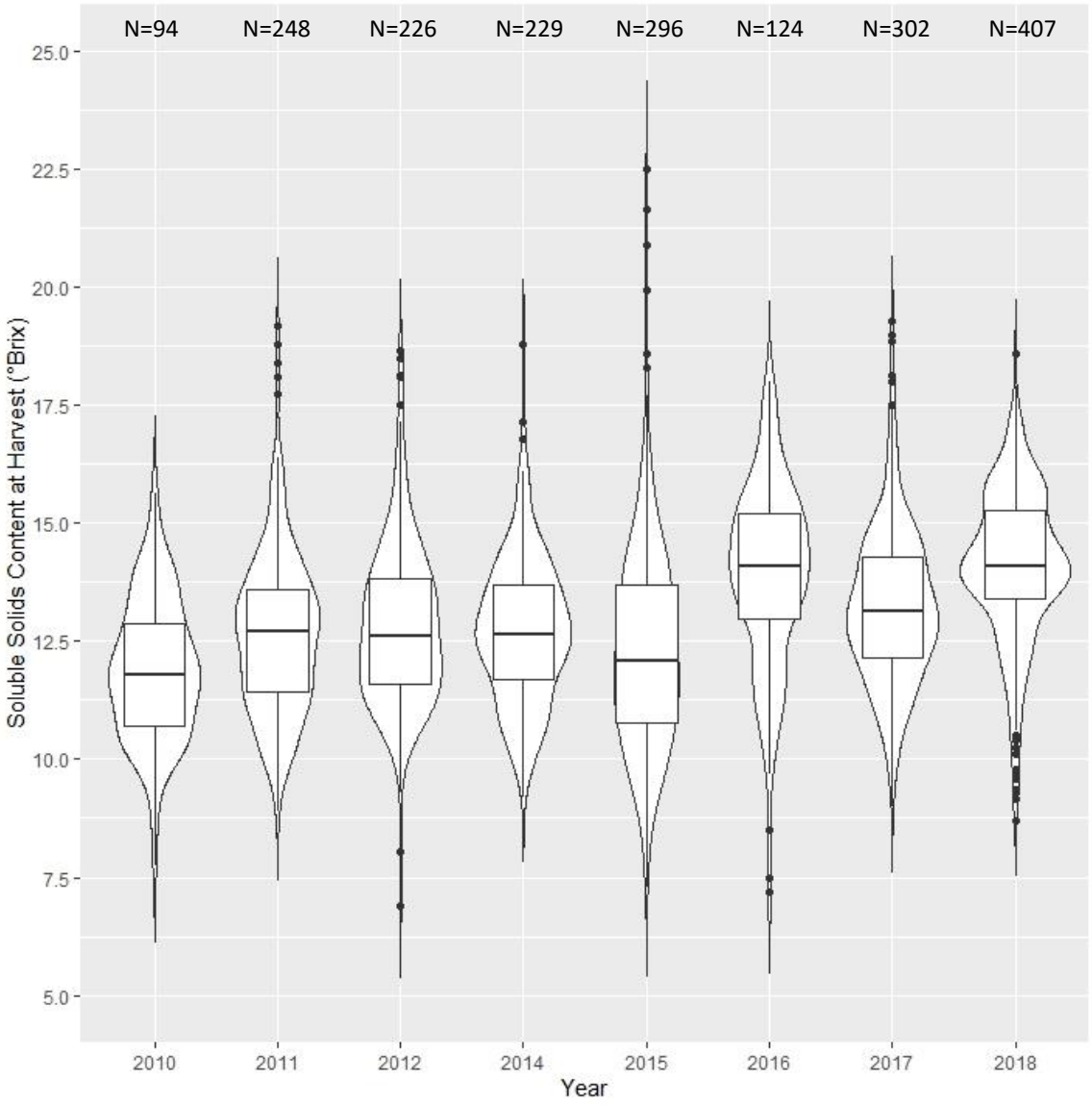


Figure 2.2. Distributions of SSC at Harvest across BLUP Models. A boxplot laid over a violin plot for adjusted SSC values at harvest using three BLUP models based on SSC data from 1) 2010-2018, 2) 2010-2012, and 3) 2014-2018. Each bolded point represents outliers. The box plots indicate the 25th, 50th, and 75th percentiles for each dataset.

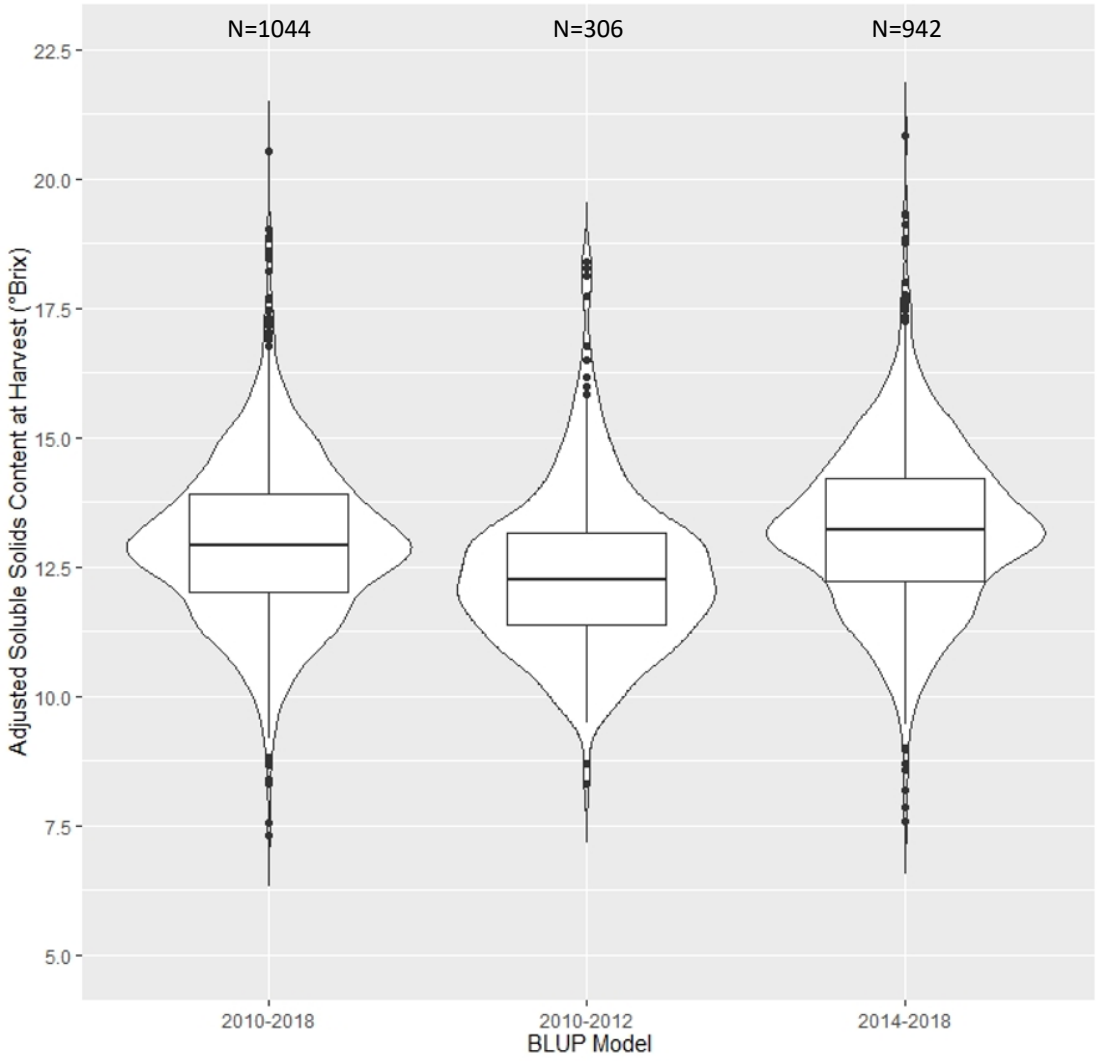


Figure 2.3. Distributions of SSC at Harvest across Families. The distributions of adjusted SSC values at harvest for the individuals using the 2014-2018 BLUP model in a histogram for each of the six families of focus in this study.

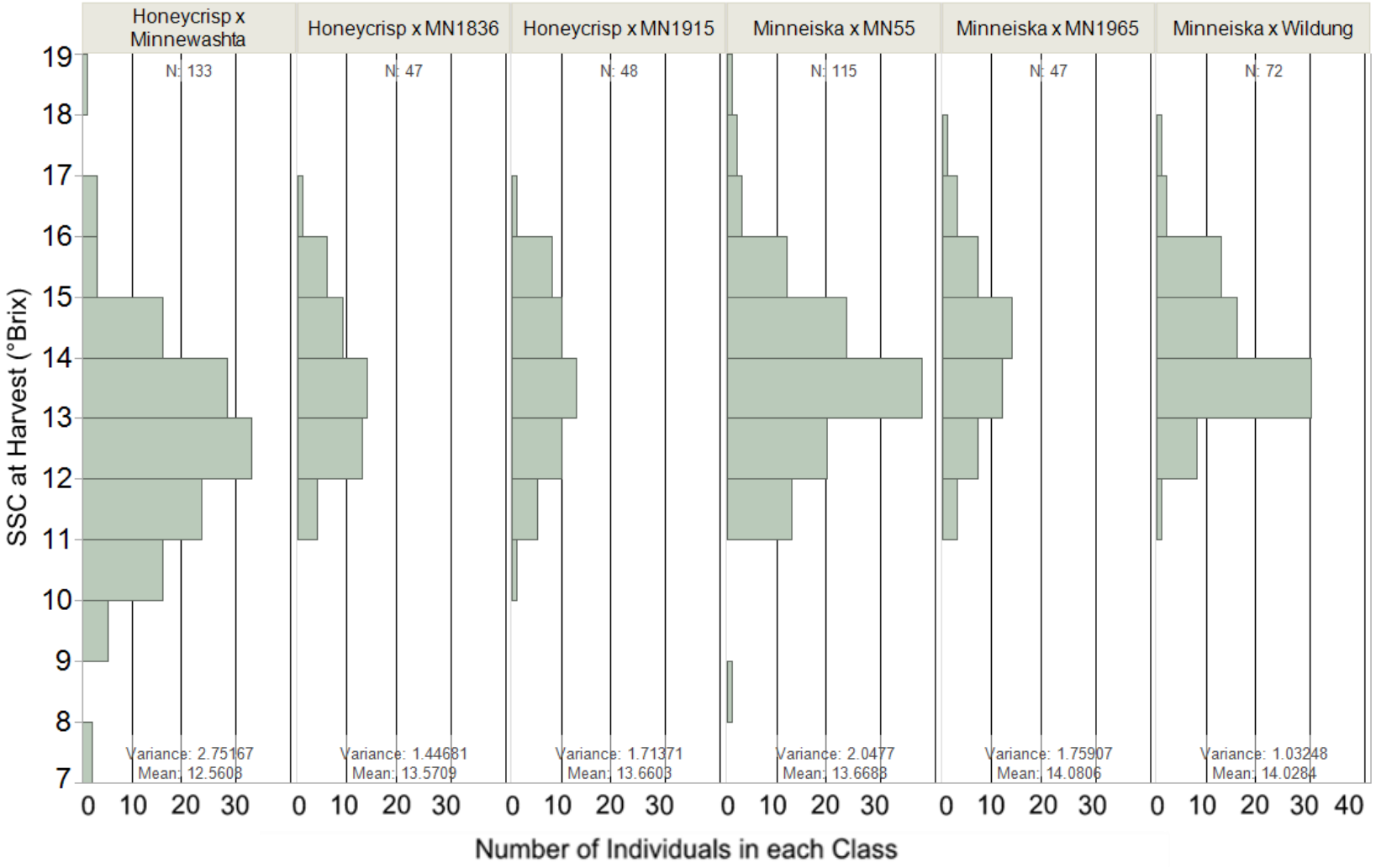


Figure 2.4. Distributions of SSC at 10-Weeks Post Storage across Families. The distributions of adjusted SSC values at 10-weeks post storage for the individuals using the 2014-2018 BLUP model in a histogram for each of the six families of focus in this study.

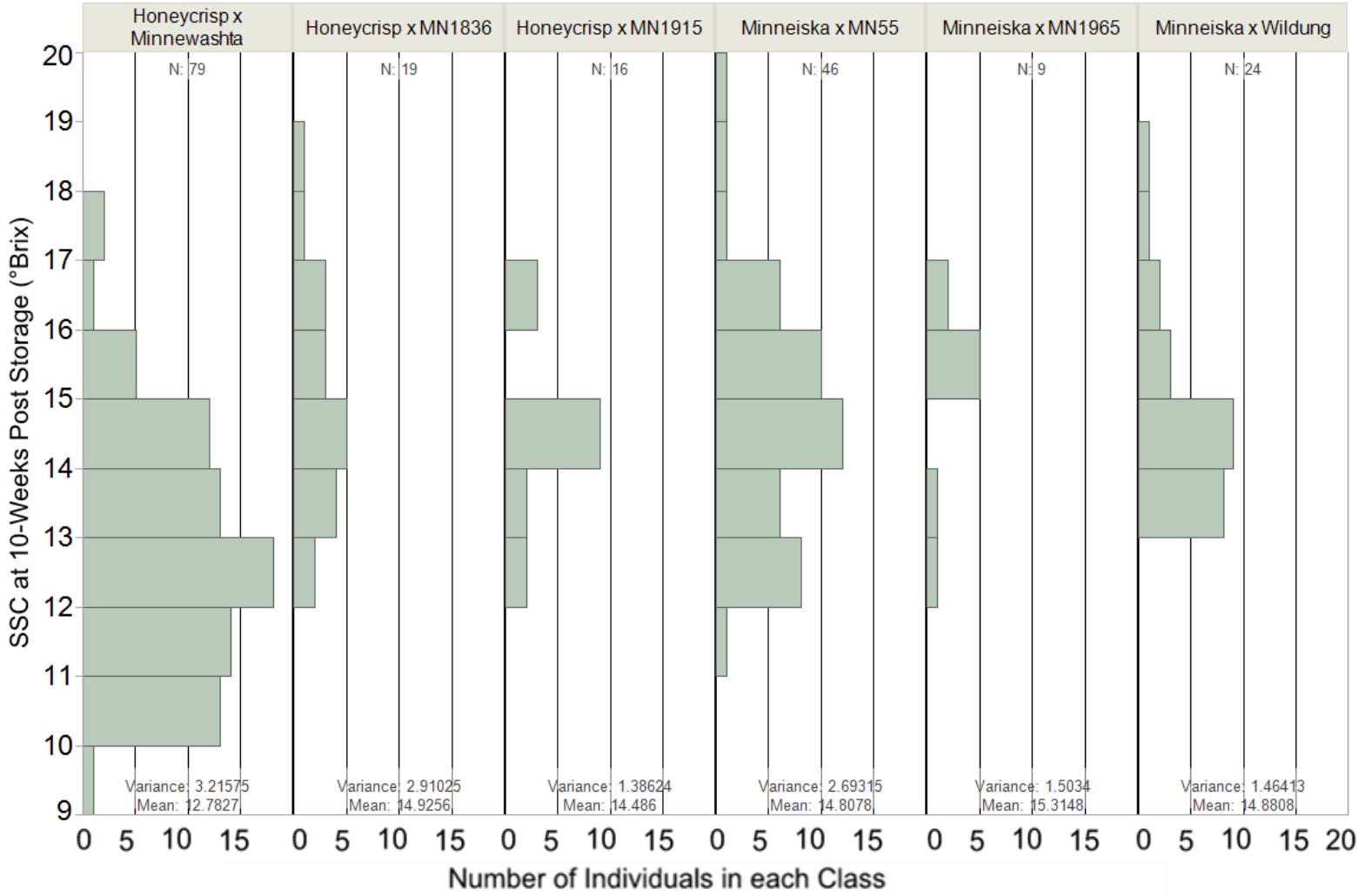


Figure 2.5. SSC at Harvest vs. 10-Weeks Post Storage. Plot of the individual values from the 2014-2018 BLUP model for SSC at harvest and at 10-weeks post storage for the six families of interest. Each point represents one individual. The total number of individuals is 171. The linear regression and coefficient of determination is significant with $p=1.97x10^{-39}$.

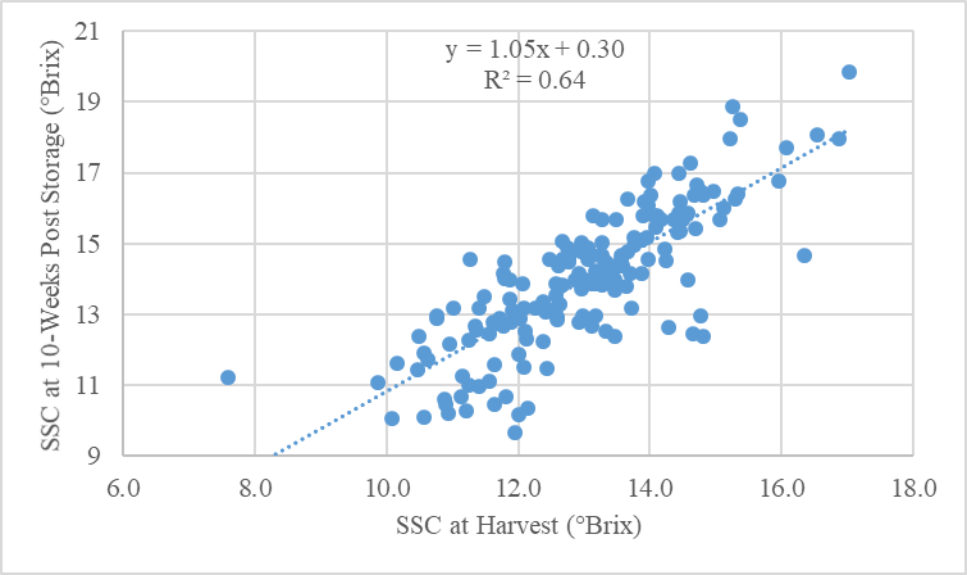
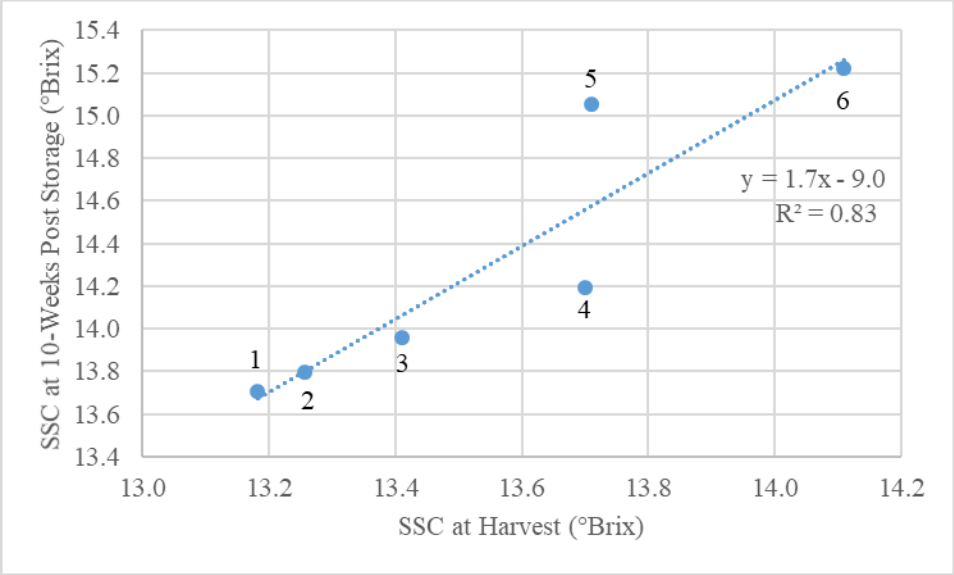


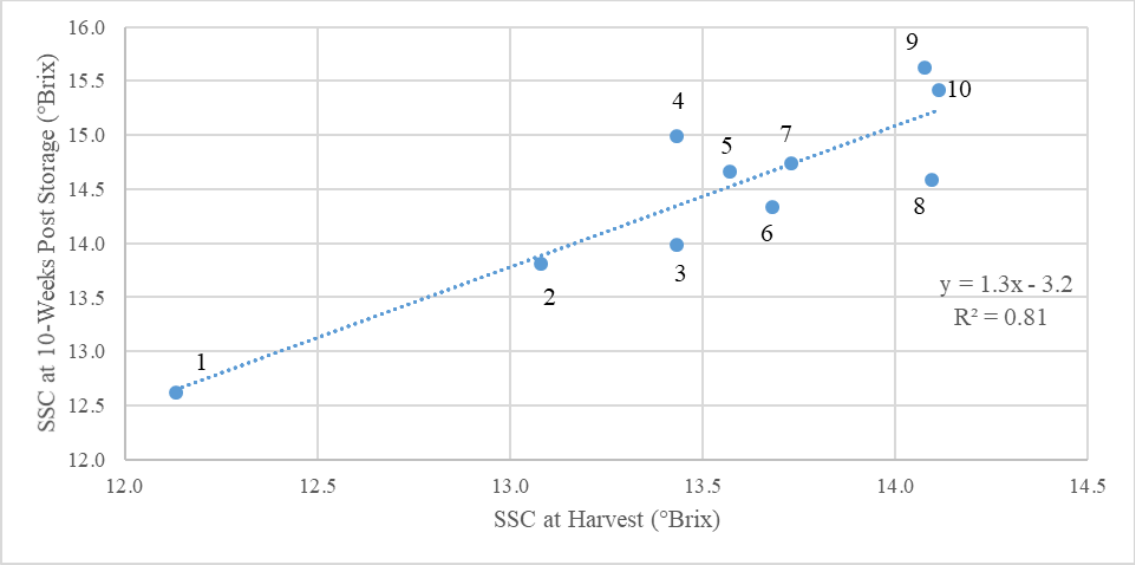
Figure 2.6. LG1 SSC Haplotypes at Harvest vs. 10-Weeks Post Storage. Plot of the groups of haplotype means from the 2014-2018 BLUP model for SSC at harvest and at 10-weeks post storage at the LG1 locus. The linear regression and coefficient of determination is significant with p=0.012.



Key for labels:

1: Frostbite1
2: Malinda2
3: NWGreening1
4: GrimesRecom
5: NorthSpy1
6: Aspa2

Figure 2.7. LG13 SSC Haplotypes at Harvest vs. 10-Weeks Post Storage. Plot of the groups of haplotype means from the 2014-2018 BLUP model for SSC at harvest and at 10-weeks post storage at the LG13 locus. The linear regression and coefficient of determination is significant with $p=8.1x10^{-4}$.



Key for labels:

1: MinSFRecom
2: Duchess2
3: Keepsake2
4:1836SFRecom
5: NorthSpy2
6: Goodland2
7: US2Ark
8: NWGreening1
9: GoldenDel2
10: Sharon2

Table 2.1. SSC QTL Detected. Transformed Bayes factors (BF) for QTL detected by FlexQTL™ software on each linkage group for SSC at harvest, at 10-weeks post storage, and at 20-weeks post storage. Analyses were performed using BLUP adjusted SSC values for the following multi-year analyses: 2010-2012, 2014-2018, and 2010-2018. BF values 0-2 indicate hardly any evidence, 2-5 are positive, 5-10 are strong, and >10 are decisive (Bink et al. 2012). N is the number of individuals utilized in each analysis.

Year	Evaluation Period	Linkage Group																	N
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
2010	Harvest	-0.4	-0.3	0.8	0.3	0	0	0.3	0.1	0	0.1	0	-0.4	-0.1	0.7	0.3	-0.4	-0.6	94
2010	10-week	-0.1	0.2	0.6	0.7	-0.1	0.1	0.5	0.1	-0.2	0.2	0.1	-0.3	-0.1	0.1	-0.2	-0.5	0.1	39
2010	20-week	-0.5	0.3	-0.5	0	-0.6	-0.6	-0.2	-0.1	-0.2	-0.5	0.4	-0.5	-0.4	-0.2	-0.2	-0.6	-0.3	30
2011	Harvest	-0.4	1.3	3.3	-0.1	-0.3	-0.8	-0.8	0.5	-0.3	0.6	-0.6	0.2	0	0	-0.1	0.2	4.8	248
2011	10-week	0.8	1.8	1.4	0	0.6	0.6	0.9	2	0.1	-0.1	-0.5	-0.2	-0.2	-0.2	0.2	0.8	3	153
2011	20-week	1.8	5.9	0.3	0.3	-0.4	-0.5	-0.2	-0.1	0.9	0	-0.9	-0.5	-0.7	0	-0.4	4.1	0.3	187
2012	Harvest	-0.1	0.5	1.3	-0.5	1.1	-0.1	0.3	0.7	1.1	1.2	-0.4	-0.3	3.4	0	1.6	0.6	-0.4	226
2012	10-week	-0.4	-0.4	-0.3	-0.4	-0.8	-0.5	-0.6	-0.5	-0.6	-0.7	-0.8	-0.7	0.5	-0.9	0.4	-0.9	0	113
2012	20-week	2.7	1.5	-0.3	0.5	-0.2	-0.3	0.7	0.5	1.5	-0.1	-0.7	-0.5	1.2	0.9	0.5	0.6	0	151
2014	Harvest	0	1.3	0.1	-0.4	0.3	1.7	0.3	-0.5	5.2	-0.2	-0.1	-0.7	0.9	0	0.7	0.7	-0.3	229
2014	20-week	0.6	-0.1	0.2	-0.1	0.1	0.2	-0.6	-1	11.7	-0.5	-0.8	1.1	0.3	-0.1	-0.4	1	-0.6	209
2015	Harvest	0.5	6.8	-0.7	0.2	2	2.3	0.3	-0.3	2.3	2.2	0.5	-1.2	1	0.5	1.9	0.7	1.8	296
2015	10-week	4.8	0.4	0.5	0.6	-0.6	3.1	-0.5	-0.3	0.9	-0.2	-0.2	0.2	3.5	0.5	0	-0.2	-0.2	121
2016	Harvest	1.1	-0.3	0.2	-0.4	-0.1	-0.8	-1	-0.1	2.2	-0.6	-0.5	1	7	0.4	0.6	-0.1	-0.4	124
2016	10-week	1.8	1.1	0.5	1.5	0.3	-1.1	0.7	-0.2	0.9	0.8	-1	-0.3	1.6	1.8	2.3	1.2	-0.2	67
2017	Harvest	0.3	1.4	-1	-0.5	-0.8	-0.4	-0.3	1.1	2.8	-0.4	1.3	5.5	1.1	-1.2	0.2	7.1	5.8	302
2018	Harvest	0.3	-1.4	2.3	-0.5	-2.5	-1.5	-1.1	2	-1.6	-0.6	-1.4	-0.9	15.1	-1.9	1.2	-2.1	-1.4	407
2018	10-week	0.1	0	-0.3	0	1.6	0.7	4.9	-0.8	-0.5	1.3	-0.4	1.2	1.2	-0.6	-1.2	-0.3	1.9	216
2010-2012	Harvest	-0.2	0.4	0.9	-0.8	-1.4	-2.1	-0.4	0.8	-0.8	-1	-1.5	-1.1	2.2	0.1	-0.7	2.7	5.6	306
2010-2012	10-week	0.5	2.5	0.7	-0.1	-0.2	-0.2	2.2	2.7	-0.8	-0.3	0.6	0.2	1.8	-0.2	0.7	0	0.5	219
2010-2012	20-week	6.4	0.7	0.1	-0.3	-1.1	-0.3	1	1.3	2	-0.4	-1.3	-1	1.8	0.2	-0.1	3.8	0.5	257
2014-2018	Harvest	7.6	4.6	-1.7	-2.6	-3.1	-1.1	-2	0.3	3.4	2.4	-1.1	1.5	15.1	-2.2	7.5	6.9	0.2	942

Table 2.1. SSC QTL Detected Cont.

2014-2018	10-week	1.9	0.8	0.8	2	0.4	1.8	2.4	-1.2	-0.5	0.6	-1.4	2.1	7.1	0.6	1.2	4.2	0.7	342
2010-2018	Harvest	14.2	4.1	-2.8	-2.2	-2.6	-1.5	-2.5	0.2	3.3	1.1	-1.8	-1.5	12	-2.8	5.6	16	1.1	1044
2010-2018	10-week	1.1	2.1	0.3	-0.3	1	6	2.9	1.4	-0.9	1.9	-0.3	1.3	4.7	1.9	2.1	3.8	-0.9	554
2010-2018	20-week	11.1	0.2	0.4	-1.4	-0.5	0.2	1.3	1	7.4	-0.2	-1.2	-1.2	2.4	0.4	-0.6	4.4	1.7	344

Table 2.2. SSC QTL Detected Detailed. QTL regions on linkage groups (LGs) in centimorgans (cM) with a transformed Bayes factor (BF) > 5 identified in analyses of SSC in germplasm evaluated in individual years or combined years. The corresponding mode (cM location of the marker that was identified most often), and probability (higher value indicates higher likelihood that the QTL is real and significant) are also provided.

Year	Evaluation Period	LG	BF	QTL Region (cM)	QTL Mode (cM)	Probability
2014	Harvest	9	5.2	14-53	25	0.691
2014	20-week	9	11.7	9-46	25	1.016
2016	Harvest	13	7	17-68	44	0.97
2017	Harvest	12	5.5	1-33	26	0.649
2017	Harvest	16	7.1	2-14	7	0.846
2017	Harvest	17	5.8	5-16	12	0.077
2017	Harvest	17		31-49	36	0.783
2018	Harvest	13	15.1	62-68	68	0.996
2010-2012	Harvest	17	5.6	12-26	20	0.263
2010-2012	Harvest	17		28-32	31	0.058
2010-2012	Harvest	17		35-47	40	0.386
2010-2012	20-week	1	6.4	37-63	55	0.845
2014-2018	Harvest	1	7.6	42-58	52	0.855
2014-2018	Harvest	13	15.1	54-68	66	0.965
2014-2018	Harvest	15	7.5	28-45	36	0.831
2014-2018	Harvest	15		50-67	54	0.756
2014-2018	Harvest	16	6.9	1-12	7	0.795
2014-2018	10-week	13	7.1	52-57	56	0.061
2014-2018	10-week	13		60-68	67	0.803
2010-2018	Harvest	1	14.1	46-58	52	0.98
2010-2018	Harvest	13	12	50-67	56	0.987
2010-2018	Harvest	15	5.6	28-41	35	0.92
2010-2018	Harvest	15		50-72	61	0.864
2010-2018	Harvest	16	16	3-11	6	0.92
2010-2018	Harvest	16		13-15	14	0.051
2010-2018	10-week	6	6	33-73	69	0.813
2010-2018	20-week	1	11.1	34-63	54	1.068
2010-2018	20-week	9	7.4	1-62	31	1.058

Table 2.3. LG1 Locus Markers and Haplotypes. SNPs and marker states used to identify segregating haplotypes at the LG1 SSC locus in six families evaluated in 2014-2018. All six SNPs and their states for each parental haplotype are shown, along with the name designation for each haplotype.

		cM Position	51.28	52.13	52.98	56.21	56.37	56.57
Parent	Parental Hap	SNP ID / Haplotype	1051	2640	2292	1084	516	723
Honeycrisp	Keepsake	Frostbite1	B	B	A	B	A	B
Honeycrisp	MN1627	GrimesRecom	A	B	B	A	B	A
Minneiska	Honeycrisp	GrimesRecom	A	B	B	A	B	A
Minneiska	Minnewashta	NWGreening1	B	A	B	A	B	B
Minnewashta	State Fair	Malinda2	B	B	A	A	B	A
Minnewashta	MN1691	NWGreening1	B	A	B	A	B	B
MN1836	Keepsake	NorthSpy1	A	B	B	A	A	A
MN1836	State Fair	Malinda2	B	B	A	A	B	A
MN1915	Sweet16	NorthSpy1	A	B	B	A	A	A
MN1915	Minnewashta	Malinda2	B	B	A	A	B	A
MN1965	Ginger Gold	Aspa2	B	B	A	B	A	A
MN1965	Honeycrisp	Frostbite1	B	B	A	B	A	B
MN55	Honeycrisp	GrimesRecom	A	B	B	A	B	A
MN55	Monark	Malinda2	B	B	A	A	B	A
Wildung	Sharon	Frostbite1	B	B	A	B	A	B
Wildung	Fireside	Malinda2	B	B	A	A	B	A

The SNP ID numbers refer to the 20K index name, detailed information for each can be found in Appendix Table 1B.

SNP states A and B refer to the base associated with each SNP marker. A and B for each SNP are unique, and can be found for each marker in Appendix Table 1B.

Table 2.4. LG13 Locus Markers and Haplotypes. SNPs and marker states used to identify segregating haplotypes at the LG13 SSC locus in six families evaluated in 2014-2018. All 18 SNPs and their states for each parental haplotype are shown, along with the name designation for each haplotype.

		cM Position	53.38	53.49	53.59	53.59	53.77	55.61	56.78	57.47	57.89
Parent	Parental Hap	SNP ID / Haplotype	903	2845	670	476	1604	3237	2601	512	70
Honeycrisp	Keepsake	Keepsake2	A	A	A	B	B	A	B	A	B
Honeycrisp	MN1627	Duchess2	A	B	B	B	A	B	B	A	B
Minneiska	Honeycrisp	Keepsake2	A	A	A	B	B	A	B	A	B
Minneiska	Minnewashta	Goodland2	A	A	A	A	B	B	B	A	B
Minnewashta	State Fair	MinSFRecom	A	B	B	B	A	B	B	B	A
Minnewashta	MN1691	Goodland2	A	A	A	A	B	B	B	A	B
MN1836	Keepsake	Keepsake2	A	A	A	B	B	A	B	A	B
MN1836	State Fair	1836SFRecom	A	A	A	B	B	A	B	B	A
MN1915	Sweet16	NorthSpy2	A	B	B	B	B	A	B	A	B
MN1915	Minnewashta	Goodland2	A	A	A	A	B	B	B	A	B
MN1965	Ginger Gold	GoldenDel2	A	A	A	B	B	A	A	B	A
MN1965	Honeycrisp	Keepsake2	A	A	A	B	B	A	B	A	B
MN55	Honeycrisp	Keepsake2	A	A	A	B	B	A	B	A	B
MN55	Monark	US2Ark	A	B	B	B	A	B	B	A	B
Wildung	Sharon	Sharon2	B	A	B	A	B	A	B	A	B
Wildung	Fireside	NWGreening1	B	A	B	A	B	A	B	A	B

The SNP ID numbers refer to the 20K index name, detailed information for each can be found in Appendix Table 1B.

SNP states A and B refer to the base associated with each SNP marker. A and B for each SNP are unique, and can be found for each marker in Appendix Table 1B.

Table 2.4. LG13 Locus Markers and Haplotypes Cont.

		cM Position	58.58	59.49	60.25	61.31	63.01	63.01	63.10	64.85	65.94
Parent	Parental Hap	SNP ID / Haplotype	2932	574	522	404	1724	298	1043	439	567
Honeycrisp	Keepsake	Keepsake2	A	A	A	B	B	B	A	B	A
Honeycrisp	MN1627	Duchess2	A	A	B	A	B	A	A	B	B
Minneiska	Honeycrisp	Keepsake2	A	A	A	B	B	B	A	B	A
Minneiska	Minnewashta	Goodland2	B	A	B	A	B	A	A	B	A
Minnewashta	State Fair	MinSFRecom	A	A	A	A	B	A	A	B	A
Minnewashta	MN1691	Goodland2	B	A	B	A	B	A	A	B	A
MN1836	Keepsake	Keepsake2	A	A	A	B	B	B	A	B	A
MN1836	State Fair	1836SFRecom	A	A	A	A	A	B	B	B	B
MN1915	Sweet16	NorthSpy2	A	A	A	B	B	B	A	B	A
MN1915	Minnewashta	Goodland2	B	A	B	A	B	A	A	B	A
MN1965	Ginger Gold	GoldenDel2	A	A	A	B	B	B	A	A	B
MN1965	Honeycrisp	Keepsake2	A	A	A	B	B	B	A	B	A
MN55	Honeycrisp	Keepsake2	A	A	A	B	B	B	A	B	A
MN55	Monark	US2Ark	B	A	B	A	B	A	A	B	A
Wildung	Sharon	Sharon2	A	A	B	A	B	A	A	B	B
Wildung	Fireside	NWGreening1	B	B	A	A	A	B	B	A	B

The SNP ID numbers refer to the 20K index name, detailed information for each can be found in Appendix Table 1B.

SNP states A and B refer to the base associated with each SNP marker. A and B for each SNP are unique, and can be found for each marker in Appendix Table 1B.

Table 2.5. SSC Loci Parental Haplotypes. Haplotypes (Hap) for the linkage group (LG) 1 and LG13 SSC loci for parents of the six families evaluated in 2014-2018. Haplotypes for the LG1 and LG13 loci are named according to the most distant genotyped ancestor (see Table 2.6).

Parent	LG1		LG13	
	Hap1	Hap2	Hap1	Hap2
Honeycrisp	Frostbite1	GrimesRecom	Keepsake2	Duchess2
Minneiska	GrimesRecom	Malinda2	Goodland2	Keepsake2
Minnewashta	Malinda2	NWGreening1	MinSFRecom	Goodland2
MN1836	Malinda2	NorthSpy1	Keepsake2	1836SFRecom
MN1915	NorthSpy1	Malinda2	NorthSpy2	Goodland2
MN1965	Aspa2	Frostbite1	GoldenDel2	Keepsake2
MN55	NWGreening1	GrimesRecom	US2Ark	Keepsake2
Wildung	Frostbite1	Malinda2	Sharon2	NWGreening1

Table 2.6. SSC Loci Haplotype Names. Haplotype name, associated linkage group (LG), and the ancestral origin of all SSC haplotypes. Haplotype names that end with a 1 indicate that the haplotype came from the unknown parent of the cultivar designated as maternal, and haplotypes that end with 2 indicate origin from the unknown parent designated as paternal.

Haplotype Name	LG	Origin of Name
Aspa2	1	Unknown parent of 'Aspa'
Frostbite1	1	Unknown parent of 'Frostbite'
GrimesRecom	1	Recombination between the haplotypes of 'Grimes Golden', inherited by 'Golden Delicious'
Malinda2	1	Unknown parent of 'Malinda'
NorthSpy1	1	Unknown parent of 'Northern Spy'
NWGreening1	1	Unknown parent of 'Northwest Greening'
1836SFRecom	13	Recombination between the haplotypes of 'State Fair', inherited by MN1836
Duchess2	13	Unknown parent of 'Duchess of Oldenburg'
Goodland2	13	Unknown parent of 'Goodland'
GoldenDel2	13	Unknown parent of 'Golden Delicious'
Keepsake2	13	Unknown parent of 'Keepsake'
MinSFRecom	13	Recombination between the haplotypes of 'State Fair', inherited by 'Minnewashta'
NorthSpy2	13	Unknown parent of 'Northern Spy'
NWGreening1	13	Unknown parent of 'Northwest Greening'
Sharon2	13	Unknown parent of 'Sharon'
US2Ark	13	Unknown parent of US2Ark, haplotypes for US2Ark at this locus are homozygous

Table 2.7. Chi-squared Values for each SSC Locus. Chi-squared values for diplotype segregation distortion for each family for the LG1 and LG13 SSC loci. N is the number of individuals in each family that were evaluated at least once between 2014 and 2018, not including individuals with recombinant haplotypes. Bolded values indicate significant distortion ($p < 0.05$).

Family	LG1	N (LG1)	LG13	N (LG13)
'Honeycrisp' x 'Minnewashta'	0.214	115	**0.002**	112
'Honeycrisp' x MN1836	0.284	40	0.158	40
'Honeycrisp' x MN1915	**0.048**	45	0.572	38
'Minneiska' x MN1965	**0.026**	44	0.164	35
'Minneiska' x 'MN55'	0.397	108	0.910	104
'Minneiska' x 'Wildung'	0.599	64	0.133	60

Table 2.8. SSC at Harvest ANOVAs by Family. ANOVA p-values for variation due to diplotype effects within each family for SSC loci on LG1 and LG13 on 2014-2018 BLUP values for SSC at harvest. N is the number of individuals in each family that were evaluated at least once between 2014 and 2018, not including individuals with recombinant haplotypes. Significant p-values ($0.01<p<0.05$) are bolded.

Family	LG1	N (LG1)	LG13	N (LG13)
'Honeycrisp' x 'Minnewashta'	0.131	113	**0.042**	112
'Honeycrisp' x MN1836	0.266	40	0.610	40
'Honeycrisp' x MN1915	0.124	45	0.841	38
'Minneiska' x MN1965	0.755	44	0.881	35
'Minneiska' x 'MN55'	0.168	107	0.418	104
'Minneiska' x 'Wildung'	0.511	64	0.991	60

Table 2.9. SSC at 10-Weeks Post Storage ANOVAs by Family. ANOVA p-values for variation due to diplotype effects within each family for SSC loci on LG1 and LG13 on 2014-2018 BLUP values for SSC at 10-weeks post storage. N is the number of individuals in each family that were evaluated at least once between 2014 and 2018, not including individuals with recombinant haplotypes. Nearly significant p-values ($0.05<p<0.10$) are bolded.

Family	LG1	N (LG1)	LG13	N (LG13)
'Honeycrisp' x 'Minnewashta'	0.581	69	0.229	72
'Honeycrisp' x MN1836	0.828	15	0.833	16
'Honeycrisp' x MN1915	**0.054**	15	0.104	14
'Minneiska' x MN1965	0.251	8	0.582	7
'Minneiska' x 'MN55'	0.849	43	0.821	45
'Minneiska' x 'Wildung'	0.364	21	0.366	18

Table 2.10. SSC Difference between Harvest and 10-Weeks ANOVAs by Family. ANOVA p-values for variation due to diplotype effects within each family for SSC loci on LG1 and LG13 on 2014-2018 BLUP values for the difference between harvest and 10-weeks post storage. N is the number of individuals in each family that were evaluated at least once between 2014 and 2018, not including individuals with recombinant haplotypes. Nearly significant p-values ($0.05<p<0.10$) are bolded.

Family	LG1	N (LG1)	LG13	N (LG13)
'Honeycrisp' x 'Minnewashta'	0.700	69	0.573	71
'Honeycrisp' x MN1836	0.879	15	0.669	16
'Honeycrisp' x MN1915	0.215	15	**0.089**	14
'Minneiska' x MN1965	0.424	8	0.770	7
'Minneiska' x 'MN55'	0.587	43	**0.091**	45
'Minneiska' x 'Wildung'	0.380	21	0.567	18

Table 2.11. SSC LG1 Locus Diplotypes. Means of all individuals having the same diplotype at the LG1 locus in the six key families for soluble solids content (SSC, °Brix) at harvest and at 10-weeks post storage based on the adjusted values from the 2014-2018 BLUP model. ANOVA indicated significant variation among diplotypes at harvest ($p= 5.7x10^{-6}$) and at 10-weeks post storage ($p= 1.1x10^{-4}$). The group column designates statistically significant groupings based on a Tukey's HSD test (p= 0.05); diplotypes with different letters are statistically significant. N= number of individuals in each diplotype group.

Hap1	Hap2	SSC at Harvest				SSC at 10-Weeks Post Storage			
		Group	Mean (°Brix)	Std Error	N	Group	Mean (°Brix)	Std Error	N
GrimesRecom	GrimesRecom	A	14.2	0.3	24	A B	15.0	0.6	10
Aspa2	GrimesRecom	A	14.2	0.4	16	A B	14.3	1.3	2
Frostbite1	GrimesRecom	A	14.2	0.3	24	A	15.4	0.5	11
Aspa2	NWGreening1	A	14.0	0.4	16	A B	15.7	0.9	4
Frostbite1	NorthSpy1	A	13.8	0.3	23	A B	15.3	0.7	7
Malinda2	NWGreening1	A	13.6	0.2	43	A	15.1	0.5	15
GrimesRecom	NorthSpy1	A B	13.6	0.4	15	A B	14.7	0.8	5
GrimesRecom	Malinda2	A	13.4	0.1	109	A B	13.9	0.3	48
GrimesRecom	NWGreening1	A B	13.4	0.2	44	A B	13.5	0.4	24
Frostbite1	NWGreening1	A B	13.1	0.2	49	A B	13.1	0.5	15
Frostbite1	Malinda2	B	12.5	0.2	50	B	13.0	0.3	30

Table 2.12. SSC LG1 Locus Haplotypes. Means of all individuals having the same haplotype at the LG1 locus in the six key families for soluble solids content (SSC, °Brix) at harvest and at 10-weeks post storage based on the adjusted values from the 2014-2018 BLUP model. ANOVA indicated significant variation among haplotypes at harvest ($p= 5.6x10^{-4}$) and at 10-weeks post storage ($p= 0.088$). The group column designates statistically significant groupings based on a Tukey's HSD test ($p= 0.05$); haplotypes with different letters are statistically significant. N= number of individuals in each haplotype group.

Haplotype	SSC at Harvest				SSC at 10-Weeks Post Storage			
	Group	Mean (°Brix)	Std Error	N	Group	Mean (°Brix)	Std Error	N
Aspa2	A	14.1	0.3	32	A	15.2	0.8	6
NorthSpy1	A B	13.7	0.2	38	A	15.1	0.6	12
GrimesRecom	A	13.7	0.1	256	A	14.2	0.2	110
NWGreening1	A B	13.4	0.1	152	A	14.0	0.3	58
Malinda2	B	13.3	0.1	202	A	13.8	0.2	93
Frostbite1	B	13.2	0.1	146	A	13.7	0.2	63

Table 2.13. SSC LG13 Locus Diplotypes. Means of all individuals having the same diplotype at the LG13 locus in the six key families for soluble solids content (SSC, °Brix) at harvest and at 10-weeks post storage based on the adjusted values from the 2014-2018 BLUP model. ANOVA indicated significant variation among diplotypes at harvest ($p= 7.4x10^{-10}$) and at 10-weeks post storage ($p= 1.7x10^{-5}$). The group column designates statistically significant groupings based on a Tukey's HSD test (p= 0.05); diplotypes with different letters are statistically significant. N= number of individuals in each diplotype group.

Hap1	Hap2	SSC at Harvest				SSC at 10-Weeks Post Storage			
		Group	Mean (°Brix)	Std Error	N	Group	Mean (°Brix)	Std Error	N
GoldenDel2	Goodland2	A	14.2	0.4	14	A B C	15.6	1.2	2
Goodland2	NWGreening1	A	14.1	0.3	20	A B C	14.3	0.8	5
Goodland2	Sharon2	A	14.1	0.3	18	A B	15.7	0.8	5
Keepsake2	US2Ark	A	14.1	0.3	25	A B	14.9	0.5	12
Keepsake2	Sharon2	A B	14.1	0.5	8	A B C	14.0	1.7	1
Keepsake2	NWGreening1	A	14.0	0.4	14	A B C	14.8	0.6	7
Duchess2	Keepsake2	A	13.9	0.4	11	A B C	14.3	0.9	2
GoldenDel2	Keepsake2	A B	13.9	0.5	7	-	-	-	-
Duchess2	NorthSpy2	A B	13.7	0.6	6	A B C	15.5	0.9	4
Duchess2	Goodland2	A	13.6	0.3	25	A B C	14.3	0.5	11
Keepsake2	Keepsake2	A	13.6	0.2	39	A	15.1	0.5	10
1836SFRecom	Duchess2	A B	13.5	0.4	15	A B	15.2	0.7	6
Keepsake2	NorthSpy2	A B	13.5	0.4	12	A B C	14.0	0.8	5
Goodland2	Keepsake2	A	13.4	0.2	66	A B C	14.0	0.3	34
Goodland2	US2Ark	A	13.4	0.3	29	A B	14.6	0.4	15
1836SFRecom	Keepsake2	A B	13.2	0.6	5	A B C	14.3	1.2	4
Duchess2	MinSFRecom	B	12.1	0.2	36	B C	12.8	0.4	22
Keepsake2	MinSFRecom	B	12.1	0.2	39	C	12.5	0.3	27

Table 2.14. SSC LG13 Locus Haplotypes. Means of all individuals having the same haplotype at the LG13 locus in the six key families for soluble solids content (SSC, °Brix) at harvest and at 10-weeks post storage based on the adjusted values from the 2014-2018 BLUP model. ANOVA indicated significant variation among haplotypes at harvest ($p= 1.4x10^{-13}$) and at 10-weeks post storage ($p= 1.4x10^{-6}$). The group column designates statistically significant groupings based on a Tukey's HSD test ($p= 0.05$); haplotypes with different letters are statistically significant. N= number of individuals in each haplotype group.

Haplotype	SSC at Harvest				SSC at 10-Weeks Post Storage			
	Group	Mean (°Brix)	Std Error	N	Group	Mean (°Brix)	Std Error	N
Sharon2	A B	14.1	0.3	26	A	15.4	0.7	6
NWGreening1	A	14.1	0.3	34	A	14.6	0.5	12
GoldenDel2	A B	14.1	0.3	21	A B	15.6	1.3	2
US2Ark	A B	13.7	0.2	54	A	14.7	0.3	27
Goodland2	A B	13.7	0.1	172	A	14.3	0.2	72
NorthSpy2	A B	13.6	0.4	18	A B	14.7	0.6	9
1836SFRecom	A B	13.4	0.3	20	A	15.0	0.6	8
Keepsake2	A B	13.4	0.1	265	A	14.0	0.2	112
Duchess2	B	13.1	0.2	93	A	13.8	0.3	47
MinSFRecom	C	12.1	0.2	75	B	12.6	0.3	49

Table 2.15. SSC Difference between Harvest and 10-Weeks for LG1 Haplotypes. Means of all individuals having the same haplotype at the LG1 locus in the six key families for soluble solids content (SSC, °Brix) difference between harvest and 10-weeks based on the adjusted values from the 2014-2018 BLUP model. ANOVA did not indicate variation among haplotypes (p= 0.113). The group column designates statistically significant groupings based on a Tukey's HSD test (p= 0.05); haplotypes with different letters are statistically significant. N= number of individuals in each haplotype group.

Haplotype	SSC Difference Between Harvest and 10-Weeks			
	Group	Mean (°Brix)	Std Error	N
NorthSpy1	A	1.9	0.3	12
Aspa2	A	1.6	0.5	6
Frostbite1	A	1.0	0.1	63
GrimesRecom	A	1.0	0.1	111
Malinda2	A	0.9	0.1	93
NWGreening1	A	0.9	0.2	58

Table 2.16. SSC Difference between Harvest and 10-Weeks for LG13 Haplotypes. Means of all individuals having the same haplotype at the LG13 locus in the six key families for soluble solids content (SSC, °Brix) difference between harvest and 10-weeks based on the adjusted values from the 2014-2018 BLUP model. ANOVA indicated significant variation among haplotypes (p= 0.0185). The group column designates statistically significant groupings based on a Tukey's HSD test (p= 0.05); haplotypes with different letters are statistically significant. N= number of individuals in each haplotype group.

Haplotype	SSC Difference Between Harvest and 10-Weeks			
	Group	Mean (°Brix)	Std Error	N
1836SFRecom	A	1.7	0.4	8
NorthSpy2	A	1.7	0.4	9
GoldenDel2	A	1.6	0.8	2
Sharon2	A	1.6	0.5	6
NWGreening1	A	1.2	0.3	12
US2Ark	A	1.2	0.2	27
Duchess2	A	1.1	0.2	47
Keepsake2	A	1.0	0.1	111
Goodland2	A	0.9	0.1	72
MinSFRecom	A	0.4	0.2	48

CHAPTER 3: GENETIC COMPONENTS OF TITRATABLE ACIDITY CONTENT IN IMPORTANT UNIVERSITY OF MINNESOTA APPLE GERMPLASM

Introduction

Consumers base their liking of a fresh fruit product on many factors. These factors include outward appearance (color, shape, size) and eating characteristics (firmness, texture, flavor) (Harker et al. 2008, Hamadziripi et al. 2014). Many fresh fruit traits in apple including skin color, uniform shape and size, and resistance to bruising have been vastly improved over the last several decades. Due to the highly improved visual appearance in the market, apple breeders are now focusing new cultivars on having more desirable eating characteristics such as texture, sweetness, sourness, and flavor.

Apple fruit are sold to the final consumer over a long period. While in the case of a pick-your-own orchard or for local markets the fruit are sold and consumed at harvest or within a few weeks, a large proportion of fruit are harvested, placed in cold storage, and slowly removed from storage to sell to the consumer throughout the year. This long marketing period poses a challenge for apple breeders when selecting desirable fruit acidity for the fresh fruit market.

Acidity levels tend to drop after long periods of refrigerated storage because malic acid, the main acid component in apple fruit, is used by the cells to maintain respiration (Kidd et al. 1951). This process of malic acid consumption by the cells occurs at varying rates for different apple cultivars (Kidd et al. 1951), therefore requiring extensive storage evaluation for cultivar development. Some apple cultivars are high in acid at harvest and,

therefore, unfit for the markets in which apples are consumed shortly after harvest. Conversely, some of those same cultivars that have high acid content at harvest may undergo a significant decrease in acid content during long periods of storage so that they taste insipid and lack flavor complexity. Identifying new cultivars that maintain desirable acidity throughout maturity and storage is a target for apple breeders.

Previous genetic architecture studies have used sensory categorizations of acidity and sweetness (Rymenants et al. 2020) and measured TA values (Verma et al. 2019) for apple.

Titratable acidity (TA) measurements are an optimal method for using instruments to describe acidity perception for several reasons. TA correlates highly with sensory perception of apple sourness (R^2 as high as 0.93) and is a better estimator of sensory sourness than pH (Harker et al. 2002), suggesting that TA is a more accurate quantification of apple juice acidity than pH. The diprotic nature of malic acid adds an additional layer of acidity perception that cannot be captured by pH values.

The research described in this chapter seeks to further understand the genetic components of variation for apple fruit TA content at harvest and after storage, as well as the difference between harvest and 10-weeeks post storage in germplasm from the University of Minnesota (UMN) apple breeding program. A better understanding of the genetic components of apple fruit acidity will allow parent and seedling selection for these traits, and make the breeding process more efficient.

Materials and Methods

Plant Material

Plant material used for this chapter is the same material described in Chapter 2.

Phenotyping Protocol

Apple fruit harvest and apple juice sample preparations were performed as described in Chapter 2. After thawing frozen juice to room temperature, juice acidity was measured by titration using an automatic titrator (Mettler Toledo T50 and InMotion Flex Autosampler). Each sample had two 5 mL samples placed into sampling cups. The samples were automatically diluted by the titrator with 50 mL of deionized water. While the sample was stirred by the machine, it automatically measured pH and subsequently added 0.1 M sodium hydroxide solution in incremental volumes until the homogenized sample achieved a pH of 8.2. The volume of 0.1 M sodium hydroxide used for the titration was then transformed into the relative amount of acid in the sample, as determined by software on the automatic titrator, using g/L malic acid as the reference.

QTL Mapping

QTL mapping strategies and software used for this chapter are the same as those described in Chapter 2.

QTL mapping was performed using observed and adjusted Best Linear Unbiased Prediction (BLUP) TA values.

The three BLUP models described in Chapter 2 (2010-2012, 2014-2018, and 2010-2018) were applied to the TA dataset. The BLUP models were utilized for QTL mapping in the same methods described in Chapter 2.

Haplotyping Protocol

Individuals sharing haplotypes with the same name (within this study, and across the Verma et al. 2019 study) have been confirmed to share the haplotypes by both state and descent. Haplotypes that are identical in state to the Verma et al. (2019) haplotypes that could not be positively identified as identical by descent are labeled with an asterisk. For haplotypes at LG8 and LG16 that were not described in Verma et al. (2019), the haplotype naming protocol is the same as the protocol used for Chapter 2; haplotypes are named based upon the most distant genotyped ancestor that could be positively identified based on genetic data (Table 3.5).

Haplotypes at the LG1 locus are defined and named the same as in Chapter 2.

Statistical Analyses of Diplotype and Haplotype Effects and Segregation Distortion

Statistical analyses and characterization of individuals used in this chapter are the same as those used in Chapter 2.

Chi-squared tests were performed for each family at each QTL to determine if segregation distortion based on predicted Mendelian outcomes was present.

ANOVA tests were performed on each family for each QTL to evaluate if variation among diplotypes is significant. ANOVA tests were also performed for the combined six families for haplotype effects and diplotype effects at each locus, and a Tukey's HSD test was performed when the ANOVA is significant ($p<0.10$) or nearly significant. A Tukey's HSD test is performed in order to identify which, if any, groups of diplotypes or haplotypes have statistically significant differences among each other.

In addition to evaluating TA at harvest and at 10 weeks post storage, the difference between harvest and 10-weeks post storage was calculated for each individual. ANOVAs and comparisons of means using Tukey's HSD tests were performed to determine the effect of haplotypes on the difference between harvest and 10-weeks post storage.

An additional statistical analysis was performed for the 'Minneiska' x 'MN55' family, which segregated for diplotype effects for TA content at harvest at both the LG1 and LG16 loci. This family was larger and thus provided greater statistical power for multi-QTL evaluation. Individuals were evaluated by their combined states at the LG1 and LG16 loci (forming one set of sixteen diplotype groups of individuals opposed to two sets of four groups of individuals if each locus is observed independently) to determine if the combined LG1-LG16 states exhibit statistically significant variation among groups for TA at harvest.

Results

Variation for Titratable Acidity Content

TA data were collected over eight harvest seasons between 2010 and 2018. Figure 3.1 shows the distributions of observed TA values for all individuals obtained in each year. 1095 individuals were evaluated at least once during these eight years. The distributions of TA values is different each year, which may be due to germplasm differences and/or environmental differences (Appendix Table 1A). The same set of germplasm was evaluated in years 2010, 2011, and 2012, but their median TA values vary. The germplasm evaluated in 2014, 2015, 2016, and 2017 have similar distributions with the exception of 2016, which included a smaller number of individuals. The germplasm evaluated in 2018 contains some of the germplasm from the previous years (2014-2017), and also includes several families that were exclusively observed in 2018, and has a distribution that appears to be bimodal.

The distributions of adjusted values for all individuals for each of the multi-year BLUP models (2010-2018, 2010-2012, and 2014-2018) can be found on Figure 3.2. The 2010-2018 and 2014-2018 BLUP models have similar distributions, whereas the 2010-2012 BLUP model has a slightly lower median (Figure 3.2). While the 2010-2018 BLUP model closely resembles the 2014-2018 BLUP model (Figure 3.2), the 2014-2018 BLUP model was selected for haplotype effect estimates because it has more consistent TA distributions across years, therefore minimizing variation within the dataset by removing the data from 2010-2012 (Figure 3.1).

The distributions of the 2014-2018 BLUP TA values at harvest for all individuals in each of the six families of focus can be found on Figure 3.3. Distributions range from approximately normal (‘Honeycrisp’ x ‘Minnewashta’), to bimodal (‘Minneiska’ x ‘MN55’), to flat and possibly multimodal (‘Honeycrisp’ x MN1836, ‘Honeycrisp’ x

MN1915, ‘Minneiska’ x MN1965, and ‘Minneiska’ x ‘Wildung’). Family means are relatively similar, ranging from 5.51 g/L (‘Minneiska’ x MN1965) to 7.01 g/L (‘Minneiska’ x ‘MN55’). Within family variance tended to vary with family size, and families with a larger number of individuals tended to have smaller variance. Distributions of the 2014-2018 BLUP TA values at 10-weeks post storage range from approximately normal (‘Honeycrisp’ x MN1836), to bimodal (‘Minneiska’ x ‘MN55’), possibly flat and multimodal (‘Honeycrisp’ x ‘Minnewashta’, ‘Honeycrisp’ x MN1915, ‘Minneiska’ x MN1965, ‘Minneiska’ x ‘Wildung’) (Figure 3.4). Family means at 10-weeks post storage are similar, ranging from 2.97 g/L (‘Minneiska’ x MN1965) to 4.97 g/L (‘Honeycrisp’ x MN1836). Within family variance was relatively consistent across families, with the exception of ‘Minneiska’ x ‘Wildung’, which had a much higher variance than the other families.

The 2014-2018 BLUP TA values for each individual within the six families for harvest and 10-weeks post storage were plotted (Figure 3.5). An overall coefficient of determination was R^2=0.41 (p= $8.1x10^{-25}$). The slope for the linear line of best fit is 0.44 and the y intercept is 0.99.

QTL Mapping

Linkage groups (LGs) containing quantitative trait loci (QTLs) for TA that were detected and mapped using FlexQTL™ in each analysis are summarized in Table 3.1, including total sample size for each evaluation period-year-model. For each linkage group with a QTL, the transformed Bayes Factor (BF) defined as [2*ln][Bayes Factor] is

presented. Bink et al. (2012) considered 5<BF<10 and BF>10 as “strong” and “decisive” evidence, respectively, for QTL in genomic regions indicated in FlexQTL™ analyses.

The region (in cM) on significant LGs (BF>5) are defined in Table 3.2. Bink et al. (2012) considered 5<BF<10 and BF>10 as “strong” and “decisive” evidence, respectively, for QTL in genomic regions indicated in FlexQTL™ analyses.

The regions on LGs that were consistent (in multiple single years/models and/or multi-year evaluation periods) and significant (have BFs>5) were used in further analyses are described below:

A QTL on LG1 was detected for variation in TA content in ten analyses (Table 3.1). Four of these QTL had a 5<BF<10 (strong evidence, analyses: 2014 at 20-weeks post storage, 2017 at harvest, 2018 at 10-weeks post storage, 2010-2018 BLUP at 10-weeks post storage). Six of these QTL had a BF>10 (decisive evidence, analyses: 2014 at harvest, 2015 at harvest, 2018 at harvest, 2014-2018 BLUP at harvest, 2010-2018 BLUP at harvest and 2010-2018 BLUP at 20 weeks post storage). The QTL region consistently identified spans 34 cM to 63 cM (Table 3.2).

A QTL on LG8 was detected for variation in TA content in twelve analyses (Table 3.1). Five of these QTL had a 5<BF<10 (strong evidence, analyses: 2014 at harvest, 2015 at 10-weeks post storage, 2010-2012 BLUP at 10-weeks post storage, 2010-2012 BLUP at 20-weeks post storage, and 2010-2018 BLUP at 20-weeks post storage). Seven of these QTL had a BF>10 (decisive evidence, analyses: 2017 at harvest, 2018 at harvest, 2018 at 10-weeks post storage, 2014-2018 BLUP at harvest, 2014-2018 BLUP at 10-weeks post storage, 2010-2018 BLUP at harvest, 2010-2018 BLUP at 10-

weeks post storage). The QTL region consistently identified spans 25 cM to 36 cM (Table 3.2).

A QTL on LG16 was detected for variation in TA content in fifteen analyses (Table 3.1). Four of these QTL had a 5<BF<10 (strong evidence, analyses: 2010 at harvest, 2015 at 10-weeks post storage, 2017 at harvest, 2010-2012 BLUP at 10-weeks post storage). Eleven of these QTL had a BF>10 (decisive evidence, analysis: 2011 at harvest, 2011 at 10-weeks post storage, 2014 at harvest, 2015 at harvest, 2018 at harvest, 2010-2012 BLUP at harvest, 2014-2018 BLUP at harvest, 2014-2018 BLUP at 10-weeks post storage, 2010-2018 BLUP at harvest, 2010-2018 BLUP at 10-weeks post storage, and 2010-2018 BLUP at 20-weeks post storage). The QTL region consistently identified spans 1 cM to15 cM (Table 3.2).

Haplotyped Loci

For the QTLs identified on LG1, LG8, and LG16, the minimum number of informative SNPs needed to define the segregating haplotypes at each locus was determined.

The LG1 TA haplotypes are defined and named the same as the LG1 SSC haplotypes referenced in Chapter 2, and the naming rationale is found in Table 3.3 and the SNP states are found in Table 3.4.

The LG8 locus was thinned to fourteen markers from 24.60 cM to 36.23 cM, which ranged 7,968,643 bp to 12,838,023 bp in reference to the ‘Golden Delicious’ doubled-haploid genome (GDDH13) (Table 3.5). Eight markers used in the Verma et al.

(2019) study were retained after thinning. The full range of the region on LG8 detected by Verma et al. (2019) was retained, with some of the markers within the region being removed due to being uninformative or redundant in this germplasm set. This region is consistent with reported literature (Jia et al. 2018, Verma et al. 2019, Rymenants et al. 2020). The naming rationale for the haplotypes can be found in Table 3.3.

The LG16 locus was thinned to six markers from 8.43 cM to 10.43 cM, which ranged 3,029,704 bp to 3,409,834 bp in reference to the GDDH13 (Table 3.6). One marker used by Verma et al. (2019) was removed due to redundancy. All markers retained were utilized in the Verma et al. (2019) study. Two candidate genes for the *Ma* locus are proposed by Khan et al. (2013) and are located in this region: MDP0000244249 and MDP0000252114 are both flanked by the third and fourth markers (IDs 2043 and 647, see Appendix Table 1B for marker details) in this region. The naming rationale for the haplotypes can be found in Table 3.3.

The parental haplotypes and associated names for all TA loci in the six key families can be found in Table 3.7.

Segregation Distortion

Chi-squared tests were performed for each family at each locus to determine if expected segregation ratios of diplotypes were observed (Table 3.8). In addition to the distortion identified at LG1 (described in Chapter 2), segregation distortion was also detected for the QTL region for TA on LG16 in ‘Honeycrisp’ x MN1836 ($p= 0.022$), ‘Honeycrisp’ x MN1915 ($p= 0.002$), and ‘Minneiska’ x ‘MN55’ ($p= 0.022$). Segregation

distortion was detected for the QTL region for TA on LG8 in ‘Honeycrisp’ x ‘Minnewashta’ ($p= 0.040$) and ‘Minneiska’ x ‘Wildung’ ($p= 0.032$). Hypotheses for causes of distortion are described later.

Diplotype Associations for LG16, LG8, and LG1 with Family as a Factor

Families were investigated individually for at each locus to determine if they segregated for significant diplotype effects for TA content at harvest, at 10-weeks post storage, and for the difference in mean acid content between harvest and 10-weeks post storage. The ANOVAs for each family at harvest can be found in Table 3.9, ANOVAs for 10-weeks post storage can be found in Table 3.10, and ANOVAs for the difference between harvest and 10-weeks post storage can be found in Table 3.11.

ANOVAs indicated significant variation among diplotypes for all families for at least one locus at harvest (Table 3.9), four families exhibited significant variation among diplotypes for at least one locus at 10-weeks post storage (Table 3.10), and three families exhibited significant variation among diplotypes at the LG1 locus for the difference between harvest and 10-weeks post storage (Table 3.11).

Diplotype Associations for LG16 at Harvest and at 10-Weeks Post Storage

Diplotypes at the LG16 locus were evaluated (Table 3.12). ANOVA utilizing individuals from all six families detected significant variation among LG16 diplotypes for TA at harvest ($p= 9.97x10^{-7}$) and at 10-weeks post storage ($p= 2.86x10^{-5}$).

Four statistically significant diplotype groupings at harvest were determined using a Tukey's HSD test (Table 3.12). The TA diplotype mean values at harvest ranged from 7.97 g/L (B*C*) to 4.23 g/L (JMinHCRecom).

Two statistically significant diplotype groupings at 10-weeks post storage were identified using a Tukey's HSD test (Table 3.12). The TA diplotype mean values at 10-weeks post storage ranged 5.76 g/L (B*MinHcRecom) to 2.43 g/L ($B_{F2/Jt}$MinHcRecom).

Haplotype Associations for LG16 at Harvest and at 10-Weeks Post Storage

Haplotypes at the LG16 locus were evaluated (Table 3.13). ANOVA utilizing individuals from all six families detected significant variation among LG16 haplotypes for TA at harvest ($p= 8.26x10^{-6}$) and at 10-weeks post storage ($p= 0.0278$).

Three statistically significant haplotype groupings were determined for TA at harvest using a Tukey's HSD test (Table 3.13). The TA haplotype mean values at harvest ranged from 6.79 g/L (C*) to 4.69 g/L (J). The lowest haplotype mean (J) was 1.21 g/L lower than the next lowest haplotype ($B_{F2/Jt}$, 5.90 g/L), indicating a substantial difference in TA values. Additionally, the four highest haplotypes (C*, A, D, and B*) had very similar effects on TA at harvest maturity, ranging from 6.78 g/L to 6.54 g/L.

A Tukey's HSD test did not indicate any haplotypes for TA at 10-weeks post storage being statistically different from the others (Table 3.13). The TA haplotype mean values at 10-weeks post storage ranged from 4.19 g/L (A) to 2.83 g/L (J).

Diplotype Associations for LG8 at Harvest and at 10-Weeks Post Storage

Diplotypes at the LG8 locus were evaluated (Table 3.14). ANOVA utilizing individuals from all six families detected significant variation among LG8 diplotypes for TA at harvest ($p= 4.4x10^{-9}$) and at 10-weeks post storage (p= 0.0004).

Three statistically significant diplotype groupings were determined for TA at harvest using Tukey's HSD test (Table3.14). The TA diplotype mean values at harvest ranged from 7.66 g/L (JUS1Ark1) to 2.75 (GI). The diplotypes GG (mean= 3.74 g/L) and GI (mean= 2.75 g/L) were drastically lower than the other diplotypes, with the next lowest diplotype having a mean of 5.79 g/L.

Two statistically significant diplotype groupings were determined for TA at 10-weeks post storage using a Tukey's HSD test (Table 3.14). The TA diplotype mean values at 10-weeks post storage ranged from 5.29 g/L (bDuchess1) to 1.23 g/L (GI). Similar to the results for LG8 diplotypes at harvest, the diplotypes GG (10-week mean= 1.69 g/L) and GI (10-week mean= 1.23 g/L) were substantially lower at 10-weeks post storage than other diplotypes. When excluding the GG and GI diplotypes, the lowest mean diplotype value for 10-weeks post storage was 3.24 g/L (bG).

Haplotype Associations for LG8 at Harvest and at 10-Weeks Post Storage

Haplotypes at the LG8 locus were evaluated (Table 3.15). ANOVA utilizing all individuals from all six families detected significant variation among LG8 haplotypes for TA at harvest ($p= 5.1x10^{-7}$) and at 10-weeks post storage (p= 0.0037).

Two statistically significant haplotype groupings were determined for TA at harvest using a Tukey's HSD test (Table 3.15). The TA haplotype mean values at harvest ranged from 7.45 g/L (US1Ark1) to 5.74 g/L (G).

Two statistically significant haplotype groupings for TA at 10-weeks post storage were determined using a Tukey's HSD test (Table 3.15). The TA haplotype mean values at 10-weeks post storage ranged from 4.91 g/L (Duchess1) to 3.38 g/L (G).

Diplotype Associations for LG1 at Harvest and at 10-Weeks Post Storage

Diplotypes at the LG1 locus were evaluated (Table 3.16). ANOVA utilizing individuals from all six families detected significant variation among LG1 diplotypes for TA at harvest ($p= 1.3x10^{-7}$) and at 10-weeks post storage ($p= 0.0282$).

Three statistically significant diplotype groupings were determined for TA at harvest using a Tukey's HSD test (Table 3.16). The TA diplotype mean values at harvest ranged from 8.05 g/L (GrimesRecomGrimesRecom) to 5.37 g/L (Frostbite1Malinda2).

A Tukey's HSD test did not indicate any diplotypes for TA at 10-weeks post storage being statistically different from the others. The TA diplotype mean values at 10-weeks post storage ranged from 5.21 g/L (GrimesRecomNorthSpy1) to 2.73 g/L (Aspa2NWGreening1).

Haplotype Associations for LG1 at Harvest and at 10-Weeks Post Storage

Haplotypes at the LG1 locus were evaluated (Table 3.17). ANOVA utilizing individuals from all six families detected significant variation among LG1 haplotypes for TA at harvest ($p= 1.26x10^{-6}$) and at 10-weeks post storage ($p= 0.0127$).

Three statistically significant haplotype groupings were determined for TA at harvest using a Tukey's HSD test (Table 3.17). The TA haplotype mean values at harvest ranged from 7.11 g/L (NorthSpy1) to 5.64 g/L (Aspa2).

Two statistically significant haplotype groupings were determined for TA at 10-weeks post storage using a Tukey's HSD test (Table 3.17). The TA haplotype mean values at 10-weeks post storage ranged from 5.02 g/L (NorthSpy1) to 3.02 g/L (Aspa2). The lowest haplotype mean (Aspa2) was 0.64 g/L lower than the next lowest mean (Frostbite1, mean=3.66 g/L). The highest haplotype mean (NorthSpy1, mean= 5.02 g/L) was 1.06 g/L higher than the next highest mean (GrimesRecom, mean= 3.97 g/L). Most haplotypes had very similar mean values for 10-weeks post storage, with the exception of the two haplotypes that had the highest and lowest averages.

'Minneiska' x 'MN55' LG1-LG16 loci interaction

The 'Minneiska' x 'MN55' family was evaluated for diplotypes at both the LG1 and LG16 loci for TA at harvest (Table 3.9). Individuals in this family were grouped according to the sixteen combined diplotype states at the LG1 and LG16 loci. ANOVA utilizing this grouping of individuals detected significant variation among LG1-LG16 groups ($p= 1.1x10^{-4}$). Three distinct groupings were indicated by a Tukey's HSD test (Table 3.18).

Haplotype Associations for LG16, LG8, and LG1 for the Difference between Harvest and 10-Weeks Post Storage

Haplotypes at the LG16, LG8, and LG1 loci were evaluated for mean TA difference between harvest and 10-weeks post storage (Table 3.19, Table 3.20, Table 3.21). ANOVA detected significant variation among LG16 haplotypes (p= 0.0918), LG8 haplotypes (p= 0.079), and LG1 haplotypes (p= 0.0003).

The Tukey's HSD test for LG16 haplotypes determined two statistically different haplotype groupings at the p=0.05 level (Table 3.19), none for the LG8 haplotypes (Table 3.20), and two for the LG1 haplotypes (Table 3.21).

The 2014-2018 BLUP TA mean haplotype values at harvest and at 10-weeks post storage for each of the six families were plotted for each locus. The predictive ability of TA at harvest to TA at 10-weeks post storage varied among the loci. For the LG16 locus (Figure 3.6) the regression had a large coefficient of determination (R^2=0.79, p= 0.003) with a slope of 0.62 and y intercept of -0.22. For the LG8 locus (Figure 3.7) the coefficient of determination was more modest (R^2=0.39, p= 0.30) with a slope of 0.36 and y intercept of 1.82. The LG1 locus (Figure 3.8) was similar to the LG16 locus in having a large coefficient of determination (R^2=0.75, p= 0.025) with a slope of 0.98 and y intercept of -2.37.

Discussion

Genomic Regions Contributing to TA Variation Determined Using QTL Mapping

Three major loci contributing to variation for TA were identified on LG1, LG8, and LG16. All loci were detected in all three of the BLUP models as well as many individual year analyses, and had BF>10 for many analyses (Table 3.1).

The LG16 locus consistently identified by QTL analyses spans the 8.44 cM to 10.43 cM region. This result is expected, as this locus was identified over 60 years ago and was named the *Ma* (malic acid) locus (Nybom 1959) and was identified in QTL studies relating to modern apple germplasm sets (Xu et al. 2012, Zhang et al. 2012, Jia et al. 2018, Verma et al. 2019). Candidate genes proposed by Khan et al. (2013) are two malate transporter genes (MDP0000244249 and MDP0000252114). These genes are found within the *Ma* locus region at 3.165 Mbp to 3.168 Mbp and 3.176 Mbp to 3.179 Mbp, respectively, and fall between two markers used in this study (located at 8.83 cM and 10.22 cM). This region is consistent with the QTL mapping results from this study, and allows for side-by-side comparison of results with the Verma et al. (2019) study.

In addition to the well-known *Ma* locus identified in this study and previous QTL detection studies, a less frequently detected QTL located on LG8, named the *Ma3* locus by Verma et al. (2019), was also identified in this germplasm set. The LG8 locus has been described in several previous studies (Zhang et al. 2012, Jia et al. 2018, Sun et al. 2015, Liebhard et al. 2003). Jia et al. (2018) suggested three unique and epistatic genes within the LG8 locus as being responsible for controlling variation in acidity content. The LG8 region identified in this study that is being utilized for further haplotype analysis is 24.61 cM to 36.23 cM and is consistent with the region investigated by Verma et al. (2019).

A third locus identified on LG1 has been reported in literature in a few studies (Xu et al. 2012, Liu et al. 2016, Rymenants et al. 2020). This locus was named the *Ma5* locus by Rymenants et al. (2020). This locus as identified in this study spans 51.28 cM to 56.57 cM and completely coincides with the QTL region described in Chapter 2 relating to SSC. Therefore, the region that was utilized for SSC haplotype analysis in Chapter 2 was also utilized to analyze TA content. By utilizing the same region, comparisons and links between SSC and TA can be investigated and will be described later.

Segregation Distortion

Evaluation of the LG16 locus uncovered significant segregation distortion in four of the six families examined in this study (Table 3.8). Zhang et al. (2012) observed segregation distortion on LG16, specifically in the region of 0 cM to 15.2 cM. They suggested that a lethal gene is likely present near the region observed. A lethal region on LG16 was positively identified by Orcheski et al. (2015). They named this region the *Pale Green Lethal* (*PGL*) locus. Upon investigation of the material in this study and the region identified by Orcheski et al. (2015), one family ('Honeycrisp' x MN1915) is likely exhibiting segregation distortion due to this lethal gene. A single diplotype (DD) is greatly underrepresented and this diplotype exhibits identical homozygous marker states in each family, consistent with *PGL* haplotype characterizations by Orcheski et al. (2015). When this diplotype is compared to the region containing the *PGL* locus, it is highly probable that the missing individuals would be homozygous for the lethal haplotype. Two individuals in 'Honeycrisp' x MN1915 appear to have two copies of the *PGL* haplotype but survived into the reproductive stage. This may due to the *PGL* locus

being closely linked to the region identified for the acidity QTL, but the acidity QTL does not contain the *PGL* locus. The gene MDP0000194211 that segregates with the *PGL* locus according to Orcheski et al. (2015) maps to 3.754 Mbp to 3.760 Mbp on chromosome 16 of the 'Golden Delicious' doubled-haploid (GDDH13) genome (Daccord et al. 2017). The most distal marker used to characterize the LG16 locus is located at 3.440 Mbp, placing the PGL candidate gene about 0.32 Mbp further downstream from the end of the LG16 haplotyped region. Therefore, it is possible that the two individuals in 'Honeycrisp' x MN1915 experienced recombination near the *Ma* and *PGL* loci.

The LG8 locus exhibited segregation distortion in two families ('Honeycrisp' x 'Minnewashta' and 'Minneiska' x 'Wildung', Table 3.8). Segregation distortion at this locus was observed by Jia et al. (2018) who suggested selective pressure against certain (undescribed) alleles as the cause of the distortion. Verma et al. (2019) also observed segregation distortion at this locus, and suggested that 'Jonathan' may carry a lethal allele about 10 cM away from the LG8 acidity QTL. As Verma et al. (2019) suggested, a study that contains many individuals with recombination in this region might be able to help identify the exact region in which distortion is occurring, and a potential cause could be investigated.

The two families that exhibited segregation distortion at the LG1 locus are described in Chapter 2.

Utilization of BLUP Models and the Effects on QTL Mapping

In addition to utilizing observed TA values, several datasets were created consisting of BLUP values. Such datasets were created in order to capture the additional information available from having multiple years of data for some individuals, since FlexQTL™ is only able to evaluate a single phenotypic value for each individual in each analysis. In addition to capturing a more accurate TA value for many individuals that have replicated data across years, these BLUP models allow combining of datasets with missing values by estimating and accounting for a year effect. Therefore, utilizing BLUP models produced datasets with larger sample sizes than any of the individual years of data.

The QTL analyses that utilized BLUP adjusted data were compared to the QTL analyses that utilized only the observed data (Table 3.1). The BLUP analyses consistently identified similar QTL that commonly were detected in the observed data analyses. The BLUP QTL analyses also identified additional QTL. For instance, a QTL was detected on LG10 in several analyses, but was not significant in any single year analysis (Table 3.1).

Haplotype Naming

While the haplotypes were traced back as far as possible using current genetic data, some ancestors have missing genetic data (such as SNPs that were unable to be characterized). Haplotypes such as those described in Verma et al. (2019) that could not be fully distinguished in this study may be later revealed to be more related than currently thought. Several haplotypes at the LG16 locus could not be positively identified as identical by descent with haplotypes from Verma et al. (2019) (Table 3.3).

Diversity and Origin of LG1 Haplotypes

The diversity and origin of LG1 haplotypes was described in detail in Chapter 2.

Diversity and Origin of LG8 Haplotypes

Several haplotypes at the LG8 locus were identified as identical by state and descent to those described by Verma et al. 2019 (Table 3.3). Five unique haplotypes are described in this study and originate from: the unknown parent (designated as maternal) of 'Duchess of Oldenburg', the unknown parent (designated as maternal) of 'Frostbite', the unknown parent (designated as paternal) of 'Malinda', the unknown parent (designated as paternal) of 'Mantet', and the unknown parent (designated as maternal) of US1Ark. The new haplotypes described in this study show diversity in germplasm between the UMN and WSU apple breeding programs.

Diversity and Origin of LG16 Haplotypes

Several haplotypes at the LG16 locus were identified as identical by state and descent to those described by Verma et al. 2019 (Table 3.3). Several haplotypes (B*, C* and H*) could not be positively identified as identical by descent compared to those described in Verma et al. (2019). The B* haplotype is identical by state to the B haplotypes described by Verma et al. (2019), but the genotyped pedigree does not provide the connecting link to the ancestors described by Verma et al. (2019).

Additionally, the C* haplotype is identical by state to the C haplotypes described by Verma et al. (2019) but the genotyped pedigree does not provide the connecting link to the ancestors described by Verma et al. (2019). Lastly, the H* haplotype is identical by state to the H and I haplotypes described by Verma et al. (2019), but missing genetic data does not allow more detailed characterization in this study. Due to the H and I haplotypes being predicted to have a "low TA" (q) state, the H* haplotype name was arbitrarily used for further referencing. A novel haplotype, MinHcRecom, was identified in this germplasm set. It originates from a recombination event in 'Honeycrisp' that was inherited to 'Minneiska'. It is not identical by state to any haplotypes described by Verma et al. (2019) and was therefore given a unique name.

Comparing Haplotypes and Diplotypes to Verma et al. 2019

Several haplotypes at the LG16 and LG8 loci in this study were consistent with those identified by Verma et al. (2019) while several others were not. The novel haplotypes are the result of germplasm differences between the UMN and WSU apple breeding programs. Additionally, at least one newly defined haplotype at the LG16 locus arose from a recent crossover event that was inherited into 'Minneiska' from 'Honeycrisp'. For the LG8 locus, five newly defined haplotypes were characterized in this study that were not present in germplasm examined by Verma et al. (2019). These haplotypes are Duchess1, Frostbite1, Malinda2, Mantet2, and US1Ark1. With these new haplotypes being reported, a more diverse array of germplasm can now be characterized for its state at the LG8 locus.

By examining the effects of shared haplotypes between Verma et al. (2019) and this study while defining and characterizing novel haplotypes, the work by Verma et al. (2019) can be evaluated for consistency across different locations and germplasm.

Most haplotypes in this study that were previously identified by Verma et al. (2019) at the LG16 locus had similar contributions to relatively higher or lower TA content. These include five haplotypes predicted to have high TA values (designated as "Q" haplotypes; A, B*, $B_{F2/Jt}$, C, and D in this study) and two haplotypes that are predicted to have low TA values (designated as "q" haplotypes; H* and J in this study). The haplotypes C*, A, D, and B* are the haplotypes with the highest mean TA values, while H*, MinHcRecom, $B_{F2/Jt}$, and J have the lowest mean TA values (Table 3.13). The relative effect of the $B_{F2/Jt}$ haplotype in this study was inconsistent with the effect predicted by Verma et al. (2019) as it did not exhibit a high TA value at harvest relative to the other "Q" haplotypes in this study. However, this haplotype had the smallest sample size at this locus (n= 24). Additionally, the range of mean TA values at harvest for the four highest haplotypes (C*, A, D, and B*) was 6.54 g/L to 6.79 g/L, suggesting similar effects to those predicted by Verma et al. (2019). The TA mean values at harvest for the four lowest haplotypes (H*, MinHcRecom, $B_{F2/Jt}$, and J) range from 4.69 g/L to 6.12 g/L, with the lowest haplotype (J, mean= 4.69 g/L) being 1.21 g/L lower than the next lowest haplotype (Table 3.13).

The two families ('Minneiska' x 'MN55' and 'Minneiska' x 'Wildung') that had significant ANOVA p-values for the LG16 locus at harvest were expected to segregate for diplotype effects based on predictions from Verma et al. (2019). No other families were predicted to have significant variation among diplotypes at the LG16 locus by

Verma et al. (2019). These results support the characterizations of the A, B*, C*, H*, J haplotypes as being split into two groups of haplotype effects (A, B* and C* having high TA, and H* and J having low TA).

The conclusion from comparing LG16 TA haplotypes between Verma et al. (2019) and this study is that the majority of haplotypes at the LG16 locus match the predictions suggested by Verma et al. (2019). The J haplotype may be an exception, as it has a substantially lower mean TA value at harvest than all other haplotypes in this study, suggesting that an additional group of "very low" TA haplotypes may be characterized in addition to the "high" and "low" groups suggested by Verma et al. (2019).

At the LG8 locus, four haplotypes from Verma et al. (2019) were identified in this set of UMN germplasm (J, I, G, B_{GG+DO}), and five new haplotypes were identified (US1Ark1, Frostbite1, Duchess1, Malinda2, Mantet2). The four haplotypes consistent with Verma et al. (2019) are all predicted to have the effect "q", or "lower" TA values. Three of these "q" effect haplotypes evaluated in this study (I, G, B_{GG+DO}) are on the lower end of the range for mean TA values, whereas J had the third highest mean TA value at harvest in this study (Table 3.15). Because this study did not include any haplotypes identified by Verma et al. (2019) as having a "Q" or "high TA" effect at the LG8 locus, the predictions of Verma et al. (2019) could not be validated or disputed. Although Verma et al. (2019) and this study had several haplotypes in common, they shared no common diplotypes at this locus. The diplotype averages at this locus suggest that some haplotypes have different effects (Table 3.14). In particular, the GG and GI diplotypes have especially low mean TA values at harvest. Additionally, the IJ diplotype has the second highest mean TA value at harvest at the LG8 locus (Table 3.14). The

observations of diplotypes at this locus suggest that the haplotypes I, G, B_{GG+DO} and J do not have similar effects on TA at harvest, as a Tukey's HSD test identifies these diplotypes containing these haplotypes in three statistically significant groupings (Table 3.14). Additionally, the G and J haplotypes are grouped into different statistically significant groups by a Tukey's HSD test when observing their values for TA at harvest (Table 3.15). Such diplotypes and observations were not made in Verma et al. (2019), due to the genetic background and segregation that occurred in their selected families.

The conclusion from comparing LG8 TA haplotypes with Verma et al. (2019) is that all families in this study carried "q" haplotypes as previously defined by Verma et al. (2019), or had haplotypes not previously identified by Verma et al. (2019). The effects of those overlapping haplotypes between the two studies could not be verified as having similar effects on TA at harvest. Rather, the characterization of the J haplotype by Verma et al. (2019) is disputed based on the results from this study, which suggests J as having a "high TA" designation in comparison to the "low TA" designations assigned to haplotypes B_{GG+DO}, I, and G.

Comparing QTL and Haplotype Effects to Rymenants et al. 2020

Rymenants et al. (2020) identified four loci contributing to variation in apple fruit acidity, including the LG1, LG8, and LG16 loci described in this study. The acidity loci identified in this current study are consistent with the loci identified by Rymenants et al. (2020) and Verma et al. (2019). While this current study and Verma et al. (2019) utilized instrumental characterizations of acidity, Rymenants et al. (2020) used a sensory panel to

characterize acidity for individuals. Additive effects are suggested by Rymenants et al. (2020) and Verma et al. (2019) as the gene action for the LG8 locus, such that acidity content for an individual can be estimated by summing the total "Q" (high acidity) alleles. Gene action for the LG16 locus was not specifically hypothesized by Rymenants et al. (2020) due to a lack of "qq" individuals in the study, which was likely due to the lethal *PGL* locus residing nearby. Similar segregation patterns in Rymenants et al. (2020) were observed for the LG16 and LG8 loci, suggesting that recessive lethal alleles are responsible for distortion of segregation ratios in some families for some loci. The LG1 locus identified by Rymenants et al. (2020) suggested two haplotypes resulting in low acidity perception and two resulting in high acidity perception. However, not all individuals in the study had been genotyped (four out of nine parental haplotypes were given TA predictions) that were observed for sensory acidity, so minimal characterizations were made for haplotypes at the LG1 locus by Rymenants et al. (2020). Additionally, the potential origin of the four characterized haplotypes do not match the origin for the LG1 haplotypes identified in this study. Therefore, predictions by Rymenants et al. (2020) could not be applied or investigated in this current study.

While Rymenants et al. (2020) had limitations for predicted haplotype effects, the study is important because a trained sensory panel was utilized to characterize individuals for perceived acidity and QTLs were identified on the same LGs as studies that have utilized instrumental measurements of acidity. This sensory characterization is consistent with mechanical characterizations and further supports the idea that these loci are in fact important for sensory perception of acidity and have relevance from a breeding perspective for sensory field evaluations.

Significant Loci for TA at 10-Weeks Post Storage

LG8 and LG16 were both identified in several families as having significant variation among diplotypes at 10-weeks post storage (Table 3.10). Families segregated for diplotype effects at the LG8 and LG16 loci for TA at harvest and at 10-weeks post storage (Table 3.9, Table 3.10), suggesting that LG8 and LG16 explain significant variation among individuals both at harvest and at 10-weeks post storage. The LG1 locus likely does not have an effect on TA at 10-weeks post storage for the families observed in this study.

Significant Loci for TA difference between Harvest and 10-Weeks Post Storage

While the LG1 locus did not have significant variation among diplotypes for TA at 10-weeks post storage for any of the families in this study, the LG1 locus exhibited significant variation among diplotypes for three of the six families ('Honeycrisp' x 'Minnewashta', 'Honeycrisp' x MN1836, and 'Minneiska' x 'MN55') when evaluated for the TA difference between harvest and 10-weeks post storage ($p= 0.041$, $p= 0.064$, $p= 4.4x10^{-3}$, respectively, Table 3.11). The LG8 and LG16 loci were not significant at the $p= 0.10$ level for the TA difference between harvest and 10-weeks post storage (Table 3.11). While the LG1 locus did not exhibit significant variation among diplotypes in any families for TA at 10-weeks post storage, the LG1 locus did exhibited significant variation among diplotypes for the TA difference between harvest and 10-weeks post storage for half of the families in this study (Table 3.11). The LG1 locus may be of

interest to breeders when considering acidity content difference between harvest and storage.

Hypothesized Locus Interactions for TA Content

While three influential loci were identified for variation in TA at harvest, they did not appear to have the same contributing effects. The LG1 and LG8 loci are mostly only significant for variation among diplotypes in families where the LG16 locus is not significant for variation among diplotypes.

No family had significant diplotype effects in ANOVAs for both LG1 and LG8 at harvest, at 10-weeks post storage, or for the difference between harvest and 10-weeks post storage (Table 3.9, Table 3.10, Table 3.11). However, 'Minneiska' x 'MN55' exhibited significant variation among diplotypes at harvest and at 10-weeks post storage, and 'Minneiska' x 'Wildung' exhibited significant variation among diplotypes at 10-weeks post storage (Table 3.9, Table 3.10). Future investigations may consider studying additional families that may have significant diplotype effects for one, two, or all three loci.

Linkage Group 1, 8, and 16 Loci Relationships of Haplotype Average TA at Harvest and at 10-Weeks Post Storage

All haplotypes at all loci exhibited a decrease in TA content at 10-weeks post storage compared to TA content at harvest (Table 3.19, Table 3.20, Table 3.21).

However, haplotypes at the LG1 and LG16 loci contributed to differential rates of acid decrease, as ANOVA and Tukey's HSD tests indicated significant variation among haplotypes at LG1 and LG16, but not LG8.

The LG16 locus haplotype TA means at harvest were highly predictive of haplotype means at 10-weeks post storage (Figure 3.6). Regressing the mean haplotype value after storage on the mean value at harvest results in linear regression slope 0.62 with a y intercept of -0.22, and with the proportion of variance predicted for 10-weeks post storage from harvest being R^2= 0.792 (p= 0.003) (Figure 3.6). ANOVA detected significant variation among haplotypes (p= 0.0918). This suggests that the LG16 locus may have an influence on acid metabolism during storage, and that it is likely proportional to the influence it has on acidity at harvest. This suggests that haplotypes with high acid levels at harvest will likely still have relatively high levels after storage (Table 3.19).

The LG8 locus haplotype TA means at harvest were not predictive of the haplotype means at 10-weeks post storage (R^2= 0.153, p= 0.297, Figure 3.7). ANOVA detected significant variation among haplotypes at the LG8 locus for the difference between harvest and 10-weeks post storage (p= 0.079), but a Tukey's HSD test did not determine any significant groupings (Table 3.20). This suggests that the LG8 locus did not have a significant influence on acid difference between harvest and storage.

The LG1 locus haplotype TA means at harvest are predictive of the haplotype means at 10-weeks post storage (Figure 3.8). Regressing the mean haplotype value after storage on the mean value at harvest results in linear regression slope 0.98 with a y intercept of -2.37, and with the proportion of variance predicted for 10-weeks post

storage from harvest being R^2= 0.7563 (p= 0.025) (Figure 3.8). ANOVA detected significant variation among haplotypes (p= 0.0003). This suggests that the LG1 locus likely had an influence on acid metabolism during storage. The ranking of haplotype averages from highest TA content to lowest TA content does not change when observed for values at harvest to values at 10-weeks post storage (with the exception of a miniscule difference of 0.01 g/L between the GrimesRecom haplotype and the NWGreening1 haplotype, Table 3.17). Therefore, haplotypes with high acid levels at harvest will likely still have high levels after storage, but mean differences vary depending on the LG1 haplotype (Table 3.21).

LG1 TA and SSC Interaction at Harvest and at 10-Weeks Post Storage

When considering the LG1 locus for its influence on both TA and SSC, most haplotypes positively correlate for SSC and TA at harvest; haplotypes that exhibited higher TA also exhibited higher SSC (Table 3.22). However, the Aspa2 haplotype exhibited high SSC but low TA, relative to the other haplotypes at the LG1 locus. The cause of this unique result may be explained with the following hypothesis:

TA and SSC have a QTL on LG1, between 51.3 cM and 56.6 cM, but are not caused by the same gene. Rather, they are two separate genes that are tightly linked in coupling phase in the majority of the haplotypes exhibiting high TA and high SSC or low TA and low SSC. However, the Aspa2 haplotype appears to be in repulsion for TA and SSC and exhibited an opposite trend, having high SSC and low TA.

While the gene(s) causing the different TA/SSC pattern exhibited by Aspa2 cannot be elucidated in this study, the pattern itself is unique and worth noting for a breeder and as a basis for future studies. Depending on the other haplotype combined with Aspa2 in breeding, one has the potential to produce seedlings with an imbalance of acid and sugar. If a cross is performed ('Wildung' x MN1965, for example) that results in seedlings inheriting both the Aspa2 and Frostbite1 haplotypes, those seedlings will have very low TA and moderate SSC. The combination of low TA and moderate SSC will result in an apple tasting "flat" and uninteresting from a flavor perspective, since the sugar/acid balance is a major component of sensory perception of flavor (Chitarrini et al. 2020). At 10-weeks post storage, the Aspa2 haplotype results in very low TA (mean = 3.02 g/L) and very high SSC (mean = 15.2 degrees brix). This combination of sugar and acid content will result in a very "flat" tasting fruit with an undesirable "sugar water" taste when fruit are stored for long periods. When considering storage ability of fruit, the Aspa2 haplotype should be paired with a haplotype that maintains a higher TA content in storage, such as the NorthSpy1 haplotype (mean = 5.02 g/L at 10-weeks post storage), to compensate for the low TA associated with Aspa2.

Conclusions

QTLs for variation in TA content were identified on LG16, LG8, and LG1 in UMN apple germplasm. The QTLs on LG16 and LG8 have been previously reported in many published papers (Liebhard et al. 2003, Xu et al. 2012, Zhang et al. 2012, Khan et al. 2013, Sun et al. 2015, Jia et al. 2018, Verma et al. 2019, Rymenants et al. 2020). The QTL detected on LG1 has been reported a few times as being a locus with major

influence on TA content and has been minimally characterized (Xu et al. 2012, Liu et al. 2016, Rymenants et al. 2020).

Several haplotypes at the LG16 and LG8 loci coincided with those described by Verma et al. (2019). The relative haplotype effects estimated by Verma et al. (2019) for LG16 are supported by the data in this study. Due to the lack of "Q" haplotypes at the LG8 locus in this study, only the LG16 locus could be compared for haplotype effects. In this study, several new haplotypes were identified and characterized for LG16 and LG8.

Several families were identified as having fixed diplotype effects at the LG16 locus and/or at the LG8 locus, and a third QTL explaining variation for TA content among diplotypes was detected on LG1. However, the LG16 and LG8 loci appear to have larger influences on TA content variation than the LG1 locus. The LG1 locus only contributed to significant variation among diplotypes in families where LG16 and LG8 were not identified as having significant effects on TA content variation, with the exception of 'Minneiska' x 'MN55' for TA at harvest.

Due to the LG16 and LG8 loci being fixed for haplotypes having similar effects in many UMN families, LG1 emerged as an additional influential locus for TA content. Within the six major families included in this study, six unique haplotypes were identified at the LG1 locus. Many of these haplotypes are present in multiple families, largely due to their relatedness because of the extensive use of 'Honeycrisp' and its progeny as breeding parents. One haplotype in particular, GrimesRecom, is found in all six of the families, and is one of the haplotypes carried by 'Honeycrisp' and 'Minneiska', and is therefore a haplotype of high interest for future breeding efforts due to its

extensive use in breeding and its contribution to high TA content at harvest and at 10-weeks post storage.

The QTLs identified for variation in TA content exhibited segregation distortion in several families. Segregation distortion based on Mendelian inheritance was observed for three of the six families at the LG16 locus. This locus contains the *Pale Green Lethal* locus (Orcheski et al. 2015) and was confirmed to be closely linked to the *Ma* locus and is likely the cause of segregation distortion in one family. Additionally, the LG1 locus exhibited segregation distortion in two families. Segregation distortion was previously reported on LG1 in close proximity to the *Vf* locus for apple scab resistance (Maliepaard et al. 1998) and the *Vf* locus was confirmed to be in close physical proximity to the LG1 locus identified in this study. Crossing designs should consider the LG1 and LG16 loci and the haplotypes that result in lethality, as the number of seedlings produced from crosses affected by these loci would be lower than the predicted number.

Desirable and undesirable haplotype combinations are suggested in this study for breeders who seek desirable TA at harvest and at 10-weeks post storage. The LG1 locus is of particular interest when seeking an optimal sugar/acid balance, as this locus appears to explain significant variation among haplotypes for both TA and SSC. The Aspa2 haplotype should be further studied to investigate the unique high SSC and low TA pattern it exhibits. A large family resulting from a parent that carries the Aspa2 haplotype may reveal where a crossover event occurs that helps elucidate the pattern exhibited by the Aspa2 haplotype.

Lastly, parental haplotype characterizations for SSC and TA can be updated for UMN germplasm to reflect the results from this study. For example, haplotypes for the

Ma locus (LG16, Verma et al. 2019) and *Ma3* locus (LG8, Verma et al. 2019), can be updated for ‘Minneiska’, and the fructose locus (LG1, Guan et al. 2015) can be updated for MN1965. ‘Minneiska’ is described by Verma et al. (2019) as having “low-low” haplotypes at the *Ma* locus, which should be updated to “high-medium” based on results from this study. The *Ma3* locus is described by Verma et al. (2019) as having “low-low” haplotypes for ‘Minneiska’, which should be updated to “medium-medium” based on results from this study. For the fructose locus (LG1, Guan et al. 2015) ‘MN1965’ was previously characterized as “very high-very high” and should be updated to “low-high” based on the results from this study. These are only a few examples of the improved haplotype characterizations for parents in UMN apple breeding germplasm. Results from this study show that previous haplotype characterizations that were developed using WSU apple breeding germplasm do not directly translate to some UMN apple breeding germplasm, and that UMN apple breeding germplasm is now more accurately described for its haplotype characterizations pertaining to acidity and sugar QTLs.

Figure 3.1. Distributions of TA at Harvest across Years. A boxplot laid over a violin plot for the observed TA values at harvest for each year in this study from 2010 to 2018. No data were collected in 2013. Each bolded point represents outliers. The box plots indicate the 25th, 50th, and 75th percentiles for each dataset.

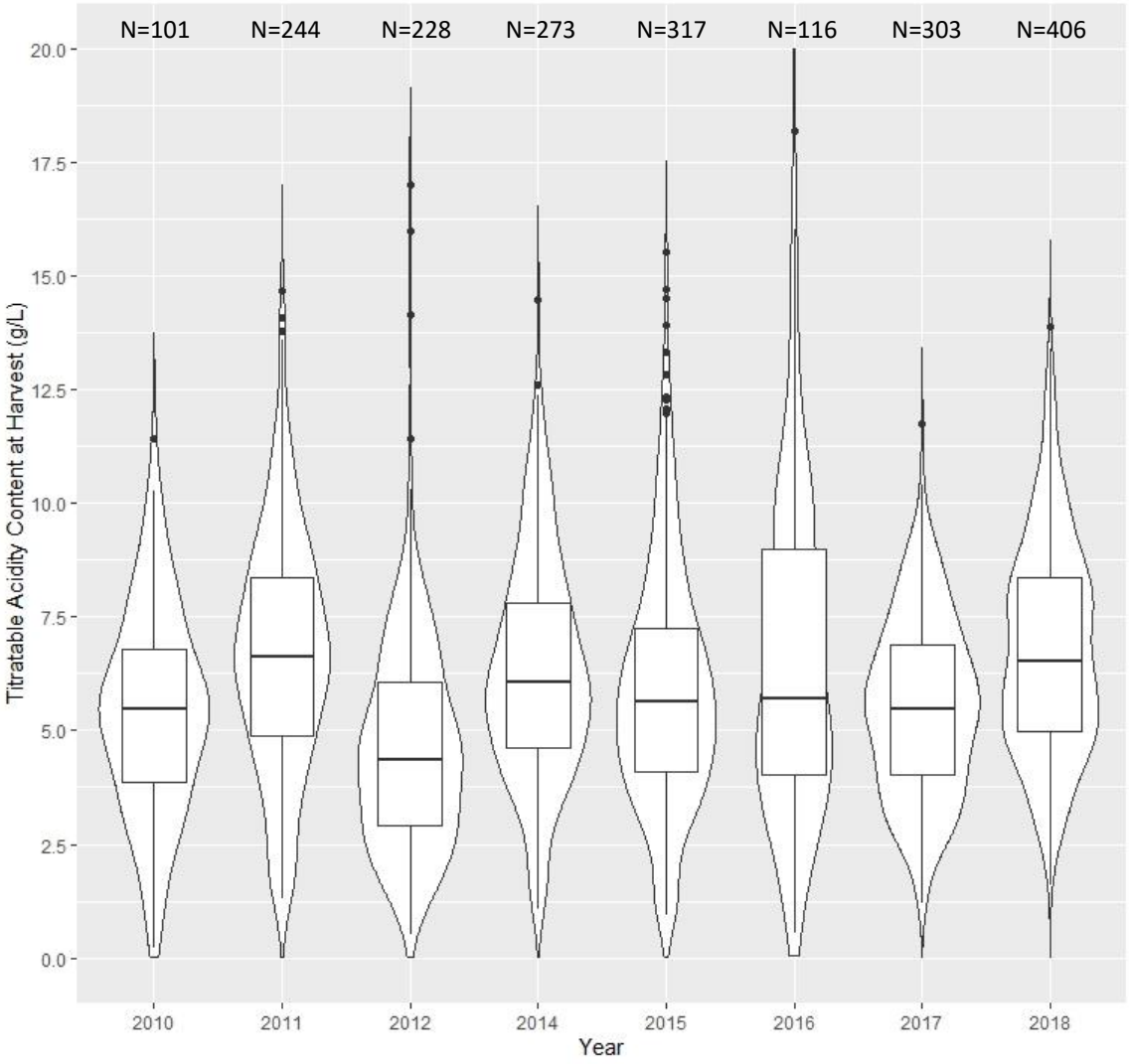


Figure 3.2. Distributions of TA at Harvest across BLUP Models. A boxplot laid over a violin plot for adjusted TA values at harvest using three BLUP models based on TA data from 1) 2010-2018, 2) 2010-2012, and 3) 2014-2018. Each bolded point represents outliers. The box plots indicate the 25th, 50th, and 75th percentiles for each dataset.

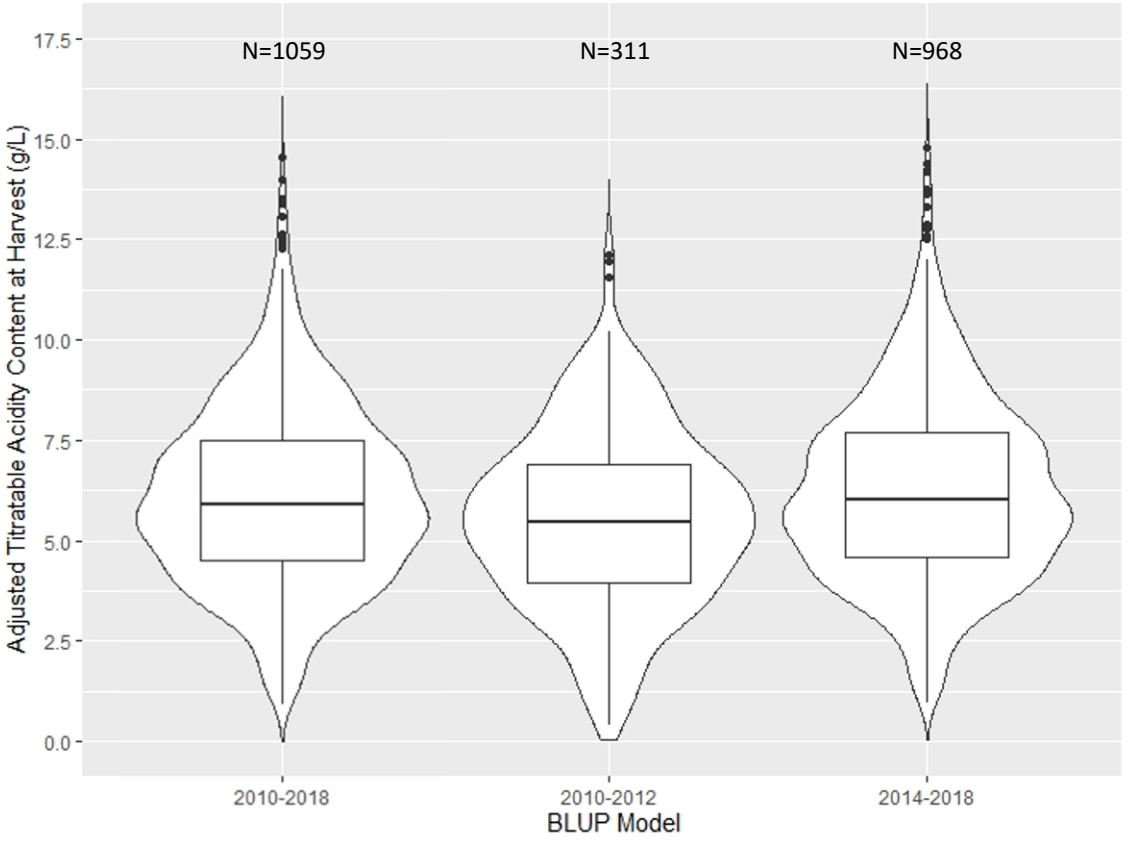


Figure 3.3. Distributions of TA at Harvest across Families. The distributions of adjusted TA values at harvest for the individuals using the 2014-2018 BLUP model in a histogram for each of the six families of focus in this study.

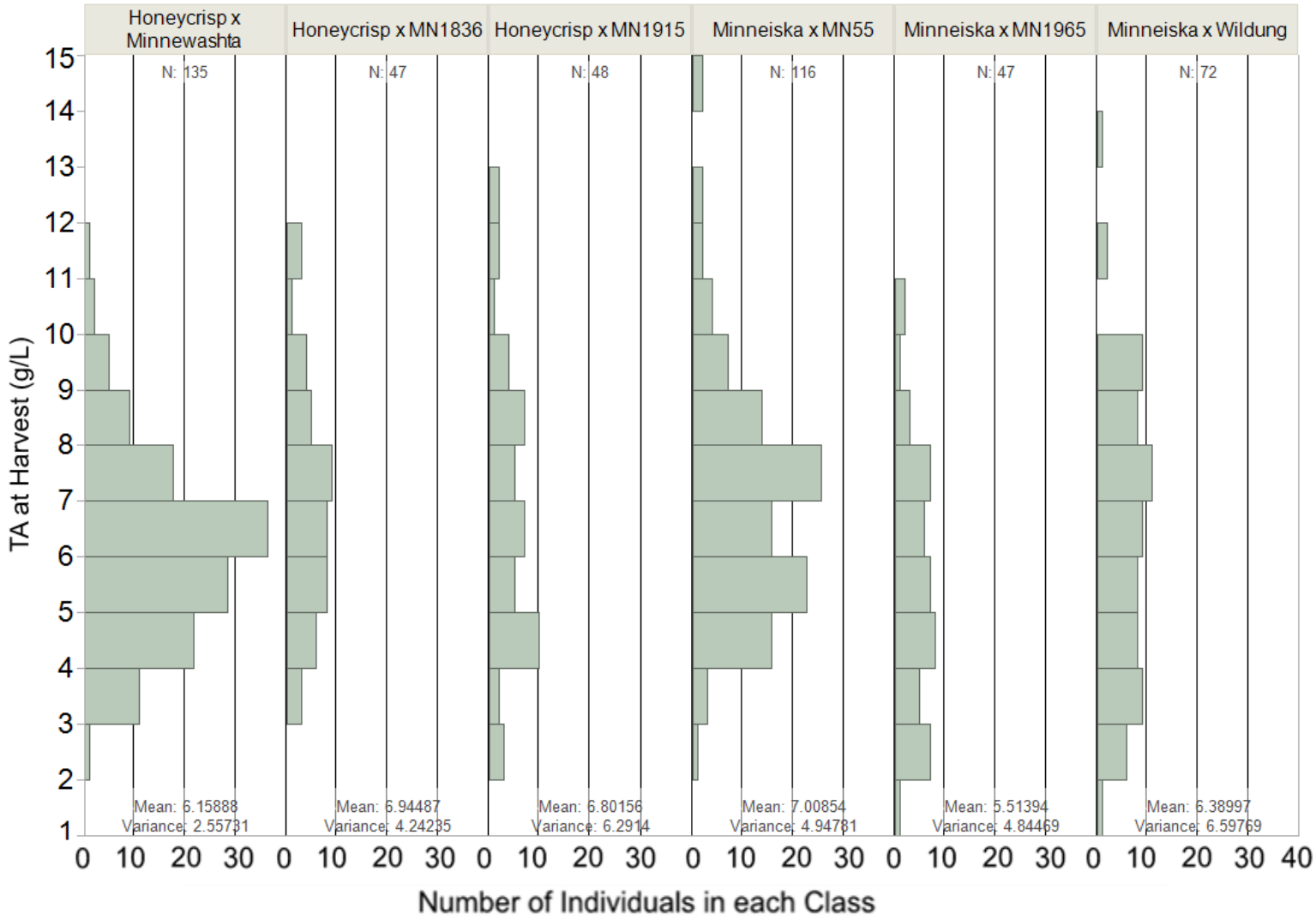


Figure 3.4. Distribution of TA at 10-Weeks Post Storage across Families. The distributions of adjusted TA values at 10-weeks post storage for the individuals using the 2014-2018 BLUP model in a histogram for each of the six families of focus in this study.

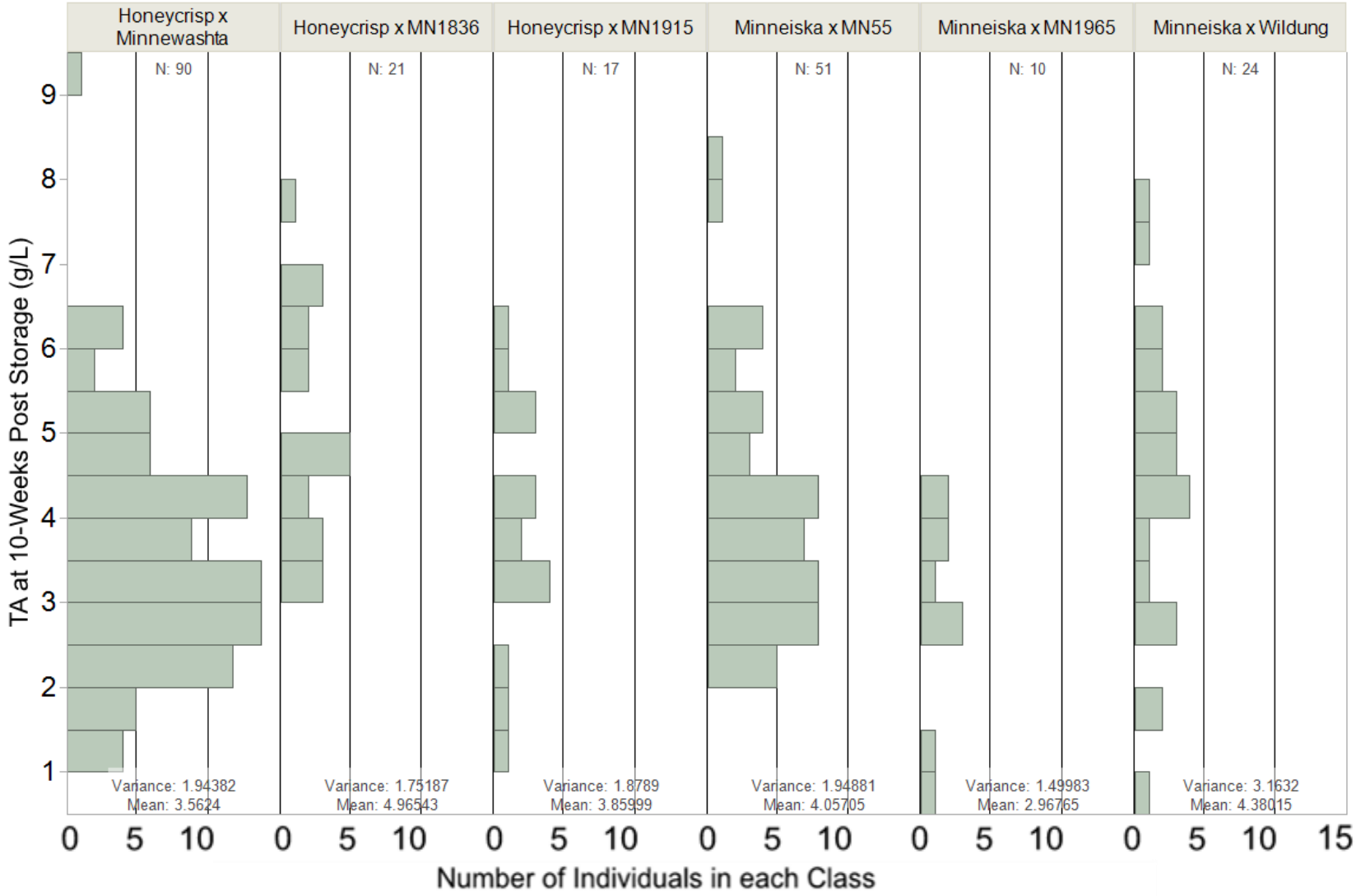


Figure 3.5. TA at Harvest vs. 10-Weeks Post Storage. Plot of the individual values from the 2014-2018 BLUP model for TA at harvest and at 10-weeks post storage for the six families of interest. Each point represents one individual. The total number of individuals is 202. The p-value for the linear regression and coefficient of determination is 8.1×10^{-25}.

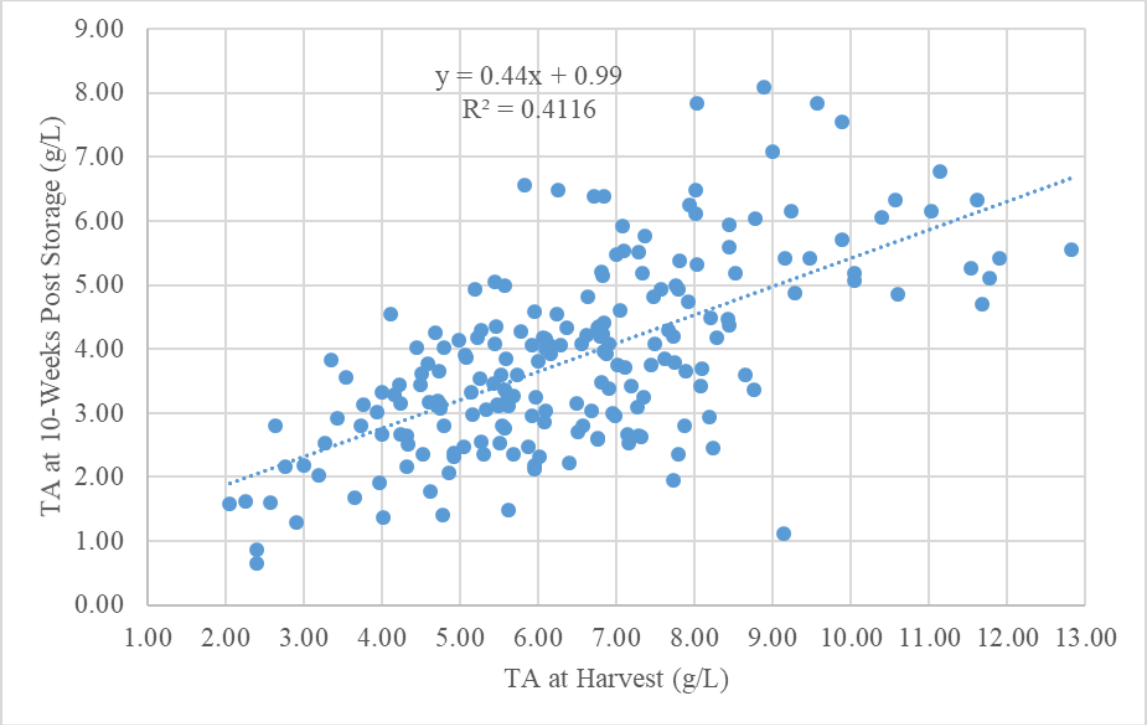
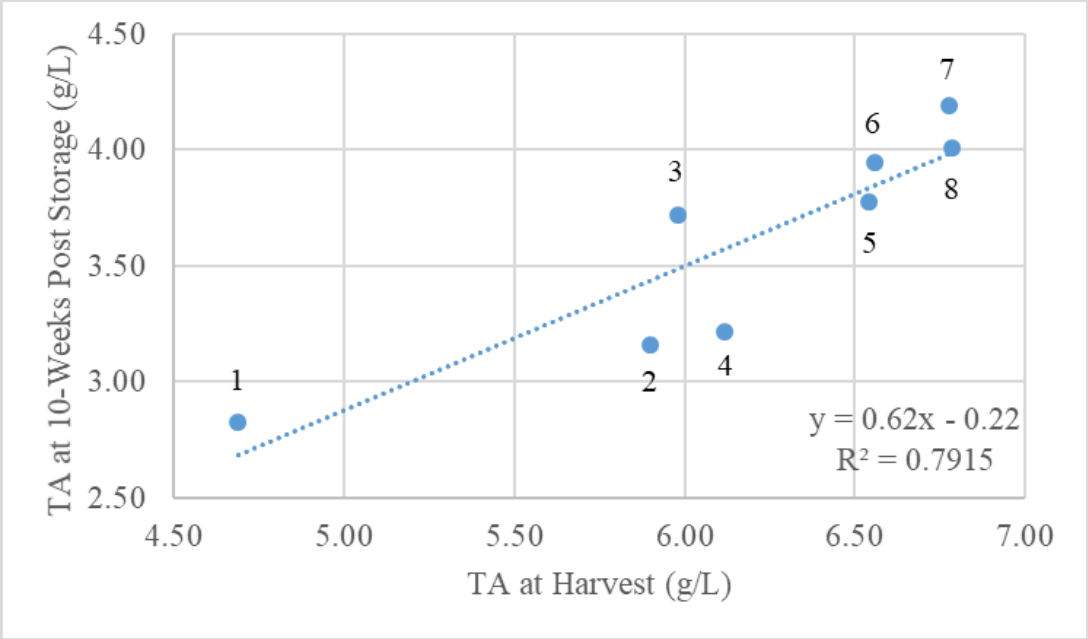


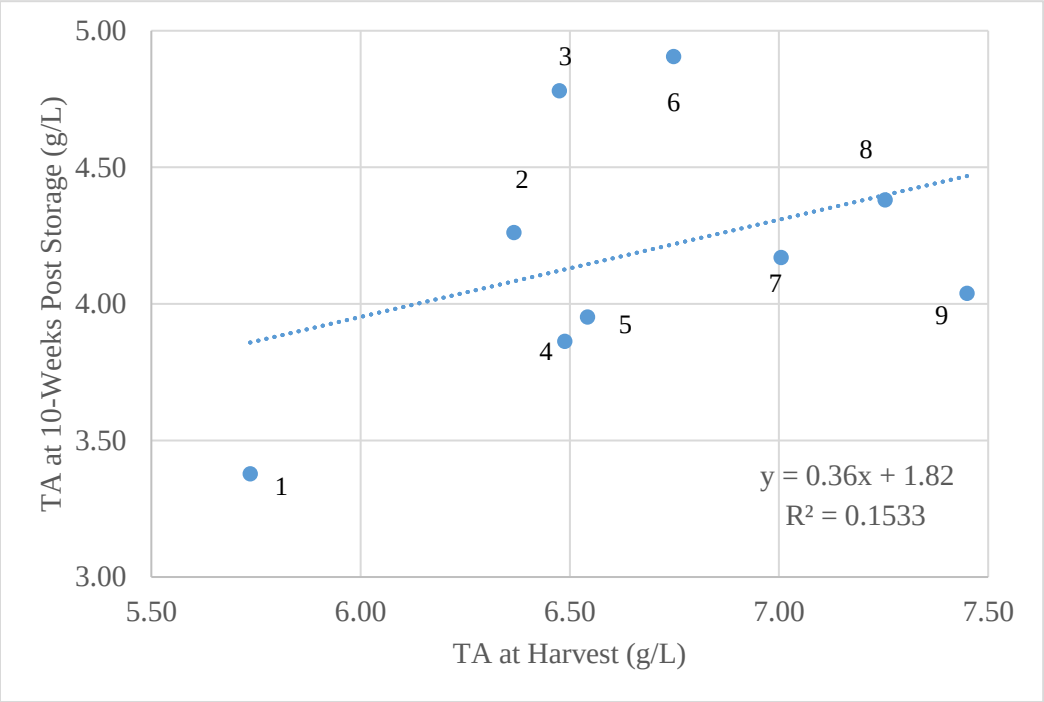
Figure 3.6. LG16 TA Haplotypes at Harvest vs. 10-Weeks Post Storage. Plot of the groups of haplotype means from the 2014-2018 BLUP model for TA at harvest and at 10-weeks post storage at the LG16 locus. The p-value for the linear regression and coefficient of determination is 0.003.



Key for labels:

1: J
2: $B_{F2/Jt}$
3: MinHcRecom
4: H*
5: B*
6: D
7: A
8: C*

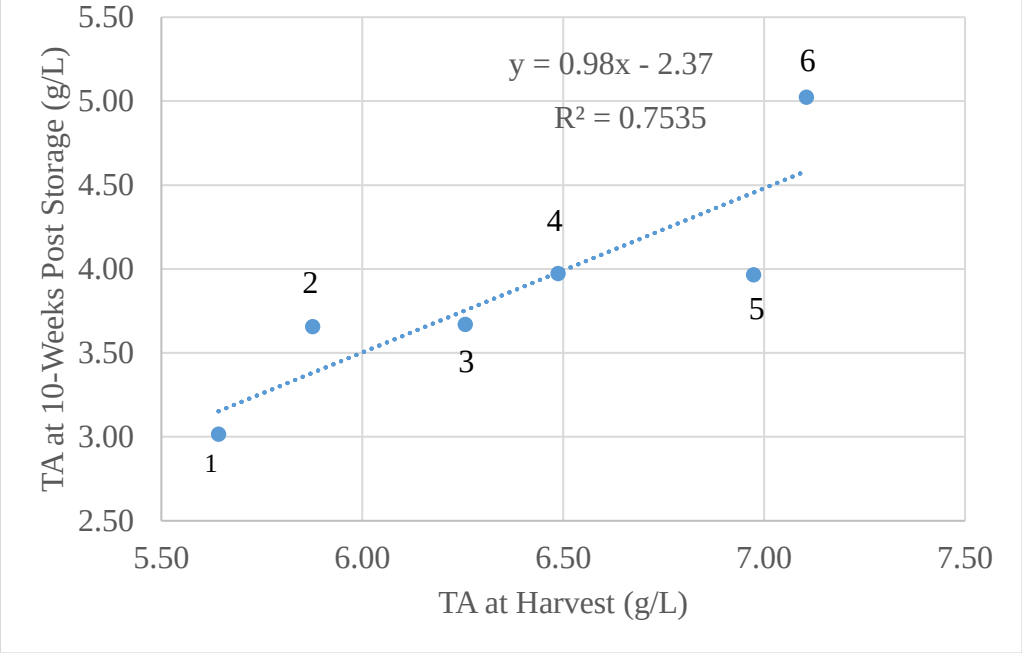
Figure 3.7. LG8 TA Haplotypes at Harvest vs. 10-Weeks Post Storage. Plot of the groups of haplotype means from the 2014-2018 BLUP model for TA at harvest and at 10-weeks post storage at the LG8 locus. The p-value for the linear regression and coefficient of determination is 0.297.



Key for labels:

1: G
2: I
3: Mantet2
4: b
5: Malinda2
6: Duchess1
7: J
8: Frostbite1
9: US1Ark1

Figure 3.8. LG1 TA Haplotypes at Harvest vs. 10-Weeks Post Storage. Plot of the groups of haplotype means from the 2014-2018 BLUP model for TA at harvest and at 10-weeks post storage at the LG1 locus. The p-value for the linear regression and coefficient of determination is 0.025.



Key for labels:

1: Aspa2
2: Frostbite1
3: Malinda2
4: NWGreening1
5: GrimesRecom
6: NorthSpy1

Table 3.1. TA QTL Detected. Transformed Bayes factor (BF) for QTL detected by FlexQTL™ software on each linkage group for TA at harvest, at 10-weeks post storage, and at 20-weeks post storage. Analyses were performed using BLUP adjusted TA values for the following multi-year analyses: 2010-2012, 2014-2018, and 2010-2018. BF values 0-2 indicate hardly any evidence, 2-5 are positive, 5-10 are strong, and >10 are decisive

Year	Evaluation Period	Linkage Group																	N
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
2010	Harvest	2.4	0.4	0	-0.6	0.9	0.2	-0.4	-0.3	2.7	0.4	-0.2	-0.1	-0.8	-0.5	0.1	5.4	0.1	101
2010	10-week	-0.8	-0.2	-0.7	-0.3	-0.7	0.1	-0.3	0.1	0.3	-0.2	-0.3	-0.1	-0.2	-0.3	-0.3	1.4	-0.4	55
2010	20-week	0.1	-0.1	0.2	-0.4	-0.3	-0.6	0.1	-0.3	0.2	-0.7	-0.7	-0.9	-0.1	0.2	-0.1	0.4	-0.2	37
2011	Harvest	-1.2	1	-0.3	-0.7	-0.8	-1.9	-0.7	0.3	-1.7	1.1	-0.9	-1.8	-0.2	-0.5	-1	13.6	-1.7	244
2011	10-week	-1.4	1.6	-0.2	-0.4	-0.2	-1.6	-0.9	-0.3	-1.2	-0.7	-0.9	-1.4	-0.2	-0.5	-0.6	12.6	-1.5	221
2011	20-week	0.7	3.3	-1.6	-0.3	0.4	-0.1	0.5	0	1.4	1.6	0.2	-0.9	-0.1	0.3	-0.8	5.9	-0.3	210
2012	Harvest	-0.9	0.3	1	-0.8	1.3	-0.4	0.9	2.6	0.6	0.9	-0.5	-0.3	1.3	0.2	-0.3	1	1.1	228
2012	10-week	-0.1	-0.1	0.3	-0.1	0	0.7	1	3.2	0	1.8	-0.5	0	-0.1	-0.4	0	0.8	0.2	156
2012	20-week	-3.8	-3.4	-4.2	-4	-0.8	-0.6	-1.5	-0.4	1	-3.3	-2.3	-3.4	-2.9	-4.1	-3.2	-1.1	10.6	180
2014	Harvest	31.3	-0.7	-1.5	0	1.9	-1.6	-0.3	8.2	-1	-0.5	1.9	0.5	-0.6	-0.3	-0.3	11*	-1.4	273
2014	20-week	5.6	0	0.8	-0.7	1.1	-0.6	-0.2	0.8	-0.1	2.1	-0.6	-0.3	-1.1	0	-1.4	4.8	4.8	254
2015	Harvest	11.4	5.5	-1.6	1.8	1	2.4	-1	0.3	2.8	-0.2	0.5	0.1	-2	-1.5	-1.2	29.9	-0.6	317
2015	10-week	3.1	0.1	-0.2	4.7	-0.1	0	-0.7	8.4	1.1	0.4	-0.5	-0.6	0.1	-0.5	0.4	5	0.7	144
2016	Harvest	1.9	2.2	1.9	-1.1	-0.6	-0.3	0.7	0.7	1.1	-0.2	-0.5	-0.5	-0.5	2.8	0.1	0.7	-0.9	116
2016	10-week	1.2	-0.1	-0.7	0.2	-0.3	-0.9	-0.3	-0.5	3.3	0.4	0.1	-0.6	0.9	0	-0.8	2.1	1.8	60
2017	Harvest	6.6	-0.6	-1	-0.8	-1.3	0	-1.6	31	-2	-0.9	1	1.4	-2.2	-0.8	-1.3	9.2	1.6	303
2018	Harvest	31.1	0.5	0.1	0.5	-0.1	-1.3	-0.7	30.8	-2	-0.2	-1.3	-1.5	3.2	-1.2	-1.1	30.8	1.2	406
2018	10-week	5.1	1.6	-0.9	1.1	0.4	-0.4	0.1	13.7	-1.7	0.1	-0.3	0.3	1.2	-0.6	-1.1	4.7	0.5	216
2010-2012	Harvest	-0.7	2.8	-1	1	-1.2	-1.3	-0.6	2	-0.4	1.9	0.1	-0.5	4.9	-1.1	0.2	14	-0.8	311
2010-2012	10-week	-0.9	2.2	-0.6	0	-0.6	-0.4	-0.1	5.2	1.1	3.6	-0.6	0.1	-0.5	-0.2	-0.5	5.5	-0.1	278

Table 3.1 Cont.

2010-2012	20-week	1.1	1.9	-1.2	-0.6	0	2.2	0.6	5.8	4.4	1	0.3	-1	0.4	-0.5	-1.2	3.6	1.4	277
2014-2018	Harvest	31.1	0	-1.9	3.5	3.6	4.9	2.2	30	-4.7	7.9	-4.6	-4.4	0.4	-4	-3.7	30.8	2.3	968
2014-2018	10-week	12.7	2.2	-1.5	8.6	-0.7	-0.9	-1.4	30.6	-1	0	-1.4	-1.7	-0.3	-1.8	-1	31.1	0.8	350
2010-2018	Harvest	31.3	0.1	-2.5	4.7	-1.3	4.7	6	29.4	-4.4	8.4	-3.2	-3.9	1.2	-4.3	-4.5	31	1.8	1059
2010-2018	10-week	9.3	6.3	-3.7	7.9	-1.9	-2.3	-1.8	28.6	-0.2	7.3	0.5	-3	-0.3	-3	-2.3	31.2	-1.2	650
2010-2018	20-week	11.5	0.9	0	-1.5	0.5	-0.9	-1.5	6.6	-0.5	5.1	-1.1	-0.4	-1.3	0.1	-1.6	11.7	4.9	378

*This value represents the 1_2 (1 QTL vs. 2 QTL) model BF value instead of the 0_1 (0 QTL vs. 1 QTL) model BF value, as the 0_1 value was reported as NA.

Table 3.2. TA QTL Detected Detailed. QTL regions on linkage groups (LGs) in centimorgans (cM) with a transformed Bayes factor (BF) > 5 identified in analyses of TA content in germplasm evaluated in individual years or combined years. The corresponding mode (cM location of the marker that was identified most often) and probability (higher value indicates higher likelihood that the QTL is real and significant) are also provided.

Year	Evaluation Period	LG	BF	QTL region (cM)	QTL Mode (cM)	Probability
2010	Harvest	16	5.4	1-27	11	0.761
2010	Harvest	16		29-56	35	0.112
2011	Harvest	16	13.6	1-12	9	1.009
2011	Harvest	16		15-28	16	0.058
2011	10-week	16	12.6	1-32	2	1.122
2011	20-week	16	5.9	2-32	18	0.875
2011	20-week	16		35-45	38	0.075
2011	20-week	16		54-66	65	0.103
2012	20-week	17	10.6	44-69	52	1.006
2014	Harvest	1	31.3	45-56	54	0.926
2014	Harvest	1		58-63	62	0.089
2014	Harvest	8	8.2	1-22	5	0.843
2014	Harvest	8		25-30	27	0.075
2014	Harvest	16	10.6	7-11	10	0.996
2014	Harvest	16		15-27	22	0.811
2014	Harvest	16		31-36	33	0.088
2014	20-week	1	5.6	40-63	48	0.712
2015	Harvest	1	11.4	30-48	38	0.892
2015	Harvest	1		50-58	53	0.108
2015	Harvest	16	29.9	1-7	6	0.285
2015	Harvest	16		11-28	12	1.031
2015	Harvest	16		59-66	65	0.129
2015	10-week	8	8.4	1-48	27	0.984
2015	10-week	8		50-56	53	0.062
2015	10-week	16	5	1-24	10	0.656
2017	Harvest	1	6.6	1-12	12	0.055
2017	Harvest	1		32-63	61	0.827
2017	Harvest	8	31	20-54	33	1.079
2017	Harvest	16	9.2	1-17	13	0.924
2018	Harvest	1	31.1	36-44	38	0.101
2018	Harvest	1		48-55	52	0.911
2018	Harvest	8	30.8	16-48	30	1.144
2018	Harvest	16	30.8	1-12	10	1.01
2018	Harvest	16		22-43	27	0.154
2018	10-week	8	13.7	14-58	23	1.327

Table 3.2 Cont.

2010-2012	Harvest	16	14	2-413	10	0.999
2010-2012	10-week	8	5.2	1-46	19	0.749
2010-2012	10-week	16	5.5	1-26	10	0.747
2010-2012	20-week	8	5.8	10-48	18	0.809
2014-2018	Harvest	1	31.1	46-60	52	1.034
2014-2018	Harvest	8	30	19-32	30	1.057
2014-2018	Harvest	8		37-47	42	0.334
2014-2018	Harvest	10	7.9	10-30	26	0.819
2014-2018	Harvest	16	30.8	2-6	4	0.129
2014-2018	Harvest	16		9-11	10	0.998
2014-2018	10-week	1	12.7	38-60	51	0.987
2014-2018	10-week	4	8.6	16-47	35	1.002
2014-2018	10-week	8	30.6	7-45	28	1.202
2014-2018	10-week	16	31.1	4-11	10	0.998
2010-2018	Harvest	1	31.9	47-54	52	0.993
2010-2018	Harvest	7	6	50-59	54	0.269
2010-2018	Harvest	7		64-72	66	0.146
2010-2018	Harvest	7		78-82	82	0.291
2010-2018	Harvest	8	29.4	25-32	29	0.994
2010-2018	Harvest	8		34-46	41	0.512
2010-2018	Harvest	10	8.4	11-14	12	0.076
2010-2018	Harvest	10		16-27	22	0.793
2010-2018	Harvest	16	31	1-10	10	1.073
2010-2018	10-week	1	9.3	31-60	50	0.938
2010-2018	10-week	2	6.3	19-50	28	0.766
2010-2018	10-week	4	7.9	14-18	15	0.176
2010-2018	10-week	4		19-28	26	0.519
2010-2018	10-week	4		31-43	36	0.211
2010-2018	10-week	8	28.6	18-49	23	1.633
2010-2018	10-week	8		51-55	53	0.069
2010-2018	10-week	10	7.3	21-34	28	0.869
2010-2018	10-week	16	31.2	5-10	9	0.998
2010-2018	20-week	1	11.5	45-63	55	0.976
2010-2018	20-week	8	6.6	0-23	12	0.247
2010-2018	20-week	8		25-44	35	0.604
2010-2018	20-week	10	5.1	16-39	32	0.705
2010-2018	20-week	16	11.7	1-12	10	0.941
2010-2018	20-week	16		14-18	16	0.077
2010-2018	20-week	16		20-28	26	0.051

Table 3.3. TA Loci Haplotype Names. Haplotype name, associated linkage group (LG), and the ancestral origin of all TA haplotypes. Haplotype names that end with a 1 indicate that the haplotype came from the unknown parent of the cultivar designated as maternal and haplotypes that end with 2 indicate origin from the unknown parent designated as paternal.

Haplotype Name	LG	Origin of Name
Aspa2	1	Unknown parent of 'Aspa'
Frostbite1	1	Unknown parent of 'Frostbite'
GrimesRecom	1	Recombination between the haplotypes of 'Grimes Golden', inherited by 'Golden Delicious'
Malinda2	1	Unknown parent of 'Malinda'
NorthSpy1	1	Unknown parent of 'Northern Spy'
NWGreening1	1	Unknown parent of 'Northwest Greening'
B_{GG+DO}	8	Name designation given by Verma et al. (2020)
Duchess1	8	Unknown parent of 'Duchess of Oldenburg'
Frostbite1	8	Unknown parent of ‘Frostbite’
G	8	Name designation given by Verma et al. (2020)
I	8	Name designation given by Verma et al. (2020)
J	8	Name designation given by Verma et al. (2020)
Malinda2	8	Unknown parent of 'Malinda'
Mantet2	8	Unknown parent of 'Mantet'
US1Ark1	8	Unknown parent of US1Ark1
A	16	Name designation given by Verma et al. (2020)
B*	16	Name designation given by Verma et al. (2020), ancestral origin unknown, identical by state to other B designations
$B_{F2/Jt}$	16	Name designation given by Verma et al. (2020)
C*	16	Name designation given by Verma et al. (2020), ancestral origin unknown, identical by state to other C designations
D	16	Name designation given by Verma et al. (2020)
H*	16	Name designation given by Verma et al. (2020), lacking information to distinguish between H and I designation
J	16	Name designation given by Verma et al. (2020)
MinHcRecom	16	Recombination between the haplotypes of 'Honeycrisp', inherited by 'Minneiska'

Table 3.4. LG1 Locus Markers and Haplotypes. SNPs and marker states used to identify segregating haplotypes at the LG1 TA locus in six families evaluated in 2014-2018. All six SNPs and their states for each parental haplotype are shown, along with the name designation for each haplotype.

		cM Position	51.28	52.13	52.98	56.21	56.37	56.57
Parent	Parental Hap	SNP ID / Haplotype	1051	2640	2292	1084	516	723
Honeycrisp	Keepsake	Frostbite1	B	B	A	B	A	B
Honeycrisp	MN1627	GrimesRecom	A	B	B	A	B	A
Minneiska	Honeycrisp	GrimesRecom	A	B	B	A	B	A
Minneiska	Minnewashta	NWGreening1	B	A	B	A	B	B
Minnewashta	State Fair	Malinda2	B	B	A	A	B	A
Minnewashta	MN1691	NWGreening1	B	A	B	A	B	B
MN1836	Keepsake	NorthSpy1	A	B	B	A	A	A
MN1836	State Fair	Malinda2	B	B	A	A	B	A
MN1915	Sweet16	NorthSpy1	A	B	B	A	A	A
MN1915	Minnewashta	Malinda2	B	B	A	A	B	A
MN1965	Ginger Gold	Aspa2	B	B	A	B	A	A
MN1965	Honeycrisp	Frostbite1	B	B	A	B	A	B
MN55	Honeycrisp	GrimesRecom	A	B	B	A	B	A
MN55	Monark	Malinda2	B	B	A	A	B	A
Wildung	Sharon	Frostbite1	B	B	A	B	A	B
Wildung	Fireside	Malinda2	B	B	A	A	B	A

The SNP ID numbers refer to the 20K index name, detailed information for each can be found in Appendix Table 1B.

SNP states A and B refer to the base associated with each SNP marker. A and B for each SNP are unique, and can be found for each marker in Appendix Table 1B.

Table 3.5. LG8 Locus Markers and Haplotypes. SNPs and marker states used to identify segregating haplotypes at the LG18 TA locus in six families evaluated in 2014-2018. All fourteen SNPs and their states for each parental haplotype are shown, along with the name designation for each haplotype. The bolded SNP markers were used by Verma et al. (2019) for QTL identified for TA at harvest and at 10 and 20-weeks post storage.

		cM Position	**24.61**	25.83	**25.83**	**25.83**	**27.93**	30.04	**30.04**	**31.84**	33.80	**33.82**	35.53	35.95	35.95	**36.23**
Parent	Parental Hap	SNP ID / Haplotype	**652**	2785	**128**	**2073**	**51**	2092	**2095**	**2811**	874	**1580**	394	549	2833	**2138**
Honeycrisp	Keepsake	J	**B**	A	**B**	**B**	**B**	B	**B**	**B**	A	**B**	A	A	B	**B**
Honeycrisp	MN1627	B_{GG+DO}	**A**	A	**B**	**B**	**A**	B	**B**	**B**	A	**B**	A	B	A	**A**
Minneiska	Honeycrisp	J	**B**	A	**B**	**B**	**B**	B	**B**	**B**	A	**B**	A	A	B	**B**
Minneiska	Minnewashta	G	**B**	A	**A**	**B**	**B**	B	**A**	**A**	A	**B**	B	A	B	**A**
Minnewashta	State Fair	Malinda2	**B**	A	**A**	**B**	**B**	A	**A**	**A**	B	**A**	A	A	A	**A**
Minnewashta	MN1691	G	**B**	A	**A**	**B**	**B**	B	**A**	**A**	A	**B**	B	A	B	**A**
MN1836	Keepsake	Frostbite1	**A**	A	**A**	**A**	**B**	A	**A**	**A**	B	**A**	A	A	A	**A**
MN1836	State Fair	Duchess1	**B**	A	**B**	**A**	**A**	B	**B**	**B**	A	**B**	A	B	A	**B**
MN1915	Sweet16	Frostbite1	**A**	A	**A**	**A**	**B**	A	**A**	**A**	B	**A**	A	A	A	**A**
MN1915	Minnewashta	G	**B**	A	**A**	**B**	**B**	B	**A**	**A**	A	**B**	B	A	B	**A**
MN1965	Ginger Gold	G	**B**	A	**A**	**B**	**B**	B	**A**	**A**	A	**B**	B	A	B	**A**
MN1965	Honeycrisp	B_{GG+DO}	**A**	A	**B**	**B**	**A**	B	**B**	**B**	A	**B**	A	B	A	**A**
MN55	Honeycrisp	B_{GG+DO}	**A**	A	**B**	**B**	**A**	B	**B**	**B**	A	**B**	A	B	A	**A**
MN55	Monark	US1Ark1	**B**	B	**A**	**B**	**B**	B	**B**	**A**	A	**B**	A	A	A	**B**
Wildung	Sharon	Mantet2	**B**	A	**A**	**B**	**B**	B	**B**	**A**	A	**B**	A	B	A	**B**
Wildung	Fireside	I	**B**	A	**A**	**B**	**B**	B	**A**	**A**	A	**A**	A	A	A	**B**

The SNP ID numbers refer to the 20K index name, detailed information for each can be found in Appendix Table 1B.

SNP states A and B refer to the base associated with each SNP marker. A and B for each SNP are unique, and can be found for each marker in Appendix Table 1B.

Table 3.6. LG16 Locus Markers and Haplotypes. SNPs and marker states used to identify segregating haplotypes at the LG16 TA locus in six families evaluated in 2014-2018. All six SNPs and their states for each parental haplotype are shown, along with the name designation for each haplotype. All SNP markers were used by Verma et al. (2019) for QTL identified for TA at harvest and at 10 and 20-weeks post storage.

		cM Position	8.44	8.44	8.83	10.22	10.22	10.43
Parent	Parental Hap	SNP ID / Haplotype	352	2017	2043	647	2783	1542
Honeycrisp	Keepsake	D	B	B	B	B	B	B
Honeycrisp	MN1627	A	A	B	B	B	A	A
Minneiska	Honeycrisp	MinHcRecom	B	B	B	B	A	A
Minneiska	Minnewashta	C*	B	A	B	A	A	B
Minnewashta	State Fair	C*	B	A	B	A	A	B
Minnewashta	MN1691	B*	B	B	A	A	A	B
MN1836	Keepsake	D	B	B	B	B	B	B
MN1836	State Fair	C*	B	A	B	A	A	B
MN1915	Sweet16	D	B	B	B	B	B	B
MN1915	Minnewashta	B*	B	B	A	A	A	B
MN1965	Ginger Gold	$B_{F2/Jt}$	B	B	A	A	A	B
MN1965	Honeycrisp	A	A	B	B	B	A	A
MN55	Honeycrisp	A	A	B	B	B	A	A
MN55	Monark	H*	A	B	B	B	B	A
Wildung	Sharon	J	A	B	B	B	A	A
Wildung	Fireside	B*	B	B	A	A	A	B

The SNP ID numbers refer to the 20K index name, detailed information for each can be found in Appendix Table 1B.

SNP states A and B refer to the base associated with each SNP marker. A and B for each SNP are unique, and can be found for each marker in Appendix Table 1B.

Table 3.7. TA Loci Parental Haplotype Names. Haplotypes (Hap) for the linkage group (LG) 1, LG8 and LG16 TA loci in parents of the six families evaluated in 2014-2018. Haplotypes for the LG1 loci were named according to the most distant genotyped ancestor (see Table 3.3). Haplotypes for the LG8 and LG16 loci are named using designations from Verma et al. (2019), when a positive haplotype identification could be made. Haplotypes at LG8 and LG16 that were not found in the Verma et al. (2019) study were named based upon the most distant genotyped ancestor.

Parent	LG1		LG8		LG16	
	Hap1	Hap2	Hap1	Hap2	Hap1	Hap2
Honeycrisp	Frostbite1	GrimesRecom	J	B_{GG+DO}	D	A
Minneiska	GrimesRecom	Malinda2	G	J	MinHcRecom	C*
Minnewashta	Malinda2	NWGreening1	Malinda2	G	C*	B*
MN1836	Malinda2	NorthSpy1	Frostbite1	Duchess1	D	C*
MN1915	NorthSpy1	Malinda2	G	Frostbite1	D	B*
MN1965	Aspa2	Frostbite1	B_{GG+DO}	G	$B_{F2/Jt}$	A
MN55	NWGreening1	GrimesRecom	B_{GG+DO}	US1Ark1	A	H*
Wildung	Frostbite1	Malinda2	I	Mantet2	J	B*

*haplotypes are identical by state to those described by Verma et al. (2019) but cannot be verified as identical by descent.

Table 3.8. Chi-squared Values for each TA Locus. Chi-squared values for diplotype segregation distortion for each family for the LG1, LG8, and LG16 TA loci. N is the number of individuals in each family that were evaluated at least once between 2014 and 2018, not including individuals with recombinant haplotypes. Bolded values indicate significant distortion ($p < 0.05$).

Family	LG1	N (LG1)	LG8	N (LG8)	LG16	N (LG16)
'Honeycrisp' x 'Minnewashta'	0.214	115	**0.040**	107	0.077	135
'Honeycrisp' x MN1836	0.284	40	0.710	37	**0.022**	48
'Honeycrisp' x MN1915	**0.048**	45	0.184	37	**0.002**	48
'Minneiska' x MN1965	**0.026**	44	0.611	44	0.936	45
'Minneiska' x 'MN55'	0.397	108	0.363	94	**0.022**	116
'Minneiska' x 'Wildung'	0.599	64	**0.032**	62	0.154	69

Table 3.9. TA at Harvest ANOVAs by Family. ANOVA p-values for variation due to diplotype effects within each family for TA loci on LG1, LG8, and LG16 on 2014-2018 BLUP values for TA at harvest. N is the number of individuals in each family that were evaluated at least once between 2014 and 2018, not including individuals with recombinant haplotypes. Significant values are bolded.

Family	LG1	N (LG1)	LG8	N (LG8)	LG16	N (LG16)
'Honeycrisp' x 'Minnewashta'	**2.9×10^{-3}**	115	0.141	107	0.526	133
'Honeycrisp' x MN1836	**7.4×10^{-4}**	40	0.950	37	0.328	46
'Honeycrisp' x MN1915	0.125	45	0.135	37	**0.083**	47
'Minneiska' x MN1965	0.707	44	**1.7×10^{-4}**	44	0.630	45
'Minneiska' x 'MN55'	**4.4×10^{-3}**	108	0.696	94	**3.3×10^{-4}**	115
'Minneiska' x 'Wildung'	0.623	64	**1.0×10^{-4}**	62	**1.2×10^{-5}**	69

Table 3.10. TA at 10-Weeks Post Storage ANOVAs by Family. ANOVA p-values for variation due to diplotype effects within each family for TA loci on LG1, LG8, and LG16 on 2014-2018 BLUP values for TA at 10-weeks post storage. N is the number of individuals in each family that were evaluated at least once between 2014 and 2018, not including individuals with recombinant haplotypes. Significant values are bolded.

Family	LG1	N (LG1)	LG8	N (LG8)	LG16	N (LG16)
'Honeycrisp' x 'Minnewashta'	0.199	73	**0.029**	68	0.122	89
'Honeycrisp' x MN1836	0.261	15	0.630	16	0.263	21
'Honeycrisp' x MN1915	0.599	15	0.567	14	0.864	16
'Minneiska' x MN1965	0.621	8	**0.073**	10	0.294	10
'Minneiska' x 'MN55'	0.264	47	0.833	44	**7.0×10^{-4}**	50
'Minneiska' x 'Wildung'	0.174	21	**0.034**	22	**0.033**	22

Table 3.11. TA Difference between Harvest and 10-Weeks ANOVAs by Family. ANOVA p-values for variation due to diplotype effects within each family for TA loci on LG1, LG8, and LG16 on 2014-2018 BLUP values for the TA difference between harvest and 10-weeks post storage. N is the number of individuals in each family that were evaluated at least once between 2014 and 2018, not including individuals with recombinant haplotypes. Significant values are bolded.

Family	LG1	N (LG1)	LG8	N (LG8)	LG16	N (LG16)
'Honeycrisp' x 'Minnewashta'	**0.041**	73	0.720	68	0.422	89
'Honeycrisp' x MN1836	**0.064**	15	0.728	16	0.184	21
'Honeycrisp' x MN1915	0.236	15	0.170	14	0.298	16
'Minneiska' x MN1965	0.925	8	0.3097	10	0.862	10
'Minneiska' x 'MN55'	**4.4×10^{-3}**	47	0.604	44	0.656	50
'Minneiska' x 'Wildung'	0.815	21	0.7521	22	0.145	22

Table 3.12. TA LG16 Locus Diplotypes. Means of all individuals having the same diplotype at the LG16 locus in the six key families for titratable acidity content (TA, g/L) at harvest and at 10-weeks post storage based on the adjusted values from the 2014-2018 BLUP model. ANOVA indicated significant variation among diplotypes at harvest ($p= 9.97x10^{-7}$) and at 10-weeks post storage ($p= 2.86x10^{-5}$). The group column designates statistically significant groupings based on a Tukey's HSD test ($p= 0.05$); diplotypes with different letters are statistically significant. N= number of individuals in each diplotype group.

Hap1	Hap2	TA at Harvest				TA at 10-Weeks			
		Group	Mean (g/L)	Std Error	N	Group	Mean (g/L)	Std Error	N
B*	C*	A	7.97	0.42	25	A B	4.61	0.41	11
A	D	A B	7.39	0.40	27	A	5.07	0.41	11
A	C*	A B	7.08	0.23	83	A	4.47	0.24	33
C*	H*	A B C	6.87	0.48	19	A B	3.35	0.41	11
B*	MHR	A B C D	6.75	0.52	16	A	5.76	0.68	4
D	D	A B C D	6.73	0.51	17	A B	5.14	0.61	5
A	MHR	A B C	6.61	0.34	39	A B	4.04	0.37	14
C*	D	A B C D	6.37	0.30	48	A B	3.78	0.25	31
A	B*	B C D	6.23	0.26	64	A B	3.72	0.23	36
$B_{F2/Jt}$	C*	A B C D	6.20	0.58	13	A B	3.59	0.61	5
B*	D	B C D	6.06	0.34	39	B	3.20	0.27	26
H*	MHR	B C D	5.57	0.41	26	A B	3.07	0.41	11
$B_{F2/Jt}$	MHR	A B C D	5.55	0.63	11	A B	2.43	0.79	3
C*	J	C D	5.03	0.52	16	A B	2.56	0.68	4
J	MHR	D	4.23	0.60	12	A B	3.18	0.79	3

*MHR refers to the MinHcRecom haplotype

Table 3.13. TA LG16 Locus Haplotypes. Means of all individuals having the same haplotype at the LG16 locus in the six key families for titratable acidity content (TA, g/L) at harvest and at 10-weeks post storage based on the adjusted values from the 2014-2018 BLUP model. ANOVA indicated significant variation among haplotypes at harvest ($p= 8.26x10^{-6}$) and at 10-weeks post storage ($p= 0.0278$). The group column designates statistically significant groupings based on a Tukey's HSD test ($p= 0.05$); haplotypes with different letters are statistically significant. N= number of individuals in each haplotype group.

Haplotype	TA at Harvest				TA at 10-Weeks			
	Group	Mean (g/L)	Std Error	N	Group	Mean (g/L)	Std Error	N
C*	A	6.78	0.15	204	A	4.01	0.15	95
A	A	6.78	0.15	213	A	4.19	0.15	94
D	A B	6.56	0.18	148	A	3.94	0.17	78
B*	A B	6.54	0.18	144	A	3.78	0.17	77
H*	A B C	6.12	0.32	45	A	3.21	0.31	22
MHR	B C	5.98	0.21	104	A	3.72	0.25	35
$B_{F2/Jt}$	A B C	5.90	0.44	24	A	3.16	0.52	8
J	C	4.69	0.41	28	A	2.83	0.55	7

*MHR refers to the MinHcRecom haplotype

Table 3.14. TA LG8 Locus Diplotypes. Means of all individuals having the same diplotype at the LG8 locus in the six key families for titratable acidity content (TA, g/L) at harvest and at 10-weeks post storage based on the adjusted values from the 2014-2018 BLUP model. ANOVA indicated significant variation among diplotypes at harvest ($p= 4.4x10^{-9}$) and at 10-weeks post storage ($p= 0.0004$). The group column designates statistically significant groupings based on a Tukey's HSD test ($p= 0.05$); diplotypes with different letters are statistically significant. N= number of individuals in each diplotype group.

Hap1	Hap2	TA at Harvest				TA at 10-Weeks			
		Group	Mean (g/L)	Std Error	N	Group	Mean (g/L)	Std Error	N
J	US1Ark1	A	7.66	0.39	29	A B	4.25	0.35	15
I	J	A	7.64	0.51	17	A	5.02	0.48	8
Frostbite1	J	A	7.63	0.43	24	A B	4.65	0.43	10
J	Mantet2	A	7.36	0.51	17	A B	4.86	0.61	5
G	US1Ark1	A	7.21	0.42	25	A B	3.77	0.39	12
B_{GG+DO}	J	A	6.99	0.36	33	A B	4.29	0.43	10
B_{GG+DO}	Frostbite1	A	6.80	0.47	20	A B	3.93	0.55	6
Duchess1	J	A B	6.76	0.70	9	A B	3.75	0.96	2
B_{GG+DO}	Duchess1	A B	6.73	0.70	9	A	5.29	0.55	6
J	Malinda2	A	6.62	0.39	29	A B	3.94	0.30	20
B_{GG+DO}	Malinda2	A	6.48	0.34	38	A B	3.97	0.28	23
G	J	A B	6.10	0.33	41	A B	3.57	0.30	20
B_{GG+DO}	G	A B	5.99	0.30	50	A B	3.24	0.27	25
G	Mantet2	A B C	5.79	0.45	22	A B	4.72	0.51	7
G	G	B C	3.74	0.60	12	B	1.69	0.78	3
G	I	C	2.75	0.85	6	B	1.22	0.96	2

Table 3.15. TA LG8 Locus Haplotypes. Means of all individuals having the same haplotype at the LG8 locus in the six key families for titratable acidity content (TA, g/L) at harvest and at 10-weeks post storage based on the adjusted values from the 2014-2018 BLUP model. ANOVA indicated significant variation among haplotypes at harvest ($p= 5.1x10^{-7}$) and at 10-weeks post storage ($p= 0.0037$). The group column designates statistically significant groupings based on a Tukey's HSD test ($p= 0.05$); haplotypes with different letters are statistically significant. N= number of individuals in each haplotype group.

Haplotype	TA at Harvest				TA at 10-Weeks			
	Group	Mean (g/L)	Std Error	N	Group	Mean (g/L)	Std Error	N
US1Ark1	A	7.45	0.30	54	A B	4.04	0.28	27
Frostbite1	A	7.25	0.33	44	A B	4.38	0.36	16
J	A	7.01	0.16	199	A	4.17	0.15	90
Duchess1	A B	6.75	0.52	18	A B	4.91	0.51	8
Malinda2	A B	6.54	0.27	67	A B	3.95	0.22	43
B_{GG+DO}	A B	6.49	0.18	150	A B	3.86	0.17	70
Mantet2	A B	6.47	0.35	39	A	4.78	0.41	12
I	A B	6.37	0.46	23	A B	4.26	0.45	10
G	B	5.74	0.17	168	B	3.38	0.17	72

Table 3.16. TA LG1 Locus Diplotypes. Means of all individuals having the same diplotype at the LG1 locus in the six key families for titratable acidity content (TA, g/L) at harvest and at 10-weeks post storage based on the adjusted values from the 2014-2018 BLUP model. ANOVA indicated significant variation among diplotypes at harvest ($p= 1.3x10^{-7}$) and at 10-weeks post storage ($p= 0.0282$). The group column designates statistically significant groupings based on a Tukey's HSD test ($p= 0.05$); diplotypes with different letters are statistically significant. N= number of individuals in each diplotype group.

Hap1	Hap2	TA at Harvest				TA at 10-Weeks			
		Group	Mean (g/L)	Std Error	N	Group	Mean (g/L)	Std Error	N
GrimesRecom	GrimesRecom	A	8.05	0.41	25	A	4.18	0.42	11
GrimesRecom	NorthSpy1	A B	7.88	0.53	15	A	5.21	0.62	5
GrimesRecom	NWGreening1	A B	7.36	0.31	44	A	4.17	0.27	26
GrimesRecom	Malinda2	A B	6.62	0.20	109	A	3.69	0.20	49
Frostbite1	NorthSpy1	A B C	6.60	0.43	23	A	4.89	0.52	7
Malinda2	NWGreening1	A B C	6.41	0.31	43	A	4.32	0.36	15
Frostbite1	NWGreening1	B C	6.10	0.30	49	A	3.65	0.33	17
Frostbite1	GrimesRecom	B C	5.84	0.42	24	A	3.79	0.42	11
Aspa2	GrimesRecom	B C	5.81	0.52	16	A	3.59	0.97	2
Aspa2	NWGreening1	B C	5.48	0.52	16	A	2.73	0.69	4
Frostbite1	Malinda2	C	5.37	0.29	52	A	3.34	0.24	32

Table 3.17. TA LG1 Locus Haplotypes. Means of all individuals having the same haplotype at the LG1 locus in the six key families for titratable acidity content (TA, g/L) at harvest and at 10-weeks post storage based on the adjusted values from the 2014-2018 BLUP model. ANOVA indicated significant variation among haplotypes at harvest ($p= 1.26x10^{-6}$) and at 10-weeks post storage ($p= 0.0127$). The group column designates statistically significant groupings based on a Tukey's HSD test ($p= 0.05$); haplotypes with different letters are statistically significant. N= number of individuals in each haplotype group.

Haplotype	TA at Harvest				TA at 10-Weeks			
	Group	Mean (g/L)	Std Error	N	Group	Mean (g/L)	Std Error	N
NorthSpy1	A B	7.11	0.35	38	A	5.02	0.40	12
GrimesRecom	A	6.97	0.13	258	A B	3.97	0.13	115
NWGreening1	A B C	6.49	0.17	152	A B	3.97	0.18	62
Malinda2	B C	6.26	0.15	204	B	3.67	0.14	96
Frostbite1	C	5.88	0.18	148	B	3.66	0.17	67
Aspa2	C	5.64	0.38	32	B	3.02	0.57	6

Table 3.18. ‘Minneiska’ x ‘MN55’ Family LG1-LG16 Loci for TA at Harvest. Means of all individuals for the ‘Minneiska’ x ‘MN55’ family having the same diplotypes at the LG1 and LG16 loci for titratable acidity content (TA, g/L) at harvest based on the adjusted values from the 2014-2018 BLUP model. ANOVA indicated significant variation among LG1-LG16 diplotypes ($p= 1.13x10^{-4}$). The group column designates statistically significant groupings based on a Tukey’s HSD test (p= 0.05); LG1-LG16 diplotypes with different letters are statistically significant. N= number of individuals in each LG1-LG16 diplotype group.

LG1 Hap1	LG1 Hap2	LG16 Hap1	LG16 Hap2	TA at Harvest			
				Group	Mean (g/L)	Std Error	N
GrimesRecom	GrimesRecom	A	C*	A	11.15	0.86	5
GrimesRecom	NWGreening1	C*	H*	A B C	9.26	1.36	2
GrimesRecom	NWGreening1	A	C*	A B	8.89	0.72	7
GrimesRecom	NWGreening1	A	MHR	A B C	8.30	0.86	5
GrimesRecom	Malinda2	A	C*	A B C	8.17	0.72	7
GrimesRecom	GrimesRecom	C*	H*	A B C	8.17	0.86	5
Malinda2	NWGreening1	A	C*	A B C	7.79	0.64	9
GrimesRecom	GrimesRecom	A	MHR	A B C	7.35	0.86	5
GrimesRecom	GrimesRecom	H*	MHR	A B C	6.69	0.96	4
GrimesRecom	Malinda2	A	MHR	B C	6.60	0.86	5
GrimesRecom	Malinda2	C*	H*	B C	6.56	0.78	6
Malinda2	NWGreening1	A	MHR	B C	6.55	0.78	6
GrimesRecom	NWGreening1	H*	MHR	B C	5.74	0.86	5
Malinda2	NWGreening1	C*	H*	B C	5.60	0.96	4
GrimesRecom	Malinda2	H*	MHR	B C	5.50	0.78	6
Malinda2	NWGreening1	H*	MHR	C	4.78	0.86	5

*MHR refers to the MinHcRecom haplotype.

Table 3.19. TA Difference between Harvest and 10-Weeks for LG16 Haplotypes. Means of all individuals having the same haplotype at the LG16 locus in the six key families for titratable acidity content (TA, g/L) difference between harvest and 10-weeks post storage based on the adjusted values from the 2014-2018 BLUP model. ANOVA indicated variation among haplotypes (p= 0.0918). The group column designates statistically significant groupings based on a Tukey's HSD test (p= 0.05); haplotypes with different letters are statistically significant. N= number of individuals in each haplotype group.

Haplotype	TA Difference between Harvest and 10-Weeks			
	Group	Mean (g/L)	Std Error	N
J	A	-0.93	0.60	7
$B_{F2/Jt}$	A B	-2.15	0.56	8
MHR	A B	-2.50	0.27	34
B*	A B	-2.51	0.18	75
A	A B	-2.58	0.17	90
D	A B	-2.64	0.18	75
C*	A B	-2.77	0.16	93
H*	B	-3.12	0.34	22

*MHR refers to the MinHcRecom haplotype

Table 3.20. TA Difference between Harvest and 10-Weeks for LG8 Haplotypes. Means of all individuals having the same haplotype at the LG8 locus in the six key families for titratable acidity content (TA, g/L) difference between harvest and 10-weeks post storage based on the adjusted values from the 2014-2018 BLUP model. ANOVA indicated variation among haplotypes (p= 0.0790). The group column designates statistically significant groupings based on a Tukey's HSD test (p= 0.05); haplotypes with different letters are statistically significant. N= number of individuals in each haplotype group.

Haplotype	TA Difference between Harvest and 10-Weeks			
	Group	Mean (g/L)	Std Error	N
Duchess1	A	-1.74	0.57	8
Mantet2	A	-1.85	0.46	12
I	A	-2.01	0.51	10
Frostbite1	A	-2.37	0.40	16
J	A	-2.63	0.17	90
G	A	-2.66	0.19	72
Malinda2	A	-2.69	0.24	43
B_{GG+DO}	A	-2.75	0.19	70
US1Ark1	A	-3.42	0.31	27

Table 3.21. TA Difference between Harvest and 10-Weeks for LG1 Haplotypes. Means of all individuals having the same haplotype at the LG1 locus in the six key families for titratable acidity content (TA, g/L) difference between harvest and 10-weeks post storage based on the adjusted values from the 2014-2018 BLUP model. ANOVA indicated variation among haplotypes (p= 0.0003). The group column designates statistically significant groupings based on a Tukey's HSD test (p= 0.05); haplotypes with different letters are statistically significant. N= number of individuals in each haplotype group.

Haplotype	TA Difference between Harvest and 10-Weeks			
	Group	Mean (g/L)	Std Error	N
Frostbite1	A	-2.00	0.18	67
NorthSpy1	A B	-2.37	0.43	12
Malinda2	A	-2.41	0.15	96
Aspa2	A B	-2.42	0.61	6
NWGreening1	A B	-2.62	0.19	62
GrimesRecom	B	-3.08	0.14	115

Table 3.22. SSC and TA at Harvest for LG1 Haplotypes. Means of all individuals having the same haplotype at the LG1 locus in the six key families for soluble solids content (SSC, °Brix) and titratable acidity content (TA, g/L) at harvest based on the adjusted values from the 2014-2018 BLUP model. ANOVA indicated significant variation among haplotypes for SSC at harvest ($p= 5.6x10^{-4}$) and among haplotypes for TA at harvest ($p= 1.26x10^{-6}$). The group column designates statistically significant groupings based on a Tukey's HSD test ($p= 0.05$); haplotypes with different letters are statistically significant. N= number of individuals in each haplotype group.

Haplotype	SSC at Harvest				TA at Harvest			
	Group	Mean (°Brix)	Std Error	N	Group	Mean (g/L)	Std Error	N
Aspa2	A	14.1	0.3	32	C	5.64	0.38	32
NorthSpy1	A B	13.7	0.5	38	A B	7.11	0.35	38
GrimesRecom	A	13.7	0.1	256	A	6.97	0.13	258
NWGreening1	A B	13.4	0.1	152	A B C	6.49	0.17	152
Malinda2	B	13.3	0.1	202	B C	6.26	0.15	204
Frostbite1	B	13.2	0.2	146	C	5.88	0.18	148

LITERATURE CITED

Beruter J. Sugar Accumulation and Changes in the Activities of Related Enzymes during Development of the Apple Fruit. Journal of Plant Physiology (1985) 121(4):331-341.

Beruter J. Carbohydrate metabolism in two apple genotypes that differ in malate accumulation. Journal of Plant Physiology (2004) 161:1011-1029.

Bianco L, Cestaro A, Sargent D J, Banchi E, Derdak S, Di Guardo M, Salvi S, Jansen J, Viola R, Gut I, Laurens F, Chagne D, Velasco R, van de Weg E, and Troggio M. Development and Validation of a 20K Single Nucleotide Polymorphism (SNP) Whole Genome Genotyping Array for Apple (Malus x domestica Borkh). Plos One (2014) 9(10):e110377.

Bink M C A M, Uimari P, Sillanpaa M J , Jans L L G, and Jansen R C. Multiple QTL mapping in related plant populations via a pedigree-analysis approach. Theor Appl Genet (2002) 104:751-762.

Bink M C A M, Anderson A D, van de Weg W E and Thompson E A. Comparison of marker-based pairwise relatedness estimators on a pedigreed plant population. Theoretical Apple Genetics (2008) 117:843-855.

Bink M C A M, Totir L, Radu, Braak, Cajo J. F. ter, Winkler, Christopher R., Boer, Martin P., and Smith, Oscar S. QTL linkage analysis of connected populations using ancestral marker and pedigree information. Theor Appl Genet (2012) 124:1097-1113.

Bink M C, Jansen J, … van de Weg W E. Bayesian QTL analyses using pedigreed families of an outcrossing species, with application to fruit firmness in apple. Theor Appl Genet (2014) 127(5):1073-1090.

Blanpied G D and Silsby K J. Predicting Harvest Date Windows for Apples. Cornell Cooperative Extension (1992) Information Bulletin 221.

Blissett E N. Genomewide Selection in Apple: Prediction and Postdiction in the University of Minnesota Apple Breeding Program (ProQuest Number 27544810) (2019) [PhD dissertation, University of Minnesota]. ProQuest.

Bus V G M, Rikkerink E H A, Caffier V, Durel C E, and Plummer K M. Revision of the Nomenclature of the Differential Host-Pathogen Interactions of Venturia inaequalis and Malus. Annual Review of Phytopathology (2011) 49:391-413.

Chagne D, Crowhurst R N, Troggio M, Davey M W, Gilmore B, Lawley C, Vanderzande S, Hellens R P, Kumar S, Cestaro A, Velasco R, Main D, Rees J D, Iezzoni A, Mockler T, Wilhelm L, Van de Weg E, Gardiner S E, Bassil N, and Peace C. Genome-Wide SNP Detection, Validation, and Development of an 8K SNP Array for Apple. PLoS ONE (2012) 7(2): e31745.

Chagne D, Krieger C, Rassam M, Sullivan M, Fraser J, Andre C, Pindo M, Troggio M, Gardiner S E, Henry R A, Allan A C, McGhie T K, and Laing W A. QTL and candidate gene mapping for polyphenolic composition in apple fruit. BMC Plant Biology (2012) 12:12.

Chagne D, Vanderzande S, Kirk C, Profitt N, Weskett R, Gardiner S E, Peace C P, Volz R K, and Bassil N V. Validation of SNP markers for fruit quality and disease resistance loci in apple (Malus x domestica Borkh.) using the OpenArray platform. Hortic Res. (2019) 6(30).

Chitarrini G, Dordevic N, Guerra W, Robatscher P, and Lozano L. Aroma Investigation of New and Standard Apple Varieties Grown at Two Altitudes Using Gas Chromatography-Mass Spectrometry Combined with Sensory Analysis. Molecules (2020) 25:3007.

Clark M D, Schmitz C A, Rosyara U R, Luby J J, and Bradeen J M. A consensus 'Honeycrisp' apple (Malus x domestica) genetic linkage map from three full-sib progeny populations. Tree Genetics & Genomes (2014) 10:627-639.

Collard B C Y and Mackill D J. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philosophical Transactions of the Royal Society B (2008) 363: 557-572.

Coombe B. The development of fleshy fruits. Annual Review of Plant Physiology (1976) 27:207-228.

Conner P J, Brown S K, Weeden N F. Randomly amplified polymorphic DNA-based genetic linkage maps of three apple cultivars. Journal of American Society of Horticultural Science (1997) 122(3):350-359.

Daccord N, Celton J-M, Linsmith G, Becker C, Choisne N, Schijlen Em, van de Geest H, Bianco L, Micheletti D, Velasco R, Di Pierro E A, Gouzy J, Rees D J G, Guerif P, Muranty H, Durel C-E, Laurens F, Lespinasse Y, Gaillard S, Aubourg S, Quesneville H, Weigel D, van de Weg E, Troggio M, and Bucher E. High-quality de novo assembly of the apple genome and methylome dynamics of early fruit development. Nature Genetics (2017) 49(7):1099-1108.

Edge-Garza D A, Luby J J, and Peace C. Decision support for cost-efficient and logistically feasible marker-assisted seedling selection in fruit breeding. Molecular Breeding (2015) 35:223.

Etienne A, Genard M, Lobit P, Mbeguie-A-Mbeguie D, and Bugaud C. What controls fleshy fruit acidity? A review of malate and citrate accumulation in fruit cells. Journal of Experimental Botany (2013) 64(6): 1451-1469.

Evans K, Guan Y, Luby J, Clark M, Schmitz C, Brown S, Orcheski B, Peace C, van de Weg E, and Iezzoni A. Large-scale standardized phenotyping of apple in RosBREED. Acta Hortic. (2012) 945:233-238.

Evans K and Peace C. Advances in marker-assisted breeding of apples. Burleigh Dodds Science Publishing Unlimited. 2017.

Gallardo R K, Hanrahan I, Yue C, McCracken V A, Luby J, McFerson J R, Ross C, and Carillo-Rodriguez L. Combining sensory evaluations and experimental auctions to assess consumers' preferences for fresh fruit quality characteristics. Agribusiness (2018) 34:407-425.

Garris, Amanda. "SnapDragon and RubyFrost are new apple varieties." Cornell Chronicle, Cornell University. Published August 1, 2013.

Ghafir S A M, Gadalla S O, Murajei B N, and El-Nady M F. Physiological and anatomical comparison between four different apple cultivars under cold-storage conditions. African Journal of Plant Science (2009) 3(6):133-138.

Gianfranceschi L, Seglias N, Tarchini R, Komjanc M, and Gessler C. Simple sequence repeats for the genetic analysis of apple. Theor Appl Genet (1998) 96:1069-1076.

Gorny J R and Kader A A. Controlled-atmosphere Suppression of ACC Synthase and ACC Oxidase in 'Golden Delicious' Apples during Long-term Cold Storage. Journal of American Horticultural Science (1996) 121(4):751-755.

Guan Y, Peace C, Rudell D, Verma S, and Evans K. QTLs detected for individual sugars and soluble solids content in apple. Mol Breeding (2015) 35:135.

Hamadziripi E T, Theron K I, Muller M, and Steyn W J. Apple Compositional and Peel Color Differences Resulting from Canopy Microclimate Affect Consumer Preference for Eating Quality and Appearance. Hort Science (2014) 49(3):384-392.

Harker F R, Marsh K B, Young H, Murray S H, Gunson F A, and Walker S B. Sensory interpretation of instrumental measurements. 2. Sweet and acid taste of apple fruit. Postharvest Biol. Technol (2002) 24(3):241-250.

Harker F R, Kupferman E M, Marin A B, Gunson F A, and Triggs C M. Eating quality standards for apples based on consumer preferences. Postharvest Biology and Technology (2008) 50:70-78.

Hemmat M, Weedon N F, Manganaris A G, and Lawson D M. Molecular Marker Linkage Map for Apple. Journal of Heredity (1994) 85(1):4-11

Howard N P, van de Weg E, Bedford D S, Peace C P, Vanderzande S, Clark M D, The S L, Cai L, and Luby J. Elucidation of the 'Honeycrisp' pedigree through haplotype analysis with a multi-family integrated SNP linkage map and a large apple (Malus x domestica) pedigree-connected SNP data set. Horticulture Research (2017) 4:17003.

Howard N P, van de Weg E, Tillman J, Tong C B S, Silverstein D A T, and Luby J J. Two QTL characterized for soft scald and soggy breakdown in apple (Malus x domestica) though pedigree-based analysis of a large population of interconnected families. Tree Genetics and Genomes (2018) 14(2).

Howard N P, Tillman J, Vanderzande S, and Luby J J. Genetics of zonal leaf chlorosis and genetic linkage to a major gene regulating skin anthocyanin production (MdMYB1) in the apple (Malus x domestica) cultivar Honeycrisp. PLoS ONE (2019) 14(3):e0213799.

Iezzoni A, Weebadde C, Luby J, Yue C, van de Weg E, Fazio G, Main D, Peace C P, Bassil N V, and McFerson J. RosBREED: Enabling Marker-Assisted Breeding in Rosaceae. Acta Hortic (2010) 859: 389-394.

Iezzoni A, Peace C, … and Yue C. RosBREED2: progress and future plans to enable DNA-informed breeding in the Rosaceae. Acta horticulturae (2017) 1172:115-118.

Iwanami H, Moriya S, Kotoda N, Mimida N, Takahashi-Sumiyoshi S, and Abe K. Mode of inheritance in fruit acidity analyzed with a mixed model of a major gene and polygenes using large complex pedigree. Plant Breeding (2012) 131:322-328.

Janick J, and Moore J N. Fruit Breeding, Volume I. Tree and Tropical Fruits. John Wiley & Sons Inc. 1996.

Jayasena V, and Cameron I. Brix/acid ratio as a predictor of consumer acceptability of crimson seedless table grapes. Journal of Food Quality (2008) 31:736-750.

Jia D, Shen F, Wang Y, Wu T, Xu X, Zhang X, and Han Z. Apple fruit acidity is genetically diversified by natural variations in three hierarchical epistatic genes: MdSAUR37, MdPP2CH and MdALMTII. The Plant Journal (2018) 95:427-443.

Kader A A, Zagory D, Kerbel E L, and Wang C Y. Modified atmosphere packaging of fruits and vegetables. Critical Reviews in Food Science and Nutrition (1989) 28(1):1-30.

Kenis K, Keulemans J, and Davey M W. Identification and stability of QTLs for fruit quality traits in apple. Tree Genetics & Genomes (2008) 4:647-661.

Khan S A, Beekwilder J, Schaart J G, Mumm T, Soriano J M., Jacobsen, and E., Schouten, H. J. Deifferences in acidity of apples are probably mainly caused by a malic acid transporter gene on LG16. Tree Genetics and Genomes (2013) 9:475-487.

Kidd F, West C, Griffiths D G, and Potter N A. Metabolism of Malic Acid in Apples. Journal of Horticultural Science (1951) 26(3):163-168.

King G J, Maliepaard C, Lynn J R, Alston F H, Durel C E, Evans K M, Griffon B, Laurens F, Manganaris A G, Schrevens E, Tartarini S, an Verhaegh J. Quantitative genetic analysis and comparison of physical and sensory descriptors relating to fruit flesh firmness in apple (Malus pumila Mill.). Theor Appl Genet (2000) 100:1074-1084.

Kondo S. Effect of Environmental Conditions on the Contents of Sugars and Ascorbic Acid in 'Senshu' Apple Fruit. Nippon Shokuhin Kogyo Gakkaishi (1992) 39(12):1112-1118.

Liebhard R, Koller B, Gianfranceschi L, and Gessler C. Creating a saturated reference map for the apple (Malus x domestica Borkh.) genome. Theor Appl Genet (2003a) 106:1497-1508.

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

Li M, Feng G, and Cheng L. Expression Patterns of Genes Involved in Sugar Metabolism and Accumulation during Apple Fruit Development. PLoS ONE (2012) 7(3):e33055

Liu Z, Bao D, Liu D, Zhang Y, Ashraf M A, and Chen X. Construction of a Genetic Linkage Map and QTL Analysis of Fruit-related Traits in an F1 Red Fuji x Hongrou Apple Hybrid. Open Life Sci. (2016) 11:487-497.

Maliepaard C, Alston F H, … King G J. Aligning male and female linkage maps of apple (Malus pumila Mill.) using multi-allelic markers. Theor Appl Genet (1998) 97:60-73.

Mangandi J, Verma S, Osorio L, Peres N A, van de Weg E, and Whitaker V M. Pedigree-Based Analysis in a Multiparental Population of Octoploid Strawberry Reveals QTL Alleles Conferring to Resistance to Phytophthora cactorum. Genes, Genomes, Genetics (2017) 7(6): 1707-1719.

Mora J R H, Micheletti D, … Aranzana J. Integrated QTL detection for key breeding traits in multiple peach progenies. BMC Genomics (2017) 18:404.

Nybom N. On the Inheritance of Acidity in Cultivated Apples. Hereditas (1959) 45:332-350.

Orcheski B, Parker R, and Brown S. Pale green lethal disorder in apple (Malus) is caused by a mutation in the PHYLLO gene which is essential for phylloquinone (vitamin K_1) biosynthesis. Tree Genetics & Genomes (2015) 11:131.

Peace C P, Luby J J, van de Weg W E, Bink M C A M, and Iezzoni A F. A strategy for developing representative germplasm sets for systematic QTL validation, demonstrated for apple, peach, and sweet cherry. Tree Genetics and Genomes (2014) 10:1679-1694.

Rodbotten M, Martinsen B K, Borge G I, Mortvedt H S, Knutsen S H, Lea P, and Naes T. A cross-cultural study of preference for apple juice with different sugar and acid contents. Food Quality and Preference (2009) 20(3):277-284.

Rymenants M, van de Weg E, Auwekerken A, De Wit I, Czech A, Nijland B, Heuven H, De Storme N and Keulemans W. Detection of QTL for apple fruit acidity and sweetness using sensorial evaluation in multiple pedigreed full-sib families. Tree Genetics and Genomes (2020) 16:71.

Shulaev V, Sargent D J, … Folta K M. The genome of the woodland strawberry (Fragaria vesca). Nature Genetics (2011) 43:109-116.

Sturm K and Stampar F. Seasonal Variation of Sugars and Organic Acids in Apple (Malus domestica Borkh.) in Different Growing Systems. Phyton (Austria) Special issue: "Plant Physiology" (1999) 39(3): 91-96.

Sun R, Chang Y, Yang F, Wang Y, Li H, Zhao Y, Chen D, Wu T, Zhang X, and Han Z. A dense SNP genetic map constructed using restriction site-associated DNA sequencing enables detection of QTLs controlling apple fruit quality. BMC Genomics (2015) 16:747.

Suni M, Nyman M, Eriksson N A, Bjork L, and Bjorck I. Carbohydrate composition and content of organic acids in fresh and stored apples. Journal of the Science of Food and Agriculture (2000) 80(10).

Ulrich J I. Organic Acids. The biochemistry of fruits and their products (1970). New York: Academic Press, 89-117.

United States Department of Agriculture. Fresh Apples, Grapes, and Pears: World Markets Trade. Updated June 2020. https://tinyurl.com/y63fwpl3

United States Department of Agriculture. Noncitrus Fruit – Production and value, price, yield per acre, and bearing acreage: Table 4-B. Updated October 31, 2019. https://tinyurl.com/y47wnu7y

United States Department of Agriculture. U.S. Apple Statistics –U.S. apple production, by variety, 1980 to date: Table 14. Updated May 1, 2012.

Vanderzande S, Howard N P, … and Peace C. 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.

Velasco R, Zharkikh A, …Viola R. The genome of the domesticated apple (Malus x domestica Borkh.) Nature Genetics (2010) 42(10): 833-839.

Verde I, Abbott A G, … Rokhsar D S. The high-quality draft genome of peach (Prunus persica) identifies unique patterns of genetic diversity, domestication and genome evolution. Nat Genet. (2013) 45(5):487-494

Verma S, Evans K, Guan Y, Luby J J, Rosyara U R, Howard N P, Bassil N, Bink M C A M, van de Weg E, and Peace C P. Two large-effect QTLs, Ma and Ma3, determine genetic potential for acidity in apple fruit: breeding insights from a multi-family study. Tree Genetics & Genomes (2019) 15:18.

Visser T, Verhaegh J J. Inheritance and selection of some fruit characters of apple. I. Inheritance of low and high acidity. Euphytica (1978) 27:753-760.

Wannemuehler, S D, Luby J J, Yue C, Bedford D S, Gallardo R K, and McCracken V A. A Cost-Benefit Analysis of DNA Informed Apple Breeding. HortScience (2019) 54(11):1998-2004.

Wilhite, Kate. “WSU announces the name for its stellar new apple.” WSU Insider, Washington State University. Published May 21, 2014.
Xu K, Wang A, and Brown S. Genetic characterization of the Ma locus with pH and titratable acidity in apple. Mol Breeding (2012) 30:899-912.
Yamaki S. Isolation of Vacuoles from Immature Apple Fruit Flesh and Compartmentation of Sugars, Organic Acids, Phenolic Compounds and Amino Acids. Plant and Cell Physiol. (1984) 25(1):151-166.
Zhang Q, Ma B, Li H, Chang Y, Han Y, Li J, Guochao W, Zhao S, Khan M A, Zhou Y, Gu C, Zhang X, Han Z, Korban S S, Li S, and Han Y. Identification, characterization, and utilization of genome-wide simple sequence repeats to identify a QTL for acidity in apple. BMC Genomics (2012) 13:537.
Zhen Q, Fang T, Peng Q, Liao L, Zhao L, Owiti A, and Han Y. Developing gene-tagged molecular markers for evaluation of genetic association of apple SWEET genes with fruit sugar accumulation. Horticulture Research (2018) 5:14.