ABSTRACT

SHEEHAN-LUST, KATELYN ELLEN. Genotype × Environment Analysis, and SNP Validation and Trait Association of *Rubus spp.* (Under the direction of Dr. Gina Fernandez)

Although a very popular fruit in retail and local markets, the diversity of raspberry and blackberry (*Rubus spp.*) cultivars has been limited by the scant germplasm pool used in breeding programs. North Carolina State University, Cornell University, University of Arkansas, BC Berry Cultivar Development Inc., and the United States Department of Agriculture (USDA) collaborated with Pairwise, and Plant Sciences Inc., in a series of research projects focused on *Rubus*, the genus that includes blackberries and raspberries.

One component of this collaboration was a genotype × environmental (G × E) study that focused on phenotyping a set of traits on five seminal genotypes in five geographically distinct environments over two years. The genotypes evaluated included the primocane-fruiting red raspberry 'Heritage', the floricane-fruiting red raspberry 'Latham', the primocane-fruiting blackberry 'Prime-Ark® 45', the floricane-fruiting blackberry 'Chester Thornless', and 'Bristol' a black raspberry. Data from all sites was collated and then analyzed by the NCSU collaborators. Fruit quality components: floricane and primocane fruit size, yield components: number of fruiting laterals and berries per lateral, and plant architecture components: dormant primocane length and nodes per 50 cm were the focus of these analyses. Floricane fruit size was significantly affected by the three-way interaction of G, E, and harvest date (H). The G × E × H interaction encouraged further investigation into individual climate variables and their effects on fruit size. An interaction between genotype and growing degree days (GDD) was found to impact floricane fruit size. Analysis of the primocane fruit size indicated a significant two-way interaction between G and E. The primocane blackberry maintained the same fruit size across all

E, while the primocane raspberry fruit size differed by E. The yield indicator components including floricane fruiting lateral counts and floricane berries per lateral were affected by interactions between G and E. The plant architecture values of dormant primocane length and node densities were also analyzed. No interaction was found, and the main effects of G and E were significant for both. In the genotype × environment study, G × E interactions were significant sources of variation for most of the traits phenotyped with the exception of plant architecture traits.

The second component of the research at NCSU focused on validating a set of SNPs and associating those validations with traits of interest. Thirty-one *Rubus spp.* genotypes sequenced by Pairwise were aligned to the black raspberry (*R. occidentalis*) genome to identify single nucleotide polymorphisms (SNP). NCSU set out to validate the presence of SNPs correlated with meaningful phenotypes. Traits evaluated that are valuable to producers include the primocane fruiting and prickle-free phenotypes. Primocane fruiting SNPs were identified by comparing nucleotide identity and fruiting phenotype. This yielded SNPs associated with 3 previously annotated genes involved in protein phosphorylation, DNA repair, and RNA splicing. SNPs associated with the presence or absence of prickles were selected after aligning the flanking sequences to red raspberry genes (*R. idaeus subsp. strigosus*) previously identified by Cornell to be associated with prickle presence. Three of these SNPs were found to be associated with prickle presence and within annotated genes determined to produce proteins associated with prickle development in other Rosaceae species. Validating the presence of meaningful SNPs within the population may facilitate the development of KASP (competitive allele-specific PCR) markers, and ultimately improve marker assisted selection.

Genotype × Environment Analysis, and SNP Validation and Trait Association of *Rubus spp.*

by
Katelyn E. Sheehan-Lust

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Horticultural Science

Raleigh, North Carolina
2022

APPROVED BY:

Dr. Gina E. Fernandez Committee Chair	Dr. Cherie Ochsenfeld External Member
Dr. Hamid Ashrafi	Dr. Marcia Gumpertz
Dr. Penelope Perkins-Veazie	

DEDICATION

To my Aunt Karen, thank you for making this possible. I would not have been able to do this without you. You are greatly missed.

BIOGRAPHY

Katelyn Ellen Sheehan-Lust was born in Ventura, California in 1994. She grew up in the Ojai Valley, surrounded by orchards and adjacent to the large berry production region that is Oxnard. Katelyn started forming an interest in food production at the end of high school, raising pigs. As she began her journey at Santa Barbara City College, the interest in animal husbandry expanded. Volunteering for milking shifts as a part of a herd share program for raw goats' milk sparked her curiosity and inspired enrollment in course work supporting an agricultural science degree. Exploring this area of study, she gradually focused in on plant science.

A summer internship with PlantHaven Inc. managing ornamental plant variety trials was her first exposure to the world of plant breeding. After completing an associate degree in math and science, Katelyn was accepted at California Polytechnic State University in San Luis Obispo. By the end of her first semester, Katelyn found the Cal Poly Strawberry Center. The center director, Dr. Gerald Holmes, hired Katelyn to work as a student intern in January 2017. This position exposed her to the world of berries, and she fell in love. With the support of the Strawberry Center, she was able to secure internships with the California Strawberry Commission and the UC Davis Strawberry Breeding Program.

In December 2018, Katelyn completed a B.S. in Agricultural and Environmental Plant Science with a concentration in Fruit and Crop Science. After graduation, she began a position as a field research specialist for Naturipe Berry Growers under the direction of Dr. Hillary Thomas. This position provided the opportunity to oversee and assist in the coordination of breeding trail management and data collection for many strawberry and caneberry breeding programs. Dr. Thomas supported Katelyn’s desire to pursue a future in berry breeding and facilitated the relationships that lead to her return to academia.

In March 2020, Katelyn received an acceptance letter from North Carolina State University's Department of Horticultural Science granting her the opportunity to study caneberries under the guidance of Dr. Gina Fernandez. Katelyn, her then boyfriend Andrew, and their dog Cruiser drove from Southern California to Raleigh, NC at the end of July 2020. This long drive was only the start of an exciting journey to further explore the world of berries.

ACKNOWLEDGMENTS

I would like to acknowledge and express gratitude to all of those who have supported me through my studies and made this research possible. First, I must thank Pairwise for spearheading this project and bringing the collaborators together. Without their efforts, this project would not have been conceived. Thank you to the head of R&D at Pairwise, Ryan Rapp, for organizing and funding the genotyping and plant procurement. Thank you, Dr. Cherie Ochsenfeld, for guiding all collaborative efforts, and Andrew Fister for developing the phenotyping protocols.

Thank you to the collaborators for not only maintaining plots and collecting data, but also for providing input and guidance throughout the process. Thank you to Clare Taylor and Abraham Corrales at Plant Sciences Inc., Dr. Margaret Worthington and Erika Henderson at University of Arkansas, Dr. Courtney Weber at Cornell University, Dr. Jill Bushakra and Dr. Nahla Bassil at the USDA in Corvallis, OR, and Dr. Michael Dossett at University of British Columbia.

Thank you to the staff at the Piedmont Research Station for all their efforts in maintaining the NC research plots and continued communications to help me stay up to date on plot conditions. A special thank you to Brianna Haynes for assisting in the establishment and data collection before my arrival and when the workload was heavy. Thank you to José Guillermo Chacón-Jiménez for answering my endless number of questions, guiding me through DNA extractions, and always being available to help.

I would also like to thank my committee members at NCSU. They provided a great deal of support and direction through every step of the process. Thank you Dr. Penelope Perkins-Veazie for building my understanding of *Rubus* postharvest physiology. Thank you to Dr.

Marcia Gumpertz for providing endless guidance through the statistical analyses. Thank you to Dr. Hamid Ashrafi for your patience in building my understanding around genomic data and its applications. Most of all, I would like to thank my advisor, Dr. Gina Fernandez. Her mentorship has been instrumental to my continued growth as a researcher. She introduced me to East Coast berry production, helped build my understanding in the art of plant breeding, and became my support system when the pandemic was isolating. I could not have done this without her and will eternally be grateful for her support and guidance. Thank you, Dr. Gina Fernandez.

Lastly, I would like to thank my friends and family for supporting me through the ups and downs of graduate school. Thank you to my husband, Andrew, for being my biggest supporter. And to our dog-baby, Cruiser, for forcing me to take walk breaks during long study days. Thank you to my parents, Paul and Sue Sheehan-Lust, for always pushing me to succeed. A special thank you to all of those who visited and called often, it made me feel like home wasn't so far away: Hillary Cooke, Taylor Young, Nikki Sheehan-Lust, Liz Golden, Bonnie Collins, Zelda Grove, Molly Roberts, and Jonas Grove.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Literature Review

Significance of *Rubus*

Caneberry crops include numerous berry species of economic, nutritional, and medicinal importance. Blackberries (*Rubus spp*.), raspberries (*R. idaeus*), and black raspberries (*R. occidentalis*) are the three groups of caneberries that are well known to the consumers and have been in growing demand, especially for the fresh fruit market. Per capita consumption of caneberries has increased 17% annually over the past decade (Clark, 2019). According to FAOSTAT, during 2016 – 2018, the United States produced 106,423 tons of the 830,544 tons or 12.8% of raspberries produced globally. This makes the US the fifth largest producer of raspberry (FAOSTAT, 2020). The United States produces an even greater proportion of blackberries. A 2005 report indicates the US produced 35,097 tons (23%) of the 154,644 tons of *Rubus spp*. grown worldwide (Strik et al., 2007).

In the southeastern US, the commercial blackberry industry started to emerge in the early 2000's when SunnyRidge Farms first contracted with growers in GA, SC and NC (Fernandez, personal communication). The industry has grown and numerous companies have established a regionally significant shipping market (Sowder, 2021). Pick-your-own and direct market farms are a smaller but growing industry in this region (Clark, 2019).

Caneberry fruit offers a sweet, low-calorie snack with essential nutrients, anthocyanins, and other beneficial phytochemicals with antioxidant properties (Rao and Snyder, 2010; Zafra-Rojas et al., 2018). Blackberries have been determined to contain greater proportions of organic acids,

phenolic compounds, anthocyanins and antioxidant capacity than prunes. Additionally, the dietary fiber effects are greater from blackberry than prune (Zafra-Rojas et al., 2018). The high fiber and fructose content of raspberries has the potential to help regulate blood sugar. Raspberries also contain a substantial amount of vitamin C in the flesh of the drupelets and fat-soluble vitamins in the seeds (Rao and Snyder, 2010). In a study utilizing rats, black raspberry fruit was found to prevent the development of esophageal tumors before formation and once abnormal cell lesions were present (Stoner et al., 2007). The nutritional value is significant, and compounds prevalent in caneberry fruit have many potential health benefits. The high concentrations of ellagitannins, polyphenols, and anthocyanins in *Rubus* species make it a good source of antioxidants (Moreno-Medina, 2018). In addition to antioxidant properties, anthocyanins also create an anti-inflammatory and anticancer effect (Seeram, 2008). The popularity and beneficial properties of caneberries drives consumption and production, but delectable fruit are not consistently found when the fruit is purchased in the grocery stores.

Rubus **breeding**

The genus *Rubus* is extremely diverse in terms of fruit and plant characteristics, but this diversity has narrowed as breeding efforts have concentrated on specific fruit and plant adaptations in breeding programs. There are more than 740 species of *Rubus* with origins from around the globe (Hummer, 1996). This global availability of wild *Rubus* may have contributed to the lack of formal improvement efforts or breeding programs focused on *Rubus* until the late 19th century. However, breeding efforts over the last century have been successful in improving numerous traits less common in the wild types of the genus (Clark et al., 2007). Blackberries and raspberries have seen great improvements in plant architecture, increased yields, fruit quality,

thornlessness, primocane fruiting crops, self-fertility, and environmental adaptability (Clark et al., 2007; Hall et al., 2009). This improvement has required the repeated use of a handful of older genotypes resulting in a loss of diversity. Of the 12-15 subgenera of *Rubus* that have been described, *Idaeobatus* and *Eubatus* are the two most commonly used in cultivar development (Hall et al., 2009). The ancestry of red raspberries is primarily dominated by the five cultivars 'Lloyd George', 'Pynes Royal', 'Preussen', 'Cuthbert' and 'Newburgh'. These are all derived from the subgenus *Idaeobatus*, while three varieties also include the subgenus *Eubatus*. This has resulted in commercial cultivars that are genetically similar (Graham and Woodhead, 2009). Although many horticultural improvements have been made in *Rubus*, improvement of traits such as fruit firmness, thornlessness, heat tolerance, pest tolerance/resistance and primocane crop yields still require improvements (Hall et al., 2009; Graham and Woodhead, 2009). Red raspberry domestication efforts began in the 1800's. Controlled crossing led to the release of 'Latham' in 1914, and the primocane-fruiting 'Heritage' in 1969 (Jennings, 1988). Formal blackberry breeding efforts were first established in the 1920's in the southern US, including stations in Arkansas, Florida, Georgia, North Carolina, Tennessee, and Virginia (Ballington, 2016). Black raspberries have experienced less domestication efforts and have retained more wild characteristics including disease tolerance, difficult growth habits, spiny canes, and small fruit size. Despite the lack of domestication, available cultivars also lack diversity, similar to red raspberry and blackberry (Hall et al., 2009). The amount of diversity available in the genus, paralleled by the reduced diversity in domesticated cultivars, creates an opportunity for great improvements.

Developing an understanding of the effects of genotype, environment, and their interactions on the expression of economically important traits will improve cultivar development and predictions of performance in a variety of production regions (Nyquist and Baker, 1991). Quantifying the effect of genetics on trait expression provides information to breeders that will aid in defining breeding goals and expectations (Gilbert et al., 2015). Incorporating genomic information from a SNP validation study will be foundational for the development of tools for modern breeding techniques. The discovery and validation of SNPs correlated with traits of interest will allow for marker development and lead to the ability to implement genomic selection (Chagné et al., 2012).

G × E evaluation

Genotype × environment (G × E) evaluations are performed to determine trait plasticity and variety performance in different cultivation locations. A range of phenotypic expressions are possible for a single genotype depending on the environment in which it is cultivated. Genotype does not necessarily provide an exact phenotype value but rather a range of possible expressions, especially for quantitative traits (Nyquist and Baker, 1991). The greater the effect of genotype, the more narrow the possible range of phenotypes is possible. This is more common in qualitative traits. Genotype variance is a measure of the variability controlled by genotype within a cultivar, indicating trait stability (Gonçalves et al, 2013). A genotype's phenotypic expression can be heavily influenced by environment. Evaluating the effect of environment requires consistent management across locations so the variability measured is not affected by cultivation practices. An environment can be a location, but it also can be a growing season or year. Weather patterns change season to season, so evaluations over multiple years should be considered as

different environments (Hill, 2010; Gonçalves et al., 2013). Consistent or predictable results across a range of environments is ideal for development of cultivars for diverse production regions. Genotype-by-environment interactions result from the assessment of a particular genotype in an individual environment (Nyquist and Baker, 1991). The G × E interaction is an account of the phenotypic variability not explained by the average performance of the genotype or the average performance of the environment independently (Hill, 2010). Identifying the proportion of effect the genotype, environment, and their interactions have on phenotype improves estimation of trait performance.

G × E studies in small fruit species

Within the Rosaceae family, a strawberry (*Fragaria x ananassa*) G × E trial conducted on nine named cultivars at three environments throughout Italy evaluated fruit quality characteristics and correlated the chemical components to tasting preferences. Climate components were incorporated into the analysis to find relationships between environment and phenotypic expression (Palmieri et al., 2017). Similar G × E evaluations of fruit quality preferences have been conducted in blueberry (*Vaccinium corymbosum* hybrids), using six cultivars, over three years and at three diverse environments in Florida, to direct breeding goals toward traits with higher proportions of genetic control over phenotypic expression (Gilbert et al., 2015). Other fruit quality evaluations in blueberry have been performed using different harvests in data collection and analysis as an additional factor to capture the differences in fruit quality throughout a growing season (Redpath et al., 2021).

Exploration of the effects of genotype, environment, and their interactions (G × E) in *Rubus* has previously focused on one commodity per evaluation and has not been as diverse in environments and traits evaluated. Most studies have involved either red or black raspberries. Yield stability has been assessed in red raspberry (*R. idaeus* L.) for three cultivars across fourteen environments (Sullivan and Prive, 2001). This study produced an ANOVA evaluating the effects of genotype, environment, and their interactions on yield. The most stable of the three cultivars was 'Heritage', which is one of the selected genotypes in the G × E evaluation used in the current study. A G × E evaluation encompassing a set of environments and traits comparable to the current study was conducted by the USDA in Oregon over a three-year period (Bushakra et al., 2016). In this study, two full-sibling populations of black raspberries (*R. occidentalis* L.) derived from USDA plant materials were used and provided information on the behavior of black raspberry crops. This same population was evaluated for drupelet counts in order to identify the most stable genotypes over the range of environments to correlate the behavior with a genic region that controls for drupelet number (Willman et al., 2020). The current G × E study includes similar traits but expands into plant architecture and evaluates multiple species of *Rubus*, providing more general information about the genus.

The aim of the evaluation in progress is to provide information on the adaptability of economically important traits and the effects of genotype, environment, and their interactions. The range of environments that encompass the primary North American public breeding programs and production regions has not been conducted in a single evaluation. There also has previously been a limited focus on the traits of interest, with minimal evaluation of plant

architecture. Comparing genotypes of different species of *Rubus* is unique, especially considering the scope of environments and traits.

Single nucleotide polymorphism validation

Comparison of single nucleotide polymorphisms (SNP) linked to traits of interests from a range of genotypes is used to determine if these polymorphisms are present with the same associations in other genotypes. SNPs are found by comparing similar DNA sequences within a population and looking for single nucleotide differences between genotypes. Individual genotypes selected for SNP discovery and validation should be representative of the range of germplasm available. The DNA sequences from the genotype are then aligned to an established whole genome sequence (Chagné et al., 2012). When a common sequence of nucleotides is found within the individuals in a population with a single nucleotide differing between genotypes, this is considered a SNP. SNPs are abundant within a genome and easy to find, especially within highly heterozygous species. They are usually bi-allelic, meaning they are less polymorphic compared to SSRs (simple sequence repeats) (Byrne, 2009). There will be millions of SNPs found but finding the SNPs that are meaningful is the challenge. Many SNPs will be found in non-coding regions of the genome, meaning there is no effect on phenotype. Other SNPs may be in coding regions but also have no effect on phenotype. Selected SNPs should come from quality reads with good coverage, be well dispersed throughout the genome, and be associated with phenotypic information of interest. A variety of filtering techniques are used to narrow down the list of SNPs from millions to thousands, and ultimately only a few are selected for validation (Bianco et al., 2014). Comparing SNPs found in one population to another population is done to validate the existence of SNPs in a broader range of genotypes.

SNP validation can assist in the construction of linkage maps by comparing the recombination frequencies between SNPs. Lower recombination rates between SNPs may be considered linked (Senthilvel et al., 2019). Validating SNPs that are correlated with traits of economic importance would also facilitate the creation of molecular markers. Having markers associated with SNPs of interest allows for predictions to be made as to how a particular genotype will perform at maturity (Chagné et al., 2012). Predictions of performance using DNA extracted from small amounts of plant tissue from young plants would allow for more efficient use of resources for breeding programs. If juvenile plants can be genotyped, then only plants with promising genetic markers can be maintained for further evaluation at maturity. This eliminates the need to maintain plants through a long juvenile period before being able to phenotype (Bus et al., 2009). A population with only the individuals determined to be superior through genotyping improves the probability the traits associated with the markers will be present.

Single nucleotide polymorphism validations in *Rubus*

SNP discovery, validation, and further explorations in *Rubus* have primarily concentrated on a single trait and usually a single species. Prickles on the canes and leaves of *Rubus* are considered impediments to pruning, training and harvest. Prickle SNPs were discovered using a segregating population (‘Joan J’ × ‘Caroline’) of *Rubus idaeus* L. The four associated SNPs were determined to be near or within annotated genes correlated with the prickle-free phenotype (Khadgi and Weber, 2020). A segregating population of raspberry has also been used to explore the primocane-fruiting characteristic. The population was phenotyped in the field for the presence of fruit at the tips of the first-year canes. The population was concurrently genotyped and the SNPs flanking regions associated with the annual fruiting trait were used to develop PCR

primers and facilitated the development of HCR markers (Jibran et al., 2019). Full sibling populations of *Rubus occidentalis* L. were evaluated for aphid resistance and 84 associated SNPs were found in three regions of the genome (Bushakra et al., 2018).

Interspecific populations of *Rubus occidentalis* L. and *Rubus idaeus* L. have been evaluated to determine the region of the genome that controls the production of an epicuticular wax on the canes. This protective wax may affect the susceptibility of the canes to disease. SNP validation and analysis led to the determination the waxy bloom trait was associated with a 2Mb region between two SNPs (Pinczinger et al., 2020). SNPs have also been used to determine levels of relatedness in *Rubus glaucus* cultivars to aid in progenitor selection. The SNPs validated were used to construct dendrogram clusters, but no relationship was found between thorn morphology or geographic origin (López et al., 2019). Alleles associated with soluble solid content were determined to be on chromosome 4 and 6 in *R. occidentalis*. This determination was made using a family-wide approach, validating SNPs using the genome assemblies from the family Rosaceae and the genera *Malus*, *Prunus*, *Fragaria*, and ultimately *Rubus* (Zurn et al., 2020)

The objective of the present study is to provide SNP validation and association information for prickle-free and primocane-fruiting traits in the genus *Rubus*. The materials selected include primarily red raspberry genotypes, but also blackberry and black raspberry genotypes. The wide range selected was representative of the public germplasm currently accessible for breeding efforts. Evaluation of SNPs will not be isolated to a single trait, but rather allelic variation will be explored for correlations with fruiting habits and prickle presence.

Research objectives

Two studies were conducted, consisting of a field-based multilocation phenotyping study and a survey of genomic information in a range of *Rubus* genotypes. The studies are a part of a large public-private research project with Pairwise, Plant Sciences Inc., North Carolina State University, Cornell University, University of Arkansas, USDA-NCGR, and BC Berry Cultivar Development Inc.

To provide information on the effects of genotype, environment, and their interaction on economically important traits, a genotype × environment study was conducted on five genotypes (G) in five environments (E). The five genotypes include two raspberries, two blackberries, and one black raspberry. The blackberry genotypes selected were the tetraploids 'Prime-Ark® 45' and 'Chester Thornless'. All the selected raspberry genotypes are diploids, including red raspberries 'Heritage' and 'Latham', and the black raspberry 'Bristol'. 'Prime-Ark® 45' and 'Heritage' are the two primocane fruiting genotypes. The difference in production of a primocane crop versus a floricane crop is significant, so the traits recorded specific to primocane production will be analyzed independent of floricane values.

These genotypes (G) were planned to be evaluated across six environments (E) maintained by public and private partners spread throughout the United States (Figure 1.1). The environments included Salisbury, North Carolina; Fayetteville, Arkansas; Ithaca, New York; Watsonville, California; Corvallis, Oregon. Unfortunately, delays caused by SARS-CoV-2 virus excluded the Vancouver, British Columbia environment from the analysis. The evaluated environments were located in significant North American caneberry production regions. Oregon, California,

Arkansas, North Carolina, and New York, These locations collectively produced 30,200 tons (86%) of the 35,097 tons of blackberries produced nationwide (Strik et al., 2007). Placement in these target environments provides relevant information for substantial production regions. Over five plant-related and fifteen fruit-related traits were quantitatively evaluated for the five genotypes at all environments, starting with the primocane crop and continuing through the first season of floricane production. Plant architecture traits evaluated included: canes per plant, erectness scoring, leaf structure, prickle score, emergence date, cane length, node count, cane diameter, and cane elasticity. Fruit quality and yield component traits evaluated included soluble solids (brix), titratable acidity, anthocyanins, berry size, drupelet count and size, berry color, firmness, seed size, time of flowering, number of flowering nodes, and inflorescence flower counts. These traits represent economically important features in caneberry production and consumer preferences. Two traits from each category were evaluated in this study, while other trait data remains available for further exploration.

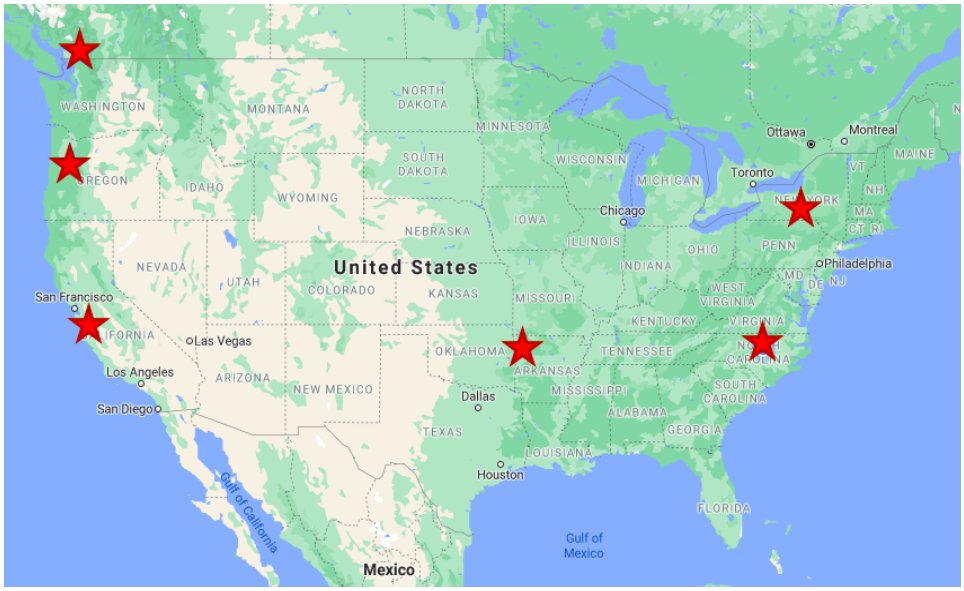

Figure 1.1. Location of the participating public breeding institutions and private companies in the genotype x environment study. These companies include Pairwise (NC), Plant Sciences Inc. (CA), North Carolina State University (NC), Cornell University (NY), University of Arkansas (AR), USDA-NCGR (OR), BC Berry Cultivar Development Inc (BC Canada).

Table 1.1. List of 31 genotypes, their genus and species and geographic origin that were utilized in the SNP discovery and validation in *Rubus spp*..

Name	Taxon	Origin
Ralitsa	*R. idaeus subsp. idaeus*	Bulgaria
Bababerry	*R. idaeus subsp. idaeus*	United States, California
R. strigosus Washington	*R. idaeus subsp. strigosus*	United States, Washington
Liberty	*R. idaeus subsp. idaeus*	United States, Iowa
Ranere	*R. idaeus subsp. idaeus*	United States, New Jersey
Viking	*R. idaeus subsp. idaeus*	Canada, Ontario
Newburgh	*R. idaeus subsp. idaeus*	United States, New York
Scepter	*R. idaeus subsp. idaeus*	United States, Maryland
Heritage	*R. idaeus subsp. idaeus*	United States, New York
Watson	*R. idaeus subsp. idaeus*	United States, New York
Comet	*R. idaeus subsp. idaeus*	Canada, Ontario
June	*R. idaeus subsp. idaeus*	Uncertain
Mammoth Red	*R. idaeus subsp. idaeus*	Uncertain
R. hirsutus 26-21 Japan	*R. hirsutus*	Japan
Bristol	*R. occidentalis*	United States, New York
R. strigosus Jarbridge River ID	*R. idaeus subsp. strigosus*	United States, Idaho
Gradina	*R. idaeus subsp. idaeus*	Former Serbia and Montenegro
Krupna Dvorodna	*R. idaeus subsp. idaeus*	Former Serbia and Montenegro
Pocahontas	*R. idaeus subsp. idaeus*	United States, Virginia
Southland	*R. idaeus subsp. idaeus*	United States, North Carolina
Preussen	*R. idaeus subsp. idaeus*	Germany
Korbfuller	*R. idaeus subsp. idaeus*	Germany
Zeva II	*R. idaeus subsp. idaeus*	Switzerland
Schoenemann	*R. idaeus subsp. idaeus*	Germany
Oregon 1030	*R. idaeus subsp. idaeus*	United States, Oregon
Lewis	*R. idaeus subsp. idaeus*	United States, Oregon
Chilcotin	*R. idaeus subsp. idaeus*	Canada, British Columbia
Cuthbert	*R. idaeus subsp. idaeus*	United States, New York
Latham	*R. idaeus subsp. idaeus*	United States, Minnesota
Prime-Ark® 45	*R. hybrid*	United States, Arkansas
Chester Thornless	*R. hybrid*	United States, Ohio

The SNP validation study was conducted to generate a concise list of alleles present in a representative group of *Rubus* genotypes that were associated with economically important traits. The 31 genotypes selected include red raspberry, black raspberry, and blackberry genotypes (Table 1.1). The individuals selected represent a range of genotypes from geographically diverse origins. The genotypes were aligned to the *R. occidentalis* genome assembly to find the initial list of SNPs. The prickle presence analysis will utilize genes previously discovered in *R. idaeus* and SNPs with flanking sequences located within these genes to determine the presence of

associated alleles in the selected genotypes. SNPs related to primocane- and floricane-fruiting habits will be evaluated by looking for correlations between the cultivar's phenotype and genotype. SNP identities with a high correlation to phenotype across the population will be selected for validation.

This research is unique in the subject matter, extent of traits evaluated, and the variety of environments tested. This is the first instance of a genotype-by-environment study and SNP validation across multiple species of *Rubus*. Quantifying the effects of genotype, environment, and their interactions on important traits will allow for increased efficiency in breeding programs and better predictions of performance in diverse growing regions. Identification and validation of SNPs associated with economically important traits promotes the potential for marker development and future marker assisted selection in *Rubus* breeding.

LITERATURE CITED

Ballington, J. R. (2016). The history of blackberry and raspberry breeding in the southern USA. *Acta Horticulturae,* (1133), 13-22. https://10.17660/ActaHortic.2016.1133.3

Bianco, L., Cestaro, A., Sargent, D. J., Guardo, D., M, Jansen, J., & Weg, V. D., W.E. (2014). Development and validation of a 20K single nucleotide polymorphism (SNP) whole genome genotyping array for apple (*Malus* × *domestica* Borkh). *PLoS ONE, 9*(10), e110377. https://10.1371/journal.pone.0110377

Bus, V., Esmenjaud, D., Buck, E., & Laurens, F. (2009). Application of genetic markers in rosaceous crops. In K. M. Folta, & S. E. Gardiner (Eds.), *Genetics and Genomics of Rosaceae* (1.th ed.,). Springer. https://10.1007/978-0-387-77491-6

Bushakra, J. M., Bradish, C. M., Weber, C. A., Dossett, M., Fernandez, G., Weiland, J., Peterson, M., Scheerens, J. C., Robbins, L., Serçe, S., Finn, C. E., & Bassil, N. V. (2016). Toward understanding genotype × environment interactions in black raspberry (*Rubus occidentalis* L.). *Acta Horticulturae,* (1117), 25-30. https://10.17660/ActaHortic.2016.1117.5

Bushakra, J. M., Dossett, M., Carter, K. A., Vining, K. J., Lee, J. C., Bryant, D. W., VanBuren, R., Lee, J., Mockler, T. C., Finn, C. E., & Bassil, N. V. (2018). Characterization of aphid resistance loci in black raspberry (*Rubus occidentalis* L.). *Molecular Breeding, 38*(7), 1-19. https://10.1007/s11032-018-0839-5

Byrne, D. H. (2009). Rose structural genomics. In K. M. Folta, & S. E. Gardiner (Eds.), *Genetics and Genomics of Rosaceae* (1.th ed.,). Springer. https://10.1007/978-0-387-77491-6

Chagné, D., Crowhurst, R. N., Troggio, M., Davey, M. W., Gilmore, B., Lawley, C., Vanderzande, S., Hellens, R. P., Kumar, S., Cestaro, A., Velasco, R., Main, D., Rees, J. D., Iezzoni, A. F., Mockler, T., Wilhelm, L., Weg, v. d., W.E, Gardiner, S. E., Bassil, N., & Peace, C. (2012). Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. *PLoS ONE, 7*(2), e31745. https://10.1371/journal.pone.0031745

Clark, J. (2019). Markets Strong for Berries. Paper presented at the *NARBA,*

Clark, J. R., Stafne, E. T., Hall, H. K., & Finn, C. E. (2007). Blackberry breeding and genetics. *Plant Breeding Reviews 29: 19-124*

Food and Agriculture Organization of the United Nations. (2020). http://www.fao.org/faostat/en/#data/QC/visualize

Gilbert, J. L., Guthart, M. J., Gezan, S. A., Pisaroglo de Carvalho, M., Schwieterman, M. L., Colquhoun, T. A., Bartoshuk, L. M., Sims, C. A., Clark, D. G., & Olmstead, J. W. (2015). Identifying breeding priorities for blueberry flavor using biochemical, sensory, and genotype by environment analyses. *PLoS ONE, 10*(9), e0138494. https://10.1371/journal.pone.0138494

Gonçalves, E., Carrasquinho, I., St. Aubyn, A., & Martins, A. (2012). Broad-sense heritability in mixed models for grapevine initial selection trials. *Euphytica, 189*(3), 379-391. https://10.1007/s10681-012-0787-9

Graham, J., & Woodhead, M. (2009). Raspberries and blackberries: The genomics of rubus. *Genetics and Genomics of Rosaceae* (pp. 507-524). Springer New York. https://10.1007/978-0-387-77491-6_24

Hall, H. K., Hummer, K. E., Jamieson, A. R., Jennings, N. S., & Weber, C. A. (2009). Raspberry breeding and genetics. *Plant Breeding Reviews 32* (pp. 39-314)

Hill, W. G. (2010). Understanding and using quantitative genetic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences, 365*(1537), 73-85. https://10.1098/rstb.2009.0203

Hummer, K. E. (1996). *Rubus Diversity* HortScience.

Jennings, D. L. (1988). *Raspberries and blackberries: Their breeding, diseases and growth*. Academic Press.

Jibran, R., Spencer, J., Fernandez, G., Monfort, A., Mnejja, M., Dzierzon, H., Tahir, J., Davies, K., Chagné, D., & Foster, T. M. (2019). Two loci, RiAF3 and RiAF4, contribute to the annual-fruiting trait in *Rubus*. *Frontiers in Plant Science, 10*, 1341. https://10.3389/fpls.2019.01341

Khadgi, A., & Weber, C. A. (2020). Genome-wide association study (GWAS) for examining the genomics controlling prickle production in red raspberry (*Rubus idaeus* L.). *Agronomy (Basel), 11*(1), 27. https://10.3390/agronomy11010027

López, A. M., Barrera, C. F., & Marulanda, M. L. (2019). Evaluation of SSR and SNP markers in *Rubus glaucus* Benth progenitors selection. *Revista Brasileira De Fruticultura, 41*(1)https://10.1590/0100-29452019081

Moreno-Medina, B. L., Casierra-Posada, F., & Cutler, J. (2018). Phytochemical composition and potential use of *Rubus* species. *Gesunde Pflanzen, 70*(2), 65-74. https://10.1007/s10343-018-0416-1

Nyquist, W.E. (Purdue University,West Lafayette, IN). (1991). Estimation of heritability and prediction of selection response in plant populations. *Critical Reviews in Plant Sciences, 10*(3), 235-322. https://10.1080/07352689109382313

Palmieri, L., Masuero, D., Martinatti, P., Baratto, G., Martens, S., & Vrhovsek, U. (2017). Genotype-by-environment effect on bioactive compounds in strawberry (*Fragaria x ananassa* duch.). *Journal of the Science of Food and Agriculture, 97*(12), 4180-4189. https://10.1002/jsfa.8290

Pinczinger, D., von Reth, M., Keilwagen, J., Berner, T., Peil, A., Flachowsky, H., & Emeriewen, O. F. (2020). Mapping of the waxy bloom gene in 'Black Jewel' in a parental linkage map of 'Black Jewel' X 'Glen Ample' interspecific population. *Agronomy (Basel), 10*(10), 1. https://10.3390/agronomy10101579

Rao, A. V., & Snyder, D. M. (2010). Raspberries and human health: A review. *Journal of Agricultural and Food Chemistry, 58*(7), 3871-3883. https://10.1021/jf903484g

Redpath, L. E., Gumpertz, M., Ballington, J. R., Bassil, N., & Ashrafi, H. (2021). Genotype, environment, year, and harvest effects on fruit quality traits of five blueberry (*Vaccinium corymbosum* L.) cultivars. *Agronomy (Basel), 11*(9), 1788. https://10.3390/agronomy11091788

Seeram, N. P. (2008). Berry fruits: Compositional elements, biochemical activities, and the impact of their intake on human health, performance, and disease. *Journal of Agricultural and Food Chemistry, 56*(3), 627-629. https://10.1021/jf071988k

Senthilvel, S., Ghosh, A., Shaik, M., Shaw, R. K., & Bagali, P. G. (2019). Development and validation of an SNP genotyping array and construction of a high-density linkage map in castor. *Scientific Reports, 9*(1), 3003. https://10.1038/s41598-019-39967-9

Sowder, A. (2021). *Southeast U.S., California Berries — Blackberries and Raspberries.* https://www.thepacker.com/markets/marketing/southeast-us-california-berries-blackberries-and-raspberries

Stoner, G. D., & Aziz, R. M. (2007). Prevention and therapy of squamous cell carcinoma of the rodent esophagus using freeze-dried black raspberries. *Acta Pharmacologica Sinica, 28*(9), 1422-1428. https://10.1111/j.1745-7254.2007.00686.x

Strik, B. C., Clark, J. R., Finn, C. E., & Banados, M. P. (2007). Worldwide blackberry production. *HortTechnology (Alexandria, Va.), 17*(2), 205-213. https://10.21273/HORTTECH.17.2.205

Sullivan, J. A., & Privé, J. P. (2001). Yield stability indices of primocane-fruiting red raspberry (*Rubus idaeus* L.). *Canadian Journal of Plant Science, 81*(2), 297-301. https://10.4141/P00-065

Willman, M., Bushakra, J. M., Bassil, N. V., Finn, C. E., Dossett, M., Perkins-Veazie, P., Bradish, C. M., Fernandez, G. E., Weber, C. A., Scheerens, J., Dunlap, L., & Fresnedo-Ramírez, J. (2020). Genetic analysis of drupelet count in black raspberry (*Rubus occidentalis*). *Acta Horticulturae,* (1277), 65-72. https://10.17660/ActaHortic.2020.1277.9

Zafra-Rojas, Q., Cruz-Cansino, N., Delgadillo-Ramírez, A., Alanís-García, E., Añorve-Morga, J., Quintero-Lira, A., Castañeda-Ovando, A., & Ramírez-Moreno, E. (2018). Organic acids, antioxidants, and dietary fiber of Mexican blackberry (*Rubus fruticosus*) residues cv. Tupy. *Journal of Food Quality, 2018*, 1-9. https://10.1155/2018/5950761

Zurn, J. D., Driskill, M., Jung, S., Main, D., Yin, M. H., Clark, M. C., Cheng, L., Ashrafi, H., Aryal, R., Clark, J. R., Worthington, M., Finn, C. E., Peace, C., Iezzoni, A., & Bassil, N. (2020). A Rosaceae family-level approach to identify loci influencing soluble solids content in blackberry for DNA-informed breeding. *G3 (Bethesda, Md.), 10*(10), 3729-3740. https://10.1534/g3.120.401449

CHAPTER 2

Genotype × Environment Analysis of *Rubus spp.* Fruit size, Plant architecture, and Yield Components

Abstract

This genotype × environmental (G × E) study focused on phenotyping a set of traits on five seminal genotypes in five geographically distinct environments over two years. The genotypes evaluated included 'Heritage', a primocane-fruiting raspberry, 'Latham' a floricane-fruiting raspberry, 'Prime-Ark® 45', a primocane-fruiting blackberry, 'Chester Thornless', a floricane fruiting blackberry, and 'Bristol' a black raspberry. Our team at NCSU collated and then analyzed the data from all sites. Floricane fruit size was significantly affected by the interaction of G, E, and harvest date (H). The three-way interaction encouraged further investigation into individual climate variable effects on fruit size. An interaction between genotype and growing degree days (GDD) was found to impact floricane fruit size. Analysis of the primocane fruit size indicated a significant interaction between G and E. The primocane blackberry maintained the same fruit size across all E, while the primocane raspberry fruit size differed by E. Floricane fruiting lateral counts were affected by an interaction between G and E. Contrasts indicated that primocane genotypes had higher counts of floricane fruiting laterals. The floricane berries per fruiting lateral analysis also found the same interaction but contrasts indicate the primocane fruiting blackberries produced significantly less fruit per lateral than the floricane genotypes. The length of dormant primocanes and their node densities were also analyzed. No interaction was found, and the main effects of G and E were significant for both. The Arkansas plots produced the longest primocanes and 'Heritage' produced significantly shorter canes than other genotypes. Node density evaluations determined a clear separation between blackberry and raspberry.

Raspberry genotypes had significantly more nodes per 50 cm segment. In the genotype × environment study, G and E were significant sources of variation for all of the traits phenotyped.

Introduction

The genus *Rubus* is genetically very diverse, and wild species can be found across North America (Hummer, 1996). However, the number of species utilized in breeding programs is limited. In general, black and red raspberries are grown in moderate temperate climates and fresh market blackberries are adapted to warmer regions. Breeding programs to date have focused on developing cultivars for these regions with a limited genetic background. Breeding for increased adaptation and important horticultural traits would increase the production capacity.

Exploration of the effects of genotype, environment, and their interactions (G × E) in *Rubus* has previously focused on one commodity per evaluation and has not been as diverse in environments and traits evaluated. The majority of this area of study has focused on red or black raspberries individually. Yield stability has been assessed in red raspberry (*R. idaeus* L.) for three cultivars across fourteen environments. This study produced an ANOVA evaluating the effects of genotype, environment, and their interactions on yield. The most stable of the three genotypes was 'Heritage', which is one of the selected genotypes in the G × E evaluation in progress (Sullivan and Prive, 2001). A G × E evaluation encompassing a set of environments and traits comparable to the current study was conducted by the USDA in Oregon over a three-year period. This study focused on two full-sibling populations of black raspberries (*R. occidentalis* L.) derived from USDA plant materials, providing information on the behavior of black raspberry crops specifically (Bushakra et al., 2016). This same population was evaluated for

drupelet counts that worked to identify the most stable genotypes over the range of environments to correlate the behavior with a genic region that controls for this trait (Willman et al., 2020). The current G × E study includes similar traits but expands into plant architecture and evaluates multiple species of *Rubus*, providing more general information about the genus.

Genotype × environment (G × E) evaluations provide information on trait plasticity and genotype performance in a range of environments. A single genotype can result in a variety of phenotypes depending on the environment in which it is cultivated. Genotype does not provide a set value of a trait but rather a range of expressions possible, especially with quantitative traits like yield (Nyquist and Baker, 1991). A narrow possible range of expression indicates a greater effect of genotype. Genotypic variance quantifies the variability caused by genotype, providing a measure of stability for a trait (Gonçalves et al., 2013). The phenotypic expression of a genotype can be heavily influenced by environment. Evaluating the effect of different environments requires consistent management across environments so the variability measured is not actually the effect of different cultivation practices. Consistent or predictable results across a range of environments is ideal for development of cultivars in diverse production regions. Genotype-by-environment interactions are determined utilizing the assessment of a particular genotype in an individual environment (Nyquist and Baker, 1991). The G × E variation accounts for the portion of phenotypic variability that is not explained by the average performance of the genotype across environments, or the average performance of the environment across genotypes (Hill, 2010). Identifying the proportion of effect genotype, environment, and their interactions have on phenotype improves estimation of trait performance in a variety of environments.

In 2020, public berry breeders at Cornell University, North Carolina State University, University of Arkansas, University of British Columbia/BC Berry Cultivar Development Inc., and the United States Department of Agriculture (USDA) joined the private companies Plant Sciences Inc and the initiator of this collaboration, Pairwise, in a collaborative effort to uncover the genetic potential in *Rubus*. Pairwise spearheaded this project to facilitate further research into these important crops. One of the projects in this public-private collaboration was to develop baseline phenomics data for a set of important traits in *Rubus* at 6 different geographic environments (Figure 2.1) in a G × E study.

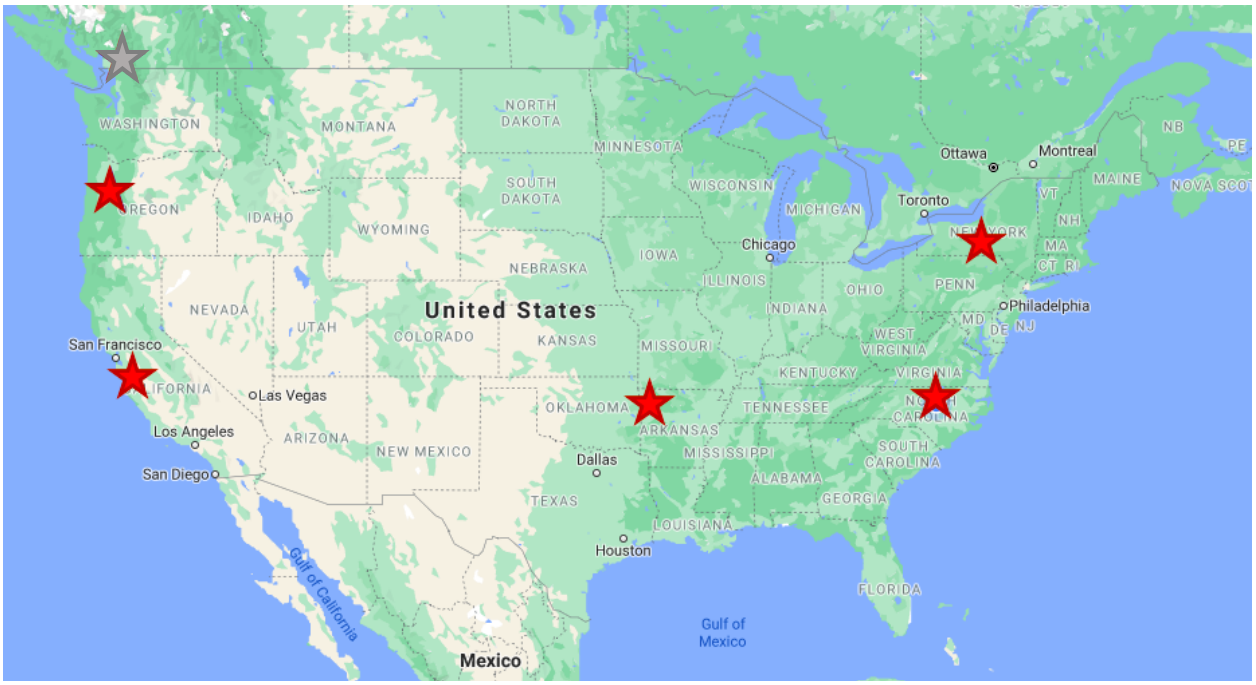

Figure 2.1. Distribution of evaluation environments participating in the genotype × environment study, although due to COVID complications the B.C. environment was removed.

The aim of the evaluation in progress is to provide information on the adaptability of economically important traits and the effects of genotype, environment, and their interactions. The range of environments that encompass the primary North American public breeding

programs and production regions has not been conducted in a single evaluation. There also has previously been a limited focus on the traits of interest, with minimal evaluation of plant architecture. Comparing genotypes of different species of *Rubus* is unique, especially considering the scope of environments and traits.

Materials and Methods

Seminal genotypes representative of the diversity of cultivated *Rubus* were selected. Two blackberries, two red raspberries, and one black raspberry were designated. Blackberries include the primocane-fruiting genotype 'Prime-Ark® 45', and the floricane-fruiting 'Chester Thornless'. Raspberries include the floricane-fruiting red raspberry 'Latham', floricane-fruiting black raspberry 'Bristol', and the primocane-fruiting red raspberry 'Heritage'. The environments evaluated were based on the presence of major breeding programs and production regions represented by collaborators. These environments included Salisbury, NC (35.6978,-80.6298), Watsonville, CA (36.9451,-121.7652), Corvallis, OR (44.5590,-123.2325), Fayetteville, AR (35.5302,-93.4011), and Geneva, NY (42.8717,-77.0410). The site in British Columbia was not included as they were unable to get plants in the ground in 2020 due to border restrictions.

The experimental design was a randomized complete block design (RCBD), and the block was each environment. Tissue cultured plants were obtained from commercial nurseries and were set in the ground in the spring as soon as local conditions allowed. Each genotype was planted in six plant plots, 18 inches apart, with variable between row spacing in each environment. Plots were replicated three times at each location in a completely randomized design (CRD). Plantings were done in 2020 on March 17th in California, April 17th in Arkansas, April 23rd in Oregon, May

12th in North Carolina, and July 7th in New York. Collaborator plots were maintained and data was collected using the same protocols across all environments. The timing of data collection was determined by crop development and phenological markers rather than calendar date. Date of data collection varied by each environment for this reason. Pest control and cultural practices for each environment were executed based on regional production guides and UC ANR recommendations (Bolda et al., 2009; Fernandez et al., 2016). All plots were grown in-ground and without coverage except for the California location where plants were grown in substrate and under hoops.

Plant architecture traits

Dormant primocane length was collected in winter at the end of the second year, after the plots were dormant and most leaves had senesced. One representative primocane was selected from each plot, avoiding edge plants and plants with less than three canes. The selected cane was cut at the base, at the soil level with pruning shears. Canes were straightened out on the ground; excess branches were removed for ease of measurement. A tape measure was used to determine the length (cm) of the cane from the base to the topmost node; this value was recorded.

Node density was collected at the same time as the dormant primocane length values, utilizing the same representative cane. The bottom 10 cm of the cane were pruned off and discarded. The next 50 cm portion was measured and a cut was made at the top of this segment, and the portion above this cut was also discarded. The resulting 50 cm segment is the portion from which the node counts were taken. Petiole or leaf scars were counted to determine the number of nodes, and this value was recorded.

Yield indicators

Floricane fruiting lateral counts were collected at the 10% ripe floricane fruit date, which was determined to be when all the plants within a plot had 1-2 ripe fruit present per plant. Lateral counts were performed on a single representative cane from each plot, avoiding edge plants. A meter stick was used to determine the basal meter, the segment from which the data was collected. The floricane fruiting lateral counts were collected from the 30 cm segment immediately above the basal meter of the cane (Figure 2.2). Laterals producing any flowers or fruit within that segment were counted and recorded. If fruit were only being produced below one meter, the uppermost fruiting lateral was identified, and counts were collected from the 30 cm below it.

Floricane berries per fruiting lateral counts were collected at the same time as the fruiting lateral counts. The same representative cane was suggested for use, although any other representative cane could have been chosen. A representative lateral would be selected from the cane, usually 3-4 laterals down from the most apical fruiting lateral (Figure 2.2). Unopened flowers, open flowers, unripe fruit, and ripe fruit were recorded as a single value of berries per lateral.

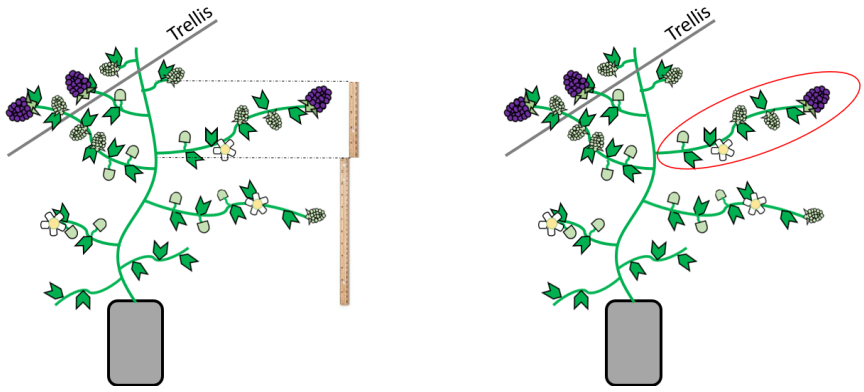

Figure 2.2. Visualization of fruiting lateral and berries per lateral identifying morphology utilized for count procedures, courtesy of Pairwise.

Fruit quality components

Fruit weights, recorded as a 10 berry mass, were collected three times throughout the season to capture early, mid, and late season fruit size. The dates of collection as well as the amount of time between collections was determined by genotype phenology. ‘Bristol’ only had a 3–4-week production window, so early, mid, and late season may have been as little as one week apart. Genotypes with longer harvest seasons had more spaced-out collection dates. At each collection date, 10 representative ripe berries were collected directly into labeled clamshells from each plot. Fruit were determined to be ripe and ready to harvest at the shiny black stage for blackberry, when red raspberries were fully red and easily detached from the torus, and when black raspberries were deep purple to black and also easily detached from the torus. The 10 berry mass (g) was collected at the time of harvest, and the date was recorded. This same procedure was conducted for the primocane fruiting genotypes; ‘Prime-Ark® 45’ and ‘Heritage’ when the primocane crop was being produced.

Preliminary analyses of fruit size encouraged investigation into climate variables and their effect on 10 berry mass. The longitude and latitudinal coordinates of each planting environment were used with the climate data collection platform (*Precisionsustainableag.org*) to download relevant climate data. Four climate variables were selected to include in the analysis: the highest recorded temperature, lowest recorded temperature, growing degree days (GDD), and chill hour accumulation (chill). The highest recorded temperature was determined to be the maximum high temperature between the planting date and the date of trait measurement. The lowest recorded temperature was defined as the minimum low temperature recorded between the vegetative bud break date and the date of trait measurement. The GDD value was calculated using the Base 50

model, starting accumulation at January 1st and ending at the date of trait measurement (Nugent, 2021; Vance et al., 2019).

Base 50 model:

$$GDD = \text{Avg daily temp} - 50 = [(\text{max temp} + \text{min temp})/2] - 50$$

GDD are cumulative, and negative values are not included in the summation (Nugent, 2021). Chill was measured using the 32-45 °F (0-7.2 °C) model starting on November 1st and ending at the recorded vegetative bud break date. The November 1st start date was selected because it is a traditionally used date that would allow for consistent calculations across environments (Glozer, 2021). Other models suggest an inception point of -2.2 °C, but that threshold was never met in the California and Oregon environments. Every hour between 0-7.2 °C added to the cumulative chill hour value. These climate values were used in the 10 berry mass analysis in place of environment to further investigate their impacts. Monthly averages for each site were recorded for each environment in Table 2.1.

Table 2.1. Climate effects monthly summary for each environment from planting date to final dormant cane collection date. Chill hours reported from November 1st, 2021 through the vegetative bud break date for the last plots at each environment.

Environment, Year	Cumulative Monthly GDD											
	January	February	March	April	May	June	July	August	September	October	November	December
North Carolina, 2020	-	-	-	-	349.6	729.1	981.8	923.2	602.7	442.4	199.1	13.5
North Carolina, 2021	16.5	25.5	169.6	268.1	508.0	780.3	920.5	951.3	681.0	496.6	77.6	152.6
Arkansas, 2020	-	-	-	129.0	486.1	782.0	969.8	861.2	644.1	339.6	165.8	14.8
Arkansas, 2021	7.7	17.1	180.8	245.5	460.7	799.9	908.7	948.0	727.1	468.6	83.8	56.9
New York, 2020	-	-	-	-	-	-	614.9	660.0	389.2	111.5	50.5	0.0
New York, 2021	0.0	0.0	14.5	72.5	230.3	571.4	595.7	730.4	419.8	258.6	0.2	0.6
Oregon, 2020	-	-	-	52.4	237.0	327.3	529.5	557.5	533.3	274.6	37.3	6.4
Oregon, 2021	0.1	0.4	4.9	107.6	219.8	488.4	622.8	642.6	493.2	150.8	68.7	-
California, 2020	-	-	49.8	321.3	522.2	653.3	729.2	867.5	820.2	795.6	280.6	151.4
California, 2021	171.3	112.8	119.1	307.1	486.8	682.8	826.5	838.5	724.4	471.9	362.3	91.4

	Cumulative Monthly Chill Hours											
	January	February	March	April	May	June	July	August	September	October	November	December
North Carolina, 2020	-	-	-	-	-	-	-	-	-	-	123	419
North Carolina, 2021	487	376	185	50	-	-	-	-	-	-	-	-
Arkansas, 2020	-	-	-	-	-	-	-	-	-	-	139	403
Arkansas, 2021	479	244	69	-	-	-	-	-	-	-	-	-
New York, 2020	-	-	-	-	-	-	-	-	-	-	379	304
New York, 2021	129	56	240	46	-	-	-	-	-	-	-	-
Oregon, 2020	-	-	-	-	-	-	-	-	-	-	254	356
Oregon, 2021	306	389	350	45	-	-	-	-	-	-	-	-
California, 2020	-	-	-	-	-	-	-	-	-	-	40	64
California, 2021	60	79	185	2	-	-	-	-	-	-	-	-

	Air Temperature (°C)																					
	January		February		March		April		May		June		July		August		September		October		November		December	
	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min
North Carolina, 2020	-	-	-	-	-	-	-	-	29.5	4.0	32.1	12.2	34.2	19.7	33.1	19.0	33.1	6.7	28.0	3.7	25.7	-0.8	18.4	-4.6
North Carolina, 2021	19.5	-2.5	23.7	-3.1	25.0	-0.9	28.3	-1.0	30.9	4.0	33.1	14.8	35.5	15.1	34.4	16.7	31.9	9.6	29.1	4.7	23.9	-3.8	22.6	-2.5
Arkansas, 2020	-	-	-	-	-	-	26.4	2.8	29.5	5.3	32.4	12.9	34.6	19.5	33.5	15.9	30.9	8.4	28.9	2.5	24.7	-2.7	19.4	-5.8
Arkansas, 2021	19.2	-2.6	21.0	-15.5	25.5	1.7	27.8	-1.0	29.1	8.0	33.2	13.4	34.3	16.9	34.2	19.1	33.5	9.9	29.6	6.5	24.6	-1.4	23.9	-1.8
New York, 2020	-	-	-	-	-	-	-	-	-	-	-	-	32.6	15.6	31.8	11.2	30.4	1.9	26.0	0.6	22.4	-3.1	7.5	-11.9
New York, 2021	4.5	13.5	3.2	-15.7	21.5	11.0	24.2	-4.8	28.9	2.0	31.9	7.1	30.2	10.5	31.5	10.7	29.6	7.2	24.6	4.5	15.6	-1.9	14.9	-4.3
Oregon, 2020	-	-	-	-	-	-	20.8	7.5	27.7	3.5	29.0	6.7	32.4	8.8	34.9	8.2	32.7	8.6	29.2	-1.6	19.9	0.1	16.1	-0.9
Oregon, 2021	15.1	-0.3	13.3	0.3	20.0	0.1	24.1	0.8	28.2	4.5	37.1	7.3	33.5	10.5	37.1	8.5	34.5	6.3	22.8	2.9	17.8	0.0	-	-
California, 2020	-	-	-	-	21.8	4.7	30.6	5.7	38.5	8.9	38.4	9.1	38.5	9.7	43.8	11.1	48.4	11.8	43.0	8.6	33.5	4.9	26.6	4.0
California, 2021	25.6	3.2	23.4	4.6	27.2	4.2	30.5	5.4	35.2	6.2	43.6	8.5	44.3	11.0	40.7	11.9	38.6	10.9	36.8	7.8	27.3	7.2	28.1	3.6

Statistical Analysis

The following models were constructed using the statistical software JMP (Raleigh, NC, USA). We used the restricted maximum likelihood (REML) method to determine if significant differences were present within our response variables in relationship to the main effects and their interactions ($\alpha = 0.05$). All the data sets recorded presented large differences in variance between genotypes, and justified Log transformations. The analysis utilized the transformed values to fit the models, but figures and tables include the untransformed means of the recorded values for easier interpretation.

Floricane fruit size analysis

The floricane 10 berry mass (g) was recorded as an indicator of fruit quality. The model below encompasses the effects of genotypes, environment, harvest, and their interactions on fruit size. Random error and error between replicates are accounted for. All five genotypes (i) and all five environments (h) are included in the analysis. The fruit size was recorded at an early, mid, and late point of the harvest season (k) for all three replicates (j) for each genotype at each location.

Genotype, environment, harvest, and all interactions:

$$Y_{hijk} = \mu + \theta_h + \alpha_i + (\theta\alpha)_{hi} + W_{(hi)j} + \tau_k + (\theta\tau)_{hk} + (\alpha\tau)_{ik} + (\theta\alpha\tau)_{hik} + E_{hijk} \quad (1)$$

$Y_{hijk} = ln($floricane 10 berry mass)
$\theta_h = h^{th}$ environment effect, $h = 1, 2, 3, 4, 5$
$\alpha_i = i^{th}$ genotype effect, $i = 1, 2, 3, 4, 5$
$j = 1, 2, 3$ replicates
$\tau_k = k^{th}$ harvest effect, $k = 1, 2, 3$

$E_{hijk} \sim^{iid} N(0, \sigma_E^2)$, E = random error for the k^{th} harvest for the hij plot
$W_{(hi)j} \sim^{iid} N(0, \sigma_W^2)$, W = plot to plot error term
} mutually independent

Preliminary analysis presented a complex 3-way interaction that was further explored to determine the effects of individual climate variables in each of these environments at the time of harvest. To further explore this interaction, a second model (2) was developed replacing environment with four climate variables: highest high, lowest low, growing degree days (GDD), and chill hours.

Genotype, harvest, climate variables, and interactions:

(2)

$$Y_{hijk} = \mu + \alpha_i + \tau_k + \beta_1 x_{hik1} + \beta_2 x_{hik2} + \beta_3 x_{hik3} + \beta_4 x_{hik4} + W_{(hi)j} + (\alpha\tau)_{ik} + \beta_{1i} x_{hik1} + \beta_{2i} x_{hik2} + \beta_{3i} x_{hik3} + \beta_{4i} x_{hik4} + \beta_{1k} x_{hik1} + \beta_{2k} x_{hik2} + \beta_{3k} x_{hik3} + \beta_{4k} x_{hik4} + \beta_{1ik} x_{hik1} + \beta_{2ik} x_{hik2} + \beta_{3ik} x_{hik3} + \beta_{4ik} x_{hik4} + E_{hijk}$$

Y_{hijk} = *ln(*floricane 10 berry mass)
h = 1, 2, 3, 4, 5 environments
α_i = i^{th} genotype effect, i = 1, 2, 3, 4, 5
j = 1, 2, 3 replicates
τ_k = k^{th} harvest effect, k = 1, 2, 3

x_{hik1} = high temp for k^{th} harvest of genotype i at environment h
x_{hik2} = low temp for k^{th} harvest of genotype i at environment h
x_{hik3} = GDD for k^{th} harvest of genotype i at environment h
x_{hik4} = chill for k^{th} harvest of genotype i at environment h
β_1 = coefficient for high temp
β_2 = coefficient for low temp
β_3 = coefficient for GDD
β_4 = coefficient for chill
β_{1i} = coefficient for high temp for genotype i
β_{2i} = coefficient for low temp for genotype i
β_{3i} = coefficient for GDD for genotype i
β_{4i} = coefficient for chill for genotype i
β_{1k} = coefficient for high temp at harvest k
β_{2k} = coefficient for low temp at harvest k
β_{3k} = coefficient for GDD at harvest k
β_{4k} = coefficient for chill at harvest k
β_{1ik} = coefficient for high temp at k^{th} harvest for genotype i
β_{2ik} = coefficient for low temp at k^{th} harvest for genotype i
β_{3ik} = coefficient for GDD at k^{th} harvest for genotype i
β_{4ik} = coefficient for chill at k^{th} harvest for genotype i

$E_{hijk} \sim^{iid} N(0, \sigma_E^2)$, E = random error for the k^{th} harvest for the hij plot
$W_{(hi)j} \sim^{iid} N(0, \sigma_W^2)$, W = plot to plot error term
} mutually independent

A lack of significance was found for most of the effects which encouraged the creation of a simplified model that eliminated all the insignificant interactions. The main effects of G, H, and GDD remained in the model along with the G × GDD two-way interaction.

Reduced genotype, harvest, climate variables, and interactions:

$$Y_{hijk} = \mu + \alpha_i + \tau_k + \beta_3 x_{hik3} + W_{(hi)j} + \beta_{3i} x_{hik3} + E_{hijk} \quad (3)$$

Y_{hijk} = *ln(*floricane 10 berry mass)
h = 1, 2, 3, 4, 5 environments
$\alpha_i = i^{th}$ genotype effect, i = 1, 2, 3, 4, 5
j = 1, 2, 3 replicates
$\tau_k = k^{th}$ harvest effect, k = 1, 2, 3

x_{hik3} = GDD for k^{th} harvest of genotype i at environment h
β_3 = coefficient for GDD
β_{3i} = coefficient for GDD for genotype i

$E_{hijk} \sim^{iid} N(0, \sigma_E^2)$, E = random error for the k^{th} harvest for the hij plot } mutually independent
$W_{(hi)j} \sim^{iid} N(0, \sigma_W^2)$, W = plot to plot error term

Primocane fruit size analysis

The primocane 10 berry mass (g) was recorded as an indicator of fruit quality. The model utilized is the same as model 1 and encompasses the effects of genotype, environment, harvest, and their interactions on primocane fruit size. Random error and error between replicates are accounted for. For this analysis Y_{hijk} = *ln(*primocane 10 berry mass), $\theta_h = h^{th}$ environment effect, h = 1, 2, 3, and $\alpha_i = i^{th}$ genotype effect, i = 1, 2. The environments included California, North Carolina, and Arkansas. Genotypes included the two primocane-fruiting cultivars: the red raspberry ‘Heritage’, and the blackberry ‘Prime-Ark® 45’. Error terms, replicates and harvests remained the same between analyses, with j = 1, 2, 3 replicates and $\tau_k = k^{th}$ harvest effect, k = 1, 2, 3.

Individual climate variables at each of the environments were again incorporated into the analysis to explore potential impacts. Because of the smaller sample size and fewer degrees of freedom, the primocane analysis could only include two of the four climate variables tested in the floricane fruit size analysis. GDD and chill were selected as the more important values to incorporate into the model based on knowledge that both are known to contribute to fruit development.

Genotype, harvest, climate variables (GDD and chill), and interactions:

$$Y_{hijk} = \mu + \alpha_i + \tau_k + \beta_3 x_{hik3} + \beta_4 x_{hik4} + W_{(hi)j} + (\alpha\tau)_{ik} + \beta_{3i} x_{hik3} + \beta_{4i} x_{hik4} + \beta_{3k} x_{hik3} + \beta_{4k} x_{hik4} + \beta_{3ik} x_{hik3} + \beta_{4ik} x_{hik4} + E_{hijk} \quad (4)$$

Y_{hijk} = *ln(*10 berry mass)
h = 1, 2, 3 environments
α_i = i^{th} genotype effect, i = 1, 2
j = 1, 2, 3 replicates
τ_k = k^{th} harvest effect, k = 1, 2, 3

x_{hik3} = GDD for k^{th} harvest of genotype i at environment h
x_{hik4} = chill for k^{th} harvest of genotype i at environment h

β_3 = coefficient for GDD
β_4 = coefficient for chill
β_{3i} = coefficient for GDD for genotype i
β_{4i} = coefficient for chill for genotype i
β_{3k} = coefficient for GDD at harvest k
β_{4k} = coefficient for chill at harvest k
β_{3ik} = coefficient for GDD at k^{th} harvest for genotype i
β_{4ik} = coefficient for chill at k^{th} harvest for genotype i

$E_{hijk} \sim^{iid} N(0, \sigma_E^2)$, E = random error for the k^{th} harvest for the *hij* plot } mutually independent
$W_{(hi)j} \sim^{iid} N(0, \sigma_W^2)$, W = plot to plot error term

Floricane fruiting lateral count analysis

Floricane fruiting lateral counts were recorded from a representative lateral from each plot as an indicator of yield potential. The model below encompasses the effects of genotype, environment

and their interaction on the number of fruiting laterals. Random error and error between replicates are accounted for. This model will continue to be the guide for analyzing the other yield components as well as the plant architecture values. Yield component analyses include all five genotypes and four environments: California, North Carolina, New York, and Arkansas.

Genotype, environment, and their interaction:

$$Y_{hij} = \mu + \theta_h + \alpha_i + (\theta\alpha)_{hi} + W_{(hi)j} + E_{hij} \quad (5)$$

Y_{hijk} = ln(lateral count)
θ_h = h^{th} environment effect, h = 1, 2, 3, 4
α_i = i^{th} genotype effect, i = 1, 2, 3, 4, 5
j = 1, 2, 3 replicates

$E_{hij} \sim^{iid} N(0, \sigma_E^2)$, E = random error for the j^{th} replicate for the hi plot
$W_{(hi)j} \sim^{iid} N(0, \sigma_W^2)$, W = plot to plot error term
} mutually independent

Floricane berries per lateral count analysis

Floricane berries per lateral counts were recorded as another indicator of yield potential. Model 5 was utilized for the analysis, where Y_{hijk} = $ln($berry count). This encompasses the effects of genotype, environment and their interaction on the number of berries per lateral. Random error and error between replicates are accounted for. The environments and genotypes included are the same as those in the number of floricane fruiting laterals analysis.

Dormant primocane length analysis

Dormant primocane lengths (cm) were recorded as an indicator of plant architecture. Model 5 was utilized for the analysis, where Y_{hijk} = $ln($cane length). This encompasses the effects of genotype, environment and their interaction on the dormant primocane length. Random error and

error between replicates are accounted for. Plant architecture component analyses include all five genotypes and four environments: North Carolina, New York, Oregon, and Arkansas.

Node count in 50 cm section of cane analysis

The number of nodes in 50 cm section of a predetermined section of the dormant cane were recorded as an indicator of plant architecture. The model utilized is constructed the same as model 5, but $Y_{hijk} = ln($node count). This encompasses the effects of genotype, environment and their interaction on the number of nodes within a 50 cm segment. Random error and error between replicates are accounted for. The environments and genotypes included are the same as those in the dormant primocane length analysis.

Results

Floricane fruit size

The analysis of model 1 generated an effects test output indicating all main effects and all but one interaction had a significant effect on floricane 10 berry mass ($\alpha = 0.05$) (Table 2.2). The 3-way interaction between environment, genotype, and harvest is the most complex and requires further investigation into the relationship between these variables and berry weight.

Table 2.2. REML analysis of the Log values for floricane fruit size found all main effects and interactions to be significant on floricane fruit size except for the Genotype*Harvest interaction (α=0.05, n=222).

Source	Nparm	DF	DFDen	F Ratio	Prob > F
Environment	4	4	44.62	36.8552	<.0001*
Genotype	4	4	44.29	630.4330	<.0001*
Environment*Genotype	16	16	44.4	7.2487	<.0001*
Harvest	2	2	101.7	51.3949	<.0001*
Environment*Harvest	8	8	101.7	7.3008	<.0001*
Genotype*Harvest	8	8	101.7	1.3948	0.2077
Environment*Genotype*Harvest	32	32	101.7	4.0849	<.0001*

A line graph was generated to visualize the difference in genotype fruit sizes throughout the harvest season at each environment using the recorded, untransformed values (Figure 2.3). There is a clear separation of blackberries 'Prime-Ark® 45' and 'Chester Thornless' consistently producing larger fruit than the raspberries 'Heritage', 'Latham', and 'Bristol'. There is variability in fruit size among the genotypes at different harvests across environments. For some genotypes there is a peak in fruit size mid-season in some environments while fruit may downsize mid-season for the others. Some genotypes also showed more linear trends in fruit size throughout the season. The degree to which the fruit size changes over the season also differed for each genotype and environment.

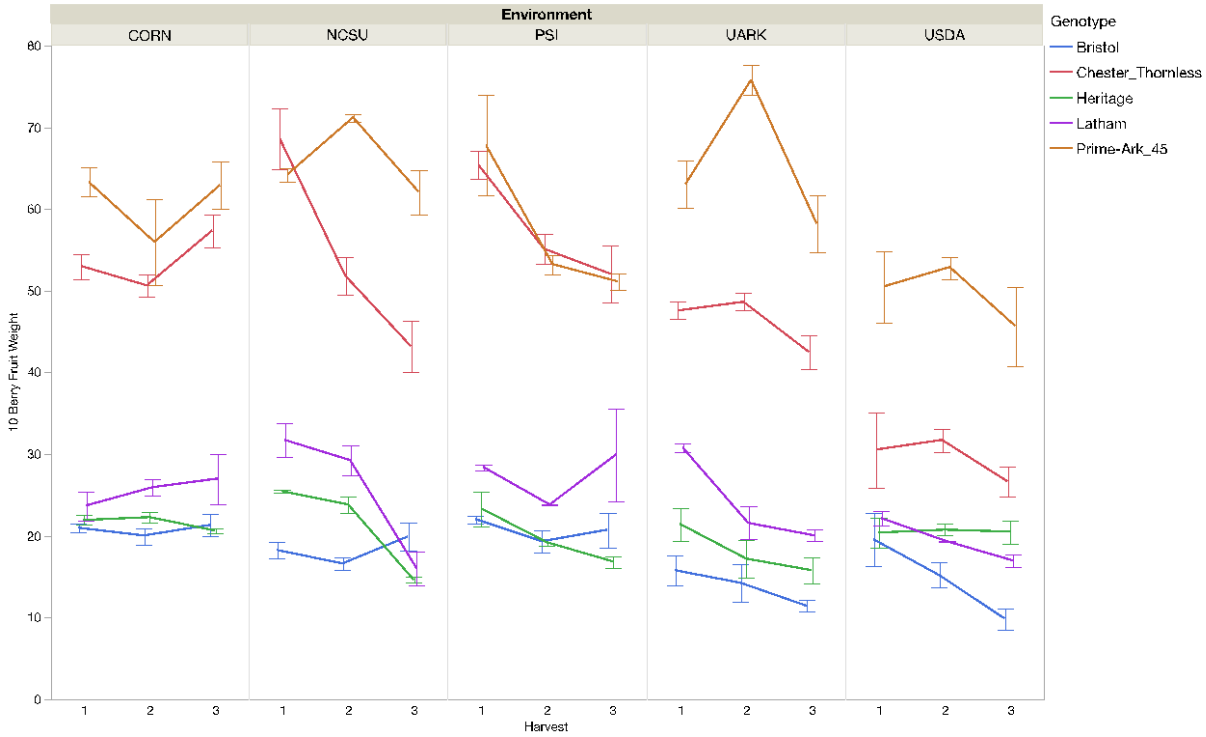

Figure 2.3. Mean 10 berry fruit size values in each environment over the three harvests. Genotype is indicated by line color with a key on the right side. Environment separates the chart into columns, and environment names are presented at the top of each column. Error bars indicate standard error (α=0.05, n=222).

The complexity present in the 3-way interaction inspired further investigation into the effects of individual climate variables at each of these environments at the time of harvest. The analysis of

model 2 produced an effects test output replacing environment with climate variables. This output indicated that the main effects of G and H were significant as well as the interaction between G and GDD (p=0.0007) (Table 2.3). The 2-way interaction between genotype and GDD is the focus while all the other interactions were considered insignificant (α = 0.05). The lack of significance encouraged the creation of a simplified model, eliminating all the insignificant interactions.

Table 2.3. REML analysis of the Log values for floricane fruit size including climate variables found the main effects of genotype and harvest to be significant for floricane fruit size as well as the interaction between genotype and GDD (α=0.05, n=222).

Source	Nparm	DF	DFDen	F Ratio	Prob > F
Genotype	4	4	75.41	43.0134	<.0001*
Harvest	2	2	134.4	5.8704	0.0036*
High	1	1	71.86	0.3817	0.5386
Low	1	1	87.5	0.6702	0.4152
GDD	1	1	116.5	2.5824	0.1108
Chill	1	1	83.5	0.0001	0.9904
Genotype*High	4	4	63.66	1.2287	0.3076
Genotype*Low	4	4	64.28	0.8836	0.4789
Genotype*GDD	4	4	76.54	5.4454	0.0007*
Genotype*Chill	4	4	65.48	2.1640	0.0828
Genotype*Harvest	8	8	135.5	1.4595	0.1777
Harvest*High	2	2	117.1	1.0291	0.3605
Harvest*Low	2	2	118.4	0.2845	0.7529
Harvest*GDD	2	2	128.6	1.1726	0.3128
Harvest*Chill	2	2	120.3	2.1459	0.1214
Genotype*Harvest*High	8	8	110.6	1.1715	0.3227
Genotype*Harvest*Low	8	8	110.8	1.2819	0.2601
Genotype*Harvest*GDD	8	8	131.3	0.5718	0.7995
Genotype*Harvest*Chill	8	8	113.9	1.0824	0.3805

The simplified model (3) generated an effects test output indicating previously significant effects remained significant. The main effects of G and H as well as the interaction between G and GDD were still considered significant (p=0.0279) (Table 2.4). The 2-way interaction prompted a post hoc analysis.

Table 2.4. REML analysis of the Log values for floricane fruit size reduced model including climate variables found the main effects of genotype and harvest to be significant for floricane fruit size as well as the interaction between genotype and GDD (α=0.05, n=222).

Source	Nparm	DF	DFDen	F Ratio	Prob > F
Genotype	4	4	63.51	123.8397	<.0001*
Harvest	2	2	166.5	15.0593	<.0001*
GDD	1	1	109.1	0.0402	0.8415
Genotype*GDD	4	4	193.5	2.7842	0.0279*

Prediction expressions generated using JMP were used to further explore the interaction between the nominal variable of genotype and the continuous variable of GDD. Creation of these expressions used the Log transformed values, providing insight into general relationship trends between different genotypes and GDD (see Appendix A). The formulas produced a negative relationship between red raspberry fruit size and GDD, indicating fruit size decreases as GDD increases. The blackberry and black raspberry genotypes fruit sizes increased as GDD increased (Figure 2.4).

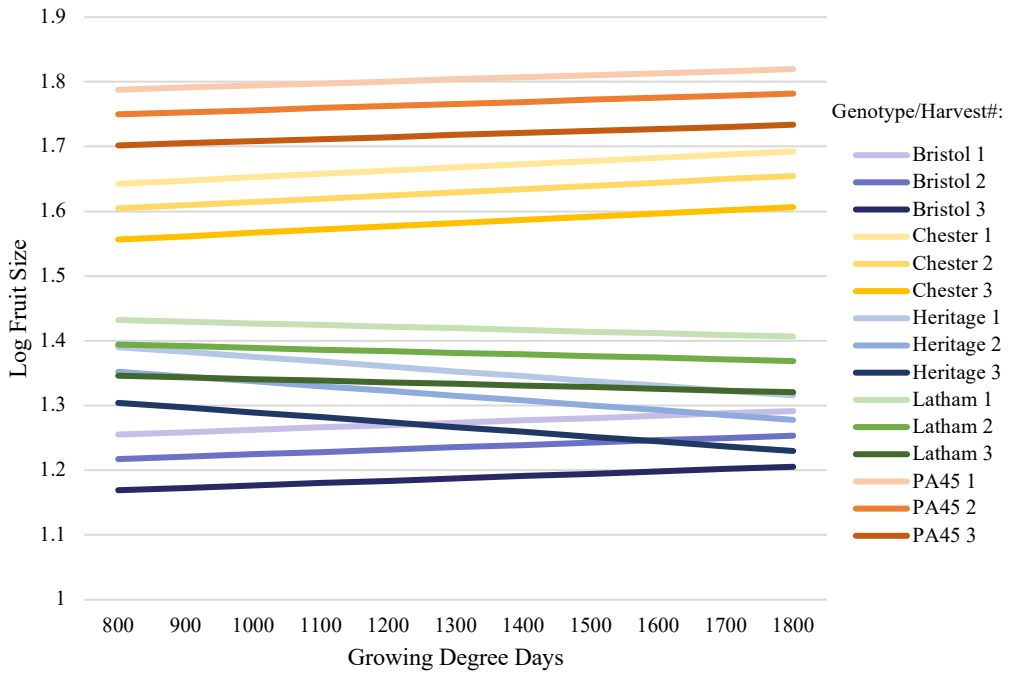

Figure 2.4. Plotting prediction estimates using transformed values produces positive slopes for blackberries ('Chester Thornless', 'Prime-Ark® 45') and black raspberry ('Bristol') genotypes, and negative slopes for red raspberries ('Heritage', 'Latham') as GDD increase.

Primocane fruit size

The analysis using model 1 produced an effects test output indicating all the main effects and the interaction between environment and genotype were considered significant for primocane fruit size ($p<0.0001$) (Table 2.5). The 2-way interaction prompted post hoc analysis to further sparse out differences.

Table 2.5. REML analysis of the Log values for primocane fruit size found all main effects to be significant as well as the interaction between environment and genotype ($\alpha=0.05$, $n=53$).

Source	Nparm	DF	DFDen	F Ratio	Prob > F
Environment	2	2	12.19	5.7100	0.0178*
Genotype	1	1	12.2	796.9943	<.0001*
Environment*Genotype	2	2	12.19	24.7658	<.0001*
Harvest	2	2	23.87	23.5873	<.0001*
Environment*Harvest	4	4	23.84	2.5476	0.0657
Genotype*Harvest	2	2	23.87	2.1230	0.1417
Environment*Genotype*Harvest	4	4	23.84	1.7119	0.1804

To dissect the interaction between environment and genotype, a Tukey HSD differences of means test was used to compare the Log values and produce a connecting letters report. Untransformed means are again utilized in determining significant differences. There is a significant separation among the genotypes. Similar to the floricane fruit size data, blackberries are consistently larger than raspberries. ‘Prime-Ark® 45’ shows consistent fruit size across all environments, while ‘Heritage’ produced significantly smaller fruit in North Carolina compared to other environments. This relationship is displayed in the line graph (Figure 2.5), showing the size difference between ‘Prime-Ark® 45’ and ‘Heritage’ at each environment. The difference in environment performance for ‘Heritage’ encouraged further analysis on the effects of individual climate variables at each of the environments.

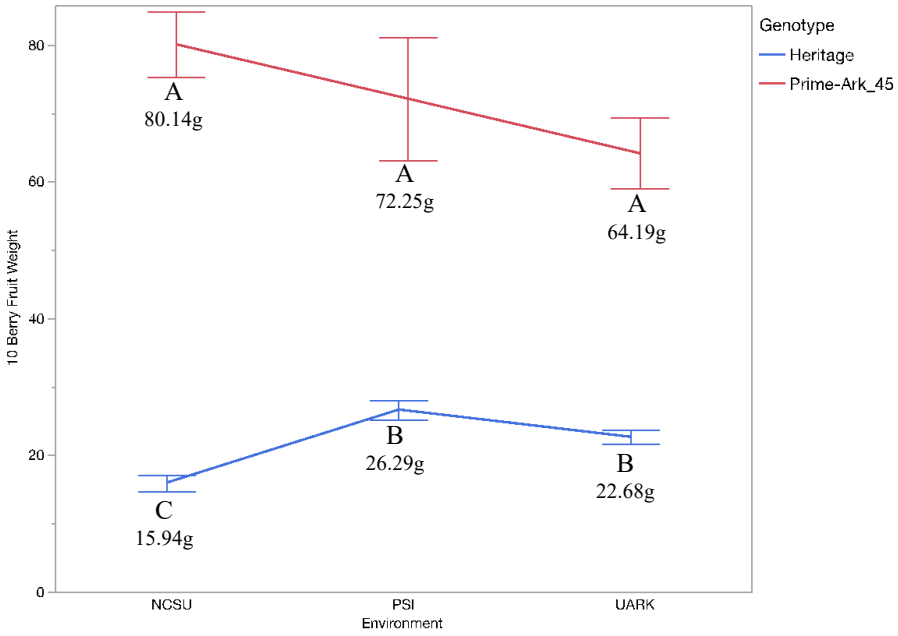

Figure 2.5. Mean 10 berry fruit size values at each environment. Genotype is indicated by line color with a key on the right side. Error bars indicate standard error. Tukey HSD post hoc analysis output for differences of means indicated by a connecting letters report, showing a separation in primocane fruit size between genotypes and environments (α=0.05, n=53).

Model 4 analysis generated an effects test output indicating the main effects of genotype and harvest were significant. The interactions between genotype and harvest as well as genotype and the climate variables were also significant (Table 2.6). The 2-way interaction required post hoc analysis.

Table 2.6. REML analysis of the Log values for primocane fruit size found the main effects of genotype and harvest to be significant for primocane fruit size as well as the interactions between genotype x harvest, GDD, and chill (α=0.05, n=53).

Source	Nparm	DF	DFDen	F Ratio	Prob > F
Genotype	1	1	34.95	150.6316	<.0001*
Harvest	2	2	29.92	10.9289	0.0003*
GDD	1	1	16.77	1.7744	0.2007
Chill	1	1	14.22	0.0033	0.9549
Genotype*GDD	1	1	16.77	8.7477	0.0089*
Genotype*Chill	1	1	14.22	21.5119	0.0004*
Genotype*Harvest	2	2	29.92	7.5329	0.0022*
Harvest*GDD	2	2	26.66	0.1772	0.8386
Harvest*Chill	2	2	24.72	0.2023	0.8182
Genotype*Harvest*GDD	2	2	26.66	1.1195	0.3413
Genotype*Harvest*Chill	2	2	24.72	1.7549	0.1938

To dissect the interaction between genotype and harvest, a Tukey HSD differences of means test was used to compare the Log values and produce a connecting letters report (Table 2.7). Means of the recorded values are included in this output but were not used in establishing the differences of means. There is a significant separation among the genotypes as expected. 'Prime-Ark® 45' mean fruit size decreased in the third harvest, while 'Heritage' maintained fruit size throughout the season.

Table 2.7. Tukey HSD post hoc analysis output for differences of means indicated by a connecting letters report, showing a separation in primocane fruit size between genotypes and harvest (α=0.05, n=53).

Level				Least Sq Mean (Log)	10 Berry Mass Mean
Prime-Ark_45,1	A			2.0736830	115.41755
Prime-Ark_45,2	A			1.8420671	70.41415
Prime-Ark_45,3		B		1.6014480	38.71901
Heritage,2			C	1.3350881	21.96623
Heritage,1			C	1.3300068	22.59898
Heritage,3			C	1.3013657	20.31832

Climate variable effects were found to be significant, indicating that GDD and chill did affect primocane 10 berry mass. In the model (4) that included the climate data, the genotype × harvest interaction effect was also considered significant, while it was insignificant in the model (1) utilizing the location effect instead of climate data. The original analysis better minimizes the variance, and therefore should be considered the superior model. Other components that would cumulatively result in the environment effect are missing, potentially other climate variables, soil factors, or field conditions. When the environment is fully accounted for, the genotype × environment interaction is significant while others are not. The original model better accounts for the differences present and is the preferred model for interpretation, while model 4 provided support that climate variables had an effect on primocane fruit size.

Floricane fruiting lateral count

Model 5 produced an effects test output indicating the main effects of environment and genotype were considered significant as well as their interaction on the number of floricane fruiting laterals (Table 2.8). The 2-way interaction prompted post hoc analysis.

Table 2.8. REML analysis of the Log values for floricane fruiting lateral counts found both main effects to be significant as well as the interaction between environment and genotype (α=0.05, n=59).

Source	Nparm	DF	DFDen	F Ratio	Prob > F
Environment	3	3	39	22.8026	<.0001*
Genotype	4	4	39	8.8589	<.0001*
Environment*Genotype	12	12	39	2.0227	0.0485*

To dissect the interaction between environment and genotype, a Tukey's HSD differences of means test was used to compare the Log values and produce a connecting letters report (Table 2.9). Again, means of the recorded, untransformed values are included in this output for easier interpretation.

Table 2.9. Tukey HSD post hoc analysis output for differences of means indicated by a connecting letters report, showing a separation in floricane fruiting lateral counts between genotypes and environments (α=0.05, n=59).

Level						Least Sq Mean (Log)	Flori Lateral Count Mean
CORN,Bristol	A					0.96	9.33
CORN,Heritage	A	B				0.90	8.00
CORN,Chester_Thornless	A	B	C			0.86	7.33
UARK,Bristol	A	B	C			0.86	7.33
NCSU,Bristol	A	B	C	D		0.83	7.00
CORN,Prime-Ark_45	A	B	C	D		0.82	6.67
CORN,Latham	A	B	C	D		0.82	6.67
NCSU,Prime-Ark_45	A	B	C	D		0.82	6.67
NCSU,Heritage	A	B	C	D		0.78	6.33
UARK,Prime-Ark_45	A	B	C	D		0.77	6.00
PSI,Heritage	A	B	C	D	E	0.73	5.33
UARK,Heritage		B	C	D	E	0.70	5.00
NCSU,Latham		B	C	D	E	0.69	5.00
NCSU,Chester_Thornless		B	C	D	E	0.67	4.67
PSI,Bristol			C	D	E	0.63	4.33
PSI,Prime-Ark_45			C	D	E	0.63	4.33
PSI,Latham			C	D	E	0.60	4.00
UARK,Latham				D	E	0.59	4.00
PSI,Chester_Thornless				D	E	0.59	4.00
UARK,Chester_Thornless					E	0.48	3.00

As there was a large amount of overlap between each genotype and environment, a series of contrasts were made to better parse out the differences. For the effect of environment, two contrasts were made determining the plots at Cornell ($p<0.0001$) and NCSU ($p=0.0003$) both had a higher number of fruiting laterals than PSI. Potentially the substantially colder winters in these environments contributed to a greater number of flowering laterals developing. Another contrast comparing the blackberry and red raspberry genotypes found no significant differences in the number of fruiting laterals ($p=0.4165$). Contrasting the primocane and floricane genotypes indicated that the primocane genotypes produced more fruiting laterals. 'Heritage' had significantly more fruiting laterals than 'Latham' ($p=0.0070$). 'Prime-Ark® 45' produced significantly more fruiting laterals than 'Chester Thornless' ($p=0.0019$).

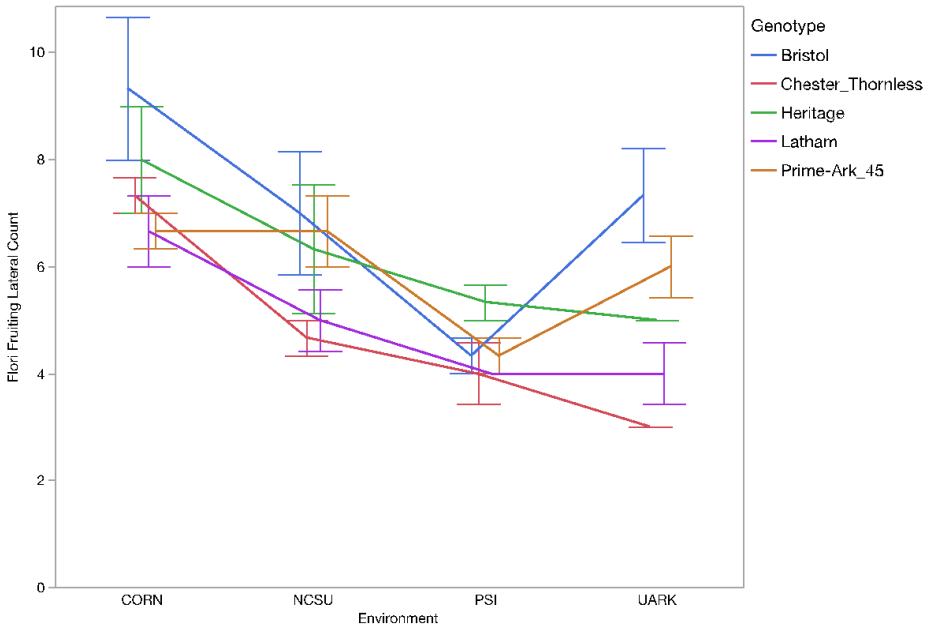

Figure 2.6. Mean number of floricane fruiting laterals at each environment. Genotype is indicated by line color with a key on the right side. Error bars indicate standard error ($\alpha=0.05$, $n=59$).

A line graph was made using the untransformed values to help visualize the general genotype trends in floricane fruiting lateral counts across environments (Figure 2.6). This supports the contrast results; the California environment (PSI) had lower fruiting lateral counts compared to

other environments. It also appears that ‘Latham’ and ‘Chester Thornless’ tend to have fewer fruiting laterals than other genotypes, further agreeing with the contrast showing that the floricane genotypes had lower counts than the primocane genotypes.

Floricane berries per lateral count

Model 5 generated an effects test output indicating the main effects of environment and genotype were considered significant as well as their interaction on the number of berries per lateral (Table 2.10). The 2-way interaction resulted in post hoc analysis.

Table 2.10. REML analysis of the Log values for floricane berries per lateral counts found both main effects to be significant as well as the interaction between environment and genotype (α=0.05, n=59).

Source	Nparm	DF	DFDen	F Ratio	Prob > F
Environment	3	3	39	52.3381	<.0001*
Genotype	4	4	39	19.6616	<.0001*
Environment*Genotype	12	12	39	3.8236	0.0007*

To dissect the interaction between environment and genotype, a Tukey HSD differences of means test was used to compare the Log values and produce a connecting letters report (Table 2.11). Again, means of the recorded values are included in this output for easier interpretation.

Table 2.11. Tukey HSD post hoc analysis output for differences of means indicated by a connecting letters report, showing a separation in floricane berries per lateral counts between genotypes and environments (α=0.05, n=59).

Level							Least Sq Mean (Log)	Berries per Lateral Mean
NCSU,Chester_Thornless	A						1.51	32.33
NCSU,Latham	A	B					1.46	28.67
UARK,Latham	A	B					1.42	27.00
UARK,Heritage	A	B	C				1.32	22.33
PSI,Latham	A	B	C				1.31	20.50
NCSU,Heritage	A	B	C				1.25	18.00
NCSU,Prime-Ark_45	A	B	C				1.25	18.33
PSI,Chester_Thornless	A	B	C	D			1.24	18.67
NCSU,Bristol	A	B	C	D			1.24	17.67
UARK,Chester_Thornless	A	B	C	D	E		1.22	16.67
PSI,Heritage	A	B	C	D	E		1.17	15.00
UARK,Bristol	A	B	C	D	E		1.14	14.00
PSI,Prime-Ark_45		B	C	D	E		1.08	12.00
CORN,Chester_Thornless			C	D	E		0.96	9.67
CORN,Heritage			C	D	E		0.95	9.00
CORN,Latham				D	E		0.87	7.33
CORN,Bristol					E		0.84	7.00
UARK,Prime-Ark_45					E		0.84	7.00
PSI,Bristol					E		0.84	7.00
CORN,Prime-Ark_45						F	0.36	2.67

There is a large amount of overlap between each genotype and environment. To better dissect the differences, a series of contrasts were made. For the effect of environment, two contrasts were made determining the plots at NCSU ($p<0.0001$) and PSI ($p<0.0001$) both produced more floricane berries per lateral than Cornell. The earlier spring in North Carolina and California may have contributed to more successful bud development. A contrast between blackberries and red raspberries indicated blackberries produced significantly fewer berries per fruiting lateral ($p<0.0001$). Contrasting primocane and floricane genotypes indicated primocane blackberries had less floricane fruit per lateral ($p<0.0001$), while primocane and floricane raspberries showed no significant difference ($p=0.1025$).

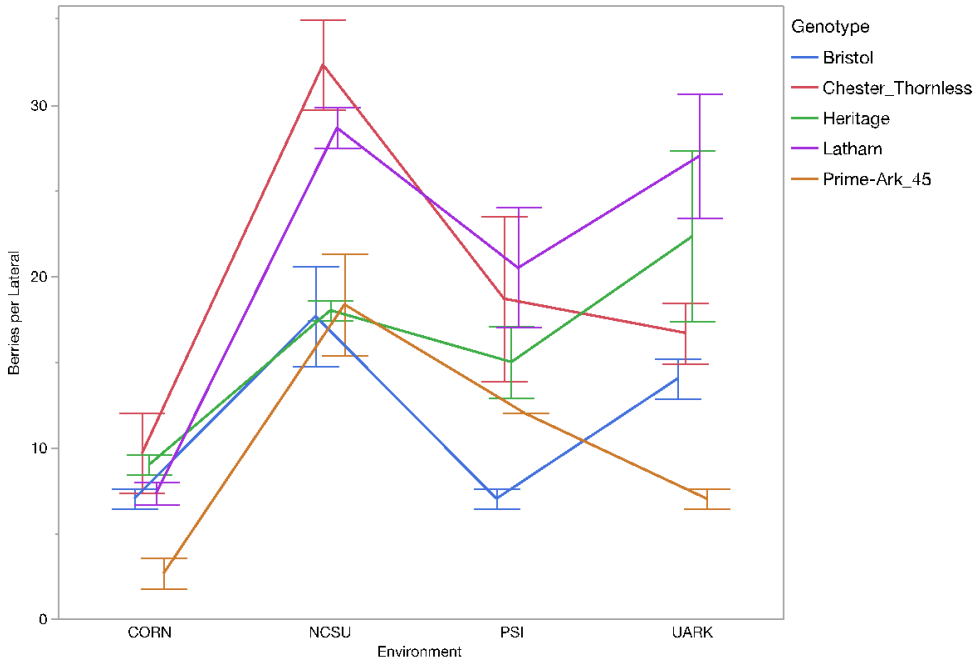

Figure 2.7. Mean number of floricane berries per laterals at each environment. Genotype is indicated by line color with a key on the right side. Error bars indicate standard error (α=0.05, n=59).

A line graph was produced using the untransformed values to help visualize the general genotype trends in the number of berries per lateral across environments (Figure 2.7). Although the contrast p-value was too large to determine a difference between berries per lateral for the primocane and floricane red raspberries, the visualization shows floricane genotypes produced more berries per lateral in all but one environment (CORN). The contrast indicating the lower number of berries per lateral at Cornell compared to NCSU and PSI is also captured in this figure.

Dormant primocane length

Model 5 generated an effects test output indicating the main effects of environment and genotype were considered significant for dormant primocane length, while the interaction was not (Table 2.12).

Table 2.12. REML analysis of the Log values for dormant cane length (cm) found both main effects to be significant but the interaction between environment and genotype was insignificant (α=0.05, n=60).

Source	Nparm	DF	DFDen	F Ratio	Prob > F
Environment	3	3	40	15.0845	<.0001*
Genotype	4	4	40	24.1135	<.0001*
Environment*Genotype	12	12	40	1.6973	0.1042

To look at the effect of environment on dormant cane length, a Tukey HSD differences of means test was used to compare the Log values and produce a connecting letters report. NCSU and University of Arkansas reported significantly longer canes than the USDA location, while Cornell produced an intermediate length cane. The Arkansas plots produced longer canes than the other locations in the test. A bar graph using the untransformed values is provided to help visualize the differences described (Figure 2.8).

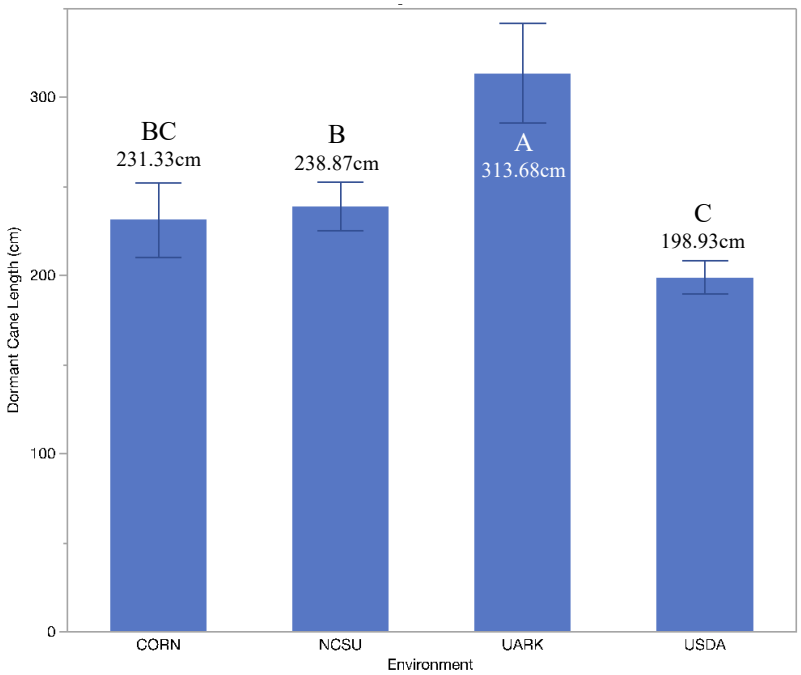

Figure 2.8. Mean primocane dormant cane length (cm) at each environment. Error bars indicate standard error. Tukey HSD post hoc analysis output for differences of means indicated by a connecting letters report, showing a separation in dormant cane length by environment (α=0.05, n=60).

To look at the effect of genotype on dormant cane length, a Tukey HSD differences of means test was used to compare the Log values and produce a connecting letters report. ‘Bristol’ and

'Chester Thornless' had significantly longer canes than 'Latham' and 'Heritage', while 'Prime-Ark® 45' produced an intermediate length cane. A bar graph using the untransformed values is provided to help visualize the differences described (Figure 2.9).

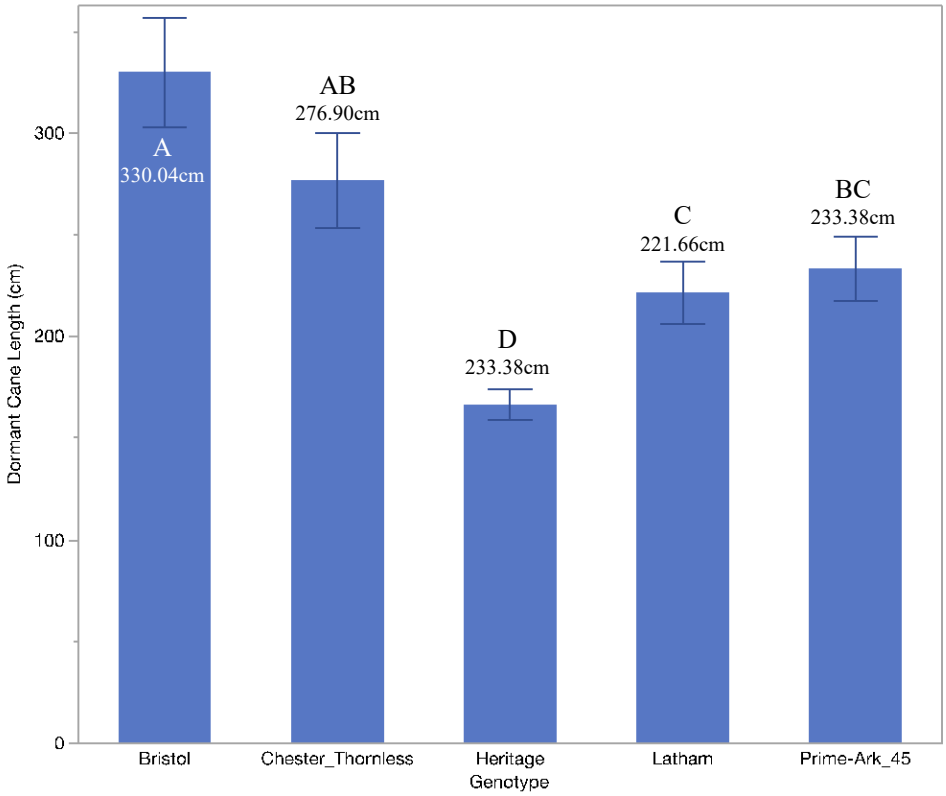

Figure 2.9. Mean primocane dormant cane length (cm) for each genotype. Error bars indicate standard error. Tukey HSD post hoc analysis output for differences of means indicated by a connecting letters report, showing a separation in dormant cane length by genotype (α=0.05, n=60).

Node count in 50 cm section of cane

Model 5 produced an effects test output; the main effects of environment and genotype were significant for node density, while the interaction was not (Table 2.13).

Table 2.13. REML analysis of the Log values for the node counts within 50cm found both main effects to be significant but the interaction between environment and genotype was insignificant (α=0.05, n=60).

Source	Nparm	DF	DFDen	F Ratio	Prob > F
Environment	3	3	40	9.3191	<.0001*
Genotype	4	4	40	26.0488	<.0001*
Environment*Genotype	12	12	40	1.3646	0.2230

To look at the effect of environment on node density, a Tukey HSD differences of means test was used to compare the Log values and produce a connecting letters report. Means of the recorded values are included in the bar graph for easier interpretation, but were not used in establishing the differences of means. The Arkansas environment recorded higher node density than USDA and NCSU, while Cornell produced an intermediate node density. A bar graph using the untransformed values is provided to help visualize the differences described (Figure 2.10).

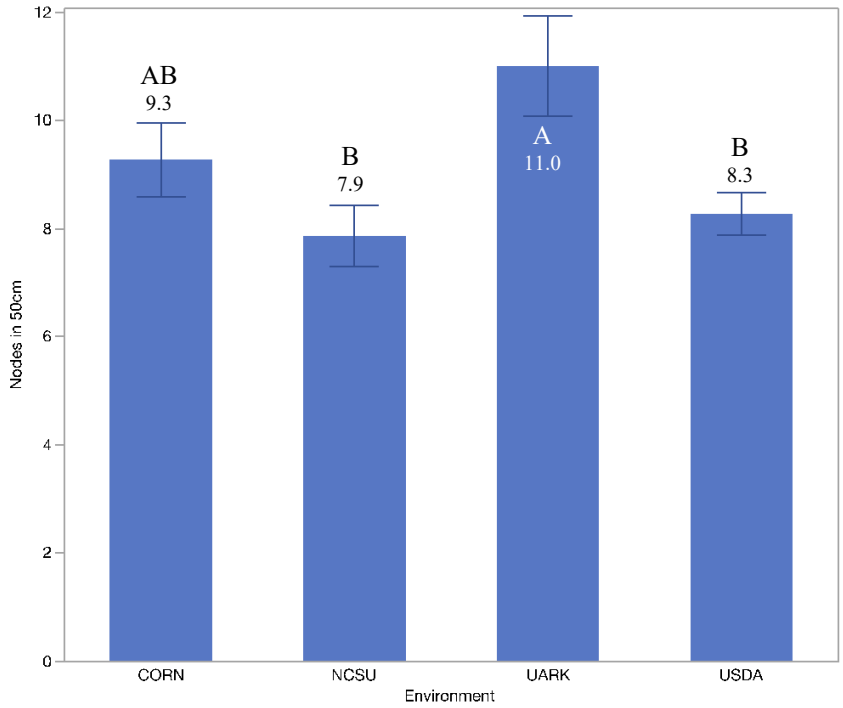

Figure 2.10. Mean node count per 50cm at each environment. Error bars indicate standard error. Tukey HSD post hoc analysis output for differences of means indicated by a connecting letters report, showing a separation in node counts by environment (α=0.05, n=60).

To look at the effect of genotype on node density, a Tukey HSD differences of means test was used to compare the Log values and produce a connecting letters report. There is a clear separation between blackberries and raspberries. ‘Bristol’, ‘Heritage’, and ‘Latham’ all had significantly more nodes in a 50cm segment than the blackberries ‘Chester Thornless’ and

'Prime-Ark® 45'. A bar graph using the untransformed values is provided to help visualize the differences described (Figure 2.11).

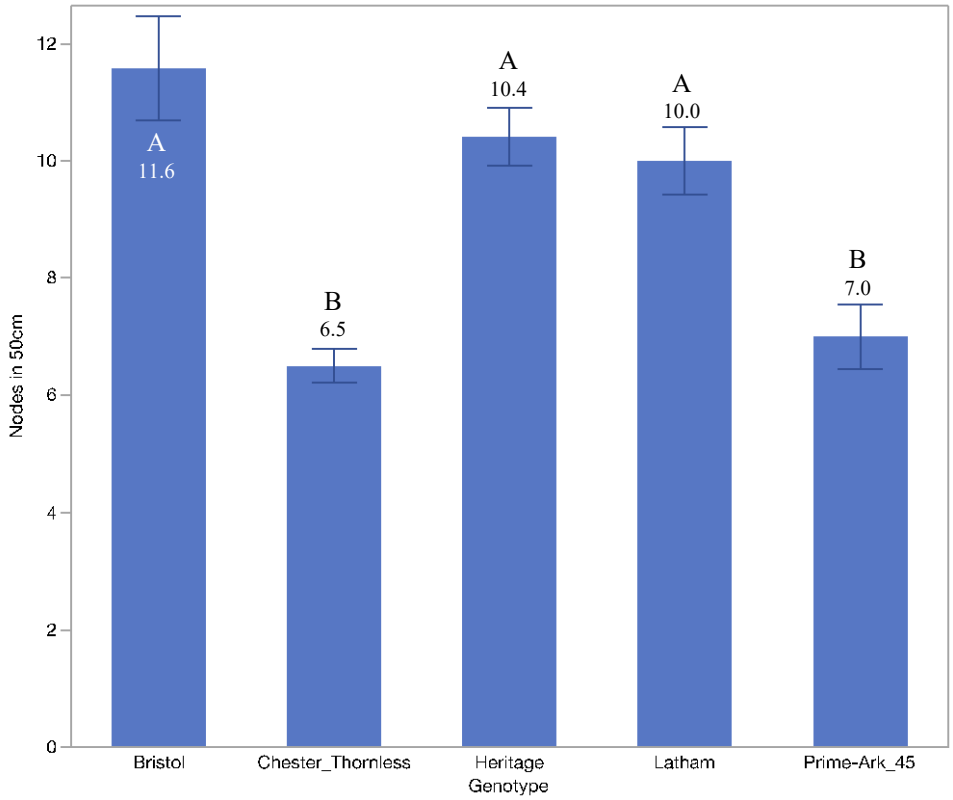

Figure 2.11. Mean nodes per 50cm for each genotype. Error bars indicate standard error. Tukey HSD post hoc analysis output for differences of means indicated by a connecting letters report, showing a separation in node counts by genotype (α=0.05, n=60).

Discussion

Production of *Rubus* crops occurs in a broad range of environments yet genotype response to environment is generally uncharacterized. The objective of this G × E study was to determine the degree to which different sources of variation affect the traits evaluated. Investigating the sources affecting phenotypic plasticity indicates the stability and the level of genetic control for traits of interest. For breeding efforts, a higher degree of consistency in performance is preferred (Bernardo, 2020). The study evaluates sources of variation relevant to major Rubus crops and

production regions to allow for generalizations of performance to be made for crop categories by growing region and provide indications of plasticity.

The evaluation of floricane fruit size that incorporated the climate variables into the model provided interesting insight into the differences between the responses of raspberry and blackberry to increased GDD. Raspberries are typically not heat tolerant and suffer from high temperatures (Ballington and Fernandez, 2008). The correlation between GDD and genotype produced prediction expressions that indicate a negative association between red raspberry fruit size and GDD. This supports the existing knowledge that raspberries lack heat tolerance. In contrast, blackberries and the black raspberry genotype were positively associated with increased GDD (see Appendix A).

Yield components analyses produced interesting results. A contrast of the number of floricane fruiting laterals between primocane and floricane genotypes indicated primocane genotypes had higher values. This initially was perplexing, because it would be reasonable to assume genotypes that only produced one crop would be more productive in that one crop than the double cropping genotypes. The evaluation of berries per node provided more insight. The floricane genotypes tended to produce more berries per fruiting lateral than the primocane fruiting genotypes. The primocane genotypes had more fruiting laterals but the laterals were less productive. The complementary nature of these two data sets is also helpful in understanding the differences found between locations. It is notable that the CA location produced fewer fruiting laterals since CA is known for its large yields. Although the number of fruiting laterals was smaller, the berries

per lateral values were intermediate. This fact coupled with the longer production season in the temperate central coast of CA may help explain the disconnect.

Plant architecture-related traits produced effects tests that found only the main effects of G and E to be significant. The dormant cane length values show the black raspberry and blackberry genotypes tended to produce longer canes than red raspberry. Arkansas produced significantly longer canes than any other location along with higher node counts per 50 cm than other locations. It is interesting to note that NC and AR have relatively similar environments, but the plant architecture values are quite different. In contrast, the fruit size and yield components between the two environments were similar. Arkansas did have a greater number of GDD in March 2021 and an early vegatative bud break date that month, while NC bud break did not occur until April. The longer growing season could potentially explain the longer canes produced in AR. The nodes per 50 cm counts were also higher in AR than NC, indicating denser foliage. This may also be related to GDD. Although AR had higher GDD and earlier vegetative bud break in March, NC had higher monthly GDD for the rest of the year. The longevity of the active growth season may be responsible for the longer canes in AR, while the sustained heat may have contributed to elongation of internode spacing in NC. It would be interesting to revisit these analyses and incorporate the climate variables utilized in the fruit size analysis.

The traits measured displayed variability in plasticity when comparing analyses. The majority of the traits evaluated produced effects tests with the highest F-values being contributed by genotype. The large F-values associated with genotype implies a large amount of genetic control for fruit size and plant architecture components. The yield components, including number of

floricane fruiting laterals and berries per lateral, had higher F-values associated with environment rather than genotype. This indicates that the environment had a larger degree of control over yield components compared to the plant architecture and fruit size values.

The G × E evaluation experiment provides baseline information on the performance of seminal cultivars for growers in these regions as well as breeding programs. Considering the influence of environment on yield components, this variability may be useful in cultivar selection for a specific location. For breeding purposes, the determination of the degree of genetic control may assist in decision making for improvement efforts. A measure of trait stability may allow for more efficient evaluation of genotypes. Evaluation of traits with a large amount of genetic control may require less replication in a range of environments. The plant architecture and fruit size components' effects outputs producing large F-values for G indicates these traits will be less variable than the yield components across environments. Evaluation of these traits in fewer environments would be sufficient for determining values for these characteristics. Yield component evaluations would benefit from replication in a range of different environments to build a true understanding of the potential range in performance.

LITERATURE CITED

Ballington, J. R. (2016). The history of blackberry and raspberry breeding in the southern USA. *Acta Horticulturae,* (1133), 13-22. https://10.17660/ActaHortic.2016.1133.3

Ballington, J. R., & Fernandez, G. E. (2008). Breeding raspberries adapted to warm humid climates with fluctuating temperatures in winter. *Acta Horticulturae,* (777), 87-90. https://10.17660/ActaHortic.2008.777.9

Bernardo, R. (2020). *Breeding for quantitative traits in plants* (3rd ed.). Stemma Press.

Bolda, M. P., Joseph, S. V., Daugovish, O., Gaskell, M., Koike, S. T., Bettiga, L. J. & Gubler, W. D. (2020). *UC IPM Pest Management Guidelines: Caneberries.* https://www2.ipm.ucanr.edu/agriculture/caneberries/

Dossett, M., Lee, J., & Finn, C. E. (2010). Variation in anthocyanins and total phenolics of black raspberry populations. *Journal of Functional Foods, 2*(4), 292-297. https://10.1016/j.jff.2010.10.004

Fernandez, G. (2021). *Pest Management Strategic Plan for Blackberry in Alabama, Arkansas, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, and Virginia.* IPM Data: https://ipmdata.ipmcenters.org/source_report.cfm?view=yes&sourceid=1444

Fernandez, G., Garcia, E. & Lockwood, D. (2016). *Southeast Regional Caneberry Production Guide.* https://content.ces.ncsu.edu/southeast-regional-caneberry-production-guide

Food and Agriculture Organization of the United Nations. (2020). http://www.fao.org/faostat/en/#data/QC/visualize

Giovanelli, G., Limbo, S., & Buratti, S. (2014). Effects of new packaging solutions on physico-chemical, nutritional and aromatic characteristics of red raspberries (*Rubus idaeus* L.) in postharvest storage. *Postharvest Biology and Technology, 98*, 72-81. https://10.1016/j.postharvbio.2014.07.002

Glozer, K. (2006). Rest-breaking alternatives for sweet cherry and an update on chill accumulation. https://ucanr.edu/sites/fruitandnut/files/67019.pdf

Hummer, K. E. (1996). *Rubus* diversity. *HortScience, 31*(2), 182-183. https://10.21273/HORTSCI.31.2.182

Kaume, L., Howard, L. R., & Devareddy, L. (2012). The blackberry fruit: A review on its composition and chemistry, metabolism and bioavailability, and health benefits. *Journal of Agricultural and Food Chemistry, 60*(23), 5716-5727. https://10.1021/jf203318p

Nugent, J.*Calculating Growing Degree Days*. Michigan State University:

Overview of the Caneberry Industry: Facts & Figures. (2017). https://www.raspberryblackberry.com/consumers/overview-of-the-caneberry-industry-facts-figures/

Precision Sustainable Ag. https://api.precisionsustainableag.org/weather

Strik, B. C., Clark, J. R., Finn, C. E., & Banados, M. P. (2007). Worldwide blackberry production. *HortTechnology (Alexandria, Va.), 17*(2), 205-213. https://10.21273/HORTTECH.17.2.205

Vance, A. J., Jones, P. A., Finn, C. E., & Strick, B. C. (2019). Fruit development in blackberry types and cultivars – impact of days and temperature from bloom to stages of ripening. *Journal of American Pomological Society, 73*(4), 227-239. https://raspberryblackberry.com/wp-content/uploads/Fruit-Development-in-Blackberry-Types-and-Cultivars-Impact-of-Days-and-Temperature-Image.pdf

Weber, C., & Hai Liu, R. (2002). Antioxidant capacity and anticancer properties of red raspberry. *Acta Horticulturae 585,* , 451-457. 10.17660/ActaHortic.2002.585.73

Worthington, M. L., & Clark, J. R. (2020). Development of blackberry cultivars with novel plant architecture. *Acta Horticulturae,* (1277), 159-164. https://10.17660/ActaHortic.2020.1277.23

Wu, Y., Zhang, C., Huang, Z., Lyu, L., Li, J., Li, W., & Wu, W. (2021). The color difference of *Rubus* fruits is closely related to the composition of flavonoids including anthocyanins. *Food Science & Technology, 149*, 111825. https://10.1016/j.lwt.2021.111825

CHAPTER 3

Rubus spp. SNP Validation and Association with Prickle-Free and Primocane-Fruiting Phenotypes

Abstract

Although popular fruits in retail and local markets, the diversity of raspberry and blackberry has been limited by the scant germplasm pool used in breeding programs. North Carolina State University, Cornell University, University of Arkansas, BC Berry Cultivar Development Inc., and the United States Department of Agriculture (USDA) collaborated with Pairwise Inc. (hereafter Pairwise), and Plant Sciences Inc. (PSI hereafter), in a series of research projects that focused on *Rubus,* the genus that includes blackberries and raspberries. A total of 31 *Rubus spp.* genotypes were previously sequenced by Pairwise. The Illumina short reads were aligned to the black raspberry (*R. occidentalis*) genome to identify single nucleotide polymorphisms (SNP) markers. A subset of SNPs was used to validate the markers in a biological context and to identify the association of the phenotypes with genotypes. Traits evaluated and considered valuable to producers include the primocane fruiting and prickle-free phenotypes. Primocane fruiting SNPs were identified by associating the nucleotide polymorphisms with the fruiting phenotype. This yielded SNPs associated with three previously annotated genes involved in protein phosphorylation, DNA repair, and RNA splicing. SNPs associated with the presence or absence of prickles were selected after aligning the flanking sequences of SNPs (101bp) to the red raspberry candidate genes (*R. idaeus subsp. strigosus*) identified by Khadgi and Weber (2020). Three out of 98 tested SNPs (~3%), were found to be linked with prickle presence-related candidate genes. The candidate genes produce proteins associated with prickle development in other Rosaceae species. Validating the presence of these SNPs within a larger

population of *Rubus* may facilitate the development of reliable KASP markers linked to the desirable trait to be used in marker-assisted selection.

Introduction

Raspberries and blackberries (*Rubus spp.*) are desirable healthy snacks and popularity and production are increasing. Caneberry consumption is linked to a large range of health benefits, many attributed to the high anthocyanin content. The high concentration of anthocyanins is associated with antioxidant, anti-inflammatory, antiviral, antiproliferative, and anticarcinogenic effects (Weber and Hai Lui, 2002; Dossett, M. et al. 2010; Kaume, L., 2011) The health benefits of consumption and snackability of these sweet fruits are coupled with increasing production. Worldwide blackberry production tripled between 2006 and 2016 (NARBA, 2017). Global raspberry production value has also been steadily increasing. The gross production value for raspberry was $228,086,000 in 1991 and valued just under $2.5 billion in 2020 (FAOSTAT, 2020). This may be attributed to increased production and even shifts from processing fruit to more fresh fruit, increasing crop value. The three largest producers of caneberries in the US are Oregon, Washington, and California, with a cumulative 2015 value of 5.7, 38.1, and $44.5 million for black raspberry, blackberry, and red raspberries, respectively. The total value of these three crops for this production range was $88.3 million, and this value has likely only increased with the trend of production growth (NARBA, 2017). Total US production of raspberries grossed $468,423,000 in 2020 (FAOSTAT, 2020). According to FAOSTAT, the United States produced 106,423 tons during 2016-2018 in the US, representing 12.8% of the 830,544 tons of global production. This makes the US the fifth largest producer of raspberries (FAOSTAT, 2020). The United States produces an even greater proportion of blackberries. A 2005 report indicates the

US produced 35,097 tons (23%) of the 154,644 tons of *Rubus spp*. grown worldwide (Strik et al, 2007). The high value of these crops and the room for improvement justifies exploration into further crop improvements and continued research into how to make production more efficient.

Formal blackberry breeding efforts were first established in the early 1900's in the southern US. (Ballington, 2016). Red raspberries, the more popular commodity, experienced earlier domestication efforts beginning in the 1800's. The first commercially successful primocane-fruiting red raspberry, 'Heritage', was released in 1969 and remained a significant cultivar worldwide through the 1990's (Jennings, 1988). Breeding for the recessive primocane-fruiting trait began immediately, starting in blackberry. The University of Arkansas primocane-fruiting blackberry cultivars produce fruit on current season canes rather than in the second year, providing a harvest a year earlier than floricane-fruiting cultivars (Clark, 2008). In addition to the earlier production, the floricanes still are capable of producing fruit. The floricane crop is produced from late spring into summer while the primocane crop starts in late summer or fall. Primocane-fruiting cultivars can be mowed to reduce labor costs on pruning, but this comes at the cost of producing a floricane crop. The different harvest window for the primocane crop allows for production in more environments. Combining traits essential for commercial viability with primocane-fruiting phenotypes can be difficult, and improvement efforts are ongoing (Clark, 2008).

Prickle-free canes are another desirable recessive phenotype that remains as a common breeding objective (Coyner et al., 2005). Although 'prickle' is the botanically correct term for the spikey hardened trichomes, they are commonly referred to as thorns (Coyner, et al,. 2005). The prickles

present in *Rubus spp.* are considered modified glandular trichomes (Kellogg et al., 2011). The presence of prickles protects that plant from physical damage and herbivory. The same characteristics that keep the plant safe can make harvest more difficult. Thorns may slow down picking crews reaching though the canopy for fruit or cause injury. Developing linked markers to prickle-free genotypes would speed up the breeding process during early generation selection. One of the most abundant types of markers are Single Nucleotide Polymorphism (SNP). With the advent of Next Generation Sequencing (NGS), discovering millions of SNPs in a diversity panel is no longer insurmountable. However, not all markers are linked to the trait of interest nor are they correctly identified during bioinformatic data processing.

Validating the presence of meaningful SNPs identified bioinformatically in a broad range of diverse biological samples of *Rubus spp.* could facilitate creation of improved cultivars. Pairwise sharing sequencing data from public accession provided the information necessary to pursue these efforts. Selecting SNPs associated with the recessive primocane-fruiting and prickle-free phenotypes could lead to marker creation allowing for identification of genotypes maintaining these genes in the recessive state. *Rubus spp*. have a base chromosome number of 7 but range widely in level of ploidy. Red raspberries are diploid while blackberries can be diploid, tetraploid, octoploid and even dodecaploid (Meng and Finn, 1999). The potential for high ploidy levels can make it difficult to achieve the fully homozygous state necessary to express thornlessness or primocane-fruiting in blackberries (Coyner et al. 2005). Having molecular markers available would allow for earlier population screening, or even support genetic engineering approaches.

Materials and Methods

SNP discovery: Thirty-one public genotypes from five *Rubus* species were selected of the more than 300 public accessions sequenced and phenotyped by Pairwise. After trimming the adapters and primer sequences from the reads, the clean reads were mapped to the black raspberry genome (*R. occidentalis*) (VanBuren, 2016) using CLC Genomics Workbench V20.0 (Qiagen, Hilden, Germany 2021). The mapping parameters were 80% identity, 80% read length, and unique mapping.

After mapping each genotype to the black raspberry genome, a deduplication algorithm in CLC Genomics WB was used to remove duplicated reads. The mapping files with duplicated reads removed were exported as bam files and transferred into a Unix server. Subsequently, SAMtools (ver 0.1.7a) was used to merge all 31 bam files to be used to call the SNP for all possible variable positions corresponding to each accession. The SAMtools pile-up command was used to generate the pile-up files for the merge and each individual accession. The pile-up files were processed by Galaxy Pileup Parser (see Appendix B).

In-house Perl Scripts were used to filter and further process the parsed pile up files. The clean pile-up files were then merged by a Perl Script to generate the genotype matrix where the rows are markers and columns are genotypes, with their corresponding base composition and coverage. The SNP discovery algorithm is written in Perl requiring at least four genotypes to be homozygous and different. For instance, two accessions at any given position should show G and two others should be A. G and A have been used as an example here, but all transition or transversion comparisons were considered when putative SNP positions were filtered out of the

master genotyping table. Subsequently, the flanking sequences of SNPs were extracted from black raspberry genome using a Perl script. Approximately 3.5 million SNPs were discovered using SAMTools and in-house Perl scripts (see Appendix B).

SNPs in Prickle-Free Candidate Genes: The candidate genes for development of prickles were previously identified by Khadgi and Weber (2020) using RNA-Seq combined with phenotyping. All the genes were determined to be on chromosome 4. (Khadgi and Weber, 2020). The 10Kb upstream and downstream sequences of the candidate genes were kindly provided by A. Khadgi (Personal Communication). The flanking sequences of SNPs (101bp) were aligned to the flanking sequences of candidate genes using a locally installed BLAST program (https://www.ncbi.nlm.nih.gov/). Of the original 3.5 million SNPs, 1285 were colocalized with the flanking sequences of the candidate gene with 99% homology and 100bp coverage (101 minus the SNP site). An arbitrary threshold of maximum 20% missing data points across the diversity panel (6 out of 31) was set to further filter the SNPs with >20% missing data. A sliding window of 50bp was selected and the SNPs with more than 25 overlapping base pairs were eliminated. The final selection of 1 to 2 SNPs per associated gene were made, prioritizing ones with less missing data points across the diversity panel. Based on all applied filters, 96 SNPs were selected for validation using KASP (competitive allele-specific PCR) assays and analysis of association with the prickle-free phenotype.

Identification of Putative SNPs Linked to Primocane-Fruiting: In the sequenced diversity panel of 31 accessions, 9 accessions were known to be primocane-fruiting and 10 accessions were determined to be floricane-fruiting. An in-house Perl script was developed to select from the

initial 3.5 M segregating SNPs across the panel, SNP positions where primocane-fruiting genotypes were homozygous and different from homozygous alternate alleles in the floricane genotype group. The flanking sequences of 177 selected SNPs (Table 3.1) were provided to AgBiotech Inc. (Monterey, CA) for KASP marker analysis in the same diversity panel in addition to 14 blackberry genotypes.

Table 3.1. Summary of SNPs selected for validation by trait and their distribution across the 7 chromosomes of *Rubus occidentalis*.

Trait	SNPs Validated	Chr1	Chr2	Chr3	Chr4	Chr5	Chr6	Chr7
PRICKLE	96	0	0	0	96	0	0	0
PRIMO/FLORI	81	2	9	17	6	1	31	15
TOTAL	177	2	9	17	102	1	31	15

Analysis

Trait-Marker Analysis: Biologically validated SNPs returned from AgBiotech Inc. were further analyzed for meaningful correlations between phenotype and allele frequencies. Trait-based analyses using allele frequency differences for each marker have commonly been used in plant breeding to find associations between the markers and traits of interest (Foolad et al., 1997; Merk et al., 2012). Population allele frequency differences were calculated for each marker, comparing the allele frequencies of alternate phenotypes for both the prickle-free and primocane-fruiting populations. The value of the difference between alternate phenotype allele frequencies were compared to their 2x and 3x standard error confidence interval. SNPs with allele frequency differences greater than the population 3x standard error were further investigated for their biological function. SNPs with less than 5 genotypes data points for each phenotype were removed. The remaining SNPs were mapped back to the positioning within the genome and those associated with genes were further explored.

Annotating and functional analysis of candidate genes: Using OmicsBox (BioBam, 2019) formerly Blast2Go software, the function of the flanking sequences of SNPs were determined.

Results

Genes previously annotated with relevant associated products lead to a heightened interest in 3 primocane-fruiting related SNPs and 3 SNPs associated with the prickle-free phenotype (Figure 3.1). The SNPs associated with prickle formation were located on chromosome 4, agreeing with the previous findings in red raspberry. (Khadgi and Weber, 2020) The primocane-fruiting SNP loci were spread throughout the genome, but a cluster of significantly associated SNPs were found on chromosome 6 within one gene referred to as 182.37 (snap_masked-Ro06-processed-gene-182.37-mRNA-1). The other two primocane-associated SNPs with annotated gene products were found to be on chromosomes 3 and 7.

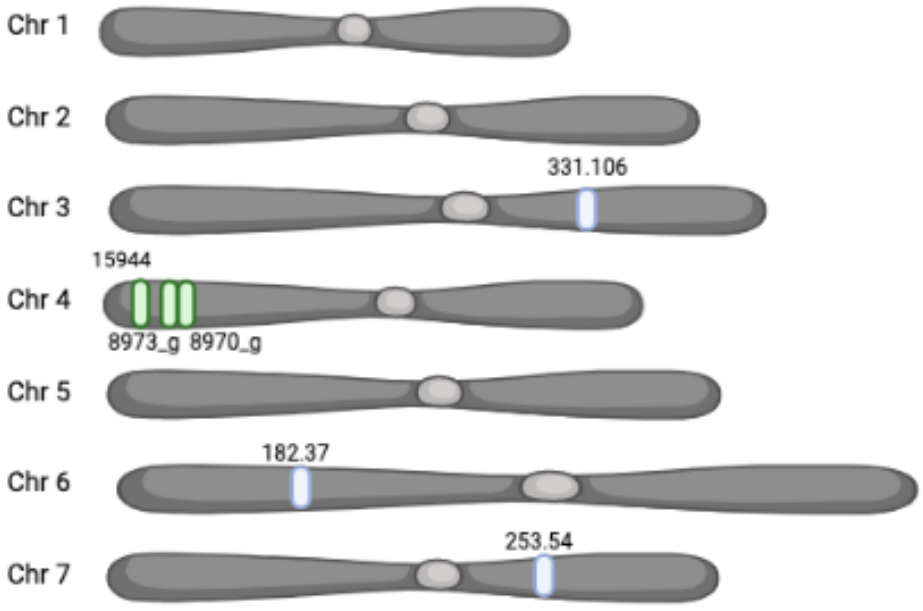

Figure 3.1. Visualization of the 7 base chromosomes in *Rubus* and the environment and genes containing selected SNPs associated with phenotypes of interest. Genes marked with green are linked to prickles, and blue indicates association with primocane fruiting.

KASP Marker Analysis for Prickle Trait: Of the 96 prickle related SNPs selected for validation and analysis, 75 markers were successfully read. Markers were considered successful when at least 5 genotypes from each phenotype had reads returned. Utilizing the known phenotypes of all 46 genotypes, 49 of the 75 successful prickle-related SNPs were determined to have allele frequency differences greater than 3 times their standard error. The SNP positions were mapped back to the genes previously identified by Cornell in red raspberry (Khadgi and Weber, 2020). Fourty-eight of these genes were annotated with associated gene products determined from research on a range of species. All but four of the SNP's associated products were determined from reference species within the family Rosaceae.

Individual products and their functions were investigated and provided a wide range of associated functions (see Appendix B). OmicsBox results indicated three of the significantly associated SNPs were located within genes attached to products previously found to determine epidermal cell identity and prickle formation (Table 3.2). The SNP identified as 1444893 was located within the gene labeled 15944. This gene sequence's inferred product was a WD40-like transcription factor established using OmicsBox. The 1469434 SNP located within 8973_g was determined to produce an MYB (myeloblastosis) transcription factor. The MYB and WD40 transcription factors along with a bHLH (basic helix-loop-helix) protein make a transcription activator-inhibitor complex which determines if epidermal cells form trichomes in a range of species (Swarnkar et al., 2021; Han et al, 2022; Ramsay and Glover, 2005). SNP 1469744 within the 8970_g gene was connected to the C2H2 family transcription factor. C2H2-like transcription factors have been connected to the initiation and development of trichomes in Arabidopsis and rose (Han et al, 2022; Swarnkar et al., 2021). The determination that three of the SNPs with

allele frequencies associated with thornlessness were located within genes identified to produce transcription factors that control prickle development is promising.

Table 3.2. The final prickle-free associated SNPs within annotated genes with products previously connected to trichome formation into prickles.

SNP name	Chr #	Gene	Associated Products	Species Referenced	Function
1444893	Ro04	15944	putative transcription factor WD40-like family	*Fragaria vesca subsp. vesca*	Cell differentiation
1469434	Ro04	8973_g	transcription factor MYB54	*Rosa chinensis*	Cell differentiation
1469744	Ro04	8970_g	putative transcription factor C2H2 family	*Rosa chinensis*	Cell differentiation

KASP Marker Analysis for Primocane-Fruiting Trait: Forty of the 81 primocane-fruiting related SNPs were successfully returned with at least 5 genotypes of each phenotype having alleles identified. Of the 40 successful SNPs associated with the known fruiting phenotype of 36 genotypes, 23 were selected for further analysis based on their allele frequency differences being greater than 3 times their standard error (see Appendix B). The SNP positions were used along with the *R. occidentalis* genome and Integrative Genomics Viewer (IGV 2.13.0) to determine if they were located within previously annotated genes. Only 9 of the 23 SNPs were positioned in or around annotated genes, and 7 of those SNPs were associated with the same gene on chromosome 6 (Table 3.3). All 3 genes have been annotated with associated products determined from *Rosa chinensis*, which is in the family Rosaceae (OmicsBox). The associated gene functions include the formation of mRNA, DNA integration and repair, and protein phosphorylation of DNA (Gish and Clark, 2011; UniProt, 2022; UniProt, 2022). The cluster of SNPs all associated with the gene labeled as 182.37 is promising. Two of the SNPs are immediately upstream of the gene, potentially affecting the promoter region. It is possible that this single gene could control the fruiting phenotype, considering primocane-fruiting has been

determined to be controlled by a single recessive gene (Lopez-Medina et al. 2000). SNP 2638853 is 6bp upstream of the annotated gene 182.37, and has perfect segregation of genotype by phenotype in raspberry. Most of the blackberry genotypes returned no allele reads for this SNP except for the floricane-fruiting 'Rio Grande', which followed the same allele pattern as the floricane-fruiting raspberries.

Table 3.3. The final primocane-fruiting associated SNPs within annotated genes and their products

SNP name	Chr #	Gene	Associated Products	Species Referenced	Function
1286318	Ro03	331.106	putative non-specific protein-tyrosine kinase RLK-Pelle-DLSV family	*Rosa chinensis*	Protein phosphorylation
2638849	Ro06	99bp before 182.37	putative ribonuclease H-like domain, reverse transcriptase zinc-binding domain-containing protein	*Rosa chinensis*	DNA integration DNA repair
2638853	Ro06	6bp before 182.37			
2638865	Ro06	182.37			
2638866	Ro06	182.37			
2638867	Ro06	182.37			
2638868	Ro06	182.37			
2638870	Ro06	182.37			
3305590	Ro07	253.54	U11/U12 small nuclear ribonucleoprotein 25 kDa protein	*Rosa chinensis*	RNA splicing

Discussion

The final SNPs of interest were selected based on phenotype associations and environment within previously annotated genes. Nearly all the prickle-associated SNPs were within previously annotated genes. This is not surprising considering the identification of the SNPs utilized genes previously linked to the trait of interest (Khadgi and Weber, 2020). The motivation for focusing on the three specific prickle SNPs is based on the associated gene products (OmicsBox). SNP 1444893 located within the gene labeled as 15944 was annotated from *Fragaria vesca subsp. Vesca*. This sequence was determined to produce a putative WD-40-like

transcription factor. WD40 repeat proteins have been found to have higher expression in *Rosa rugose* genotypes with dense thorns, and mutation of this transcription factor in Arabidopsis will result in near complete loss of prickles (Feng et al., 2015). SNP 1469434 within the 8973_g gene produces an MYB transcription factor. WD40 and MYB transcription factor along with a bHLH protein form a protein complex found to control multiple epidermal cell differentiation pathways in a wide range of plants (Ramsay and Glover, 2005). In addition to the MYB-bHLH-WD40 complex, C2H2 transcription factors have been linked to initiation of prickle formation in rose (Swarnkar et al., 2021). The 8970_g gene containing SNP 1469744 was linked to a putative transcription factor considered to be in the C2H2 family. C2H2-type proteins have been found to negatively regulate trichome initiation through GA signaling in Arabidopsis and positively regulate trichome development in tobacco (Han, et al, 2022). The trichomes in Arabidopsis are non-glandular, while tobacco produces glandular trichomes like those present in *Rubus spp*. This indicates that regardless of type of trichome, this protein plays a role in prickle formation. The three SNPs within these genes producing relevant products have superior association values in the raspberry genotypes compared to the blackberry genotypes. This may be attributed to different mechanisms being responsible for prickle formation in the different species. Morphological studies of prickle development have indicated differences in developmental progression in blackberry compared to raspberry and rose (Kellogg et al. 2011). The lack of continuity between raspberry and blackberry allele identities with prickle phenotype may be contributed to different mechanisms required for their formation.

The SNPs determined to be connected to primocane-fruiting were less successful in being matched to previously annotated genes. This is not to say the SNPs lack function, this

information is just yet to be explored. The 9 SNPs located in or around genes have associated products with broad applications. The gene identified as 331.106 is associated with a putative non-specific protein-tyrosine kinase (PTK) in the RLK-Pelle family, subfamily DLSV. RLK-Pelle family of proteins is important in development, organ identity, and defense response, and the DLSV subfamily more specifically most often plays a role in biotic stress responses (Lehti-Shiu and Shiu, 2012). This family of proteins was established early in land plant evolution, likely for the purpose of fighting early, fast-evolving pathogens (Lehti-Shiu et al., 2009). RLKs act through phosphorylation of other proteins, turning them on or off. (Liu et al., 2020) PTK phosphorylates tyrosine and moderates enzymatic activity and downstream signaling. This process occurs during embryogenesis and adult tissue maintenance (Hubbard and Till, 2000). The ERECTA-family of RLKs has been found to control organ growth and flower development in Arabidopsis (Shpak, et al, 2004). The broad range of functions in RLKs supports the indication that the DLSV protein could potentially be playing a role in flowering. The PTK protein also being associated with the processes of maintaining adult tissues and formation of embryos could support flower development. Healthy adult tissue is a requirement for flower development, and embryogenesis is a bloom's ultimate fate.

The 253.54 gene produces a U11/U12-25k small nuclear ribonucleoprotein (snRNP). U11/U12 snRNPs bind to pre-mRNA to remove introns necessary for the formation of mature mRNA (Will, et al, 2004). There are different forms of this complex. The functions of this complex is better understood in humans than plants, although the 65k form has been determined to be essential for normal plant development (Park et al., 2016). The U11/U12-31k protein was found

to affect growth form, leaf and flower morphology, but not flower timing in Arabidopsis (Kim et al., 2010).

The gene containing and adjacent to the majority of significant primocane-fruiting SNPs was 182.37. Gene 182.37 was connected to a putative ribonuclease H-like (RNHL) domain and reverse transcriptase zinc-binding domain products. The RNHL superfamily are almost all nucleases and are responsible for RNA or DNA binding. It is also common for RNHL to have co-occurring domains such as a zinc-finger domain (Majorek et al. 2014). Most genes encoding reverse transcriptase are part of transposable elements and/or inserted retroviral genomes, although some have been endogenized during plant evolution. A zinc finger containing reverse transcriptase in Arabidopsis has been endogenized into the genome, although their exact function is not well understood (Galván-Gordillo et al.). The lack of information of their function further increases interest in this cluster of associated SNPs, especially the one with the perfect segregation pattern with phenotype (2638853). The SNPs labeled 2638853 and 2638849 are both upstream of the annotated gene within 100bp, indicating they may be affecting the promoter region (Promoters, 2021). Further research into the molecular mechanisms of these gene products would be beneficial in determining if they affect flowering.

The SNPs focused on in this study were located on chromosomes 3, 4, 6, and 7, while clusters of SNPs or associated genes were primarily on chromosomes 4 and 6. Certain chromosomes have been found to contain loci controlling traits of interest at a higher frequency than others. Another study found alleles associated with soluble solid content to be located on chromosome 4 and 6 as well (Zurn et al., 2020). Linkage groups containing QTLs and candidate genes associated with

fruit color in red raspberry have been identified on chromosomes 2, 3, 4, and 6 (McCallum et al., 2010). QTLs associated with fruit development were also found on chromosomes 1, 2, 3, 5, and 6 in red raspberry (Graham et al., 2009). Chromosome 1 was also found to have a large QTL containing hundreds of genes associated with control over drupelet count and other fruit size related traits in black raspberry (Willman et al., 2020; Willman et al., 2022). A gene controlling waxy bloom on the canes of black and red raspberry was also identified on chromosome 2 (Pinczinger et al., 2020). The prickle-related SNPs focused on in the current study were all found on chromosome 4. Previous work in black raspberry determined chromosome 4 to contain 3 QTLs associated with titratable acidity (Willman et al., 2022). The connections found between chromosome 4 and fruit quality related traits in other analyses shows potential that this region of the genome may be heavily involved in variable phenotypic expression. A gene identified in apple to produce bHLH was also mapped to chromosome 1 in red raspberry along with QTLs associated with anthocyanin production (Graham et al., 2009). Although this was not captured in the current study, bHLH is an essential part of the complex found to be responsible for prickle formation (Swarnkar et al., 2021; Han, et al, 2022; Ramsay and Glover, 2005). The other components of the complex were products of genes containing meaningfully associated SNPs determined here. The QTL on chromosome 1 may contribute to the formation of this complex along with the regions identified on chromosome 4. The majority of the SNPs associated with primocane-fruiting were clustered within one gene on chromosome 6. QTLs controlling fruit size in black raspberry have also been found on chromosome 6 (Willman et al., 2022). These results indicate potential to find other genes controlling aspects of fruit development in the same region.

The alleles analyzed for both primocane-fruiting and prickle-free phenotypes provided superior results in raspberry genotypes than blackberry, indicating the information provided here is better applied to raspberry. Utilizing the validated SNPs and their associations for marker creation would allow for identification of these alleles within juvenile populations, facilitating greater efficiency in cultivar improvement.

LITERATURE CITED

Ballington, J. R. (2016). The history of blackberry and raspberry breeding in the southern USA. *Acta Horticulturae,* (1133), 13-22. https://10.17660/ActaHortic.2016.1133.3

Clark, J. R. (2008). Primocane-fruiting blackberry breeding. *HortScience, 43*(6), 1637-1639. https://10.21273/HORTSCI.43.6.1637

Coyner, M. A., Skirvin, R. M., Norton, M. A., & Otterbacher, A. G. (2005). Thornlessness in blackberries: A review. *Small Fruits Review, 4*(2), 83-106. https://10.1300/J301v04n02_09

Dossett, M., Lee, J., & Finn, C. E. (2010). Variation in anthocyanins and total phenolics of black raspberry populations. *Journal of Functional Foods, 2*(4), 292-297. https://10.1016/j.jff.2010.10.004

Feng, L., Luan, X., Wang, J., Xia, W., Wang, M., & Sheng, L. (2015). Cloning and expression analysis of transcription factor RrTTG1 related to prickle development in rose (*Rosa rugosa*). *Archives of Biological Sciences, 67*(4), 1219-1225. https://10.2298/ABS150310098F

Food and Agriculture Organization of the United Nations. (2020). http://www.fao.org/faostat/en/#data/QC/visualize

Foolad, M.R. (Pennsylvania State Univ., University Park,PA (USA).Dept.of Horticulture), Stoltz, T., Dervinis, C., Rodriguez, R. L., & Jones, R. A. (1997). Mapping QTLs conferring salt tolerance during germination in tomato by selective genotyping. *Molecular Breeding, 3*(4), 269-277. https://10.1023/A:1009668325331

Galván-Gordillo, S. V., Concepción Martínez-Navarro, A., Xoconostle-Cázares, B., & Ruiz-Medrano, R. (2016). Bioinformatic analysis of Arabidopsis reverse transcriptases with a zinc-finger domain. *Biológia, 71*(11), 1223-1229. https://10.1515/biolog-2016-0145

Gish, L. A., & Clark, S. E. (2011). The RLK/pelle family of kinases. *The Plant Journal : For Cell and Molecular Biology, 66*(1), 117-127. https://10.1111/j.1365-313X.2011.04518.x

Graham, J., Hackett, C. A., Smith, K., Woodhead, M., Hein, I., & McCallum, S. (2009). Mapping QTLs for developmental traits in raspberry from bud break to ripe fruit. *Theoretical and Applied Genetics, 118*(6), 1143-1155. https://10.1007/s00122-009-0969-6

Han, G., Li, Y., Yang, Z., Wang, C., Zhang, Y., & Wang, B. (2022). Molecular mechanisms of plant trichome development. *Frontiers in Plant Science, 13*, 910228. https://10.3389/fpls.2022.910228

Hubbard, S. R., & Till, J. H. (2000). Protein tyrosine kinase structure and function. *Annual Review of Biochemistry, 69*(1), 373-398. https://10.1146/annurev.biochem.69.1.373

Integrative Genomics Viewer [computer software] (2022). Broad Institute and the Regents of the University of California: CA, USA.

Jennings, D. F. (1988). *Raspberries and blackberries: Their breeding, diseases and growth*. Academic Press London UK.

Kaume, L., Howard, L. R., & Devareddy, L. (2012). The blackberry fruit: A review on its composition and chemistry, metabolism and bioavailability, and health benefits. *Journal of Agricultural and Food Chemistry, 60*(23), 5716-5727. https://10.1021/jf203318p

Kellogg, A. A., Branaman, T. J., Jones, N. M., Little, C. Z., & Swanson, J. (2011). Morphological studies of developing *Rubus* prickles suggest that they are modified glandular trichomes. *Botany, 89*(4), 217-226. https://10.1139/b11-008

Khadgi, A., & Weber, C. A. (2020). RNA-seq analysis of prickled and prickle-free epidermis provides insight into the genetics of prickle development in red raspberry (*Rubus ideaus* L.). *Agronomy (Basel), 10*(12), 1904. https://10.3390/agronomy10121904

Kim, W. Y., Jung, H. J., Kwak, K. J., Kim, M. K., Oh, S. H., Han, Y. S., & Kang, H. (2011). The Arabidopsis U12-type spliceosomal protein U11/U12-31K is involved in U12 intron splicing via RNA chaperone activity and affects plant development. *The Plant Cell, 22*(12), 3951-3962. https://10.1105/tpc.110.079103

Lehti-Shiu, M. D., & Shiu, S. (2012). Diversity, classification and function of the plant protein kinase superfamily. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1602), 2619-2639. https://10.1098/rstb.2012.0003

Lehti-Shiu, M. D., Zou, C., Hanada, K., & Shiu, S. (2009). Evolutionary history and stress regulation of plant receptor-like kinase/pelle genes. *Plant Physiology (Bethesda), 150*(1), 12-26. https://10.1104/pp.108.134353

Liu, H., Qu, W., Zhu, K., & Cheng, Z. (. M. (2020). The wild strawberry kinome: Identification, classification and transcript profiling of protein kinases during development and in response to gray mold infection. *BMC Genomics, 21*(1), 635. https://10.1186/s12864-020-07053-4

Majorek, K. A., Dunin-Horkawicz, S., Steczkiewicz, K., Muszewska, A., Nowotny, M., Ginalski, K., & Bujnicki, J. M. (2014). The RNase H-like superfamily: New members, comparative structural analysis and evolutionary classification. *Nucleic Acids Research, 42*(7), 4160-4179. https://10.1093/nar/gkt1414

McCallum, S., Woodhead, M., Hackett, C. A., Kassim, A., Paterson, A., & Graham, J. (2010). Genetic and environmental effects influencing fruit colour and QTL analysis in raspberry. *Theoretical and Applied Genetics, 121*(4), 611-627. https://10.1007/s00122-010-1334-5

Meng, R., & Finn, C. (1999). Using flow cytometry to determine ploidy level in *Rubus*. *Acta Horticulturae 505,* , 223-230. https://doi.org/10.17660/ActaHortic.1999.505.28

Merk, H. L., Ashrafi, H., & Foolad, M. R. (2012). Selective genotyping to identify late blight resistance genes in an accession of the tomato wild species *Solanum pimpinellifolium*. *Euphytica, 187*(1), 63-75. https://10.1007/s10681-012-0729-6

OmicsBox – Bioinformatics Made Easy [computer software] (2019). BioBam Bioinformatics: https://www.biobam.com/omicsbox

Overview of the Caneberry Industry: Facts & Figures. (2017). https://www.raspberryblackberry.com/consumers/overview-of-the-caneberry-industry-facts-figures/

Park, S. J., Jung, H. J., Nguyen Dinh, S., & Kang, H. (2016). Structural features important for the U12 snRNA binding and minor spliceosome assembly of Arabidopsis U11/U12-small nuclear ribonucleoproteins. *RNA Biology, 13*(7), 670-679. https://10.1080/15476286.2016.1191736

Pinczinger, D., von Reth, M., Keilwagen, J., Berner, T., Peil, A., Flachowsky, H., & Emeriewen, O. F. (2020). Mapping of the waxy bloom gene in 'Black Jewel' in a parental linkage map of 'Black Jewel' x 'Glen Ample' (*Rubus*) interspecific population. *Agronomy (Basel), 10*(10), 1. https://10.3390/agronomy10101579

Promotors. (2021). https://www.addgene.org/mol-bio-reference/promoters/#:~:text=Promoters%20are%20about%20100%2D1000,to%20the%20mRNA%20transcript%20produced

Putative Ribonuclease H-like Domain-Containing Protein. (2022). https://www.uniprot.org/uniprotkb/A0A2P6QBX0/entry

Ramsay, N. A., & Glover, B. J. (2005). MYB–bHLH–WD40 protein complex and the evolution of cellular diversity. *Trends in Plant Science, 10*(2), 63-70. https://10.1016/j.tplants.2004.12.011

Shpak, E. D., Berthiaume, C. T., Hill, E. J., & Torii, K. U. (2004). Synergistic interaction of three ERECTA-family receptor-like kinases controls Arabidopsis organ growth and flower development by promoting cell proliferation. *Development (Cambridge), 131*(7), 1491-1501. https://10.1242/dev.01028

Strik, B. C., Clark, J. R., Finn, C. E., & Banados, M. P. (2007). Worldwide blackberry production. *HortTechnology (Alexandria, Va.), 17*(2), 205-213. https://10.21273/HORTTECH.17.2.205

Swarnkar, M. K., Kumar, P., Dogra, V., & Kumar, S. (2021). Prickle morphogenesis in rose is coupled with secondary metabolite accumulation and governed by canonical MBW transcriptional complex. *Plant Direct, 5*(6), e00325-n/a. https://10.1002/pld3.325

U11/U12 Small Nuclear Ribonucleoprotein 25 KDA Protein. (2022). https://www.uniprot.org/uniprotkb/Q9BV90/entry

VanBuren, R., Bryant, D., Bushakra, J. M., Vining, K. J., Edger, P. P., Rowley, E. R., Priest, H. D., Michael, T. P., Lyons, E., Filichkin, S. A., Dossett, M., Finn, C. E., Bassil, N. V., & Mockler, T. C. (2016). The genome of black raspberry (*Rubus occidentalis*). *The Plant Journal: For Cell and Molecular Biology, 87*(6), 535-547. https://10.1111/tpj.13215

Weber, C., & Hai Liu, R. (2002). Antioxidant capacity and anticancer properties of red raspberry. *Acta Horticulturae 585*, 451-457. https://doi.org/10.17660/ActaHortic.2002.585.73

Will, C. L., Schneider, C., Hossbach, M., Urlaub, H., Rauhut, R., Elbashir, S., Tuschl, T., & Luhrmann, R. (2004). The human 18S U11/U12 snRNP contains a set of novel proteins not found in the U2-dependent spliceosome. *RNA (Cambridge), 10*(6), 929-941. https://10.1261/rna.7320604

Willman, M., Bushakra, J. M., Bassil, N. V., Finn, C. E., Dossett, M., Perkins-Veazie, P., Bradish, C. M., Fernandez, G. E., Weber, C. A., Scheerens, J., Dunlap, L., & Fresnedo-Ramírez, J. (2020). Genetic analysis of drupelet count in black raspberry (*Rubus occidentalis*). *Acta Horticulturae,* (1277), 65-72. https://10.17660/ActaHortic.2020.1277.9

Willman, M. R., Bushakra, J. M., Bassil, N., Finn, C. E., Dossett, M., Perkins-Veazie, P., Bradish, C. M., Fernandez, G. E., Weber, C. A., Scheerens, J. C., Dunlap, L., & Fresnedo-Ramirez, J. (2022). Analysis of a multi-environment trial for black raspberry (*Rubus occidentalis* L.) quality traits. *Genes, 13*(3), 418. https://10.3390/genes13030418

Zurn, J. D., Driskill, M., Jung, S., Main, D., Yin, M. H., Clark, M. C., Cheng, L., Ashrafi, H., Aryal, R., Clark, J. R., Worthington, M., Finn, C. E., Peace, C., Iezzoni, A., & Bassil, N. (2020). A Rosaceae family-level approach to identify loci influencing soluble solids content in blackberry for DNA-informed breeding. *G3 (Bethesda, Md.), 10*(10), 3729-3740. https://10.1534/g3.120.401449

APPENDICES

Appendix A

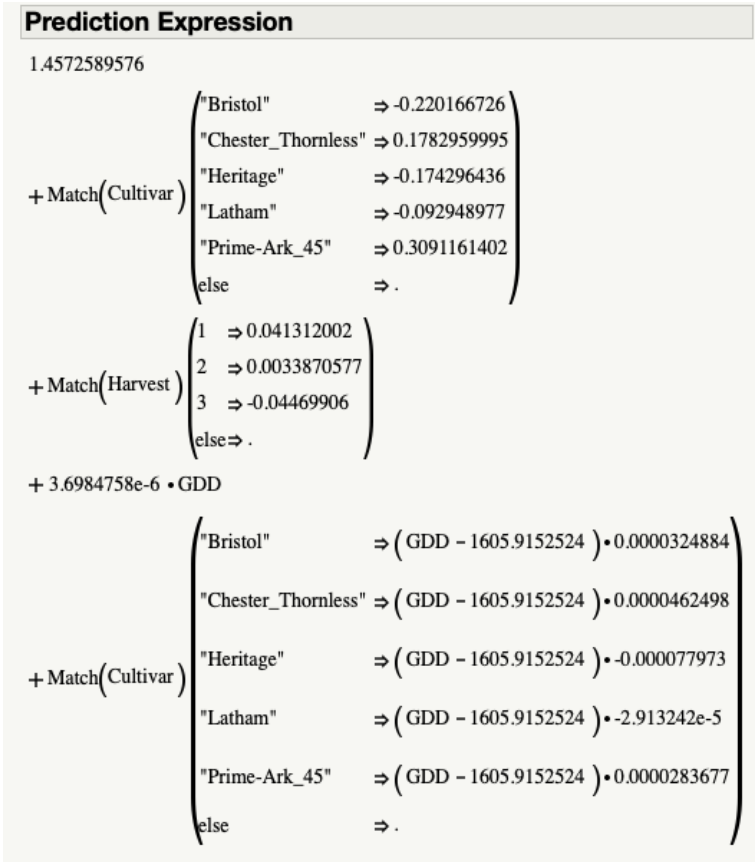

Figure A.1. Prediction expressions for floricane 10 berry mass log values derived from the model incorporating climate data. Used to predict fruit size using the interaction between genotype and GDD, displaying general trends in fruit size.

Appendix B

Table B.1. Sequencing report data for the 31 *Rubus spp.* SNP discovery.

Name	Taxon	Origin	Total Reads	Total Bases Before Trimming	Reads After Trimming	Total Bases After Trimming	Total Reads Mapped	Percent Mapped	Mapped Deduplication Reads To Black Raspberry		Reference Length	Percent of Ref. Covered	Total Read Length Covered the Genome
Ralitsa	R. idaeus subsp. idaeus	Bulgaria	40,301,854	6,045,278,100	40,223,308	5,477,992,231	24,401,884	61%	22,781,552	57%	290,824,797	70%	3,334,543,771
Bababerry	R. idaeus subsp. idaeus	United States, California	56,612,178	8,491,826,700	56,518,678	7,693,416,682	34,082,624	60%	31,631,011	56%	290,824,797	67%	4,657,595,050
R. strigosus Washington	R. idaeus subsp. strigosus	United States, Washington	85,360,530	12,804,079,500	85,145,212	11,512,494,463	51,285,218	60%	48,379,849	57%	290,824,797	64%	7,001,912,488
Liberty	R. idaeus subsp. idaeus	United States, Iowa	72,046,330	10,806,949,500	71,866,908	9,718,433,051	44,273,260	61%	41,690,665	58%	290,824,797	68%	6,037,299,085
Ranere	R. idaeus subsp. idaeus	United States, New Jersey	58,052,934	8,707,940,100	57,899,990	7,811,901,161	35,610,109	61%	33,781,807	58%	290,824,797	66%	4,853,905,857
Viking	R. idaeus subsp. idaeus	Canada, Ontario	73,047,544	10,957,131,600	72,867,534	9,855,173,811	50,326,009	69%	42,745,004	59%	290,824,797	69%	6,811,372,614
Newburgh	R. idaeus subsp. idaeus	United States, New York	69,131,992	10,369,798,800	68,957,198	9,335,823,253	43,434,462	63%	40,845,095	59%	290,824,797	68%	5,925,282,232
Scepter	R. idaeus subsp. idaeus	United States, Maryland	62,471,344	9,370,701,600	62,165,022	8,384,654,378	42,259,956	68%	36,027,942	58%	290,824,797	68%	5,705,531,143
Heritage	R. idaeus subsp. idaeus	United States, New York	80,776,688	12,116,503,200	80,568,728	10,885,084,076	53,280,317	66%	45,496,120	56%	290,824,797	69%	7,203,831,457
Watson	R. idaeus subsp. idaeus	United States, New York	73,157,262	10,973,589,300	72,957,496	9,882,225,865	46,176,786	63%	43,184,580	59%	290,824,797	68%	6,298,026,341
Comet	R. idaeus subsp. idaeus	Canada, Ontario	62,551,086	9,382,662,900	62,405,106	8,427,723,247	38,536,078	62%	36,372,715	58%	290,824,797	67%	5,252,882,626
June	R. idaeus subsp. idaeus	Uncertain	111,970,932	16,795,639,800	111,681,684	15,064,344,327	69,894,906	62%	66,801,120	60%	290,824,797	73%	9,528,251,110
Mammoth Red	R. idaeus subsp. idaeus	Uncertain	89,385,692	13,407,853,800	89,149,252	12,053,972,435	55,246,626	62%	52,063,817	58%	290,824,797	68%	7,535,892,703
R. hirsutus 26-21 Japan	R. hirsutus	Japan	83,167,004	12,475,050,600	82,882,816	11,217,520,008	52,022,256	63%	42,927,467	52%	290,824,797	58%	7,046,385,144
Bristol	R. occidentalis	United States, New York	63,236,726	9,485,508,900	63,067,424	8,537,641,804	44,037,553	70%	34,187,450	54%	290,824,797	89%	5,968,344,651
R. strigosus Jarbridge River ID	R. idaeus subsp. strigosus	United States, Idaho	62,466,184	9,369,927,600	62,126,178	8,411,114,410	40,213,746	64%	41,846,145	67%	290,824,797	67%	5,448,245,747
Gradina	R. idaeus subsp. idaeus	Former Serbia and Montenegro	73,642,690	11,046,403,500	73,441,810	9,921,811,979	49,064,386	67%				68%	6,633,656,196
Krupna Dvorodna	R. idaeus subsp. idaeus	Former Serbia and Montenegro	67,918,658	10,187,798,700	67,765,734	9,168,395,394	42,557,976	58%	40,046,948	54%	290,824,797	66%	5,805,044,043
Pocahontas	R. idaeus subsp. idaeus	United States, Virginia	60,863,864	9,129,579,600	60,709,652	8,210,178,787	38,322,128	63%	48,085,499	79%	290,824,797	68%	5,226,479,705
Southland	R. idaeus subsp. idaeus	United States, North Carolina	86,256,364	12,938,454,600	80,035,499	11,566,447,687	50,082,559	58%	39,802,854	46%	290,824,797	68%	6,834,002,239
Preussen	R. idaeus subsp. idaeus	Germany	67,974,860	10,196,229,000	67,827,150	9,155,315,551	42,268,026	62%	36,707,987	54%	290,824,797	67%	5,759,421,009
Korbfuller	R. idaeus subsp. idaeus	Germany	66,838,850	10,025,827,500	66,639,662	8,955,911,354	38,517,960	58%	43,573,691	65%	290,824,797	66%	5,242,054,519
Zeva II	R. idaeus subsp. idaeus	Switzerland	73,966,766	11,095,014,900	73,777,040	9,925,835,355	51,177,040	69%	43,491,194	59%	290,824,797	68%	6,900,352,697
Schoenemann	R. idaeus subsp. idaeus	Germany	76,269,120	11,440,368,000	76,060,510	10,238,829,620	45,561,094	60%	40,475,156	53%	290,824,797	66%	6,202,380,560
Oregon 1030	R. idaeus subsp. idaeus	United States, Oregon	69,419,214	10,412,882,100	69,152,106	9,327,699,585	47,486,815	68%	42,706,860	62%	290,824,797	69%	6,411,330,667
Lewis	R. idaeus subsp. idaeus	United States, Oregon	73,851,192	11,077,678,800	73,684,145	9,937,945,605	50,073,194	68%	40,403,573	55%	290,824,797	68%	6,766,404,003
Chilcotin	R. idaeus subsp. idaeus	Canada, British Columbia	69,976,982	10,496,547,300	69,806,316	9,419,947,741	42,939,245	61%	41,202,095	59%	290,824,797	67%	5,852,498,369
Cuthbert	R. idaeus subsp. idaeus	United States, New York	77,821,646	11,673,246,900	77,681,157	10,589,461,875	45,100,819	58%	27,657,876	36%	290,824,797	69%	6,162,161,246
Latham	R. idaeus subsp. idaeus	United States, Minnesota	48,121,294	7,218,194,100	47,939,243	6,511,983,640	29,715,027	62%	36,812,503	76%	290,824,797	69%	4,058,919,271
Prime-Ark 45	R. hybrid	United States, Arkansas	104,551,244	15,682,686,600	104,206,664	14,171,498,686	50,416,451	48%	27,826,312	27%	290,824,797	64%	6,886,448,555
Chester Thornless	R. hybrid	United States, Ohio	75,572,564	11,335,884,600	75,363,171	10,257,981,370	37,561,186	50%			290,824,797	64%	5,131,173,640

Table B.2. Prickle presence SNPs within previously annotated genes and their associated products and the roles of such compounds.

SNP name	Chr #	Gene	Associated Products	Species Referenced	Function
TLN1456332	Ro04	2733	PREDICTED: purple acid phosphatase 8-like	*Fragaria vesca subsp. vesca*	Increase phosphate availability
TLN1456646	Ro04	2736	protein NSP-INTERACTING KINASE 3 isoform X2	*Malus baccata*	ATP binding
TLN1456857	Ro04	2739	ATP-citrate synthase alpha chain protein 1	*Citrus unshiu*	Fatty-acid elongation
TLN1457466	Ro04	2743	CRS2-associated factor 2, chloroplastic	*Rosa chinensis*	mRNA processing
TLN1457502	Ro04	2745	cytochrome b-c1 complex subunit 8	*Rosa chinensis*	Mitochondrial electron transport
TLN1457526	Ro04	2746	pyrrolidone-carboxylate peptidase isoform X2	*Prunus armeniaca*	Neuropeptide degrading enzyme
TLN1457698	Ro04	2748	probable polygalacturonase At1g80170	*Fragaria vesca subsp. vesca*	Carbohydrate metabolic process
TLN1457698	Ro04	2750	protein C2-DOMAIN ABA-RELATED 9	*Rosa chinensis*	Regulates ABA sensitivity
TLN1457754	Ro04	2752	U-box domain-containing protein 35	*Rosa chinensis*	ATP binding
TLN1457848	Ro04	2754	MLP-like protein 34	*Rosa chinensis*	ABA response
TLN1458007	Ro04	2756	PREDICTED: kirola-like	*Fragaria vesca subsp. vesca*	Uncharacterized, RNA-binding
TLN1453204	Ro04	3772	CRAL-TRIO domain-containing protein YKL091C isoform X1	*Prunus yedoensis var. nudiflora*	Unreviewed
TLN1453130	Ro04	3773	two-component response regulator ORR21	*Rosa chinensis*	Cytokinin-activated signaling pathway
TLN1452547	Ro04	3782	pentatricopeptide repeat-containing protein At2g01740, putative tetratricopeptide-like helical domain-containing protein	*Rosa chinensis*	Mitochondrial DNA expression
TLN1452131	Ro04	3787	putative UDP-3-O-acyl-N-acetylglucosamine deacetylase	*Rosa chinensis*	Lipid biosynthesis
TLN1451911	Ro04	3789	protein VAC14 homolog	*Fragaria vesca subsp. vesca*	Cell membrane formation
TLN1451774	Ro04	3791	ankyrin repeat-containing protein At2g01680	*Fragaria vesca subsp. vesca*	Protein dephosphorylation
TLN1451571	Ro04	3797	histone-lysine N-methyltransferase setd3 isoform X2	*Fragaria vesca subsp. vesca*	Cell differentiation
TLN1456430	Ro04	10474	purple acid phosphatase 3, putative Acid phosphatase	*Rosa chinensis*	Increase phosphate availability, Catalyze cellular chemical reactions
TLN1457652	Ro04	13978	---NA---		
TLN1457768	Ro04	13980	U-box domain-containing protein 35	*Rosa chinensis*	ATP binding
TLN1446484	Ro04	15927	ubiquitin-like-specific protease 1D isoform X2	*Fragaria vesca subsp. vesca*	Cell division
TLN1446390	Ro04	15929	equilibrative nucleotide transporter 1	*Rosa chinensis*	Mediate nucleoside transport
TLN1446331	Ro04	15930	probable serine/threonine-protein kinase PBL1	*Rosa chinensis*	Plant growth and defense
TLN1446321	Ro04	15932	probable xyloglucan endotransglucosylase/hydrolase protein 33	*Rosa chinensis*	Cell wall biosynthesis
TLN1446252	Ro04	15933	aldehyde dehydrogenase family 2 member B7, mitochondrial-like	*Fragaria vesca subsp. vesca*	Catalyzes NADH formation
TLN1445641	Ro04	15936	putative DNA (cytosine-5-)-methyltransferase	*Rosa chinensis*	DNA methylation
TLN1445230	Ro04	15940	potassium transporter 6	*Rosa chinensis*	Potassium transport regulation
TLN1445046	Ro04	15941	probable alpha,alpha-trehalose-phosphate synthase [UDP-forming] 10	*Fragaria vesca subsp. vesca*	Trehalose formation
TLN1444988	Ro04	15942	putative six-bladed beta-propeller, TolB	*Carya illinoinensis*	Contribute to larger protein structures
TLN1444893	Ro04	15944	putative transcription factor WD40-like family	*Fragaria vesca subsp. vesca*	Cell differentiation
TLN1444893	Ro04	15945	AP-1 complex subunit gamma-2	*Trifolium medium*	Intracellular protein transport
TLN1444535	Ro04	15949	WAT1-related protein At3g28050-like	*Prunus avium*	Cell wall thickness
TLN1444398	Ro04	15950	WAT1-related protein At3g28050	*Rosa chinensis*	Cell wall thickness
TLN1444388	Ro04	15951	WAT1-related protein At1g70260	*Fragaria vesca subsp. vesca*	Cell wall thickness
TLN1444296	Ro04	15953	putative serine/threonine-protein kinase PBL28	*Prunus yedoensis var. nudiflora*	ATP binding
TLN1453142	Ro04	28406	two-component response regulator ORR21	*Rosa chinensis*	Cytokinin-activated signaling pathway
TLN1471950	Ro04	8946_g	hypothetical protein IFM89_017185	*Coptis chinensis*	Hypothetical
TLN1471918	Ro04	8947_g	transmembrane emp24 domain-containing protein p24delta9-like	*Fragaria vesca subsp. vesca*	Intracellular protein transport
TLN1471790	Ro04	8948_g	hypothetical protein DVH24_040668	*Malus domestica*	Hypothetical
TLN1471225	Ro04	8951_g	RmlC-like cupins superfamily protein	*Prunus dulcis*	Seed development
TLN1471691	Ro04	8952_g	phosphate transporter PHO1 homolog 3-like	*Fragaria vesca subsp. vesca*	Phosphate ion transport
TLN1471056	Ro04	8954_g	galactoside 2-alpha-L-fucosyltransferase	*Rosa chinensis*	Cell wall biosynthesis
TLN1470928	Ro04	8955_g	putative glycosyl hydrolase, five-bladed beta-propellor domain-containing protein	*Rosa chinensis*	Catalyzes hydrolysis
TLN1470618	Ro04	8956_g	putative glycosyl hydrolase, five-bladed beta-propellor domain-containing protein	*Rosa chinensis*	Catalyzes hydrolysis
TLN1470172	Ro04	8960_g	anaphase-promoting complex subunit 2	*Pyrus ussuriensis x Pyrus communis*	Cell division
TLN1469744	Ro04	8970_g	putative transcription factor C2H2 family	*Rosa chinensis*	Environmental stress response
TLN1469627	Ro04	8971_g	EXP4 protein	*Rosa hybrid cultivar*	Nuclear exporter
TLN1469434	Ro04	8973_g	transcription factor MYB54	*Rosa chinensis*	Cell differentiation

Table B.3. Primocane-fruiting related SNPs within previously annotated genes and their associated products and the functions of such compounds.

SNP name	Chr #	Gene	Associated Products	Species Referenced	Function
488579	Ro02		Not Annotated		
779032	Ro02		Not Annotated		
1195818	Ro03		Not Annotated		
1286217	Ro03		Not Annotated		
1286318	Ro03	331.106	putative non-specific protein-tyrosine kinase RLK-Pelle-DLSV family	*Rosa chinensis*	Protein phosphorylation
1290881	Ro03		Not Annotated		
2475612	Ro06		Not Annotated		
2475630	Ro06		Not Annotated		
2475632	Ro06		Not Annotated		
2638849	Ro06	99bp before 182.37			
2638853	Ro06	6bp before 182.37			
2638865	Ro06	182.37			
2638866	Ro06	182.37	putative ribonuclease H-like domain, reverse transcriptase zinc-binding domain-containing protein	*Rosa chinensis*	DNA integration DNA repair
2638867	Ro06	182.37			
2638868	Ro06	182.37			
2638870	Ro06	182.37			
2638922	Ro06		Not Annotated		
2639007	Ro06		Not Annotated		
2639023	Ro06		Not Annotated		
3133120	Ro07		Not Annotated		
3305590	Ro07	253.54	U11/U12 small nuclear ribonucleoprotein 25 kDa protein	*Rosa chinensis*	RNA splicing
3398606	Ro07		Not Annotated		
3416099	Ro07		Not Annotated		

Table B.4. All 3.5 million SNPs identified available through Dryad: https://doi.org/10.5061/dryad.zkh1893cp