University of Arkansas, Fayetteville
ScholarWorks@UARK

Graduate Theses and Dissertations

12-2021

Genetic Control of Sweetness, Acidity, and Seediness in Blackberry

Carly Elizabeth Godwin
University of Arkansas, Fayetteville

Follow this and additional works at: https://scholarworks.uark.edu/etd

 Part of the Fruit Science Commons, Horticulture Commons, and the Plant Breeding and Genetics Commons

Citation

Godwin, C. E. (2021). Genetic Control of Sweetness, Acidity, and Seediness in Blackberry. *Graduate Theses and Dissertations* Retrieved from https://scholarworks.uark.edu/etd/4286

This Thesis is brought to you for free and open access by ScholarWorks@UARK. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of ScholarWorks@UARK. For more information, please contact scholar@uark.edu.

Genetic Control of Sweetness, Acidity, and Seediness in Blackberry

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Horticulture

by

Carly Elizabeth Godwin
University of Oklahoma
Bachelor of Science in Botany, 2016
Oklahoma State University
Bachelor of Science Agriculture in Plant and Soil Science, 2019

December 2021
University of Arkansas

This thesis is approved for recommendation to the Graduate Council.

_____
Margaret Worthington, Ph.D.
Committee Chair

_____
Renee Threlfall, Ph.D.
Committee Member

_____
Leandro Mozzoni, Ph.D.
Committee Member

_____
Jacquelyn Lee, Ph.D.
Committee Member

ABSTRACT

The global blackberry (*Rubus* L. subgenus *Rubus* Watson) industry has experienced rapid growth during the past 15 years. Even so, many industry stakeholders report complaints from consumers and grocers stating blackberries are often too tart, too seedy, or not sweet enough for their liking. The development of molecular markers for high sweetness, low acidity, and reduced seediness would allow breeding programs to expeditiously make selection and crossing decisions in the early stages of the breeding pipeline. The objective of this study was to use a Genome-Wide Association Study (GWAS) to identify marker-trait associations, locate quantitative trait loci (QTL), and find possible candidate genes related to sweetness, acidity, and seediness in autotetraploid blackberries. A panel of 307 commercially-available cultivars and University of Arkansas System Division of Agriculture (UA) breeding selections grown at the Fruit Research Station in Clarksville, Arkansas, was phenotyped for soluble solids content (SSC) and pH in 2019, 2020, and 2021. Samples from a subset of 277 cultivars and breeding selections harvested during the summers of 2019 and 2020 were also evaluated for titratable acidity (TA), 100-seed weight, seed width to length (WL) ratio, seed area, and seeds per berry. The *Rubus argutus* Link. reference genome was used to design 35,054 Capture-Seq probes distributed across the genome, which were used to genotype the GWAS panel. Heritability estimates for flavor attributes concluded SSC had the lowest broad sense entry mean heritability at 61%, while pH and TA had heritabilities of 67% and 69%, respectively. Seediness attribute heritability estimates were 91% for 100-seed weight and WL ratio, 89% for seed area, and 70% for seeds per berry. Association analysis was performed in GWASpoly with 124,564 single nucleotide polymorphisms (SNPs) generated by Capture-Seq genotyping and a total of six QTL were identified. Three of the six QTL were related to flavor attributes; one for TA on Ra01, one for SSC on Ra02, and a shared

QTL for TA and pH on Ra05. The remaining three QTL were related to seediness; one for WL ratio on Ra01, one for 100-seed weight on Ra03, and one for seed area on Ra05. No significant markers or QTL were identified for seeds per berry. Ten possible candidate genes for blackberry flavor attributes were identified, including an H^+-ATPase 8 and a vacuolar proton-translocating pyrophosphatase associated with the QTL for TA on Ra01, a sucrose binding protein for the SSC QTL on Ra02, and two ALMT9 proteins, a MYB1, a PEPC, and three malate synthase genes for the shared TA and pH QTL on Ra05. Six seediness candidates were identified consisting of genes annotated for BRI1, CKX2, and GG3 associated with the WL ratio QTL on Ra01, an RPT2A and an AP2C1 in the QTL for 100-seed weight on Ra03, and a DA1 protein in the seed area QTL on Ra05. These results will be used to develop diagnostic markers for attributes related to sweetness, acidity, and seediness and as a training population for genomic selection.

©2021 by Carly Elizabeth Godwin
All Rights Reserved

ACKNOWLEDGEMENTS

First, I would like to thank Dr. Margaret Worthington, my advisor, for giving me the privilege and opportunity to work under her advisement during such a formative point in her career. Dr. Worthington taught me about several aspects of fruit breeding including everything from day-to-day proper plant care to making crosses and filing patents for new varieties.

Appreciation is also owed to Dr. Renee Threlfall, who guided me through my juice analysis process, Dr. Jacquelyn Lee, who directed operations at the fruit research station to ensure I was able to obtain berry samples, and Dr. Leandro Mozzoni who provided me with knowledge and practical application on the topic of plant breeding and analysis. Without these four individuals, my project would not have flourished as it did.

Additional appreciation is owed to the crew located at the Fruit Research Station in Clarksville, without whom obtaining healthy berry samples for my project would not have been possible. Hortifrut North America, Inc. provided financial assistance throughout my master's program, allowing me to focus solely on my research and education during this time, for which I am very grateful. To my work family: Carmen, Lacy, Kenneth, and Mason, thank you for all your positive words, encouragement, and summers full of memories that I will never forget.

Finally, I would like to thank everyone who helped support me mentally and emotionally from near and far throughout this process. The guidance, encouragement, and support given to me during this time meant the world to me. Mom and Dad, I know you both had struggles of your own in life and health during my time in graduate school, but this never stopped either of you from building me up, keeping me motivated, and loving me unconditionally. Amanda and Megan, without you two I would not be the woman I am today. Thank you for always being there for me and for making me the luckiest sister and aunt ever.

DEDICATION

To Ada,

Week after week, you devoted your time to help me grow into being a better student, daughter, sister, friend, and person. You taught me how to find my inner strength and never give up. Without you, I would not have made it to this day.

TABLE OF CONTENTS

LIST OF TABLES

Chapter I **Page**

Chapter II

LIST OF FIGURES

Chapter I **Page**

Chapter II

LITERATURE REVIEW

Botanical information

The blackberry is classified within the *Rubus* L. genus of the Rosaceae family (Clark et al., 2007; Ryu et al., 2014). Rosaceae is also known as the Rose family and is one of the largest plant families, exhibiting a wide variety of plant types. A few well-known Rosaceae crops include blackberries, raspberries (*Rubus idaeus* L.), apples [*Malus* × *sylvestris* (L.) Mill. var. *domestica* (Borkh.) Mansf.], and peaches [*Prunus persica* (L.) Batsch]. There are many species of blackberry found in the wild with varying levels of ploidy. Since cultivated blackberries are derived from many years of crosses between multiple wild species they are described as *Rubus* subgenus *Rubus* Watson, with no specific species epithet.

The *Rubus* genus has a center of origin in Asia, however, the majority of species in this region are raspberries in subgenus *Idaeobatus* (Clark et al., 2007). The flowers on *Rubus* plants have multiple ovaries, stigmas, and styles attached to the central receptacle, and upon maturation, the ovaries form into an aggregate fruit (Clark et al., 2007; Hummer, 2018). The *Rubus* genus consists of twelve subgenera, most of which are wild species (Foster et al., 2019). Blackberries (subgenus *Rubus*) and red and black raspberries (subgenus *Idaeobatus*) are the two commercial subgenera, and these two groups exhibit substantial horticultural differences from one another (Clark et al., 2007; Foster et al., 2019). When blackberries are picked, the torus is retained within the center of the aggregate fruit cluster. However, the raspberry fruit detaches from the torus when picked, leaving it on the plant.

Although blackberries are mentioned in gardening books as far back as the 1660s, the cultivation of blackberries was delayed because they were so readily available in the wild (Clark et al., 2007; Clark and Finn, 2011). Blackberry plants are perennial with typically biennial canes

requiring a period of dormancy before fruiting (Clark et al., 2007). Recently, new primocane-fruiting cultivars have been released, eliminating the need for a dormant period and allowing fruit production to occur on first-year canes. Primocane-fruiting blackberries possess several advantages such as increased yield; later season fruiting, effectively extending the harvest season; reduction of pruning costs; expansion of geographic area; and avoidance of winter damage (Clark, 2008, 2016).

Blackberry Breeding

The first public blackberry breeding program was established in 1908 at the Texas Agricultural Experiment Station, in College Station at Texas A & M University (Clark et al., 2007). Private blackberry breeding is noted in literature as far back as the 1880s. Judge James H. Logan was the first noted to breed and release a cultivar from a formal breeding effort when he created the 'Loganberry' (Clark et al., 2007). The oldest continuously active blackberry breeding program in the world is the USDA-ARS program located in Corvallis, Oregon, which was initiated in 1928 (Clark et al., 2007). Other blackberry breeding programs include public breeding programs at John Innes in Scotland, Agriculture and Agri-food of Canada, New Zealand HortResearch, New York State Agricultural Experiment Station at Cornell University, North Carolina State University, and the University of Arkansas System Division of Agriculture (UA), as well as private breeding companies such as Driscoll's Strawberry Associates, Planasa, and Hortifrut North America, Inc. (Clark et al., 2007).

Between the years 1985 and 2016, there were over 60 new cultivars of blackberries released from breeding programs in the United States (Clark, 2016). Blackberry production has increased, in part, due to breeding programs enhancing desirable traits like flavor and texture while reducing undesired traits in the fruits and plants such as bitterness and thorniness (Finn and

Clark, 2017). There are different architectures seen in blackberry plants including erect, trailing, and semi-erect. Erect blackberries produce new primocanes from buds on the crown or buds on roots and do not require a trellis for support, whereas trailing and semi-erect blackberries only produce primocanes from the crown and require the use of a trellis (Clark, 2005).

The blackberry breeding program at UA began in 1964 under the direction of James N. Moore (Clark et al., 2007). In the Eastern United States, the UA breeding program has been the main producer of unique and improved blackberry cultivars (Clark, 2008). The UA breeding program released the first thorny, erect blackberry cultivars exhibiting high yield and good fruit quality; these cultivars are 'Cherokee', 'Comanche', and 'Cheyenne' (Clark et al., 2007). The sources of erect growth habit and thornlessness used in the UA breeding program were associated with several negative traits, including very tart flavor (Clark, 2005). However, these initial barriers were overcome by persistent crossing and selection over many years. The breeding program also developed the first primocane-fruiting cultivars (Clark, 2008). Current breeding objectives include enhanced quality and flavor and reduced bitterness, red drupelet reversion, and seediness.

Economic Value of Blackberries

Blackberries are on the rise in the world market, due in part to growing consumer interest in nutritious foods. Blackberries have higher levels of anthocyanins than many other fruit crops (Clark et al., 2007). Anthocyanins were studied along with other flavonoids in epidemiological and clinical studies which suggested the consumption of these natural compounds could decrease the risk of obesity, coronary heart disease, degenerative conditions, and various types of cancers (Kaume et al., 2012). The growth of the global blackberry industry, however, is limited by their postharvest quality. Blackberries are considered one of the most difficult fruits to transport and

sell due to their tendency to soften and leak during postharvest storage (Perkins-Veazie et al., 1999; Salgado and Clark, 2016). Blackberries rapidly deteriorate in storage and only last approximately one week when stored in a refrigerator (Edgley et al., 2019).

Blackberries are produced across the globe in North, South, and Central America, especially in Mexico, as well as Australia, New Zealand, Asia, Europe, and South Africa (Clark et al., 2007). The expansion of blackberry production to new regions has facilitated year-round product availability in many markets around the world. Often in areas where red raspberries are produced, blackberry production is a simple transition. Red raspberries are expensive to produce because they require frequent replanting and heavy pest control inputs. Switching to blackberry production eliminates the need to replant as frequently and reduces inputs (Clark et al., 2007). In 2005, worldwide production of cultivated blackberries was estimated at 20,000 ha with an additional 8,000 ha of wild plants harvested (Strik et al., 2007). Commercial blackberry production had reached an estimate of 154,578 tons (1.4×10^8 kg) annually worldwide (Strik, 2007). From August 2020 to August 2021, approximately 46,473 tons (4.2×10^7 kg) of blackberries were sold in the United States alone with Americans spending just over $671 million on blackberries (California Strawberry Commission, 2021).

Blackberry Flavor Composition

The flavor of blackberries is determined by the chemistry of different compounds found in these fruits. Blackberries contain three principal sugars, glucose, fructose, and sucrose, with the most prominent sugars being glucose and fructose. The levels of these sugars present within a given berry change throughout the ripening process (Kafkas et al., 2006; Fan-Chiang and Wrolstad, 2010). The primary organic acid found in blackberries is citric acid (Finn and Clark, 2017), but some research has also noted the main acid in blackberries as malic rather than citric

(Kafkas et al., 2006). Several other organic and phenolic acids are also detected in blackberries, such as ascorbic, lactoisocitric, and isocitric acids (Fan-Chiang and Wrolstad, 2010).

Blackberries also contain flavonols, which increase bitterness and affect the flavor experienced by consumers (Lawless et al., 2012). Flavonols exist in blackberries and blueberries as nine quercetin and three kaempferol derivatives as well as two acylated quercetin-derived compounds (Cho et al., 2005; Kaume et al., 2012). Flavonols play a major role in both plant and human protection against oxidative stress promoted by free radical species. Since this stress plays a major role in many chronic diseases, increasing consumption of foods rich in flavonols could reduce incidence and mortality rates of chronic diseases (Cho et al., 2005). However, bitterness was shown to cause a decrease in the overall liking of blackberries in a study where participants drank 100% blackberry juice along with other juice mixture treatments (Lawless et al., 2012).

Sweetness is one attribute known to play a large role in the overall consumer experience and acceptance of blackberries (Threlfall et al., 2016; Dunteman et al., 2018). A study conducted at UA suggested fresh-market blackberries having a medium-level ratio of sweetness-to-sourness rather than a low- or high-level ratio were preferred, although more consumers than expected preferred blackberries with extreme ratios (Dunteman et al., 2018). An earlier study showed that blackberry pH had a negative correlation with titratable acidity ($r = -0.91$), while soluble solids were not correlated with pH or titratable acidity (Threlfall et al., 2016).

Volatile organic compounds (VOCs) are also known to have an impact on consumer perceptions and preferences of blackberry flavor. Unlike sweetness, acidity, and bitterness, which are perceived by taste receptors on the tongue, VOCs are perceived through smell detected by olfactory receptors of the mouth and nose (Threlfall et al., 2016; Klee and Tieman, 2018). The

olfactory system and odor thresholds of individuals vary widely, making olfactory perception a difficult trait to quantify (Hasin-Brumshtein et al., 2009). While many different VOCs affect blackberry flavor, it is not yet clear which specific compounds are the main drivers of consumer preferences (Klee and Tieman, 2018). A study conducted in blueberry showed lipid-derived volatiles explained 15% of overall liking scores in a sensory panel and the carotenoid/terpene compound group explained 21% of the overall liking score (Colantonio et al., 2020). Because of the complexity and cost associated with GC/MS and sensory analysis, VOCs are not a current focus in blackberry genetic studies.

Blackberry flavor, while strongly genetic, differs depending on the growing region and environmental conditions (Clark et al., 2007). These factors affect sweetness, acidity, and flavor volatiles. 'Chickasaw' grown in Arkansas had a sweet, slightly grassy, and bitter flavor, whereas the same cultivar grown in Oregon was intensely aromatic with a sweet/acid balance (Wang et al., 2005). The most important volatile compounds found in the Oregon-grown fruit were ethyl 2-methylpropanoate, methyl butanoate, ethyl hexanoate, 2,3-butanedione, α-pinene, limonene, and hexanal (Wang et al., 2005). Eastern blackberries have a very different flavor volatile profile from the Pacific Northwest blackberries such as 'Marion' which contain germplasm from *R. ursinus* (Clark, 2005).

Seediness in Blackberries

Seed size as well as the overall feel of seediness during consumption is a common attribute for breeders to focus on when evaluating blackberries (Clark et al., 2007). Blackberries are aggregate fruits composed of a cluster of drupelets surrounding a soft tissue receptacle, also called a torus. Each drupelet is comprised of an exocarp, a mesocarp, and a lignified endocarp (or pyrene) which encloses a single seed (Tomlik-Wyremblewska et al., 2010). Seed size varies

widely across the *Rubus* genus. Wada and Reed (2008) found a high level of morphological and micromorphological diversity in *Rubus* seeds collected across 56 taxa and 10 subgenera. The size of *Rubus* seeds ranged from dimensions of 1.4×1.2×0.81 mm to 6.0×3.5×2.35 mm (Wada and Reed, 2008). Seed size in *Rubus* is likely influenced by genetics (Hummer and Peacock, 1994).

While blackberry seeds tend to be larger than raspberry seeds, substantial variation exists within the *Rubus* and *Idaeobatus* subgenera. Wada et al. (2011) compared three species each from the *Rubus* and *Idaeobatus* subgenera and found variable 100-seed weight in subgenus *Rubus* species *R. caesius* L. (0.37 g), *R. georgicus* Focke (0.26 g), and *R. ursinus* Cham. & Schltdl. (0.12 g) and subgenus *Idaeobatus* species *R. occidentalis* (0.19 g), *R. coreanus* (0.10 g), and *R. hoffmeisterianus* (0.04 g). In another study, seeds of the American blackberry species *R. allegheniensis* and *R. ursinus* and the European blackberries *R. procerus* Muller and *R. caesius* L. were found to be larger than seeds of the European raspberry from subgenus *Idaeobatus* (Hummer and Peacock, 1994). Wada et al. (2011) also observed that seed hardness and seed coat thickness was variable within subgenus *Rubus* and subgenus *Idaeobatus.* Hardness was rated from 1 to 5 described as follows: 1 = soft; 2 = slightly hard; 3 = hard; 4 = very hard; and 5 = extremely hard, with subgenus *Rubus* species *Rubus caesius*, *R. georgicus,* and *R. ursinus* having ratings of 3, 4, and 5, and subgenus *Idaeobatus* species, *R. coreanus*, *R. hoffmeisterianus*, and *R. occidentalis*, having ratings of 2, 2, and 5 respectively (Wada et al., 2011).

Rubus seed coats have three distinct layers of heavily lignified sclerenchymatous cells (Wada et al., 2011). The endotesta of the seeds have uniformly globular, isodiametric, macrosclereid layers running perpendicularly to the mesotesta, and the outer mesotesta is composed of irregularly shaped macrosclerids (Wada et al., 2011). The complexity and thickness of the outer epidermis (exotesta) and the alternating orientation of the macrosclerenchyma

provide strength to the seed coat. The species with the hardest seed coat, *R. occidentalis*, had a thick, multilayered exotesta and the most heavily lignified and intricate sclereid composition, whereas the soft seeded *R. hoffmeisterianus* had an exotesta only two cell layers thick (Wada et al., 2011).

Seed attributes also vary across cultivated blackberries. Generally, trailing blackberries developed in the Western United States have seeds presenting as ellipsoidal and small compared to Eastern semi-erect blackberries, which have "clam shaped" seeds (Clark et al., 2007). Western blackberries are often perceived as "seedless" or having a lower number of seeds compared to Eastern blackberries, and progenies of crosses between Eastern and Western blackberries tend to show a range of seediness (Clark et al., 2007). Seeds of wild diploid *R. allegheniensis* blackberries were also lighter than the seeds from many closely related Eastern semi-erect cultivated tetraploid breeding selections (Hummer and Peacock, 1994), indicating that ploidy may affect seed size.

Variation in seediness and seed attributes also exists within Eastern erect and semi-erect breeding selections. A descriptive panel rated seediness of 22 commercially-available blackberry cultivars and UA breeding selections (genotypes) grown at the UA Fruit Research Station (FRS) on a 10-point scale (0 = no seeds to 9 = extremely seedy) (Sebesta et al., 2013). The descriptive panelists did not find significant differences among the genotypes but rated seediness of UA blackberries as ranging from 4.38 (A-2416 and 'Tupy') to 7.25 (A-2418). Cultivars falling in the higher end of the seediness scale were 'Prime-Ark®45' at 6.63, and 'Natchez' which was rated 6.75. 'Natchez' had the highest pyrenes per berry at 131, whereas 'Navaho' and 'Tupy' had the lowest average at 53 pyrenes per berry (Sebesta et al., 2013). Berry weight of the 22 genotypes varied from 5.1 g (A-2252) to 9.6 g (A-2434), while dry pyrene weight of the same evaluated

genotypes ranged from 160 mg (‘Tupy’) to 491 mg (‘Natchez’) (Sebesta et al., 2013). The pyrene weight to total berry weight ratio may have been noticeable to consumers as ‘Prime-Ark® 45’ had a high proportion of pyrene weight to berry weight and was scored by the descriptive sensory panel among the highest for overall seediness (Sebesta et al., 2013).

In 2016, a group of UA blackberry genotypes were evaluated by a descriptive sensory panel and a consumer panel. The descriptive sensory panel evaluated different texture attributes, including size and amount of seeds in A-2416, A-2418, A-2434, A-2450, A2453, A-2491, ‘Natchez’, ‘Osage’, ‘Ouachita’, ‘Prime-Ark® 45’, and ‘Prime-Ark® Traveler’ (Threlfall et al., 2016). Similar to Sebesta et al. (2013), the descriptive panelists in this study were also unable to identify differences in the size of pyrenes, amount of pyrenes, or loose particles across the genotypes evaluated (Threlfall et al., 2016). The genotypes evaluated in this study differed significantly in pyrenes per berry as well as berry weight. A-2453 had the lowest berry weight at 6.01g, and ‘Natchez’ had the highest at 14.26 g, while pyrenes per berry ranged from 51 to 115 in A-2453 and ‘Natchez’, respectively (Threlfall et al., 2016). A-2453, the smallest genotype in the study, was considered too seedy by 36% of consumer panelists, whereas, ‘Natchez’, the largest berry in the study, was rated as ‘Just About Right’ (JAR) by 90% of panelists (Threlfall et al., 2016). These findings indicate the impact of berry size concerning the overall perception of seediness with consumers perceiving larger berries as being less seedy than smaller berries with similar seediness attributes. According to a principal component analysis (PCA), amount of seeds was a negative driver of the overall liking of blackberries in the consumer panel (Threlfall et al., 2016).

Genomic Resources for Blackberry Breeding

Blackberries are highly heterozygous and have a ploidy ranging between 2n=2x=14 and 2n=14x=98 (Clark et al., 2007). While Eastern US blackberry cultivars are mostly tetraploids, Western cultivars typically range in ploidy from 6x to 9x (Clark et al., 2007). Blackberries are autopolyploids, which can make breeding efforts more difficult. Autopolyploidy and polysomic inheritance can be challenging for applied breeding programs because of complicated segregation ratios for Mendelian traits and reduced fertility among crosses between parents with different ploidy levels (Meng and Finn, 2002). Autopolyploidy also complicates genetic research and molecular breeding in blackberry (Foster et al., 2019). The chromosome-scale genome assembly of black raspberry along with a new sequencing initiative with two diploid relatives within the *Rubus* subgenus (*R. argutus* and *R. ulmifolius*) will allow for rapid advances in blackberry genetic resources (VanBuren et al., 2018; Worthington et al., 2020). The available *Rubus* resources paired with the development of new software for genetic analysis in polyploid plants will allow for blackberry researchers to conduct quantitative trait loci (QTL) mapping and genome-wide association studies (GWAS) for important breeding traits (Foster et al., 2019; Worthington et al., 2020).

Several molecular marker systems have been implemented in *Rubus*. Many studies have applied Polymerase Chain Reaction (PCR)-based markers such as randomly amplified polymorphisms (RAPDs), amplified fragment length polymorphisms (ALFPs), expressed sequence tags (ESTs), and simple sequence repeats (SSRs) as well as multiplexed fingerprinting assays, and next-generation sequencing (NGS) technology for *de novo* marker development (Graham et al., 1997; Dossett et al., 2012; Miyashita et al., 2015; Molina-Bravo et al., 2019). Molecular markers have been used in the past in blackberries to identify cultivars, validate

pedigree records, and construct genetic linkage maps (Stafne et al., 2003; Castro et al., 2013; Bushakra et al., 2015). The development of closely related reference genomes has also facilitated high throughput genotyping strategies including genotyping-by-sequencing (GBS) and target capture sequencing (Foster et al., 2019).

Polyploids, unlike their diploid counterparts, carry more than two copies of each chromosome and require specialized tools to properly call allele dosage and carry out genetic studies (Bourke et al., 2018). Genetic studies such as a GWAS are a powerful tool for breeders to identify marker-trait associations and have become commonplace in diploid species. A specialized software package named GWASpoly has recently been created for the implementation of GWAS in autopolyploids (Rosyara et al., 2016). The purpose of conducting a GWAS is to understand the variation of complex traits by relating genotypic data consisting of a large number of markers, typically single nucleotide polymorphisms (SNPs), to measured phenotypic data while controlling for population structure (Ball, 2013). With any GWAS, the standard approach is to select and genotype a large sample size of your tested species on which several phenotypic measurements will be obtained. Each genetic locus contributes a varying amount of control on a phenotype, typically exhibiting a large or small effect. Having a large sample increases the number of polymorphisms observed in the genotypic data of the sample set, therefore increasing the likelihood of relationships being found between genotypic composition and phenotypic outcome, especially when dealing with small-effect loci (Flint, 2013).

There are many approaches to choosing a GWAS panel, one is using a “star-like design” including geographically distant individuals (Korte and Ashley, 2013). This geographically distant approach, however, comes with some setbacks as this might introduce genetic heterogeneity because of factors such as adaptation to growing regions (Korte and Ashley,

2013). A second approach often used for choosing a panel for a GWAS study is to sample a local population known to have a wide range of phenotypic diversity (Korte and Ashley, 2013).

Genetic Control of Flavor and Seediness

Compared to other Rosaceae crops, there are few genomic and genetic resources available for the improvement of *Rubus* species (Foster et al., 2019; Worthington et al., 2020). Very little research has been conducted on the genetic control of sweetness, acidity, and seediness in *Rubus.* Fortunately, there is a high degree of genome conservation between taxa across the Rosaceae family, and information on QTL controlling these traits in other fruit crops can be applied to blackberries.

Strawberries (*Fragaria vesca*, and *F.* x *ananassa*) serve as one of many model species for Rosaceae crops due to their short generation time, ease of vegetative propagation, and small size as compared to widely-grown woody species such as apple or peach (Shulaev et al., 2011). In strawberries, a single gene – *FaFAD1*, which is associated with "peachy flavor" specifically, was isolated and shown to likely be a key factor in controlling flavor volatility (Chambers et al., 2014). Studies in red raspberry also found QTL associated with flavor volatiles (Paterson et al., 2013).

Malic acid (*Ma*), the predominant acid in apples, is controlled by the *Ma* gene, initially mapped to the distal end of linkage group L16 (Maliepaard et al., 1998). Further studies fine mapped the *Ma* locus in apple using a sequence-based approach mapping the *Ma* gene to a chromosomal region no larger than 150 kb containing 44 predicted genes on LG 16 (Xu et al., 2012), with an additional mapping study refining the locus to a 65kb region (Bai et al., 2012). The *Ma* locus accounts for 17% to 42% of the variation in fruit acidity in apples (Maliepaard et al., 1998; Liebhard et al., 2003; Xu et al., 2012; Li et al., 2020). Data from the fine-mapping

study suggested apple pH and TA were under the control of the same major *Ma* gene. The *Ma* allele associated with high/medium acidity fruit had incomplete dominance over the recessive low-acid *ma* allele (Xu et al., 2012). Two aluminum-activated malate transporter (ALMT)-like genes, *Ma1* and *Ma2,* located in the *Ma* locus on LG 16 were identified as strong candidate genes for fruit acidity (Bai et al., 2012). Expression of *Ma1* was correlated significantly with acidity levels, whereas *Ma2* was more consistently expressed at low levels in both high and low acidity fruit (Bai et al., 2012). A later study of apples also identified and located an additional locus, *Ma3*, on chromosome 8 exhibiting additive allele dosage with the *Ma1* gene, together explaining approximately 66% of the variation in apple fruit acidity (Verma et al., 2019).

Plant ALMTs transport malate across vacuolar cell membranes through facilitated diffusion of an electrical potential gradient ($\Delta\Psi$) regulating vacuolar storage and concentration of malate and citrate in cells, with citrate transport occurring far slower than malate (Etienne et al., 2013). Malate transport relies on the presence of a $\Delta\Psi$ to occur, with the $\Delta\Psi$ expected to decrease as vacuolar pH decreases, effectively closing the transport channels at a low vacuolar pH to prevent over-acidification of already highly acidic vacuoles (Etienne et al., 2013). The non-truncated *Ma1* (also called *Ma1G*) allele in apple codes for a full-length ALMT9 transporter, while the truncation of *Ma1* to the recessive *ma1* allele causes low acidity in apple, indicating the importance of ALMTs in the control of fruit acidity (Li et al., 2020). A study of *Vitis* (grapes) acidity also showed the ability of VvALMT9 to mediate malate accumulation in the vacuole and counteract excessive organic acid decompartmentation during maturation (De Angeli et al., 2013).

The genetic control of acidity was also studied in peaches, with a major QTL localized to a 0.4 cM genetic interval corresponding to a 100kb region and named the *D* locus (D for ‘Doux’,

the French word for sweet) located at the proximal end of LG 5 (Boudehri et al., 2009). The *D* allele, which is associated with reduced malic and citric acid content in peaches is dominant (Boudehri et al., 2009). An encoding auxin efflux carrier family protein, *Prupe*_5G004300/*ppa006339m*, in the *D* locus of peach was more highly expressed in mature stages of acidic fruits than non-acidic fruits, making it a candidate gene for the control of fruit acidity (Cao et al., 2016). Another possible candidate gene in the *D* locus was the pyruvate small protine *Prupe_5G008400*, which showed consistency between gene expression profile and fruit acidity and was designated as a pH regulator in peach (Wang et al., 2021). *Prupe_5G008400* was posited as a stronger candidate for pH regulation in peach than *Prupe_5G004300* because the transient ectopic expression of *Prupe_5G004300* had no impact on pH or organic acid accumulation (Wang et al., 2021). Results from the transient ectopic expression test along with the discovery of a new strong candidate gene suggest further studies are necessary to fully understand the genetic control of fruit acidity in peach (Wang et al., 2021).

While genetic loci associated with major differences in fruit acidity have been mapped in apple and peach, the QTL associated with sweetness in Rosaceae crops tend to have a smaller effect on phenotypic variation and are often inconsistent across environments. Studies of fruit acidity and sweetness show that acidity is much more heritable than sweetness, which is generally known to have low heritability estimates with QTLs explaining 35% or less of phenotypic variation (Lerceteau-Köhler et al., 2012; Shaw, 2019; Zurn et al., 2020).

Individual sugars in apple were controlled by several QTL across the genome, with fructose QTL mapping to LGs 1, 3, and 15, glucose QTL located on LGs 1, 2, 3, 15, and 16, and sucrose QTL on LGs 1, 3, 4, 9, and 12. The QTL on LG 1 was particularly promising and consistent across harvest years for glucose, fructose, and sucrose (Guan et al., 2015). A study by

Zhen et al. (2018) identified 25 sugar transporters called *SWEETs* in the apple genome. The study identified two likely candidates involved with sugar accumulation in apple fruit; *MdSWEET9b* located on LG4 located in the region of a QTL for individual sugar content, and *MdSWEET15a* located on LG16 in a locus with several overlapping BRIX, sorbitol, and fructose QTL (Zhen et al., 2018). In peaches, QTL for individual sugars have been detected in a biparental population segregating for sugar content. A sucrose QTL on LG 5 and glucose and fructose QTLs on LG 4 were consistent across harvest years (Zeballos et al., 2016).

Zurn et al. (2020) identified the first environmentally stable sweetness QTL in blackberry. In this study, sugar-related genes from conserved QTL associated with sweetness-related traits for *Fragaria*, *Malus*, and *Prunus* and 798 unique genes sugar-associated genes from apple (Quilot et al., 2004; Lerceteau-Köhler et al., 2012; Bushakra et al., 2015; Guan et al., 2015; Li et al., 2016; Jung et al., 2019) were used to design HybSeq probes and sequence a set of 20 high (>11.5 ºBrix) and 20 low (≤ 11.5 ºBrix) SSC genotypes from the UA breeding program and USDA-ARS Horticultural Crops Research Unit (HCRU) program. Sequences associated with variation in acidity in the initial set of 40 genotypes were used to design KASP primers and genotype biparental populations from the UA and USDA-ARS HCRU breeding programs, which were evaluated for SSC over two years. Most alleles identified in this study had a negative influence on SSC, with 48 significant alleles mapping to 16 regions on all chromosomes except chromosome 3. Markers associated with SSC were found in two syntenic regions with other Rosaceae sweetness QTL on chromosomes 4 and 6 (Zurn et al., 2020). Three significant markers, BBS_SNP45, BBS_INDL31, and BBS_SNP46, were found on chromosome 1 in a 736 bp region. The QTL associated with these markers, named qSCC-Ruh-ch1.1 was significant in three of four environments and accounted for a 1.46 ºBrix difference in SSC. Two overlapping

genes were present in the qSCC-Ruh-ch1.1 gene space, marker-Ro01-snap-gene-149.62, and marker_Ro01_snap_gene-149.66. The marker_Ro01_snap_gene-149.66 gene had homology with glycosyl-transferase and sucrose synthase genes, which have previously been implicated in the accumulation of sugars and starches in other plants (Zurn et al., 2020).

Very little has been published on the genetic control of seediness and pyrene attributes in *Rubus* or other Rosaceae crops. Genetic variability for pyrene size exists within cultivated blackberry germplasm, and Moore et al. (1975) found pyrene size in blackberries was highly heritable with partial dominance for small pyrene size. With this knowledge, it should be possible to identify QTL for pyrene attributes in blackberry. A study published in 2020 found that drupelet count in black raspberry had a strong genetic influence, and the genotype × environment interaction was an unimportant variance component (Willman et al., 2020). Elimination of the genotype × environment component in the genetic control of drupelet count suggests breeders can select for this trait regardless of the environment in which a plant is grown. Significant marker-trait associations were found for drupelet count on chromosomes 1, 4, 5, and 7, with the most highly significant marker located on chromosome 1 (Willman et al., 2020).

With new genetic resources under development and synteny with other Rosaceae crops, it should be possible to identify QTL or candidate genes for flavor and fruit quality traits and seed attributes in blackberry. Genomic breeding for fruit quality, fruit flavor, and seediness attributes can allow selections to be made in the early stages of the breeding process resulting in expedited development of more flavorful and less seedy new cultivars.

Literature Cited

Bai, Y., L. Dougherty, M. Li, G. Fazio, L. Cheng, and K. Xu. 2012. A natural mutation-led truncation in one of the two aluminum-activated malate transporter-like genes at the *Ma* locus is associated with low fruit acidity in apple. Mol. Genet. Genomics 287:663–678.

Ball, R.D. 2013. Designing a GWAS: power, sample size, and data structure. Methods in Mol. Biol. vol. 1019. Clifton, NJ. 37–98.

Boudehri, K., A. Bendahmane, G. Cardinet, C. Troadec, A. Moing, and E. Dirlewanger. 2009. Phenotypic and fine genetic characterization of the *D* locus controlling fruit acidity in peach. BMC Plant Biol. 9:59.

Bourke, P.M., R.E. Voorrips, R.G.F. Visser, and C. Maliepaard. 2018. Tools for genetic studies in experimental populations of polyploids. Front. Plant Sci. 9:513.

Bushakra, J.M., D.W. Bryant, M. Dossett, K.J. Vining, R. VanBuren, B.S. Gilmore, J. Lee, T.C. Mockler, C.E. Finn, and N.V. Bassil. 2015. A genetic linkage map of black raspberry (*Rubus occidentalis*) and the mapping of Ag_4 conferring resistance to the aphid *Amphorophora agathonica*. Theor. Appl. Genet. 128:1631–1646.

California Strawberry Commission. 2021. U.S. Retail Category Trends. 23 August 2021. <https://www.calstrawberry.com/en-us/market-data/retail-category-trends/>.

Cao, K., Z. Zhou, Q. Wang, J. Guo, P. Zhao, G. Zhu, W. Fang, C. Chen, X. Wang, X. Wang, Z. Tian, and L. Wang. 2016. Genome-wide association study of 12 agronomic traits in peach. Nat. Commun. 7:13246.

Castro, P., E.T. Stafne, J.R. Clark, and K.S. Lewers. 2013. Genetic map of the primocane-fruiting and thornless traits of tetraploid blackberry. Theor. Appl. Genet. 126:2521–2532.

Chambers, A.H., J. Pillet, A. Plotto, J. Bai, V.M. Whitaker, and K.M. Folta. 2014. Identification of a strawberry flavor gene candidate using an integrated genetic-genomic-analytical chemistry approach. BMC Genomics 15:217.

Cho, M.J., L.R. Howard, R.L. Prior, and J.R. Clark. 2005. Flavonol glycosides and antioxidant capacity of various blackberry and blueberry genotypes determined by high-performance liquid chromatography/mass spectrometry. J. Sci. Food Agric. 85:2149–2158.

Clark, J.R. 2005. Intractable traits in eastern U.S. blackberries. HortScience 40:1954–1955.

Clark, J.R. 2008. Primocane-fruiting blackberry breeding. HortScience 43:1637–1639.

Clark, J.R. 2016. Breeding southern US blackberries, idea to industry. Acta Hort. 1133:3–11.

Clark, J.R., and C.E. Finn. 2011. Blackberry breeding and genetics. Fruit, Veg. Cereal Sci. Biotechnol. 5:27–43.

Clark, J.R., E.T. Stafne, H.K. Hall, and C.E. Finn. 2007. Blackberry breeding and genetics. Plant Breeding Rev. 29:19–144.

Colantonio, V., L.V. Felipe Ferrão, D. Tieman, N. Bliznyuk, C. Sims, H. Klee, P.R. Munoz, and M.F. R Resende Jr. 2020. Metabolomic selection for enhanced fruit flavor. BioRxiv.

De Angeli, A., U. Baetz, R. Francisco, J. Zhang, M.M. Chaves, and A. Regalado. 2013. The vacuolar channel VvALMT9 mediates malate and tartrate accumulation in berries of *Vitis vinifera*. Planta 238:283–291.

Dossett, M., N.V. Bassil, K.S. Lewers, and C.E. Finn. 2012. Genetic diversity in wild and cultivated black raspberry (*Rubus occidentalis* L.) evaluated by simple sequence repeat markers. Genet. Resour. Crop Evol. 59:1849–1865.

Dunteman, A.N., R.T. Threlfall, J.R. Clark, and M.L. Worthington. 2018. Evaluating consumer censory and composition attributes of Arkansas-grown fresh-market blackberries. Discov. Student J. Dale Bump. Coll. Agric. Food Life Sci. 19:16–23.

Edgley, M., D.C. Close, and P.F. Measham. 2019. Effects of climatic conditions during harvest and handling on the postharvest expression of red drupelet reversion in blackberries. Sci. Hortic. 253:399–404.

Etienne, A., M. Génard, P. Lobit, D. Mbeguié-A-Mbéguié, and C. Bugaud. 2013. What controls fleshy fruit acidity? A review of malate and citrate accumulation in fruit cells. J. Exp. Bot. 64:1451–1469.

Fan-Chiang, H.J., and R.E. Wrolstad. 2010. Sugar and nonvolatile acid composition of blackberries. J. AOAC Int. 93:956–965.

Finn, C.E., and J.R. Clark. 2017. Cultivar development and selection, p. 63–92. In: H.K. Hall, R.C. Funt (eds.). Blackberries and their hybrids. CABI, Boston, MA.

Flint, J. 2013. GWAS. Curr. Biol. 23:R265–R266.

Foster, T.M., N.V. Bassil, M. Dossett, M.L. Worthington, and J. Graham. 2019. Genetic and genomic resources for Rubus breeding: a roadmap for the future. Hortic. Res. 6:116.

Graham, J., G.R. Squire, B. Marshall, and R.E. Harrison. 1997. Spatially dependent genetic diversity within and between colonies of wild raspberry Rubus idaeus detected using RAPD markers. Mol. Ecol. 6:1001–1008.

Guan, Y., C. Peace, D. Rudell, S. Verma, and K. Evans. 2015. QTLs detected for individual sugars and soluble solids content in apple. Mol. Breed. 35(135):1–13.

Hasin-Brumshtein, Y., D. Lancet, and T. Olender. 2009. Human olfaction: from genomic variation to phenotypic diversity. Trends Genet. 25:178–184.

Hummer, K.E. 2018. Blackberries: an introduction, p. 1–16. In: H.K. Hall, R.C. Funt (eds.). Blackberries and their hybrids. CABI, Boston, MA.

Hummer, K.E., and D.N. Peacock. 1994. Seed dimension and weight of selected Rubus species. HortScience 29:1034–1036.

Jung, S., T. Lee, C.H. Cheng, K. Buble, P. Zheng, J. Yu, J. Humann, S.P. Ficklin, K. Gasic, K. Scott, M. Frank, S. Ru, H. Hough, K. Evans, C. Peace, M. Olmstead, L.W. DeVetter, J. McFerson, M. Coe, J.L. Wegrzyn, M.E. Staton, A.G. Abbott, and D. Main. 2019. 15 years of GDR: New data and functionality in the genome database for Rosaceae. Nucleic Acids Res. 47:D1137–D1145.

Kafkas, E., M. Koşar, N. Türemiş, and K.H.C. Başer. 2006. Analysis of sugars, organic acids and vitamin C contents of blackberry genotypes from Turkey. Food Chem. 97:732–736.

Kaume, L., L.R. Howard, and L. Devareddy. 2012. The blackberry fruit: A review on its composition and chemistry, metabolism and bioavailability, and health benefits. J. Agric. Food Chem. 60:5716–5727.

Klee, H.J., and D.M. Tieman. 2018. The genetics of fruit flavour preferences. Nat. Rev. Genet. 19:347–356.

Korte, A., and F. Ashley. 2013. The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9:29.

Lawless, L.J.R., R.T. Threlfall, L.R. Howard, and J.F. Meullenet. 2012. Sensory, compositional, and color properties of nutraceutical-rich juice blends. Amer. J. Enol. Vitic. 63:529–537.

Lerceteau-Köhler, E., A. Moing, G. Guérin, C. Renaud, A. Petit, C. Rothan, and B. Denoyes. 2012. Genetic dissection of fruit quality traits in the octoploid cultivated strawberry highlights the role of homoeo-QTL in their control. Theor. Appl. Genet. 124:1059–1077.

Li, C., L. Dougherty, A.E. Coluccio, D. Meng, I. El-Sharkawy, E. Borejsza-Wysocka, D. Liang, M.A. Piñeros, K. Xu, and L. Cheng. 2020. Apple ALMT9 requires a conserved C-terminal domain for malate transport underlying fruit acidity. Plant Physiol. 182:992–1006.

Li, M., D. Li, F. Feng, S. Zhang, F. Ma, and L. Cheng. 2016. Proteomic analysis reveals dynamic regulation of fruit development and sugar and acid accumulation in apple. J. Exp. Bot. 67:5145–5157.

Liebhard, R., B. Koller, A. Patocchi, M. Kellerhals, W. Pfammatter, M. Jermini, and C. Gessler. 2003. Mapping quantitative field resistance against apple scab in a ‘Fiesta’ × ‘Discovery’ progeny. Phytopathology 93:493–501.

Maliepaard, C., F.H. Alston, G. van Arkel, L.M. Brown, E. Chevreau, F. Dunemann, K.M. Evans, S. Gardiner, P. Guilford, A.W. van Heusden, J. Janse, F. Laurens, J.R. Lynn, A.G. Manganaris, A.P.M. den Nijs, N. Periam, E. Rikkerink, P. Roche, C. Ryder, S. Sansavini,

H. Schmidt, S. Tartarini, J.J. Verhaegh, M. Vrielink-van Ginkel, and G.J. King. 1998. Aligning male and female linkage maps of apple (*Malus pumila* Mill.) using multi-allelic markers. Theor. Appl. Genet. 97:60–73.

Meng, R., and C.E. Finn. 2002. Determining ploidy level and nuclear DNA content in *Rubus* by flow cytometry. J. Amer. Soc. Hortic. Sci. 127:767–775.

Miyashita, T., H. Kunitake, N. Yotsukura, and Y. Hoshino. 2015. Assessment of genetic relationships among cultivated and wild *Rubus* accessions using AFLP markers. Sci. Hortic. 193:165–173.

Molina-Bravo, R., M.L. Worthington, and G.E. Fernandez. 2019. Advances and challenges in raspberry and blackberry breeding, p.1-33. In: G.A. Lang (ed.). Achieving sustainable cultivation of temperate zone tree fruits and berries. vol. 2. Burleigh Dodds, Cambridge, UK.

Moore, J.N., C. Lundergan, and E.D. Brown. 1975. Inheritance of seed size in blackberry. J. Amer. Soc. Hortic. Sci. 100:377–379.

Paterson, A., A. Kassim, S. Mccallum, M. Woodhead, K. Smith, D. Zait, and J. Graham. 2013. Environmental and seasonal influences on red raspberry flavour volatiles and identification of quantitative trait loci (QTL) and candidate genes. Theor. Appl. Genet. 126:33–48.

Perkins-Veazie, P., J.K. Collins, and J.R. Clark. 1999. Shelf-life and quality of 'Navaho' and 'Shawnee' blackberry fruit stored under retail storage conditions. J. Food Qual. 22:535–544.

Quilot, B., B.H. Wu, J. Kervella, M. Génard, M. Foulongne, and K. Moreau. 2004. QTL analysis of quality traits in an advanced backcross between *Prunus persica* cultivars and the wild relative species *P. davidiana*. Theor. Appl. Genet. 109:884–897.

Rosyara, U.R., W.S. De Jong, D.S. Douches, and J.B. Endelman. 2016. Software for genome-wide association studies in autopolyploids and its application to potato. Plant Genome 9(2):1–10.

Ryu, J., B.-K. Ha, D. sub Kim, J.-B. Kim, S.H. Kim, J.-W. Ahn, I.Y. Jeong, H.-J. Jo, E.-Y. Kim, and S.-Y. Kang. 2014. Genetic diversity and relationship assessment based on AFLP analysis in blackberry (*Rubus fructicosus* L.) mutant lines. Plant Breed. Biotechnol. 2:386–395.

Salgado, A.A., and J.R. Clark. 2016. "Crispy" blackberry genotypes: a breeding innovation of the university of Arkansas blackberry breeding program. HortScience 51:468–471.

Sebesta, B., J.R. Clark, R.T. Threlfall, and L.R. Howard. 2013. Characterization of seediness attributes of blackberry genotypes. Discov. Student J. Dale Bump. College Agr. Food and Life Sci. 14:72–79.

Shaw, D.V. 2019. Response to selection and associated changes in genetic variance for soluble solids and titratable acids contents in strawberries. J. Amer. Soc. Hortic. Sci. 115:839–843.

Shulaev, V., D.J. Sargent, R.N. Crowhurst, T.C. Mockler, O. Folkerts, A.L. Delcher, P. Jaiswal, K. Mockaitis, A. Liston, S.P. Mane, P. Burns, T.M. Davis, J.P. Slovin, N.V. Bassil, R.P. Hellens, C. Evans, T. Harkins, C. Kodira, B. Desany, O.R. Crasta, R.V. Jensen, A.C. Allan, T.P. Michael, J.C. Setubal, J.M. Celton, D.J.G. Rees, K.P. Williams, S.H. Holt, J.J.R. Rojas, M. Chatterjee, B. Liu, H. Silva, L. Meisel, A. Adato, S.A. Filichkin, M. Troggio, R. Viola, T.L. Ashman, H. Wang, P. Dharmawardhana, J. Elser, R. Raja, H.D. Priest, D.W. Bryant, S.E. Fox, S.A. Givan, L.J. Wilhelm, S. Naithani, A. Christoffels, D.Y. Salama, J. Carter, E.L. Girona, A. Zdepski, W. Wang, R.A. Kerstetter, W. Schwab, S.S. Korban, J. Davik, A. Monfort, B. Denoyes-Rothan, P. Arús, R. Mittler, B. Flinn, A. Aharoni, J.L. Bennetzen, S.L. Salzberg, A.W. Dickerman, R. Velasco, M. Borodovsky, R.E. Veilleux, and K.M. Folta. 2011. The genome of woodland strawberry (*Fragaria vesca*). Nat. Genet. 43:109–116.

Stafne, E.T., J.R. Clark, and A.L. Szalanski. 2003. Nuclear ribosomal ITS region sequences for differentiation of *Rubus* genotypes. J. Ark. Acad. Sci. 57:176–180.

Strik, B.C. 2007. Worldwide berry crop production, p. 5-15. In: Y. Zhao (ed.). Berry fruit value-added products for health promotion, 1st ed. CRC Press, Boca Raton, FL.

Strik, B.C., J.R. Clark, C.E. Finn, and M.P. Bañados. 2007. Worldwide blackberry production. HortTechnology 17:205–213.

Threlfall, R.T., O.S. Hines, J.R. Clark, L.R. Howard, C.R. Brownmiller, D.M. Segantini, and L.J.R. Lawless. 2016. Physiochemical and sensory attributes of fresh blackberries grown in the southeastern United States. HortScience 51:1351–1362.

Tomlik-Wyremblewska, A., J. Zieliński, and M. Guzicka. 2010. Morphology and anatomy of blackberry pyrenes (*Rubus* L., Rosaceae) elementary studies of the European representatives of the genus *Rubus* L. Flora Morphol. Distrib. Funct. Ecol. Plants 205:370–375.

VanBuren, R., C.M. Wai, M. Colle, J. Wang, S. Sullivan, J.M. Bushakra, I. Liachko, K.J. Vining, M. Dossett, C.E. Finn, R. Jibran, D. Chagné, K. Childs, P.P. Edger, T.C. Mockler, and N.V. Bassil. 2018. A near complete, chromosome-scale assembly of the black raspberry (*Rubus occidentalis*) genome. Gigascience 7:1–9.

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

Wada, S., and B.M. Reed. 2008. Morphological analysis of *Rubus* seed. Acta Hort. 782:67–74.

Wada, S., J.A. Kennedy, and B.M. Reed. 2011. Seed-coat anatomy and proanthocyanidins contribute to the dormancy of *Rubus* seed. Sci. Hortic. 130:762–768.

Wang, L., X. Jiang, L. Zhao, F. Wang, Y. Liu, H. Zhou, H. He, and Y. Han. 2021. A candidate PpRPH gene of the *D* locus controlling fruit acidity in peach. Plant Mol. Biol. 105:321–332.

Wang, Y., C.E. Finn, and M.C. Qian. 2005. Impact of growing environment on Chickasaw blackberry (*Rubus* L.) aroma evaluated by gas chromatography olfactometry dilution analysis. J. Agric. Food Chem. 53:3563–3571.

Willman, M., J.M. Bushakra, N.V. Bassil, C.E. Finn, M. Dossett, P. Perkins-Veazie, C.M. Bradish, G.E. Fernandez, C.A. Weber, J. Scheerens, L. Dunlap, and J. Fresnedo-Ramírez. 2020. Genetic analysis of drupelet count in black raspberry (*Rubus occidentalis*). Acta Hort. 1277:65–72.

Worthington, M.L., R. Aryal, N.V. Bassil, D. Mead, G.E. Fernandez, J.R. Clark, F. Fernández-Fernández, C.E. Finn, K.E. Hummer, and H. Ashrafi. 2020. Development of new genomic resources and tools for molecular breeding in blackberry. Acta Hort. 1277:39–46.

Xu, K., A. Wang, and S. Brown. 2012. Genetic characterization of the *Ma* locus with pH and titratable acidity in apple. Mol. Breed. 30:899–912.

Zeballos, J.L., W. Abidi, R. Giménez, A.J. Monforte, M.Á. Moreno, and Y. Gogorcena. 2016. Mapping QTLs associated with fruit quality traits in peach [*Prunus persica* (L.) Batsch] using SNP maps. Tree Genet. Genomes 12:37.

Zhen, Q., T. Fang, Q. Peng, L. Liao, L. Zhao, A. Owiti, and Y. Han. 2018. Developing gene-tagged molecular markers for evaluation of genetic association of apple *SWEET* genes with fruit sugar accumulation. Hortic. Res. 5:14.

Zurn, J.D., M. Driskill, S. Jung, D. Main, M.H. Yin, M.C. Clark, L. Cheng, H. Ashrafi, R. Aryal, J.R. Clark, M.L. Worthington, C.E. Finn, C. Peace, A. Iezzoni, and N.V. Bassil. 2020. A Rosaceae family-level approach to identify loci influencing soluble solids content in blackberry for DNA-Informed breeding. Genes Genomes Genet. 10:3729-3740.

CHAPTER I

GENETIC CONTROL OF SWEETNESS AND ACIDITY IN BLACKBERRY

Abstract

The global blackberry (*Rubus* L. subgenus *Rubus* Watson) industry has experienced rapid growth during the past 15 years. Even so, many industry stakeholders report complaints from consumers and grocers stating that blackberries are often too tart or not sweet enough for their palate. Studies have shown most consumers prefer sweet blackberries with relatively low acidity. Marker-assisted selection (MAS) for high sweetness and low acidity would allow breeding programs to expeditiously make selection and crossing decisions in the early stages of the cultivar development pipeline. The objective of this study was to identify marker-trait associations and quantitative trait loci (QTL) related to sweetness and acidity in autotetraploid blackberry using a Genome-Wide Association Study (GWAS). A panel consisting of 307 commercially-available cultivars and University of Arkansas System Division of Agriculture (UA) breeding selections was phenotyped at the Fruit Research Station in Clarksville, Arkansas in 2019, 2020, and 2021. Soluble solids content (SSC), pH, and titratable acidity (TA) were measured using juice from ten ripe berries of each genotype (selection or cultivar) collected across two separate harvest dates in each season. The *Rubus argutus* reference genome was used to design 35,054 Capture-Seq (RAPiD Genomics, Gainesville, FL) probes distributed across the genome, and annotated genes associated with sugar and acidity in other species were especially targeted for probe development. The average SSC, pH, and TA for the blackberry genotypes evaluated in this study over both harvest years were 10.9%, 3.66, and 0.81%, respectively. The phenotypic ranges for each flavor attribute were 8.1% to 13.9% for SSC, 3.39 to 4.05 for pH, and 0.52% to 1.25% for TA. Soluble solids content had the lowest broad sense entry mean

heritability at 61%, while pH and TA had heritabilities of 67% and 69%, respectively. The genotype and phenotype data were analyzed using GWASpoly software to identify marker-trait associations. Significant marker-trait associations for all three traits were found on three of the seven chromosomes, Ra01, Ra02, and Ra05, with the strongest peak located on chromosome Ra05 for TA and pH. A total of ten possible candidate genes were identified in this study, with seven candidates on Ra05 (three malate synthesis genes, two vacuolar malate transporters, a MYB transcription factor, and a phosphoenolpyruvate carboxylase PEPC) possibly associated with fruit acidity.

Introduction

As our world becomes more health-conscious, consumers are changing their diet to meet a higher nutritional standard. Due to this shift in health-conscious eating, certain fresh fruits are being consumed at an unprecedented rate (Molina-Bravo et al., 2019). One fruit with increasing fresh-market sales in recent years is blackberry. The popularity of blackberries in the United States is steadily increasing, with Americans spending just over $671 million on blackberries from Aug. 2020 to Aug. 2021, an increase of 7.3% from the year prior (California Strawberry Commission, 2021). Even with the rise in blackberry consumption and sales, industry stakeholders report complaints from both consumers and grocers claiming blackberries are often too tart or not sweet enough for their liking (Worthington et al., 2020a). The success of fruit cultivars depends upon their consumer appeal, which emphasizes flavor as a very important breeding target (Migicovsky, 2020). Studies have shown that most consumers preferred blackberries with a balance between sweet and sour tastes and a medium SSC to TA ratio of 8.93 over fruit with low or high SSC to TA ratios of 6.25 or 10.92 (Dunteman et al., 2018). Therefore,

selecting for consistently sweet flavor and balanced acidity is a major objective of fresh-market blackberry breeding programs.

The blackberry is classified within the *Rubus* genus of the Rosaceae family and is highly heterozygous with ploidy levels ranging from 2n=2x=14 to 2n=14x=98 (Clark et al., 2007). Historically, fresh-market blackberry breeding goals included thornlessness, erect growth habit, high yield, and fruit quality (Clark et al., 2007). Combining thornlessness and erect growth habit with good flavor was challenging, as the original sources of thornlessness and erect growth habits used in U.S. fresh-market blackberry breeding programs were very tart (Clark, 2005). Despite these challenges, new cultivars with reduced acidity and improved flavor have been developed through years of persistent crossing and selection. Until recently, very few genomic and genetic resources have been available for the improvement of *Rubus* species (Foster et al., 2019; Worthington et al., 2020b). Most tools for genomic studies have been developed and optimized for diploid species. However, most fresh-market blackberry cultivars throughout the world are autotetraploids (Clark et al., 2007). Autopolyploids have three or more copies of each chromosome and therefore require specialized tools to properly call allele dosage and carry out genetic studies (Bourke et al., 2018). Furthermore, very little research has been conducted on the genetic control of sweetness and acidity in *Rubus*. Fortunately, there is a high degree of genome conservation between taxa across the Rosaceae family, and blackberry researchers can build upon genetic research conducted in other diploid fruit crops to accelerate the development of molecular breeding tools (Zurn et al., 2020).

Blackberry flavor is determined by the chemistry of sugars, acids, and volatile organic compounds (VOCs) found in the fruit. Blackberries contain three principal sugars, glucose, fructose, and sucrose, with the most prominent sugars being glucose and fructose. The levels of

these sugars present within a given berry change throughout the ripening process (Fan-Chiang and Wrolstad, 2010; Kafkas et al., 2006). The primary organic acids found in blackberries are citric acid (Finn and Clark, 2017) and malic acid (Kafkas et al., 2006). Malic acid (*Ma*) content in apples is controlled by the *Ma* gene in a 65 kb region of the distal end of linkage group L16 that codes for an aluminum-activated malate transporter 9 (ALMT9) (Bai et al., 2012). The ALMT9 protein facilitates diffusion of malate and citrate through vacuolar cell membranes to regulate pH within the vacuole (Etienne et al., 2013). In both *Malus* and *Vitis*, ALMT9 plays a role in the control of fruit acidity (De Angeli et al., 2013; Etienne et al., 2013; Li et al., 2020). A major QTL for acidity in peaches named the *D* locus was mapped to a 0.4 cM genetic interval corresponding to a 100 kb region at the proximal end of LG 5 (Boudehri et al., 2009). An encoding auxin efflux carrier family protein, *Prupe*-5G004300/*ppa006339m*, in the *D* locus was found to be more highly expressed in mature fruit from acidic genotypes than non-acidic genotypes, making it a candidate gene for the control of fruit acidity in peach (Cao et al., 2016).

While genetic loci associated with major differences in fruit acidity have been mapped in apple and peach, the QTL associated with sweetness in Rosaceae crops tend to have a smaller effect on phenotypic variation and are often inconsistent across environments (Shaw, 2019). Studies of fruit acidity and sweetness show that acidity is much more heritable than sweetness, which is generally known to have low heritability estimates with QTLs explaining 35% or less of phenotypic variation (Lerceteau-Köhler et al., 2012; Shaw, 2019; Zurn et al., 2020). Zhen et al. (2018) identified 25 sugar transporters called *SWEETs* in the apple genome that likely play a role in determining SSC and sweetness. Two of these *SWEET* sugar transporters, *MdSWEET15a* and *MdSWEET9b*, were located in QTL regions associated with variation in SSC and individual sugar content. In peaches, glucose and fructose QTLs on LG 4 and a sucrose QTL on LG 5 were

consistently detected across harvest years in a biparental population (Zeballos et al., 2016). A recent study in blackberry identified loci influencing SSC on six of the seven chromosomes, with a QTL identified on *Rubus occidentalis* L. chromosome Ro01 accounting for a 1.46% difference in SSC across several small biparental populations (Zurn et al., 2020). A gene in the QTL region on Ro01, *qSSC-Ruh-Ch1.1*, had homology with glycosyl-transferase and sucrose synthase genes that have previously been implicated in the accumulation of sugars and starches in other plants (Zurn et al., 2020).

Genome-wide association studies are a powerful tool for breeders to identify marker-trait associations in diverse populations, and these studies have become commonplace in diploid species. The purpose of conducting a GWAS is to understand the genetic control of traits by relating genotypic data consisting of a large number of markers, typically single nucleotide polymorphisms (SNPs), to measured phenotypic data while controlling for population structure (Ball, 2013). A specialized R package "GWASpoly" has been developed specifically for GWAS in autopolyploid species (Rosyara et al., 2016). Determining the genetic control of sweetness and acidity in blackberry through genetic studies such as a GWAS will allow breeders to develop genetic markers that can be used during seedling selection and to optimize crossing plans.

Objectives

The objectives of this study were to:

1. Evaluate flavor attributes of titratable acidity, pH, and soluble solids content in a large panel of commercially available fresh-market blackberry cultivars and UA breeding selections.
2. Conduct a GWAS to identify marker-trait associations, QTL, and possible candidate genes associated with sweetness and acidity in blackberry.

Materials and Methods

Plant Material and Sample Collection

A panel of 307 genotypes consisting of 29 commercially available fresh-market blackberry cultivars and 278 UA breeding selections were chosen for GWAS analysis (Supplementary Table 1 of Appendix A). Two hundred and thirty-six, 237, and 205 genotypes from this panel were harvested and phenotyped in the summers of 2019, 2020, and 2021, respectively. The 307 genotypes in the panel were grown and maintained in 6 m plots at the UA Fruit Research Station (FRS) located in Clarksville, Arkansas. The FRS is located in the U.S. Department of Agriculture (USDA) hardiness zone 7b (USDA, 2021) with Linker fine sandy loam type soil (Salgado and Clark, 2016). The research plots received regular maintenance of pruning, tipping, integrated pest management, and irrigation.

Blackberry samples were collected from floricane fruit during the summers of 2019, 2020, and 2021 with two harvest dates obtained per genotype each year. The harvest season began in mid-June and continued through the first week of July each year. One pint-sized clamshell (240 g) was collected for each sample. Ripe, marketable berries were harvested at the shiny-black stage and had no discolored or damaged drupelets. On rare occasions, poor fruit quality caused by insect pressure prohibited a second harvest for some genotypes. Harvest was delayed at least 24 h after rain events with 5 mm or more of rain, as precipitation can influence postharvest quality.

Each clamshell was stored at 5 ºC with 90% relative humidity for seven days. The 2020 and 2021 samples were subjected to a 30 min shaking cycle at 10 Hz frequency with approximately 2 mm of displacement at room temperature before refrigeration. This step was implemented to simulate transportation for a study focused on red drupelet reversion (RDR) and

was used with the intention to promote RDR phenotype expression. After refrigeration, 10-berry samples for each genotype were bagged and frozen in a -12 ºC walk-in freezer in preparation for juicing. Frozen 10-berry samples were thawed and squeezed by hand in cheesecloth over a 50-mL centrifuge tube to extract the juice and then frozen again for flavor attribute analysis.

Evaluation of Soluble Solids Content, pH, and Titratable Acidity

Samples collected in 2019 were evaluated for SSC (expressed as %) using a tabletop Reichert ABBE Mark II digital refractometer (Reichert Inc., Buffalo, NY). In 2020 and 2021 soluble solids content was analyzed using a handheld PAL-BX/Acid F5 meter for multiple fruits (ATAGO, Tokyo, Japan). Titratable acidity was measured with a Metrohm 862 Compact Titrosampler (Metrohm AG, Herisau, Switzerland) standardized to buffer solutions with a pH of 2.0, 4.0, 7.0, and 10.0 prior to analysis. In 2019, TA was determined by diluting 6 g of juice in 50 mL deionized, degassed water by titration with 0.1 N of sodium hydroxide (NaOH) to a pH endpoint of 8.2. A similar protocol was followed in 2020, using 1 g instead of 6 g of juice diluted in 50 mL deionized, degassed water. TA was not measured for samples collected in 2021. A citric acid standard was used to calculate TA with a milliequivalent factor of 0.064 following the equation:

$$TA\ (\%) = ((mL\ NaOH \times milliequivalent\ factor \times 100) / grams\ of\ sample)$$

In 2019, pH was measured with a pH meter fitted to the Metrohm 862 Compact Titrosampler, while pH measurement was conducted with a PH700 Benchtop pH Meter (APERA instruments, Columbus, OH) in 2020 and 2021.

Analysis of Heritability and Trait Correlations

Best linear unbiased predictions (BLUPs) of SSC, pH, and TA across all years and replicates were calculated for each genotype using the MIXED procedure in SAS (SAS Institute,

Inc., Cary, NC). Broad-sense heritability (H) for each trait was calculated using harmonic means for year (*y*) and replicate (*r*) as well as genotypic variance (σ^2_g), genotype by year variance (σ^2_{gy}), and residual effects (σ^2). Broad-sense heritability was calculated as: $H = \sigma^2_g / [\sigma^2_g + (\sigma^2_{gy}/y) + (\sigma^2/yr)]$. Harmonic means for years and replicates were calculated using the "dplyr" package in R, while variance and residual effects were obtained through the Type3 method of the MIXED procedure in SAS. Pearson correlation coefficients were calculated to measure the strength and direction of linear relationships between flavor attributes using the "GGally" and "ggplot2" packages in R.

CaptureSeq Genotyping

A modified cetyltrimethylammonium bromide (CTAB) protocol was used to extract DNA collected from young leaves of each genotype (Porebski et al., 1997). Quantification was completed using the dsDNA assay kit by Qubit (Invitrogen, Carlsbad, CA, USA) and samples were standardized to 40 ng/µl. Sequence capture genotyping was performed at RAPiD Genomics (Gainesville, FL). The *Rubus argutus* 'Hillquist' V1 reference genome (Worthington et al., 2020b) was used to custom-design 35,054 biotinylated 120-mer Capture-Seq probes consisting of 17,100 probes in candidate gene regions, including genes associated with sugar and acidity in other species, and 17,954 probes scattered across the genome to cover linkage disequilibrium (LD). Paired end sequencing was performed on Illumina HiSeq2000. Raw data from sequencing was cleaned and trimmed then aligned to the 'Hillquist' genome using MOSAIK (Lee et al., 2014). Variant calling was performed using Freebayes (Garrison and Marth, 2012). SNPs were filtered in VCFtools (Danecek et al., 2011) to produce a file with biallelic markers with a minor allele frequency of ≥ 0.01 and average read depth between three and 750 per sample. Filtered

data was then converted to tetraploid SNP calls using the multidog function in UpDog (Gerard et al., 2018).

GWAS Analysis

Association analysis was performed with the GWASpoly (Rosyara et al., 2016) package in R, using SSC, pH, and TA BLUPs calculated for each genotype along with the genotypic data generated from Capture-Seq. Data were analyzed using a random effect model with a kinship matrix constructed with the leave-one-chromosome-out (LOCO) method. Additive, simplex dominant alternative (simplex-dom-alt), and simplex dominant reference (simplex-dom-ref) gene action models were tested. The SNP effect in the additive model is proportional to the dosage of the minor allele, whereas in the two simplex dominant models all three heterozygotes are equivalent to one of the homozygotes (Rosyara et al., 2016). Significance thresholds were calculated using the Bonferroni method with $\alpha = 0.05$.

Identification of Possible Candidate Genes

Quantitative trait loci regions of each chromosome containing significant markers, including a 1 Mb region flanking either side of the significant makers, were investigated for possible candidate genes with functions related to sweetness or acidity. The 'Hillquist' V1 genome assembly was functionally annotated with the Swiss-Prot database, now UniProt (The UniProt Consortium, 2021), and Araport11 (Cheng et al., 2017) blastsets using the predicted protein-coding sequences of 38,503 genes as queries. Annotated genes coding for factors and proteins known to play a role in the acidity of related species within regions significantly associated with SSC, pH, or TA were considered as potential candidate genes.

Results

Phenotypic Results

Data from 1268 replicates collected during 2019, 2020, and 2021 were used to calculate BLUPs for 307 genotypes evaluated for SSC and pH. Best linear unbiased predictions for TA were calculated using 865 replicates from 272 genotypes evaluated in 2019 and 2020 (Supplementary Table 1 of Appendix A). Total rainfall and average weekly high and low temperatures were tracked starting one week before blackberry harvest began and continuing through the end of the harvest season (Fig. 1). The average high and low temperatures for each week were determined by taking the highest and lowest temperatures recorded each day for seven days and calculating the means. The highest weekly average temperatures were 31.6 ºC, 31.8 ºC, and 32.5 ºC in 2019, 2020, and 2021, respectively, while the lowest weekly averages were 16.4 ºC, 18.3 ºC, and 18.3 ºC. All three summers had similar highest and lowest weekly average temperatures, with 2019 being slightly cooler than the following two years. Rainfall during harvest was 486 mm in 2019, 218 mm in 2020, and 346 in 2021.

Overall genotypic BLUPs for SSC and pH were calculated using data obtained in 2019, 2020, and 2021, while BLUPs for TA were calculated with data from the summers of 2019 and 2020. The average normalized BLUPs were 10.9% for SSC, 3.66 for pH, and 0.81% for TA (Table 1). The BLUPs for SSC ranged from 8.1% to 13.9%, with A-2752T having the lowest SSC and A-2806T the highest. The range of BLUPs for pH was from 3.39 to 4.05, with the least and most acidic genotypes being APF-447T and A-2620T, respectively. The genotype with the lowest BLUP for TA was A-2525TN at 0.52%, while the genotype with the highest TA was A-2416T at 1.25%. Titratable acidity and pH were highly correlated (r = -0.75), while SSC was

more weakly correlated with pH (r = 0.21) and TA (r = -0.22) (Fig. 2). The broad-sense heritabilities of juice attributes were 61% for SSC, 67% for pH, and 69% for TA.

Association Analysis

A total of 124,564 polymorphic markers were discovered from Capture-Seq genotyping and passed initial data filtering steps in VCFtools and UpDog. Three hundred and two of the 307 blackberry cultivars and selections phenotyped in this research were genotyped with these markers and passed quality filtering steps. The markers and the phenotype data for in the blackberry GWAS panel were analyzed in GWASpoly to create QQ-plots, Manhattan plots, and to identify significant SNPs to determine marker-trait associations. Inspection of QQ-plots did not show any evidence of systemic bias for the traits and models evaluated in this study (Supplemental Fig. 1). Likelihood of odds (LOD) thresholds of 6.39 for the additive model and 6.23 for the simplex dominant models were determined using the Bonferroni method with $\alpha = 0.05$ (Fig. 3).

An important overlapping QTL for TA and pH was found on Ra05 between 3384682 and 6149829 bp (Fig. 3). This approximately 2.8 Mb region contained 155 significant markers produced with both the additive and simplex dominant models. Of the 155 markers, 150 were associated with TA and 149 were associated with pH, with both traits sharing 144 significant markers. Significant markers for TA in the QTL on Ra05 consisted of 150 simplex dominant and 134 additive markers, all significant additive markers were also significant in the simplex dominant model. Significant markers for pH consisted of 148 simplex dominant and 130 additive significant markers with an overlap of 129 markers significant for both models. Two peak markers located at 4448123 and 4448155 bp on Ra05 had LOD scores of 8.85 for pH and 7.71 for TA in the simplex dominant model (Fig. 3). At both of these peak markers, 253 of the 302

genotypes in the GWAS panel were nulliplex, 47 had simplex allele dosage, and two had duplex allele dosages. None of the genotypes in the GWAS panel had triplex or quadruplex dosage for the dominant allele at either peak position. Therefore, it is uncertain if the underlying acidity QTL in this region on Ra05 has additive or dominant gene action. A QTL consisting of one significant marker for TA was located on chromosome Ra01 at 723480 bp with a LOD of 6.97 as determined with the additive model. Two significant markers produced a QTL for SSC on chromosome Ra02 located at the 29208288 and 29284319 bp positions with LOD scores of 6.5 and 6.62 obtained through the additive model.

Possible Candidate Genes

One possible candidate gene for sweetness and nine possible candidate genes for acidity in fresh-market blackberries were identified in this study (Table 2). A gene coding for a sucrose binding protein (Ra_g7910) was located 480 kb distally of the SSC QTL at 29208288 to 29284319 bp on Ra02. Two genes with functional annotations coding for an H^+-ATPase 8 (Ra_g141) and a vacuolar proton-translocating pyrophosphatase protein (H^+-PPase) (Ra_g225) were located within the candidate gene region of the QTL containing a single significant marker for TA on Ra01 located at 723480 bp. The Ra_g141 gene was located 17 kb proximally of the significant marker while Ra_g225 was located 331 kb distally of the marker. Seven possible candidate genes with annotations related to fruit acidity in other species were located within the region of interest around the QTL for pH and TA between 3384682 and 6149829 bp on chromosome Ra05. Two genes functionally annotated for ALMT9 proteins (Ra_g20630 and Ra_g21398) were located 341 kb proximally and 949 kb distally of the QTL on Ra05. The gene Ra_g20727, functionally annotated for a MYB1 transcription factor, was located within the QTL on Ra05 at 3560200 bp. A phosphoenolpyruvate carboxylase (PEPC) gene (Ra_g20750) was

located at 3685932 bp, and three genes coding for malate synthase (Ra_g21026, Ra_g21027, and Ra_g21028) were located at 5081414 bp, 5082698 bp, and 5083558 bp within the TA and pH QTL on Ra05.

Discussion

The ranges and means for SSC, TA, and pH obtained in this study were similar to previously published values for fresh-market blackberries. The genotypic BLUPs for SSC reported here ranged from 8.1% to 13.9%, within the range of previous research showing UA germplasm had SSC values from 4.6% – 16.2% (Zurn et al., 2020). The mean SSC value reported in this study was 10.9%, above the U.S. minimum standard of 10.0% for blackberries used in the production of beverages containing fruit juice (FDA, 2020) and comparable to previously reported SSC means for UA fresh-market blackberry populations at 9.9% (Zurn et al., 2020) and mixed fresh-market and processing cultivars at 10.8% (Fan-Chiang and Wrolstad, 2010).

The genotypic BLUPs for pH observed in our fresh-market blackberries ranged from 3.39 to 4.05 with an average of 3.66. A study conducted on processing and fresh-market blackberry cultivars reported a pH range from 2.65 to 3.61 with a mean of 3.19, lower than the pH values observed in our germplasm (Fan-Chiang and Wrolstad, 2010). Values obtained for pH of 11 genotypes in the UA germplasm were reported as ranging from 3.0 to 3.6 (Threlfall et al., 2016), more closely aligned with the values we observed. Analysis of TA in our study resulted in BLUPs ranging from 0.52% to 1.25% with an average of 0.81%. Fan-Chiang and Wrolstad (2010) measured TA of varying blackberry germplasm that showed a range from 0.52% to 2.24% reporting an average of 1.35%, while a study of only UA germplasm reported TA ranging from 0.7% to 1.4% (Threlfall et al., 2016). Temperature and precipitation must be taken into

consideration when evaluating and comparing flavor attributes between harvest years, as environmental factors can affect the acidity and sweetness of blackberry juice (Etienne et al., 2013). Precipitation levels were variable among the three harvest seasons in this study, possibly explaining phenotypic differences between harvest years within our study and similar studies. The beginning of the harvest season in 2019 was rainier and cooler than in 2020 and 2021, which could explain why 2019 samples were slightly more acidic and less sweet than the following two years.

Our study presents the first reported broad-sense heritability estimates for SSC, pH, and TA in blackberry with values of 61%, 67%, and 69%, respectively. Heritabilities for these traits have been reported in related Rosaceae crops. In peaches, broad-sense heritability calculations for SSC and TA were 76% and 93% (Rawandoozi et al., 2020), far above our estimates of 61% and 69%, respectively. Broad-sense heritability estimates in strawberry were closer to blackberry, with pH and TA estimated at 53% and 65%, respectively (Lerceteau-Köhler et al., 2012).

Until now, a GWAS for fruit flavor traits has never been performed in any *Rubus* species. The majority of significant markers were associated with TA and pH and were located on chromosome Ra05 between 3384682 and 6149829 bp. This QTL on Ra05 had a maximum LOD value of 8.85 and contained 155 of the 158 significant markers identified in this study. Of the 302 genotypes in the panel, 253 were nulliplex, having four copies of the allele associated with high acidity at the peak LOD positions of 4448123 and 4448155 bp. Forty-seven genotypes were simplex, with one copy of the low acid allele, and only two genotypes were duplex, with two copies each of the low and high acid alleles. Thus, we were unable to determine whether the gene action of the underlying allele was additive or dominant. It is unclear why there were no

genotypes with three or more copies of the low acidity allele in the panel, especially considering that reduced acidity is a highly desired trait in fresh-market breeding programs. One possibility that should be further investigated is that there could be a lethal or otherwise deleterious effect of the low acid allele at higher dosages.

Seven possible candidate genes for the genetic control of blackberry acidity were identified within a 1 Mb region flanking the acidity QTL on Ra05, three for malate synthase, two for ALMT9, and one each for MYB1 and PEPC. The proximal ALMT9 gene, Ra_g20630, was found 341 kb upstream from the QTL, with a second ALMT9 gene, Ra_g21398, located 949 kb distal of the QTL. These ALMT9 genes were considered possible candidate genes for pH and TA in fresh-market blackberries as the candidate gene, *Ma1*, in the apple acidity locus codes for ALMT9 (Etienne et al., 2013; Liu and Zhou, 2018; Li et al., 2020). Plant ALMTs transport malate across vacuolar cell membranes through facilitated diffusion of an electrical potential gradient ($\Delta\Psi$) regulating vacuolar storage and concentration of malate and citrate in cells, with citrate transport occurring far slower than malate (Etienne et al., 2013). Malate transport relies on the presence of a $\Delta\Psi$ to occur, with the $\Delta\Psi$ expected to decrease as vacuolar pH decreases, effectively closing the transport channels at a low vacuolar pH to prevent over-acidification of already highly acidic vacuoles (Etienne et al., 2013). In apple, the recessive low acid allele, *ma1,* results in a truncated protein with a lost C-terminal domain essential for malate transport activity. However, not all ALMTs transport malate exclusively and it was recently discovered that the ALMT10 protein primarily transports chloride and nitrate rather than malate (Moreno Racero, 2020). For this reason, an ALMT10 coding gene within the QTL on Ra05 was excluded as a possible candidate gene for fruit acidity.

Another possible candidate gene, Ra_g20727, in the acidity QTL region on Ra05 codes for the MYB1 transcription factor. MYB1 regulates the expression of genes that encode vacuolar proton pump subunits, an anthocyanin transporter, and a malate transporter in apples, modulating anthocyanin and malate accumulation in the vacuole (Hu et al., 2016, 2017). A gene, Ra_g20750, encoding PEPC is also located within the QTL region. The PEPC gene was considered a possible candidate for acidity because it is a key enzyme in the pathway responsible for the initial formation of organic acids. Specifically, PEPC catalyzes the reaction causing carboxylation of phosphoenolpyruvate in the cytosol, producing oxaloacetate, which can then be reduced to malate by the cytosolic NAD-dependent malate dehydrogenase (Etienne et al., 2013). Three contiguous genes (Ra_g21026, Ra_g21027, and Ra_g21028) in the QTL region on Ra05 code for malate synthase. Malate synthase is one of five key enzymes involved in the glyoxylate cycle, which facilitates conversion between tri- and dicarboxylates in the glyoxysome (Etienne et al., 2013) and are therefore considered possible candidates for the control of blackberry acidity. On chromosome Ra01, a single significant marker for TA was located at 723480 bp with a LOD of 6.97. Genes encoding an H^+ -ATPase protein (Ra_g141) and an H^+-PPase protein (Ra_g255) were located 17 kb proximal and 331 kb distal to the significant marker. These two proteins are involved in the establishment of an electrochemical proton gradient across plant vacuolar membranes and acidifying the vacuolar lumen (Lin et al., 2012; Nakanishi and Maeshima, 1998), which makes Ra_g141 and Ra_g255 possible candidates for blackberry acidity.

In this study, a single QTL associated with SSC was identified on chromosome Ra02 with two significant markers with a peak LOD of 6.62. Fruit acidity is generally more heritable than sweetness, and QTL for SSC and individual sugars tend to have smaller effects and are often inconsistent across environments (Lerceteau-Köhler et al., 2012; Shaw, 2019; Zurn et al.,

2020). Quantitative trait loci associated with variation in SSC have been previously mapped in the *Rubus* (Zurn et al., 2020), *Prunus* (Rawandoozi et al., 2020), *Malus* (Guan et al., 2015), and *Fragaria* (Lerceteau-Köhler et al., 2012) genera of the Rosaceae family. An earlier study conducted with blackberry populations from the UA and USDA HCRU breeding programs found 48 significant alleles for SSC mapped to six of the seven base chromosomes of the *Rubus occidentalis* L. genome (Zurn et al., 2020). However, only six of these 48 alleles were significantly associated with sugar content in the UA populations in either year, five on Ro01 and one on Ro04 (Zurn et al., 2020). A QTL, *qSSC-Ruh-ch1.1*, on chromosome Ro01 in the region of a sucrose synthase gene accounted for a 1.46% difference in SSC in the combined UA and USDA-HCRU populations over both years of the study. Our study was not able to validate or duplicate these results as the only significant QTL identified for SSC consisted of two significant markers on chromosome Ra02. A single possible candidate gene (Ra_g7190) coding for a sucrose-binding protein was identified 479 kb away from the closest significant marker. The sucrose binding protein is strongly associated with cells that actively transport sucrose to young sink leaves and the cotyledon sink cells of soybean [*Glycine max* (L.) Merr.], however, it is involved in a number of different cell types and possibly functions as a monitor of sucrose concentration and uptake in sink tissues (Grimes et al., 1992; Ayre, 2011).

Further studies need to be conducted on the genetic control of sweetness and acidity in fresh-market blackberries to validate the results found in this GWAS and develop diagnostic markers for breeding. Isolating and phenotyping each of the individual sugars and acids would be a useful approach to refine these results and validate or identify additional QTL for sweetness and acidity. Candidate gene sequences should also be compared to identify any nonsynonymous mutations that might distinguish low and high acid genotypes. Data from this research can be

used to help create a training population for the implementation of genomic selection (GS) studies. Unlike association mapping, GS uses all available molecular markers to predict the performance and breeding values of candidates for selection (Crossa et al., 2017). Genomic selection is particularly powerful as a predictive approach to identify the best candidates in a population for low to moderate heritability traits such as SSC, which are controlled by many small-effect QTL (Bernardo, 2016).

Conclusion

In summary, the phenotypic and heritability estimates for SSC, pH, and TA reported in this study are consistent with previously published studies in blackberries and related *Rubus* and Rosaceae species. Ten possible candidate genes associated with fruit sweetness and acidity were discovered in three QTL regions on three of the seven *R. argutus* chromosomes, Ra01, Ra02, and Ra05. One QTL for TA was located on Ra01, one for SSC was located on Ra02, and a QTL for TA and pH was located on Ra05. Nine possible candidate genes for acidity were identified, two on Ra01 and seven on Ra05, while one possible candidate gene for sweetness was identified on Ra02. Data from this research will be used to design and validate diagnostic molecular markers and to train genomic selection models for enhanced sweetness and reduced acidity in blackberries. Further studies can also be conducted using the same methods with a different panel of genotypes or with GWAS for individual acids and sugars. A better understanding of the genetic control of sweetness and acidity in fresh-market blackberries will allow the UA breeding program to expedite the production and release of blackberry cultivars with improved flavor.

Literature Cited

Ayre, B.G. 2011. Membrane-transport systems for sucrose in relation to whole-plant carbon partitioning. Mol. Plant 4:377–394.

Bai, Y., L. Dougherty, M. Li, G. Fazio, L. Cheng, and K. Xu. 2012. A natural mutation-led truncation in one of the two aluminum-activated malate transporter-like genes at the *Ma* locus is associated with low fruit acidity in apple. Mol. Genet. Genomics 287:663–678.

Ball, R.D. 2013. Designing a GWAS: power, sample size, and data structure. Methods in Mol. Biol. vol. 1019. Clifton, N.J. 37–98.

Bernardo, R. 2016. Bandwagons I, too, have known. Theor. Appl. Genet. 129:2323–2332.

Boudehri, K., A. Bendahmane, G. Cardinet, C. Troadec, A. Moing, and E. Dirlewanger. 2009. Phenotypic and fine genetic characterization of the *D* locus controlling fruit acidity in peach. BMC Plant Biol. 9:59.

Bourke, P.M., R.E. Voorrips, R.G.F. Visser, and C. Maliepaard. 2018. Tools for genetic studies in experimental populations of polyploids. Front. Plant Sci. 9:513.

California Strawberry Commission. 2021. U.S. retail category trends. 23 August 2021. <https://www.calstrawberry.com/en-us/market-data/retail-category-trends/>

Cheng, C.Y., V. Krishnakumar, A.P. Chan, F. Thibaud-Nissen, S. Schobel, and C.D. Town. 2017. Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. Plant J. 89:789–804.

Clark, J.R. 2005. Intractable traits in eastern U.S. blackberries. HortScience 40:1954–1955.

Clark, J.R., E.T. Stafne, H.K. Hall, and C.E. Finn. 2007. Blackberry breeding and genetics. Plant Breeding Rev. 29:19–144.

Crossa, J., P. Pérez-Rodríguez, J. Cuevas, O. Montesinos-López, D. Jarquín, G. de los Campos, J. Burgueño, J.M. González-Camacho, S. Pérez-Elizalde, Y. Beyene, S. Dreisigacker, R. Singh, X. Zhang, M. Gowda, M. Roorkiwal, J. Rutkoski, and R.K. Varshney. 2017. Genomic selection in plant breeding: methods, models, and perspectives. Trends Plant Sci. 22:961–975.

Danecek, P., A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, R.E. Handsaker, G. Lunter, G.T. Marth, S.T. Sherry, G. McVean, and R. Durbin. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158.

De Angeli, A., U. Baetz, R. Francisco, J. Zhang, M.M. Chaves, and A. Regalado. 2013. The vacuolar channel VvALMT9 mediates malate and tartrate accumulation in berries of *Vitis vinifera*. Planta 238:283–291.

Dunteman, A.N., R.T. Threlfall, J.R. Clark, and M.L. Worthington. 2018. Evaluating consumer censory and composition attributes of Arkansas-grown fresh-market blackberries. Discov. Student J. Dale Bump. Coll. Agric. Food Life Sci. 19:16–23.

Etienne, A., M. Génard, P. Lobit, D. Mbeguié-A-Mbéguié, and C. Bugaud. 2013. What controls fleshy fruit acidity? A review of malate and citrate accumulation in fruit cells. J. Exp. Bot. 64:1451–1469.

Fan-Chiang, H.J., and R.E. Wrolstad. 2010. Sugar and nonvolatile acid composition of blackberries. J. AOAC Int. 93:956–965.

FDA, U. 2020. Food for human consumption. CFR - code Fed. Regul. title 21. 9 June 2021. <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm?fr=101.30/>

Finn, C.E., and J.R. Clark. 2017. Cultivar development and selection. In H.K. Hall, R.C. Funt (eds.). Blackberries and their hybrids. CABI, Boston, MA. 63–92.

Foster, T.M., N.V. Bassil, M. Dossett, M.L. Worthington, and J. Graham. 2019. Genetic and genomic resources for *Rubus* breeding: a roadmap for the future. Hortic. Res. 6:116.

Garrison, E., and G. Marth. 2012. Haplotype-based variant detection from short-read sequencing. Genomics 2:1–9.

Gerard, D., L.F.V. Ferrão, A.A.F. Garcia, and M. Stephens. 2018. Genotyping polyploids from messy sequencing data. Genet. 210:789–807.

Grimes, H.D., P.J. Overvoorde, K. Ripp, V.R. Franceschi, and W.D. Hitz. 1992. A 62-kD sucrose binding protein is expressed and localized in tissues actively engaged in sucrose transport. Plant Cell 4:1561–1574.

Guan, Y., C. Peace, D. Rudell, S. Verma, and K. Evans. 2015. QTLs detected for individual sugars and soluble solids content in apple. Mol. Breed. 35(135):1–13.

Hu, D.G., C.H. Sun, Q.J. Ma, C.X. You, L. Cheng, and Y.J. Hao. 2016. MdMYB1 regulates anthocyanin and malate accumulation by directly facilitating their transport into vacuoles in apples. Plant Physiol. 170:1315–1330.

Hu, D.G., Y.Y. Li, Q.Y. Zhang, M. Li, C.H. Sun, J.Q. Yu, and Y.J. Hao. 2017. The R2R3-MYB transcription factor MdMYB73 is involved in malate accumulation and vacuolar acidification in apple. Plant J. 91:443–454.

Kafkas, E., M. Koşar, N. Türemiş, and K.H.C. Başer. 2006. Analysis of sugars, organic acids and vitamin C contents of blackberry genotypes from Turkey. Food Chem. 97:732–736.

Lee, W.P., M.P. Stromberg, A. Ward, C. Stewart, E.P. Garrison, and G.T. Marth. 2014. MOSAIK: A hash-based algorithm for accurate next-generation sequencing short-read mapping. PLoS One. 9:3.

Lerceteau-Köhler, E., A. Moing, G. Guérin, C. Renaud, A. Petit, C. Rothan, and B. Denoyes. 2012. Genetic dissection of fruit quality traits in the octoploid cultivated strawberry highlights the role of homoeo-QTL in their control. Theor. Appl. Genet. 124:1059–1077.

Li, C., L. Dougherty, A.E. Coluccio, D. Meng, I. El-Sharkawy, E. Borejsza-Wysocka, D. Liang, M.A. Piñeros, K. Xu, and L. Cheng. 2020. Apple ALMT9 requires a conserved C-terminal domain for malate transport underlying fruit acidity. Plant Physiol. 182:992–1006.

Lin, S.M., J.Y. Tsai, C.D. Hsiao, Y.T. Huang, C.L. Chiu, M.H. Liu, J.Y. Tung, T.H. Liu, R.L. Pan, and Y.J. Sun. 2012. Crystal structure of a membrane-embedded H^+-translocating pyrophosphatase. Nature 484:399–403.

Liu, J., and M. Zhou. 2018. The *ALMT* gene family performs multiple functions in plants. Agron. 8:2.

Migicovsky, Z. 2020. Tasting improvement in fruit flavor using genomics. New Phytol. 226:1539–1540.

Molina-Bravo, R., M.L. Worthington, and G.E. Fernandez. 2019. Advances and challenges in raspberry and blackberry breeding, p.1-33. In: G.A. Lang (ed.). Achieving sustainable cultivation of temperate zone tree fruits and berries. vol. 2. Burleigh Dodds, Cambridge, UK.

Moreno Racero, F.J. 2020. Functional characterization of the anion channel AtALMT10 (in Spanish). Universidad de Sevilla. MS Thesis.

Nakanishi, Y., and M. Maeshima. 1998. Molecular cloning of vacuolar H^+-pyrophosphatase and its developmental expression in growing hypocotyl of mung bean. Plant Physiol. 116:589–597.

Porebski, S., L.G. Bailey, and B.R. Baum. 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Mol. Biol. Rptr. 15:8–15.

Rawandoozi, Z.J., T.P. Hartmann, S. Carpenedo, K. Gasic, C. Da Silva Linge, L. Cai, E. Van De Weg, and D.H. Byrne. 2020. Identification and characterization of QTLs for fruit quality traits in peach through a multi-family approach. BMC Genomics 21:522.

Rosyara, U.R., W.S. De Jong, D.S. Douches, and J.B. Endelman. 2016. Software for genome-wide association studies in autopolyploids and its application to potato. Plant Genome 9(2):1–10.

Salgado, A.A., and J.R. Clark. 2016. “Crispy” blackberry genotypes: a breeding innovation of the university of Arkansas blackberry breeding program. HortScience 51:468–471.

Shaw, D. V. 2019. Response to selection and associated changes in genetic variance for soluble solids and titratable acids contents in strawberries. J. Amer. Soc. Hortic. Sci. 115:839–843.

Threlfall, R.T., O.S. Hines, J.R. Clark, L.R. Howard, C.R. Brownmiller, D.M. Segantini, and L.J.R. Lawless. 2016. Physiochemical and sensory attributes of fresh blackberries grown in the southeastern United States. HortScience 51:1351–1362.

The UniProt Consortium. 2021. UniProt: The universal protein knowledgebase in 2021. Nucleic Acids Res. 49:D1.

U.S. Department of Agriculture. 2021. USDA plant hardiness zone map. Agricultural Research Service. 23 August 2021. <https://planthardiness.ars.usda.gov/pages/view-maps/>

Worthington, M.L., R. Aryal, N.V. Bassil, D. Mead, G.E. Fernandez, J.R. Clark, F. Fernández-Fernández, C.E. Finn, K.E. Hummer, and H. Ashrafi. 2020b. Development of new genomic resources and tools for molecular breeding in blackberry. Acta Hort. 1277:39–46.

Worthington, M.L., M. Coe, A. McWhirt, and L. Herrera. 2020a. Research and extension priorities of the U.S. blackberry industry identified through a national stakeholder survey. Fayetteville, A.R.

Zeballos, J.L., W. Abidi, R. Giménez, A.J. Monforte, M.Á. Moreno, and Y. Gogorcena. 2016. Mapping QTLs associated with fruit quality traits in peach [*Prunus persica* (L.) Batsch] using SNP maps. Tree Genet. Genomes 12:37.

Zhen, Q., T. Fang, Q. Peng, L. Liao, L. Zhao, A. Owiti, and Y. Han. 2018. Developing gene-tagged molecular markers for evaluation of genetic association of apple *SWEET* genes with fruit sugar accumulation. Hortic. Res. 5:14.

Zurn, J.D., M. Driskill, S. Jung, D. Main, M.H. Yin, M.C. Clark, L. Cheng, H. Ashrafi, R. Aryal, J.R. Clark, M.L. Worthington, C.E. Finn, C. Peace, A. Iezzoni, and N.V. Bassil. 2020. A rosaceae family-level approach to identify loci influencing soluble solids content in blackberry for DNA-Informed breeding. Genes Genomes Genet. 10:3729–3740.

Tables and Figures

Table 1. Minimum, maximum, and mean values for soluble solids content (SSC), pH, and titratable acidity (TA) as normalized best linear unbiased predictions (BLUPs) for fresh-market blackberries grown in Clarksville, AR. Overall BLUPs for SSC and pH were calculated using data collected in 2019, 2020, and 2021, and the overall TA BLUPs were calculated with data from only 2019 and 2020.

	Overall			2019			2020			2021	
Trait	SSC (%)	pH	TA (% citric)	SSC (%)	pH	TA (% citric)	SSC (%)	pH	TA (% citric)	SSC (%)	pH
Maximum	13.9	4.05	1.25	11.3	3.83	1.23	15.1	4.16	1.22	15.1	4.02
Minimum	8.1	3.39	0.52	9.8	3.37	0.56	8.1	3.46	0.45	8.2	3.39
Mean	10.9	3.66	0.81	10.6	3.57	0.85	11.2	3.74	0.76	10.8	3.74

Table 2. Possible candidate genes for titratable acidity (TA), soluble solids content (SSC), and pH in blackberry in three QTL regions with their location and annotated functions.

Trait	Gene ID	Chromosome	Start bp	Stop bp	Functional annotation
TA	Ra_g141	Ra01	701664	706107	H^+-ATPase 8
TA	Ra_g225	Ra01	1054638	1060512	Vacuolar proton-translocating pyrophosphatase
SSC	Ra_g7910	Ra02	29763951	29765632	Sucrose-binding protein
TA and pH	Ra_g20630	Ra05	3043192	3046890	Vacuolar malate transporter 9 - ALMT9
TA and pH	Ra_g20727	Ra05	3560200	3560987	Transcription factor protein - MYB1
TA and pH	Ra_g20750	Ra05	3685932	3686993	Phosphoenolpyruvate carboxylase - PEPC
TA and pH	Ra_g21026	Ra05	5081414	5082275	Malate synthase
TA and pH	Ra_g21027	Ra05	5082698	5083511	Malate synthase
TA and pH	Ra_g21028	Ra05	5083558	5084695	Malate synthase
TA and pH	Ra_g21398	Ra05	7098716	7102494	Vacuolar malate transporter 9 - ALMT9

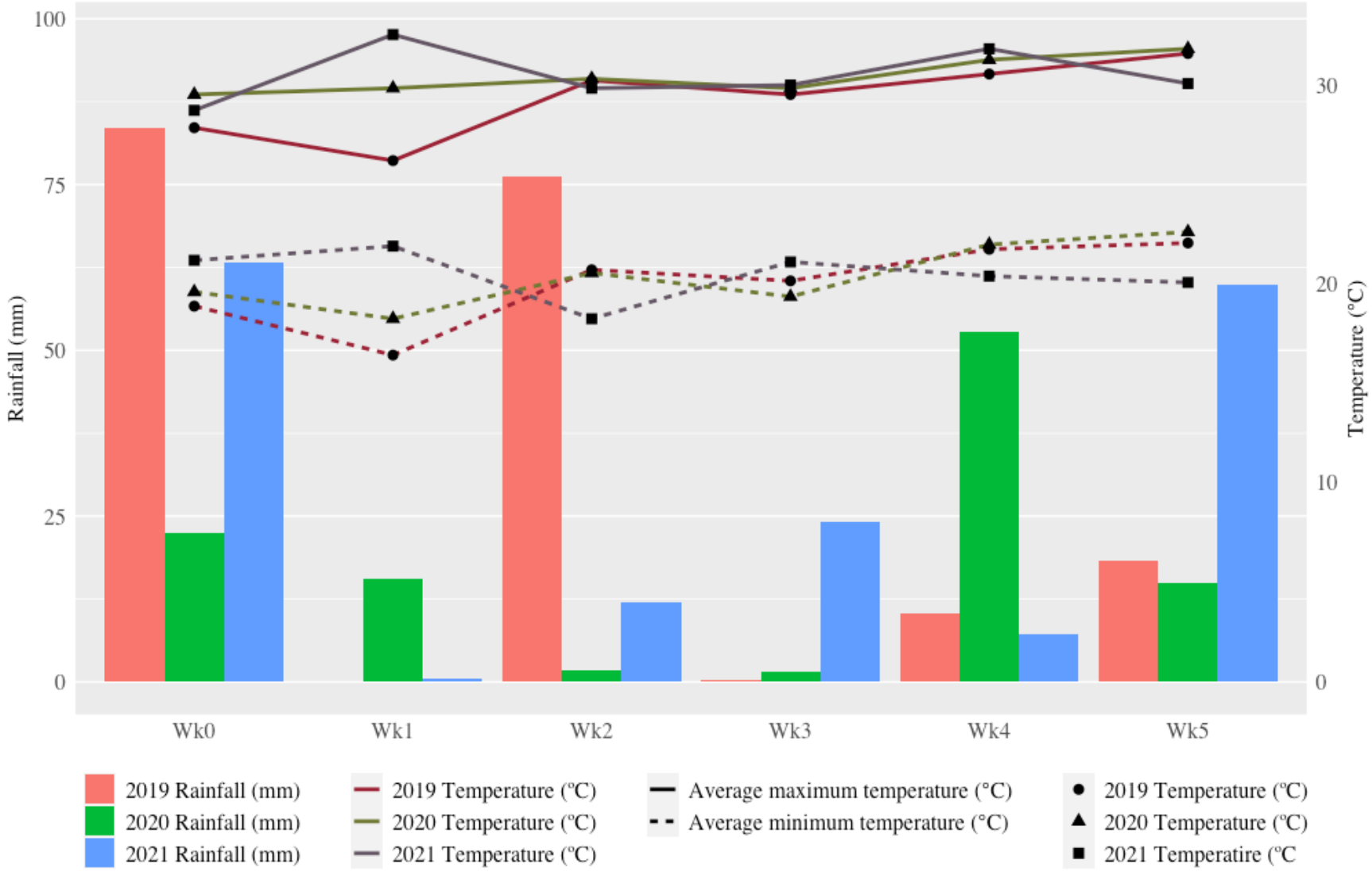


Fig. 1. Weather conditions, including weekly high and low average temperatures and rainfall accumulation, at the University of Arkansas System Division of Agriculture Fruit Research Station in Clarksville, AR during blackberry harvest seasons in 2019, 2020, and 2021. Records start on week zero with first harvest occurring on week one.

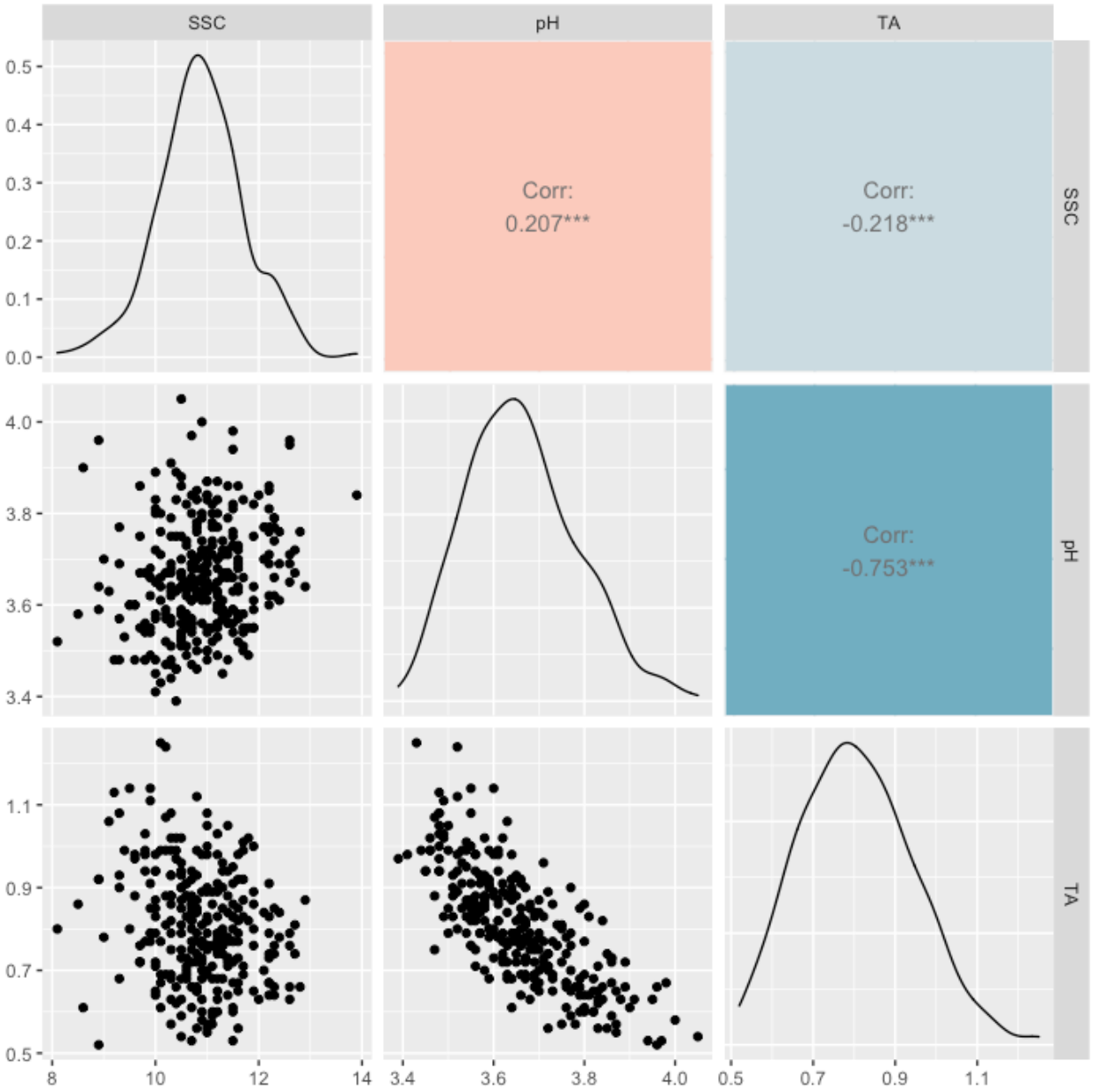


Fig. 2. Distribution and correlation of soluble solids content (SSC), pH, and titratable acidity (TA), with TA evaluated in 2019 and 2020 and SSC and pH evaluated in 2019, 2020, and 2021. Diagonal plots show the distributions best linear unbiased predictions (BLUPs) calculated over all years, scatter plots below the diagonal show relationships between traits, and Pearson's correlation coefficients are above the diagonal.

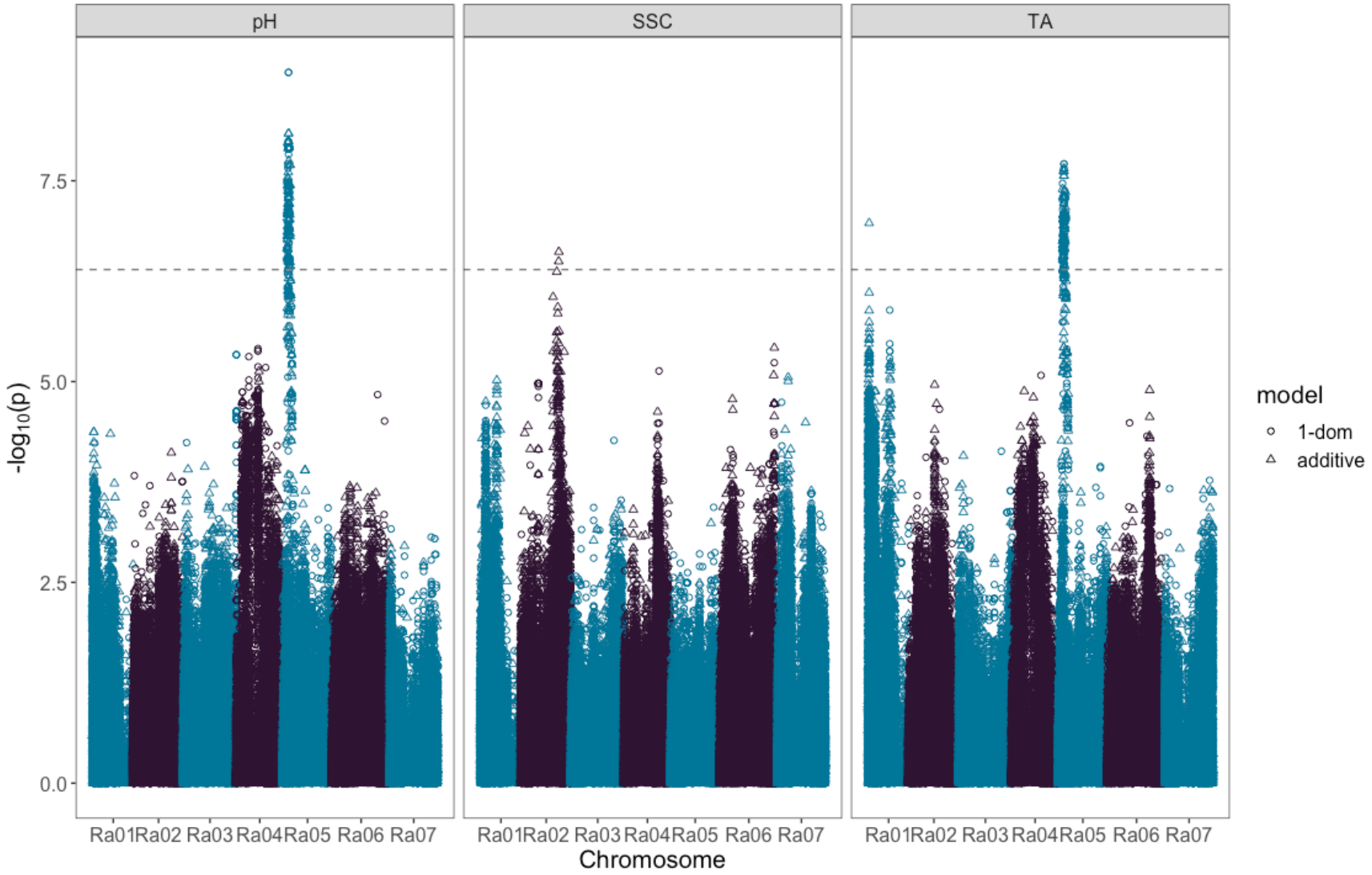


Fig. 3. Manhattan plot showing results from a fresh-market blackberry genome-wide association study. Soluble solids content (SSC) and pH phenotypes were calculated using data obtained during 2019, 2020, and 2021, while titratable acidity (TA) phenotypes were calculated using data obtained during 2019 and 2020. The dashed line indicates the Bonferroni threshold of $\alpha = 0.05$.

CHAPTER II

GENETIC CONTROL OF SEEDINESS IN BLACKBERRY

Abstract

Sales of fresh-market fruits, including blackberries (*Rubus* L. subgenus *Rubus* Watson), have been on the rise in recent years. Even still, many consumers complain that they dislike the large and noticeable seeds in most fresh-market blackberry cultivars. Breeding goals in blackberry include reducing seed size so their presence is less noticeable to consumers. An efficient approach to reducing seediness in blackberries would be to determine the genetic regions controlling seed size, shape, and number and to apply this knowledge to develop genetic markers to aide in more deliberate decisions while conducting crosses and choosing selections. With these long-term breeding goals in mind, we implemented a genome-wide association study (GWAS) as the first step towards determining the genetic control of seediness attributes in blackberry. For this study, we genotyped and phenotyped a panel of 277 fresh-market blackberry cultivars and breeding selections during the summers of 2019 and 2020. Genotyping for this study was conducted with Capture-Seq technology and association analysis was performed with GWASpoly. Seediness attributes evaluated included 100-seed weight, width to length (WL) ratio, seed area, and seeds per berry. Broad sense heritability estimates were 91% for 100-seed weight and WL ratio, 89% for seed area, and 70% for seeds per berry. Five hundred and five significant markers located on chromosomes Ra01, Ra03, and Ra05, were associated with seed WL ratio, weight, and area, respectively. Three possible gene candidates for WL ratio on Ra01 code for a serine carboxypeptidase 24 BRI SUPPRESSOR 1 (BRS1), a cytokinin oxidase 2 (CKX2), and G-protein subunit γ (GG3). Two candidates for 100-seed weight on Ra03 code for

an AAA-ATPase 2A RPT2A and a protein phosphatase AP2C1, and a possible candidate related to seed area on Ra05 codes for a DA1 protein.

Introduction

Consumption of fresh-market blackberries (*Rubus* L. subgenus *Rubus* Watson) has been growing in recent years, with sales in the United States from Aug. 2020 to Aug. 2021 totaling approximately 46,473 tons and Americans spending just over $671 million on blackberries (California Strawberry Commission, 2021). This increase in fresh fruit sales is likely due to consumers being more mindful of the foods they eat, the health benefits associated with the foods, and where those foods originate (Molina-Bravo et al., 2019). Grocers and consumers often complain that blackberries have too many seeds that are too big and too hard, and request for seeds to be softer, smaller, and less numerous (Coe, 2020). How the fibers of the berry are attached to the seed is another trait known to affect overall "seediness" perception (Darrow and Sherwood, 1931). Therefore, seed size, hardness, and the overall feel of seediness during consumption are common attributes breeders focus on when evaluating blackberries (Clark et al., 2007).

Blackberries are aggregate fruits composed of a cluster of drupelets surrounding a soft tissue receptacle, also called a torus. Each drupelet is comprised of an exocarp, a mesocarp, and a lignified endocarp (or pyrene) which encloses a single seed (Tomlik-Wyremblewska et al., 2010). *Rubus* seed coats have three distinct layers of heavily lignified sclerenchymatous cells (Wada et al., 2011). The endotesta of the seeds have uniformly globular, isodiametric, macrosclereid layers running perpendicularly to the mesotesta, and the outer mesotesta is composed of irregularly shaped macrosclerids (Wada et al., 2011). The complexity and thickness

of the outer epidermis (exotesta) and the alternating orientation of the macrosclerenchyma provide strength to the seed coat.

The *Rubus* genus consists of 12 subgenera, most of which only contain wild species (Foster et al., 2019). The two commercial subgenera are *Rubus* and *Idaeobatus*, which contain blackberries and red (*R. idaeus* L.) and black raspberries (*R. occidentalis* L.), respectively (Clark et al., 2007; Foster et al., 2019). A 2008 study found a high level of morphological diversity in *Rubus* seeds observed across 56 taxa and 10 subgenera, with seed size ranging from L × W × H dimensions of 1.4 × 1.2 × 0.81 mm to 6.0 × 3.5 × 2.35 mm (Wada and Reed, 2008). The seeds of red raspberries are generally not noticed by consumers compared to blackberries because of their smaller ratio of seed size to berry size. Seeds of the American blackberry species, *R. allegheniensis* Porter and *R. ursinus* Cham. & Schldl, and the European blackberries, *R. procerus* Muller and *R. caesius* L., were larger than seeds of the European red raspberry (Hummer and Peacock, 1994). Studies show seed size varies widely across the *Rubus* genus and within the *Rubus* subgenus, with these attributes likely influenced by genetics (Hummer and Peacock, 1994).

Seed attributes across cultivated blackberries vary widely, with trailing blackberries developed in the Western United States generally having ellipsoidal and smaller seeds compared to Eastern semi-erect blackberries, which have “clam shaped” seeds (Clark et al., 2007). Western U.S. blackberries, grown primarily for processing, are hexaploids and other higher order polyploids derived mostly from the wild species *R. ursinus*, while tetraploid fresh-market Eastern U.S. blackberries were mostly derived from crosses with the wild species *R. argutus* and *R. allegheniensis* (Clark and Finn, 2011; Meng and Finn, 2002). Western blackberries are often perceived as “seedless” or having a lower number of seeds compared to Eastern blackberries,

and progenies of crosses between Eastern and Western blackberries tend to show a range of seediness (Clark et al., 2007).

A descriptive panel rated seediness of 22 Eastern erect and semi-erect blackberry cultivars and breeding selections (genotypes) grown at the University of Arkansas System Division of Agriculture (UA) Fruit Research Station (FRS) on a 10-point scale (0 = no seeds to 9 = extremely seedy) (Sebesta et al., 2013). No significant differences among the genotypes were found by the panelists even though the mean seediness ratings ranged from 4.38 (A-2416 and 'Tupy') to 7.25 (A-2418). Cultivars observed as higher on the seediness scale were 'Prime-Ark®45' and 'Natchez', which were rated at 6.63 and 6.75, respectively. 'Natchez' had significantly higher pyrenes per berry with 131, compared to 'Navaho' and 'Tupy' with the lowest average pyrenes per berry at 53 (Sebesta et al., 2013). The proportion of seed weight to the total weight of the berry is more important than overall seed size or seed number (Darrow and Sherwood, 1931). The percent of berry weight constituted by seeds in the 22 UA fresh-market blackberry genotypes ranged from 2.7% to 5.4% (Sebesta et al., 2013). In a subsequent study, a descriptive sensory panel and a consumer panel evaluated a group of UA blackberry genotypes to evaluate seediness and other fruit quality attributes. The descriptive sensory panel evaluated different texture attributes, including size and amount of seeds in 11 UA genotypes (Threlfall et al., 2016). As in the 2013 study, the descriptive panelists identified differences in the size of pyrenes, amount of pyrenes, or loose particles across the genotypes evaluated. However, according to a principal component analysis (PCA), the amount of seeds was a negative driver of the overall liking of blackberries in the consumer panel (Threlfall et al., 2016).

No research has been conducted on the genetics of seed attributes in *Rubus* or Rosaceae species. However, seed attributes have been studied in species such as *Arabidopsis thaliana* (L.)

Heynh., soybean [*Glycine max* (L.) Merr.], and rice (*Oryza sativa* L.) as seed size is an important grain yield component. Angiosperm seeds consist of an endosperm and embryo wrapped in a seed coat (Li et al., 2019). Seed development occurs in two phases; a morphogenesis phase where cell division occurs during endosperm, embryo, and cotyledon development, followed by a maturation phase where the embryo grows and accumulates seed storage materials, such as proteins, starches, lipids, and nutrients, and the seed desiccates (Savadi, 2018). Seed growth is controlled by zygotic tissues of the endosperm and embryo, as well as sporophytic integuments of the maternal tissue that develop into the seed coat after fertilization (Li et al., 2019). Simultaneously while the seed coat is forming from the maternal tissue, the embryo and endosperm develop from the zygotic tissue (Li et al., 2019). Growth regulation of maternal tissues impacts seed size and is under the control of several signaling pathways, including the ubiquitin-proteasome pathway, G-protein signaling, mitogen-activated protein kinase (MAPK) signaling, perception and homeostasis of phytohormones, and some transcriptional factors (Li et al., 2019). Seed size can also be impacted by the growth and development of zygotic tissues. In *A. thaliana* and most other dicots, growth of the triploid endosperm precedes embryo growth, and the embryo is still in the globular stage and occupies only a small part of the seed when the seed has nearly reached its final size (Li et al., 2019). Therefore, the endosperm plays a crucial role in determining seed size. Endosperm development and seed size are influenced by the HAIKU (IKU) genetic pathway, which is regulated by abscisic acid and brassinosteroid hormones (Garcia et al., 2003; Li et al., 2013). Other plant hormones including auxin, cytokinin, brassinolide, and gibberellic acid also affect endosperm development (Savadi, 2018).

Genome-wide association studies can be used to find associations between phenotypic measurements and single nucleotide polymorphisms (SNPs) in an organism's DNA and identify

possible quantitative trait loci (QTL) associated with the control of these phenotypes. Blackberries have ploidies ranging between 2n=2x=14 and 2n=14x=98, with Eastern U.S. cultivars being mostly autotetraploid (Clark et al., 2007). Autopolyploids require specialized tools to properly call allele dosage and carry out genetic studies (Bourke et al., 2018). The R package 'GWASpoly' utilized in this study was created specifically for GWAS in polyploid species (Rosyara et al., 2016). This study focuses on seed attributes of tetraploid fresh-market blackberries including 100-seed weight, width to length (WL) ratio, seed area, and seeds per berry. Using a GWAS to determine the genetic control of seediness attributes in blackberry will allow for the development of genetic markers. These markers can be used for making crossing plans as well as in the early stages of seedling development during selection in the breeding process.

Objectives

The objectives of this study were to:

3. Evaluate seediness attributes including 100-seed weight, WL ratio, seed area, and seeds per berry using the seeds collected from a large panel of commercially available fresh-market blackberry cultivars and UA breeding selections.
4. Conduct a GWAS to identify marker-trait associations, QTL, and possible candidate genes associated with seediness attributes in blackberry.

Materials and Methods

Plant Material

A panel of 277 blackberry genotypes consisting of 29 commercially available cultivars and 248 UA breeding selections was chosen for GWAS analysis. Plants used for this study were grown and maintained in 6 m plots at the UA FRS located in Clarksville, Arkansas. The FRS is

located in the U.S. Department of Agriculture (USDA) hardiness zone of 7b (USDA, 2021) and on Linker fine sandy loam type soil (Salgado and Clark, 2016). The research plots received regular maintenance of pruning, tipping, and irrigation. Management of the plots consisted of fungicide applications at budbreak for anthracnose, bloom for botrytis, and primocane emergence for orange rust. Insecticide applications occurred weekly for spotted wing drosophila, with periodic insecticide applications as needed to control for strawberry clipper, Japanese beetle, and broad mite throughout the growing season.

Sample Collection

Sample collection began in mid-June and continued through the first week of July during the summers of 2019 and 2020. During the 2019 harvest season 236 genotypes were phenotyped, while 237 genotypes were phenotyped in 2020 (Supplementary Table 1 of Appendix B). Two replicates were harvested one week apart for each genotype during both harvest seasons, resulting in a total of 881 samples for analysis. One pint-sized clamshell of floricane fruit was collected at the shiny black stage, avoiding berries with discolored or damaged drupelets to ensure a high-quality sample. A 10-berry sample was squeezed to remove the juice and produce a pulp of berry flesh and seeds. The pulp from each sample was retained in individual zip-top bags and frozen at -10 ºC for this study.

Seed Extraction

Seeds were isolated from samples by first thawing each bag of berry pulp, then adding 20 mL of DI water and 200 µl of pectinase (Carolina Biological Supply Company, Burlington, NC). Samples were left to process for 48 h at 5 ºC to ensure sufficient pectin catabolism and reduce molding. To isolate seeds after processing, the pulp solution was transferred to a fine mesh strainer and lixiviated with water until minimal berry flesh remained. Rinsed samples were

dried in a vent hood for an additional 24 h at room temperature (21 ºC). Once dried, the seeds had flecks of berry flesh which were removed through an additional step consisting of gentle abrasion of the seeds with a fine mesh strainer (2 mm). All samples were transferred to paper coin envelopes and dried at 60 ºC for 24 h before analysis.

Evaluation of Seediness Phenotypes

Weights for all the seeds from the 10-berry sample and 100-count seed subsamples were quantified using an OHAUS Pioneer PX analytical balance (H & C Weighing Systems, Columbia, MD). The estimated number of seeds per berry was calculated by dividing the full 10-berry sample seed weight from the 100-seed weight and then multiplying the quotient by ten. To validate the estimated seeds per berry, the full number of seeds for 30% of the 10-berry samples were hand-counted. Values from the estimated seeds per sample and hand-counted seeds per sample were compared with a Pearson's correlation coefficient. Seed area and WL ratio were measured on the 100-seed samples using WinSEEDLE (Regent Instruments Inc., Québec, Canada) software.

Analysis of Heritability and Trait Correlations

Best linear unbiased predictions (BLUPs) of 100-seed weight, WL ratio, area, and seeds per berry were calculated for each genotype over both years and all replicates using the Type 3 method of the MIXED statement in SAS v. 9.4 (SAS Institute, Inc., Cary, NC). Harmonic means were calculated using the "dplry" package in R, while variance components and residual effects were obtained through the Type3 method of PROC MIXED in SAS. Broad-sense heritability (H) for each trait was calculated using the harmonic means for years (y) and replicates (r), genotypic variance (σ^2_g), genotype by year variance (σ^2_{gy}), and residual effects (σ^2). Broad-sense heritability was calculated as: $H = \sigma^2_g / [\sigma^2_g + (\sigma^2_{gy}/y) + (\sigma^2/yr)]$. Pearson correlation coefficients

were calculated using the “ggplot2” and “GGally” packages in R to measure the strength and direction of linear relationships between seediness attributes.

CaptureSeq Genotyping

A modified cetyltrimethylammonium bromide (CTAB) protocol was used to extract DNA collected from young leaves of each genotype (Porebski et al., 1997). Quantification was completed using the dsDNA assay kit by Qubit (Invitrogen, Carlsbad, CA). Sequence capture genotyping was performed at RAPiD Genomics (Gainesville, FL). The *Rubus argutus* ‘Hillquist’ V1 reference genome (Worthington et al., 2020) was used to custom-design 35,054 biotinylated 120-mer Capture-Seq probes consisting of 17,100 probes in candidate gene regions and 17,954 probes scattered across the genome to cover linkage disequilibrium (LD). Paired-end sequencing was performed on Illumina HiSeq2000. Raw data from sequencing were cleaned, trimmed, and aligned to the ‘Hillquist’ genome using MOSAIK (Lee et al., 2014). Variant calling was performed using Freebayes (Garrison and Marth, 2012). SNPs were filtered in VCFtools (Danecek et al., 2011) to produce a file with biallelic markers with a minor allele frequency of ≥ 0.01 and an average read depth between 3 and 750 per sample. Filtered data was then converted to tetraploid SNP calls using the multidog function in UpDog (Gerard et al., 2018).

GWAS Analysis

Association analysis was performed in GWASpoly (Rosyara et al., 2016) using 100-seed weight, WL ratio, area, and seeds per berry BLUPs calculated for each genotype along with the genotypic data generated from Capture-Seq. Data were analyzed using a random effect model with a kinship matrix constructed with the leave-one-chromosome-out (LOCO) method. Additive, simplex dominant alternative (simplex-dom-alt), and simplex dominant reference (simplex-dom-ref) gene action models were tested. The SNP effect in the additive model is

proportional to the dosage of the minor allele, whereas in the two simplex dominant models all three heterozygotes are equivalent to one of the homozygotes (Rosyara et al., 2016). Significance thresholds were calculated using the Bonferroni method with $\alpha = 0.05$.

Identification of Candidate Genes

The region of the chromosome containing significant markers as well as 1 Mb extended past either side of the markers were regions of consideration for the location of possible candidate genes. Queries of 38,503 genes of the 'Hillquist' V1 genome assembly were functionally annotated using the Swiss-Prot, now Uniprot (The UniProt Consortium, 2021) database and the Araport11 (Cheng et al., 2017) blastsets to predict protein-coding sequences. Genes within the region of consideration having functional annotations related to seediness were considered as possible candidate genes.

Results

Phenotypic Results

Seediness attributes were measured for 881 samples from 277 genotypes collected in the summers of 2019 and 2020 (Supplementary Table 1 of Appendix B). Normalized BLUPs for 100-seed weight ranged from 194.7 mg to 443.5 mg with an average of 318.9 mg across both years (Table 1). The genotype with the highest 100-seed weight was A-2444T, while the genotype with the lowest 100-seed weight was 'Choctaw' (Fig. 1). Seed area ranged from 3.93 mm^2 to 6.71 mm^2 with an average of 5.24 mm^2 (Table 1). A-2676T had the largest seed area. The genotype with the smallest seed area was 'Choctaw', which also had the lowest 100-seed weight (Fig. 1). Width to length ratio ranged from 0.56 to 0.82, with APF-PBB1 and A-2484T having the smallest and largest ratios, respectively (Table 1, Fig. 2). The average WL ratio across both harvest years was 0.69.

Ten-berry samples were unavailable for 106 replicates from 2019 and 35 replicates from 2020. As a result, these samples were excluded from analyses related to number of seeds per berry, leaving 267 genotypes and 739 replicates for analysis. The estimates of seeds per berry determined through manual counts of full samples and calculations based on total seed weight and 100-seed weights were strongly correlated (r = 0.97, Fig. 3). Berries had an average of 80.9 seeds with a range of 43.6 to 123.1 seeds per berry. The genotype with the lowest number of seeds per berry was APF-318, and the genotype with the most seeds per berry was A-2473T. Traits exhibited similar phenotypic ranges and averages across both harvest years, though seeds were slightly heavier and larger in 2019 than in 2020 (Table 1).

Seed area and 100-seed weight were strongly correlated (r = 0.83) (Fig. 4). Significant negative correlations were also identified between seeds per berry and 100-seed weight (r = -0.20) as well as seeds per berry and WL ratio (r = -0.34). No significant correlations were found between WL ratio and 100-seed weight (r = 0.03), WL ratio and area (r = -0.11), or area and seeds per berry (r = 0.01). Broad-sense heritability estimates were 91% for 100-seed weight and WL ratio, 89% for seed area, and 70% for seeds per berry.

Association Analysis

Sequencing data from 272 genotypes along with their phenotype data were analyzed in GWASpoly with a Bonferroni threshold set at $\alpha = 0.05$ to analyze 124,546 markers across the *Rubus* genome. The threshold set for the additive model had a LOD of 6.39, while the two simplex models, 1-dom-alt and 1-dom-ref, had a LOD threshold of 6.23. A total of 504 significant markers were discovered during analysis. Four of the significant markers were found using the simplex dominant model while the remaining 500 markers were additive, indicating allele dosage is likely important in the genetic control of seediness. Three QTLs were identified,

one each for WL ratio, 100-seed weight, and seed area with peaks located on chromosomes Ra01, Ra03, and Ra05, respectively (Fig. 5). The QTL for seed area consisted of two significant markers, both in the additive model, located on Ra05 at 22305960 and 27929457 bp with LOD scores of 7.02 and 6.56. The QTL for 100-seed weight was made up of 15 significant markers on chromosome Ra03 from 37205067 to 39519728 bp with a peak LOD of 7.24 and additive gene action. A peak for seed area was also observed in this same region on chromosome Ra03; however, none of the markers were significant after Bonferroni correction (Fig. 5).

The trait with the most highly significant QTL was WL ratio. A QTL for WL ratio consisting of 484 significantly associated markers was discovered on Ra01 between 1045454 and 5435208 bp with a peak LOD score of 9.9 at 2972427 bp. Four hundred and eighty-three markers in this region were significantly associated with WL ratio considering only the additive model, one marker was significant using only the simplex dominance model, and three markers were significant according to both models. The additive model produced higher LOD scores than the simplex dominance model for all three markers that were significant according to both metrics. Consistent significant peaks associated with WL ratio were discovered on chromosome Ra01 when 2019 and 2020 data were analyzed separately and in the analysis of both years combined (Fig. 6). No markers were significantly associated with the number of seeds per berry.

Possible Candidate Genes

To determine possible candidate genes, significant markers obtained from GWAS analysis were aligned with annotated genes from the 'Hillquist' reference genome. Genes within the significant QTL region as well as 1 Mb on either side of the QTL were considered possible candidates. A total of six possible candidate genes previously shown to control seed size and shape in *A. thaliana* and rice were identified within the candidate region, three on chromosome

Ra01, two on chromosome Ra03, and one on Ra05 (Table 2). The possible candidate genes in the QTL region for WL ratio on chromosome Ra01 were Ra_g83, Ra_g689, and Ra_g820. These genes were annotated as a serine carboxypeptidase 24 BRI SUPPRESSOR 1 (BRS1), a cytokinin oxidase 2 (CKX2), and a G-protein subunit γ (GG3), respectively. Ra_g83 was located 640 kb upstream of the QTL, while Ra_g689 and Ra_g820 were located within the QTL region. Possible gene candidates within the QTL for 100-seed weight on Ra03 were an AAA-ATPase 2A (RPT2A) annotated gene (Ra_g14549) and a protein phosphatase (AP2C1) coding gene (Ra_g14574). A gene candidate for the DA1 protein (Ra_g23187) was located within the candidate region 183 kb upstream from the QTL for seed area on Ra05.

Discussion

Seed attribute measurements from our study were comparable to those reported in previous studies on seediness in *Rubus*. Our study reports 100-seed weights ranging from 194.7 mg to 443.5 mg with an average of 318.9 mg. A study evaluating seed size in wild species from subgenus *Rubus* found a range of 100-seed weight from 159 mg to 386 mg (Hummer and Peacock, 1994). In 1931, a study comparing 40 Eastern erect and Western trailing blackberry genotypes found 100-seed weights from 108 mg to 382 mg (Darrow and Sherwood, 1931). Another study found 100-seed weight from 41 UA genotypes ranged from 120 mg and 430 mg (Moore et al., 1974). Our observed correlation between 100-seed weight and area ($r = 0.83$) was expected, as seeds with a larger area are likely to have a higher overall weight. The genotype with the lowest 100-seed weight in both years and the smallest area was ‘Choctaw’, consistent with the strong correlation between the two traits.

The number of seeds per berry in the genotypes evaluated in this study ranged from 43.6 to 123.1 with an average of 80.9. Fourteen of the genotypes included in this study were also

phenotyped for seeds per berry by Sebesta et al. (2013). Of those 14 overlapping genotypes, eight had discrepancies of 10 or fewer seeds per berry between the two studies, and the remaining six genotypes had a discrepancy of 30 or fewer seeds per berry between the two studies (Sebesta et al., 2013). Eleven genotypes evaluated in a study of advanced UA breeding selections blackberry and cultivars had pyrenes per berry ranging from 51 to 115 (Threlfall et al., 2016). Another study of 41 UA blackberry genotypes collected in 1970 reported a range of seeds per berry from 27 to 74 (Moore et al., 1974).

This study is the first to use WinSEEDLE to analyze area and WL ratio data on a large number of *Rubus* genotypes; however, other studies have measured seed shape in a smaller set of genotypes. Variation in seed shape between blackberry genotypes has been assessed visually, with seeds described as long, short, narrow, wide, concave, convex, or straight (Sebesta et al., 2013; Wada et al., 2010). The effect of seed shape attributes, such as WL ratio, have yet to be included in assessments of consumer perception of blackberry seediness. Blackberries from our study had WL ratios ranging from 0.56 to 0.82 with an average of 0.69. Western trailing blackberries usually have long and skinny seeds and are often perceived as less seedy than Eastern erect and semierect cultivars (Clark and Finn, 2011). The popular Western trailing blackberry cultivar ‘Obsidian’ had a 100-seed weight of 260 mg (Wada et al., 2010), within the range reported in this study for Eastern erect and semierect fresh-market genotypes. However, the WL ratio for ‘Obsidian’ was 0.47 (Wada et al., 2010), below the minimum WL ratio reported for the GWAS panel, which suggests that seed WL ratio could a play role in the overall perception of seediness.

Heritability estimates from our analysis were 91% for 100-seed weight and WL ratio and 89% for seed area. These estimates are slightly lower than the 97%, 98%, 98%, and 96%

estimates of heritability for seed weight previously reported in four different biparental populations from the UA breeding program (Moore et al., 1975). This same research noted that seed size showed little variation across several years and in varying locations, indicating the environment has little effect on this trait (Moore et al., 1975). The heritability estimate for seeds per berry was the lowest observed in this study at 70%. The lower heritability estimate for seeds per berry is not entirely unexpected and could be due in part to the smaller, more unbalanced sample size for seeds for berry compared to other seediness attributes. The number of seeds per berry is also dependent on pollination and weather (Strik et al., 1996). Furthermore, the heritability of seeds per berry could be affected by variable levels of self-pollination and partial self-incompatibility in some UA germplasm (Lopez-Medina and Moore, 2001; Perry and Moore, 1985) which could cause uneven drupelet set.

Three QTLs associated with seediness were identified through GWAS analysis on Ra01 for WL ratio, Ra03 for 100-seed weight, and Ra05 for seed area. Each QTL region, including 1 Mb on either side of the QTL, were investigated for genes associated with seed attributes. Candidate genes were chosen based on results from studies of other angiosperms, as no genetic studies on seed attributes have been published on any *Rubus* or Rosaceae species. The QTL for WL ratio between 1045454 and 5435208 bp on Ra01 consisted of 484 significantly associated markers with a peak LOD score of 9.9. Three possible candidate genes for WL ratio were identified in the region of the QTL. Ra_g83, a gene annotated as a BRI SUPPRESSOR 1 (BRS1), located approximately 600 kb proximal of the QTL for WL ratio was chosen a potential candidate because of its involvement in brassinosteroid signaling through its interaction with BRASSINSTEROID INSENSITIVE 1 (BRI1) (Jiang et al., 2013; Li et al., 2001). In *A. thaliana bri1-5* mutants with weak expression of BRI1 produced seeds with slightly lower weight and

increased WL ratio compared with wild-type controls. Furthermore, *GS5*, a gene encoding putative serine carboxypeptidase protein with high sequence similarity to BRS1 was identified as a candidate gene associated with a QTL for grain width and weight in rice (Li et al., 2011; Xu et al., 2015). Overexpression of BRS1 in rice can increase grain width, strongly suggesting that BRS1 may function similarly to GS5 (Xu et al., 2015). Another possible candidate gene for WL ratio, Ra_g820, is located within the QTL region on Ra01 and codes for a GG3. The GTP-binding proteins (G-proteins) of plants are signal transduction components that consist of α (alpha), ß (beta), and γ (gamma) subunits, and play a part in organ growth (Chakravorty et al., 2011). The G-protein subunit γ (AGG3) of *A. thaliana* is known to play a role in determining organ size and shape (Li et al., 2012). A candidate gene for a major grain size QTL in rice, *GS3*, codes for a protein homologous with AGG3. The null allele of *GS3* is associated with a long grain phenotype that is highly favored by rice breeders (Fan et al., 2006; Mao et al., 2010; Takano-Kai et al., 2009). A third possible candidate gene identified within the QTL for WL ratio on Ra01 is Ra_g689, which codes for a CKX2. Cytokinin oxidase 2 is an enzyme involved in cytokinin catabolism (Frébortová et al., 2007) that is directly activated by the IKU transcription factor WRKY10 to promote endosperm growth (Li et al., 2013; Su et al., 2021). Overexpression of *CKX2* in *A. thaliana* results in a phenotype with larger and fewer seeds (Werner et al., 2003). Although *CKX2* mutants result in differentiated seed size in *A. thaliana*, they have not yet been associated with variation in seed shape (Li et al., 2013).

Genes coding for RPT2A (Ra_g14549) and the protein phosphatase AP2C1 (Ra_g14574) were identified as potential candidates in the 100-seed weight QTL on Ra03. The AAA-ATPase 2A (RPT2A) is one of two homologous genes encoding the 19S regulatory particle of the 26S proteasome, which degrades ubiquitinated proteins, effectively removing abnormal proteins and

short-lived regulatory proteins (Smalle and Vierstra, 2004). *A. thaliana* plants with a loss of function of *AtRPT2A* had reduced cell number and increased in cell size associated with enlarged leaves, stems, flowers, fruits, seeds, and embryos (Kurepa et al., 2009). Ra_g14574, the gene coding for the PP2C phosphatase AP2C1, was chosen as a possible candidate for 100-seed weight. Studies have shown AP2C1 is a stress signal regulator shown to negatively regulate MAPK4 and MAPK6 activity in *A. thaliana* (Schweighofer et al., 2007). MAPK6 is an important component of the signaling pathway for maternal control of embryogenesis, and *mpk6* mutant *A. thaliana* plants produced seeds with burst-out embryos or wrinkled seed coats (Zhang et al., 2017). OsMAPK6 activity was also positively associated with grain size in rice (Xu et al., 2018b). Although AP2C1 has not been directly associated with variation in seed size, it has significant functional redundancies with DSP-type MAPK phosphatase 1 (MKP1) genes through their shared action as negative regulators of MAPK6. Loss of function mutations in *OsMKP1* resulted in large grains, while overexpression of *OsMKP1* led to small grains phenotypes in rice (Ayatollahi et al., 2021; Xu et al., 2018a). Genes coding for other proteins of the PP2C family located in the 100-seed weight QTL region on Ra03, including PP2C-24, -6, and -55, were not selected as candidates because they are not in the same clade as the PP2C-1 proteins known to control seed weight (Schweighofer et al., 2004). A gene coding for a PP2C26 protein found within the peak region was also excluded as a possible candidate gene for the genetic control of 100-seed weight because its function is related to photosynthesis and the plant immune system (Akimoto-Tomiyama et al., 2018).

Ra_g23187, the only possible candidate gene identified for seed area, coded for a DA1 protein and was located approximately 180 kb upstream of the QTL on chromosome Ra05. DA1 proteins are known to play a part in the genetic control of seed size in *A. thaliana* by working

with ubiquitin receptor proteins to regulate the activity or binding capacity of other proteins initiating a signal cascade to negatively regulating seed size by restricting integument cell proliferation (Adamski et al., 2009; Li et al., 2008; Li and Li, 2014; Savadi, 2018). DA1 mutants in *A. thaliana*, such as *da1-1*, have restricted integument growth, resulting in phenotypes with large seeds and other organs (Li et al., 2019).

Conclusion

In this study, we present the first GWAS of traits related to seediness in blackberry. Significant marker-trait associations were discovered for 100-seed weight, WL ratio, and area, and six possible gene candidates related to seediness were identified. All of the six possible candidate genes are known to affect seed attributes in *A. thaliana* and rice. Results from this study are the first steps in determining the genetic control of seediness in blackberry. This data can also be used as a training set to develop genomic selection (GS) models for moderate heritability traits without major effect QTL such as seeds per berry. Unlike association mapping studies, GS uses all available molecular markers to predict the performance and breeding values of candidates for selection (Crossa et al., 2017).

Literature Cited

Adamski, N.M., E. Anastasiou, S. Eriksson, C.M. O'Neill, and M. Lenhard. 2009. Local maternal control of seed size by KLUH/CYP78A5-dependent growth signaling. Proc. Natl. Acad. Sci. U.S. 106:20115–20120.

Akimoto-Tomiyama, C., S. Tanabe, H. Kajiwara, E. Minami, and H. Ochiai. 2018. Loss of chloroplast-localized protein phosphatase 2Cs in *Arabidopsis thaliana* leads to enhancement of plant immunity and resistance to *Xanthomonas campestris* pv. *campestris* infection. Mol. Plant Pathol. 19:1184–1195.

Ayatollahi, Z., V. Kazanaviciute, V. Shubchynskyy, B. Mueller-roeber, I. Meskiene, M.P. Labs, V. Biocenter, A. Life, K. Kvederaviciute, M. Schwanninger, W. Rozhon, M. Stumpe, F. Mauch, S. Bartels, R. Ulm, S. Balazadeh, B. Mueller-roeber, I. Meskiene, and A. Schweighofer. 2021. Dual control of MAPK activities by AP2C1 and MKP1 MAPK phosphatases regulates defence responses in *Arabidopsis*. BioRxiv. Preprint.

Bourke, P.M., R.E. Voorrips, R.G.F. Visser, and C. Maliepaard. 2018. Tools for genetic studies in experimental populations of polyploids. Front. Plant Sci. 9:513.

California Strawberry Commission. 2021. U.S. retail category trends. 23 August 2021. <https://www.calstrawberry.com/en-us/market-data/retail-category-trends/>

Chakravorty, D., Y. Trusov, W. Zhang, B.R. Acharya, M.B. Sheahan, D.W. McCurdy, S.M. Assmann, and J.R. Botella. 2011. An atypical heterotrimeric G-protein β-subunit is involved in guard cell K^+-channel regulation and morphological development in *Arabidopsis thaliana*. Plant J. 67:840–851.

Cheng, C.Y., V. Krishnakumar, A.P. Chan, F. Thibaud-Nissen, S. Schobel, and C.D. Town. 2017. Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. Plant J. 89:789–804.

Clark, J.R., and C.E. Finn. 2011. Blackberry breeding and genetics. Fruit, Veg. Cereal Sci. Biotechnol. 5:27–43.

Clark, J.R., E.T. Stafne, H.K. Hall, and C.E. Finn. 2007. Blackberry breeding and genetics. Plant Breeding Rev. 29:19–144.

Coe, M., 2020. Status and priorities of the U.S. blackberry industry: preliminary findings from the 2019 U.S. blackberry industry survey., Portland, O.R.

Crossa, J., P. Pérez-Rodríguez, J. Cuevas, O. Montesinos-López, D. Jarquín, G. de los Campos, J. Burgueño, J.M. González-Camacho, S. Pérez-Elizalde, Y. Beyene, S. Dreisigacker, R. Singh, X. Zhang, M. Gowda, M. Roorkiwal, J. Rutkoski, and R.K. Varshney. 2017. Genomic selection in plant breeding: methods, models, and perspectives. Trends Plant Sci. 22:961–975.

Danecek, P., A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, R.E. Handsaker, G. Lunter, G.T. Marth, S.T. Sherry, G. McVean, and R. Durbin. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158.

Darrow, G.M., and H. Sherwood. 1931. Seed and berry size of cane fruits. Proc. Amer. Soc. Hortic. Sci. 28:194–199.

Fan, C., Y. Xing, H. Mao, T. Lu, B. Han, C. Xu, X. Li, and Q. Zhang. 2006. GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theor. Appl. Genet. 112:1164–1171.

Foster, T.M., N.V. Bassil, M. Dossett, M.L. Worthington, and J. Graham. 2019. Genetic and genomic resources for *Rubus* breeding: a roadmap for the future. Hortic. Res. 6:116.

Frébortová, J., P. Galuszka, T. Werner, T. Schmülling, and I. Frébort. 2007. Functional expression and purification of cytokinin dehydrogenase from *Arabidopsis thaliana* (AtCKX2) in *Saccharomyces cerevisiae*. Biol. Plant. 51:673–682.

Garcia, D., V. Saingery, P. Chambrier, U. Mayer, G. Jürgens, and F. Berger. 2003. *Arabidopsis* haiku mutants reveal new controls of seed size by endosperm. Plant Physiol. 131:1661–1670.

Garrison, E., and G. Marth. 2012. Haplotype-based variant detection from short-read sequencing. Genomics 2:1–9.

Gerard, D., L.F.V. Ferrão, A.A.F. Garcia, and M. Stephens. 2018. Genotyping polyploids from messy sequencing data. Genetics 210:789–807.

Hummer, K.E., and D.N. Peacock. 1994. Seed dimension and weight of selected *Rubus* species. HortScience 29:1034–1036.

Jiang, W.B., H.Y. Huang, Y.W. Hu, S.W. Zhu, Z.Y. Wang, and W.H. Lin. 2013. Brassinosteroid regulates seed size and shape in *Arabidopsis*. Plant Physiol. 162:1965–1977.

Kurepa, J., S. Wang, Y. Li, D. Zaitlin, A.J. Pierce, and J.A. Smalle. 2009. Loss of 26S proteasome function leads to increased cell size and decreased cell number in *Arabidopsis* shoot organs[1[C][W][OA]]. Plant Physiol. 150:178–189.

Lee, W.P., M.P. Stromberg, A. Ward, C. Stewart, E.P. Garrison, and G.T. Marth. 2014. MOSAIK: A hash-based algorithm for accurate next-generation sequencing short-read mapping. PLoS One. 9:3.

Li, J., K. a Lease, F.E. Tax, and J.C. Walker. 2001. BRS1, a serine carboxypeptidase, regulates BRI1 signaling in *Arabidopsis* thaliana. Proc. Natl. Acad. Sci. U.S. 98:1–6.

Li, J., X. Nie, J.L.H. Tan, and F. Berger. 2013. Integration of epigenetic and genetic controls of seed size by cytokinin in *Arabidopsis*. Proc. Natl. Acad. Sci. U.S. 110:15479–15484.

Li, N., and Y. Li. 2014. Ubiquitin-mediated control of seed size in plants. Front. Plant Sci. 5:1–6.

Li, N., R. Xu, and Y. Li. 2019. Molecular networks of seed size control in plants. Annu. Rev. Plant Biol. 70:435–463.

Li, S., Y. Liu, L. Zheng, L. Chen, N. Li, F. Corke, Y. Lu, X. Fu, Z. Zhu, M.W. Bevan, and Y. Li. 2012. The plant-specific G protein γ subunit AGG3 influences organ size and shape in *Arabidopsis thaliana*. New Phytol. 194:690–703.

Li, Y., C. Fan, Y. Xing, Y. Jiang, L. Luo, L. Sun, D. Shao, C. Xu, X. Li, J. Xiao, Y. He, and Q. Zhang. 2011. Natural variation in *GS5* plays an important role in regulating grain size and yield in rice. Nat. Genet. 43:1266–1269.

Li, Y., L. Zheng, F. Corke, C. Smith, and M.W. Bevan. 2008. Control of final seed and organ size by the *DA1* gene family in *Arabidopsis thaliana*. Genes Dev. 22:1331–1336.

Lopez-Medina, J., and J.N. Moore. 2001. Drupelet set on 'Arapaho' thornless blackberry flowers protected and nonprotected from cross pollination and assessment of resulting seedlings. HortScience 36:174.

Mao, H., S. Sun, J. Yao, C. Wang, S. Yu, C. Xu, X. Li, and Q. Zhang. 2010. Linking differential domain functions of the GS3 protein to natural variation of grain size in rice. Proc. Natl. Acad. Sci. U.S. 107:19579–19584.

Meng, R., and C.E. Finn. 2002. Determining ploidy level and nuclear DNA content in *Rubus* by flow cytometry. J. Amer. Soc. Hortic. Sci. 127:767–775.

Molina-Bravo, R., M.L. Worthington, and G.E. Fernandez. 2019. Advances and challenges in raspberry and blackberry breeding, p.1-33. In: G.A. Lang (ed.). Achieving sustainable cultivation of temperate zone tree fruits and berries. vol. 2. Burleigh dodds, Cambridge, UK.

Moore, J.N., G.R. Brown, and E. Brown. 1974. Relationships between fruit size and seed number and size in blackberries. Fruit Var. J. 40–45.

Moore, J.N., C. Lundergan, and E.D. Brown. 1975. Inheritance of seed size in blackberry. J. Amer. Soc. Hortic. Sci. 100:377–379.

Perry, J. l., and J.N. Moore. 1985. Self and cross compatability in tetraploid blackberry cultivars. HortScience 20:738–739.

Porebski, S., L.G. Bailey, and B.R. Baum. 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Mol. Biol. Reporteriology Report 15:8–15.

Rosyara, U.R., W.S. De Jong, D.S. Douches, and J.B. Endelman. 2016. Software for genome-wide association studies in autopolyploids and its application to potato. Plant Genome 9(2):1–10.

Salgado, A.A., and J.R. Clark. 2016. “Crispy” blackberry genotypes: a breeding innovation of the university of Arkansas blackberry breeding program. HortScience 51:468–471.

Savadi, S. 2018. Molecular regulation of seed development and strategies for engineering seed size in crop plants. Plant Growth Regul. 84:401–422.

Schweighofer, A., V. Kazanaviciute, E. Scheikl, M. Teige, R. Doczi, H. Hirt, M. Schwanninger, M. Kant, R. Schuurink, F. Mauch, A. Buchala, F. Cardinale, and I. Meskiene. 2007. The PP2C-type phosphatase AP2C1, which negatively regulates MPK4 and MPK6, modulates innate immunity, jasmonic acid, and ethylene levels in *Arabidopsis*. Plant Cell 19:2213–2224.

Schweighofer, A., H. Hirt, and I. Meskiene. 2004. Plant PP2C phosphatases: Emerging functions in stress signaling. Trends Plant Sci. 9:236–243.

Sebesta, B., J.R. Clark, R.T. Threlfall, and L.R. Howard. 2013. Characterization of seediness attributes of blackberry genotypes. Discov. Student J. Dale Bump. Coll. Agric. Food Life Sci. 14:72–79.

Smalle, J., and R.D. Vierstra. 2004. The ubiquitin 26S proteasome proteolytic pathway. Annu. Rev. Plant Biol. 55:555–590.

Strik, B., J. Mann, and C.E. Finn. 1996. Percent drupelet set varies among blackberry genotypes. J. Amer. Soc. Hortic. Sci. 121:371–373.

Su, L., S. Wan, J. Zhou, Q.S. Shao, and B. Xing. 2021. Transcriptional regulation of plant seed development. Physiol. Plant. 1–13.

Takano-Kai, N., J. Hui, T. Kubo, M. Sweeney, T. Matsumoto, H. Kanamori, B. Padhukasahasram, C. Bustamante, A. Yoshimura, K. Doi, and S. McCouch. 2009. Evolutionary history of *GS3*, a gene conferring grain length in rice. Genetics 182:1323–1334.

Threlfall, R.T., O.S. Hines, J.R. Clark, L.R. Howard, C.R. Brownmiller, D.M. Segantini, and L.J.R. Lawless. 2016. Physiochemical and sensory attributes of fresh blackberries grown in the southeastern United States. HortScience 51:1351–1362.

Tomlik-Wyremblewska, A., J. Zieliński, and M. Guzicka. 2010. Morphology and anatomy of blackberry pyrenes (*Rubus* L., Rosaceae) Elementary studies of the European representatives of the genus *Rubus* L. Flora Morphol. Distrib. Funct. Ecol. Plants 205:370–375.

The UniProt Consortium. 2021. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 49:D1.

Wada, S., and B.M. Reed. 2008. Morphological analysis of *Rubus* seed. Acta Hort. 782:67–74.

Wada, S., H. Nonogaki, and B.M. Reed. 2010. Identifying blackberry cultivars by seed structure. Corvallis, OR, 1–22 pp.

Wada, S., J.A. Kennedy, and B.M. Reed. 2011. Seed-coat anatomy and proanthocyanidins contribute to the dormancy of *Rubus* seed. Sci. Hortic. 130:762–768.

Werner, T., V. Motyka, V. Laucou, R. Smets, H. Van Onckelen, and T. Schmülling. 2003. Cytokinin-deficient transgenic *Arabidopsis* plants show multiple developmental alterations indicating opposite functions of cytokinins in the regulation of shoot and root meristem activity. Plant Cell. 15:2532–2550.

Worthington, M.L., R. Aryal, N.V. Bassil, D. Mead, G.E. Fernandez, J.R. Clark, F. Fernández-Fernández, C.E. Finn, K.E. Hummer, and H. Ashrafi. 2020. Development of new genomic resources and tools for molecular breeding in blackberry. Acta Hort. 1277:39–46.

Xu, C., Y. Liu, Y. Li, X. Xu, C. Xu, X. Li, and J. Xiao, Q. Zhang. 2015. Differential expression of *GS5* regulates grain size in rice. J. Exp. Bot. 66:2611–2623.

Xu, R., H. Yu, J. Wang, P. Duan, B. Zhang, J. Li, Y. Li, J. Xu, J. Lyu, N. Li, T. Chai, and Y. Li. 2018a. A mitogen-activated protein kinase phosphatase influences grain size and weight in rice. Plant J. 95:937–946.

Xu, R., P. Duan, H. Yu, Z. Zhou, B. Zhang, R. Wang, J. Li, G. Zhang, S. Zhuang, J. Lyu, N. Li, T. Chai, Z. Tian, S. Yao, and Y. Li. 2018b. Control of grain size and weight by the OsMKKK10-OsMKK4-OsMAPK6 signaling pathway in rice. Mol. Plant 11:860–873.

Zhang, M., H. Wu, J. Su, H. Wang, Q. Zhu, Y. Liu, J. Xu, W. Lukowitz, and S. Zhang. 2017. Maternal control of embryogenesis by MPK6 and its upstream MKK4/MKK5 in *Arabidopsis*. Plant J. 92:1005–1019.

Tables and Figures

Table 1. Minimum, maximum, and average values for 100-seed weight, width to length (WL) ratio, area, and seeds per berry as normalized BLUPs generated from data collected on blackberries across two harvest years (2019 and 2020).

	Both Years			
	100-seed weight (mg)	WL ratio	Area (mm^2)	Seeds/berry
Maximum	443.5	0.82	6.71	123.1
Minimum	194.7	0.56	3.93	43.6
Mean	318.9	0.69	5.24	80.9
	2019			
	100-seed weight (mg)	WL ratio	Area (mm^2)	Seeds/berry
Maximum	452.7	0.83	6.90	113.0
Minimum	199.2	0.57	3.94	36.5
Mean	325.1	0.70	5.33	81.9
	2020			
	100-seed weight (mg)	WL ratio	Area (mm^2)	Seeds/berry
Maximum	427.6	0.82	6.61	137.5
Minimum	211.1	0.55	4.03	49.6
Mean	312.7	0.69	5.16	80.3

Table 2. Possible candidate genes for width to length (WL) ratio, 100-seed weight, and seed area in blackberry.

QTL in *Rubus*	Gene ID	Chromosome	Start bp	Stop bp	Functional Annotation
WL ratio	Ra_g83	Ra01	400657	405027	Serine carboxypeptidase 24 - BRS1
WL ratio	Ra_g689	Ra01	3429820	3431139	Cytokinin dehydrogenase 2 - CKX2
WL ratio	Ra_g820	Ra01	3976825	3979900	Guanine nucleotide-binding protein γ 3 - GG3
100-seed weight	Ra_g14549	Ra03	37573130	37574461	26S regulatory particle AAA-ATPase 2A - RPT2A
100-seed weight	Ra_g14574	Ra03	37794408	37796107	PP2C protein phosphatase - AP2C1
Area	Ra_g23187	Ra05	22120088	22122719	DA1 Protein

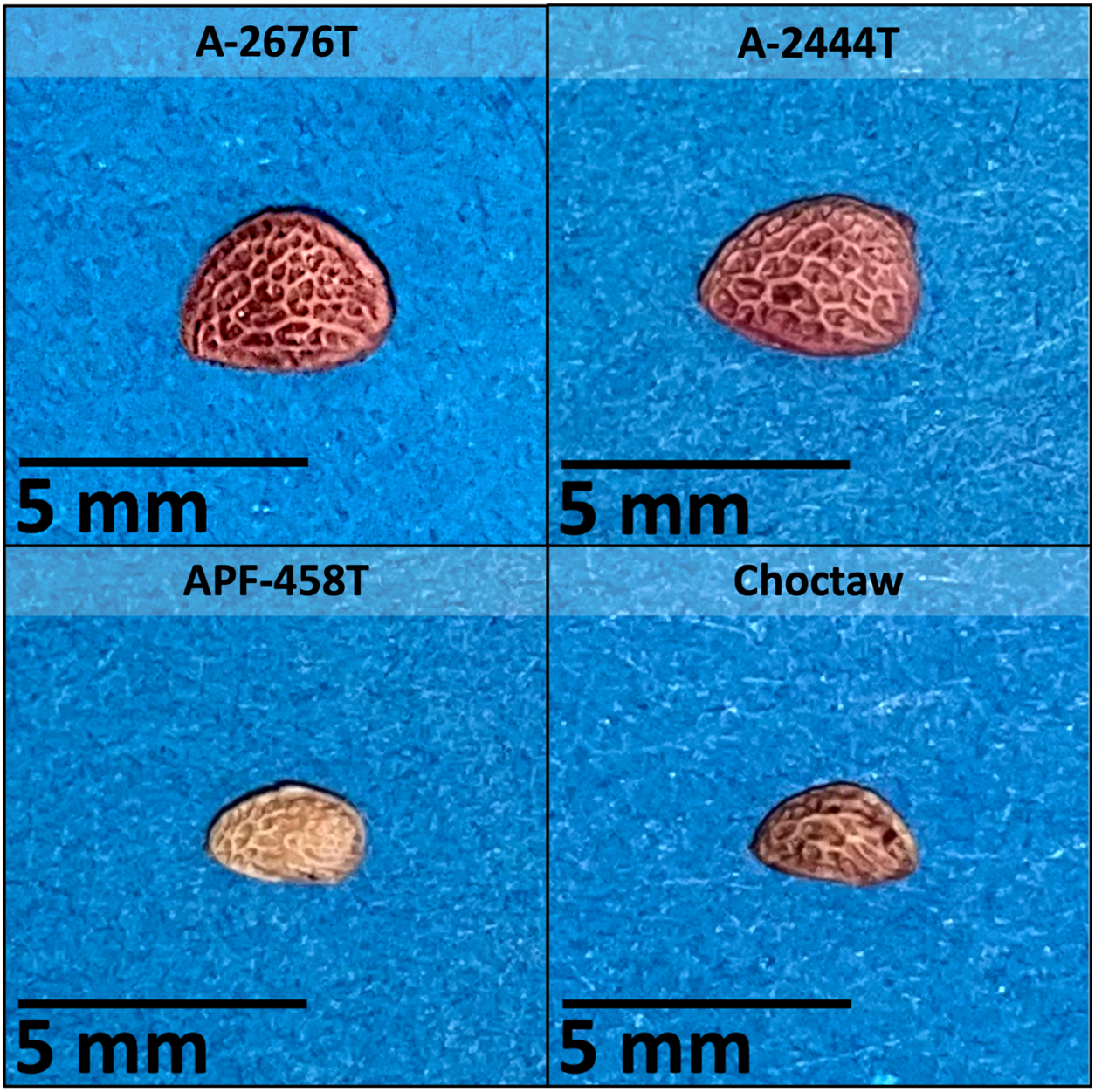


Fig. 1. Visualization of seed size diversity in fresh-market blackberries, showing two of the largest (A-2676T and A-2444T) genotypes and two of the smallest (APF-458T and ‘Choctaw’) 100-seed weight and seed area genotypes included in the study.

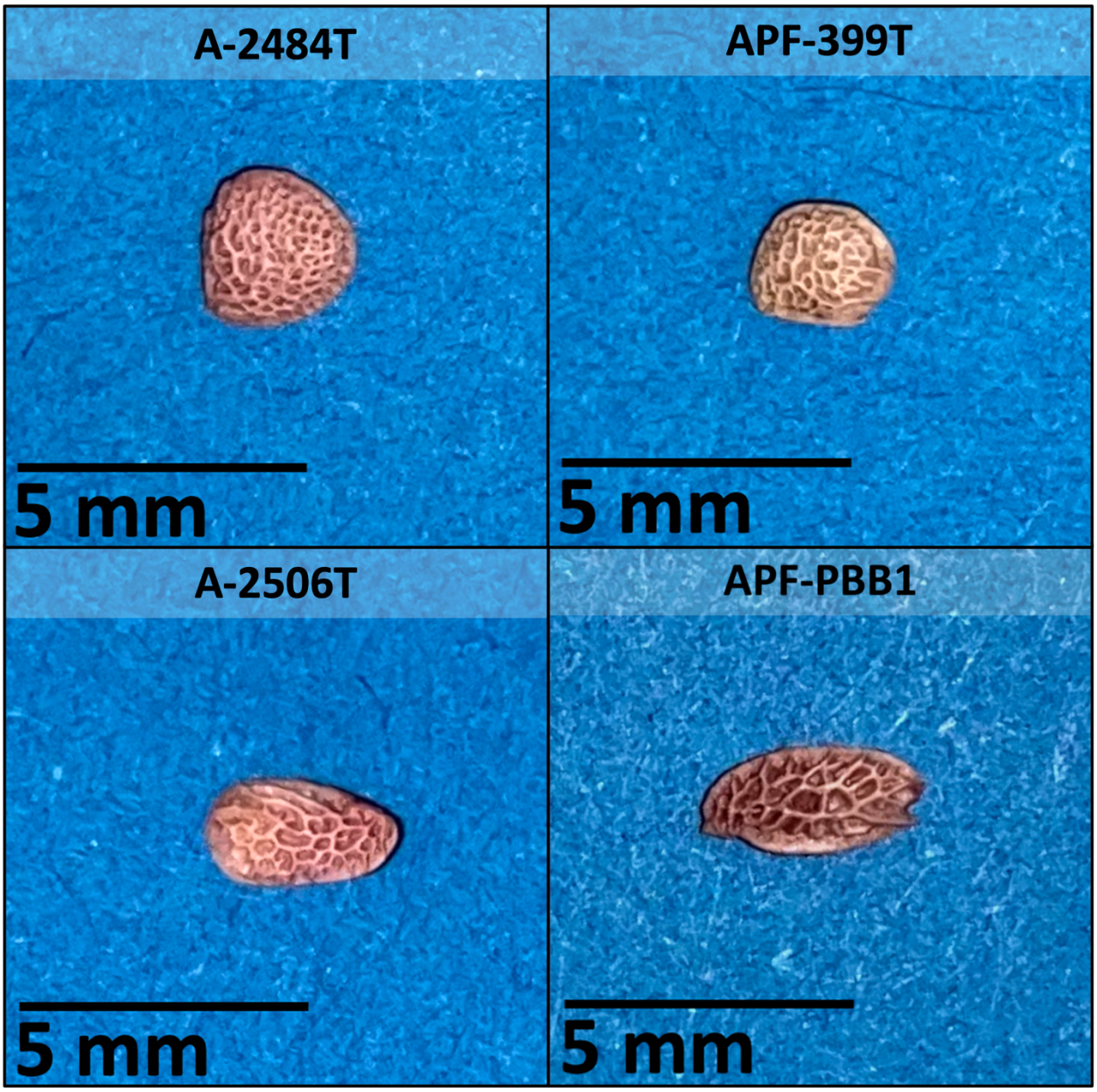


Fig. 2. Visualization of the diversity in seed shape of fresh-market blackberries as determined by width to length (WL) ratio. Shown are genotypes with the highest (A-2484T and APF-399T) and lowest (A-2506 and APF-PBB1) WL ratios.

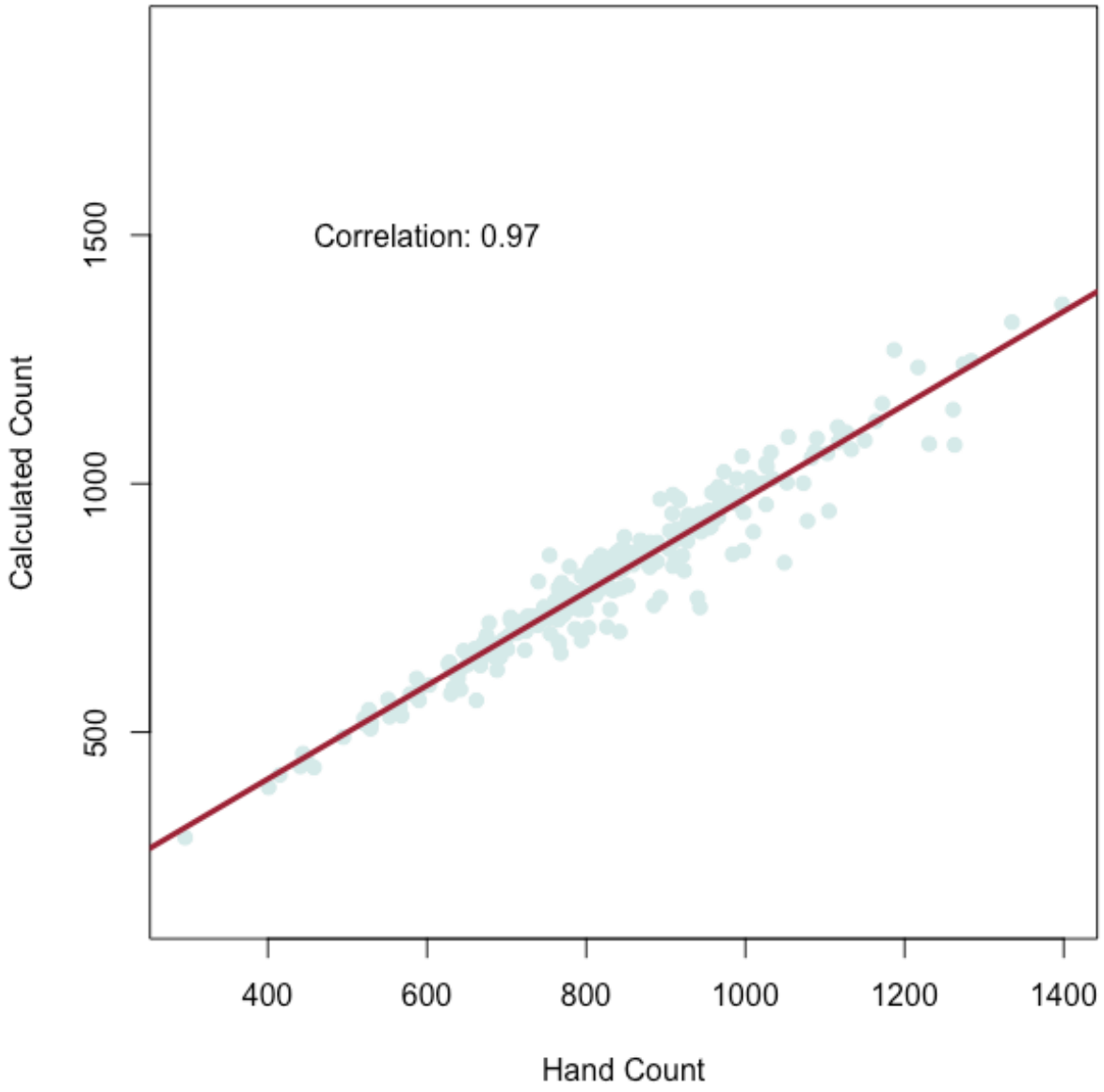


Fig. 3. Pearson correlation coefficient between the hand counted blackberry seed number in a 10-berry sample and the calculated count (10-berry seed weight/100-seed weight) × 100.

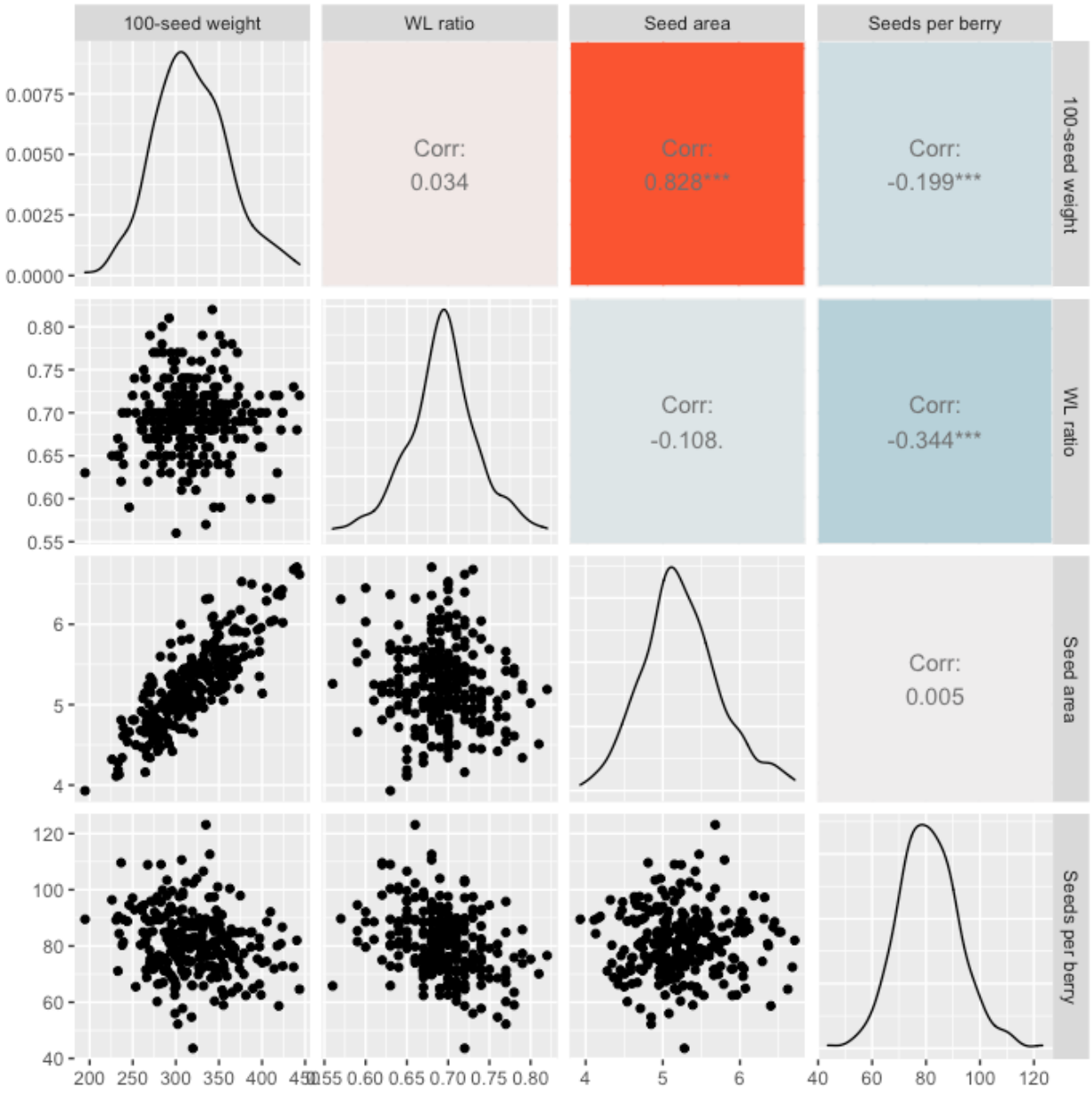


Fig. 4. Distribution and correlation of blackberry 100-seed weight, width to length (WL) ratio, area, and seeds per berry evaluated in 2019 and 2020. Distributions of best linear unbiased predictions for each trait over two years are graphed in the diagonal, scatter plots below the diagonal show relationships between traits, and Pearson's correlation coefficients are indicated above the diagonal.

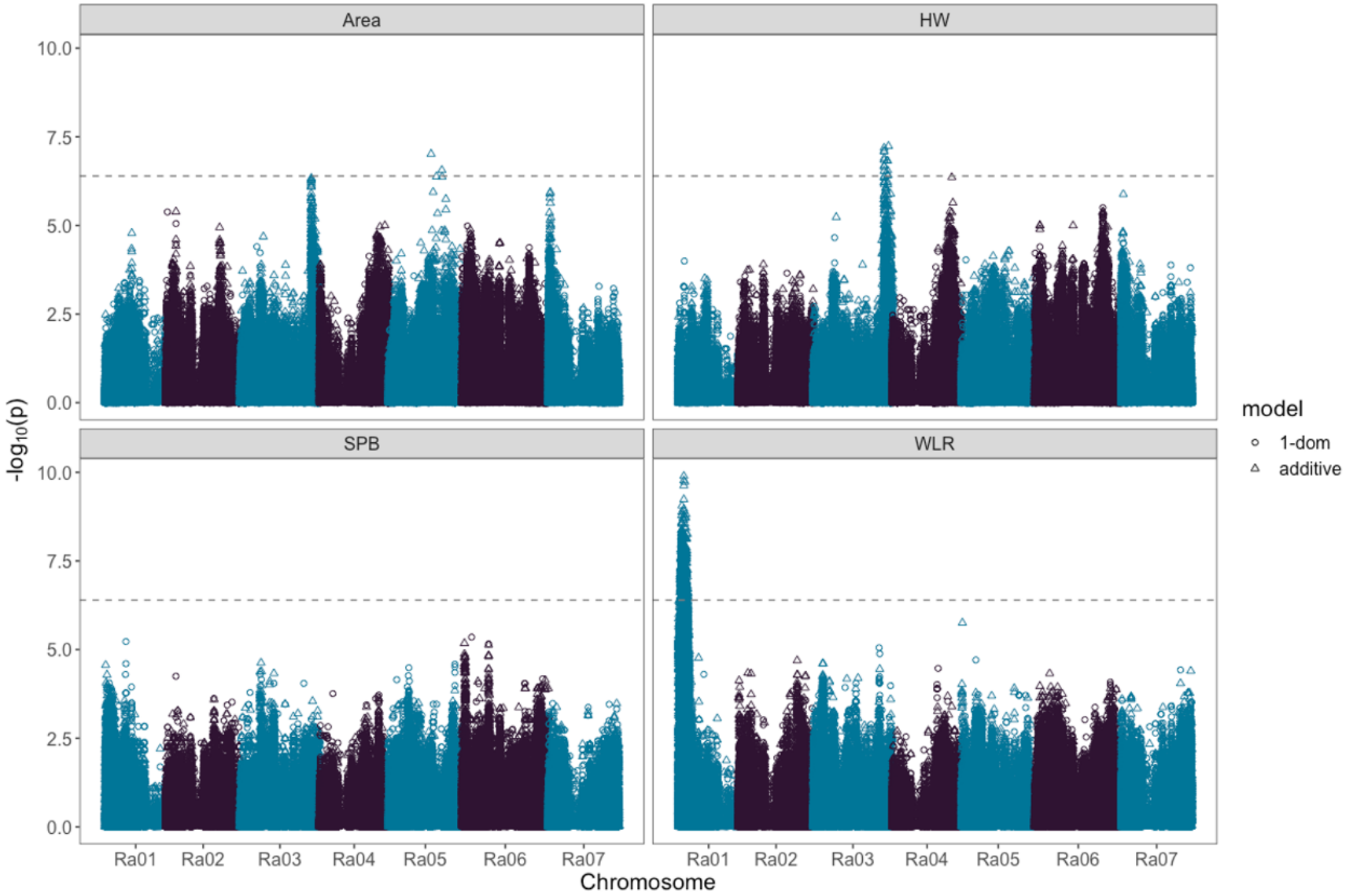


Fig. 5. Manhattan plots showing peaks from GWAS analysis of area, 100-seed weight (HW), seeds per berry (SPB), and width to length ratio (WLR) using data obtained from harvest samples of blackberry collected during 2019 and 2020.

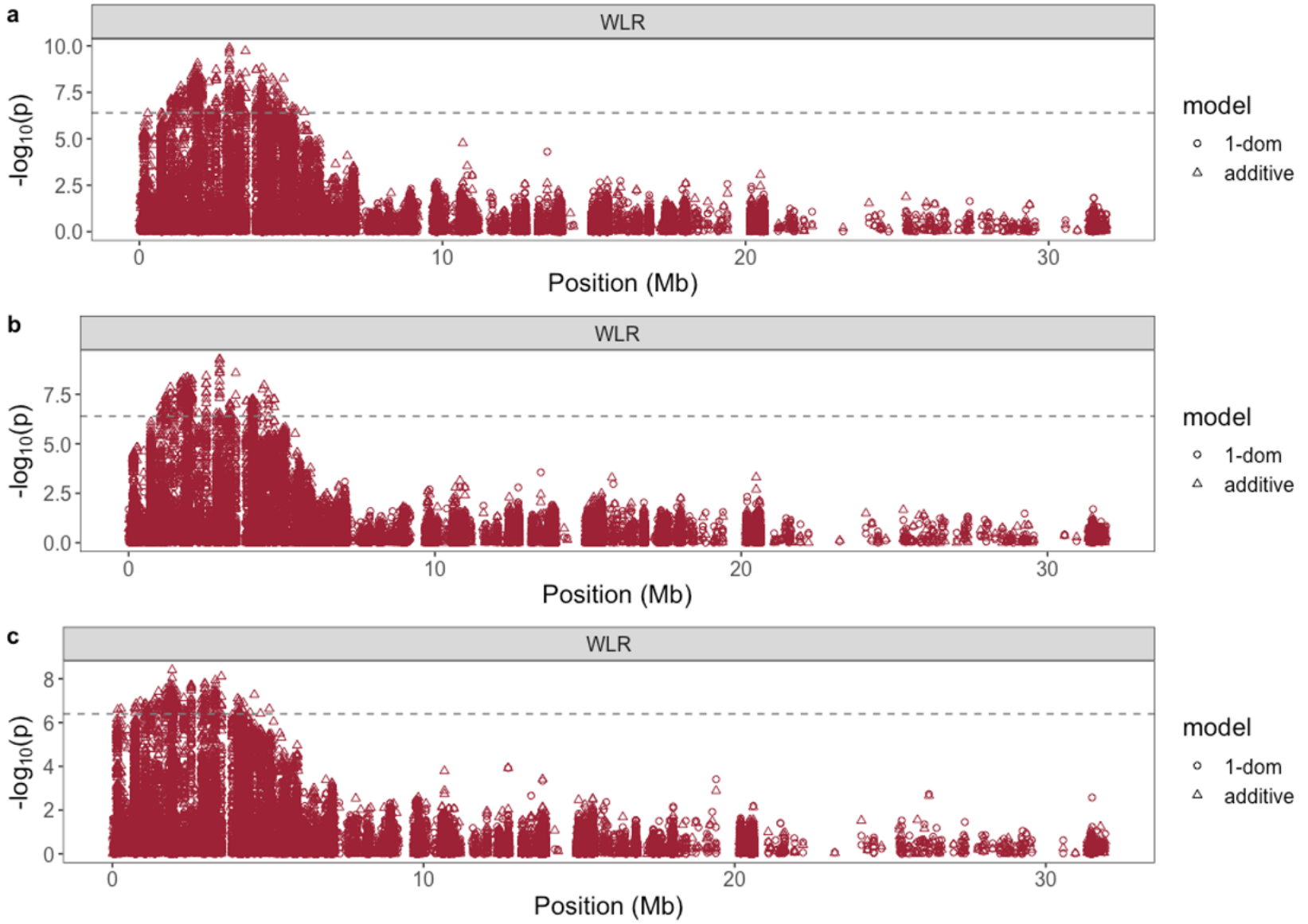


Fig. 6. Manhattan plots focused on chromosome Ra01 showing a consistently mapped peak for blackberry width to length ratio (WLR) across (A) both years, (B) 2019, and (C) 2020.

OVERALL CONCLUSION

While blackberry fresh-market sales have increased in recent years, there are still genetic improvements needed to enhance consumers' experience. Until now, a very limited number of genetic resources have been made available for the improvement of *Rubus* species regarding attributes of sweetness, acidity, and seediness. By implementing genetic studies, genetic markers can be developed related to sweetness, acidity, and seediness. With these genetic tools, blackberry breeders can confidently select parents for crosses and make informed decisions regarding selections earlier in the breeding process, effectively saving time, labor, and money. Increased turnaround time for the release of improved blackberry cultivars also has the potential to expand the customer base of an already rapidly growing market.

A panel of 307 fresh-market blackberry genotypes were phenotyped for soluble solids content (SSC) and pH during the summers of 2019, 2020, and 2021, while a subset of 277 genotypes from the panel were phenotyped for titratable acidity (TA), 100-seed weight, width to length (WL) ratio, area, and seeds per berry during the summers of 2019 and 2020. Genotype data were sequenced and aligned to 35,054 Capture-Seq (RAPiD Genomics, Gainesville, FL) probes of annotated genes distributed across the *Rubus argutus* Link. reference genome. Heritability estimates were calculated for each trait with SSC having the lowest broad sense entry mean heritability at 61%, while pH and TA had heritabilities of 67% and 69%, respectively. For seediness attributes, heritability estimates were 91% for 100-seed weight and WL ratio, 89% for seed area, and 70% for seeds per berry. A Genome-Wide Association Study (GWAS) was conducted using the 'GWASpoly' R package to analyze 124,564 single nucleotide polymorphisms (SNPs) and determine marker-trait associations representing possible quantitative trait loci (QTL) for use in the identification of possible candidate genes.

In these studies, we report the first possible marker-trait associations and QTL associated with sweetness, acidity, and seediness in fresh-market blackberries. A total of six QTL were identified on four of the seven *Rubus* chromosomes; one for WL ratio and one for TA on Ra01, one for SSC on Ra02, one for 100-seed weight on Ra03, and one for seed area and another shared QTL for TA and pH on Ra05. The QTL for TA on Ra01 had a LOD of 6.77. The QTL on Ra02 associated with SSC had a peak LOD of 6.65. The QTL for seed area located on Ra05 had a peak LOD of 7.02. On Ra03, the QTL for 100-seed weight had a peak LOD of 7.24. The overlapping QTL for TA and pH on Ra05 had peak LOD scores of 7.71 for TA and 8.85 for pH. The most significant QTL found in these studies was for WL ratio, which was located on Ra01 with the highest LOD score observed at 9.9. Sixteen possible candidate genes with functions related to acidity, sweetness, or seed development were identified within the QTL regions identified through GWAS analysis. Of these 16 possible candidate genes, 10 were associated with sweetness and acidity, and six were associated with seediness attributes.

The goal of this research was to determine the heritability of sweetness, acidity, and seediness attributes in fresh-market blackberry germplasm and identify genetic loci contributing to useful variation in these traits. Based on the marker-trait associations discovered in this research, we hope to develop molecular markers for use in marker-assisted selection (MAS) during the breeding process. Adding MAS to the breeding process will expedite the release of cultivars with desired traits for fresh-market sales. Future research can be conducted to continue fine-mapping and validating the genetic control of these traits within *Rubus* and to use these data as a training population for genomic selection (GS).

APPENDICES

Appendix A

Supplementary Table 1. Number of replicates of each blackberry genotype harvested at the University of Arkansas System Fruit Research Station in Clarksville, AR during 2019, 2020, and 2021. Normalized best linear unbiased predictions (BLUPs) for soluble solids content (SSC) and pH were calculated from data obtained in all three years, while titratable acidity (TA) BLUPs were calculated from samples collected in 2019 and 2020.

Genotype	2019 replicates	2020 replicates	2021 replicates	Total replicates	SSC (%)	pH	TA (% citric)
Comanche	1	1	2	4	8.9	3.59	0.92
Cherokee	2	1	2	5	10.0	3.81	0.83
Cheyenne	1	2	---	3	10.0	3.89	0.72
Shawnee	2	2	2	6	9.6	3.60	0.98
Choctaw	2	1	2	5	9.3	3.77	0.90
Navaho	1	2	2	5	11.2	3.55	0.83
Kiowa	2	2	2	6	9.3	3.48	1.08
Arapaho	2	2	---	4	10.7	3.61	0.82
Chickasaw	2	---	---	2	10.2	3.59	0.99
A-1790	2	2	2	6	11.5	3.73	0.86
Apache	2	1	2	5	10.5	3.58	1.02
Ouachita	2	2	2	6	11.7	3.55	0.87
A-1960T	2	1	2	5	11.0	3.61	0.82
Natchez	2	2	2	6	10.0	3.45	0.94
A-2252T	2	2	2	6	10.5	3.86	0.59
Stella	2	2	2	6	11.5	3.76	0.61
A-2316T	2	2	2	6	10.2	3.52	1.24
Osage	2	2	2	6	11.2	3.83	0.63
A-2416T	2	2	2	6	10.1	3.43	1.25
A-2418T	2	2	2	6	10.3	3.44	0.99

Genotype	2019 replicates	2020 replicates	2021 replicates	Total replicates	SSC (%)	pH	TA (% citric)
A-2421	2	1	2	5	11.5	3.59	0.80
Sweet-Ark™ Caddo	2	2	2	6	11.2	3.59	0.78
A-2435T	1	---	---	1	11.6	3.70	0.67[z]
A-2444T	2	2	2	6	11.8	3.55	0.82
A-2453T	2	2	2	6	11.1	3.82	0.60
A-2454T	2	2	2	6	11.5	3.94	0.53
A-2473T	2	2	2	6	10.3	3.47	0.75
A-2484T	2	2	2	6	10.5	3.75	0.81
A-2491T	2	2	2	6	12.3	3.66	0.64
A-2496	2	2	2	6	12.2	3.77	0.79
A-2500TN	2	2	2	6	10.7	3.55	0.85
A-2506T	2	2	2	6	10.3	3.56	0.88
A-2518T	2	---	---	2	10.6	3.65	0.88
A-2520T	2	2	2	6	12.2	3.60	0.73
A-2522T	2	2	2	6	10.1	3.66	0.77
A-2524T	2	2	2	6	10.5	3.67	0.73
A-2525TN	2	2	2	6	8.9	3.96	0.52
A-2526T	2	2	4	8	11.0	3.63	0.69
A-2527T	2	2	---	4	11.2	3.60	0.89
A-2528T	2	2	2	6	10.5	3.51	0.90
Sweet-Ark™ Ponca	2	2	2	6	12.2	3.76	0.67
A-2541T	2	2	2	6	11.5	3.82	0.60
A-2545	2	---	---	2	11.0	3.87	0.55
A-2547T	1	2	4	7	10.3	3.67	0.82
A-2551TN	1	2	---	3	9.7	3.55	0.76
A-2560T	2	---	---	2	11.4	3.56	0.84
A-2563TN	2	2	---	4	10.6	3.58	0.70

Genotype	2019 replicates	2020 replicates	2021 replicates	Total replicates	SSC (%)	pH	TA (% citric)
A-2568T	2	---	---	2	9.9	3.49	1.11
A-2571T	2	2	2	6	10.9	3.74	0.81
A-2572T	2	2	2	6	10.9	3.62	0.89
A-2575T	2	2	---	4	10.1	3.80	0.69
A-2578T	2	2	---	4	11.5	3.56	0.83
A-2580T	2	1	2	5	10.5	3.52	0.94
A-2583T	2	---	---	2	11.2	3.65	0.74
A-2587T	2	2	2	6	9.9	3.54	0.79
A-2596T	2	2	2	6	12.0	3.84	0.63
A-2597T	2	---	---	2	10.3	3.91	0.63
A-2598T	2	---	---	2	11.9	3.55	1.00
A-2601T	2	---	---	2	11.3	3.57	0.79
A-2602T	2	1	2	5	12.4	3.69	0.84
A-2604T	2	1	1	4	11.4	3.67	0.73
A-2605T	2	---	---	2	11.0	3.67	0.90
A-2606T	2	2	2	6	12.1	3.77	0.70
A-2609T	1	1	---	2	12.6	3.65	0.79
A-2610T	2	2	2	6	11.2	3.56	0.98
A-2611T	1	---	---	1	10.7	3.66	0.80
A-2613T	2	1	1	4	11.0	3.84	0.82
A-2614T	1	1	2	4	10.8	3.63	0.77
A-2615T	1	1	2	4	10.0	3.41	0.98
A-2616T	2	1	2	5	11.4	3.56	0.94
A-2617T	2	2	2	6	11.3	3.61	0.78
A-2620T	2	2	2	6	10.4	3.39	0.97
A-2624T	2	2	2	6	11.9	3.64	0.89
A-2625T	2	1	2	5	11.6	3.64	0.92

Genotype	2019 replicates	2020 replicates	2021 replicates	Total replicates	SSC (%)	pH	TA (% citric)
A-2628T	2	2	2	6	11.7	3.58	0.92
A-2636T	2	1	2	5	12.2	3.81	0.64
A-2637T	2	2	2	6	12.6	3.69	0.76
A-2639T	2	---	---	2	11.1	3.74	0.76
A-2642T	2	1	2	5	10.6	3.82	0.68
A-2644T	2	2	---	4	10.0	3.83	0.65
A-2651T	2	---	---	2	11.5	3.66	0.83
A-2652T	2	2	2	6	11.7	3.83	0.71
A-2656T	2	2	2	6	11.2	3.68	0.77
A-2657T	2	---	---	2	10.8	3.78	0.66
A-2658T	2	2	2	6	11.3	3.59	0.80
A-2660T	2	---	2	4	10.9	3.65	0.77
A-2663T	2	2	2	6	10.7	3.64	0.77
A-2665T	2	2	---	4	10.0	3.65	0.79
A-2668T	2	2	---	4	10.9	4.00	0.58
A-2669T	2	2	---	4	12.7	3.72	0.74
A-2672T	2	2	2	6	10.6	3.73	0.69
A-2673T	2	2	2	6	11.2	3.53	0.91
A-2676T	2	---	---	2	10.8	3.65	0.93
A-2677T	2	2	2	6	12.6	3.96	0.66
A-2678T	---	1	1	2	11.5	3.98	0.67
A-2680T	2	2	---	4	10.0	3.80	0.64
A-2683T	2	---	---	2	11.5	3.54	0.95
A-2684T	2	2	2	6	11.5	3.73	0.77
A-2685T	2	2	2	6	11.9	3.75	0.80
A-2686T	2	2	2	6	11.0	3.69	0.76
A-2687T	2	2	2	6	10.1	3.76	0.67

Genotype	2019 replicates	2020 replicates	2021 replicates	Total replicates	SSC (%)	pH	TA (% citric)
A-2688T	2	---	---	2	11.0	3.64	0.98
A-2700T	2	2	2	6	11.0	3.70	0.70
A-2701T	2	2	2	6	10.4	3.89	0.66
A-2708T	2	2	2	6	12.3	3.79	0.85
A-2709T	2	2	2	6	11.6	3.58	0.82
A-2710T	2	2	2	6	12.7	3.67	0.81
A-2711T	2	2	---	4	12.2	3.86	0.67
A-2713T	2	2	2	6	10.5	3.54	0.87
A-2716T	2	2	2	6	10.9	3.66	0.76
A-2717T	2	2	2	6	12.2	3.85	0.74
A-2718T	2	2	2	6	10.5	3.53	0.96
A-2720T	2	2	---	4	11.0	3.63	0.76
A-2722T	2	2	2	6	10.8	3.62	0.79
A-2723T	2	2	2	6	11.2	3.80	0.60
A-2724T	2	2	2	6	11.6	3.57	0.79
A-2725T	2	2	---	4	12.1	3.70	0.84
A-2726TN	2	---	---	2	10.8	3.77	0.65
A-2727T	1	---	---	1	9.9	3.68	0.79
A-2728T	2	---	---	2	10.5	3.56	0.75
A-2730T	2	2	---	4	11.3	3.71	0.96
A-2731T	2	2	---	4	11.6	3.52	0.98
A-2733T	2	1	2	5	11.3	3.58	0.85
A-2734T	2	---	---	2	11.3	3.72	0.74
A-2735T	2	---	---	2	11.6	3.68	0.85
A-2736T	2	---	---	2	10.9	3.71	0.73
A-2738T	2	2	---	4	9.9	3.62	0.77
A-2739T	2	2	---	4	10.0	3.58	0.87

Genotype	2019 replicates	2020 replicates	2021 replicates	Total replicates	SSC (%)	pH	TA (% citric)
A-2740T	2	2	2	6	10.7	3.69	0.72
A-2742	2	---	---	2	11.4	3.71	0.72
A-2743T	2	2	2	6	11.4	3.72	0.71
A-2745T	2	2	---	4	10.6	3.55	0.91
A-2746T	2	2	2	6	10.5	3.53	0.85
A-2747T	2	2	---	4	9.6	3.48	0.97
A-2749T	1	2	2	5	10.5	3.64	0.68
A-2751T	1	1	2	4	11.9	3.82	0.65
A-2752T	2	2	2	6	8.1	3.52	0.80
A-2753T	2	2	---	4	10.8	3.70	0.87
A-2755T	2	2	---	4	11.4	3.66	0.77
A-2756T	2	2	2	6	10.8	3.65	0.72
A-2757T	2	2	2	6	11.0	3.87	0.57
A-2758T	---	1	---	1	11.0	3.50	1.05
A-2759T	2	2	2	6	11.7	3.51	0.83
A-2760T	2	2	---	4	9.9	3.55	1.14
A-2761T	2	2	---	4	10.3	3.68	0.86
A-2762T	2	2	---	4	11.3	3.66	0.76
A-2763T	2	2	---	4	10.7	3.97	0.53
A-2764T	2	---	---	2	10.7	3.66	0.92
A-2765T	2	2	2	6	11.4	3.86	0.73
A-2766T	2	2	---	4	10.4	3.83	0.62
A-2767T	2	2	2	6	12.3	3.77	0.76
A-2768T	2	2	2	6	10.7	3.80	0.61
A-2770T	1	2	---	3	10.7	3.84	0.66
A-2771T	2	2	2	6	10.5	3.70	0.79
A-2772T	2	2	---	4	12.4	3.76	0.78

Genotype	2019 replicates	2020 replicates	2021 replicates	Total replicates	SSC (%)	pH	TA (% citric)
A-2773T	2	2	2	6	8.6	3.90	0.61
A-2774T	2	2	2	6	12.6	3.95	0.63
A-2775T	2	---	---	2	10.6	3.57	0.91
A-2777T	2	1	---	3	9.0	3.70	0.78
A-2778T	2	2	2	6	11.1	3.80	0.86
A-2779T	2	2	2	6	11.0	3.55	1.08
A-2780T	2	1	2	5	10.3	3.57	0.93
A-2781T	---	2	2	4	11.7	3.50	0.99
A-2782T	---	2	2	4	11.4	3.67	0.78
A-2783T	---	1	2	3	10.9	3.72	0.70
A-2784T	---	2	2	4	10.6	3.49	0.83
A-2785T	---	2	2	4	11.4	3.79	0.64
A-2786T	---	2	---	2	11.1	3.71	0.71
A-2787T	---	2	2	4	10.9	3.63	0.68
A-2788T	---	2	2	4	11.5	3.63	0.72
A-2789TN	---	2	2	4	10.9	3.79	0.60
A-2790TN	---	2	2	4	10.8	3.70	0.65
A-2791TN	---	2	2	4	10.8	3.85	0.56
A-2792TN	---	2	2	4	11.6	3.72	0.56
A-2793TN	---	2	2	4	10.6	3.56	0.71
A-2794TN	---	2	2	4	10.5	3.88	0.63
A-2795TN	---	2	2	4	10.2	3.67	0.69
A-2796TN	---	2	2	4	11.9	3.55	0.76
A-2797T	---	1	2	3	10.7	3.54	0.86
A-2798T	---	1	2	3	9.2	3.48	1.13
A-2799T	---	2	1	3	11.4	3.48	1.05
A-2801T	---	2	2	4	10.0	3.57	0.85

Genotype	2019 replicates	2020 replicates	2021 replicates	Total replicates	SSC (%)	pH	TA (% citric)
A-2802T	---	2	2	4	10.2	3.58	0.88
A-2803T	---	---	2	2	11.6	3.63	---
A-2804T	---	---	2	2	12.4	3.61	---
A-2805T	---	---	2	2	10.2	3.57	---
A-2806T	---	---	2	2	13.9	3.84	---
A-2807T	---	---	2	2	12.3	3.74	---
A-2808T	---	---	2	2	10.4	3.75	---
A-2809T	---	---	2	2	11.2	3.62	---
A-2810T	---	---	2	2	11.0	3.84	---
A-2811T	---	---	2	2	10.8	3.83	---
A-2813TN	---	---	2	2	11.9	3.59	---
A-2814TN	---	---	2	2	11.1	3.55	---
A-2815T	---	---	2	2	10.6	3.74	---
A-2816T	---	---	2	2	11.5	3.75	---
A-2817T	---	---	2	2	10.6	3.63	---
A-2818T	---	---	2	2	11.2	3.77	---
A-2819T	---	---	2	2	10.7	3.71	---
A-2820T	---	---	2	2	12.6	3.71	---
A-2823T	---	---	2	2	10.9	3.56	---
A-2824T	---	---	2	2	11.2	3.63	---
Prime Jim	1	1	---	2	10.1	3.61	0.99
Madeline	2	2	2	6	11.2	3.87	0.65
Sharon's Delight	2	2	2	6	9.7	3.67	0.72
Prime-Ark® Freedom	2	1	2	5	10.3	3.75	0.57
Prime-Ark® Traveler	2	2	2	6	11.1	3.78	0.57
Black Gem	2	2	1	5	11.2	3.59	0.68
Baby Cakes	2	2	2	6	8.9	3.64	0.92

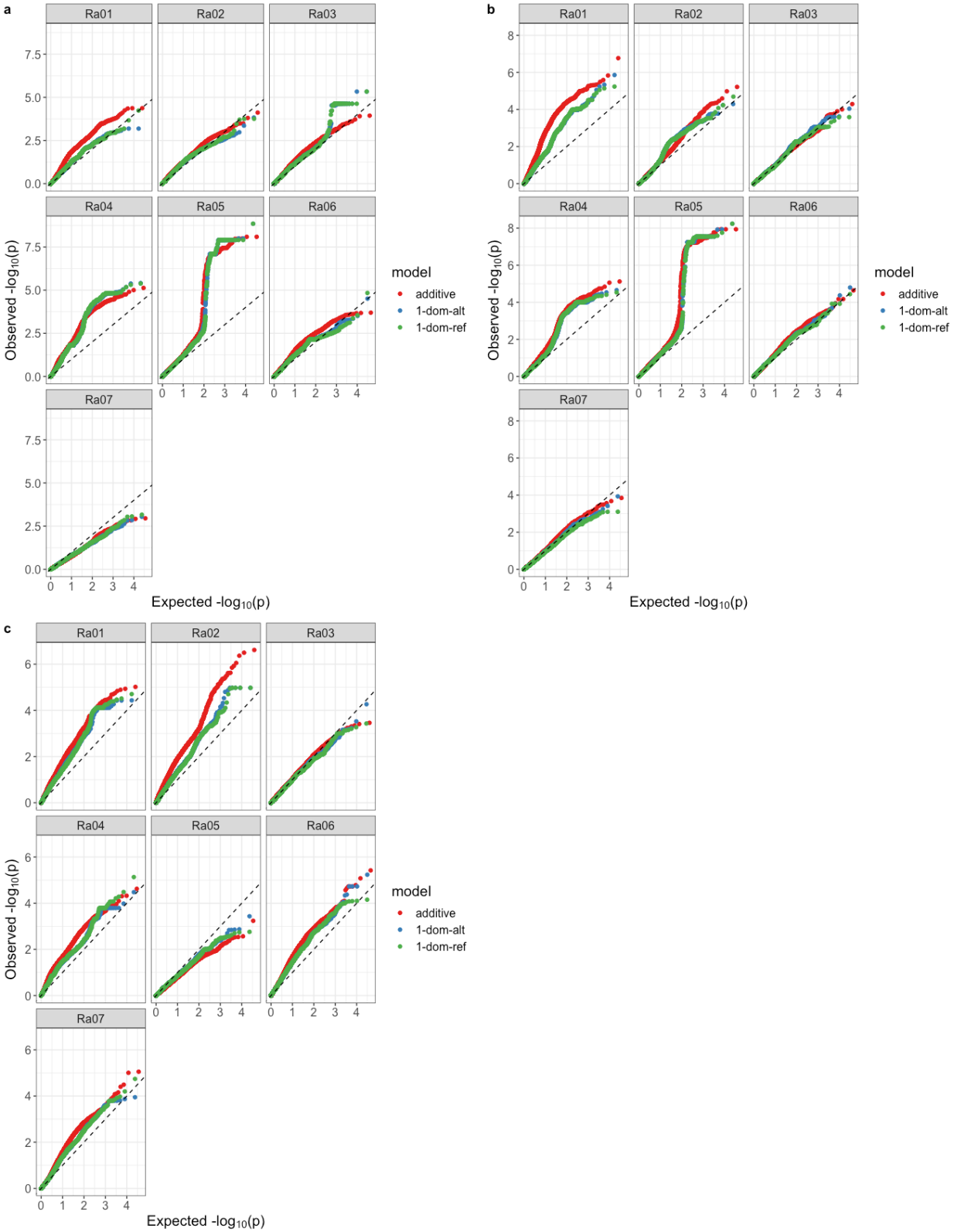
Genotype	2019 replicates	2020 replicates	2021 replicates	Total replicates	SSC (%)	pH	TA (% citric)
APF-238T	2	---	2	4	12.3	3.62	0.74
APF-259TN	2	2	---	4	9.5	3.60	0.80
Prime-Ark® Horizon	2	1	2	5	10.6	3.51	0.88
APF-276TN	2	2	2	6	10.8	3.46	0.99
APF-298TN	2	2	2	6	10.4	3.46	0.99
APF-318	2	---	2	4	12.9	3.64	0.87
APF-328	2	2	---	4	10.0	3.72	0.71
APF-334T	2	2	2	6	11.6	3.73	0.65
APF-335T	2	2	2	6	9.8	3.56	0.98
APF-338	2	---	---	2	12.2	3.62	0.91
APF-341TN	2	2	---	4	10.8	3.63	0.84
APF-345T	2	---	---	2	10.5	3.58	0.93
APF-355TN	2	2	2	6	12.8	3.76	0.66
APF-366T	2	2	2	6	10.3	3.63	0.88
APF-370T	2	2	2	6	11.1	3.77	0.59
APF-372T	1	1	---	2	10.7	3.61	0.94
APF-373T	2	2	---	4	11.2	3.60	0.79
APF-377TN	2	2	---	4	10.7	3.61	0.87
APF-379TN	1	2	---	3	11.0	3.83	0.56
APF-386TN	2	1	2	5	11.4	3.58	0.88
APF-388T	2	---	---	2	10.3	3.79	0.68
APF-389TN	2	2	2	6	10.7	3.47	0.88
APF-392TN	2	2	---	4	9.6	3.60	0.88
APF-394T	2	---	---	2	9.8	3.48	1.03
APF-399T	2	---	---	2	9.3	3.57	0.93
APF-400T	1	2	2	5	10.7	3.56	0.82
APF-404T	2	2	2	6	11.2	3.70	0.75

Genotype	2019 replicates	2020 replicates	2021 replicates	Total replicates	SSC (%)	pH	TA (% citric)
APF-405T	2	2	---	4	11.0	3.65	0.82
APF-406TN	2	2	1	5	11.7	3.54	1.01
APF-407TN	2	1	---	3	11.6	3.86	0.73
APF-409T	2	2	2	6	11.0	3.50	0.79
APF-410T	2	2	---	4	11.9	3.61	0.86
APF-414TN	2	2	2	6	10.2	3.67	0.81
APF-415T	1	1	2	4	9.8	3.54	0.99
APF-423TN	2	---	---	2	10.8	3.76	0.74
APF-424TN	2	2	---	4	9.9	3.66	0.91
APF-425T	1	2	2	5	11.1	3.73	0.68
APF-426T	2	2	2	6	11.7	3.66	0.69
APF-427T	---	1	---	1	10.8	3.52	1.12
APF-428T	2	2	2	6	12.3	3.79	0.66
APF-432T	---	2	1	3	10.3	3.48	1.08
APF-434T	2	---	---	2	10.7	3.57	0.84
APF-435T	---	2	---	2	11.5	3.59	0.91
APF-437T	2	2	---	4	11.2	3.49	1.03
APF-439TN	---	---	2	2	11.5	3.81	---
APF-44	1	2	2	5	9.3	3.69	0.68
APF-440	---	1	2	3	10.7	3.68	0.75
APF-441	---	1	2	3	11.6	3.69	0.77
APF-443T	1	1	---	2	10.2	3.47	1.07
APF-444TN	2	2	2	6	9.9	3.66	0.85
APF-447T	2	2	---	4	10.5	4.05	0.54
APF-448T	2	2	2	6	11.3	3.71	0.77
APF-449TN	2	2	2	6	10.1	3.64	0.61
Prime-Ark® 45	2	2	2	6	10.7	3.69	0.66

Genotype	2019 replicates	2020 replicates	2021 replicates	Total replicates	SSC (%)	pH	TA (% citric)
APF-450TN	2	2	2	6	9.7	3.75	0.72
APF-451T	1	2	---	3	10.9	3.62	0.72
APF-452T	2	2	---	4	11.0	3.54	0.83
APF-457TN	2	2	2	6	11.0	3.69	0.75
APF-458T	1	---	2	3	9.5	3.60	1.14
APF-459T	---	2	2	4	9.8	3.67	0.77
APF-460	---	---	2	2	10.5	3.67	---
APF-464TN	---	1	---	1	10.0	3.48	0.94
APF-465TN	2	2	---	4	9.8	3.48	0.94
APF-466T	2	---	---	2	10.5	3.64	0.90
APF-467T	2	2	---	4	10.3	3.62	1.02
APF-468T	2	1	2	5	10.7	3.71	0.83
APF-469T	1	2	2	5	11.0	3.48	1.00
APF-470T	1	1	2	4	12.2	3.69	0.83
APF-471T	1	2	2	5	10.9	3.80	0.77
APF-472T	2	1	1	4	11.8	3.49	1.02
APF-473T	1	---	---	1	10.9	3.61	0.98
APF-474T	1	2	---	3	10.3	3.64	0.88
APF-475T	2	1	---	3	10.4	3.46	1.02
APF-477T	---	2	2	4	8.5	3.58	0.86
APF-478TN	---	1	2	3	10.9	3.68	0.63
APF-479TN	---	2	2	4	11.3	3.59	0.90
APF-480TN	---	2	2	4	10.7	3.56	0.86
APF-481TN	---	2	2	4	9.7	3.86	0.72
APF-482TN	---	1	---	1	10.6	3.62	0.73
APF-483TN	---	1	---	1	11.3	3.67	0.68
APF-484TN	---	2	2	4	10.8	3.50	0.85

Genotype	2019 replicates	2020 replicates	2021 replicates	Total replicates	SSC (%)	pH	TA (% citric)
APF-485T	---	1	1	2	11.0	3.72	0.89
APF-487TN	---	---	2	2	10.4	3.63	---
APF-488T	---	---	2	2	10.1	3.71	---
APF-491TN	---	---	2	2	10.5	3.66	---
APF-492T	---	---	2	2	11.3	3.74	---
APF-493	---	---	2	2	10.5	3.59	---
APF-494	---	---	2	2	12.2	3.71	---
APF-495T	---	---	1	1	10.9	3.71	---
APF-496T	---	---	2	2	11.1	3.64	---
APF-497T	---	---	1	1	10.8	3.68	---
APF-77	2	2	2	6	11.3	3.45	0.94
APF-PBB1	2	1	---	3	9.1	3.63	1.06
Eclipse	---	2	---	2	11.1	3.52	0.89
Galaxy	1	1	---	2	10.3	3.51	0.93
PBB-APF-1801	---	1	---	1	9.4	3.53	0.99
Tupy	1	2	2	5	10.3	3.64	0.79
Von	1	2	---	3	10.3	3.69	0.69
Sample total	440	425	403	1268			
Genotype Total	236	237	205	307			

[z]Genotypes A-2435T, A-2727T, A-2758T, A-2779T, and APF-473T were excluded from association analyses because they failed to pass DNA quality filtering thresholds.



Supplementary Fig. 1. QQ-plots for (a) pH, (b) titratable acidity (TA), and (c) soluble solids content (SSC) showing the observed versus expected distribution of p-values for markers on each *R. argutus* chromosome.

Appendix B

Supplemental Table 1 - Replicates of each blackberry genotype harvested in 2019, 2020, and both years alongside normalized BLUPs for seediness attributes of 100-seed weight, width to length (WL) ratio, area, and seeds per berry obtained from both harvest years.

Genotype	2019 replicates	2020 replicates	Total replicates	100-Seed weight (mg)	WL ratio	Area (mm^2)	Seeds per berry
Comanche	2	1	3	273.8	0.70	4.74	71.7
Cherokee	2	2	4	266.2	0.67	4.36	68.8
Cheyenne	1	2	3	232.8	0.67	4.28	71.1
Shawnee	2	2	4	317.2	0.63	4.89	80.5
Choctaw	2	1	3	194.7	0.63	3.93	89.5
Navaho	2	2	4	296.0	0.76	4.42	68.6
Kiowa	2	2	4	339.9	0.68	5.38	88.6
Arapaho	2	2	4	265.3	0.74	4.45	71.5
Chickasaw	2	---	2	318.4	0.64	5.58	88.7
A-1790	2	2	4	398.8	0.70	5.96	62.6
Apache	2	2	4	368.5	0.71	5.95	84.7
Ouachita	2	2	4	308.8	0.71	4.95	75.0
A-1960T	2	2	4	313.9	0.62	5.15	75.0
Natchez	2	2	4	339.4	0.68	5.47	112.6
A-2252T	2	2	4	297.6	0.70	5.12	68.2
Stella	2	2	4	362.6	0.63	5.68	100.4
A-2316T	2	2	4	306.6	0.61	5.05	88.6
Osage	2	2	4	334.9	0.70	5.50	76.2
A-2416T	1	2	3	294.3	0.70	5.06	---
A-2418T	2	2	4	405.6	0.60	6.45	88.6
A-2421	2	1	3	305.8	0.73	4.91	77.1
Sweet-Ark™ Caddo	2	2	4	368.5	0.69	5.70	79.2
A-2435T	1	---	1	243.8	0.70	4.54	88.8[z]

Genotype	2019 replicates	2020 replicates	Total replicates	100-Seed weight (mg)	WL ratio	Area (mm^2)	Seeds per berry
A-2444T	2	2	4	443.5	0.72	6.62	64.6
A-2453T	2	2	4	318.3	0.76	4.85	54.7
A-2454T	2	2	4	325.6	0.74	5.38	80.1
A-2473T	2	2	4	334.9	0.66	5.68	123.1
A-2484T	2	2	4	342.3	0.82	5.19	76.6
A-2491T	2	2	4	335.4	0.74	5.37	79.1
A-2496	2	2	4	397.1	0.66	5.66	79.8
A-2500TN	2	2	4	245.9	0.59	4.66	94.6
A-2506T	2	2	4	308.0	0.67	5.23	99.2
A-2518T	2	---	2	397.1	0.70	5.35	82.8
A-2520T	2	2	4	364.5	0.65	5.68	80.5
A-2522T	2	2	4	358.2	0.71	5.23	83.5
A-2524T	2	2	4	346.8	0.65	5.81	92.7
A-2525TN	2	2	4	400.3	0.66	5.14	74.0
A-2526T	2	2	4	374.0	0.69	5.91	71.9
A-2527T	2	2	4	344.6	0.73	5.62	97.5
A-2528T	2	2	4	359.7	0.64	5.68	80.5
Sweet-Ark™ Ponca	2	2	4	309.9	0.70	5.18	81.2
A-2541T	2	2	4	335.0	0.69	5.52	78.1
A-2545	2	---	2	279.5	0.67	4.96	---
A-2547T	2	2	4	264.5	0.72	4.16	89.9
A-2551TN	1	2	3	308.0	0.62	5.19	98.1
A-2560T	2	---	2	334.8	0.57	6.31	89.7
A-2563TN	2	2	4	328.2	0.69	5.16	72.8
A-2568T	2	---	2	293.8	0.69	5.00	94.1
A-2571T	2	2	4	329.1	0.76	5.05	80.6
A-2572T	2	2	4	351.9	0.59	5.53	85.4

Genotype	2019 replicates	2020 replicates	Total replicates	100-Seed weight (mg)	WL ratio	Area (mm^2)	Seeds per berry
	2	2	4	319.1	0.67	5.55	92.3
A-2578T	2	2	4	414.0	0.72	6.05	71.9
A-2580T	1	1	2	374.8	0.69	6.18	97.9
A-2583T	2	---	2	316.1	0.74	5.82	82.0
A-2587T	2	2	4	307.8	0.73	5.42	93.1
A-2596T	2	2	4	311.4	0.72	5.35	102.7
A-2597T	2	---	2	330.8	0.79	5.19	85.8
A-2598T	2	---	2	331.2	0.68	5.54	78.4
A-2601T	2	---	2	307.0	0.74	5.11	---
A-2602T	2	1	3	270.1	0.79	4.34	73.8
A-2604T	2	1	3	343.5	0.70	6.10	86.4
A-2605T	2	---	2	323.2	0.61	5.25	82.3
A-2606T	2	2	4	252.1	0.74	4.81	92.9
A-2609T	1	2	3	323.2	0.64	5.46	87.0
A-2610T	2	2	4	332.3	0.70	5.61	87.1
A-2611T	1	---	1	344.8	0.72	5.99	85.3
A-2613T	2	2	4	310.1	0.72	4.95	69.3
A-2614T	2	2	4	274.0	0.77	4.64	95.8
A-2615T	1	1	2	329.1	0.71	5.07	71.6
A-2616T	2	1	3	299.7	0.68	5.20	76.5
A-2617T	2	2	4	329.1	0.68	5.51	77.0
A-2620T	2	2	4	306.7	0.68	5.80	110.6
A-2624T	2	2	4	326.8	0.71	5.44	89.9
A-2625T	2	1	3	352.0	0.69	5.46	85.0
A-2628T	2	2	4	236.5	0.62	4.81	109.6
A-2636T	2	1	3	329.2	0.67	5.48	79.2
A-2637T	2	2	4	318.8	0.70	5.39	82.1

Genotype	2019 replicates	2020 replicates	Total replicates	100-Seed weight (mg)	WL ratio	Area (mm^2)	Seeds per berry
A-2639T	1	---	1	348.6	0.70	5.16	68.8
A-2642T	2	2	4	345.5	0.70	5.56	89.3
A-2644T	2	2	4	331.1	0.70	5.49	87.5
A-2651T	2	---	2	318.8	0.74	4.95	66.1
A-2652T	2	2	4	436.6	0.73	6.68	72.5
A-2656T	2	2	4	249.9	0.72	4.81	97.1
A-2657T	1	---	1	316.2	0.72	5.07	74.2
A-2658T	2	2	4	267.4	0.62	5.24	108.9
A-2660T	2	---	2	350.9	0.79	5.25	77.3
A-2663T	2	2	4	292.0	0.74	4.95	92.5
A-2665T	2	2	4	366.3	0.68	5.91	75.8
A-2668T	2	2	4	355.1	0.78	5.39	59.0
A-2669T	2	2	4	353.6	0.71	5.19	64.3
A-2672T	2	2	4	260.6	0.69	4.47	92.7
A-2673T	2	2	4	276.7	0.67	4.58	73.0
A-2676T	2	---	2	440.4	0.68	6.71	82.0
A-2677T	2	2	4	344.9	0.71	5.23	60.4
A-2678T	---	1	1	273.0	0.70	4.60	75.7
A-2680T	2	2	4	366.2	0.71	5.36	77.3
A-2683T	2	---	2	319.3	0.63	5.27	65.9
A-2684T	2	2	4	389.4	0.69	6.07	64.9
A-2685T	2	2	4	307.7	0.77	5.42	71.4
A-2686T	2	2	4	303.7	0.66	5.17	74.9
A-2687T	2	2	4	424.1	0.70	6.02	72.9
A-2688T	2	---	2	350.9	0.64	5.99	87.5
A-2700T	2	2	4	405.2	0.70	6.29	74.5
A-2701T	2	2	4	386.5	0.69	6.05	84.7

Genotype	2019 replicates	2020 replicates	Total replicates	100-Seed weight (mg)	WL ratio	Area (mm^2)	Seeds per berry
A-2708T	2	2	4	352.1	0.71	5.51	76.3
A-2709T	2	2	4	348.4	0.73	5.89	71.8
A-2710T	2	2	4	329.8	0.71	5.18	78.2
A-2711T	2	2	4	239.1	0.64	4.72	81.4
A-2713T	2	2	4	357.7	0.69	5.33	82.7
A-2716T	2	2	4	328.4	0.69	5.50	87.8
A-2717T	2	2	4	339.5	0.72	5.29	72.7
A-2718T	2	2	4	356.2	0.68	5.66	71.6
A-2720T	2	2	4	388.1	0.70	6.50	85.0
A-2722T	1	2	3	264.4	0.69	4.88	81.1
A-2723T	2	2	4	259.1	0.69	4.68	80.4
A-2724T	2	2	4	349.7	0.69	5.37	78.1
A-2725T	2	2	4	419.4	0.72	6.40	58.7
A-2726TN	2	---	2	365.0	0.78	5.45	63.5
A-2727T	1	---	1	285.5	0.77	4.67	67.0
A-2728T	2	---	2	284.2	0.80	5.02	75.3
A-2730T	2	2	4	322.7	0.71	5.26	87.4
A-2731T	2	2	4	318.0	0.72	5.47	74.5
A-2733T	2	2	4	316.1	0.72	5.47	79.0
A-2734T	2	---	2	286.4	0.70	4.74	71.2
A-2735T	2	---	2	313.8	0.74	5.39	74.5
A-2736T	2	---	2	282.1	0.67	5.60	76.4
A-2738T	2	2	4	299.7	0.69	5.34	76.3
A-2739T	2	2	4	317.2	0.67	5.49	88.2
A-2740T	2	2	4	315.8	0.65	5.07	82.0
A-2742	2	---	2	294.2	0.69	5.25	92.0
A-2743T	2	2	4	304.3	0.68	4.98	74.1

Genotype	2019 replicates	2020 replicates	Total replicates	100-Seed weight (mg)	WL ratio	Area (mm^2)	Seeds per berry
A-2745T	2	2	4	378.8	0.69	5.60	77.4
A-2746T	2	2	4	378.9	0.67	5.72	69.9
A-2747T	2	2	4	417.7	0.63	6.37	80.0
A-2749T	2	2	4	336.4	0.72	5.72	80.4
A-2751T	2	2	4	298.8	0.69	5.03	62.5
A-2752T	2	2	4	394.8	0.68	5.93	75.4
A-2753T	2	2	4	319.9	0.66	5.23	102.3
A-2755T	2	2	4	295.4	0.68	5.59	88.0
A-2756T	2	2	4	352.5	0.70	5.58	91.3
A-2757T	2	2	4	312.9	0.71	5.00	70.6
A-2758T	---	1	1	338.8	0.66	6.32	97.3
A-2759T	2	2	4	281.1	0.66	4.70	89.3
A-2760T	2	2	4	331.9	0.65	5.42	106.6
A-2761T	1	2	3	261.9	0.68	5.08	84.9
A-2762T	2	2	4	273.2	0.69	5.25	89.0
A-2763T	2	2	4	282.8	0.63	5.15	109.0
A-2764T	1	---	1	295.5	0.72	5.38	87.9
A-2765T	2	2	4	253.5	0.70	4.50	65.5
A-2766T	2	2	4	326.7	0.64	5.23	72.4
A-2767T	2	2	4	269.4	0.71	4.50	78.6
A-2768T	2	2	4	302.6	0.73	5.08	92.2
A-2770T	2	2	4	284.0	0.78	4.61	75.9
A-2771T	2	2	4	234.2	0.65	4.13	84.4
A-2772T	2	2	4	289.5	0.65	5.05	82.6
A-2773T	1	2	3	291.5	0.69	5.34	85.2
A-2774T	2	2	4	346.4	0.71	5.18	79.7
A-2775T	2	---	2	381.2	0.71	5.90	62.4

Genotype	2019 replicates	2020 replicates	Total replicates	100-Seed weight (mg)	WL ratio	Area (mm^2)	Seeds per berry
A-2777T	2	1	3	298.2	0.75	4.74	79.2
A-2778T	2	2	4	343.6	0.69	5.77	86.8
A-2779T	2	2	4	423.4	0.70	6.43	86.7
A-2780T	2	1	3	387.1	0.60	5.63	79.6
A-2781T	---	2	2	345.0	0.68	5.22	69.9
A-2782T	---	2	2	330.4	0.68	5.42	83.4
A-2783T	---	1	1	264.2	0.74	4.65	---
A-2784T	---	2	2	305.9	0.68	6.00	89.2
A-2785T	---	2	2	294.5	0.77	5.12	69.7
A-2786T	---	2	2	305.4	0.67	5.30	85.1
A-2787T	---	2	2	286.4	0.74	4.96	71.5
A-2788T	---	2	2	294.0	0.69	5.76	77.8
A-2789TN	---	2	2	286.7	0.70	4.98	84.9
A-2790TN	---	2	2	332.1	0.71	5.37	65.8
A-2791TN	---	2	2	290.3	0.67	4.58	71.2
A-2792TN	---	2	2	304.1	0.70	4.70	97.6
A-2793TN	---	2	2	340.9	0.75	5.17	68.5
A-2794TN	---	2	2	332.3	0.72	4.96	80.2
A-2795TN	---	2	2	373.2	0.73	5.20	74.7
A-2796TN	---	2	2	345.1	0.70	5.22	72.6
A-2797T	---	1	1	303.1	0.72	5.13	---
A-2798T	---	2	2	326.3	0.68	5.04	69.8
A-2799T	---	2	2	376.4	0.70	6.53	85.0
A-2801T	---	2	2	267.9	0.70	4.80	84.6
A-2802T	---	2	2	280.2	0.73	4.80	73.2
Prime Jim	1	1	2	233.1	0.65	4.19	90.5
Madeline	2	2	4	279.5	0.73	4.46	76.6

Genotype	2019 replicates	2020 replicates	Total replicates	100-Seed weight (mg)	WL ratio	Area (mm^2)	Seeds per berry
Sharon's Delight	2	2	4	289.6	0.67	5.04	97.8
Prime-Ark® Freedom	2	1	3	294.3	0.70	4.75	93.9
Prime-Ark® Traveler	2	2	4	316.2	0.71	5.33	81.6
Black Gem	2	2	4	282.7	0.68	4.48	95.6
Baby Cakes	2	2	4	310.4	0.69	4.70	80.1
APF-238T	2	---	2	278.4	0.77	4.54	60.2
APF-259TN	2	2	4	263.7	0.64	5.14	81.2
Prime-Ark® Horizon	2	2	4	371.3	0.77	5.55	84.8
APF-276TN	2	2	4	270.9	0.65	5.15	95.9
APF-298TN	2	2	4	260.0	0.68	4.92	94.0
APF-318	2	---	2	319.8	0.72	5.28	43.6
APF-328	2	2	4	281.2	0.73	4.76	90.0
APF-334T	2	2	4	353.4	0.70	5.28	91.0
APF-335T	2	2	4	367.3	0.71	5.56	77.0
APF-338	2	---	2	307.1	0.68	4.98	81.6
APF-341TN	2	2	4	360.1	0.68	5.48	77.4
APF-345T	2	---	2	279.4	0.65	4.44	90.8
APF-355TN	2	2	4	305.5	0.70	5.18	74.2
APF-366T	2	2	4	350.4	0.71	5.86	88.0
APF-370T	2	2	4	310.5	0.74	5.05	82.5
APF-372T	1	1	2	324.3	0.69	4.97	104.0
APF-373T	2	2	4	352.8	0.70	5.39	82.1
APF-377TN	2	2	4	282.9	0.69	4.73	89.5
APF-379TN	1	2	3	307.8	0.74	4.71	57.8
APF-386TN	2	2	4	287.0	0.74	5.03	75.7
APF-388T	2	---	2	262.8	0.75	4.67	94.5
APF-389TN	2	2	4	341.0	0.69	5.06	73.1

Genotype	2019 replicates	2020 replicates	Total replicates	100-Seed weight (mg)	WL ratio	Area (mm^2)	Seeds per berry
APF-392TN	2	2	4	293.5	0.63	4.93	83.8
APF-394T	2	---	2	306.9	0.69	5.01	71.2
APF-399T	2	---	2	292.2	0.81	4.51	70.1
APF-400T	1	2	3	345.6	0.71	5.71	79.6
APF-404T	2	2	4	409.8	0.60	6.03	92.1
APF-405T	2	2	4	289.8	0.72	4.90	103.4
APF-406TN	2	2	4	312.7	0.74	5.17	87.1
APF-407TN	2	1	3	355.3	0.69	5.05	73.6
APF-409T	2	2	4	292.8	0.71	4.97	80.8
APF-410T	2	2	4	342.0	0.68	6.08	88.6
APF-414TN	2	2	4	359.9	0.70	5.64	83.2
APF-415T	1	2	3	299.9	0.69	4.93	84.1
APF-423TN	2	---	2	279.1	0.70	4.42	66.4
APF-424TN	2	2	4	371.9	0.67	5.44	65.3
APF-425T	1	2	3	346.5	0.77	5.66	69.6
APF-426T	2	2	4	275.4	0.68	4.85	92.1
APF-427T	---	2	2	287.9	0.65	5.06	90.5
APF-428T	2	2	4	306.6	0.64	5.26	85.1
APF-432T	---	2	2	309.9	0.66	5.45	89.0
APF-434T	2	---	2	272.9	0.69	5.30	74.4
APF-435T	---	2	2	304.1	0.66	4.89	75.6
APF-437T	2	2	4	396.9	0.72	5.79	66.4
APF-44	1	2	3	348.7	0.64	5.45	101.0
APF-440	---	1	1	359.4	0.74	6.04	69.2
APF-441	---	1	1	355.4	0.75	5.69	71.7
APF-443T	1	1	2	273.2	0.70	4.83	76.0
APF-444TN	2	2	4	295.6	0.65	5.13	99.6

Genotype	2019 replicates	2020 replicates	Total replicates	100-Seed weight (mg)	WL ratio	Area (mm^2)	Seeds per berry
APF-447T	2	2	4	265.5	0.70	4.78	82.5
APF-448T	2	2	4	298.5	0.68	5.08	81.4
APF-449TN	2	2	4	268.8	0.67	5.01	88.7
Prime-Ark® 45	1	2	3	316.4	0.69	5.04	86.9
APF-450TN	2	2	4	284.5	0.64	4.99	99.9
APF-451T	1	2	3	355.5	0.67	5.53	62.4
APF-452T	2	2	4	299.1	0.73	5.08	56.0
APF-457TN	2	2	4	421.9	0.68	6.36	80.8
APF-458T	1	---	1	230.8	0.65	4.11	89.2
APF-459T	---	2	2	307.3	0.65	5.00	79.1
APF-464TN	---	2	2	333.9	0.63	5.75	72.9
APF-465TN	2	2	4	364.3	0.71	6.12	86.0
APF-466T	2	---	2	225.9	0.65	4.32	96.4
APF-467T	2	2	4	237.8	0.70	4.34	80.3
APF-468T	2	1	3	298.8	0.71	5.04	67.4
APF-469T	2	2	4	314.5	0.70	5.22	82.6
APF-470T	1	1	2	302.3	0.77	4.85	52.2
APF-471T	1	2	3	290.4	0.68	4.80	74.3
APF-472T	2	1	3	313.6	0.67	5.04	74.9
APF-473T	1	---	1	268.7	0.67	4.70	71.9
APF-474T	1	2	3	294.5	0.68	4.96	85.0
APF-475T	2	2	4	299.5	0.76	4.96	77.5
APF-477T	---	2	2	344.0	0.59	5.77	81.6
APF-478TN	---	1	1	371.8	0.72	5.44	65.1
APF-479TN	---	2	2	298.8	0.69	5.15	64.8
APF-480TN	---	2	2	270.4	0.67	5.34	75.3
APF-481TN	---	2	2	320.1	0.70	4.65	88.1

Genotype	2019 replicates	2020 replicates	Total replicates	100-Seed weight (mg)	WL ratio	Area (mm^2)	Seeds per berry
APF-482TN	---	1	1	347.9	0.69	5.69	81.9
APF-483TN	---	1	1	327.2	0.71	5.05	75.9
APF-484TN	---	2	2	337.0	0.64	5.33	76.6
APF-485T	---	2	2	320.8	0.73	4.96	83.2
APF-77	2	2	4	278.4	0.67	4.63	97.5
APF-PBB1	2	1	3	300.4	0.56	5.26	65.8
Eclipse	---	2	2	281.5	0.70	4.60	66.6
Galaxy	1	1	2	330.0	0.67	5.04	84.1
PBB-APF-1801	---	1	1	238.8	0.66	4.61	89.5
Tupy	1	2	3	281.0	0.69	4.76	76.4
Von	1	2	3	273.2	0.65	4.93	85.2
Sample Total	439	442	881				
Genotype Total	236	237	277				

[z]Genotypes A-2435T, A-2727T, A-2758T, A-2779T, and APF-473T were excluded from association analyses because they failed to pass DNA quality filtering thresholds.