# Advancing Genomics Assisted Breeding in Apple and Cannabis

by

Sophie Watts

Submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
August 2023

Dalhousie University is located in Mi'kma'ki, the
ancestral and unceded territory of the Mi'kmaq.
We are all Treaty people.

*Dedicated to my Great Grandmother, Anna McNeil (Dalhousie Class of 1923), my Grandmother, Elizabeth Yates (Dalhousie Class of 1952), my Great Aunt, Dorothy Yates (Dalhousie Class of 1953), and my mother, Jennifer Watts (Dalhousie Class of 1996).*

*Thank you for showing me the value of an education.*

# TABLE OF CONTENTS

*List of Figures* ..... *vi*

*Abstract* ..... *x*

*List of Abbreviations Used* ..... *xi*

*Acknowledgments* ..... *xiii*

*Chapter 1: Introduction* ..... *1*

**Introduction** ..... **1**

**Literature Review** ..... **4**

Apple breeding ..... 4
Apple genome and domestication history ..... 5
Cannabis breeding ..... 6
Cannabis genome and domestication history ..... 8
Genomic assisted breeding ..... 9
Genome-wide association studies ..... 11
Genotyping and Phenotyping ..... 13
GWAS Challenges ..... 15

**Concluding remarks** ..... **16**

*Chapter 2: Quantifying apple diversity: a phenomic characterization of Canada's Apple Biodiversity Collection* ..... *18*

**Introduction** ..... **20**

**Materials and Methods** ..... **22**

The Apple Biodiversity Collection ..... 22
Phenology ..... 24
Fruit quality ..... 25
Apple classification ..... 28
Statistical analysis ..... 29

**Results** ..... **30**

Phenology and fruit quality traits ..... 31
Fruit quality after storage ..... 34
Relationships among traits ..... 36
Phenotypes according to apple classifications ..... 38
Phenotypic changes over time ..... 40

**Discussion** ..... **41**

**Acknowledgements** ..... **49**

*Chapter 3: Large-scale GWAS in apple reveals NAC18.1 as a master regulator of ripening traits* ..... *51*

**Introduction** ..... **52**

**Results** ..... **55**

**Discussion** ..... **63**

Genetic structure of the ABC ..... 63
Phenolic content is associated with loci on chromosome 15 and 16 ..... 64
NAC18.1 is major locus associated with multiple ripening traits ..... 65

Allelic effects of NAC18.1 on ripening ..... 66
**Materials and Methods ..... 73**
Phenotype data ..... 73
Genetic data ..... 73
Principal components analysis ..... 74
Genome-wide association studies ..... 75
**Acknowledgements ..... 76**
***Chapter 4: Cannabis labelling is associated with genetic variation in terpene synthase genes ..... 78***
**Introduction ..... 78**
**Materials and Methods ..... 80**
Samples ..... 80
Gas chromatography ..... 80
Genomic analysis ..... 82
Genetic and chemical analysis ..... 83
Genome-wide association ..... 84
**Results and Discussion ..... 85**
**Acknowledgments ..... 91**
***Chapter 5: From GWAS to CRISPR: a direct route to the gene-edited orchard. ..... 92***
**Introduction ..... 92**
**Genetic Mapping ..... 98**
Linkage mapping ..... 98
GWAS ..... 100
Short-listing candidate targets ..... 103
**Gene editing challenges ..... 105**
In vitro tissue culture ..... 105
Gene editing technology ..... 107
**Gene editing advantages ..... 109**
Transgene-free editing ..... 109
Introducing novel variation ..... 111
Target Specificity ..... 112
Multiplexing ..... 114
**Evaluate ..... 115**
**Conclusion ..... 116**
***Chapter 6: Conclusion ..... 118***
**Summary of Findings ..... 118**
**Future directions ..... 122**
***References ..... 125***
***Appendix I: Quantifying apple diversity: a phenomic characterization of Canada's Apple Biodiversity Collection (Chapter 2) ..... 155***
***Appendix II: Large-scale GWAS in apple reveals NAC18.1 as a master regulator of ripening traits (Chapter 3) ..... 162***

*Appendix III: Cannabis labelling is associated with genetic variation in terpene synthase genes (Chapter 4)* ..... *164*

*Appendix IV: Copyright release* ..... *169*

## List of Figures

**Figure 2-1.** World map showing geographic origins of apple accessions from Canada's Apple Biodiversity Collection…….…….…….…….31

**Figure 2-2.** Distribution of apple phenotypes from Canada's Apple Biodiversity Collection. Shaded grey areas represent the range of values observed across the "top 9" cultivars grown in the USA.…….…….……...…32

**Figure 2-3.** Fruit quality measurements before and after 3 months of cold storage. Correlations of fruit quality measures taken before and after storage are found in the left column (A, C, E, G). Distributions of the percent change of each trait during storage are shown in the right column (B, D, F, H). The shaded grey areas represent the range occupied by the “top 9” cultivars.…….…….…….…...35

**Figure 2-4.** Heat map of pairwise correlations among phenology, fruit quality, and storage traits. Significant correlations ($p < .01$) after correcting for multiple comparisons are indicated with an asterisk.…….…….……37

**Figure 2-5. Figure 2-5.** Correlation of firmness at harvest with (a) flowering date, (b) harvest date, and (c) time to ripen. Correlation of firmness loss during storage with (d) flowering date, (e) harvest date, and (f) time to ripen……..……38

**Figure 2-6.** Comparisons of fruit quality traits among species, geographic origins, and uses. The arrows indicate phenotypes where the difference between groups was significant. The arrow points to the group with the higher value.….…….…….……..39

**Figure 2-7.** Apple phenolic content decreased over the past 200 years. Each dot represents a named cultivar. A cultivar's "year of release" refers to its year of commercialization, release to the public, or initial mention in historical records.…….41

**Figure 3-1.** Population genomic analysis of 1,054 diverse apple accessions. A) Principal components analysis of genome-wide SNP data, with accessions coloured by harvest date and with shape representing species. B) Boxplot of the PC1 values between *M. sieversii* and *M. domestica*. C) Pearson correlation of PC 2 values with harvest date.…….55

**Figure 3-2.** Genome-wide association study for phenolic content. A) Manhattan plot of GWAS results for phenolic content. The horizontal red line represents the significance threshold. B) Boxplot of the distribution of phenolic content measurements across the genotypes of the top SNPs on chromosomes 15 and 16, with a black line indicating the median phenolic content measurement for each genotype.…….56

**Figure 3-3.** Genome-wide association studies of ripening related traits. Manhattan plots of GWAS results for A) harvest date, B) juiciness, C) firmness at harvest, D) firmness after storage, E) change in SSC during storage and F) SSC after storage. The horizontal red lines in each Manhattan plot represent the significance threshold. The distributions of trait measurements across the *NAC18.1* D5Y genotypes are also shown adjacent to each Manhattan plot with median values for each genotype indicated by a horizontal line…..58

**Figure 3-4.** Genome-wide association study for softening during storage. A) Manhattan plot for a GWAS for softening during storage. The horizontal red line represents the significance threshold. B) A Manhattan plot showing only the genomic region around the GWAS peak from panel (A) on chromosome 10. The purple SNP is the most significant SNP associated with softening. The *PG1* SNP is shown in pink. The horizontal dotted line represents the significance threshold. C) Distributions of softening measurements across *PG1* genotypes. D) Distributions of softening measurements across genotypes of the most significant SNP from the GWAS. Horizontal black lines indicate the median value within each genotype.…….60

**Figure 3-5.** A summary of the allelic effects of *NAC18.1* D5Y genotypes on ripening phenotypes, both at harvest and after 3 months of cold storage. A) Median measures of firmness at harvest and firmness after storage for apple accessions with different D5Y genotypes. B) The median soluble solids content at harvest and after storage for apple accessions with different D5Y genotypes.…….61

**Figure 4-1.** A) Genome-wide SNP data. B) Terpenes and cannabinoids. Each dot represents a *Cannabis* sample and is coloured by the labelling scale ranging from 100% Sativa to 100% Indica. C) The percent variance explained by PCs from the genome-wide SNP data (blue), from the terpene and cannabinoid data (green) and from both the genetic and chemical data (purple). The *y* axis shows the percent variance explained as PCs are added to linear models where the Sativa–Indica labelling scale is the dependent variable..86

**Figure 4-2.** A) The percent variance of the five-point Sativa–Indica labelling scale that is explained by terpene and cannabinoid concentrations from Pearson correlations. The *P* values were Bonferroni-adjusted for multiple comparisons. The asterisks denote chemicals with tentative identifications. GWAS results are shown for chemicals highlighted in grey. C) Manhattan plots of mixed linear model (MLM) GWAS for myrcene on chromosome 5 B) Manhattan plots of mixed linear model (MLM) GWAS for guaiol, γ-eudesmol and β-eudesmol on chromosome 6. The significance thresholds from the MLM are shown as horizontal dashed lines. Significant SNPs from the MLMM GWAS are red. Terpene synthase gene clusters are green. Below the Manhattan plots are heat maps of the pairwise LD ($R^2$) between pairs of SNPs that appear in the Manhattan plots……..88

**Figure 5-1.** GWAS to CRISPR approach. A) Establish a large, diverse population to map traits of interest. B) Conduct genome-wide association studies. C) Identify putatively causal alleles. In this example, three different genes have been identified. D) Design CRISPR sgRNAs to edit and introduce novel alleles. Edits are represented by the blue shapes (ie. insertions, deletions or alleles swaps). E) Introduce CRISPR sgRNAs into multiple genetic backgrounds and regenerate plantlets from tissue culture. Here, one edit

is introduced into one genetic background, however this step can be multiplexed to introduce edits into multiple genes within a genetic background to stack traits or target a gene family. In addition, this process can be performed across multiple genetic backgrounds. F) Evaluate edited germplasm to determine if edits generated the desirable phenotypes. G) Select improved accessions with desirable phenotypes for commercialization…….97

## Abstract

Apple and cannabis are two economically and culturally important crops grown around the world for a diversity of end uses. Both apple and cannabis stand to benefit immensely from genomics assisted breeding. Discovering the genetic control of key traits is necessary to advance genomics assisted breeding of both apple and cannabis. The objective of this thesis was to characterize genomic and phenotypic diversity across apple and cannabis and leverage insights from genotype-phenotype associations to uncover the genetic control of key traits. To accomplish the outlined objectives, I first undertook a phenomic characterization of Canada's Apple Biodiversity Collection (ABC) to quantify apple phenotypic diversity across 39 fruit quality and phenology traits from over 20,000 individual apples sampled from 1,000 accessions. I measured a wide range of phenotypic variation that can be leveraged for future apple improvement. For example, apples varied 61-fold in weight, 18-fold in acidity, and 100-fold in phenolic content. I measured dramatic changes to apple physiology that occurred during 3 months of cold storage: on average, apples lost 39% of their firmness, 31% of their acidity, and 9% of their weight, but gained 7% in soluble solids. Next, I paired the apple phenotype data with over 260,000 genetic markers sequenced across the ABC and conducted genome-wide association studies (GWAS) to uncover genetic markers associated with fruit quality and phenology traits. I identified novel associations for phenolic content and an association for softening that overlapped with a previously reported locus. In addition, I identified that allelic variation at the *NAC18.1* transcription factor was associated with numerous ripening related traits including harvest date, firmness at harvest, and firmness after storage. Next, I used over 100,000 genetic markers across more than 137 Cannabis samples to conduct GWAS for 40 chemical compounds. In addition, I examined the genetic and chemical basis for cannabis labelling. I found that overall Indica/Sativa labelled samples were genetically and chemically indistinct. However, I identified variation in the content of three sesquiterpenes that was associated with Indica/Sativa labels and was also controlled by genetic variation at tandem arrays of terpene synthase genes. Finally, in Chapter 5, I use apple as a case study to outline a pathway for how putatively causal genetic variants identified from GWAS can be leveraged to develop new cultivars through gene editing. While cannabis and apple each face unique challenges in their cultivation and breeding, the research presented here provides a solid foundation for the genomics-assisted improvement of new cultivars in both crops.

**List of Abbreviations Used**

| | |
|---|---|
| ABC | Apple Biodiversity Collection |
| CBDA | Cannabidiolic acid |
| CRISPR | Clustered regularly interspaced short palindromic repeats. |
| GBS | Genotype-by-sequencing |
| GC-MS | Gas chromatography mass spectrometry |
| GWAS | Genome-wide association study |
| LD | Linkage disequilibrium |
| MAF | Minor allele frequency |
| MAS | Marker-assisted selection. |
| MLM | Mixed linear model |
| MLMM | Multi-locus mixed model |
| NAC | NAM, ATAF1,2 and CUC2 |
| PAM | Protospacer-adjacent motif |
| PG | Polygalacturonase |
| PC | Principal component |
| PCA | Principal component analysis |
| QTL | Quantitative trait locus |
| REML | Restricted Maximum Likelihood |
| sgRNA | Single guide RNA |
| SNP | Single-nucleotide polymorphism. |
| SSC | Soluble solids content |

| | |
|---|---|
| TALENs | Transcription activator-like effector nucleases. |
| TASSEL | Trait analysis by association, evolution and linkage |
| TF | Transcription factor |
| THCA | Δ9-tetrahydrocannabinolic acid |
| WGS | Whole-genome sequencing. |
| ZFNs | Zinc-finger nucleases. |

## Acknowledgments

Thank you Dr. Sean Myles, for your mentorship, guidance and support. I have learned so much from you over the years and I feel incredibly fortunate to have you as a mentor. Thank you to my co-authors and collaborators for your contributions that made this work possible. Thank you to my committee members, Dr. Vasantha Rupasinghe and Dr. Lord Abbey, for your support and feedback. Thank you Dr. Zoë Migicovsky, for the amazing mentorship and encouragement over the years. Thank you to my fellow lab members Thomas Davies and Tayab Soomro, for your friendship and comradery. Thank you to the Natural Sciences and Engineering Research Council (Vanier Canada Graduate Scholarship) and Dalhousie University (Killam Predoctoral Scholarship and President's Award) for the financial support that allowed me to do this work. Thank you to AAFC farm services for all their work maintaining the Apple Biodiversity Collection. Thank you, Dr. Pat Brown, Dr. Allan Andrew and Dr. Richard Espley, for the opportunity to visit and learn from your research groups. Thank you to my friends and family for all the ways in which you've supported me over the years. A very big thank you to my mom and dad, who sparked my passion for studying agriculture and have always encouraged me to follow my interests.

# Chapter 1: Introduction

## Introduction

Over the course of the 10,000 years of agriculture, humans have domesticated and improved numerous crop species for food, fiber, fuel, and medicines. The improvement of crops, through plant breeding, continues today to meet changing climates, taste preferences and evolving pest and disease pressures. The global population is expected to reach 9 billion by 2050 and with this growing population comes increasing demands for high quality food and plant products, all while overcoming the threats posed by climate change (Tian et al. 2021). Plant breeding requires new tools and innovations to rapidly develop improved cultivars that can meet the growing challenges faced by agriculture. Since the sequencing of the first plant genome, the Arabidopsis genome in 2000 (Arabidopsis Genome Initiative 2000), a new era of plant breeding began that aimed to capitalize upon the insights provided by the study of plant genomics (Y. Sun et al. 2021). Genomics assisted breeding is one such innovation that involves using information regarding an organism's genome to make informed decisions through the breeding process to facilitate faster and more efficient breeding (Varshney et al. 2021). Although over two decades have passed since the first plant genome was sequenced, not all crops have reaped the benefits of genomics assisted breeding. To secure a future of sustainable crop production, emphasis should be placed on advancing genomic resources across diverse crop species to enable efficient breeding of improved cultivars. Two such crops that can benefit from the application of genomics assisted breeding, are apple and cannabis.

The apple (*Malus domestica*) has been cultivated for over 3,000 years across many temperate regions around the world (Zohary, Hopf, and Weiss 2012; Cornille et al. 2014). The apple is one of the most widely produced and consumed fruits in the world, ranking second in economic value among fruit crops with an estimated global production of $67 billion US in 2021 (FAOSTAT 2021). Apples are valued for their numerous end-uses, including for cider production, processing, and fresh consumption. Breeding improved apple cultivars is necessary to meet demands fruit quality fruit, especially in the face of climate change. Breeding new apple cultivars is challenging due to the long juvenile phase of 3-7 years. For example, it can take over 25 years for breeding programs to go from an initial cross to commercial release (Peil et al. 2008). Furthermore, the factors that control fruit quality and phenology traits are complex and our understanding of these traits is limited to only a few elite varieties that account for most of the world's apple production. For example, in 2018, over half of apple production in the USA was accounted for by only 4 cultivars ("US Apple and Pear Forecast" n.d.). Furthermore, genetic analysis has revealed that many of the commercially available apples are closely related to one another (Migicovsky, Gardner, et al. 2021). Understanding how fruit quality and phenology quality traits vary across the diversity of *Malus domestica* is vital to identifying unique phenotypes that can be bred into new cultivars. Overall, the elucidation of the genetic control of fruit quality and phenology traits is essential for the successful application of genomics assisted breeding for apple cultivar development.

Cannabis (*Cannabis sativa*) is another important crop that has been cultivated since Neolithic times and is grown around the world (G. Ren et al. 2021). Cannabis is a diverse plant species that is grown for food, fiber, medicine and most notably for its use

as a psychoactive substance. People have been consuming cannabis for its psychoactive effects for over 2,500 years (Lawler 2019). However, in modern times cannabis cultivation has been prohibited due to its popular use as a recreational drug. In recent years, the genetic study of cannabis has increased due to the more favourable regulations surrounding its cultivation and due to the renewed interested in fiber-type cultivars for fiber and grain (Fike 2016). In a small number of countries, legalization of cannabis products for recreational and medicinal use has occurred in recent years, leading to growing global market that has an estimated value of over US$340 billion (Naville 2019). Along with cannabis legalization, increased focused is being placed on determining the genetic control of quality traits, such as aroma, for the targeted breeding of drug-type cannabis. However, studies on cannabis have for the most part been limited to the examination of fiber-type cannabis (Petit, Salentijn, Paulo, Denneboom, and Trindade 2020; Petit, Salentijn, Paulo, Denneboom, van Loo, et al. 2020). Despite the growing commercial market, cannabis genomics research continues to lag behind others crops due to the historical and present day restrictions on its cultivation, even for research purposes (G. Ren et al. 2021). Overall, cannabis has arguably been left behind in the genomics revolution compared to other crops (Hesami et al. 2020).

Cannabis and apple are similar in that they are both valued for the diversity in their quality attributes and end uses. To advance genomics assisted breeding in apples and cannabis, the genetic control of commercially relevant traits must be characterized. Genetic mapping, specifically genome-wide association studies (GWAS), has served as an important approach for identifying genes controlling traits by uncovering genotype-

phenotype associations. Cannabis and apple both stand to benefit from increased examination of the genetic control of commercially relevant traits through GWAS. Given the importance placed upon quality attributes that are sought after and expected from apple and cannabis cultivars, elucidating the genetic control of these traits, and targeting them using genomics assisted breeding is valuable for the development of improved cultivars.

## Literature Review

### *Apple breeding*

Apple breeding is expensive and time-consuming: large orchards are required, and it can take over 25 years from seedling to commercial release (Peil et al. 2008). In addition to taking a long time to reach maturity, perennial crops such as the apple are costly to maintain and evaluate for desirable traits (Migicovsky and Myles 2017). In traditional apple breeding programs, evaluation for desirable traits cannot take place until after the long juvenile phase of 3-7 years. Thousands of seedlings may be produced from a cross, but only a few will be commercially viable at the end of decades of quality evaluation (Myles 2013). To meet changing climates and consumer tastes there is a need to improve the efficiency of the apple breeding process.

The focus of many apple breeding programs has been on developing cultivars with improved disease and pest resistance as well fruit quality. Apple fruit quality is determined by numerous different factors, such as taste, texture, aroma, nutritional value

and appearance (Musacchi and Serra 2018). Consumers value tasty, well textured apples, and producers value apples that maintain fruit quality year-round in storage. Overall, consumers' "likeness" of fruit cultivars is strongly associated with texture traits such as fruit firmness (Szczesniak 2002). However, the factors that contribute to fruit apple quality are complex and under-characterized (Amyotte et al. 2017). Given that apples are often consumed after spending time in storage, it is essential that they maintain their texture and overall fruit quality until they are sold (Ben Sadok et al. 2015). Therefore, breeding for improved fruit quality both at harvest and post-storage is a high priority for apple breeders. Due to the close relationship between fruit quality and phenology traits, such as harvest time and ripening, phenology traits are also important targets for breeding. Furthermore, breeding for specific phenology traits is essential in order to adapt cultivars to shifting growing seasons due to climate change (Gottschalk and van Nocker 2013). As one of the most valuable fruit crops grown globally, advancing the apple breeding process is essential to efficiently produce new cultivars with improved quality, nutrition, and climate resilience.

*Apple genome and domestication history*

The center of origin of the domesticated apple is thought to be in Central Asia (Velasco et al. 2010). There are 25 – 47 recognized species of *Malus*, one of which is *Malus domestica.* The *Malus sieversii* species*,* found in Kazakhstan, is thought to be the main wild progenitor of the domesticated apple. However, other *Malus* species have also contributed to the genetic pool of the domesticated apple (Velasco et al. 2010). It has been proposed that *M. sieversii* was carried west along the Silk Road and hybridized

along the way with other *Malus* species as it evolved into the domesticated apple (Wang et al. 2018). In 2010, the assembly of the first reference genome for apple was completed for 'Golden Delicious' (Velasco et al. 2010). The results from this assembly suggested that the apple genome underwent a whole genome duplication which resulted in the domesticated apple having a distinct 17 chromosomes (Troggio et al. 2012). Cultivated apples are diploid (2x = 34), however there are some triploid (3x = 51) cultivars that spontaneously arise (Brown and Maloney 2003). The release of additional apple genomes has further spurred research into the genetic control of commercially relevant traits (Daccord et al. 2017; L. Zhang et al. 2019; X. Sun et al. 2020; A. Khan et al. 2022). In apple, fruit quality traits have been characterized at the genetic level, such as fruit firmness, acidity and ripening (Kenis, Keulemans, and Davey 2008; Xu, Wang, and Brown 2012; Costa et al. 2005). However, there are advances still to be made by examining fruit quality and phenology across diverse apple populations to fully characterize the genetic control of previously identify genes and to identify additional genetic elements that could be impacting phenotypic variation.

*Cannabis breeding*

As one of the first plants to be cultivated by humans, cannabis has been used for thousands of years for fiber, food, and medicine. For example, in China, cannabis has been grown for textiles for over 6000 years and consumed for its psychoactive properties for over 2500 years (H.-L. Li 1973; M. Ren et al. 2019). Selective breeding of cannabis has resulted in the differentiation of cannabis based on the content of two major

cannabinoid compounds, cannabidiolic acid (CBDA) and Δ9-tetrahydrocannabinolic acid (THCA). Fiber-type cannabis (often referred to as hemp) is usually classified based on a THC content of less than 0.3%, while drug-type cannabis has higher levels of CBD and THC (E. Small and Beckstead 1973; de Meijer et al. 2003). Beyond the over 150 cannabinoid compounds that are responsible for psychoactive effects, cannabis is also well known for its diversity of terpene compounds that impart distinctive aromas (Booth and Bohlmann 2019; Mudge, Brown, and Murch 2019). Aromas provided by terpene compounds have been shown to drive consumer preferences for cannabis strains (Gilbert and DiVerdi 2018). Furthermore, there is increasing evidence to suggest that terpenes may interacted with cannabinoids to mediate psychoactive effects through what is called an 'entourage' effect (Russo 2011; Koltai and Namdar 2020). Identifying the control of cannabinoid and terpene compounds is important to advance our understand of their psychoactive effects and for targeted breeding of cultivars with desirable aroma profiles.

The uncertainty surrounding classification of cannabis has been viewed as one of the major challenges that breeding programs face (Barcaccia et al. 2020; Hurgobin et al. 2020). The clandestine nature of cannabis breeding has allowed numerous label and naming conventions to be used with little oversight to ensure consistency (Schwabe and McGlaughlin 2019). For example, the labels 'Indica' and 'Sativa' have routinely been applied to cannabis strains to attempt to capture differences in cannabinoid content, psychoactive effects, plant morphologies and aromas. However, the degree to which 'Indica' and 'Sativa' labels accurately differentiate cannabis strains on a genetic level appears to be inconsistent (Sawler et al. 2015; Lynch et al. 2017). To complicate cannabis

nomenclature further, the use of strain names is not reserved for genetically unique cultivars (Sawler et al. 2015). Unlike other clonally propagated crops, like apples, strain names are still assigned to cannabis plants derived from seed. Studies have shown that neither reported ancestry or strain names can reliably predict the genetic identity of cannabis strains (Sawler et al. 2015). As cannabis shifts away from clandestine cultivation, consistent labelling and naming conventions are required to provide accurate product differentiation. Genomic analyses can help resolve cannabis labelling practices that have for too long remained convoluted due to clandestine breeding practices. Given its use as a medicine and recreational drug, understanding the genetic control of the chemical compounds produced by cannabis is essential to advance breeding for desirable phenotypes.

*Cannabis genome and domestication history*

For many years the domestication history of cannabis remained unclear due to the lack of formal and documented breeding and the tendency of domesticated cultivars to feralize and potentially admix with wild populations (G. Ren et al. 2021; Ernest Small 2015). However, Recent whole-genome resequencing of over 110 cannabis accessions from around the world, showed that cannabis was domesticated during Neolithic times in East Asia (G. Ren et al. 2021). As mentioned previously, cannabis cultivars are often distinguished based on ‘Indica’/‘Sativa’ labels. The use of the vernacular ‘Indica’/‘Sativa’ labels has arisen from the view that they represent distinct species. For example, studies have proposed the existence of multiple species within the Cannabis

genus, including *Cannabis sativa*, *Cannabis indica*, *Cannabis ruderalis* (Hillig 2005). However, the prevailing consensus based on recent studies is that *Cannabis sativa* is a monotypic species that displays a large amount of phenotypic and genetic diversity (Guy and McPartland 2017; Hurgobin et al. 2020).

Cannabis is a dioecious crop, with occasional monoecious forms arising, and has a diploid genome (2n = 20) with 9 pairs of autosomes and one pair of sex chromosomes (Sakamoto et al. 1998; Hurgobin et al. 2020). In 2011, the first draft cannabis genome was published and since then multiple new genomes have been released with increasing quality (van Bakel et al. 2011; Grassa et al. 2018; Laverty et al. 2019; S. Gao et al. 2020). Following the release of the genome references for cannabis, attention has largely been focused on identifying the genetic control of cannabinoid compounds, while the genetic control of terpene compounds has remained largely unexplored (Booth, Page, and Bohlmann 2017; Zager et al. 2019; Günnewich et al. 2007; Livingston et al. 2020).

*Genomic assisted breeding*

Genomics assisted breeding techniques, such as marker assisted selection (MAS), uses genetic information to accelerate the breeding process (K. A. McClure et al. 2014; Ru et al. 2015). MAS is a technique used to genotype plants at the seedling stage for genetic markers associated with desirable traits, and only the plants with desirable genetic profiles are grown to full maturity, while the rest are discarded (Ru et al. 2015). Markers for MAS have been developed for many annual and perennial crops, including apple and cannabis, positioning MAS as a promising approach for making breeding more efficient.

In apple, numerous markers have already been developed for fruit quality traits (Baumgartner et al. 2016). The application of MAS for storability within the Washington State apple breeding program was estimated to lead to a 60% reduction in costs compared to conventional selection methods (Edge-Garza, Peace, and Zhu 2010). In cannabis, markers have been developed to distinguish flower sex and drug- versus fiber- type cannabis (Sawler et al. 2015; Mendel et al. 2016; Borin et al. 2021). Despite the existence of markers for traits in apple and cannabis, there are still numerous commercially relevant traits for which markers have not been identified. Furthermore, in certain cases existing markers have been found to hav variable success in predicting phenotypes. For example, studies in apple have shown that existing markers for firmness-related traits displayed low predictive power for firmness phenotypes (K. A. McClure et al. 2018; Migicovsky, Yeats, et al. 2021; Costa et al. 2013; Nybom et al. 2013). Therefore, continuing to query the apple and cannabis genomes for genetic markers associated with relevant traits is crucial for the development and improvement of markers that can be leveraged for MAS.

Although the development of genetic markers for MAS has been a key approach for genomic assisted breeding, advances in genome editing now allow for the direct, targeted manipulation of genomes. CRISPR-Cas9 is a relatively new and powerful gene-editing tool that can be used to edit the DNA of any organism (Doudna and Charpentier 2014). Gene editing enables the introduction of numerous forms of targeted mutations that have otherwise been difficult to achieve with trans- and cis-genic approaches. For example, gene editing allows direct alterations of the genomes without the introduction of transgenes and can induce these alterations with high specificity. Despite successful

applications of gene editing to alter genes in apples and cannabis, there remain numerous technical challenges that prevent the routine use of gene editing for the development of improved cultivars (Osakabe et al. 2018; Nishitani et al. 2016; X. Zhang et al. 2021). These challenges include but are not limited to, difficulties in regeneration, editing efficiency and overall limitations of the current editing tools. Despite the challenges, gene-editing tools, like CRISPR, hold promise to accelerate the breeding process by bypassing years of bi-parental crossing by instead editing genes directly to achieve a desired phenotype. To advance genomics assisted breeding in apple and cannabis, whether for the development of gene editing targets or markers for MAS, it is necessary to identify genomic regions associated with key traits.

*Genome-wide association studies*

Genetic mapping is a useful approach that aims to locate quantitative trait loci (QTLs) that control phenotypic variation. There are several methods to conduct genetic mapping, such as linkage mapping and association mapping. Association mapping takes advantage of historical recombination events to link genome fragments to phenotypes (C. Zhu et al. 2008). An increasingly popular association mapping technique is genome-wide association studies (GWAS) (C. Zhu et al. 2008). GWAS makes use of genetic and phenotypic data from diverse populations in which relatedness is not controlled to identify associations between genetic variants and phenotypic traits (Myles et al. 2009). These associations then may allow for causal genetic loci of the traits to be identified (Myles et al. 2009). In comparison to other genetic mapping approaches, GWAS are able

to query a large amount of genetic diversity, thereby capturing numerous recombination events and thus providing high mapping resolution (M. A. Khan and Korban 2012).

An important concept in GWAS is linkage and linkage disequilibrium (LD). Linkage is the physical connection of loci on a chromosome that results in correlated inheritance. This connection can be broken between successive generations by recombination that breaks the chromosomal segments apart resulting in linkage decay (Bush and Moore 2012). Linkage disequilibrium is an allelic relationship within a population whereby the alleles at two or more loci are correlated and inherited together, in other words it is the degree of non-random association of alleles at different loci (Bush and Moore 2012). GWAS aims to exploit LD to identify marker-trait associations of SNPs that are either in LD with the gene that causes the trait or is the causal variant itself so that the genetic cause of an observed phenotype can be elucidated (A. Khan and Korban 2012). GWAS allows for higher mapping resolution and power to detect genetic effect for traits compared to linkage mapping, because it makes use of accumulated recombination events that have taken place over time in a population (C. Zhu et al. 2008). The higher resolution is due to the fact that over successive generations, LD decays due to recombination and the markers used within the GWAS are able to more closely tag causal genetic elements (Walsh and Jannink, n.d.).

There are numerous different methods employed to conduct GWAS, one such method that has risen to popularity is the unified mixed linear model (MLM) (Tibbs Cortes, Zhang, and Yu 2021). The MLM was developed to take into account population structure

and relatedness amongst individuals within the studied population (J. Yu et al. 2006). Accounting for population structure and relatedness is important in GWAS, because traits that correlate with the underlying population structure will contribute to higher numbers of false positives and detection of spurious associations (Bradbury et al. 2007) (Hall, Tegström, and Ingvarsson 2010). The MLM therefore includes population structure (Q) as a fixed effect and relatedness (K) as random effect within the model to account for multiple levels of relatedness (J. Yu et al. 2006). The bioinformatic program TASSEL (Trait Analysis by Association, Evolution and Linkage) is widely used to execute unified MLM Q+K model for the purpose of genetic mapping in plants (Bradbury et al. 2007). The MLM is employed a single-locus test, however, recent advances have been made to develop multi-locus mixed models (MLMM) that are capable of testing multiple markers simultaneously as covariates within the model (Tibbs Cortes, Zhang, and Yu 2021). Segura et al. developed the first MLMM that is especially useful for GWAS on traits suspected to have complex genetic architectures that could be controlled by multiple large-effect loci (Segura et al. 2012). New tools and software are rapidly being developed for conducting GWAS, and careful consideration of the examined population as well as the trait of interest dictates which methodologies are most appropriate.

*Genotyping and Phenotyping*

To capitalize on the power of GWAS to detect genotype-phenotype associations and unravel the genetic architecture of key traits, sufficient high-quality genetic and phenotypic data must be collected. Fortunately, genome-wide marker data are becoming

more economical to collect and are increasingly easily available due to advances made with next-generation sequencing technologies. Genotype-by-sequencing (GBS) is an approach that uses restriction enzymes to generate reduced representation libraries that can enable fast and cost-effective next-generation sequencing of diverse species (Elshire et al. 2011). Unlike SNP arrays, which have commonly been used for genetic mapping purposes, GBS avoids ascertainment bias and therefore is highly amenable approach for sequencing high diversity species (Peace et al. 2019; Elshire et al. 2011). SNPs make up one of the most common forms of genetic variation, however, structural variation is increasingly being shown to control phenotypic variation in plants (Saxena, Edwards, and Varshney 2014). Whole genome sequencing (WGS) not only allows for the exhaustive identification SNPs but also structural variants. Although cost of WGS has been decreasing, and is expected to continue to do so (Nordborg and Weigel 2008; C. Zhu et al. 2008), WGS currently remains prohibitively expensive especially for GWAS aimed at sequencing large populations.

A major barrier to high-powered GWAS is high-quality trait data from suitable populations. This limitation is often referred to as “the phenotyping bottleneck” because the collection of trait data will remain expensive and time consuming, while the cost of collecting DNA sequence data continues to drop (Furbank and Tester 2011). To overcome the phenotyping bottleneck many breeders have turned to high-throughput methods, using robotics and imaging technologies to speed up phenotyping (Rouf et al. 2019). In addition, existing phenotyping databases that contain historical phenotype data have been leveraged for use in GWAS of multiple apple traits (Migicovsky et al. 2016).

Similarly, there exists publicly available databases that have catalogued phenotype data for numerous cannabis strains and recent studies have proposed using these databases for the purpose of conducting GWAS (Aardema and DeSalle 2021). Despite the advancement in phenotyping technologies and strategies, challenges still exist, especially when attempting to measure a large number in phenotypes from diverse germplasm collections (Rouf et al. 2019).

*GWAS Challenges*

Due to the large number of markers that are involved in GWAS and the subsequent high number of statistical tests between markers and traits, the rate of false positives is high (Bush and Moore 2012). It is therefore important to correct for multiple testing to reduce the rate type I errors from GWAS results (Bush and Moore 2012). One method to correct for multiple testing in GWAS is with the Bonferroni correction that adjusts the p-values based on the number of SNPs tested. However, the Bonferroni correction tends to be conservative and leads to increased false negative results (de Kluiver et al. 2018). SNPs are generally considered significant with a P-value $<5 \times 10^{-8}$ (Schaid, Chen, and Larson 2018). However, a SNP with a significant p-value in association with a trait does not confirm that it is the causal variant for the phenotypic trait (Schaid, Chen, and Larson 2018). This is one of the major pitfalls of GWAS because the causal variant can remain elusive do to the fact that the causal SNP for the phenotype might not have been genotyped, rather a SNP that is in LD with the causal SNP is identified, thus only an indirect association is demonstrated (Bush and Moore 2012). Due to the possibility of

identifying only an indirect association between a SNP and the phenotype, further work fine-mapping is often required to find the causal genetic variant responsible for the phenotype (Schaid, Chen, and Larson 2018). Fine-mapping involves searching the region in which the indirect SNP is located to find the causal genetic loci based on the region's LD structure and other known genes in the area (Schaid, Chen, and Larson 2018).

In apple and cannabis, the development of genomic resources and thus the ability to perform GWAS has not been without shared challenges. A particular challenge is the high levels of heterozygosity present within the apple and cannabis genomes (X. Sun et al. 2020; Hurgobin et al. 2020). Both the apple and cannabis genomes are highly heterozygous due to their outcrossing nature that typically prevents selfing, and this heterozygosity renders the assembly of reference genomes and SNP calling challenging (Velasco et al. 2010; Xuewei Li et al. 2016; Hurgobin et al. 2020). Despite the challenges, the development of additional reference genomes holds promise to advance the genomic resources available for undertaking high-powered GWAS of these economically and culturally important crops.

**Concluding remarks**

Apple and cannabis share unique similarities, they both possess a diversity of end uses, are commonly clonally propagated, and widely cultivated around the world. This thesis aimed to advance genomics assisted breeding in apple and cannabis by characterizing genomic and phenotypic diversity and conducting GWAS to identify the genetic control of key traits in both crops. In Chapter 2, I carried out a phenomic characterization of a

diverse collection of apples to quantify apple diversity across fruit quality and phenology traits both at harvest and after three months of storage. In Chapter 3, I paired the phenotypic data from Chapter 2 with over 260,000 genetic markers sequenced across over 1000 accessions in the Apple Biodiversity Collection (ABC) to conduct GWAS to uncover genetic markers associated with fruit quality and phenology traits. In Chapter 4, I used over 100,000 genetic markers across over a 100 cannabis samples to conduct GWAS for over 40 chemical compounds and examine the genetic and chemical basis for cannabis labelling. Finally, in Chapter 5, I used apple as a case study to outline a pathway of how putatively causal genetic variants from genetic mapping studies can be acted upon to develop new cultivars through gene editing. While cannabis and apple each face unique challenges in their cultivation and breeding, the development of new cultivars, in both crops, can benefit immensely from the advancement of genomics resources.

# Chapter 2: Quantifying apple diversity: a phenomic characterization of Canada's Apple Biodiversity Collection

Watts, S., Migicovsky, Z., McClure, K.A., Cindy, H.J., Amyotte, B., Baker, T., Bowlby, D., Burgher-MacLellan, K., Butler, L., Donald, R., Fan, L., Fillmore, S., Flewelling, J., Gardner, K., Hodges, M., Hughes, T., Jagadeesan, V., Lewis, N., MacDonell, E., MacVicar, L., McElroy, M., Money, D., O'Hara, M., Ong, Q., Campbell Palmer, L., Sawler, J., Vinqvist-Tymchuk, M., Rupasinghe, HP V., DeLong, J.M., Forney, C.F., Song, J., Myles, S. (2021). Quantifying Apple Diversity: Phenomic Characterization of Canada's Apple Biodiversity Collection. Plants, People, Planet, 3, 6. https://doi.org/10.1002/ppp3.10211

**Author contributions**

Sophie Watts conducted the harvest in 2017 and coordinated the phenotyping of the collection. Sophie conducted the data analysis and wrote and edited the manuscript.

**Abstract**

A future with a secure and safe food supply requires humanity to preserve and exploit the vast variation available across agricultural plant species. Apples are one of the most widely consumed fruits and provide significant nutritional value worldwide. Here, we characterize key agricultural traits in a diverse collection of apples to provide a foundation for future apple improvement. We show that commercially successful apple varieties capture only a small fraction of apple diversity and demonstrate that significant

improvement is possible by tapping into existing genetic diversity. Here we present a comprehensive evaluation of apple diversity through phenotyping of Canada's Apple Biodiversity Collection (ABC) which contains over 1000 apple accessions. We assessed, over a 4-year period, more than 20,000 individual apples and quantified variation across 39 phenotypes, including phenology and fruit quality both at harvest and after 3 months of cold storage. We observe that apples in the ABC display a wide range of phenotypic variation that may prove useful for future apple improvement. For example, apples can differ by nearly 61-fold in weight, 18-fold in acidity, and 100-fold in phenolic content. We quantified the dramatic changes to apple physiology that occur during 3 months of cold storage: on average, apples lost 39% of their firmness, 31% of their acidity, and 9% of their weight, but gained 7% in soluble solids. Harvest date, flowering date, and time to ripen were all positively correlated with firmness, which suggests that the developmental pathways that drive phenological events throughout the growing season may play a role in determining an apple's texture. Finally, we show that apple breeding has selected for a significant decline in phenolic content over the past 200 years: apple cultivars released after 1940 had a 30% lower median phenolic content than cultivars released before 1940. The data and analyses presented here not only provide a comprehensive quantification of the range across, and relationships among diverse apple phenotypes, but they also enable genetic mapping studies that will provide the foundation for future apple improvement via genomics-assisted breeding.

## Introduction

The domesticated apple (*Malus domestica*) is thought to have been cultivated for over 3,000 years (Zohary, Hopf, and Weiss 2012). Genetic evidence suggests that the main progenitor species of the domesticated apple is *Malus sieversii* from Central Asia (Velasco et al. 2010); however, significant gene flow from *Malus sylvestris* has also been detected (X. Sun et al. 2020; Duan et al. 2017; Cornille et al. 2012). Red color, reduced acidity, larger fruit, and firmness appear to have been under selection during apple domestication and improvement (Ma et al. 2015; Migicovsky, Gardner, et al. 2021). The apple's widespread geographic distribution and long-standing popularity have resulted in over 10,000 named apple cultivars with a fascinating diversity of phenotypes (Liang et al. 2015). Despite this tremendous diversity, a small number of cultivars make up a significant proportion of production. For example, in 2018, only four cultivars accounted for over 50% of apple production in the United States ("US Apple and Pear Forecast" n.d.). In addition, commercial apples have been shown to be closely related to each other genetically (Migicovsky, Gardner, et al. 2021; Noiton and Alspach 1996). Reliance on a limited number of closely related cultivars means that consumers experience an extremely limited fraction of the apple's genetic diversity. However, the extent of phenotypic diversity captured by the top commercial cultivars remains to be quantified.

Fruit firmness is strongly associated with consumers "likeness" of cultivars and is, therefore, a key target for breeders (Szczesniak 2002; Nybom et al. 2013). In addition, breeding apples that retain their firmness after long-term storage have been a key breeding target (Kouassi et al. 2009). Not only is it important for breeders to select for

firmness and firmness retention, but it is also crucial to breed for the phenological traits associated with firmness and firmness retention. This is especially important given that the optimal phenological breeding targets are expected to shift over time due to climate change. Characterizing how fruit texture and phenological traits are associated with each other can enable the development of new cultivars that adapt to climate change and meet consumer preferences.

Germplasm collections serve as important reservoirs of genetic diversity for crop improvement. They contain the high levels of diversity that are essential for identifying valuable phenotypes that can be leveraged to develop improved cultivars (Bramel and Volk 2019). Extensive germplasm resources are already available for major crops such as maize and rice which can be stored as seeds, but resources for perennial fruit crops are lacking as they are expensive to establish and maintain (Flint-Garcia et al. 2005; Jackson 1997; Migicovsky and Myles 2017). Despite the high cost, living germplasm collections are essential for woody perennials which are difficult to conserve in seed banks or tissue culture. There is increasing interest in using germplasm collections to not only preserve biodiversity and exploit valuable phenotypes but also to perform genetic mapping (Migicovsky et al. 2019). The high genetic diversity and sample sizes present in many germplasm collections can result in well-powered genetic mapping studies that identify genetic markers useful for genomics-assisted breeding (C. Zhu et al. 2008). In addition, germplasm collections often contain wild relatives with novel traits that can be introgressed into elite cultivars (Migicovsky and Myles 2017).

There are numerous apple germplasm collections that have been established worldwide (e.g., USA (Gross et al. 2013), China (Yuan Gao et al. 2015), New Zealand (Kumar et al. 2010), and Europe (Jung et al. 2020)). Within Canada, there are apple collections located across the country, for example, in Ontario and British Columbia (Hampson et al. 2009; Ward 1978). Most recently, Canada's Apple Biodiversity Collection (ABC) was established in Nova Scotia, Canada, with over 1,000 accessions that includes trees primarily belonging to the domesticated apple, *M. domestica*, and its primary wild ancestor, *M. sieversii*. The ABC was established as a dual-purpose orchard. The first purpose of the ABC is to preserve and maintain potentially valuable apple genetic and phenotypic diversity. Second, the ABC is specifically designed to enable accurate measurements of phenological and fruit quality traits primarily for the purposes of genetic mapping. The trees in the ABC were grafted at the same time to the same rootstock and planted in duplicate in a randomized block design to control for positional effects in the orchard. The result is an apple population that is ideally suited to accomplish a phenome-wide characterization of apples. Here we present a comprehensive evaluation of Canada's ABC through phenotyping of phenological traits and fruit quality traits both at harvest and after 3 months of cold storage.

## Materials and Methods

### *The Apple Biodiversity Collection*

The Apple Biodiversity Collection (ABC) is located at the Agriculture and Agri-Food Canada (AAFC) Kentville Research and Development Centre in Nova Scotia, Canada

(45.071767, −64.480466). The ABC contains 1,119 apple accessions that were grafted to M.9 rootstock in August 2011 and allowed to grow outdoors until November 2012 when they were removed from the orchard and stored in moist sawdust at 2°C until planting. On May 31, 2013, we planted each of the 1,119 accessions in duplicate in a 5-acre orchard that was tile drained and fumigated with Telone® soil fumigant. The trees were spaced 1.5 m within rows and 5 m between rows. The trees were trained to a trellis system with wires at 1.5 and 2.4 meters above the ground. Soil amendments, training, thinning, and pruning were performed to industry standards.

The ABC consists of apple accessions from the United States Department of Agriculture (USDA) Plant Genetic Resources Unit apple germplasm collection in Geneva, New York, USA; commercial cultivars from the Nova Scotia Fruit Growers' Association Cultivar Evaluation Trial; and advanced breeding material from the AAFC Kentville breeding program. The collection contains mostly *M. domestica* accessions including cider, dessert, processing, heritage, and elite cultivars. The orchard also contains 78 accessions of the wild progenitor species *M. sieversii* from Central Asia. It is possible that pairs of accessions within the ABC may be clonally related in some cases. Here we treat each accession as a unique sample in downstream analyses and future genetic investigation will reveal the degree of clonal relatedness in the collection. The trees in the ABC are not available for propagation as most of the material was imported from the USDA under a section 43 import permit from the Canadian Food Inspection Agency (Permit #P-2011–00222) which prohibits the sale or distribution of the germplasm.

Experimental design

The ABC was planted in an incomplete randomized block design with a North and South block (Appendix I: Figure I-I). Each block is divided into grids that include nine trees within a row across three adjacent rows. Within each grid is a "check tree" of "Ambrosia", "SweeTango", or "Honeycrisp", amounting to 18 check trees of each of these cultivars across the orchard. These "check trees" are used as standards in the restricted maximum likelihood (REML) model that accounts for phenotypic variation due to position in the orchard. We used the following model to generate phenotype values:

$$phenotype \sim accession + (1|Block) + (1|rGrid) + (1|cGrid) + (1\,|cGrid{:}rGrid)$$

This mixed-model accounts for fixed effects of an accession and the random effects of position depending on Block (north/south), north-to-south position within the block (rGrid), east-to-west position within the block (cGrid), and the interactions between these random effects. The measurements resulted in 39 phenotypic variables collected from 2014 to 2018, which are summarized in Appendix I: Table I-I.

*Phenology*

All phenology traits that were recorded as dates (e.g., harvest date and flowering date) were converted into Julian days. Harvest date was recorded for the 2016 and 2017 harvest seasons. During both seasons, 20 apples were picked randomly from each tree. Trees ready for harvest were flagged at the beginning of the week and harvested over the

subsequent days of that week. Due to the diversity of the accessions and the variation in ripening time, a variety of methods were used to determine when to harvest. Dropped apples or changes in background skin color were indicators of harvestable trees (Watkins 2003). In addition, a sample apple was taken from each tree and touched to assess firmness, tasted to assess starch and sweetness, cut in half to check browning of seeds, and sprayed with iodine solution to evaluate starch content (Blanpied and Silsby 1992). The combination of these indicators was used to determine whether an accession was ready to be harvested. Flowering date was measured as the date when the youngest wood displayed >80% of flowers at the king bloom stage (K. McClure 2017). Flowering date was recorded in 2016, with trees being assessed every 3 days during this period. Time to ripen was calculated as the time (in days) between flowering date and harvest date. Precocity was recorded as the year in which a tree first bloomed after establishment in the orchard and converted into a score: a score of 1 corresponded to 2014, a score of 2 corresponded to 2015, a score of 3 corresponded to 2016, and a score of 4 indicated that the tree had not yet bloomed as of 2016.

*Fruit quality*

The fruit quality traits measured included weight (g), firmness ($kg/cm^2$), acidity (g/L malic acid), soluble solids content (SSC) (°Brix), and juiciness ((weight of juice/total fruit weight) × 100). In 2017, measurements were taken from 5 apples, while a sample of 10 apples was used in 2016 whenever possible and a mean measurement for each tree was calculated. Measurements were not recorded for trees with fewer than three

representative apples available for assessment. In 2016, an automated phenotyping machine was used (the Pimprenelle, Setop Giraud Technologie, France—**https://www.setop.eu/fr/produit/pimprenelle**) to collect fruit quality data. In 2017, measurements were collected manually as described below. Since the measuring techniques differed slightly between 2016 and 2017, we assessed the correlation between years for each trait (Appendix I: Figure I-II).

Total weight (g) of all the apples in a sample were measured (Mettler-Toledo, MS3002S), and an average weight per apple was calculated by dividing the recorded total weight by the number of apples in the sample. Firmness measurements ($kg/cm^2$) for five individual apples were recorded using a penetrometer (Guss Fruit Texture Analyser, GS-14). Using a vegetable peeler, a small section of skin along the side of the apple was removed to allow the penetrometer to enter the apple flesh. Each apple was then placed on the penetrometer platform so that the piston entered the middle of the apple where the skin had been shaved off. There was no preference for the side of the apple from which the measurement was taken. The firmness measurements were averaged to get a mean value for the tree.

To measure acidity in a sample of apples, a quarter of each apple was juiced using a handheld garlic press. The juice from the apples belonging to a single tree was combined to make a composite juice sample. Titratable acidity was measured using the 865 Dosimat Plus (Metrohm). This involved titrating 1 ml of the composite juice sample with 0.1 M of NaOH. Soluble solids content (SSC) was measured on the composite juice

sample using the Pocket Refractometer (Atago, PAL-1). The SSC/acid ratio was calculated by dividing SSC measurements by acidity measurements.

After harvest, apples that were not phenotyped were placed in cold storage (4°C). After 3 months in storage, these apples were removed and placed at room temperature until fruit quality could be measured on the following day using the same methods described above. Thus, apples from each accession were assessed for fruit quality both at harvest and after 3 months of storage. Post-storage measurements were taken from a minimum of 3 and a maximum of 10 apples per accession.

During the 2016 harvest, whole fruit samples were collected and ground into a fine powder using liquid nitrogen and stored at −80°C. Frozen ground samples (approximately 0.5 g each) were extracted in duplicate for the quantification of phenols using sonication in an 80% methanol solution in two 2 ml microcentrifuge tubes. A ferric-reducing antioxidant power (FRAP) assay for total antioxidant capacity (TAC) was performed to detect the reduction potential of compounds by measuring the intensity of the violet-blue color produced. FRAP reagents were prepared as described previously (Benzie and Strain 1996; Rupasinghe et al. 2008). The ferric reducing antioxidant powers of apple extracts were reported in micromolar of Trolox equivalents (TE) per gram fresh weight of an apple. The Folin–Ciocalteu assay estimates the total phenolic content (TPC) present in plant foods and was carried out as previously described (Huber and Rupasinghe 2009; Singleton, Orthofer, and Lamuela-Raventós 1999). TPC of apple extracts was reported in micromolar of gallic acid equivalents (GAE) per gram fresh

weight of an apple. TPC and TAC were strongly correlated ($\rho = 0.948$; $p < 1 \times 10^{-15}$; Appendix I: Figure I-III), so we only present TPC in the main body of the manuscript.

*Apple classification*

Apple accessions were classified into binary categories based on species (*M. domestica* vs. *M. sieversii*), geographic origin (new world vs. old world), and end use (cider vs. dessert) using information retrieved primarily from the USDA Germplasm Resource Information Network (GRIN; **https://www.ars-grin.gov/**), as well as online sources. Accessions classified as *M. pumila* (N = 5) in the GRIN database were considered *M. domestica* in our analyses since these terms are synonymous. Accessions classified by GRIN as *M. sylvestris* (N = 3) or hybrids (N = 5) were excluded from analysis of comparisons between species. All accessions that were listed as dessert, cooking, or eating were classified as "dessert" (N = 804) while only those labeled as cider were classified as "cider" (N = 104). Accessions labeled as wild (N = 78), crab (N = 5), and rootstock (N = 1) cultivars were excluded from the analysis of comparisons between cider and dessert. For geographic origin, accessions that originated in Europe or Asia were classified as "old world" (N = 450), while those that originated in North America, South America, South Africa, Australia, and New Zealand were classified as "new world" (N = 529). The year of release was also retrieved from the GRIN database or from online sources for 343 accessions that were introduced as named cultivars.

*Statistical analysis*

All data curation, handling, and analysis were performed in R (R Core Team 2018). Each accession was planted in duplicate, but if one of an accession's two trees died or did not produce fruit, then the data came from only a single tree. The phenology and fruit trait measurements before storage were adjusted for their location in the orchard by running the REML model (see above) using the "lme4" package (Bates et al. 2015) in R, which resulted in one adjusted mean measurement per accession. Phenolic measures (TPC and TAC) were not adjusted since, most often, only a single tree was sampled per accession. Fruit quality measurements taken after storage were also not adjusted for location in the orchard because, for each accession, the fruit from the two duplicate trees was combined before being placed in cold storage. Thus, post-storage measurements are simply the mean value across apples within an accession. To evaluate how fruit quality changed during storage, the difference between the measurements before storage (adjusted) and after storage (unadjusted) was calculated for each accession.

Linear correlations between phenotypes were assessed using Pearson correlations via the "cor.test" function in R. To assess how the most commercially successful cultivars differ from the remainder of the collection, we compared phenotypes from 9 of the top 10 cultivars produced in the USA ("US Apple and Pear Forecast" n.d.) that are found in the ABC to the phenotypes of the remaining accessions in the ABC. The 9 of the top 10 cultivars that are present in the ABC are as follows: "Red Delicious," "Gala," "Fuji," "Granny Smith," "Honeycrisp," "Golden Delicious," "McIntosh," "Cripps Pink," and

"Empire." The median, standard error, minimum, and maximum trait measurements were calculated for the "Top 9" and for the other ABC accessions. Mann–Whitney U tests were performed to determine whether measurements of the "top 9" differed significantly from the measurements of the other ABC accessions. The proportion of the ABC's range in trait measurements encompassed by the "top 9" cultivars was calculated by dividing the range in measurements of the "Top 9" by the range of the entire ABC. Mann-Whitney U tests were also performed to determine whether phenotypic traits differed significantly among *Malus* species, geographic origin, and end use classifications. Standard error was calculated using "std.error" function from the plotrix package in R (Lemon 2006). The Bonferroni correction was applied to account for multiple testing where appropriate.

**Results**

Canada's Apple Biodiversity Collection (ABC) aims to capture much of the world's diversity of apples and contains 1,119 accessions from 47 countries, with a strong emphasis on apples from Canada (N = 278) and the USA (N = 234) (Figure 2-1). The measurements from over 20,000 individual fruit from more than 1,000 unique apple accessions across 2 years are presented in Appendix I: Table I-I. Numerous phenotypes were measured in both 2016 and 2017. With the exception of flowering date, juiciness, and precocity, all other phenotypes presented in the main body of the manuscript are from 2017 due to the larger sample size in that year. The correlations between years for phenotypes ranged from 0 to 0.69, and 13 of 17 of the phenotypes had significant ($p < 0.05$) between-year correlations (Appendix I: Figure I-II). The weakest correlations were

for the phenotypic changes observed during storage. For example, the weakest correlation was for change in weight during storage ($R^2 = 1.2 \times 10^{-4}$, $p = 1$). The strongest correlation between years was for harvest date ($R^2 = 0.691$, $p < 1 \times 10^{-15}$).

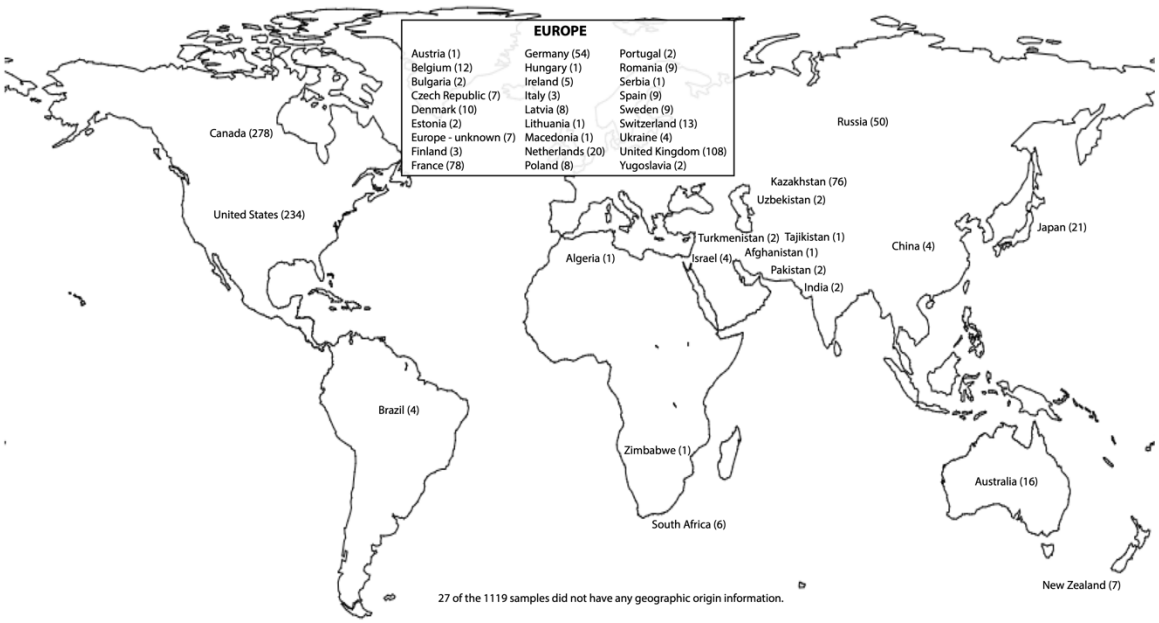


**Figure 2-1.** World map showing geographic origins of apple accessions from Canada's Apple Biodiversity Collection.

*Phenology and fruit quality traits*

The distributions of phenology and fruit quality traits are shown in Figure 2-2. The distributions of these traits for the year 2016 are provided in Appendix I: Figure I-IV. The summary statistics for each of the phenotypes are provided in Appendix I: Table I-II. Flowering date spanned 21 days with "Kaz 95 18-02P-33," a *M. sieversii* accession, flowering first on May 22nd and "Frostproof," a *M. domestica* cultivar, flowering last on

June 12th. Harvest spanned 65 days; more than three times as long as flowering. The earliest harvested accession was “C.P. Close” on August 12th and the latest harvested accession was “Red Spy” on October 17th. Time to ripen spanned 68 days. “Lunost” had the shortest time to ripen (71 days) and “KAZ 7” had the longest time to ripen (139 days).

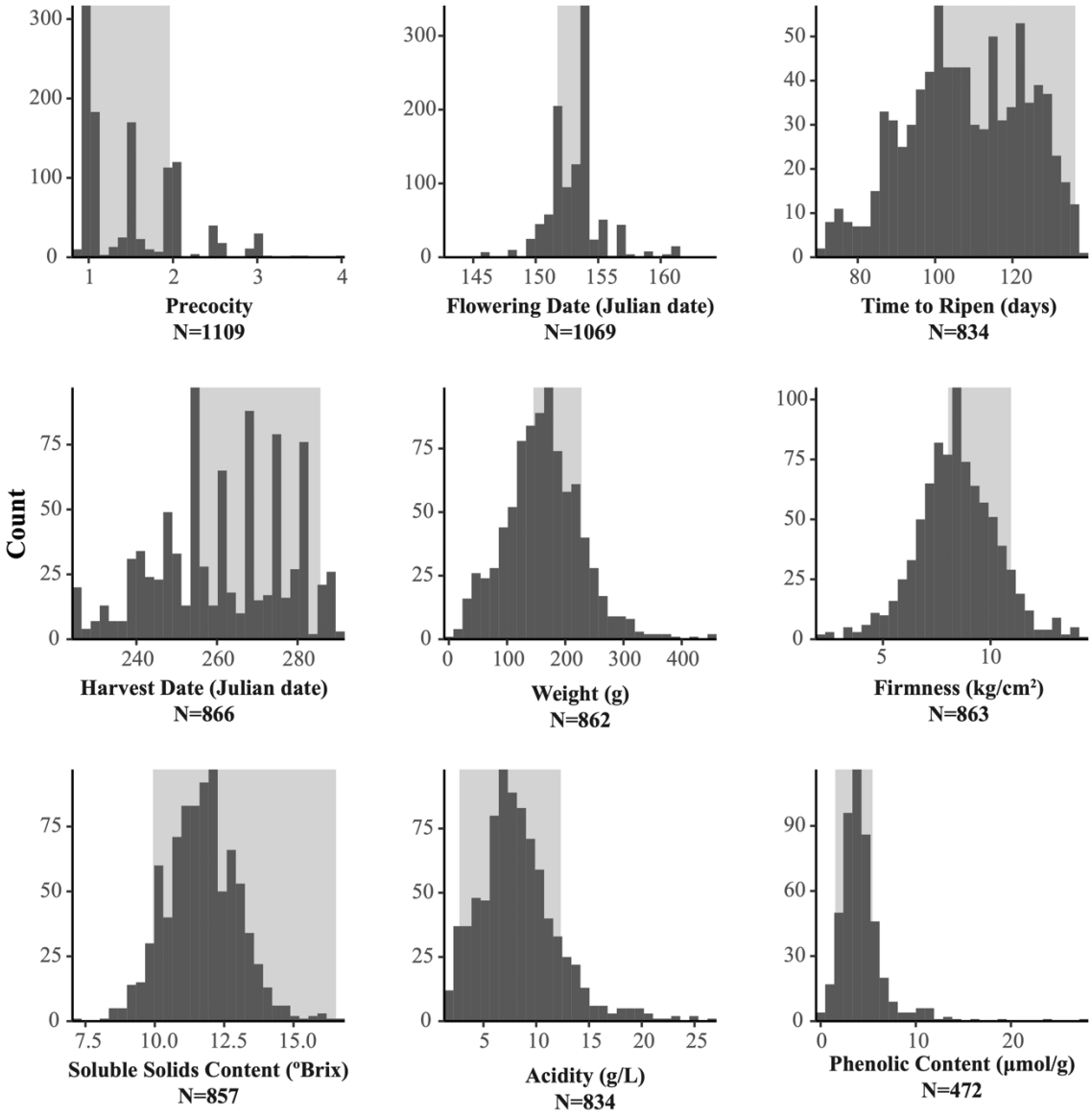

**Figure 2-2.** Distribution of apple phenotypes from Canada's Apple Biodiversity Collection. Shaded grey areas represent the range of values observed across the "top 9" cultivars grown in the USA.

We observed a 61-fold difference in apple weight among accessions, ranging from 7.57 g to 460 g. The lightest accession was “Uralskoje Nalivnoje” and the heaviest was “Bietigheimer.” Acidity varied by 18-fold: the least acidic (1.41 g/L) accession was “Dunning” and the most acidic (26.35 g/L) was “Barenhecke 3 Klipphausen” a *M. sylvestris* accession. There was a sevenfold difference in firmness among accessions at harvest: the softest accession was “Miron Sacharanij” with a measurement of 1.99 kg/cm$^2$, whereas the firmest accession was “Oekonomierat Echter-meyer” at 14.21 kg/cm$^2$. The least sweet accession was “Julyred” at 7.04 °Brix, and the sweetest accession, “Golden Delicious,” was more than twice as sweet at 16.4 °Brix. Phenolic content varied by nearly two orders of magnitude from 0.293 to 27.9 μmol/g. The two accessions with the lowest phenolic content were “Zestar” and the advanced breeding line “S19–23–52,” while the cider apple cultivar “Marachal” had the highest phenolic content. None of the phenotypes of the “top 9” cultivars differed significantly from the rest of the accessions in the ABC after accounting for multiple tests. The phenotype measurements of the “top 9” cultivars capture between 10.6% and 70.6% of the total variation in the ABC across phenotypes. For phenolic content, the “top 9” cultivars are found close to the median of the ABC and capture only 14.2% of the ABC’s total variation in phenolic content. In contrast, 70.6% of the total variation in soluble solids is captured by the “top 9”, and values for the “top 9” tend to be above the median of the ABC. Precocity ranged

from 1 year to over 4 years across all accessions, but all of the "top 9" cultivars fruited in the first 2 years after establishment in the orchard.

*Fruit quality after storage*

Significant correlations were observed between measurements taken before and after storage for all fruit quality traits (Figure 2-3, Appendix I: Figure I-V). On average, apples lost 31% of their acidity, 39.9% of their firmness, and 9.1% of their weight, but increased in soluble solids by 7.1% during 3 months of cold storage (Appendix I: Table I-II). The change in acidity, firmness, soluble solids, and weight for the "top 9" cultivars did not differ significantly from the rest of the accessions. However, the “top 9” cultivars tended to experience a less severe loss of firmness (29%) and acidity (25.9%) compared to the rest of the ABC.

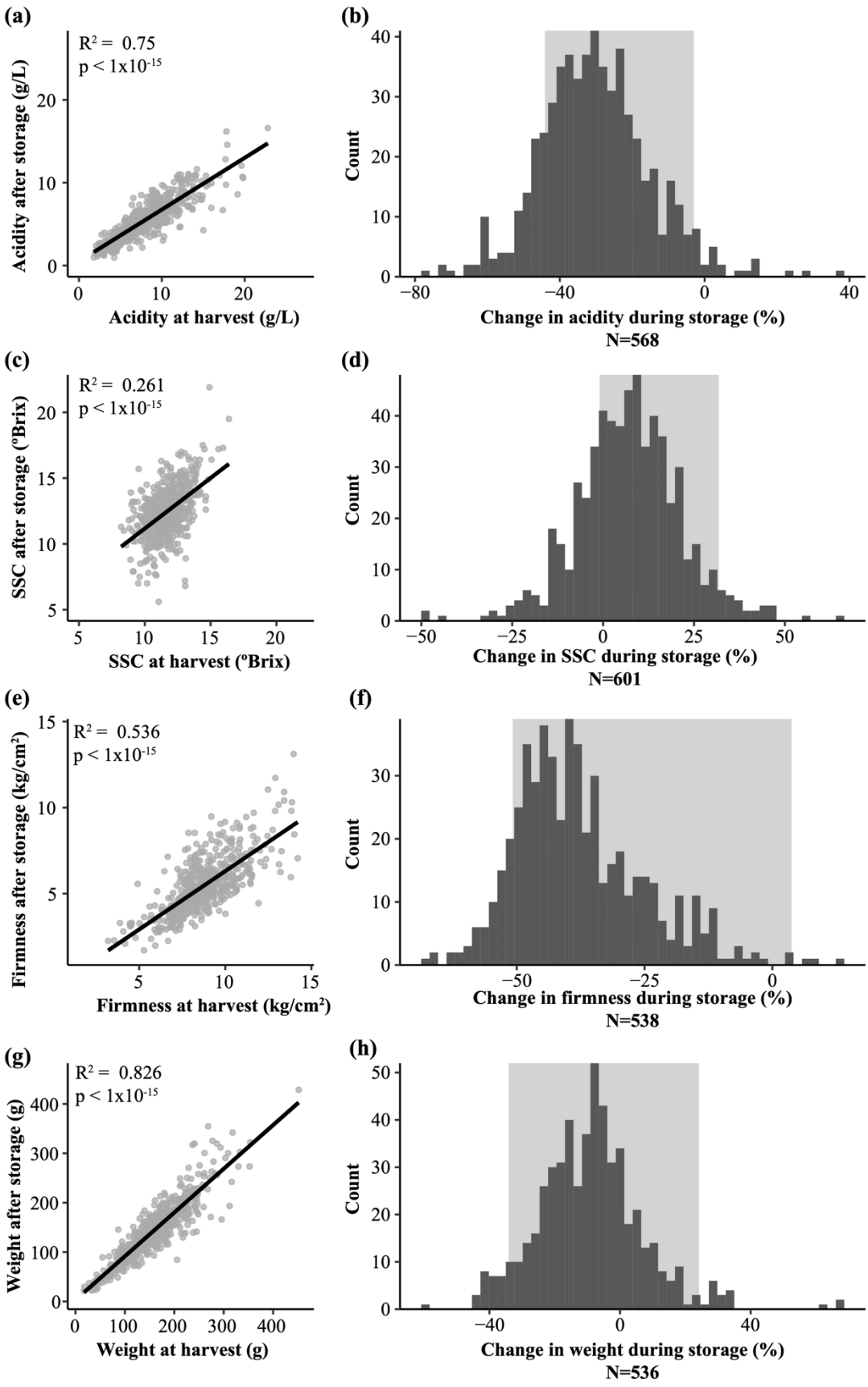
(a)
R² = 0.75
p < 1x10-15
Acidity after storage (g/L)
Acidity at harvest (g/L)
(b)
Count
Change in acidity during storage (%)
N=568
(c)
R² = 0.261
p < 1x10-15
SSC after storage (ºBrix)
SSC at harvest (ºBrix)
(d)
Count
Change in SSC during storage (%)
N=601
(e)
R² = 0.536
p < 1x10-15
Firmness after storage (kg/cm²)
Firmness at harvest (kg/cm²)
(f)
Count
Change in firmness during storage (%)
N=538
(g)
R² = 0.826
p < 1x10-15
Weight after storage (g)
Weight at harvest (g)
(h)
Count
Change in weight during storage (%)
N=536

**Figure 2-3.** Fruit quality measurements before and after 3 months of cold storage. Correlations of fruit quality measures taken before and after storage are found in the left column (a, c, e, g). Distributions of the percent change of each trait during storage are shown in the right column (b, d, f, h). The shaded grey areas represent the range occupied by the "top 9" cultivars.

*Relationships among traits*

We calculated correlations among all the fruit quality, phenology, and storage traits (Figure 2-4, Appendix I: Table I-III). Correlations for the 2016 phenotype data are provided in Appendix I: Figure I-VI and Table I-IV. Of the 105 pairwise comparisons, 32 were significant after correcting for multiple comparisons ($p < .01$). Of those significant correlations, 28 were positive and 4 were negative. Harvest date was significantly correlated with weight ($R^2 = 0.11$, $p < 1 \times 10^{-15}$), juiciness ($R^2 = 0.04$, $p = 1.85 \times 10^{-4}$), percent change in SSC ($R^2 = 0.33$, $p < 1 \times 10^{-15}$), and percent change in acidity ($R^2 = 0.18$, $p = 1.14 \times 10^{-3}$). Juiciness and weight at harvest were significantly correlated ($R^2 = 0.06$, $p = 2.05 \times 10^{-6}$). Phenolic content and weight at harvest were negatively correlated ($R^2 = 0.06$, $p = 1.1 \times 10^{-4}$). Acidity and flowering date were negatively correlated ($R^2 = 0.04$, $p = 4.4 \times 10^{-7}$). Firmness at harvest was correlated with phenology traits, including flowering date ($R^2 = 0.0292$, $p = 5 \times 10^{-7}$), harvest date ($R^2 = 0.246$, $p < 1 \times 10^{-15}$), and time to ripen ($R^2 = 0.259$, $p < 1 \times 10^{-15}$) (Figure 2-5). Change in firmness during storage was positively associated with flowering date

($R^2 = 0.0263$, $p = 1.7 \times 10^{-4}$), harvest date ($R^2 = 0.0847$, $p = 5.8 \times 10^{-12}$), and time to ripen ($R^2 = 0.0939$, $p = 7.8 \times 10^{-13}$) (Figure 2-5).

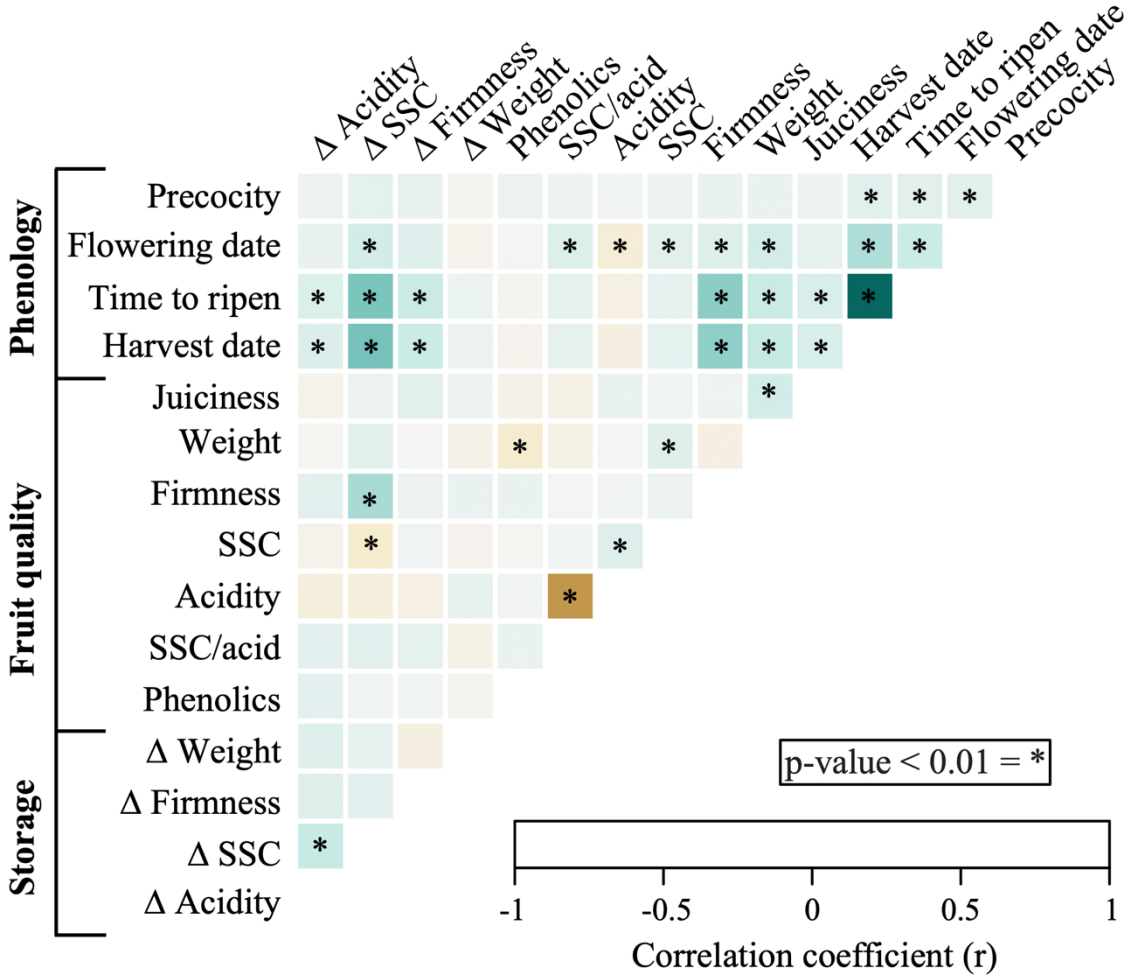


**Figure 2-4.** Heat map of pairwise correlations among phenology, fruit quality, and storage traits. Significant correlations ($p < .01$) after correcting for multiple comparisons are indicated with an asterisk.

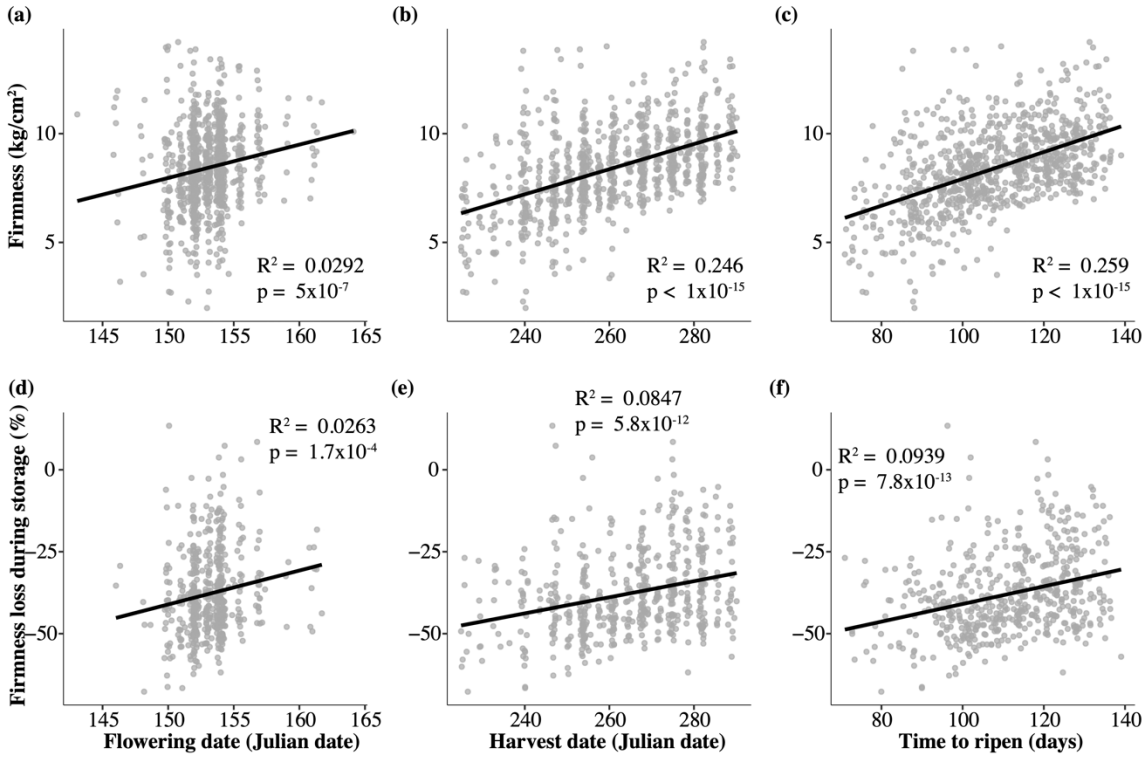


**Figure 2-5.** Correlation of firmness at harvest with (a) flowering date, (b) harvest date, and (c) time to ripen. Correlation of firmness loss during storage with (d) flowering date, (e) harvest date, and (f) time to ripen.

*Phenotypes according to apple classifications*

For 9 of the 15 phenotypes, we observed significant differences between apples classified according to species, geographic origin, and/or use (Figure 2-6). When comparing median values between groups, *M. domestica* accessions flowered 3 days later ($p < 1 \times 10^{-15}$), ripened 14 days later ($p = 1.78 \times 10^{-9}$), were harvested 15 days later ($p = 5.35 \times 10^{-13}$), were 250% heavier ($p < 1 \times 10^{-15}$), and had 67% less phenolic content ($p = 5.63 \times 10^{-3}$) than *M. sieversii* accessions (Appendix I: Table I-V). New world apples flowered 0.76 days earlier ($p = 1.39 \times 10^{-4}$), were 20% heavier ($p = 7.45 \times 10^{-3}$), and had

22% less phenolic content ($p = 3.20 \times 10^{-4}$) than old world apples. Finally, dessert apples flowered 0.75 days earlier ($p = 6.36 \times 10^{-6}$), were 43% heavier ($p = 3.59 \times 10^{-9}$), were 11% softer ($p = 1.74 \times 10^{-4}$), and had 34% less phenolic content than cider apples ($p = 1.12 \times 10^{-7}$).

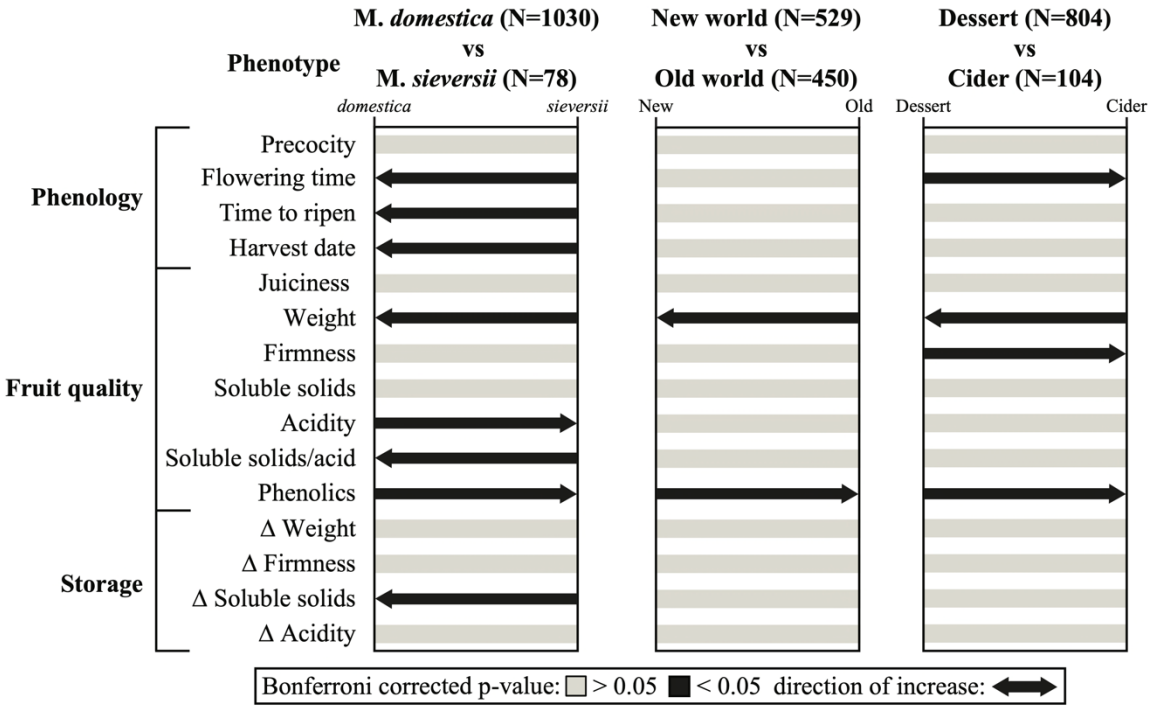


**Figure 2-6.** Comparisons of fruit quality traits among species, geographic origins, and uses. The arrows indicate phenotypes where the difference between groups was significant. The arrow points to the group with the higher value. Apple accessions were classified into binary categories based on species (*M. domestica* vs. *M. sieversii*), geographic origin (new world vs. old world), and end use (cider vs. dessert) using information retrieved primarily from the USDA Germplasm Resource Information Network, as well as online sources.

*Phenotypic changes over time*

Phenolic content was the only phenotype of the 39 tested that showed a significant change over historical time after correcting for multiple comparisons: the year of release of named cultivars was negatively correlated with phenolic content ($R^2 = 0.12$; $p = 1.0 \times 10^{-4}$) (Appendix I: Table I-VI). Apple cultivars released after 1940 had a median phenolic content of 3.38 μmol/g, which is 30% lower than cultivars released before 1940 (4.86 μmol/g) (Figure 2-7).

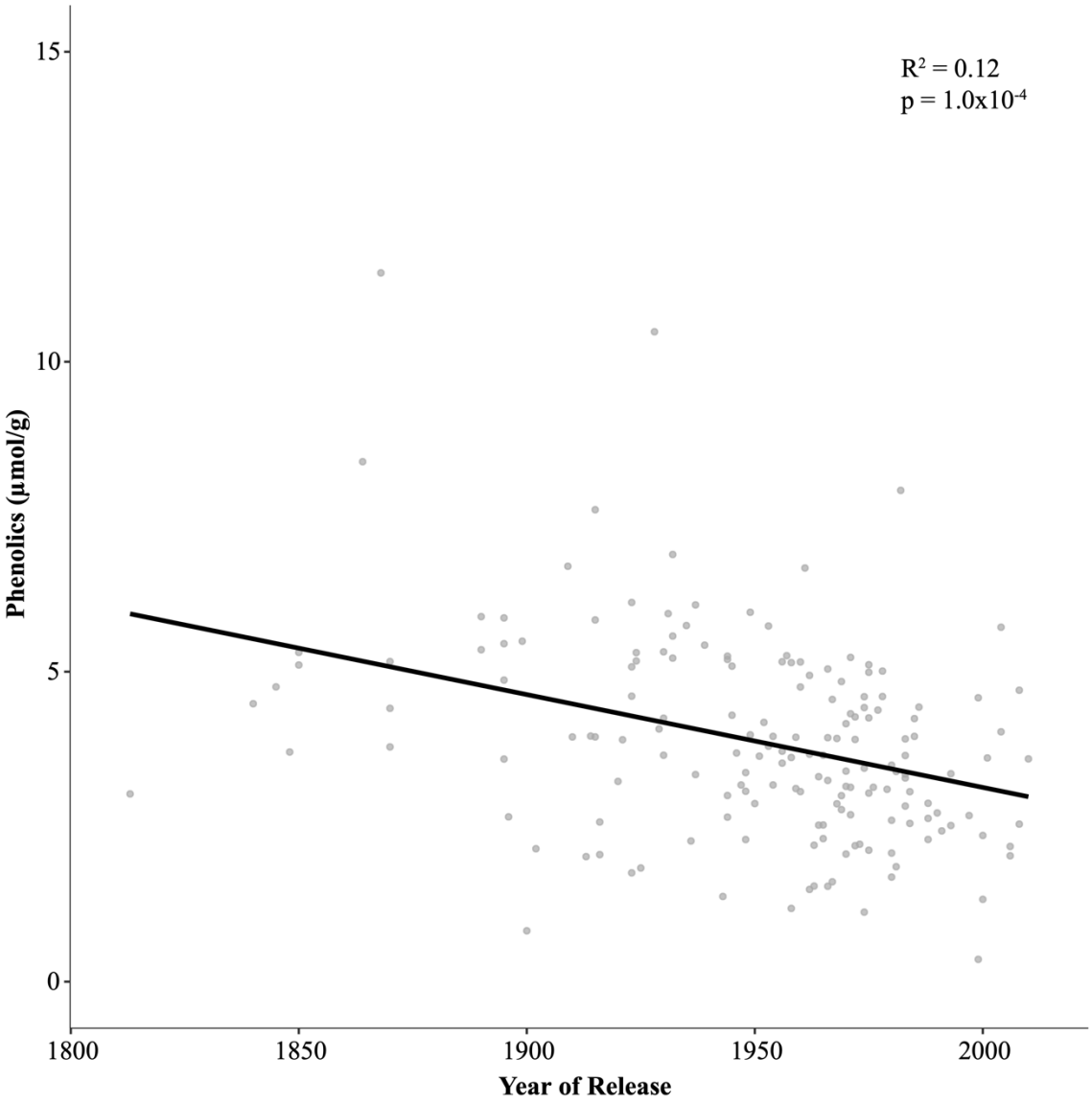


**Figure 2-7.** Apple phenolic content decreased over the past 200 years. Each dot represents a named cultivar. A cultivar's "year of release" refers to its year of commercialization, release to the public, or initial mention in historical records.

## Discussion

Humanity's food security, and the long-term sustainability of agriculture, relies on the preservation and characterization of germplasm collections that house the genetic

diversity required for future crop improvement. Detailed quantification of the phenotypes within these germplasm collections is vital to enabling breeders to more rapidly and efficiently generate improved plant varieties. To this end, our comprehensive apple phenome evaluation involved phenotyping of Canada's Apple Biodiversity Collection (ABC), which contains over 1000 apple accessions. The degree of phenotypic variation in the ABC was substantial for some traits (Figure 2-2). For example, apples can differ by 61 -fold in weight, 18-fold in acidity, and nearly 100 -fold in phenolic content. Phenological traits, such as harvest date, also varied considerably across the orchard, spanning 66 days. In contrast, flowering took place within a 21-day window. Our findings are in line with observations from diverse apples planted throughout Europe where harvest season was also approximately three times longer than the flowering period (Jung et al. 2020). Cultivars that are prone to early floral development under warm temperatures are at increased risk of damage from spring frosts, and late flowering may, therefore, be a desirable breeding target (Gottschalk and van Nocker 2013). Time to ripen is also a critical phenotype for apple producers and we observed a nearly twofold difference in the time to ripen between the cultivar with the shortest (71 days) and longest (139 days) times to ripen. There is a strong preference among producers for fruit that ripens quickly: the longer fruit hangs on the tree exposed to potentially adverse weather events or pests and disease, the higher the risk of fruit damage and yield loss. Therefore, there is interest in breeding late flowering and fast ripening cultivars to reduce the risks imposed by adverse environmental events (Gottschalk and van Nocker 2013; Seeley and Anderson 2003). Our observations provide not only an overall view of the levels of

variation across apple phenotypes but also quantifies phenotypic variation that is informative for breeding improved apple cultivars.

Widening the apple breeding pool to include lesser-known cultivars, as well as wild relatives such as *M. sieversii*, can lead to the development of elite cultivars with improved health benefits, tastes, and fruit quality. The "top 9" cultivars sold in the USA did not differ significantly from the entire collection as a whole for any phenotype, suggesting that the most popular apples in the USA are not outliers in terms of their phenotypes compared to the variation found within the ABC. However, there is considerable room for improvement through conventional breeding to move toward more desirable values for numerous phenotypes. For instance, there are apples within the ABC with acidity levels, phenolic content, ripening times, and flowering dates that lie beyond the values observed within the "top 9" cultivars. With the exception of soluble solids, where "Golden Delicious" was the sweetest accession in the ABC, none of the "top 9" cultivars had extreme values for fruit quality and phenology measurements. In addition, the range of trait measurements for the "top 9" spanned a median of 37% of the variation displayed by the broader collection. "Marachal" had the highest phenolic content, a value five times higher than "Granny Smith," which had the highest phenolic content among the "top 9" cultivars. In addition, the "top 9" only captured 10.6% of the range in flowering date observed across the entire collection, and the "top 9" cultivars took a median of 13 days longer to ripen than the average apple in the orchard. This indicates that the most popular apples in the USA capture only a fraction of the phenotypic diversity in the ABC, and that most phenotypic variation remains largely unavailable to

consumers and producers. In addition, recent genetic analysis of the USDA apple collection showed that over half the collection is related through a series of first-degree relationships due to the prolific use of the most popular cultivars in breeding programs (Migicovsky, Gardner, et al. 2021). A review of the US apple genetic resources indicates that, despite the relatively robust germplasm resources available, the apple industry remains vulnerable due to the limited number of cultivars employed in commercial production (Volk et al. 2015). Although apples did not experience a diversity-reducing domestication bottleneck like many crops, if breeding programs continue to make use of only a small fraction of apple diversity, it could lead to reductions in genetic and phenotypic diversity in the future (Gross et al. 2014; Noiton and Alspach 1996). Exploiting the diversity of *M. domestica* and *M. sieversii* is not only useful to avoid narrowing the genetic base, but it is also critical for introducing novel traits into new cultivars that can meet apple growers' challenges and the demands of consumers.

Storability is a critically important target for breeders because consumers generally eat apples that have been stored for months. Accessions within the ABC that performed well during storage with minimal losses of firmness, acidity, and weight may serve as useful breeding material to develop cultivars with superior storability. When examining how acidity, soluble solids, firmness, and weight changed over 3 months of storage, we found that the direction of change in each trait was not universal across cultivars. For example, while firmness loss is expected during storage, a small number of cultivars appeared to gain firmness during storage (Figure 2-3). These rare observations can be explained by the destructive nature of our assays: the apples from a particular accession measured at

harvest were destroyed, so the apples measured after storage from that accession were not the same apples as those measured at harvest. However, the overall direction of phenotypic change during storage was consistent with previous studies: firmness, acidity, and weight decrease during storage, while SSC increases (Guerra, Sanz, and Casquero 2010). For example, apples generally lost about a third (31%) of their acidity during storage, which is consistent with previous work (Kouassi et al. 2009; Verma et al. 2019). Apple flavor is strongly influenced by acidity, most predominantly malic acid content, and the loss of acidity during storage leads to decreased fruit quality (Harker, Gunson, and Jaeger 2003; Y. Zhang, Li, and Cheng 2010). One of the other key determinants of fruit quality both at harvest and after months of storage is fruit firmness (Guerra, Sanz, and Casquero 2010; Johnston, Hewett, and Hertog 2010). Despite the apple's superior ability to retain its firmness during storage compared to many other horticultural crops, there is still tremendous opportunity to improve apples by breeding novel cultivars that soften very little during extended periods of storage. We observed significant softening during storage: on average, apples lost 39% of their firmness over 3 months of storage at 4°C. However, there were 14 accessions that experienced ≤10% firmness loss (Figure 2-3). Our results suggest that apple breeders' use of cultivars with superior acid and firmness retention during storage could result in novel cultivars with superior eating quality after storage.

Post-storage fruit quality is key to an apple cultivar's commercial success and its biggest threat is loss of firmness due to softening. Here, we focused on the relationship between firmness and phenological traits, such as flowering date, time to ripen, and harvest date.

Previous work has found that late flowering and late harvested apples tend to be firmer (Migicovsky et al. 2016; Nybom et al. 2013; Oraguzie et al. 2004), and we confirmed this relationship here. We found that later harvested apples are firmer at harvest and soften less over 3 months of storage than early harvested apples (Figures 2-4, 2-5). We determined that, on average, for every week of on-tree ripening, an apple experiences a 25% increase in firmness at harvest. It has been observed that early harvested apples have larger cells with greater intercellular space and, therefore, weaker tissue, which may account for their softer texture both at harvest and after storage (A. A. Khan and Vincent 1990; Johnston et al. 2002). In addition, early harvested cultivars produce higher levels of ethylene than late-harvested cultivars, which likely contributes to their softer texture after storage (Watkins 2003). Firmness at harvest is also a strong predictor of firmness after storage ($R^2 = 0.54$, $p < 1 \times 10^{-15}$), indicating that the texture of an apple after storage is largely dictated by how firm it was when picked.

The time it takes an apple to ripen is the period of time between its flowering date and its harvest date. While long ripening periods may appear beneficial because they result in firm fruit that soften less, they also result in increased exposure to disease pressures and potential adverse weather events. In addition, late flowering cultivars are desirable because they are less prone to frost damage (Mehlenbacher and Voordeckers 1991). While growers may desire cultivars with short ripening periods, selection for rapid ripening will likely result in apples with poor texture attributes that make them undesirable to consumers. Apple cultivars with short ripening periods (i.e., late-flowering and early-harvested cultivars) would certainly be selected for by breeders if they had

desirable texture. It is possible that a few of the outlier cultivars we analyzed here that ripen relatively quickly but also possess desirable firmness attributes (e.g., “Tayeshnoe” and “Ivan”) may hold the key to decoupling ripening time from firmness. Genomic techniques that enable the relationship between ripening time and firmness to be broken could deliver novel apple cultivars in the future that ripen quickly, are harvested early, and are firm both at harvest and after storage.

Comparisons between domesticated crops and their wild ancestors can reveal which traits were targeted during domestication and improvement (Meyer, DuVal, and Jensen 2012). We found that the domesticated accessions (*M. domestica*) flower later, ripen slower, are harvested later, are heavier, less acidic, higher in soluble solids, and lower in phenolics than accessions belonging to the apple's primary wild progenitor, *M. sieversii*. These findings are consistent with studies of apple domestication showing that larger fruit, low acidity, and high SSC content appear to have been selected for during apple domestication and improvement (Duan et al. 2017; Ma et al. 2015; Miller and Gross 2011). Examples of fruit enlargement over the course of domestication can be seen in other crops as well, such as tomato, peach, and pear (Frary et al. 2000; Xiaolong Li et al. 2019; Y. Li et al. 2019). Our results suggest that a decrease in phenolics and later flowering and harvest times may have also been targets of selection during apple improvement.

Consistent with previous work (Migicovsky et al. 2016), we determined that New World cultivars are larger and have lower phenolic content than Old World cultivars. These

differences may be due to a difference in breeding targets between regions. It is well known that consumer preferences for fruit quality can be influenced by cultural factors like geography, as seen in peach, for instance (Y. Li et al. 2019), and this may explain the divergent apple phenotypes between different geographic regions observed here.

While most of the apples we evaluated are intended for fresh eating and are, thus, classified as “dessert apples”, about 10% of the ABC consists of “cider apples” that have attributes that make them desirable for fermenting into cider. We found that dessert apples flower earlier, are heavier, are less firm, and have lower phenolic content than cider apples. Higher phenolic content in cider apples compared to dessert apples is consistent with previous work (Sanoner et al. 1999; Valois, Merwin, and Padilla-Zakour 2006) and reflects the desire of cider makers to use apples with high astringency, a juice quality imparted by phenolics (Thompson-Witrick et al. 2014). Furthermore, genomic regions associated with flavonoid biosynthesis in dessert apples showed signatures of recent positive selection, suggesting that there was perhaps intentional breeding for reduced phenolic content in dessert apples (Migicovsky, Gardner, et al. 2021). Previous work is also consistent with our observation that dessert apples are generally larger than cider apples (Migicovsky et al. 2016).

We investigated whether the traits we measured changed over historical time and determined that, over the past 200 years, commercial apples have experienced a reduction in phenolic content. Apple cultivars released after 1940 have 30% lower phenolic content than those released prior to 1940. Phenolic compounds impart many health benefits and

apples are one of the major sources of phenolics in the human diet. It is estimated that apples provide 22% of the phenolics that North Americans consume, therefore, a reduction in these compounds during modern cultivar development may be a health concern (Vinson et al. 2001). Breeding apples with increased phenolic content is predicted to be relatively simple since phenolic content is highly heritable and has a simple genetic architecture (K. McClure et al. 2019). Similar to a previous study (Kschonsek et al. 2018), we hypothesize that the observed decline in phenolic content of newly released apple cultivars is due to active selection against astringent taste and enzymatic browning. Enzymatic browning is driven by high phenolic content (Amiot et al. 1992; Holderbaum et al. 2010; Murata, Noda, and Homma 1995) and is associated with undesirable color and flavor, and reduced browning has been a target for selection by apple breeders (Toivonen 2006). In fact, the only genetically modified apple on the market today has been engineered to be non-browning (Armstrong and Lane 2014). It is, thus, possible that the decrease in phenolics over the past 200 years is due to selection for reduced browning, a phenotype that we did not measure but that is correlated with phenolic content. However, breeding can be both the problem as well as the solution, and the present study provides the foundation for the development of novel apple cultivars that forfeit neither nutrition nor quality.

**Acknowledgements**

The authors gratefully acknowledge the Nova Scotia Fruit Growers' Association and the Farm Services team at AAFC-Kentville for establishing and maintaining the apple trees studied here. We also thank the following people for their help in establishing and/or

maintaining the collection: Charles Embree, Casi Sutherland, Emily Anderson, Dylan Troop, Scott Cann, Laura Porter-Muntz, Andrew Jamieson, Christiane DesLauriers, Thomas Gregoire, Benoit Girard, Leah Beveridge, Michael Jordan, Gavin Douglas, Samantha Lawrence, Jason Sawler, Melanie Eelman, Emma Gillis, Kerstin Surgenor, Conny Bishop, David Baldwin, Nancy Raymond, Cherie Collins, Tamara Dondi, and Leticia Reis.

# Chapter 3: Large-scale GWAS in apple reveals NAC18.1 as a master regulator of ripening traits

Watts, S., Migicovsky, Z., Myles, S. Large-scale GWAS in apple reveals NAC18.1 as a master regulator of ripening traits. 2023. Submitted to Fruit Research.

## Author Contributions

Sophie performed the data analysis and wrote and edited the manuscript.

## Abstract

Apple quality traits such as fruit texture, sugar content, and firmness retention during storage are key targets for breeders. Understanding the genetic control of fruit quality traits can enable the development of genetic markers, useful for marker-assisted breeding of new apple cultivars. We made use of over 260,000 single nucleotide polymorphisms (SNPs) genotyped across 1,054 apple accessions from Canada's Apple Biodiversity Collection to perform genome-wide association for 21 fruit quality and phenology traits. We identified two loci on chromosome 15 and 16 associated with phenolic content and a locus on chromosome 10 associated with softening. In addition, we determined that allelic variation at the *NAC18.1* transcription factor was associated with numerous traits including harvest date, firmness at harvest, and firmness after storage. Our analyses suggest that *NAC18.1* independently acts as a high level regulator of multiple ripening related traits and we propose a model for the allelic effects at *NAC18.1* on apple ripening and softening.

## Introduction

Apple is an important perennial fruit crop with a broad distribution worldwide: it ranks second in economic value among fruit crops with an estimated global production of $67 billion US in 2021 (FAOSTAT 2021). Developing new apple cultivars that are disease resistant and have improved fruit quality has been a major focus of apple breeding programs. Due to a prolonged juvenile stage, a highly heterozygous genome and self-incompatibility, apple breeding is a challenging and laborious process. Genomics assisted breeding offers a way to more efficiently develop improved apple cultivars. One such approach is the implementation of marker-assisted selection (MAS) in breeding programs. MAS can inform what progeny of a cross possess specific alleles that are associated with a trait so that only the germplasm likely to express the desired phenotype are moved forward in the breeding pipeline.

One of the most important traits to be targeted using genomics-assisted breeding in apple is fruit firmness (Laurens et al. 2018). Consumers value apples that are firm at harvest and soften minimally during storage. Softening is one of the hallmarks of apple ripening and is just one of many ripening related traits that indicates when a cultivar is ready to harvest. Apple ripening is characterized by a climacteric rise that leads to numerous biochemical changes such as softening, expression of aroma, change in color and carbohydrate metabolism (Johnston et al. 2009). The climacteric rise involves an increase in respiration and synthesis of the gaseous plant hormone ethylene (G.-S. Liu et al. 2022). During fruit maturation, ethylene production is inhibited by feedback until the climacteric

rise (Klee and Clark 2010). After the climacteric rise, autocatalytic ethylene production begins due to the introduction of 1-aminocyclopropane-1-carboxylic acid synthase (ACS). This initiates the onset of ripening and oxidation of ACC by ACC-oxidase (ACO), which leads to the production of ethylene (Klee and Clark 2010; Pech, Latché, and Bouzayen 2010). The desirable alleles of ACO1 and ACS1 have been developed into markers for MAS for firmness related phenotypes (Baumgartner et al. 2016). In addition to ACO1 and ACS1, polygalacturonase (PG1) has been identified as a key gene controlling apple firmness (Atkinson et al. 2012). PG1 encodes an enzyme that degrades pectin within cell walls and the middle lamella of apple fruit leading to loss of firmness (Atkinson et al. 2012). Although multiple markers have been developed for ripening and firmness related traits, their utility across diverse germplasm remains questionable (Migicovsky, Yeats, et al. 2021). Additional genetic mapping of key traits in apple can further advance the ability to efficiently select germplasm with promising fruit quality phenotypes.

Recent genome-wide association studies (GWAS) have identified additional candidate genes associated with ripening and firmness related traits in apple. Migicovsky et. al. (2016) first identified an association between a SNP in a NAM, ATAF1,2 and CUC2 (NAC) transcription factor, NAC18.1, and fruit firmness and harvest date. This SNP has been named the D5Y mutation since it involves a non-synonymous mutation from aspartate (D) to tyrosine (Y) at the fifth amino acid position of the NAC18.1 protein sequence. Since then, GWAS in other apple collections have uncovered associations between variants in and around NAC18.1 and fruit firmness and harvest date (Urrestarazu

et al. 2017; Larsen et al. 2019; Jung et al. 2020). NAC transcription factors (TF) are involved in a wide array of biological processes and form one of the largest TF families within plants (G.-S. Liu et al. 2022). Specifically, NAC TFs are associated with the regulation of ripening processes, such a sugar accumulation and softening, in numerous fruit species (Forlani, Mizzotti, and Masiero 2021) including tomato (Ying Gao et al. 2018), strawberry (Martín-Pizarro et al. 2021), banana (B. Li et al. 2020), peach (Pirona et al. 2013; Tan et al. 2021), kiwifruit (Nieuwenhuizen et al. 2015), sweet cherry (Qi et al. 2022) and apricot (García-Gómez et al. 2019). NAC18.1 has been shown to be ethylene independent and the D5Y marker within NAC18.1 has higher predictive power for firmness related traits compared to ACO1, ACS1 and PG1 markers (Migicovsky, Yeats, et al. 2021). Markers that can identify breeding lines with desirable firmness and ripening phenotypes can aid breeding programs that are focused on developing new apple cultivars with superior quality.

Here, we performed a large-scale GWAS in apple with a particular focus on ripening and firmness related traits using Canada's Apple Biodiversity Collection (ABC). The ABC is an orchard of over 1,000 apple accessions that was established in order to enable high resolution GWAS of fruit quality traits. The phenotypic diversity across the ABC was previously quantified by Watts et al. (2021) through the phenotyping of over 1,000 apple accessions for 39 fruit quality traits. The ABC was genotyped using genotype-by-sequencing, resulting in 278,231 genetic markers (Migicovsky, Douglas, and Myles 2022). By pairing the genotypic and phenotypic data collected from the ABC, we

identified genetic markers associated with commercially relevant fruit quality traits to advance genomics assisted breeding in apple.

## Results

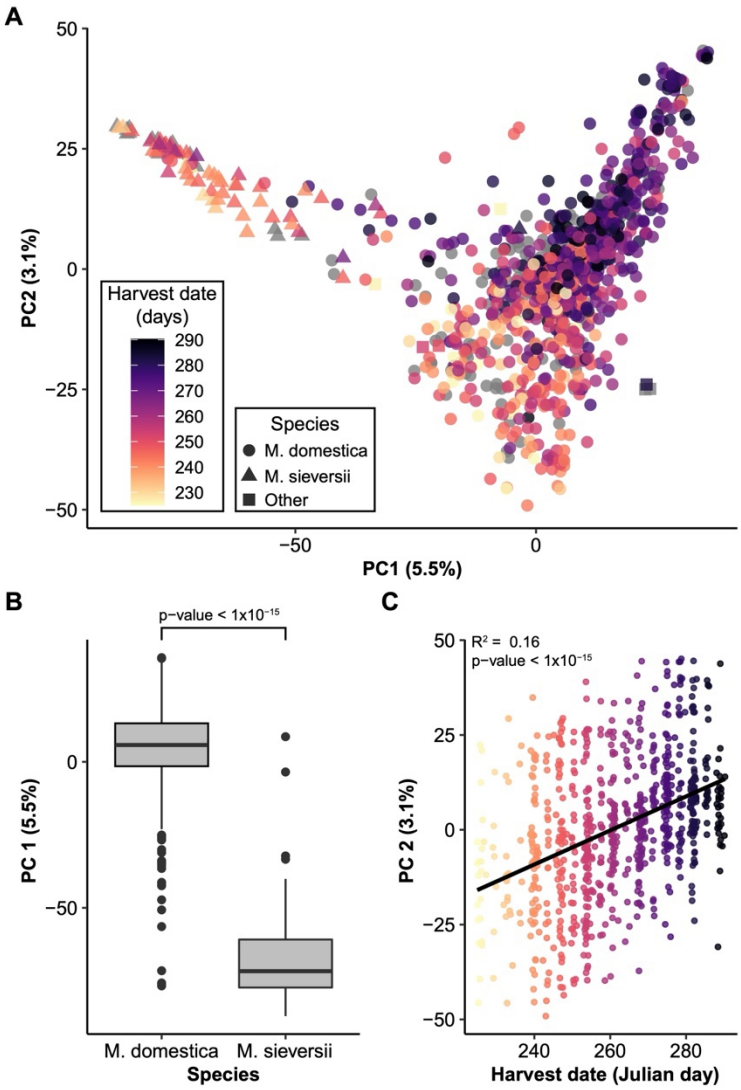


**Figure 3-1.** Population genomic analysis of 1054 diverse apple accessions. A) Principal components analysis of genome-wide SNP data, with accessions coloured by harvest date and with shape representing species. B) Boxplot of the PC 1 values between *M. sieversii* and *M. domestica*. C) Pearson correlation of PC 2 values with harvest date.

Principal components analysis (PCA) of the genome-wide SNP data revealed that PC1 explained 5.5% and PC2 explained 3.1% of the overall genetic variance (Figure 3-1A). We observed separation of *M. domestica* and *M. sieversii* along PC1: the two species had significantly different PC1 values (W = 69789, $P<1x10^{-15}$) (Figure 3-1B, Appendix II: Table II-I). We investigated the degree to which population structure correlated with phenotypes and determined that PC2 was strongly correlated with harvest date ($R^2$=0.16, $P<1x10^{-15}$) (Figure 3-1C, Appendix II: Table II-II).

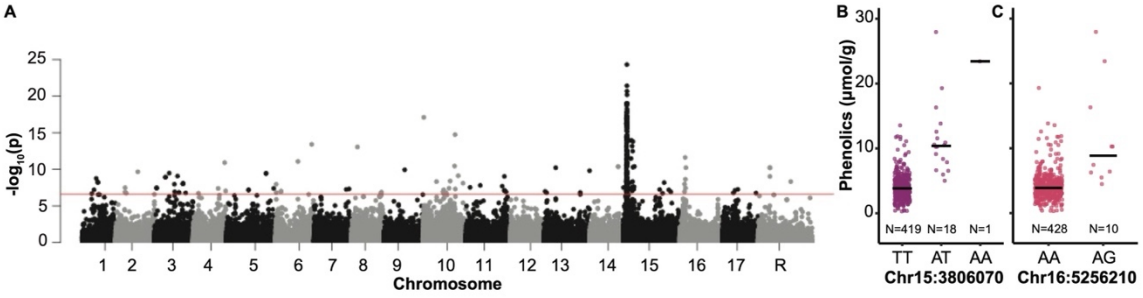


**Figure 3-2.** Genome-wide association study for phenolic content. A) Manhattan plot of GWAS results for phenolic content. The horizontal red line represents the significance threshold. (B) Boxplot of the distribution of phenolic content measurements across the genotypes of the top SNPs on chromosomes 15 and 16, with a black line indicating the median phenolic content measurement for each genotype.

GWAS were conducted for 21 fruit quality and phenology traits using a standard mixed-linear model (Appendix II: Figure II-I). The top SNP from the GWAS for phenolic content was Chr15:3806070 ($P=2.22x10^{-26}$) and it explained 27.3% of the phenotypic variation (Figure 3-2A, Appendix II: Table II-III). A second peak was observed on

chromosome 16, with the top SNP at Chr16:5256210 ($P=2.41x10^{-12}$) which explained 11.9% of the phenotypic variation (Figure 3-2A, Appendix II: Table II-III).

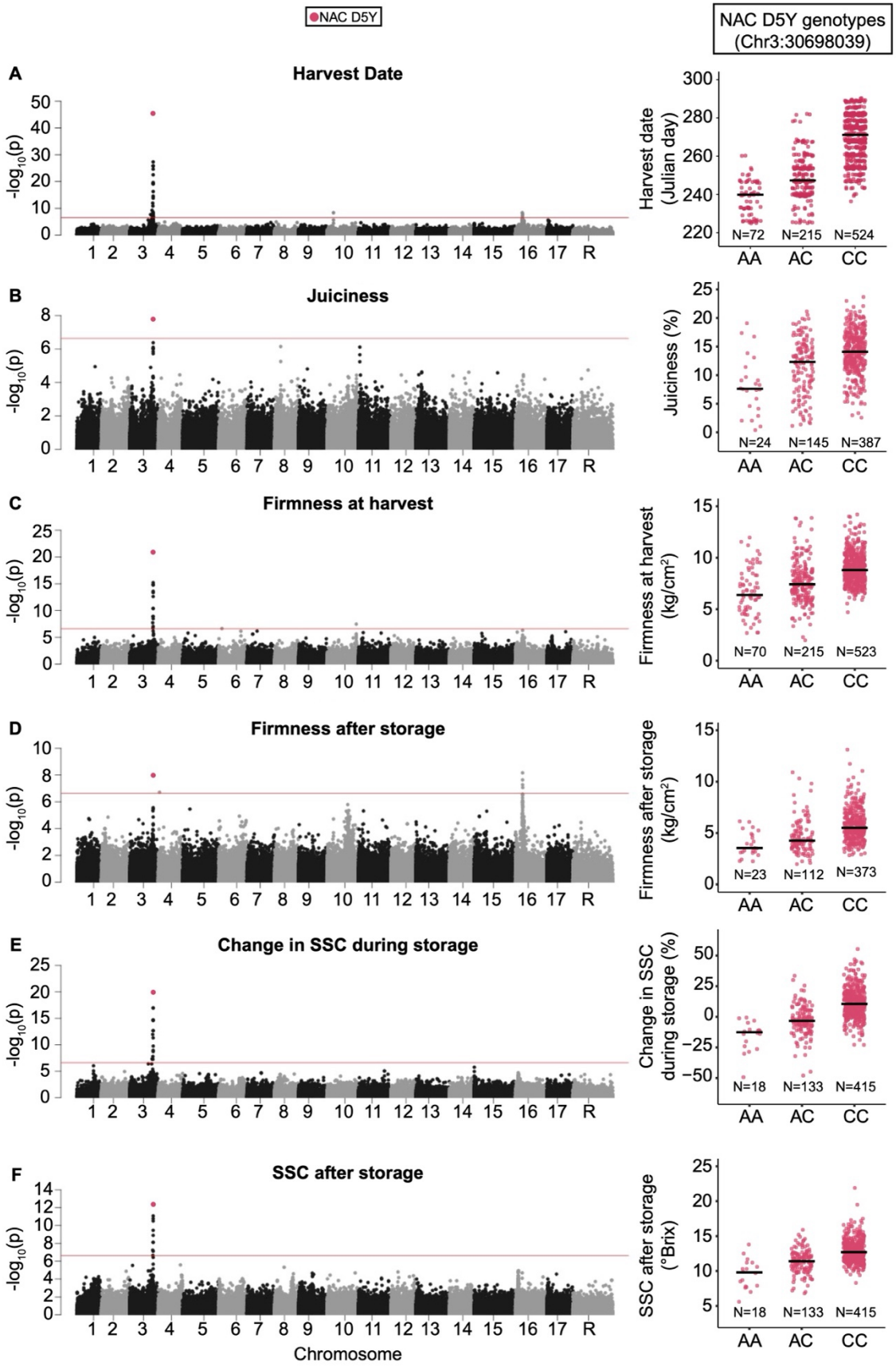
NAC D5Y
NAC D5Y genotypes (Chr3:30698039)
A
Harvest Date
-log10(p)
Harvest date (Julian day)
N=72 N=215 N=524
AA AC CC
B
Juiciness
Juiciness (%)
N=24 N=145 N=387
C
Firmness at harvest
Firmness at harvest (kg/cm2)
N=70 N=215 N=523
D
Firmness after storage
Firmness after storage (kg/cm2)
N=23 N=112 N=373
E
Change in SSC during storage
Change in SSC during storage (%)
N=18 N=133 N=415
F
SSC after storage
SSC after storage (°Brix)
N=18 N=133 N=415
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 R
Chromosome

**Figure 3-3.** Genome-wide association studies of ripening related traits. Manhattan plots of GWAS results for (A) harvest date, (B) juiciness, (C) firmness at harvest, (D) firmness after storage, (E) change in SSC during storage and (F) SSC after storage. The horizontal red lines in each Manhattan plot represent the significance threshold. The distributions of trait measurements across the *NAC18.1* D5Y genotypes are also shown adjacent to each Manhattan plot with median values for each genotype indicated by a horizontal line.

The SNP with the strongest association with harvest date was the D5Y SNP within the *NAC18.1* gene ($P=3.24x10^{-46}$; Figure 3-3A). There was an additional peak on chromosome 16, with the top SNP being Chr16:8852649 ($P=4.56x10^{-9}$). The D5Y SNP was also the top hit for juiciness, firmness at harvest, SSC during storage and SSC after storage (Figures 3-3B, C, E, F). For firmness after storage, the top GWAS hit was found on chromosome 16 (Chr:16:9235229; $P=6.85x10^{-9}$), while the second most significant association was with D5Y ($P=1.05x10^{-8}$; Figure 3-3D). We observed no significant associations for SSC at harvest (Appendix II: Figure II-I).

D5Y genotypes explained 18.1%, 7.8%, 11.9%, 6.1%, 13.3% and 10.3% of variation in harvest date, juiciness, firmness at harvest, firmness after storage, change in SSC during storage and SSC after storage, respectively (Appendix II: Table II-III). Harvest date, which was the trait most significantly associated with D5Y, was also correlated with the other 5 traits associated with D5Y (Appendix II: Figure II-II). We therefore re-ran the GWAS with harvest date as a covariate for all the other phenotypes with significant hits at D5Y (Appendix II: Figure II-III). After accounting for variation in harvest date, D5Y

remained the top SNP associated with firmness at harvest and change in SSC during storage, but was no longer significantly associated with juiciness, SSC after storage and firmness after storage (Appendix II: Figure II-III).

We did not find any noteworthy associations with the other traits measured except for malic acid. The well-characterised SNP within the MA1 gene (Bai et al. 2012) was the most significant SNP from the GWAS for acidity at harvest ($P=1.97x10^{-40}$) and acidity after storage ($P=3.97x10^{19}$) (Appendix II: Figure II-I).

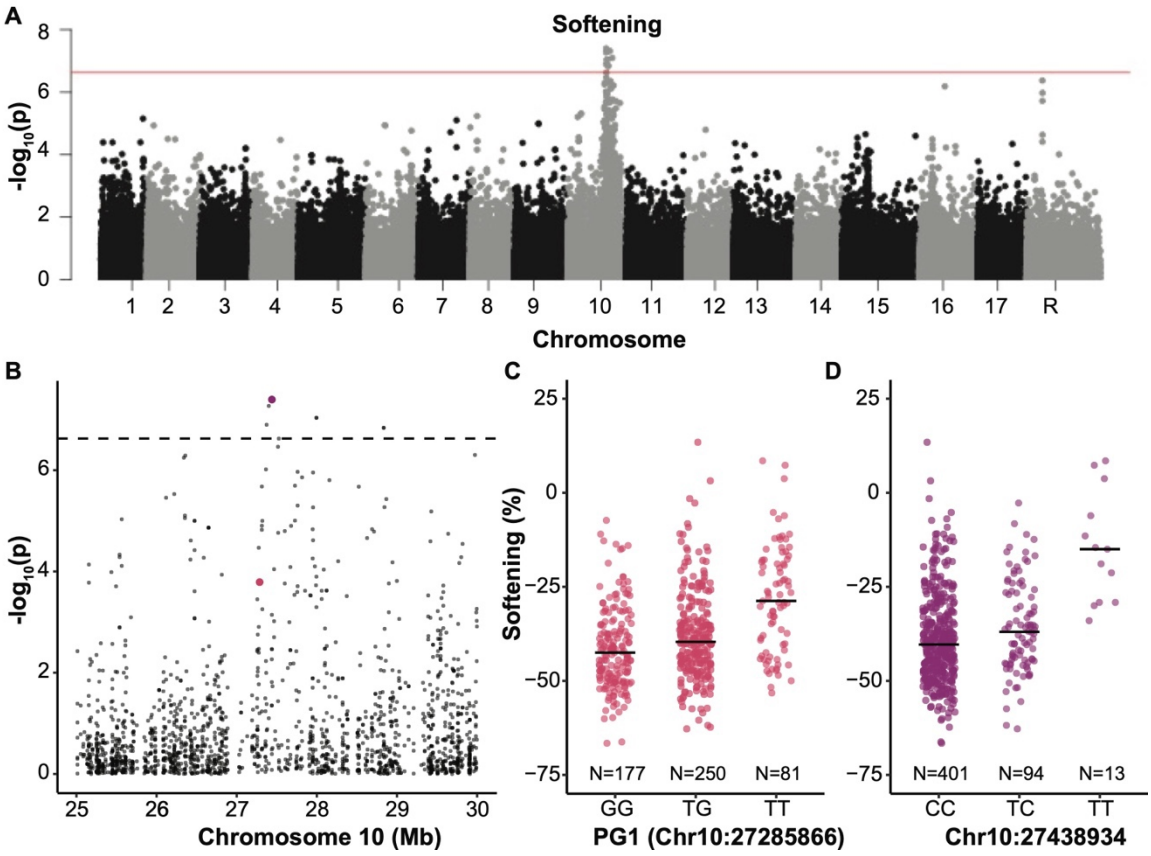


**Figure 3-4.** Genome-wide association study for softening during storage. (A) Manhattan plot from a GWAS for softening during storage. The horizontal red line represents the significance threshold. (B) A Manhattan plot showing only the genomic region around

the GWAS peak on chromosome 10. The purple SNP is the most significant SNP associated with softening. The *PG1* SNP is shown in pink. The horizontal dotted line represents the significance threshold. (C) Distributions of softening measurements across *PG1* genotypes. (D) Distributions of softening measurements across genotypes of the most significant SNP from the GWAS. Horizontal black lines indicate the median value within each genotype.

The top SNP from the GWAS for softening during storage was Chr10:27438934 ($P=4.02x10^{-8}$) and it explained 7.3% of the phenotypic variation (Figure 3-4A, Appendix II: Table II-III). The commonly used marker for apple softening, *PG1* (Chr10:27285866), was not significantly associated with softening ($P=1.62x10^{-4}$) and was 153,068 bp downstream of the top GWAS hit observed here (Figure 3-4B).

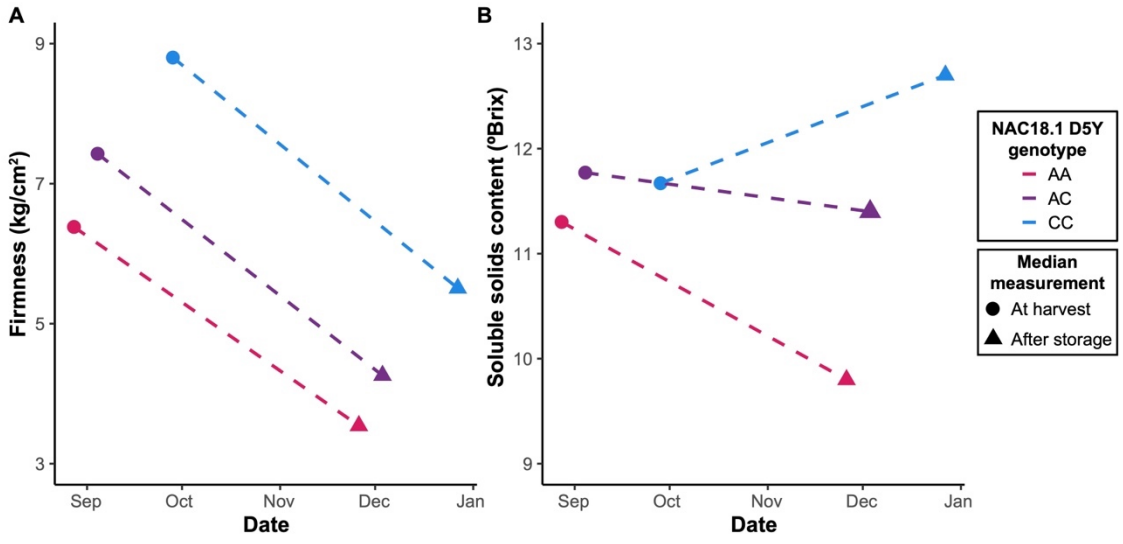


**Figure 3-5.** A summary of the allelic effects of *NAC18.1* D5Y genotypes on ripening phenotypes, both at harvest and after 3 months of cold storage. (A) Median measures of firmness at harvest and firmness after storage for apple accessions with different D5Y

genotypes. (B) The median soluble solids content at harvest and after storage for apple accessions with different D5Y genotypes.

The median SSC and firmness measurements, at harvest and after storage, were calculated across the three D5Y genotypes. D5Y genotypes had no effect on softening during storage: all three genotypic classes experienced similar degrees of softening during storage. However, the at harvest and after storage firmness measurements did vary across the genotypic classes at D5Y. Accessions homozygous for the C allele at D5Y had the highest median firmness measurements both at harvest (8.8 kg/cm$^2$) and after storage (5.5 kg/cm$^2$) (Figure 3-5A, Appendix II: Table II-IV). The D5Y heterozygotes were slightly more firm at harvest (7.4 kg/cm$^2$) and after storage (4.2 kg/cm$^2$) compared to the AA homozygotes, which were the accessions with the lowest firmness measurements at harvest (6.4 kg/cm$^2$) and after storage (3.5 kg/cm$^2$) (Figure 3-5A, Appendix II: Table II-IV).

Accessions that were homozygous for the A allele or heterozygous for D5Y declined in SSC during storage (Figure 3-5B). SSC decline was less severe, however, for D5Y heterozygotes compared to AA homozygotes, with median losses of 12.6% and 3.3% SSC, respectively (Appendix II: Table II-IV). In contrast, accessions homozygous for the C allele at D5Y gained SSC during storage, with a median increase of 10.5% SSC (Appendix II: Table II-IV).

## Discussion

### *Genetic structure of the ABC*

The genetic structure of the ABC is strongly influenced by the inclusion of the apple's primary wild progenitor species, *M. sieversii* (Figure 3-1). A recent study of the same collection studied here found that several phenotypes differ between the two species, and that domestication and breeding likely resulted in apples today that are larger, less acidic and less phenolic than their wild progenitors (Davies et al. 2022). The use of wild progenitors in crop breeding is important because it can expand the phenotypic diversity available for cultivar improvement (Migicovsky and Myles 2017). Wild relatives of domesticated crops may possess improved fruit quality and disease resistance traits, so the GWAS population studied here intentionally includes dozens of *M. sieversii* accessions to capture genetic and phenotypic diversity that may be unavailable within *M. domestica*. Figure 3-1 also shows that there is a correlation between population structure and harvest date, which has been previously shown in a similar apple population (Migicovsky et al. 2016). The correlation between population structure and harvest date likely reflects the fact that harvest date is a proxy for geographic ancestry, which is frequently a major determinant of population genetic structure (Novembre et al. 2008). Overall, our population genomic analysis suggests that the genetic structure in this apple population is primarily shaped by the inclusion of two different species and diverse cultivated germplasm that ripens over a considerable 65 day window.

*Phenolic content is associated with loci on chromosome 15 and 16*

Future apple improvement may benefit from an augmentation of bioactive compounds that impart health benefits. Phenolic compounds are a large family of secondary metabolites commonly found in plants that possess antioxidant and antimicrobial properties (Aires 2017). Consumption of phenolic compounds from apples has been shown to impart health benefits due to their antidiabetic properties (C. H. J. Yu et al. 2023). From our GWAS, we identified two significant peaks associated with phenolic content on chromosomes 15 and 16 (Figure 3-2). We did not identify any noteworthy candidate genes within 50 kb of the top hit on chromosome 15. However, within the 50 kb surrounding the top SNP on chromosome 16 is a *bHLH* protein gene and *TCP* family transcription factor gene which have been implicated in plant flavonoid biosynthesis (Appendix II: Table II-V;(Shutian Li and Zachgo 2013). Previous GWAS and linkage mapping studies have identified a quantitative trait locus (QTL) on chromosome 16 for polyphenols within which the leucoanthocyanidin reductase (*LAR1*) gene is located (K. McClure et al. 2019; Chagné et al. 2012; S. A. Khan et al. 2012). *LAR1,* which converts leucocyanidin to catechin, has been proposed as a candidate gene involved in phenolic content variation in apple. Despite the presence of markers in close proximity to the LAR1 gene within our GWAS SNP set, the top SNP on chromosome 16 from our GWAS was 1.8 Mb from *LAR1*.

We suspect that the inability to identify signals near previously reported QTLs for phenolics may be due to different methods for measuring phenolics that makes it

challenging to compare results across studies. Here, total phenolic content was measured using the Folin–Ciocalteu assay (FCA) which uses a reduction reaction to estimate the total amount of phenolics and phenolic acids within a sample (Singleton, Orthofer, and Lamuela-Raventós 1999). In contrast to chromatography methods, the FCA lacks specificity to distinguish among classes of phenolics and instead groups phenolic compounds into one total measurement (Samara, Nasser, and Mingelgrin 2022). The FCA can also reduce other compounds, such as sugars and ascorbic acid, thereby contributing to an inaccurate measurement of total phenolic content (Aires 2017). Despite the differences in methods used to measure phenolics, we identify signals on chromosome 15 and 16 associated with total phenolic content. Further work should include measurements of individual phenolic compounds using chromatography methods to discern the specific compounds driving the associations between phenolic content and the loci identified here.

*NAC18.1 is major locus associated with multiple ripening traits*

Apple ripening involves the coordination of various biochemical changes such as softening, aroma development, color change and carbohydrate metabolism that render the fruit ready for harvest (Johnston et al. 2009). Previous apple GWAS, across multiple germplasm collections, have identified strong signals near and within *NAC18.1* for harvest date and firmness (Migicovsky et al. 2016; Urrestarazu et al. 2017; Larsen et al. 2019; Jung et al. 2020). We observed a strong signal at the D5Y SNP within *NAC18.1* for harvest date, juiciness, firmness at harvest, firmness after storage, SSC change during

storage and SSC after storage. These observations suggest that the allelic effects at the *NAC18.1* locus are highly pleiotropic.

Harvest date, the trait most strongly associated with variation at *NAC18.1*, was correlated with the other phenotypes that show a GWAS signal at *NAC18.1*, such as juiciness, SSC after storage, SSC change during storage, firmness at harvest and firmness after storage (Appendix II: Figure II-II). We reasoned that the signal observed at *NAC18.1* across other traits may be driven entirely by the effect of *NAC18.1* on harvest date. After controlling for harvest date, however, the D5Y SNP in *NAC18.1* was still the most significantly associated SNP with firmness at harvest and percent change in SSC during storage (Appendix II: Figure II-III). This suggests that variation at *NAC18.1* independently affects multiple ripening related traits.

*Allelic effects of NAC18.1 on ripening*

A hallmark of apple ripening is the conversion of starch to sugar that signals fruit is ready to be picked (Brookfield et al. 1997). Within apple fruit, carbohydrates are converted to starch and stored within amyloplasts of developing fruit (Noronha et al. 2018). During the later stages of fruit maturation and at the onset of ripening, starch is degraded leading to an accumulation of sugar (Brookfield et al. 1997). Previous studies have uncovered significant associations for individuals sugars, sucrose and fructose, with a locus on chromosome 1 (Larsen et al. 2019; Guan et al. 2015). While other studies examining SSC, rather than individual sugars, have found associations on chromosomes 8 and 3

(Liebhard et al. 2003; Amyotte et al. 2017; Jung et al. 2022). We did not identify any significant SNP associations with SSC, similar to other GWAS in apple (Kumar et al. 2014). Our inability to detect significant associations for sugar content compared to other studies is likely due to low marker density surrounding key loci and differences in measurement of sugar content (individual sugars vs. SSC). Future investigation of the genetic control of sugar content in apple will likely benefit from improved genomic data and more precise measurement of individual sugars.

The degradation of starch continues once the fruit is picked and provides a source of carbohydrate for respiration and synthesis of aroma compounds and leads to an increase in sugar. Although we did not identify any significant SNPs associated with SSC at harvest, we did identify significant associations between D5Y and change in SSC during storage and SSC after storage. We observed a mostly co-dominant pattern for the effects of D5Y genotypes on SSC change during and after storage. During storage, accessions homozygous AA at D5Y lost the largest amount of soluble solids, heterozygotes lost slightly less, while accessions homozygous CC at D5Y gained soluble solids. We hypothesise that, although apples were harvested at similar SSC levels, accessions with an A allele at D5Y continue on an accelerated ripening path post-harvest. Once starch is converted to sugar, accessions on the accelerated ripening path metabolise available sugars in order to further fuel metabolic processes, leading to an observable decline in SSC during 3 months of cold storage. In contrast, we hypothesise that accessions that are homozygous CC at D5Y convert starch to sugars at a slower rate: during 3 months of storage, SSC gradually increases in these accessions as conversion of starch to sugar is

still underway. Additionally, the demand for sugar substrates could be lower in the homozygous CC accessions compared to the homozygous AA accessions due to a lower demand of substrates for respiration, as has been suggested to occur in slow ripening peach accessions (Giné-Bordonaba et al. 2020). An evaluation of our proposed model can be performed in the future by measuring starch content and SSC through a pre-harvest and post-harvest time series.

In addition to changes in SSC, changes in texture are a key aspect of fruit ripening. We identified significant associations between genetic variation at *NAC18.1* and firmness at harvest and firmness after storage (Figures 3-3C,D). The early harvested AA genotypes at D5 were less firm than the late harvested CC genotypes. The association with firmness after storage is likely driven primarily by the correlation with firmness at harvest: a firm apple at harvest will be firmer after three months of storage than a soft apple at harvest. No significant association was identified between D5Y and softening, indicating that ethylene-mediated post-harvest softening likely occurs independently of genetic variation at the *NAC18.1* locus (Figure 3-4). This is consistent with previous observations that NAC18.1 transcript levels were affected neither by exposure to ethylene, nor to an ethylene inhibitor, 1-MCP (Migicovsky, Yeats, et al. 2021). Thus, our observations suggest that alleles at *NAC18.1* mediate apple firmness via an ethylene independent on-tree ripening program rather than a post-harvest, ethylene-dependent softening mechanism.

In addition to the critical role of NAC18.1 variants in determining firmness, we found that genetic variation on chromosome 10 mediated the degree to which apples softened during storage (Figure 3-4). Previous studies have also identified significant associations between firmness and markers on chromosome 10, most of which suggest that the signal arises from variants at the polygalacturonase (*PG1*) gene, whose expression is ethylene-dependent (Kumar et al. 2013; Di Guardo et al. 2017). A functional marker for *PG1* has been developed and employed in breeding to predict softening, and this *PG1* marker was genotyped and included in our GWAS (Costa et al. 2010; Baumgartner et al. 2016; Migicovsky, Douglas, and Myles 2022). We did not find a significant association between softening and the *PG1* marker. Nor did we find significant associations with markers in the ACO1 and ACS1 genes used to predict firmness, consistent with recent work questioning the utility of these markers across diverse germplasm (Migicovsky, Yeats, et al. 2021). Our top SNP for softening was located 153 kb downstream of PG1 Results from a recent pooled-sequencing GWAS identified a signal upstream of PG1 for softening (Davies et al. 2023). Overall, these results suggest that functional genetic variants affecting softening on chromosome 10 may lie in long-range regulatory elements affecting the PG1 expression as hypothesised by Di Guardo et al. (2017).

Although *ACO1*, *ACS1*, and *PG1* have been extensively validated as key genes involved in controlling aspects of apple ripening, our observations suggest that NAC18.1 may be a high-level regulator orchestrating the overall tempo of ripening. Lü et al. (2018) proposed a model for peach ripening where a NAC TF impacts a cascade of ripening traits by binding to the promoters of important ripening genes, such as *ACO1* and *ACS1* (Lü et al.

2018). In a slow ripening peach cultivar, it has been suggested that the ripening cascade is disrupted by loss-of-function deletions in two NAC TFs (Giné-Bordonaba et al. 2020). Given these observations in peach, we hypothesise that genetic variation at *NAC18.1* in apple affects transcriptional regulation of its downstream targets and thereby modulates both firmness related traits and SSC. We suggest that *NAC18.1* functions as a ripening "throttle", and alleles of *NAC18.1* differentially control the timing of ripening and rate at which an apple moves through the biochemical changes associated with ripening. Accessions that are homozygous AA at the *NAC18.1* D5Y SNP are "rapid ripeners" as they move through the stages of ripening at a faster pace than their homozygous CC counterparts, the "slow ripeners". Heterozygous accessions fall between these two extremes, suggesting a predominantly co-dominant model of inheritance at this locus (Figure 3-5A). The rapid ripeners tend to be harvested earlier in the season, are softer at harvest and post-storage, lose SSC during storage and have lower levels of SSC post-storage compared to the slow ripeners. Here we demonstrated that genetic variation at *NAC18.1* is correlated with variation in multiple ripening related traits, however the mechanism by which this variation affects these traits remains unknown. Although the D5Y may be considered a putatively causal *NAC18.1* variant, recent results suggest that it is more likely mutations in the promoter region of *NAC18.1* that mediate ripening (Davies et al. 2023). This suggests that our GWAS signal at *NAC18.1* arises because of cis-regulatory variation affecting NAC expression, which subsequently impacts the expression of downstream target genes that mediate the ripening process. Future work will include an exhaustive search for genetic elements impacting the expression of *NAC18.1* using whole-genome sequencing of this diverse collection of apples. In

addition, the downstream genomic target(s) of NAC18.1 must be identified in order to fully elucidate *NAC18.1*'s role in the apple ripening process.

Although NAC18.1 clearly stood out in the present work as a large effect locus for ripening traits, it is likely that other loci also contribute to variation in ripening traits. As demonstrated in other GWAS studies (Zhao et al. 2011), it is possible we could be missing associations with loci of small effect due to the way in which mixed model association analyses correct for population structure and relatedness. This is particularly applicable here since harvest date is strongly associated with population structure (Figure 3-1A,C). In addition to the signal at DY5, we uncovered a significant, albeit weaker, signal on chromosome 16 for harvest date and firmness after storage (Figure 3-3A,D). The top SNP for harvest date was at Chr16:8852649 ($P=4.56x10^{-9}$) and the top SNP for firmness after storage was 383 kb away at Chr16:9235229 ($P=4.56x10^{-9}$). Previous GWAS studies have identified significant associations for harvest date and fruit firmness on chromosome 16 that co-locate to the same region of chromosome 16 that we identified here (Jung et al. 2022; Urrestarazu et al. 2017; Chagné et al. 2014). To date, the involvement of the chromosome 16 locus on harvest date and firmness traits remains undefined and should be further examined along with other small effect loci that could be impacting in ripening traits.

Identifying both large and small effect loci impacting key traits is especially important for breeding purposes. Although *NAC18.1* appears to be a large effect locus impacting apple ripening, there are two important considerations for targeting this locus with MAS

for breeding improved cultivars. The first concerns the allele frequencies at D5Y SNP across existing elite apple germplasm. A study of the apples used here showed that 66% of accessions, including the majority of elite commercial cultivars, are homozygous for the late-ripening C allele at D5Y, and this desirable allele may have been selected for by generations of apple breeders (Migicovsky, Yeats, et al. 2021). Therefore, the genomics-assisted improvement of firmness across elite germplasm will likely require the identification of additional loci that are not already fixed within elite germplasm. The second consideration is the pleiotropic nature of the NAC18.1 locus. For example, harvest date is strongly correlated with firmness at harvest. Therefore, the development of an earlier ripening cultivar by selecting for the early A allele at D5Y will likely lead to unfavourable changes such as softer fruit at harvest. The fixation of the favourable alleles at *NAC18.1* across elite germplasm and the pleiotropic effects of the *NAC18.1* locus further emboldens the search for additional loci involved in controlling ripening traits.

Here, we confirmed previous associations between genetic loci and ripening related traits, and identified new variants associated with polyphenol content and softening. Uncovering the causal genetic variant(s) controlling variation across key fruit quality and phenology traits in apple requires improved genetic mapping resolution. Future work should incorporate whole-genome sequence data of large samples in order to increase the resolution necessary to move beyond the development of genetic markers for MAS and towards the identification of causal alleles. Given the importance of health compounds, fruit texture and ripening traits both for production management and fruit quality, there is

value in continuing to pursue a comprehensive characterization of the genetic variation underlying commercially relevant traits in apple.

## Materials and Methods

### *Phenotype data*

The phenotype data were collected from 1,1119 accessions in the ABC and are described in Watts et al. (2021). In short, 39 traits were measured from 2014-2017. These traits included phenology traits such as flowering time, harvest date, time to ripen and precocity. The fruit quality traits measured at harvest included titratable acidity, soluble solids content (SSC) (degrees Brix), SSC/acidity ratio, firmness, weight, total phenolic content and juiciness. In addition, acidity, SSC, SSC/acidity ratio, firmness and weight were also measured after 3 months of cold storage and the percent change in these phenotypes during storage were calculated. Apple accessions from the ABC were also classified based on their reported species (*M. domestica* or *M. sieversii*), release year, end use (cider or dessert) and country of origin.

### *Genetic data*

The ABC was genotyped using genotyping-by-sequencing (GBS; (Elshire et al. 2011) and the methods and resulting genetic data were recently published (Migicovsky, Douglas, and Myles 2022). In short, DNA from young leaf tissue was extracted from

accessions in the ABC and then GBS libraries were prepared using two restriction enzymes and libraries were then sequenced using Illumina Hi-Seq 2000. The final SNP set after imputation and pooling across SNP callers was 278,231 SNPs from 1,175 unique accessions (Migicovsky, Douglas, and Myles 2022). An additional 8 markers were genotyped using high resolution melting (HRM) and were added to the SNP set using the --merge function in PLINK (Purcell et al. 2007). The HRM markers included Ma1 for acidity (Bai et al. 2012) and three markers for scab resistance, Rvi2, Rvi6, Rv15 (Jänsch et al. 2015). In addition, four texture markers were genotyped with HRM, these included *PG1*, *ACO1*, *ACS1* and *NAC18.1* as described previously (Migicovsky, Yeats, et al. 2021). The HRM genotype call for *NAC18.1* was retained, while the GBS SNP for *NAC18.1* was removed using PLINK. The final SNP set from Migicovsky et al. (Migicovsky, Douglas, and Myles 2022) included 278,231 SNPs from 1,1175 accessions and the data can be found Dryad: https://datadryad.org/stash/dataset/doi:10.5061/dryad.zkh1893cd.

The final SNP set was filtered to only include the 1,054 accessions which had data for at least one phenotype (Watts et al. 2021). SNPs with heterozygosity >90% and a minor allele frequency <0.01 were removed, resulting in a final SNP set of 260,399 SNPs across 1,054 apple accessions. All filtering steps were performed using PLINK (Purcell et al., 2007).

*Principal components analysis*

For principal components analysis (PCA), only SNPs anchored to an assembled chromosome were considered and LD pruning was then performed using PLINK (-indep-pairwise 10 3 0.5) (Purcell et al. 2007), resulting in 164,992 SNPs that were used to perform PCA in TASSEL (version 5.0). The correlation of principal component (PC) 1 values with harvest date was calculated in R using 'cor.test' and p-values were multiple test corrected with a Bonferroni correction. Wilcoxon test statistics of the comparison of PC2 values across *M. domestica* and *M. sieversii* were calculated in R using the 'wilcox.test' function.

*Genome-wide association studies*

GWAS was performed on trait data from 2017 due to the larger sample sizes, unless the trait data was only collected in 2016, which resulted in analysis of 21 of the 39 traits from Watts et al. 2021 (See supplementary file). For each phenotype's filtered genotype file, kinship matrices and PCA were run using TASSEL (version 5.0). GWAS was then run using the 'mlmm' R package (Segura et al. 2012). We present GWAS results from the single-locus model where no SNPs are included as cofactors (equivalent to a mixed-linear model GWAS). Kinship matrices and the first 5 PCs were used to account for relatedness. Quantile–quantile plots were created using the 'qqman' package in R (D. Turner 2018). Manhattan plots for the 21 phenotypes were plotted using the `mlmm` package and are available in the supplementary material. However, the Manhattan plots presented in the main manuscript (Figure 3-2,3-3, 3-4) were plotted using the 'qqman' package in R. All unassembled contigs were concatenated into a single chromosome labeled "R"

throughout the manuscript. The effective number of independent tests (*M*eff) was calculated to be 211,156 based on the 260,399 SNP set using the simpleM package in R. The *M*eff value was then used to calculate the threshold for significance for all GWAS using -log($\alpha$/*M*eff), where $\alpha = 0.05$.

Jitter plots of the distribution of trait measurements across the genotype classes of top GWAS SNP hits were visualized using the `ggplot` package in R. Gene annotations from the GDDH genome were retrieved for regions within 50 kilobases of either side of the top GWAS hits. The distribution of trait measurements across the genotypes of the SNPs of interest from the GWAS were visualized using the 'ggplot'. Pearson correlation tests between phenotypes were run and visualized using the 'ggpairs' function from the GGally R package. The variation explained by the top SNPs from the GWAS was calculated by subtracting the R2 from a linear model with only the top 5 PCs included as explanatory variables from the R2 from a linear model including the top 5 PCs and the top GWAS SNP as explanatory variables. The difference in R2 between the two models was reported as the variance of a trait explained by a particular SNP.

**Acknowledgements**

SW was supported by funding from a Vanier Scholarship from the National Sciences and Engineering Research Council of Canada. ZM was supported by funding from the Canada Research Chairs program. The authors gratefully acknowledge the Nova Scotia Fruit Growers' Association and the Farm Services team at AAFC-Kentville for establishing and maintaining the apple trees studied here.

# Chapter 4: Cannabis labelling is associated with genetic variation in terpene synthase genes

Watts, S., McElroy, M., Migicovsky, Z., Maassen, H., van Velzen, R., Myles, S. (2021). Cannabis labelling is associated with genetic variation in terpene synthase genes. Nature Plants, 10, 7. https://doi.org/10.1038/s41477-021-01003-y

## Author Contributions

Sophie Watts conducted the data analysis and wrote and edited the manuscript.

## Abstract

Analysis of over 100 *Cannabis* samples quantified for terpene and cannabinoid content and genotyped for over 100,000 single nucleotide polymorphisms indicated that Sativa- and Indica-labelled samples were genetically indistinct on a genome-wide scale. Instead, we found that *Cannabis* labelling was associated with variation in a small number of terpenes whose concentrations are controlled by genetic variation at tandem arrays of terpene synthase genes.

## Introduction

*Cannabis* has been consumed for its psychoactive properties for over 2,500 years, and its estimated global market value is US$340 billion (Lawler 2019; Naville 2019; Bonini et al. 2018). Because it is a widely used drug that is increasingly being legalized for

medicinal and recreational use, it is critical that *Cannabis*'s genetic and chemical variation be accurately quantified and communicated. The vernacular labels Sativa and Indica (not to be confused with the taxonomic names *C. sativa sativa* L. and *C. sativa indica* Lam.) are routinely assigned to *Cannabis* cultivars by breeders, retailers and users to describe a cultivar's morphology, aromas and/or psychoactive effects (Guy and McPartland 2017). However, it is unclear whether these labels capture meaningful information about *Cannabis* genetic and chemical variation.

*Cannabis* genomics research has thus far largely focused on the characterization of genes underlying the production of the cannabinoids cannabidiol (CBD) and tetrahydrocannabinol (THC) (Laverty et al. 2019; McKernan et al. 2020; Vergara et al. 2019; Grassa et al. 2021). However, *Cannabis* produces hundreds of aromatic terpenes that drive consumer preference and are frequently associated with Sativa and Indica labels (Guy and McPartland 2017; Gilbert and DiVerdi 2018). In addition, there is evidence to suggest that a cultivar's terpene profile affects its psychoactive properties (Russo 2011; Koltai and Namdar 2020). To date, various terpene synthase genes have been identified in *Cannabis*; however, the genetic control of terpene variation across *Cannabis* cultivars remains largely unexplored (Booth, Page, and Bohlmann 2017; Zager et al. 2019; Günnewich et al. 2007; Livingston et al. 2020).

Here we re-analysed 297 samples of drug-type *Cannabis* that were previously quantified for 40 terpenes and cannabinoids using gas chromatography–mass spectrometry (GC–MS) (Hazekamp, Tekalova, and Papadimitriou 2016) (Appendix III: Table III-I Figure

IV-I), and we paired these data with 116,296 newly generated single nucleotide polymorphisms (SNPs) from 137 of these samples from which sufficient high-quality DNA could be extracted. We determined the degree to which the genomic and GC–MS data corresponded to a five-point labelling scale ranging from 1 (100% Sativa) to 5 (100% Indica) as reported by sample sources.

## Materials and Methods

### *Samples*

The samples come from a previous study of 460 *Cannabis* chemotypes (Hazekamp, Tekalova, and Papadimitriou 2016). The samples were collected from Bedrocan International BV ($n = 37$), HempFlax ($n = 205$) and Dutch 'coffee shops' either directly or indirectly through the TRIMBOS Institute ($n = 55$). Samples labelled as 'Hemp' were excluded from the analysis. We retained and analysed 297 samples that were classified along a five-point scale according to ancestries reported by the sources: 'Sativa' (100% Sativa), 'Hybrid-Sativa' (75% Sativa, 25% Indica), 'Hybrid' (50% Sativa, 50% Indica), 'Hybrid-Indica' (25% Sativa, 75% Indica) and 'Indica' (100% Indica). These five groups were encoded as 1 (100% Sativa) to 5 (100% Indica) for the statistical analyses described below.

### *Gas chromatography*

A total of 297 samples were previously quantified for terpene and cannabinoid content, and we conduct a re-analysis of these data here. The chemical analyses of the samples are described in detail in ref. Hazekamp, Tekalova, Papadimitriou (2016). Briefly, for each sample, 500 mg of ground homogenized dried flower material was mixed with 40 ml of ethanol, agitated for 10 minutes and centrifuged. The supernatant was collected, and the process was repeated twice more on the pellet. An internal standard consisting of 200 µl of 1% solution of 1-octanol was added to the combined supernatant, the volume was adjusted to 100 ml with ethanol and the combined sample was centrifuged again. The combined sample was analysed using an Agilent GC 6890 series (Agilent Technologies) equipped with a 7683 autosampler and a flame ionizing detector. The instrument was equipped with a DB-5 column (length, 30 m; internal diameter, 0.25 mm; film thickness, 0.25 µm; J&W Scientific). Peaks from the sample chromatograms were manually integrated, and the peak area was recorded with correction for the internal standard peak area. Peak identification was conducted by analysing selected samples using gas chromatography mass spectrometry (GC-MS) and then comparing compounds' mass spectra and retention times with authentic standards and literature reports as described in ref. Hazekamp, Tekalova, Papadimitriou (2016). Compounds without authentic standards are marked with an asterisk in the figures to indicate that they were tentative identifications. Peak areas of monoterpenes, sesquiterpenes and cannabinoids were quantified (in mg per g of plant material) using calibrated standards of β-pinene, α-humulene and CBD, respectively. We re-assessed the compound identifications in Hazekamp, Tekalova, Papadimitriou (2016) and in certain cases we renamed compounds on the basis of the inability to distinguish stereoisomers using a DB-5 column. For

example, in the case of the compound listed by Hazekamp, Tekalova, Papadimitriou (2016) as '(−)-linalool', we renamed this to 'linalool'. There are also two compounds that could not be reliably identified; they are listed as 'unidentified compounds' (Appendix III: Table III-II). THC, δ-8-THC and CBN were combined into a single value, 'Total THC', because δ-8-THC and CBN are degradation products of THC. Peaks of *R*-limonene and β-phellandrene were indistinguishable and were therefore combined into a single value and reported as 'limonene'. Thymoquinone, geraniol, thymol and carvacrol were removed because they were not present in any samples, and cineol was removed because it was present in only one sample. Pearson correlations were calculated between each pair of chemicals using the cor.test function in R v.3.5.1 (R Core Team 2018). According to previous work (Ernest Small, H. D. Beckstead, and Chan 1975), the samples analysed here were nearly all drug-type *Cannabis* (that is, type I) (Appendix III: Figure III-I), except nine samples with THC > 0.3% and CBD > 0.5% (that is, type II).

*Genomic analysis*

Whole-genome DNA was extracted using a NucleoSpin 96 Plant II kit (Machery-Nagel) and quantified using the QuantiFluor dsDNA System and the GloMax-Multi + Microplate Multimode Reader with Instinct (Promega). Genotyping-by-sequencing libraries were prepared using the restriction enzyme ApeKI (Elshire et al. 2011), and the libraries were sequenced on two lanes of an Illumina Hi-Seq 4000 (Illumina). The DNA sequence data are available as NCBI BioProject PRJNA713792. Calling of SNPs was performed in TASSEL (v.5.0)(Bradbury et al. 2007) by aligning to the CBDRx reference genome

(Grassa et al. 2021). SNP calling was performed before the implementation of the new chromosome numbering of the CBDRx genome in April 2020. Chromosomes were recoded for analyses to reflect the new chromosome numbering system. We used VCFtools (v.0.1.15) (Danecek et al. 2011) to retain only bi-allelic SNPs and samples with <70% missing data, which resulted in 155 remaining samples and 284,988 SNPs. Genotype imputation was performed using LinkImputeR (Money et al. 2017) with a minor allele frequency threshold of 0.01, a minimum read depth for masking of 20 and the number of masked genotypes set to 5,000. We chose to impute with a minimum read count of 2 and a maximum missingness threshold of 70%, which resulted in an imputation accuracy of 92.88%. After imputation, 149 samples remained. An additional 12 samples were removed because they had no phenotype data. This resulted in a final set of 137 samples with both genetic and chemical data. The SNP data were filtered using PLINK (v.1.90) (Purcell et al. 2007) to exclude SNPs with a minor allele frequency less than 0.05 and SNPs with excess heterozygosity resulting in Hardy–Weinberg *P* values less than $1 \times 10^{-5}$. The final SNP dataset used for GWAS consisted of 116,296 SNPs from 137 samples. For PCA, 1,257 unanchored SNPs were removed, and the remaining 115,039 SNPs were LD-pruned using PLINK (command: –indep-pairwise 10 3 0.5), resulting in 80,939 SNPs.

*Genetic and chemical analysis*

The chemical distance between cultivars was calculated as the Euclidean distance using the ‘dist’ function in R from the matrix of metabolomic data—that is, 40 terpenes and

cannabinoids quantified across 297 samples. The genetic similarity between samples was calculated as an inverse identity-by-state matrix generated in PLINK. The correlations between the matrices were computed using a Mantel test in R (R Core Team 2018) by first reducing the chemical matrix to the 137 samples with both chemical and genetic datasets. PCA was performed on the scaled genetic and chemical data using the prcomp function in R. To calculate the variance in labelling explained by the chemical and genetic data, linear models including the top ten PCs from the genetic data, the chemical data and both the chemical and genetic datasets together were performed. Pearson correlations between chemical concentration and the 1-to-5 Sativa–Indica scale were performed with the cor.test function in R. A Bonferroni correction was applied to the *P* values from the correlation test between chemical concentration and the Sativa–Indica scale.

*Genome-wide association*

We performed GWAS for 40 terpene and cannabinoid phenotypes, using both normalized and non-normalized data. Normalizing was conducted to generate values for a chemical concentration in a sample relative to the total abundance of its chemical class (that is, monoterpene, sesquiterpene or cannabinoid) in that sample. Thus, a sample's myrcene content was divided by the total concentration of all monoterpenes in that sample to generate a normalized value for myrcene. GWAS was performed using an MMLM (Segura et al. 2012) accounting for relatedness using a kinship matrix created in TASSEL (v.5.0) (Bradbury et al. 2007). The MLMM incorporates significant SNPs as cofactors

using stepwise regression (maxsteps = 10), and the optimal model was chosen on the basis of the extended Bayesian information criterion. We also present the first step of the MLMM, which is equivalent to an MLM where relatedness is accounted for but no SNPs are included as cofactors. Using the simpleM40 package in R, the effective number of independent tests ($M_{eff}$) was generated, and the threshold for significance was then calculated using $-\log_{10}(\alpha/M_{eff})$, where $\alpha = 0.05$. Quantile–quantile and Manhattan plots were created using the qq function in R. Genomic regions with significant GWAS hits were explored, and the physical locations of genes within these regions were retrieved using annotations from the CBDRx reference genome (Grassa et al. 2021) in Geneious Prime (v.2020.1.2). The GWAS results and LD regions of interest were visualized using code adapted from ref. (Hu et al. 2019).

**Results and Discussion**

Principal component analysis (PCA) of the genomic data showed no clear clustering according to sample labels (Figure 4-1A). Even though PC1 and PC2 were significantly correlated with the Sativa–Indica scale, the variance explained by the primary PCs was low (PC1: $R^2 = 0.12$, $P = 2.1 \times 10^{-5}$; PC2: $R^2 = 0.12$, $P = 1.8 \times 10^{-5}$). Furthermore, the overall genetic structure (captured by including the first ten PCs of the genomic data in a linear model) explained only 37% of the variance in labelling (Figure 4-1C). Sativa–Indica labels thus do not accurately reflect genetic relatedness, which is consistent with previous work (Sawler et al. 2015; Lynch et al. 2017). In addition, we determined that pairs of samples with identical cultivar names (for example, OG Kush) were often as genetically and chemically distant from each other as pairs of samples with different

names (Appendix III: Figure III-II). This is consistent with previous studies indicating that cultivar names were not reliable indicators of a sample's genetic or chemical identity (Sawler et al. 2015; Henry et al. 2020; Schwabe and McGlaughlin 2019; Smith et al. 2021).

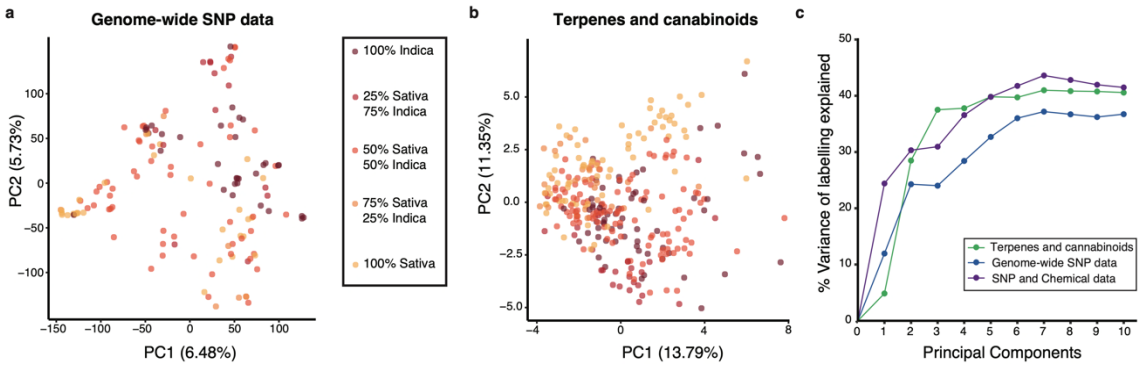


**Figure 4-1.** PCA. A) Genome-wide SNP data. B) Terpenes and cannabinoids. Each dot represents a *Cannabis* sample and is coloured by the labelling scale ranging from 100% Sativa to 100% Indica. C) The percent variance explained by PCs from the genome-wide SNP data (blue), from the terpene and cannabinoid data (green) and from both the genetic and chemical data (purple). The *y* axis shows the percent variance explained as PCs are added to linear models where the Sativa–Indica labelling scale is the dependent variable.

Similar to the PCA of the genome-wide SNP data, the PCA of the terpene and cannabinoid profiles provided poor separation of samples according to their Sativa–Indica labels (Figure 4-1B). Nevertheless, we observed significant correlations between the first two PCs and the Sativa–Indica scale (PC1: $R^2 = 0.049$, $P = 7.5 \times 10^{-5}$; PC2: $R^2 = 0.24$, $P = 3.7 \times 10^{-19}$). Including the first ten PCs from the terpene and cannabinoid profiles in a linear model accounted for only 41% of the variance in labelling

(Figure 4-1C). The pairwise genetic and chemical relatedness matrices were correlated (Mantel $r = 0.21$, $P = 1 \times 10^{-3}$, Appendix III: Figure III-III), and a linear model including the first ten PCs from both the genomic and chemical profiles captured only 41% (Figure 4-1C; $P = 3.1 \times 10^{-10}$) of the variance in labelling. Since the overall patterns of genetic and chemical relatedness could not fully account for the labels applied to *Cannabis* samples, we aimed to determine which individual chemicals were the strongest predictors of Sativa–Indica labelling.

Of the 40 measured terpenes and cannabinoids, 12 (30%) were correlated with the Sativa–Indica scale at $P < 0.01$ (Figure 4-2A and Appendix III: Figure III-IV). Sativa content was positively correlated with the concentrations of bergamotene ($R^2 = 0.12$, $P = 9.26 \times 10^{-8}$) and farnesene ($R^2 = 0.11$, $P = 1.09 \times 10^{-7}$), which impart tea-like and fruity aromas, respectively ("Compound Card for α-Trans-Bergamotene" 2020; Russo and Marcu 2017). This is consistent with descriptions of Sativa cultivars as having a 'sweet' or 'herbal' aroma (Guy and McPartland 2017; Gilbert and DiVerdi 2018). The strongest correlation was between Indica content and myrcene, whose concentration explained 21.2% of the variation in labelling ($P = 2.29 \times 10^{-15}$; Figure 4-2A). The sedative effect and earthy aroma attributed to high myrcene content are often reported by recreational users to be characteristic of Indica cultivars (Russo 2011; Pearce, Mitsouras, and Irizarry 2014; Temple and Leikin 2020; Hartsel et al. 2016). We also observed significant positive correlations between Indica labelling and three sesquiterpenes: guaiol ($R^2 = 0.18$, $P = 7.7 \times 10^{-13}$), γ-eudesmol ($R^2 = 0.11$, $P = 3.8 \times 10^{-7}$) and β-eudesmol ($R^2 = 0.21$, $P = 8.2 \times 10^{-15}$). Hillig (Hillig 2004) found that these three sesquiterpenes

were associated with plants from Afghanistan, which is considered the region of origin for Indica cultivars.

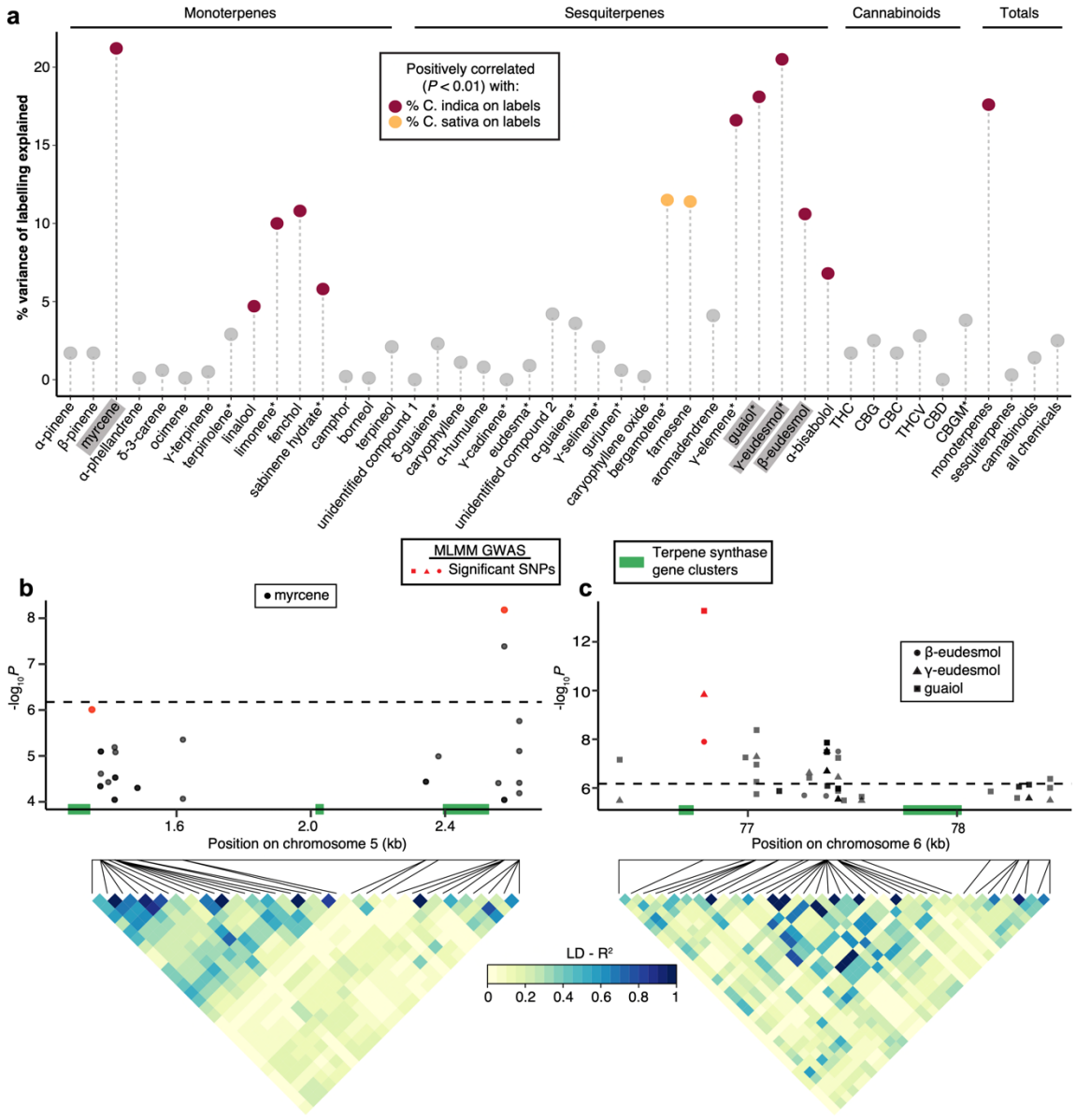


**Figure 4-2.** A) The percent variance of the five-point Sativa–Indica labelling scale that is explained by terpene and cannabinoid concentrations from Pearson correlations. The *P* values were Bonferroni-adjusted for multiple comparisons. The asterisks denote chemicals with tentative identifications. GWAS results are shown for chemicals

highlighted in grey. B) Manhattan plots of mixed linear model (MLM) GWAS for myrcene on chromosome 5 (C) Manhattan plots of mixed linear model (MLM) GWAS for guaiol, γ-eudesmol and β-eudesmol on chromosome 6. The significance thresholds from the MLM are shown as horizontal dashed lines. Significant SNPs from the MLMM GWAS are red. Terpene synthase gene clusters are green. Below the Manhattan plots are heat maps of the pairwise LD ($R^2$) between pairs of SNPs that appear in the Manhattan plots.

Previous chemical analyses of *Cannabis* have suggested that the distinction between Sativa and Indica is best explained by differences in the concentrations of specific monoterpenes and sesquiterpenes (Henry et al. 2020; Elzinga et al. 2015; Casano et al. 2011; Fischedick et al. 2010). In addition, the contrasting aromas that have been associated with Sativa (that is, sweet) and Indica (that is, earthy) were key discriminators in a sensory evaluation of *Cannabis* cultivars and mediated customers' perceptions of potency and quality (Gilbert and DiVerdi 2018). As a previous study suggested (Mudge, Brown, and Murch 2019), we hypothesize that *Cannabis* growers and breeders have been assigning labels to cultivars primarily on the basis of aroma profiles and purported effects, rather than genetic ancestry or overall chemical similarity. The primary differences between cultivars labelled as Sativa and Indica may thus be driven by a small set of genomic regions controlling the concentrations of a small number of contrasting aromas. To examine this, we conducted a genome-wide association study (GWAS) of the 40 chemicals examined here (Appendix III: Figure III-V and Table III-III).

We identified three regions of the Cannabis genome associated with the four terpenes most strongly associated with Sativa–Indica labelling (Figure 4-2). The optimal model from the multilocus mixed linear model (MLMM) GWAS for myrcene identified two significantly associated SNPs 1.2 megabases apart that tag independent blocks of linkage disequilibrium (LD) on the proximal end of chromosome 5 (Figure 4-2B). The first SNP (chr5:1348048) is located 6.4 kilobases (kb) from a block of terpene synthase genes composed of four copies of *TPS30*, which is known to encode myrcene synthase (Booth, Page, and Bohlmann 2017) (Appendix III: Table III-III). The second SNP (chr5:2576403) is 46.7 kb from another tandem array of terpene synthase genes spanning ~200 kb (Appendix III: Table III-III). Within this gene cluster are two sequences highly similar to the myrcene synthase gene, *TPS3* (Booth, Page, and Bohlmann 2017; Zager et al. 2019). These observations suggest that myrcene synthesis is mediated by genetic variants at two independent terpene synthase gene clusters on chromosome 5. The other three sesquiterpenes (guaiol, β-eudesmol and γ-eudesmol) strongly associated with Sativa–Indica labelling are correlated with each other (Appendix III: Figure III-VI) and share a common GWAS hit on chromosome 6: the single SNP identified from the MLMM (chr6:76790611) is 51.9 kb from a gene cluster comprising sesquiterpene synthase genes related to TPS7FN (δ-selinene synthase), TPS8FN (γ-eudesmol/valencene synthase) (Booth, Page, and Bohlmann 2017) and TPS20CT (Zager et al. 2019) (hedycaryol synthase) (Figure 4-2C and Appendix III: Table III-III).

Our results demonstrate that the Sativa–Indica scale currently used to label *Cannabis* poorly captures overall genomic and metabolomic

variation. *Cannabis* labelling is instead probably driven primarily by a small number of key terpenes whose concentrations contribute to the characteristic aromas commonly associated with Sativa and Indica and whose variation we genetically mapped to tandem arrays of terpene synthase genes on chromosomes 5 and 6. While the vernacular labels 'Sativa' and 'Indica' are derived from taxonomic names that were originally used to categorize plants according to ancestry (Guy and McPartland 2017), these terms have been co-opted by contemporary *Cannabis* culture and now probably reflect locus-specific genetic variation affecting terpene synthesis. Our results suggest that a practical and reliable classification system for *Cannabis* that is consistent with contemporary understanding of the terms 'Sativa' and 'Indica' may be achievable by quantifying a small number of terpenes and/or genotyping genetic markers associated with key *Cannabis* aromas.

**Acknowledgments**

We thank A. Hazekamp, M. E. Schranz and F. F. M. Becker for their contributions to this work. We thank C. F. Forney and T. Soomro for their assistance. This work was funded primarily by Bedrocan but was also supported by the National Science Foundation Plant Genome Research Programme grant no. 154686 to Z.M. and a Vanier Scholarship from the National Sciences and Engineering Research Council of Canada to S.W.

# Chapter 5: From GWAS to CRISPR: a direct route to the gene-edited orchard.

## Introduction

Sustainable and resilient food systems rely on the cultivation of a diversity of agricultural crops (McCouch et al. 2013). Although the majority of the calories consumed globally come from a small number of annual crops (i.e. corn, rice and wheat), perennial crops remain important contributors to nutritious diets and healthy agroecosystems (Vinson et al. 2001). Due to their multi-year lifecycle, perennial crops sequester more carbon and have lower risk of soil erosion compared to annuals (Kreitzman et al. 2020; Vallebona, Mantino, and Bonari 2016). Furthermore, perennial crops fill important nutritional niches and are key economic sources for small scale farmers (Kreitzman et al. 2020). Unfortunately, perennial fruit crops are especially susceptible to the detrimental effects of climate change due to specific temperature requirements and prolonged life cycles (Winkler et al. 2013; Leisner 2020).

Apple is an important perennial crop with a broad distribution worldwide: it ranks second in economic value among fruit crops with an estimated global production of $98 billion US in 2018 (FAOSTAT http://faostat.fao.org). Apple breeding has traditionally been a long and labour intensive process: it takes upwards of 25 years to commercialize new cultivars from seed (Peil et al. 2008). The slow breeding pipeline for apples is in part due to its self-incompatibility, prolonged juvenile phase and costly cultivation. Although traditional breeding has resulted in the improvement of apple flavour and texture, much

improvement remains to be made for traits such as disease and pest resistance, abiotic stress, storability and nutrition. The development of new apple cultivars with a combination of these desirable traits is crucial to ensuring sustainable production in the future.

Genomic technologies hold promise to accelerate crop improvement in order to enhance agricultural sustainability and resiliency. Over the past decade, next generation sequencing has enabled genomics assisted breeding of numerous agricultural crops leading to the development of new cultivars with superior traits. The first reference genome for apple was published in 2010 (Velasco et al. 2010). Since then, numerous assembled apple genomes have become available (Peace et al. 2019) and a major outcome of genomic analyses has been the development of markers useful for marker-assisted selection (MAS) (Chagné et al. 2019, 2016; Jänsch et al. 2015; Baumgartner et al. 2016). These markers are linked to quantitative trait loci (QTL) for key fruit quality and disease resistance traits and their use can help reduce the cost of apple improvement by allowing breeders to select for traits of interest at an early juvenile phase (Luby and Shaw 2001). However, while MAS can help breeders identify promising new varieties at lower costs, its use has little effect on the time it takes to commercialize an apple from seed: apple breeding still remains a long process that is both labour intensive and costly. The fact that the apple is recalcitrant to rapid improvement even when employing modern MAS methods suggests that additional approaches are needed for the apple to truly benefit from the genomics revolution.

As an alternative genomic tool to MAS, gene editing is capable of directly introducing deletions, insertions and other alterations to any organism's DNA sequence (Doudna and Charpentier 2014). By inducing precise alterations to plant genomes, gene editing has already been used to develop higher yielding rice varieties and wheat with improved disease resistance (Zaidi et al. 2019). The basis of many gene editing systems relies on the use of programmable nucleases that are capable of targeting specific regions of a genome to introduce double stranded breaks (DSBs). Although numerous editing systems exists, like zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs), the most popular system due to its simplicity and efficiency is the RNA guided clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated protein (Cas) system (CRISPR/Cas) (Hsu, Lander, and Zhang 2014; Doudna and Charpentier 2014). A crucial component of the CRISPR-Cas system is the single guide RNA (sgRNA) that is designed to be complementary to the target genomic region and therefore determines the specificity of the induced edits. The sgRNA directs the Cas enzyme to the target region and the Cas enzyme introduces a DSB in the DNA at the desired location. Then, through repair mechanisms, mutations are introduced at the DSB site (Doudna and Charpentier 2014). The rapid development of gene editing systems has allowed for targeted editing of numerous perennial fruit crops including apple as well as citrus, grape, banana, pomegranate, cacao and kiwifruit (Jia et al. 2016; C. Ren et al. 2016; Pompili et al. 2020; Naim et al. 2018; Chang, Wu, and Tian 2019; Fister et al. 2018; Varkonyi-Gasic et al. 2019). Thus far, in apple, studies have demonstrated successful editing of genes involved in fire blight resistance, phytoene desaturase, shoot regeneration, flowering and herbicide resistance (Nishitani et al. 2016;

Malnoy et al. 2016; Charrier et al. 2019; Pompili et al. 2020; Malabarba et al. 2020; H. Li et al. 2023). By introducing novel mutations directly into the apple genome, gene editing is a promising tool for apple improvement that could dramatically accelerate the time it takes to commercialize novel apple cultivars.

With CRISPR now widely available, researchers are rapidly exploring traits that can be improved with this promising new technology. However, to expand the list of phenotypic changes that can be made with gene editing in apples, there are numerous technical barriers that must be overcome. The key challenge that arguably precedes all others is identifying causal genetic variants that impact phenotypic traits. Nasti and Voytas (2021) described this dilemma rather succinctly: "An editor, of course, has no value unless there is text to edit." Editable text is elucidated through genotyping of an organism's genome and determining the causal genetic variants that underlie phenotypic variation. This sounds simple in principle, but pinpointing causal genetic variants for relevant phenotypes in crop genomes remains a major challenge in agricultural genomics. The most common methods to detect genotype-phenotype associations in apple have resulted in the identification of numerous QTL, but very few causal genetic variants have been discovered. Very few mapping studies have yielded actionable gene editing targets because experimental designs and populations used to perform most genetic mapping in apple were not established with the purpose of identifying causal genetic variants. Instead, these "mapping" populations were usually breeding populations: breeders generated bi-parental crosses to generate novel cultivars and then subsequently used data from the offspring to identify markers that could be used for MAS. This has led to an

abundance of low-resolution genetic mapping experiments in apple: most studies result in large QTL regions containing thousands of potential causal genetic variants (Kenis, Keulemans, and Davey 2008; K. A. McClure et al. 2016; Kostick et al. 2021). Researchers must then generally perform years of fine mapping, candidate gene guesswork, and laborious functional assays in order to narrow their search from a large QTL interval to a shortlist of potentially causal genetic variants. Much of the effort put into identifying causal genetic variants could be avoided with careful experimental design: a single, optimally-designed mapping population can deliver the mapping resolution required to efficiently identify gene editing targets.

Given the intense focus of recent research on developing gene editing tools, the value of the precise genomic coordinates of genetic variants that result in desirable phenotypes has increased dramatically. Here I present a framework for rapidly advancing apple breeding using gene editing (Figure 5-1). I propose that large GWAS populations maximizing species diversity should undergo whole-genome sequencing (WGS) and comprehensive phenotyping to identify causal alleles that are responsible for phenotypic variation. From GWAS, I propose directly undertaking CRISPR mutagenesis based on a small number of candidate functional variants across multiple genetic backgrounds to both validate genotype-phenotype relationships and create novel cultivars in a single step. Although this approach requires a large investment up front, if designed correctly, it may enable a fast pipeline to shortlist candidate causal genetic variants that can be introduced into germplasm and field tested. Below is a description and discussion of the steps involved in

this approach, from the discovery of putatively causal genetic variants to the commercialization of gene edited apples.

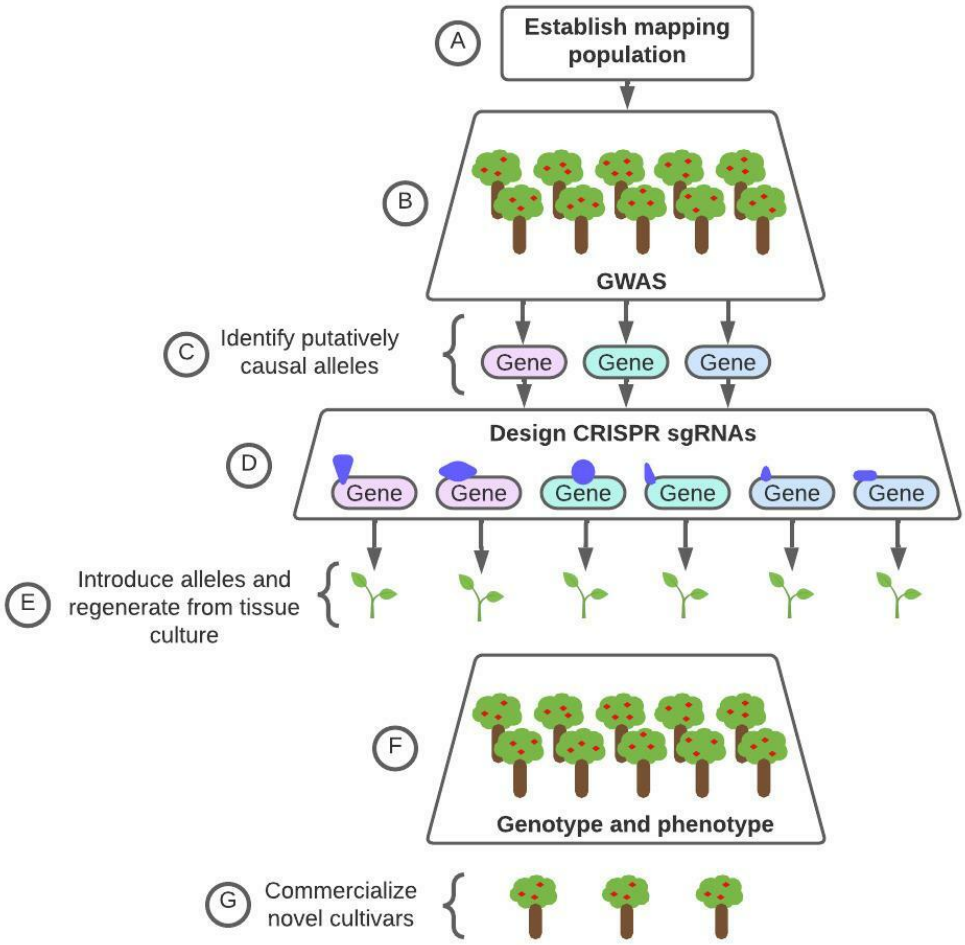


**Figure 5-1.** GWAS to CRISPR approach. A) Establish a large, diverse population to map traits of interest. B) Conduct genome-wide association studies. C) Identify putatively causal alleles. In this example, three different genes have been identified. D) Design CRISPR sgRNAs to edit and introduce novel alleles. Edits are represented by the blue shapes (ie. insertions, deletions or alleles swaps). E) Introduce CRISPR sgRNAs into

multiple genetic backgrounds and regenerate plantlets from tissue culture. Here, one edit is introduced into one genetic background, however this step can be multiplexed to introduce edits into multiple genes within a genetic background to stack traits or target a gene family. In addition, this process can be performed across multiple genetic backgrounds. F) Evaluate edited germplasm to determine if edits generated the desirable phenotypes. G) Select improved accessions with desirable phenotypes for commercialization.

## Genetic Mapping

The purpose of much research undertaken in agricultural genomics since high-throughput genotyping methods were developed has been to link phenotypes to genotypes. To edit the genome of a crop to express a specific phenotype, the causal alleles conferring the desired phenotype must be identified. The initial mapping approach has been described as the "rate-limiting" step in causal allele identification (Weigel and Nordborg 2005). Therefore, it is necessary to consider what experimental designs and mapping populations are needed to best exploit a future where genome editing will be ubiquitous. Within this section, I compare different mapping approaches and finish by providing a streamlined mapping approach for identifying gene editing targets in apple.

### *Linkage mapping*

Linkage mapping makes use of populations of closely related individuals to identify genetic loci associated with phenotypic traits. The experimental designs adopted to identify phenotype-genotype relationships through linkage mapping were often established with the intention of finding markers useful for MAS. In apple, numerous linkage mapping studies have been conducted to identify QTL for fruit quality, disease resistance and phenology traits (Verma et al. 2019; Kostick et al. 2021; Chagné et al. 2014). Most populations used for linkage mapping are derived from bi-parental crosses where limited recombination events have occurred leading to extended linkage disequilibrium (LD) (Myles et al. 2009). The resulting extended LD limits the mapping resolution and results in the identification of large genomic intervals associated with the measured phenotype, and these intervals are often millions of base pairs in length. Attempting to find the causal genetic variants within these large intervals frequently involves considerable effort. Thus, although linkage mapping is sufficient for identifying markers associated with a phenotype, it is unlikely to pinpoint the causal genetic variant(s) underlying that phenotype. For example, a recent linkage mapping study identified three QTL associated with fire blight resistance (Kostick et al. 2021). The QTL intervals ranged from 4.8 to 11.7 Mbp long and each had over 70 genes that were annotated as disease resistance genes. To find gene editing targets from long QTL intervals, additional fine mapping is required to further reduce the size of the QTL interval, and this is time consuming and, in many cases, impractical. Overall, linkage mapping has been useful in the development of markers useful for MAS, but higher resolution genetic mapping methods are needed to identify robust gene editing targets.

*GWAS*

Genome-wide association studies (GWAS) make use of diverse, largely unrelated, individuals to uncover significant associations between phenotypic traits and genetic variants. GWAS have been used to map numerous traits in apple (Migicovsky et al. 2016; K. A. McClure et al. 2018; Larsen et al. 2019; Urrestarazu et al. 2017; Kumar et al. 2013; Jung et al. 2022). In contrast to linkage mapping, GWAS query a far greater amount of genetic diversity, which means that GWAS captures more recombination events and thus provides higher mapping resolution than linkage mapping (M. A. Khan and Korban 2012). In addition, due to the genotypic and phenotypic diversity within GWAS populations, a single population can be used to identify multiple genetic variants underlying multiple traits (C. Zhu et al. 2008). In contrast, linkage mapping is usually limited to assessing only QTL that segregate between the parents of the initial cross. Therefore, to identify gene editing targets for multiple traits, association mapping is more efficient and cost-effective than linkage mapping. However, GWAS has limited power to detect associations when a trait is highly correlated with population structure or when the trait is low in frequency in the population (Yang et al. 2014). Therefore, allele frequencies and the genetic architecture of a trait should be considered when designing an association mapping experiment.

At present, only a few single large effect loci have been identified using GWAS in apple. These include loci for harvest date and acidity, which have been mapped to chromosomes 3 and 16, respectively. The chromosome 3 locus for harvest date was a novel association

first identified through GWAS of 689 diverse apple accessions (Migicovsky et al. 2016). It was found that at the chromosome 3 locus, a single-nucleotide polymorphism (SNP) in a NAC transcription factor was the most significant variant associated with harvest date (Migicovsky et al. 2016). The large effect of this locus on harvest date was further demonstrated in a larger population where the SNP at NAC explained over 40% of the variance in harvest date (Migicovsky, Yeats, et al. 2021). The harvest date locus on chromosome 3 has since been mapped in multiple germplasm collections around the world using GWAS (Urrestarazu et al. 2017; Larsen et al. 2019; Jung et al. 2022). The chromosome 3 locus was missed by previous linkage mapping studies due to the fact that most of the popular cultivars used in breeding programs, which are the parents of most linkage mapping populations, are fixed for the late-harvested allele at this locus (Migicovsky, Yeats, et al. 2021; Jung et al. 2022). If all accessions within a mapping population have the same genotype at a QTL, then the QTL cannot be mapped using that population. Thus, it was the highly diverse nature of the mapping population in the initial GWAS for harvest date that enabled the identification of the chromosome 3 locus. This underscores the ability of GWAS to exploit the apple's tremendous natural genetic diversity to uncover novel genotype-phenotype associations controlling key traits.

While GWAS is a preferred method due to the high mapping resolution it can offer, in some cases linkage mapping has led to the identification of causal genetic variants. For example, the involvement of the Ma1 locus in malic acid content in apples was first identified in 1998 through linkage mapping (Maliepaard et al. 1998). Subsequent fine mapping refined the Ma1 locus to a 150kb region within the Golden Delicious genome

(Xu, Wang, and Brown 2012). The Ma1 locus was further reduced to a 65kb region within which a SNP in the malate transporter-like gene was identified as the causal variant leading to variation in malic acid content (Bai et al. 2012). It is worth noting that the Ma1 locus was first identified using linkage mapping in 1998, but it took 14 years to fine-map the locus and identify the causal mutation (Maliepaard et al. 1998; Xu, Wang, and Brown 2012). In contrast, the causal allele at the Ma1 locus was easily identified in a single GWAS experiment that used WGS of 497 diverse apple accessions (Liao et al. 2021). This highlights the power of GWAS to directly identify causal variants when coupled with WGS. WGS has only started to become feasible for hundreds or sometimes thousands of samples, but it is likely that the price of sequencing will continue to decrease, making GWAS an even more cost-efficient method for identifying causal alleles.

These two examples demonstrate that GWAS of large, diverse populations coupled with WGS data is a mapping approach that is ideally suited to identify regions of the apple genome that can be edited to introduce key phenotypic changes. It is important to note that WGS enables the discovery of not only SNPs, but structural variants as well. Structural variants are increasingly being found to control phenotypic variation in plants (Saxena, Edwards, and Varshney 2014). A recent GWAS used WGS data from 417 peach accessions to uncover three causal structural variants controlling fruit shape, harvest date and flower morphology, for example (Tan et al. 2021). To identify gene editing targets in apple, exhaustive WGS of large numbers of accessions should be conducted in order to exhaustively catalog causal genetic variation that can be targeted with gene editing.

*Short-listing candidate targets*

If the mapping approach, whether linkage mapping or GWAS, is unable to pinpoint the exact locations of causal alleles, then functional analysis may be conducted to try to validate putatively causal variants. Functional analyses may involve, for example, transgenic complementation, functional studies of different alleles, knock-out experiments or mutational analysis (Weigel and Nordborg 2005). The default procedure is to continue amassing evidence until one can conclude definitively that a genetic variant has a phenotypic effect (Ogura and Busch 2015; Weigel and Nordborg 2005). In apple, where a trait may need to be evaluated at the adult fruiting stage, these experiments can take between 5 and 10 years (Maliepaard et al. 1998; Xu, Wang, and Brown 2012). Even testing transgenic complementation in a rapidly growing model organism, such as tomato, requires considerable time and effort and may not reveal the precise causal genetic variant or molecular mechanism leading to phenotypic variation (Migicovsky, Yeats, et al. 2021). The functional validation of putatively causal genetic variants for a particular trait is difficult and time consuming even in crops with advanced genomic resources, and it is particularly burdensome in a long-lived woody perennial like the apple.

Although the gold standard would be to identify causal alleles directly from GWAS, even with large and diverse mapping populations, high quality reference genomes, reliable phenotype data, and WGS, it may not be possible to identify causal alleles with absolute

certainty. Wherever possible, *in silico* analysis should be conducted to assess the potential function of putatively causal variants identified via GWAS. *In silico* analyses make use of available databases to prioritize causal alleles and typically involve surveying functional annotation data or analyzing cross-species annotation data (Hassani-Pak and Rawlings 2017). Evidence provided by *in silico* analyses is orthogonal to the GWAS mapping results, and can thus lead to a refinement, and an improved ranking, of the shortlist of candidate targets (Broekema, Bakker, and Jonkers 2020).

With a shortlist of putatively causal variants in hand, one can then skip the traditional functional analyses and conceivably begin gene editing immediately by producing edits in various combinations, targeting multiple phenotypes across diverse genetic backgrounds (Figure 5-1). The main purpose of this approach is to move as quickly as possible to gene editing and this is beneficial for multiple reasons. First, for many desirable fruit quality traits in apple, the strongest piece of evidence confirming causality will be the assessment of the fruit. Therefore, establishing edited trees in the ground as soon as possible is critical because it will take 5 years for those trees to produce fruit for commercial evaluation. Second, the route of identifying targets and then trying to validate the mechanism of phenotypic change before planting is simply delaying the inevitable need for field validation. If a gene edited apple cultivar is to be commercialized, end users (e.g. growers, nurseries, variety rights management companies) will require the trait improvement to be quantified and demonstrated in the orchard over the course of multiple years in any case. Therefore, in the context of advancing crop improvement, there is potentially a distinct advantage to forgoing functional validation for the sake of

understanding the molecular basis of a trait in favor of immediately generating orchards of gene edited trees for commercial evaluation.

## Gene editing challenges

To move rapidly towards the commercialization of gene edited apple trees, it is critical that short-lists of targets identified via genetic mapping are introduced quickly into multiple desirable genetic backgrounds. As outlined in the previous section, one of the primary challenges with realizing the potential of gene editing in apple is the identification of reliable editing targets that elicit desirable phenotypic changes. However, with reliable editing targets in hand, there remains several technical hurdles to introducing edits into germplasm and regenerating plants. Here, we focus on two of the technical challenges that follow the identification of editing targets; difficulties with regeneration of edited plants through tissue culture and the current limitations of editing technologies in apples.

### *In vitro tissue culture*

The foundation of any transformation system, whether it be for gene editing or trans/cis-genesis, is a tissue culture system that can successfully regenerate transformed plants. Unfortunately many clonally propagated tree species are difficult to regenerate, and successful regeneration of one cultivar does not mean that all cultivars are amenable to regeneration (Birch 1997; Nagle et al. 2018; Goralogia, Redick, and Strauss 2021).

Although there are numerous examples of successful regeneration of transformed apple tissue, the regeneration efficiency varies significantly between cultivars (Aldwinckle and Malnoy, n.d.). However, innovative ways to improve regeneration in apple are being developed. For example, a recent study used gene editing to introduce mutations within the *MdSPL6II* gene in apple which led to improved shoot regeneration (H. Li et al. 2023). The transformation of plant tissue with developmental regulators, such as BABYBOOM (*BBM*), WUSCHEL (*WUS*), and GROWTH REGULATOR FACTOR 4 (*GRF4*), can help improve regeneration (Debernardi et al. 2020; Gallois et al. 2002; Lian et al. 2022; J. Chen et al. 2022). Recently, the transformation of apple tissue with *BBM* led to significantly higher rates of transformed apple shoot production (J. Chen et al. 2022). However, the insertion of developmental regulator genes into the genome can become problematic for regulations of gene edited plants further down the line (discussed below in *Transgene-free editing*) (Goralogia, Redick, and Strauss 2021). Furthermore, without careful control of expression, developmental regulators can lead to adverse effects in later developmental stages (Nagle et al. 2018; Gallois et al. 2002).

There are also growing concerns surrounding the random mutations that can be introduced into germplasm through the process of *in vitro* tissue culture. A recent study of a fire blight resistant cis-genic apple line revealed that the cis-genic and wild type lines that underwent tissue culture experienced a loss of red skin colour compared to the wild type lines that were not propagated through tissue culture (Schlathölter et al. 2023). Schlathölter et al. suggested that the loss of red skin colour was likely due to somaclonal variation within the apple flavonoid pathway (2023). Overall, the mechanism through

which *in vitro* tissue culture can lead to unintended effects is an important issue that warrants further investigation.

*Gene editing technology*

Since its discovery in 2012, CRISPR technology has enabled us to introduce diverse types of changes to plant genomes (Doudna and Charpentier 2014). The majority of gene editing in plants involves the introduction of knock-out mutations since gain-of-function allele swaps and large insertions are more challenging (Nasti and Voytas 2021). The DSBs introduced by Cas proteins are repaired most commonly through non-homologous end joining (NHEJ). NHEJ results in the incorporation of small, random insertions or deletions into the targeted region and is therefore often used to induce knockout mutations. To date, published examples of gene editing in apple have primarily involved inducing knockout mutations that cause albino phenotypes (*MdPDS*), early flowering (*MdTF1*) and fire blight resistance (MdDIPM) (Pompili et al. 2020; Nishitani et al. 2016; Charrier et al. 2019; Dalla Costa et al. 2020; Malnoy et al. 2016).

It is worth noting that at present, gene editing cannot be used to alter all traits because of the nature of the underlying genetic changes that are required to alter the phenotype. For example, many sought after traits require gain-of-function mutations that depend on the insertion of new sequence segments or changes in nucleotide sequence. Achieving disease resistance, for example, often requires the insertion of resistance genes from closely related species. In order to achieve this, a donor DNA template can be supplied

and homology directed repair (HDR) can incorporate the desired substitution and insertion. However, there are few instances of HDR being used to edit plants, as HDR produces edits only at relatively low frequency and is therefore inefficient (Shaoya Li and Xia 2020; Nasti and Voytas 2021). Furthermore, HDR can only handle insertions of up to 50 nucleotides, which severely limits its ability to introduce sequences that will lead to phenotypic changes. Recent progress has been made with twin prime editing to introduce larger deletions, substitutions and insertions, but its efficacy in plants has yet to be demonstrated (Anzalone et al. 2021).

Although insertions are often necessary to induce many forms of gain-of-function mutations, there is also a need to perform single nucleotide changes that lead to phenotypic changes. The ability to make single nucleotide changes is critical for editing genes that are essential for the organism to live: these genes cannot be completely disrupted via a knockout but may produce desirable phenotypic changes when altered slightly by a single nucleotide change. For example, gene editing could be used to introduce precise alterations, rather than knockouts, to ripening genes to create apples with varied ripening times. CRISPR base editors, composed of adenine and cytosine deaminases, have been shown to introduce precise base alterations without the need for DSBs. To date, only one report of base editing has been published in apple: a single nucleotide change was introduced in an *ALS* gene that confers herbicide resistance (Malabarba et al. 2020). At present, base editors are only capable of inducing transition mutations (i.e. A↔G, or C↔T) which limits their ability to introduce the full spectrum of nucleotide changes (Anzalone et al. 2019). Prime-editors have further expanded the scope

of gene editing by enabling all 12 types of nucleotides substitutions without the need for donor DNA or DSBs, but their use has only been demonstrated in a small number of crops such as rice, tomato and maize (Hao, Pu, and Song 2021) .

The difficulty in inducing some forms of mutation using gene editing begs the question: why not just use transgenic methods to induce phenotypic changes in apple? Although examples of transgenic apples exist (e.g. the non-browning Arctic® apple (Armstrong and Lane 2014), a red-fleshed apple (Espley et al. 2013), and an early flowering apple (Kotoda et al. 2006)), there is potential value in attempting to recreate these phenotypes using gene editing rather than relying on transgenic approaches. Below I highlight four benefits of gene editing that can advance apple improvement.

**Gene editing advantages**

*Transgene-free editing*

One of the major benefits of gene editing is that it allows for genomes to be altered without the permanent introduction of transgenes into the recipient genome (Gu, Liu, and Zhang 2021). In countries like the United States, Canada and Brazil, it is possible for transgene-free gene-edited crops to bypass stringent regulations that often plague genetically modified crops. This is because these countries have adopted regulations based on the final product (i.e. the new cultivar), rather than the process used to develop the new cultivar (Turnbull, Lillemo, and Hvoslef-Eide 2021; Entine et al. 2021). The foundation of product-based regulation is the premise that the resulting gene-edited plant

could have, in theory, arisen from random genetic mutations (Turnbull, Lillemo, and Hvoslef-Eide 2021). While gene editing can produce plants that are transgene-free, it is the method of delivering the gene editing machinery that determines how the resulting plant will ultimately be classified by regulators.

For decades, biolistic bombardment and agrobacterium have been the primary methods used for stable transformation of plant genomes, but these methods lead to the stable integration of transgenes that must be segregated out via selfing or crossing to achieve DNA-free edited plants (Altpeter et al. 2016; C. Gao 2021). The removal of transgenes in this manner is not feasible with many perennial fruit crops due to self-incompatibility and long juvenile phases (Gu, Liu, and Zhang 2021). Methods have been developed to excise CRISPR machinery and selectable markers after stable agrobacterium transformation. For example, removal of the T-DNA expressing the CRISPR/Cas9 cassette has been successfully demonstrated in apple using a heat-inducible FLP/FRT system resulting in transgene-free apple lines with reduced susceptibility to fire blight (Pompili et al. 2020; Dalla Costa et al. 2020). An additional method that has been tested in apple is to place Cas9 cleavage target sites on the borders of the T-DNA insertion so that the Cas9 enzyme excises the T-DNA cassette after making the desired target site edit (Dalla Costa et al. 2020). These methods for excising T-DNA after stable integration still require refinement in order to be routinely used to generate transgene free edited apples.

An alternative to stable transformation is transient expression of editing machinery that does not involve integration into the plant genome. One method of transient

transformation is through the delivery of preassembled Cas proteins and sgRNAs that are delivered as ribonucleoproteins (RNPs). RNPs are delivered directly into protoplast cells. In apple, Malnoy et al. (2016) targeted fire blight susceptibility genes, by delivering CRISPR machinery using RNPs. Although delivery of editing reagents through RNPs is advantageous for developing transgene free edited cells, work still remains to successfully regenerate full plants from protoplast in apple (Malnoy et al. 2016). Agrobacterium-mediated transient transformation is another method to generate transgene-free edited plants and can be achieved by various methods that favour transient expression over stable integration (Goralogia, Redick, and Strauss 2021). Charrier et al. (2019) successfully regenerated transgene-free *PDS* edited apple lines via transient agrobacterium transformation, however the editing rate was extremely low at 0.26%. Although the generation of transgene-free edited apples remains a major challenge, it is likely that these technical hurdles will be overcome given the magnitude of the efforts currently dedicated to solving these challenges across diverse agricultural species.

*Introducing novel variation*

Gene editing has expanded the repertoire of plant variation beyond the confines of natural variation and conventional breeding methods (Nasti and Voytas 2021). Through CRISPR-driven mutagenesis, it is possible to introduce novel genetic changes (i.e. not observed in nature) into plant genomes to determine if they alter phenotypes in desirable ways (Q. Li, Sapkota, and Knaap 2020). Cis-regulatory elements, including non-coding elements like enhancers and promoters, have been identified as ideal targets for CRISPR

mutagenesis as they control the timing and patterns of gene expression (Wittkopp and Kalay 2011). Altering cis-regulatory elements is an avenue to fine tune the expression of genes, an approach Rodiguez-Leal et al. (2017) described as an “untapped resource for expanding allelic diversity for breeding”. In tomato, Rodiguez-Leal et al. (2017) used CRISPR to introduce multiple edits into the promoter regions of genes for fruit size, inflorescence and architecture, and these changes resulted in a continuum of phenotypic variation for these traits. The phenotypic changes induced by the edits to the tomato promoters were conducted without detailed knowledge of the promoter's structure or regulatory pathway, so researchers aptly named this method “promoter bashing” (Rodríguez-Leal et al. 2017). Similarly, cis-regulatory elements were targeted with CRISPR in rice to increase yield (L. Liu et al. 2021). This supports the notion that full validation of function is not always necessary to successfully gene edit a genetic element and achieve a desirable phenotypic change. The power of gene editing comes from its ability to not only introduce variation that has been previously identified within a species, but to also induce novel genetic variation that does not exist in nature. Therefore, CRISPR-driven mutagenesis could be used to edit regions of the apple genome that are likely to control key traits to generate a continuum of phenotypic variation that can be harnessed for apple improvement.

*Target Specificity*

In comparison to other forms of genetic transformation, gene editing has the potential to target a very specific region of the genome. CRISPR-Cas9 targeting specificity is in part

due to the requirement of the NGG protospacer-adjacent motif (PAM) to be present at the 3' end of the 20 nucleotide target sequence in order for Cas9 to cleave the DNA (Jinek et al. 2012; K. Chen et al. 2019). Although the NGG PAM for Cas9 aids in target specificity, it also contributes to a more constrained editing space (Endo et al. 2019). For instance, if the target sequence is not located near a required CRISPR-Cas9 PAM, then that region is inaccessible to the Cas9 protein (Endo et al. 2019). Fortunately, recent developments have expanded the targetable regions of genomes through the development of Cas proteins with expanded PAM recognition, or proteins that do not require a PAM at all (Q. Ren et al. 2021). Beyond the popular Cas enzymes that have been used extensively for editing, there is a large number of currently unutilized enzymes with similar editing capabilities and potentially novel capabilities that could expand the gene editing toolbox (Rousset and Sorek 2021).

Another important consideration is that CRISPR specificity is not always faultless and it's possible for the sgRNAs to edit regions of the genome that were not the intended target due to sequence similarity (H. Zhang et al. 2017). To overcome this, there are numerous software programs that can help design guide RNAs with high specificity (Xie et al. 2017; H. Liu et al. 2017). In addition, selection of different Cas9 variants can also aid in reducing off-targets (Mao et al. 2019). Furthermore, target specificity can be increased by designing sgRNAs specifically for the genetic background in which editing is desired. However, this requires the availability of sufficient genomic data for the cultivar so that the target sequence can be assessed. The framework suggested here of moving from GWAS to CRISPR addresses this issue because it recommends collecting

WGS for the purposes of GWAS, and this WGS can then later be further leveraged to inform the design of guides that are specific to any cultivar in the initial GWAS population. The proposed dual use of the WGS data can thereby make the process of moving from mapping to edited orchard more efficient and cost-effective.

*Multiplexing*

A major advantage of gene editing over previous methods of genetic transformation is the ability to multiplex CRISPR's sgRNA system to target multiple traits at once (Mao et al. 2019). Multiplexing involves targeting multiple genomic regions using multiple sgRNAs that are derived from one CRISPR array (Cong et al. 2013). Due to challenges that remain in introducing edits into plant species, multiplexing is particularly promising to efficiently "stack" multiple traits by targeting multiple genes in one CRISPR array. Multiplexing is also useful when attempting to alter a trait that is controlled by multiple genes within a gene family. For example, the apple genome contains multiple *PPO* genes that vary slightly in genetic sequence, but each contribute to enzymatic browning of apple flesh. Therefore, a gene-edited non-browning apple, similar to the transgenic non-browning Arctic® apple (Armstrong and Lane 2014), could be developed by targeting multiple *PPO* genes with one CRISPR construct containing sgRNA for each *PPO*. Here I propose taking multiplexing a step further: I aim to consider multiplexing across multiple genes and across multiple cultivars so that potentially causal genetic mutations can be introduced and tested across diverse genetic backgrounds.

**Evaluate**

The final stage of the proposed approach is to evaluate the gene edited plants in the field (Figure 5-1F). Similar to CRISPR libraries that are developed to screen the potential of sgRNAs to alter a gene and subsequent phenotype (H. Zhu, Li, and Gao 2020), I propose assessing a refined CRISPR library *in vivo* within an orchard. Assessment of the edited orchard will require comprehensive phenotyping of not only the targeted traits, but many others to ensure there are no unintended effects on other important traits (i.e. pleiotropic effects). For example, if plants are edited to produce red-fleshed apples, phenolic content and flesh browning should be measured in the edited plants in addition to flesh colour since all of these traits have been shown to be correlated with one another (Espley et al. 2019). For certain targeted traits, like fruit quality, it may take up to 6 years before reliable phenotype data can be acquired, while it may be possible to phenotype disease resistance and abiotic stress tolerance at earlier stages of development.

Ideally, the outcome of evaluating the gene edited orchard is the identification of cultivars with novel traits along with a validation of the underlying genetic variation causing each novel trait. With phenotype data from the gene edited population, it is possible to make comparisons back to the original mapping population. In a sense, verifying that the edits introduced into the orchard resulted in desirable phenotypes is somewhat like genetic mapping: it involves associating genotypes with phenotypes. Thus, while at first genetic mapping is used to enable gene editing, this final evaluation can be understood as the reverse process: using gene editing to enable genetic mapping. This

final evaluation step serves to validate causal alleles, and these then represent a refined list of editing targets that can be repeatedly introduced into other cultivars in the future. The approach proposed here therefore does not end at the evaluation stage, but instead is the first step in a future of iterative improvement through multiple rounds of mapping, editing and apple improvement.

## Conclusion

The goal of every apple breeding program is to rapidly develop improved cultivars that positively impact farmers and consumers through increased sustainability and quality. Here I propose a model of accelerating apple improvement by exploiting two recent technological innovations: next-generation DNA sequencing and CRISPR gene editing. By pairing high-resolution genetic mapping with modern gene editing tools, I propose a framework that accelerates breeders through the steps from genomic targets, to edits, to field trials. The proposed approach attempts to shorten the interval between GWAS and CRISPR by forgoing functional validation in favour of directly editing a short-list of putatively causal alleles. The success of this approach relies upon the appropriate design of genetic mapping populations that capture large amounts of phenotypic and genetic diversity to deliver high-resolution gene editing targets using GWAS. This approach demands that CRISPR gene editing not be relegated to the final piece of the breeding pipeline, but instead be used throughout the process to introduce novel variation and induce diverse mutations to advance apple improvement. The

proposed GWAS to CRISPR approach provides a direct route to a gene edited orchard with the goal of advancing apple breeding in the era of genome editing.

# Chapter 6: Conclusion

## Summary of Findings

The objective of this thesis was to advance genomics assisted breeding in apple and cannabis by using GWAS to uncover the genetic underpinnings of key traits in both crops. The approach employed in the thesis was to first demonstrate the utility of germplasm collections for quantifying apple diversity and for the collection of phenotypic data necessary for GWAS. Following a phenotypic characterization of apple diversity, GWAS was conducted for fruit quality and phenology traits in apple to identify phenotype-genotype associations. Then, I applied a similar GWAS approach to identify associations between chemical content and genetic markers across the cannabis genome. Finally, the thesis concluded by detailing a framework for how insights gathered from GWAS can be capitalized upon to develop gene edited cultivars.

In Chapter 2, I characterized phenotypic diversity across over 20,000 apples from over 1,000 apple accessions within Canada's Apple Biodiversity Collection (ABC). This phenotypic characterization involved measuring phenology and fruit quality traits both at harvest and after 3 months of cold storage. I found that trait measurements across apples within the ABC varies considerably, with apples displaying a 61-fold difference in weight, 18-fold difference in acidity, and a 100-fold difference in phenolic content. I found that 3 months of cold storage had significant impacts on fruit quality with apples losing 39% of their firmness, 31% of their acidity, and 9% of their weight, but gained 7% in soluble solids during the storage period. Firmness was positively correlated with

numerous phenological traits such as harvest date, flowering time and time to ripen which indicates that apple texture is impacted by developmental pathways that cause phenological changes throughout the growing season. In addition, I also gathered meta data information on each apple accession, such as species, year of release and end use. I found that phenolic content of apples has declined by 30% over the last 200 years of apple breeding apples, likely due to selection by breeders for non-browning flesh and lower tannin levels. The phenotypic characterization of the ABC highlighted germplasm with exceptional fruit quality and phenology phenotypes that could be of value in breeding programs to enable the introduction of novel traits into new cultivars. The phenotypic data collected within this chapter is foundational for the GWAS conducted in the subsequent chapter.

In Chapter 3, I aimed to identify genetic markers associated with fruit quality and phenology traits in the ABC. Using genetic data comprised of over 260,000 SNPs sequenced using GBS, I conducted GWAS for 21 fruit quality and phenology traits across 1,054 apple accessions. I identified two novel loci associated with phenolic content on chromosome 15 and 16. I identified a significant locus associated with softening on chromosome 10. The top SNP from the softening GWAS was located approximately 153 kb downstream of a well characterized candidate gene for softening, *PG1,* confirming that there is a large effect locus on chromosome 10 for softening. In addition, I determined that allelic variation at the *NAC18.1* transcription factor was associated with numerous traits including harvest date, firmness at harvest, and firmness after storage. Given our results I proposed a model for the allelic effects on variation at *NAC18.1* on

ripening related traits. The results indicate that *NAC18.1* acts a high-level regulator of multiple ripening traits by impacting the timing and rate at which apples move through the ripening process. In addition to the GWAS results, analysis of population structure within the ABC revealed that population structure is primarily shaped by harvest date and species. The results of this chapter broaden our understanding of the genetic control of multiple fruit quality and phenology traits in apple that can help inform breeding of improved apple cultivars.

In Chapter 4, I aimed to identify genetic markers associated with chemical content across cannabis samples. In addition, I also aimed to examine the chemical and genetic differences between Indica/Sativa labels that are commonly used to distinguish cannabis strains. I analyzed a genetic dataset comprised of over 100,000 SNPs that were genotyped across 100 cannabis samples using GBS and a chemical dataset of over 40 chemical compounds measured using GC-MS. Using the chemical and genetic data, I performed GWAS and identified two loci on chromosome 5 that were associated with myrcene content and a locus on chromosome 6 associated with three sesquiterpenes: guaiol, beta-eudesmol and gamma-eudesmol. By examining the genetic and chemical profiles of cannabis strains, I found that Indica/Sativa labelled samples were genetically and chemically indistinct. However, I identified significant correlations between terpenes with significant GWAS hits and cannabis labels. The results of this chapter indicate that cannabis labels are associated with variation in a small number of terpene concentrations that are driven by genetic variation at tandem arrays of terpene synthase genes. This work

was the first published GWAS conducted on drug-type cannabis and therefore lays the foundation for future genetic mapping studies in this important crop.

In Chapter 5, I proposed a framework for harnessing the results of GWAS to develop improved, gene edited apple cultivars. The framework consists of conducting high-powered GWAS to identify a short list of causal alleles responsible for phenotypic variation and then immediately targeting the short list of genes using gene editing. I outlined the challenges associated with a GWAS to CRISPR approach, which mainly involves issues with tissue culture and the current capabilities of gene editing technologies in apple. I argued that despite the present challenges associated with gene editing in apple, there are numerous benefits that this approach has over traditional genetic modification. These benefits include the ability to perform transgene-free editing, introduce novel variation, edit a gene target with high specificity and target multiple genes with relatively ease. The framework I proposed involves pairing high-resolution genetic mapping with modern gene editing tools to develop improved apple cultivars more rapidly.

In conclusion, the phenotypic and genetic data analyzed here, across apple and cannabis, furthers our basic understanding of variation across these important agricultural crops. Given the widespread importance of cannabis and apple cultivation it's crucial that these crops can capitalize upon the advancements being made in genomics to develop new improved cultivars. Here, I demonstrated that GWAS is a powerful method to identify loci of large affect controlling key traits in both apple and cannabis. The research

presented here contributes to our understanding of the genetic control of key quality traits in both apple and cannabis, which is necessary to develop markers for MAS and elucidate causal genetic variants that can be targeted with gene editing. While cannabis and apple each face unique challenges in their cultivation and breeding, the research presented here provides a solid foundation for the genomics-assisted improvement of new cultivars in both crops.

**Future directions**

The results of this thesis confirm previously identified phenotype-genotype associations and uncover novel associations with key traits both in apple and cannabis. However, from both GWAS with apple and cannabis presented within this thesis, the genetic resolution is limited due to the relatively sparse genetic data collected using GBS. Given the decreasing costs of sequencing, future genetic mapping work in apple and cannabis should undertake comprehensive sequencing, such as whole-genome sequencing, to improve the ability to develop superior markers for MAS and detect causal genetic variants for gene editing. Improved genetic resolution is necessary to further dissect the genetic control of key traits in apple and cannabis and capitalize on the power of genomics-assisted breeding.

Within this thesis a large amount of genotypic and phenotypic data was collected for the purpose of undertaking GWAS. However, both the cannabis and apple datasets each revealed additional insights into the biology of these crops. For example, one of the key findings from the analysis of the apple phenotype data in Chapter 2 was the significant

decline in phenolic content over the last 200 years of apple breeding. Given the benefits that phenolics provide to human health, future work should examine the complex trade-offs that exist between key traits to minimize unintended effects during breeding. The findings from Chapter 4 reveal important insights regarding strain labelling practices currently employed throughout the cannabis industry. We found that Indica labelled samples were largely indistinguishable from Sativa labelled samples based on their overall genetics and chemistry. This insight is valuable as it helps to clarify the labelling uncertainty which has been highlighted as major challenge for cannabis breeding and the industry as a whole (Barcaccia et al. 2020). Therefore, the value of these datasets to uncover additional insights between genotype-phenotype associations cannot be understated. As publicly available datasets, the genotype and phenotype data generated from this thesis can be continually leveraged to advance our understanding of apple and cannabis biology.

The focus for many years of genomics assisted breeding has been the development of genetic markers linked to desirable phenotypes that can be selected for using MAS. With the advent of gene editing technology, is it now possible to extend the genomics-assisted toolbox to include technologies that allow us to make precise change to the genomes of plants. Within this thesis, I proposed an approach for identifying short lists of gene editing targets from GWAS to rapidly develop new apple cultivars with desirable phenotypes. Although I focused on describing this framework for apple, it also is a potentially useful approach for enabling the genetic improvement of cannabis via gene editing and arguably could be more rapidly applied given the shorter live cycle of

cannabis. As outlined in chapter 5, challenges remain for the routine application of gene editing in apple that can deliver improved apple cultivars to producers and consumers. Given that numerous high-quality apple cultivars are currently available on the market, the role of gene editing in apple is likely not to "re-invent the wheel" but rather to introduce a small number of gene edits to generate variation in select traits across existing elite germplasm. In conclusion, the application of gene editing technology for plant improvement should focus on "piggy-backing" on what centuries of traditional breeding has already achieved in order to improve existing cultivars.

In conclusion, to fully realize the potential of genomics assisted breeding it is necessary to continue the advancement of multiple genomic approaches. There is value in the continued development of markers for MAS in conjunction with the identification of causal genetic variants for gene editing. Given the multitude of breeding challenges, the future of apple and cannabis breeding will benefit from a multi-pronged approach that leverages multiple genomics tools and approaches to generate improved cultivars.

## References


Aardema, Matthew L., and Rob DeSalle. 2021. “Can Public Online Databases Serve as a Source of Phenotypic Information for Cannabis Genetic Association Studies?” *PloS One* 16 (2): e0247607. https://doi.org/10.1371/journal.pone.0247607.

Aires, Alfredo. 2017. “Phenolics in Foods: Extraction, Analysis and Measurements.” In *Phenolic Compounds*, edited by Marcos Soto-Hernandez, Mariana Palma-Tenango, and Maria del Rosario Garcia-Mateos. IntechOpen. https://www.intechopen.com/chapters/53539.

Aldwinckle, and Malnoy. n.d. “Plant Regeneration and Transformation in the Rosaceae.” *Transgenic Plant Journal*. https://openpub.fmach.it/handle/10449/20031.

Altpeter, Fredy, Nathan M. Springer, Laura E. Bartley, Ann E. Blechl, Thomas P. Brutnell, Vitaly Citovsky, Liza J. Conrad, et al. 2016. “Advancing Crop Transformation in the Era of Genome Editing.” *The Plant Cell* 28 (7): 1510–20. https://doi.org/10.1105/tpc.16.00196.

Amiot, M. J., M. Tacchini, S. Aubert, and J. Nicolas. 1992. “Phenolic Composition and Browning Susceptibility of Various Apple Cultivars at Maturity.” *Journal of Food Science* 57 (4): 958–62. https://doi.org/10.1111/j.1365-2621.1992.tb14333.x.

Amyotte, Beatrice, Amy J. Bowen, Travis Banks, Istvan Rajcan, and Daryl J. Somers. 2017. “Mapping the Sensory Perception of Apple Using Descriptive Sensory Evaluation in a Genome Wide Association Study.” *PloS One* 12 (2): e0171710. https://doi.org/10.1371/journal.pone.0171710.

Anzalone, Andrew V., Xin D. Gao, Christopher J. Podracky, Andrew T. Nelson, Luke W. Koblan, Aditya Raguram, Jonathan M. Levy, Jaron A. M. Mercer, and David R. Liu. 2021. “Programmable Deletion, Replacement, Integration and Inversion of Large DNA Sequences with Twin Prime Editing.” *Nature Biotechnology*, December. https://doi.org/10.1038/s41587-021-01133-w.

Anzalone, Andrew V., Peyton B. Randolph, Jessie R. Davis, Alexander A. Sousa, Luke W. Koblan, Jonathan M. Levy, Peter J. Chen, et al. 2019. “Search-and-Replace Genome Editing without Double-Strand Breaks or Donor DNA.” *Nature* 576 (7785): 149–57. https://doi.org/10.1038/s41586-019-1711-4.

Arabidopsis Genome Initiative. 2000. “Analysis of the Genome Sequence of the Flowering Plant Arabidopsis Thaliana.” *Nature* 408 (6814): 796–815. https://doi.org/10.1038/35048692.

Armstrong, John, and Willia Lane. 2014. Genetically modified reduced-browning fruit-producing plant and produced fruit thereof, and method of obtaining such. USPTO 2014/0041079 A1. *US Patent*, filed October 9, 2013, and issued February 6, 2014.

Atkinson, Ross G., Paul W. Sutherland, Sarah L. Johnston, Kularajathevan Gunaseelan, Ian C. Hallett, Deepali Mitra, David A. Brummell, Roswitha Schröder, Jason W. Johnston, and Robert J. Schaffer. 2012. "Down-Regulation of POLYGALACTURONASE1 Alters Firmness, Tensile Strength and Water Loss in Apple (Malus x Domestica) Fruit." *BMC Plant Biology* 12 (August): 129. https://doi.org/10.1186/1471-2229-12-129.

Bai, Yang, Laura Dougherty, Mingjun Li, Gennaro Fazio, Lailiang Cheng, and Kenong Xu. 2012. "A Natural Mutation-Led Truncation in One of the Two Aluminum-Activated Malate Transporter-like Genes at the Ma Locus Is Associated with Low Fruit Acidity in Apple." *Molecular Genetics and Genomics: MGG* 287 (8): 663–78. https://doi.org/10.1007/s00438-012-0707-7.

Bakel, Harm van, Jake M. Stout, Atina G. Cote, Carling M. Tallon, Andrew G. Sharpe, Timothy R. Hughes, and Jonathan E. Page. 2011. "The Draft Genome and Transcriptome of Cannabis Sativa." *Genome Biology* 12 (10): R102. https://doi.org/10.1186/gb-2011-12-10-r102.

Barcaccia, Gianni, Fabio Palumbo, Francesco Scariolo, Alessandro Vannozzi, Marcello Borin, and Stefano Bona. 2020. "Potentials and Challenges of Genomics for Breeding Cannabis Cultivars." *Frontiers in Plant Science* 11 (September): 573299. https://doi.org/10.3389/fpls.2020.573299.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using Lme4." *Journal of Statistical Software* 67 (October): 1–48. https://doi.org/10.18637/jss.v067.i01.

Baumgartner, Isabelle O., Markus Kellerhals, Fabrizio Costa, Luca Dondini, Giulia Pagliarani, Roberto Gregori, Stefano Tartarini, Lucie Leumann, François Laurens, and Andrea Patocchi. 2016. "Development of SNP-Based Assays for Disease Resistance and Fruit Quality Traits in Apple (Malus × Domestica Borkh.) and Validation in Breeding Pilot Studies." *Tree Genetics & Genomes* 12 (3). https://doi.org/10.1007/s11295-016-0994-y.

Ben Sadok, Inès, Aline Tiecher, Didiana Galvez-Lopez, Marc Lahaye, Pauline Lasserre-Zuber, Maryline Bruneau, Sylvain Hanteville, Roland Robic, Raphael Cournol, and François Laurens. 2015. "Apple Fruit Texture QTLs: Year and Cold Storage Effects on Sensory and Instrumental Traits." *Tree Genetics & Genomes* 11 (6): 119. https://doi.org/10.1007/s11295-015-0947-x.

Benzie, I. F., and J. J. Strain. 1996. "The Ferric Reducing Ability of Plasma (FRAP) as a Measure of 'Antioxidant Power': The FRAP Assay." *Analytical Biochemistry* 239 (1): 70–76. https://doi.org/10.1006/abio.1996.0292.

Birch, R. G. 1997. “PLANT TRANSFORMATION: Problems and Strategies for Practical Application.” *Annual Review of Plant Physiology and Plant Molecular Biology* 48 (June): 297–326. https://doi.org/10.1146/annurev.arplant.48.1.297.

Blanpied, G. D., and Kenneth J. Silsby. 1992. “Predicting Harvest Date Windows for Apples.” Cornell University Library. July 1992. https://ecommons.cornell.edu/handle/1813/3299.

Bonini, Sara Anna, Marika Premoli, Simone Tambaro, Amit Kumar, Giuseppina Maccarinelli, Maurizio Memo, and Andrea Mastinu. 2018. “Cannabis Sativa: A Comprehensive Ethnopharmacological Review of a Medicinal Plant with a Long History.” *Journal of Ethnopharmacology* 227 (December): 300–315. https://doi.org/10.1016/j.jep.2018.09.004.

Booth, Judith K., and Jorg Bohlmann. 2019. “Terpenes in Cannabis Sativa – From Plant Genome to Humans.” *Plant Science: An International Journal of Experimental Plant Biology* 284 (March): 67—72. https://doi.org/10.1016/j.plantsci.2019.03.022.

Booth, Judith K., Jonathan E. Page, and Jörg Bohlmann. 2017. “Terpene Synthases from Cannabis Sativa.” *PloS One* 12 (3): e0173911. https://doi.org/10.1371/journal.pone.0173911.

Borin, Marcello, Fabio Palumbo, Alessandro Vannozzi, Francesco Scariolo, Gio Batta Sacilotto, Marco Gazzola, and Gianni Barcaccia. 2021. “Developing and Testing Molecular Markers in Cannabis Sativa (Hemp) for Their Use in Variety and Dioecy Assessments.” *Plants* 10 (10). https://doi.org/10.3390/plants10102174.

Bradbury, Peter J., Zhiwu Zhang, Dallas E. Kroon, Terry M. Casstevens, Yogesh Ramdoss, and Edward S. Buckler. 2007. “TASSEL: Software for Association Mapping of Complex Traits in Diverse Samples.” *Bioinformatics* 23 (19): 2633–35. https://doi.org/10.1093/bioinformatics/btm308.

Bramel, Paula, and Gayle M. Volk. 2019. “A Global Strategy for the Conservation and Use of Apple Genetic Resources.” https://doi.org/10.13140/RG.2.2.34072.34562.

Broekema, R. V., O. B. Bakker, and I. H. Jonkers. 2020. “A Practical View of Fine-Mapping and Gene Prioritization in the Post-Genome-Wide Association Era.” *Open Biology* 10 (1): 190221. https://doi.org/10.1098/rsob.190221.

Brookfield, Paul, Peter Murphy, Roger Harker, and Elspeth MacRae. 1997. “Starch Degradation and Starch Pattern Indices; Interpretation and Relationship to Maturity.” *Postharvest Biology and Technology* 11 (1): 23–30. https://doi.org/10.1016/S0925-5214(97)01416-6.

Brown, S., and K. Maloney. 2003. “Genetic Improvement of Apple: Breeding, Markers, Mapping and Biotechnology.” In *Apples: Botany, Production and Uses*, edited by D. Ferree and I. Warrington, 32–43. Cambridge: CABI Publishing.

Bush, William S., and Jason H. Moore. 2012. “Chapter 11: Genome-Wide Association Studies.” *PLoS Computational Biology* 8 (12): e1002822. https://doi.org/10.1371/journal.pcbi.1002822.

Casano, S., G. Grassi, V. Martini, and M. Michelozzi. 2011. “Variations in Terpene Profiles of Different Strains of Cannabis Sativa L.” *Acta Horticulturae* 925 (925): 115–21. https://doi.org/10.17660/ActaHortic.2011.925.15.

Chagné, David, Daya Dayatilake, Robert Diack, Murray Oliver, Hilary Ireland, Amy Watson, Susan E. Gardiner, Jason W. Johnston, Robert J. Schaffer, and Stuart Tustin. 2014. “Genetic and Environmental Control of Fruit Maturation, Dry Matter and Firmness in Apple (Malus × Domestica Borkh.).” *Horticulture Research* 1 (September): 14046. https://doi.org/10.1038/hortres.2014.46.

Chagné, David, Chris Kirk, Natalie How, Claire Whitworth, and Ignasi Iglesias. 2016. “A Functional Genetic Marker for Apple Red Skin Coloration across Different Environments.” *Tree Genetics & Genomes* 12 (4). https://doi.org/10.1007/s11295-016-1025-8.

Chagné, David, Célia Krieger, Maysoon Rassam, Mike Sullivan, Jenny Fraser, Christelle André, Massimo Pindo, et al. 2012. “QTL and Candidate Gene Mapping for Polyphenolic Composition in Apple Fruit.” *BMC Plant Biology* 12 (1): 12. https://doi.org/10.1186/1471-2229-12-12.

Chagné, David, Stijn Vanderzande, Chris Kirk, Natalie Profitt, Rosemary Weskett, Susan E. Gardiner, Cameron P. Peace, Richard K. Volz, and Nahla V. Bassil. 2019. “Validation of SNP Markers for Fruit Quality and Disease Resistance Loci in Apple (Malus × Domestica Borkh.) Using the OpenArray® Platform.” *Horticulture Research* 6 (March): 30. https://doi.org/10.1038/s41438-018-0114-2.

Chang, Lijing, Sheng Wu, and Li Tian. 2019. “Effective Genome Editing and Identification of a Regiospecific Gallic Acid 4-O-Glycosyltransferase in Pomegranate (Punica Granatum L.).” *Horticulture Research* 6 (November): 123. https://doi.org/10.1038/s41438-019-0206-7.

Charrier, Aurélie, Emilie Vergne, Nicolas Dousset, Andréa Richer, Aurélien Petiteau, and Elisabeth Chevreau. 2019. “Efficient Targeted Mutagenesis in Apple and First Time Edition of Pear Using the CRISPR-Cas9 System.” *Frontiers in Plant Science* 10: 40. https://doi.org/10.3389/fpls.2019.00040.

Chen, Jiajing, Sumathi Tomes, Andrew P. Gleave, Wendy Hall, Zhiwei Luo, Juan Xu, and Jia-Long Yao. 2022. “Significant Improvement of Apple (Malus Domestica Borkh.) Transgenic Plant Production by Pre-Transformation with a Baby Boom Transcription Factor.” *Horticulture Research*, January. https://doi.org/10.1093/hr/uhab014.

Chen, Kunling, Yanpeng Wang, Rui Zhang, Huawei Zhang, and Caixia Gao. 2019. "CRISPR/Cas Genome Editing and Precision Plant Breeding in Agriculture." *Annual Review of Plant Biology* 70 (April): 667–97. https://doi.org/10.1146/annurev-arplant-050718-100049.

"Compound Card for α-Trans-Bergamotene." 2020. Cannabis Database. March 19, 2020. https://cannabisdatabase.ca/compounds/CDB000306.

Cong, Le, F. Ann Ran, David Cox, Shuailiang Lin, Robert Barretto, Naomi Habib, Patrick D. Hsu, et al. 2013. "Multiplex Genome Engineering Using CRISPR/Cas Systems." *Science* 339 (6121): 819–23. https://doi.org/10.1126/science.1231143.

Cornille, Amandine, Tatiana Giraud, Marinus J. M. Smulders, Isabel Roldán-Ruiz, and Pierre Gladieux. 2014. "The Domestication and Evolutionary Ecology of Apples." *Trends in Genetics: TIG* 30 (2): 57–65. https://doi.org/10.1016/j.tig.2013.10.002.

Cornille, Amandine, Pierre Gladieux, Marinus J. M. Smulders, Isabel Roldán-Ruiz, François Laurens, Bruno Le Cam, Anush Nersesyan, et al. 2012. "New Insight into the History of Domesticated Apple: Secondary Contribution of the European Wild Apple to the Genome of Cultivated Varieties." *PLoS Genetics* 8 (5): e1002703. https://doi.org/10.1371/journal.pgen.1002703.

Costa, Fabrizio, Luca Cappellin, Elena Zini, Andrea Patocchi, Markus Kellerhals, Matteo Komjanc, Cesare Gessler, and Franco Biasioli. 2013. "QTL Validation and Stability for Volatile Organic Compounds (VOCs) in Apple." *Plant Science: An International Journal of Experimental Plant Biology* 211 (October): 1–7. https://doi.org/10.1016/j.plantsci.2013.05.018.

Costa, Fabrizio, Cameron P. Peace, Sara Stella, Sara Serra, Stefano Musacchi, Micaela Bazzani, Silviero Sansavini, and W. Eric Van de Weg. 2010. "QTL Dynamics for Fruit Firmness and Softening around an Ethylene-Dependent Polygalacturonase Gene in Apple (Malus x Domestica Borkh.)." *Journal of Experimental Botany* 61 (11): 3029–39. https://doi.org/10.1093/jxb/erq130.

Costa, Fabrizio, Sara Stella, W. Eric Van de Weg, Walter Guerra, Michela Cecchinel, Joseph Dallavia, Bernie Koller, and Silviero Sansavini. 2005. "Role of the Genes Md-ACO1 and Md-ACS1 in Ethylene Production and Shelf Life of Apple (Malus Domestica Borkh)." *Euphytica/ Netherlands Journal of Plant Breeding* 141 (1): 181–90. https://doi.org/10.1007/s10681-005-6805-4.

D. Turner, Stephen. 2018. "Qqman: An R Package for Visualizing GWAS Results Using Q-Q and Manhattan Plots." *Journal of Open Source Software* 3 (25): 731. https://doi.org/10.21105/joss.00731.

Daccord, Nicolas, Jean-Marc Celton, Gareth Linsmith, Claude Becker, Nathalie Choisne, Elio Schijlen, Henri van de Geest, et al. 2017. "High-Quality de Novo Assembly of the Apple Genome and Methylome Dynamics of Early Fruit Development." *Nature Genetics* 49 (7): 1099–1106. https://doi.org/10.1038/ng.3886.

Dalla Costa, Lorenza, Stefano Piazza, Valerio Pompili, Umberto Salvagnin, Alessandro Cestaro, Loredana Moffa, Lorenzo Vittani, Claudio Moser, and Mickael Malnoy. 2020. “Strategies to Produce T-DNA Free CRISPRed Fruit Trees via Agrobacterium Tumefaciens Stable Gene Transfer.” *Scientific Reports* 10 (1): 20155. https://doi.org/10.1038/s41598-020-77110-1.

Danecek, Petr, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo, H, et al. 2011. “The Variant Call Format and VCFtools.” *Bioinformatics* 27 (15): 2156—2158. https://doi.org/10.1093/bioinformatics/btr330.

Davies, Sophie Watts, Kendra McClure, Zoë Migicovsky, and Sean Myles. 2022. “Phenotypic Divergence between the Cultivated Apple (Malus Domestica) and Its Primary Wild Progenitor (Malus Sieversii).” *PloS One* 17 (3): e0250751. https://doi.org/10.1371/journal.pone.0250751.

Debernardi, Juan M., David M. Tricoli, Maria F. Ercoli, Sadiye Hayta, Pamela Ronald, Javier F. Palatnik, and Jorge Dubcovsky. 2020. “A GRF-GIF Chimeric Protein Improves the Regeneration Efficiency of Transgenic Plants.” *Nature Biotechnology* 38 (11): 1274–79. https://doi.org/10.1038/s41587-020-0703-0.

Di Guardo, Mario, Marco C. A. M. Bink, Walter Guerra, Thomas Letschka, Lidia Lozano, Nicola Busatto, Lara Poles, et al. 2017. “Deciphering the Genetic Control of Fruit Texture in Apple by Multiple Family-Based Analysis and Genome-Wide Association.” *Journal of Experimental Botany* 68 (7): 1451–66. https://doi.org/10.1093/jxb/erx017.

Doudna, Jennifer A., and Emmanuelle Charpentier. 2014. “The New Frontier of Genome Engineering with CRISPR-Cas9.” *Science* 346 (6213): 1258096. https://doi.org/10.1126/science.1258096.

Duan, Naibin, Yang Bai, Honghe Sun, Nan Wang, Yumin Ma, Mingjun Li, Xin Wang, et al. 2017. “Genome Re-Sequencing Reveals the History of Apple and Supports a Two-Stage Model for Fruit Enlargement.” *Nature Communications* 8 (1): 249. https://doi.org/10.1038/s41467-017-00336-7.

Edge-Garza, D. A., C. P. Peace, and Yanmin Zhu. 2010. “Enabling Marker-Assisted Seedling Selection in the Washington Apple Breeding Program.” *Acta Horticulturae* 859 (December 2014): 369–74. https://doi.org/10.17660/ActaHortic.2010.859.44.

Elshire, Robert J., Jeffrey C. Glaubitz, Qi Sun, Pol, Jesse A., Ken Kawamoto, Edward S. Buckler, and Sharon E. Mitchell. 2011. “A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species.” *PloS One* 6 (5): 1—10. https://doi.org/10.1371/journal.pone.0019379.

Elzinga, Sytze, Justin Fischedick, Richard Podkolinski, and Jeffrey C. Raber. 2015. "Cannabinoids and Terpenes as Chemotaxonomic Markers in Cannabis." *Nat. Prod. Chem. Res* 3 (81): 10–4172. https://www.academia.edu/download/51273986/Natural_Products_Chemistry__Research20170109-28127-sgi414.pdf.

Endo, Masaki, Masafumi Mikami, Akira Endo, Hidetaka Kaya, Takeshi Itoh, Hiroshi Nishimasu, Osamu Nureki, and Seiichi Toki. 2019. "Genome Editing in Plants by Engineered CRISPR-Cas9 Recognizing NG PAM." *Nature Plants* 5 (1): 14–17. https://doi.org/10.1038/s41477-018-0321-8.

Entine, Jon, Maria Sueli S. Felipe, Jan-Hendrik Groenewald, Drew L. Kershen, Martin Lema, Alan McHughen, Alexandre Lima Nepomuceno, et al. 2021. "Regulatory Approaches for Genome Edited Agricultural Plants in Select Countries and Jurisdictions around the World." *Transgenic Research* 30 (4): 551–84. https://doi.org/10.1007/s11248-021-00257-8.

Espley, Richard V., Arnaud Bovy, Christina Bava, Sara R. Jaeger, Sumathi Tomes, Cara Norling, Jonathan Crawford, et al. 2013. "Analysis of Genetically Modified Red-fleshed Apples Reveals Effects on Growth and Consumer Attributes." *Plant Biotechnology Journal* 11 (4): 408–19. https://doi.org/10.1111/pbi.12017.

Espley, Richard V., Davin Leif, Blue Plunkett, Tony McGhie, Rebecca Henry-Kirk, Miriam Hall, Jason W. Johnston, et al. 2019. "Red to Brown: An Elevated Anthocyanic Response in Apple Drives Ethylene to Advance Maturity and Fruit Flesh Browning." *Frontiers in Plant Science* 10: 1248. https://doi.org/10.3389/fpls.2019.01248.

FAOSTAT. 2021. "Food and Agriculture Organization of the United Nations." 2021. http://www.fao.org/faostat/en.

Fike, John. 2016. "Industrial Hemp: Renewed Opportunities for an Ancient Crop." *Critical Reviews in Plant Sciences* 35 (5–6): 406–24. https://doi.org/10.1080/07352689.2016.1257842.

Fischedick, Justin Thomas, Arno Hazekamp, Tjalling Erkelens, Young Hae Choi, and Rob Verpoorte. 2010. "Metabolic Fingerprinting of Cannabis Sativa L., Cannabinoids and Terpenoids for Chemotaxonomic and Drug Standardization Purposes." *Phytochemistry* 71 (17–18): 2058–73. https://doi.org/10.1016/j.phytochem.2010.10.001.

Fister, Andrew S., Lena Landherr, Siela N. Maximova, and Mark J. Guiltinan. 2018. "Transient Expression of CRISPR/Cas9 Machinery Targeting TcNPR3 Enhances Defense Response in Theobroma Cacao." *Frontiers in Plant Science* 9 (March): 268. https://doi.org/10.3389/fpls.2018.00268.

Flint-Garcia, Sherry A., Anne-céline Thuillet, Jianming Yu, Gael Pressoir, Susan M. Romero, Sharon E. Mitchell, John Doebley, Stephen Kresovich, Major M. Goodman, and Edward S. Buckler. 2005. "Maize Association Population: A High-resolution Platform for Quantitative Trait Locus Dissection." *The Plant Journal: For Cell and Molecular Biology* 44 (6): 1054–64. https://doi.org/10.1111/j.1365-313x.2005.02591.x.

Forlani, Sara, Chiara Mizzotti, and Simona Masiero. 2021. "The NAC Side of the Fruit: Tuning of Fruit Development and Maturation." *BMC Plant Biology* 21 (1): 238. https://doi.org/10.1186/s12870-021-03029-y.

Frary, A., T. C. Nesbitt, S. Grandillo, E. Knaap, B. Cong, J. Liu, J. Meller, R. Elber, K. B. Alpert, and S. D. Tanksley. 2000. "Fw2.2: A Quantitative Trait Locus Key to the Evolution of Tomato Fruit Size." *Science* 289 (5476): 85–88. https://doi.org/10.1126/science.289.5476.85.

Furbank, Robert T., and Mark Tester. 2011. "Phenomics – Technologies to Relieve the Phenotyping Bottleneck." *Trends in Plant Science* 16 (12): 635–44. https://doi.org/10.1016/j.tplants.2011.09.005.

Gallois, Jean-Luc, Claire Woodward, G. Venugopala Reddy, and Robert Sablowski. 2002. "Combined SHOOT MERISTEMLESS and WUSCHEL Trigger Ectopic Organogenesis in Arabidopsis." *Development* 129 (13): 3207–17. https://doi.org/10.1242/dev.129.13.3207.

Gao, Caixia. 2021. "Genome Engineering for Crop Improvement and Future Agriculture." *Cell* 184 (6): 1621–35. https://doi.org/10.1016/j.cell.2021.01.005.

Gao, Shan, Baishi Wang, Shanshan Xie, Xiaoyu Xu, Jin Zhang, Li Pei, Yongyi Yu, Weifei Yang, and Ying Zhang. 2020. "A High-Quality Reference Genome of Wild Cannabis Sativa." *Horticulture Research* 7 (1): 73. https://doi.org/10.1038/s41438-020-0295-3.

Gao, Ying, Wei Wei, Xiaodan Zhao, Xiaoli Tan, Zhongqi Fan, Yiping Zhang, Yuan Jing, et al. 2018. "A NAC Transcription Factor, NOR-Like1, Is a New Positive Regulator of Tomato Fruit Ripening." *Horticulture Research* 5 (1): 75. https://doi.org/10.1038/s41438-018-0111-5.

Gao, Yuan, Fengzhi Liu, Kun Wang, Dajiang Wang, Xin Gong, Lijun Liu, Christopher M. Richards, Adam D. Henk, and Gayle M. Volk. 2015. "Genetic Diversity of Malus Cultivars and Wild Relatives in the Chinese National Repository of Apple Germplasm Resources." *Tree Genetics & Genomes* 11 (5): 106. https://doi.org/10.1007/s11295-015-0913-7.

García-Gómez, Beatriz Ester, Juan Alfonso Salazar, Luca Dondini, Pedro Martínez-Gómez, and David Ruiz. 2019. "Identification of QTLs Linked to Fruit Quality Traits in Apricot (Prunus Armeniaca L.) and Biological Validation through Gene Expression Analysis Using QPCR." *Molecular Breeding: New Strategies in Plant Improvement* 39 (2): 28. https://doi.org/10.1007/s11032-018-0926-7.

Gilbert, Avery N., and Joseph A. DiVerdi. 2018. "Consumer Perceptions of Strain Differences in Cannabis Aroma." *PloS One* 13 (2): e0192247. https://doi.org/10.1371/journal.pone.0192247.

Giné-Bordonaba, Jordi, Iban Eduardo, Pere Arús, and Celia M. Cantín. 2020. "Biochemical and Genetic Implications of the Slow Ripening Phenotype in Peach Fruit." *Scientia Horticulturae* 259 (January): 108824. https://doi.org/10.1016/j.scienta.2019.108824.

Goralogia, Greg S., Thomas P. Redick, and Steven H. Strauss. 2021. "Gene Editing in Tree and Clonal Crops: Progress and Challenges." *In Vitro Cellular & Developmental Biology - Plant* 57 (4): 683–99. https://doi.org/10.1007/s11627-021-10197-x.

Gottschalk, Chris, and Steve van Nocker. 2013. "Diversity in Seasonal Bloom Time and Floral Development among Apple Species and Hybrids." *Journal of the American Society for Horticultural Science. American Society for Horticultural Science* 138 (5): 367–74. https://doi.org/10.21273/JASHS.138.5.367.

Grassa, Christopher J., George D. Weiblen, Jonathan P. Wenger, Clemon Dabney, Shane G. Poplawski, S. Timothy Motley, Todd P. Michael, and C. J. Schwartz. 2021. "A New Cannabis Genome Assembly Associates Elevated Cannabidiol (CBD) with Hemp Introgressed into Marijuana." *The New Phytologist* 230 (4): 1665–79. https://doi.org/10.1111/nph.17243.

Grassa, Christopher J., Jonathan P. Wenger, Clemon Dabney, Shane G. Poplawski, S. Timothy Motley, Todd P. Michael, C. J. Schwartz, and George D. Weiblen. 2018. "A Complete Cannabis Chromosome Assembly and Adaptive Admixture for Elevated Cannabidiol (CBD) Content." *BioRxiv*, 458083. https://doi.org/10.1101/458083.

Gross, Briana L., Adam D. Henk, Christopher M. Richards, Gennaro Fazio, and Gayle M. Volk. 2014. "Genetic Diversity in Malus ×domestica (Rosaceae) through Time in Response to Domestication." *American Journal of Botany* 101 (10): 1770–79. https://doi.org/10.3732/ajb.1400297.

Gross, Briana L., Gayle M. Volk, Christopher M. Richards, Patrick A. Reeves, Adam D. Henk, Philip L. Forsline, Amy Szewc-McFadden, Gennaro Fazio, and C. Thomas Chao. 2013. "Diversity Captured in the USDA-ARS National Plant Germplasm System Apple Core Collection." *Journal of the American Society for Horticultural Science. American Society for Horticultural Science* 138 (5): 375–81. https://doi.org/10.21273/jashs.138.5.375.

Gu, Xiaoyong, Lijing Liu, and Huawei Zhang. 2021. “Transgene-Free Genome Editing in Plants.” *Frontiers in Genome Editing* 3: 36. https://doi.org/10.3389/fgeed.2021.805317.

Guan, Yingzhu, Cameron Peace, David Rudell, Sujeet Verma, and Kate Evans. 2015. “QTLs Detected for Individual Sugars and Soluble Solids Content in Apple.” *Molecular Breeding: New Strategies in Plant Improvement* 35 (6): 135. https://doi.org/10.1007/s11032-015-0334-1.

Guerra, Marcos, Miguel A. Sanz, and Pedro A. Casquero. 2010. “Influence of Storage Conditions on the Sensory Quality of a High Acid Apple.” *International Journal of Food Science & Technology* 45 (11): 2352–57. https://doi.org/10.1111/j.1365-2621.2010.02410.x.

Günnewich, Nils, Jonathan E. Page, Tobias G. Köllner, Jörg Degenhardt, and Toni M. Kutchan. 2007. “Functional Expression and Characterization of Trichome-Specific (-)-Limonene Synthase and (+)-α-Pinene Synthase from Cannabis Sativa.” *Natural Product Communications* 2 (3): 1934578X0700200301. https://doi.org/10.1177/1934578X0700200301.

Guy, Geoffrey W., and John M. McPartland. 2017. “Models of Cannabis Taxonomy, Cultural Bias, and Conflicts between Scientific and Vernacular Names.” *The Botanical Review; Interpreting Botanical Progress*, no. 1859: 1–55. https://doi.org/10.1007/s12229-017-9187-0.

Hall, David, Carolina Tegström, and Pär K. Ingvarsson. 2010. “Using Association Mapping to Dissect the Genetic Basis of Complex Traits in Plants.” *Briefings in Functional Genomics* 9 (2): 157–65. https://doi.org/10.1093/bfgp/elp048.

Hampson, Cheryl R., R. McNew, J. Cline, C. Embree, and K. Wilson. 2009. “Regional Differences in Performance of Canadian-Bred Apple Cultivars and Implications for Breeding.” *Canadian Journal of Plant Science. Revue Canadienne de Phytotechnie* 89 (1): 81–91. https://doi.org/10.4141/CJPS08005.

Hao, Lijun, Xiangdong Pu, and Jingyuan Song. 2021. “Introduction of Mutations in Plants with Prime Editing.” *Methods* 194 (October): 83–93. https://doi.org/10.1016/j.ymeth.2021.03.014.

Harker, F. Roger, F. Anne Gunson, and Sara R. Jaeger. 2003. “The Case for Fruit Quality: An Interpretive Review of Consumer Attitudes, and Preferences for Apples.” *Postharvest Biology and Technology* 28 (3): 333–47. https://doi.org/10.1016/S0925-5214(02)00215-6.

Hartsel, Joshua A., Joshua Eades, Brian Hickory, and Alexandros Makriyannis. 2016. “Chapter 53 - Cannabis Sativa and Hemp.” In *Nutraceuticals*, edited by Ramesh C. Gupta, 735–54. Boston: Academic Press. https://doi.org/10.1016/B978-0-12-802147-7.00053-X.

Hassani-Pak, Keywan, and Christopher Rawlings. 2017. “Knowledge Discovery in Biological Databases for Revealing Candidate Genes Linked to Complex Phenotypes.” *Journal of Integrative Bioinformatics* 14 (1). https://doi.org/10.1515/jib-2016-0002.

Hazekamp, Amo, Katerina Tekalova, and Stelios Papadimitriou. 2016. “Cannabis : From Cultivar to Chemovar II - A Metabolomics Approach to Cannabis Classification.” *Cannabis and Cannabinoid Research* 1 (1): 1–15. https://doi.org/10.1089/can.2016.0017.

Henry, Philippe, Surender Khatodia, Karan Kapoor, Britni Gonzales, Alexis Middleton, Kevin Hong, Aaron Hilyard, et al. 2020. “A Single Nucleotide Polymorphism Assay Sheds Light on the Extent and Distribution of Genetic Diversity, Population Structure and Functional Basis of Key Traits in Cultivated North American Cannabis.” *Journal of Cannabis Research* 2 (1): 26. https://doi.org/10.1186/s42238-020-00036-y.

Hesami, Mohsen, Marco Pepe, Milad Alizadeh, Aida Rakei, Austin Baiton, and Andrew Maxwell Phineas Jones. 2020. “Recent Advances in Cannabis Biotechnology.” *Industrial Crops and Products* 158 (December): 113026. https://doi.org/10.1016/j.indcrop.2020.113026.

Hillig, Karl W. 2004. “A Chemotaxonomic Analysis of Terpenoid Variation in Cannabis.” *Biochemical Systematics and Ecology* 32 (10): 875–91. https://doi.org/10.1016/j.bse.2004.04.004.

———. 2005. “Genetic Evidence for Speciation in Cannabis (Cannabaceae).” *Genetic Resources and Crop Evolution* 52 (2): 161—180. https://doi.org/10.1007/s10722-003-4452-y.

Holderbaum, Daniel Ferreira, Tomoyuki Kon, Tsuyoshi Kudo, and Miguel Pedro Guerra. 2010. “Enzymatic Browning, Polyphenol Oxidase Activity, and Polyphenols in Four Apple Cultivars: Dynamics during Fruit Development.” *HortScience: A Publication of the American Society for Horticultural Science* 45 (8): 1150–54. https://doi.org/10.21273/HORTSCI.45.8.1150.

Hsu, Patrick D., Eric S. Lander, and Feng Zhang. 2014. “Development and Applications of CRISPR-Cas9 for Genome Engineering.” *Cell* 157 (6): 1262–78. https://doi.org/10.1016/j.cell.2014.05.010.

Hu, Zhenbin, Marcus O. Olatoye, Sandeep Marla, and Geoffrey P. Morris. 2019. “An Integrated Genotyping-by-Sequencing Polymorphism Map for Over 10,000 Sorghum Genotypes.” *The Plant Genome* 12 (1). https://doi.org/10.3835/plantgenome2018.06.0044.

Huber, G. M., and H. P. V. Rupasinghe. 2009. “Phenolic Profiles and Antioxidant Properties of Apple Skin Extracts.” *Journal of Food Science* 74 (9): C693-700. https://doi.org/10.1111/j.1750-3841.2009.01356.x.

Hurgobin, Bhavna, Muluneh Tamiru-Oli, Matthew T. Welling, Monika S. Doblin, Antony Bacic, James Whelan, and Mathew G. Lewsey. 2020. "Recent Advances in Cannabis Sativa Genomics Research." *The New Phytologist*, December. https://doi.org/10.1111/nph.17140.

Jackson, Michael T. 1997. "Conservation of Rice Genetic Resources: The Role of the International Rice Genebank at IRRI." *Plant Molecular Biology* 35 (1–2): 61–67. https://doi.org/10.1023/a:1005709332130.

Jänsch, Melanie, Giovanni A. L. Broggini, Juliane Weger, Vincent G. M. Bus, Susan E. Gardiner, Heather Bassett, and Andrea Patocchi. 2015. "Identification of SNPs Linked to Eight Apple Disease Resistance Loci." *Molecular Breeding: New Strategies in Plant Improvement* 35 (1): 45. https://doi.org/10.1007/s11032-015-0242-4.

Jia, Hongge, Vladimir Orbovic, Jeffrey B. Jones, and Nian Wang. 2016. "Modification of the PthA4 Effector Binding Elements in Type I CsLOB1 Promoter Using Cas9/SgRNA to Produce Transgenic Duncan Grapefruit Alleviating XccΔpthA4:DCsLOB1.3 Infection." *Plant Biotechnology Journal* 14 (5): 1291–1301. https://doi.org/10.1111/pbi.12495.

Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier. 2012. "A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity." *Science* 337 (6096): 816–21. https://doi.org/10.1126/science.1225829.

Johnston, Jason W., Kularajathaven Gunaseelan, Paul Pidakala, Mindy Wang, and Robert J. Schaffer. 2009. "Co-Ordination of Early and Late Ripening Events in Apples Is Regulated through Differential Sensitivities to Ethylene." *Journal of Experimental Botany* 60 (9): 2689–99. https://doi.org/10.1093/jxb/erp122.

Johnston, Jason W., Errol W. Hewett, Maartem L. A. T. M. Hertog, and F. Roger Harker. 2002. "Harvest Date and Fruit Size Affect Postharvest Softening of Apple Fruit." *The Journal of Horticultural Science & Biotechnology* 77 (3): 355–60. https://doi.org/10.1080/14620316.2002.11511505.

Johnston, Jason W., Errol W. Hewett, and Maarten L. A. T. M. Hertog. 2010. "Postharvest Softening of Apple (Malus Domestica) Fruit: A Review." *New Zealand Journal of Crop and Horticultural Science* 30 (3): 145–60. https://doi.org/10.1080/01140671.2002.9514210.

Jung, Michaela, Beat Keller, Morgane Roth, Maria José Aranzana, Annemarie Auwerkerken, Walter Guerra, Mehdi Al-Rifaï, et al. 2022. "Genetic Architecture and Genomic Predictive Ability of Apple Quantitative Traits across Environments." *Horticulture Research* 9 (February). https://doi.org/10.1093/hr/uhac028.

Jung, Michaela, Morgane Roth, Maria José Aranzana, Annemarie Auwerkerken, Marco Bink, Caroline Denancé, Christian Dujak, et al. 2020. “The Apple REFPOP—a Reference Population for Genomics-Assisted Breeding in Apple.” *Horticulture Research* 7 (1): 189. https://doi.org/10.1038/s41438-020-00408-8.

Kenis, Katrien, Johan Keulemans, and Mark W. Davey. 2008. “Identification and Stability of QTLs for Fruit Quality Traits in Apple.” *Tree Genetics & Genomes* 4 (4): 647–61. https://doi.org/10.1007/s11295-008-0140-6.

Khan, Ali A., and Julian F. V. Vincent. 1990. “Anisotropy of Apple Parenchyma.” *Journal of the Science of Food and Agriculture* 52 (4): 455–66. https://doi.org/10.1002/jsfa.2740520404.

Khan, Awais, Sarah B. Carey, Alicia Serrano, Huiting Zhang, Heidi Hargarten, Haley Hale, Alex Harkess, and Loren Honaas. 2022. “A Phased, Chromosome-Scale Genome of ‘Honeycrisp’ Apple (Malus Domestica).” *GigaByte (Hong Kong, China)* 2022 (September): gigabyte69. https://doi.org/10.46471/gigabyte.69.

Khan, Awais, and Schuyler Korban. 2012. “Association Mapping in Forest Trees and Fruit Crops.” *Journal of Experimental Botany* 63 (2): 695–709. https://doi.org/10.1093/jxb/err313.

Khan, M. Awais, and Schuyler S. Korban. 2012. “Association Mapping in Forest Trees and Fruit Crops.” *Journal of Experimental Botany* 63 (11): 4045–60. https://doi.org/10.1093/jxb/ers105.

Khan, Sabaz Ali, Pierre-Yves Chibon, Ric C. H. de Vos, Bert A. Schipper, Evert Walraven, Jules Beekwilder, Thijs van Dijk, et al. 2012. “Genetic Analysis of Metabolites in Apple Fruits Indicates an MQTL Hotspot for Phenolic Compounds on Linkage Group 16.” *Journal of Experimental Botany* 63 (8): 2895–2908. https://doi.org/10.1093/jxb/err464.

Klee, Harry J., and David G. Clark. 2010. “Ethylene Signal Transduction in Fruits and Flowers.” In *Plant Hormones: Biosynthesis, Signal Transduction, Action!*, edited by Peter J. Davies, 377–98. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-1-4020-2686-7_18.

Kluiver, Hilde de, Florence Vorspan, Emmanuel Curis, Eske M. Derks, Cynthia Marie-Claire, Sven Stringer, and Andries T. Marees. 2018. “A Tutorial on Conducting Genome-Wide Association Studies: Quality Control and Statistical Analysis.” *International Journal of Methods in Psychiatric Research* 27 (2): e1608. https://doi.org/10.1002/mpr.1608.

Koltai, Hinanit, and Dvora Namdar. 2020. “Cannabis Phytomolecule ‘Entourage’: From Domestication to Medical Use.” *Trends in Plant Science* 25 (10): 976–84. https://doi.org/10.1016/j.tplants.2020.04.007.

Kostick, Sarah A., Soon Li Teh, John L. Norelli, Stijn Vanderzande, Cameron Peace, and Kate M. Evans. 2021. "Fire Blight QTL Analysis in a Multi-Family Apple Population Identifies a Reduced-Susceptibility Allele in 'Honeycrisp.'" *Horticulture Research* 8 (1): 28. https://doi.org/10.1038/s41438-021-00466-6.

Kotoda, Nobuhiro, Hiroshi Iwanami, Sae Takahashi, and Kazuyuki Abe. 2006. "Antisense Expression of MdTFL1, a TFL1-like Gene, Reduces the Juvenile Phase in Apple." *Journal of the American Society for Horticultural Science. American Society for Horticultural Science* 131 (1): 74–81. https://doi.org/10.21273/JASHS.131.1.74.

Kouassi, Abou Bakari, Charles-Eric Durel, Fabrizio Costa, Stefano Tartarini, Eric van de Weg, Kate Evans, Felicidad Fernandez-Fernandez, et al. 2009. "Estimation of Genetic Parameters and Prediction of Breeding Values for Apple Fruit-Quality Traits Using Pedigreed Plant Material in Europe." *Tree Genetics & Genomes* 5 (4): 659–72. https://doi.org/10.1007/s11295-009-0217-x.

Kreitzman, Maayan, Eric Toensmeier, Kai M. A. Chan, Sean Smukler, and Navin Ramankutty. 2020. "Perennial Staple Crops: Yields, Distribution, and Nutrition in the Global Food System." *Frontiers in Sustainable Food Systems* 4: 216. https://doi.org/10.3389/fsufs.2020.588988.

Kschonsek, Josephine, Theresa Wolfram, Annette Stöckl, and Volker Böhm. 2018. "Polyphenolic Compounds Analysis of Old and New Apple Cultivars and Contribution of Polyphenolic Profile to the In Vitro Antioxidant Capacity." *Antioxidants (Basel, Switzerland)* 7 (1). https://doi.org/10.3390/antiox7010020.

Kumar, Satish, Dorian J. Garrick, Marco Cam Bink, Claire Whitworth, David Chagné, and Richard K. Volz. 2013. "Novel Genomic Approaches Unravel Genetic Architecture of Complex Traits in Apple." *BMC Genomics* 14 (June): 393. https://doi.org/10.1186/1471-2164-14-393.

Kumar, Satish, Pierre Raulier, David Chagné, and Claire Whitworth. 2014. "Molecular-Level and Trait-Level Differentiation between the Cultivated Apple (Malus× Domestica Borkh.) and Its Main Progenitor Malus Sieversii." *Plant Genetic Resources: Characterization and Utilization* 12 (3): 330–40. https://doi.org/10.1017/S1479262114000136.

Kumar, Satish, Richard K. Volz, Peter A. Alspach, and Vincent G. M. Bus. 2010. "Development of a Recurrent Apple Breeding Programme in New Zealand: A Synthesis of Results, and a Proposed Revised Breeding Strategy." *Euphytica/ Netherlands Journal of Plant Breeding* 173 (2): 207–22. https://doi.org/10.1007/s10681-009-0090-6.

Larsen, Bjarne, Zoë Migicovsky, Anne Aae Jeppesen, Kyle M. Gardner, Torben Bo Toldam-Andersen, Sean Myles, Marian Ørgaard, Mikael Agerlin Petersen, and Carsten Pedersen. 2019. "Genome-Wide Association Studies in Apple Reveal Loci for Aroma Volatiles, Sugar Composition, and Harvest Date." *The Plant Genome* 12 (2). https://doi.org/10.3835/plantgenome2018.12.0104.

Laurens, Francois, Maria José Aranzana, Pere Arus, Daniele Bassi, Marco Bink, Joan Bonany, Andrea Caprera, et al. 2018. "An Integrated Approach for Increasing Breeding Efficiency in Apple and Peach in Europe." *Horticulture Research* 5 (March): 11. https://doi.org/10.1038/s41438-018-0016-3.

Laverty, Kaitlin U., Jake M. Stout, Mitchell J. Sullivan, Hardik Shah, Navdeep Gill, Larry Holbrook, Gintaras Deikus, et al. 2019. "A Physical and Genetic Map of Cannabis Sativa Identifies Extensive Rearrangements at the THC/CBD Acid Synthase Loci." *Genome Research* 29 (1): 146–56. https://doi.org/10.1101/gr.242594.118.

Lawler, Andrew. 2019. "Mountain High: Oldest Clear Signs of Pot Use." *Science* 364 (6445): 1018. https://doi.org/10.1126/science.364.6445.1018.

Leisner, Courtney P. 2020. "Review: Climate Change Impacts on Food Security- Focus on Perennial Cropping Systems and Nutritional Value." *Plant Science: An International Journal of Experimental Plant Biology* 293 (April): 110412. https://doi.org/10.1016/j.plantsci.2020.110412.

Lemon, J. 2006. "Plotrix: A Package in the Red Light District of R." *R-News*.

Li, Bin, Ruiyi Fan, Qiaosong Yang, Chunhua Hu, Ou Sheng, Guiming Deng, Tao Dong, et al. 2020. "Genome-Wide Identification and Characterization of the NAC Transcription Factor Family in Musa Acuminata and Expression Analysis during Fruit Ripening." *International Journal of Molecular Sciences* 21 (2). https://doi.org/10.3390/ijms21020634.

Li, Hui-Lin. 1973. "An Archaeological and Historical Account of Cannabis in China." *Economic Botany* 28 (4): 437–48. https://doi.org/10.1007/BF02862859.

Li, Huixia, Haochen Sun, Haiqiang Dong, Shiyao Wang, Xingqiang Fan, Yi Li, Lailiang Cheng, et al. 2023. "Genome Editing of Apple SQUAMOSA PROMOTER BINDNG PROTEIN-LIKE 6 Enhances Adventitious Shoot Regeneration." *Plant Physiology* 191 (2): 840–43. https://doi.org/10.1093/plphys/kiac570.

Li, Qiang, Manoj Sapkota, and Esther van der Knaap. 2020. "Perspectives of CRISPR/Cas-Mediated Cis-Engineering in Horticulture: Unlocking the Neglected Potential for Crop Improvement." *Horticulture Research* 7 (1): 36. https://doi.org/10.1038/s41438-020-0258-8.

Li, Shaoya, and Lanqin Xia. 2020. “Precise Gene Replacement in Plants through CRISPR/Cas Genome Editing Technology: Current Status and Future Perspectives.” *ABIOTECH* 1 (1): 58–73. https://doi.org/10.1007/s42994-019-00009-7.

Li, Shutian, and Sabine Zachgo. 2013. “TCP3 Interacts with R2R3-MYB Proteins, Promotes Flavonoid Biosynthesis and Negatively Regulates the Auxin Response in Arabidopsis Thaliana.” *The Plant Journal: For Cell and Molecular Biology* 76 (6): 901–13. https://doi.org/10.1111/tpj.12348.

Li, Xiaolong, Lun Liu, Meiling Ming, Hongju Hu, Mingyue Zhang, Jing Fan, Bobo Song, Shaoling Zhang, and Jun Wu. 2019. “Comparative Transcriptomic Analysis Provides Insight into the Domestication and Improvement of Pear (P. Pyrifolia) Fruit.” *Plant Physiology* 180 (1): 435–52. https://doi.org/10.1104/pp.18.01322.

Li, Xuewei, Ling Kui, Jing Zhang, Yinpeng Xie, Liping Wang, Yan Yan, Na Wang, et al. 2016. “Improved Hybrid de Novo Genome Assembly of Domesticated Apple (Malus x Domestica).” *GigaScience* 5 (1): 35. https://doi.org/10.1186/s13742-016-0139-0.

Li, Yong, Ke Cao, Gengrui Zhu, Weichao Fang, Changwen Chen, Xinwei Wang, Pei Zhao, et al. 2019. “Genomic Analyses of an Extensive Collection of Wild and Cultivated Accessions Provide New Insights into Peach Breeding History.” *Genome Biology* 20 (1): 36. https://doi.org/10.1186/s13059-019-1648-9.

Lian, Zhaoyuan, Chi Dinh Nguyen, Li Liu, Guiluan Wang, Jianjun Chen, Songhu Wang, Ganjun Yi, et al. 2022. “Application of Developmental Regulators to Improve in Planta or in Vitro Transformation in Plants.” *Plant Biotechnology Journal*, May. https://doi.org/10.1111/pbi.13837.

Liang, Wei, Luca Dondini, Paolo De Franceschi, Roberta Paris, Silviero Sansavini, and Stefano Tartarini. 2015. “Genetic Diversity, Population Structure and Construction of a Core Collection of Apple Cultivars from Italian Germplasm.” *Plant Molecular Biology Reporter / ISPMB* 33 (3): 458–73. https://doi.org/10.1007/s11105-014-0754-9.

Liao, Liao, Weihan Zhang, Bo Zhang, Ting Fang, Xiao-Fei Wang, Yaming Cai, Collins Ogutu, et al. 2021. “Unraveling a Genetic Roadmap for Improved Taste in the Domesticated Apple.” *Molecular Plant* 14 (9): 1454–71. https://doi.org/10.1016/j.molp.2021.05.018.

Liebhard, R., M. Kellerhals, W. Pfammatter, M. Jertmini, and C. Gessler. 2003. “Mapping Quantitative Physiological Traits in Apple (Malus x Domestica Borkh.).” *Plant Molecular Biology* 52 (3): 511–26. https://doi.org/10.1023/a:1024886500979.

Liu, Gang-Shuai, Hong-Li Li, Donald Grierson, and Da-Qi Fu. 2022. "NAC Transcription Factor Family Regulation of Fruit Ripening and Quality: A Review." *Cells* 11 (3): 525. https://doi.org/10.3390/cells11030525.

Liu, Hao, Yuduan Ding, Yanqing Zhou, Wenqi Jin, Kabin Xie, and Ling-Ling Chen. 2017. "CRISPR-P 2.0: An Improved CRISPR-Cas9 Tool for Genome Editing in Plants." *Molecular Plant* 10 (3): 530–32. https://doi.org/10.1016/j.molp.2017.01.003.

Liu, Lei, Joseph Gallagher, Edgar Demesa Arevalo, Richelle Chen, Tara Skopelitis, Qingyu Wu, Madelaine Bartlett, and David Jackson. 2021. "Enhancing Grain-Yield-Related Traits by CRISPR-Cas9 Promoter Editing of Maize CLE Genes." *Nature Plants* 7 (3): 287–94. https://doi.org/10.1038/s41477-021-00858-5.

Livingston, Samuel J., Teagen D. Quilichini, Judith K. Booth, Darren C. J. Wong, Kim H. Rensing, Jessica Laflamme-Yonkman, Simone D. Castellarin, Joerg Bohlmann, Jonathan E. Page, and A. Lacey Samuels. 2020. "Cannabis Glandular Trichomes Alter Morphology and Metabolite Content during Flower Maturation." *The Plant Journal: For Cell and Molecular Biology* 101 (1): 37–56. https://doi.org/10.1111/tpj.14516.

Lü, Peitao, Sheng Yu, Ning Zhu, Yun-Ru Chen, Biyan Zhou, Yu Pan, David Tzeng, et al. 2018. "Genome Encode Analyses Reveal the Basis of Convergent Evolution of Fleshy Fruit Ripening." *Nature Plants* 4 (10): 784–91. https://doi.org/10.1038/s41477-018-0249-z.

Luby, James J., and Douglas V. Shaw. 2001. "Does Marker-Assisted Selection Make Dollars and Sense in a Fruit Breeding Program?" *HortScience: A Publication of the American Society for Horticultural Science* 36 (5): 872–79. https://doi.org/10.21273/hortsci.36.5.872.

Lynch, Ryan C. ;., Daniela Vergara, Silas Tittes, Kristin White, C. J. Schwartz, Matthew J. Gibbs, Travis C. Ruthenburg, et al. 2017. "Genomic and Chemical Diversity in Cannabis." https://doi.org/10.1080/07352689.2016.1265363.

Ma, Baiquan, Jie Chen, Hongyu Zheng, Ting Fang, Collins Ogutu, Shaohua Li, Yuepeng Han, and Benhong Wu. 2015. "Comparative Assessment of Sugar and Malic Acid Composition in Cultivated and Wild Apples." *Food Chemistry* 172: 86–91. https://doi.org/10.1016/j.foodchem.2014.09.032.

Malabarba, Jaiana, Elisabeth Chevreau, Nicolas Dousset, Florian Veillet, Julie Moizan, and Emilie Vergne. 2020. "New Strategies to Overcome Present CRISPR/Cas9 Limitations in Apple and Pear: Efficient Dechimerization and Base Editing." *International Journal of Molecular Sciences* 22 (1). https://doi.org/10.3390/ijms22010319.

Maliepaard, C., F. H. Alston, G. van Arkel, L. M. Brown, E. Chevreau, F. Dunemann, K. M. Evans, et al. 1998. "Aligning Male and Female Linkage Maps of Apple (Malus Pumila Mill.) Using Multi-Allelic Markers." *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik* 97 (1): 60–73. https://doi.org/10.1007/s001220050867.

Malnoy, Mickael, Roberto Viola, Min-Hee Jung, Ok-Jae Koo, Seokjoong Kim, Jin-Soo Kim, Riccardo Velasco, and Chidananda Nagamangala Kanchiswamy. 2016. "DNA-Free Genetically Edited Grapevine and Apple Protoplast Using CRISPR/Cas9 Ribonucleoproteins." *Frontiers in Plant Science* 7: 1904. https://doi.org/10.3389/fpls.2016.01904.

Mao, Yanfei, Jose Ramon Botella, Yaoguang Liu, and Jian-Kang Zhu. 2019. "Gene Editing in Plants: Progress and Challenges." *National Science Review* 6 (3): 421–37. https://doi.org/10.1093/nsr/nwz005.

Martín-Pizarro, Carmen, José G. Vallarino, Sonia Osorio, Victoriano Meco, María Urrutia, Jeremy Pillet, Ana Casañal, et al. 2021. "The NAC Transcription Factor FaRIF Controls Fruit Ripening in Strawberry." *The Plant Cell*, February. https://doi.org/10.1093/plcell/koab070.

McClure, Kendra. 2017. "Exploring Horticulturally Important Traits in Apple Using Genetic Mapping in Populations of Differing Size and Composition."

McClure, Kendra A., Kyle M. Gardner, Gavin M. Douglas, Jun Song, Charles F. Forney, John DeLong, Lihua Fan, et al. 2018. "A Genome-Wide Association Study of Apple Quality and Scab Resistance." *The Plant Genome* 11 (1). https://doi.org/10.3835/plantgenome2017.08.0075.

McClure, Kendra A., Kyle M. Gardner, Peter M. A. Toivonen, Cheryl R. Hampson, Jun Song, Charles F. Forney, John DeLong, Istvan Rajcan, and Sean Myles. 2016. "QTL Analysis of Soft Scald in Two Apple Populations." *Horticulture Research* 3 (1): 16043. https://doi.org/10.1038/hortres.2016.43.

McClure, Kendra A., Jason Sawler, Kyle M. Gardner, Daniel Money, and Sean Myles. 2014. "Genomics: A Potential Panacea for the Perennial Problem." *American Journal of Botany* 101 (10): 1780–90. https://doi.org/10.3732/ajb.1400143.

McClure, Kendra, Yuihui Gong, Jun Song, Melinda Vinqvist-Tymchuk, Leslie Campbell Palmer, Lihua Fan, Karen Burgher-MacLellan, et al. 2019. "Genome-Wide Association Studies in Apple Reveal Loci of Large Effect Controlling Apple Polyphenols." *Horticulture Research* 6 (1): 107. https://doi.org/10.1038/s41438-019-0190-y.

McCouch, Susan, Gregory J. Baute, James Bradeen, Paula Bramel, Peter K. Bretting, Edward Buckler, John M. Burke, et al. 2013. "Agriculture: Feeding the Future." *Nature* 499 (7456): 23–24. https://doi.org/10.1038/499023a.

McKernan, Kevin J., Yvonne Helbert, Liam T. Kane, Heather Ebling, Lei Zhang, Biao Liu, Zachary Eaton, et al. 2020. "Sequence and Annotation of 42 Cannabis Genomes Reveals Extensive Copy Number Variation in Cannabinoid Synthesis and Pathogen Resistance Genes." *Cold Spring Harbor Laboratory*. https://doi.org/10.1101/2020.01.03.894428.

Mehlenbacher, Shawn A., and Anna M. Voordeckers. 1991. "Relationship of Flowering Time, Rate of Seed Germination, and Time of Leaf Budbreak and Usefulness in Selecting for Late-Flowering Apples." *Journal of the American Society for Horticultural Science. American Society for Horticultural Science* 116 (3): 565–68. https://doi.org/10.21273/JASHS.116.3.565.

Meijer, Etienne P. M. de, Manuela Bagatta, Andrea Carboni, Paola Crucitti, V. M. Cristiana Moliterni, Paolo Ranalli, and Giuseppe Mandolino. 2003. "The Inheritance of Chemical Phenotype in Cannabis Sativa L." *Genetics* 163 (1): 335–46. https://doi.org/10.1093/genetics/163.1.335.

Mendel, P., Ajinkya Bharat Lalge, T. Vyhnánek, Václav, Trojan, P. Kalousek, Hugo Maassen, and L. Havel. 2016. "PROGRESS IN EARLY SEX DETERMINATION OF CANNABIS PLANT BY DNA MARKERS." https://www.semanticscholar.org/paper/f974e878decf5eac94bd194297bf3d249afc1672.

Meyer, Rachel S., Ashley E. DuVal, and Helen R. Jensen. 2012. "Patterns and Processes in Crop Domestication: An Historical Review and Quantitative Analysis of 203 Global Food Crops." *The New Phytologist* 196 (1): 29–48. https://doi.org/10.1111/j.1469-8137.2012.04253.x.

Migicovsky, Zoë, Gavin M. Douglas, and Sean Myles. 2022. "Genotyping-by-Sequencing of Canada's Apple Biodiversity Collection." *Frontiers in Genetics* 13 (August): 934712. https://doi.org/10.3389/fgene.2022.934712.

Migicovsky, Zoë, Kyle M. Gardner, Daniel Money, Jason Sawler, Joshua S. Bloom, Peter Moffett, C. Thomas Chao, et al. 2016. "Genome to Phenome Mapping in Apple Using Historical Data." *The Plant Genome* 9 (2): 0. https://doi.org/10.3835/plantgenome2015.11.0113.

Migicovsky, Zoë, Kyle M. Gardner, Christopher Richards, C. Thomas Chao, Heidi R. Schwaninger, Gennaro Fazio, Gan-Yuan Zhong, and Sean Myles. 2021. "Genomic Consequences of Apple Improvement." *Horticulture Research* 8 (1): 9. https://doi.org/10.1038/s41438-020-00441-7.

Migicovsky, Zoë, and Sean Myles. 2017. "Exploiting Wild Relatives for Genomics-Assisted Breeding of Perennial Crops." *Frontiers in Plant Science* 8 (April): 1–16. https://doi.org/10.3389/fpls.2017.00460.

Migicovsky, Zoë, Emily Warschefsky, Laura L. Klein, and Allison J. Miller. 2019. "Using Living Germplasm Collections to Characterize, Improve, and Conserve Woody Perennials." *Crop Science* 59 (6): 2365–80. https://doi.org/10.2135/cropsci2019.05.0353.

Migicovsky, Zoë, Trevor H. Yeats, Sophie Watts, Jun Song, Charles F. Forney, Karen Burgher-MacLellan, Daryl J. Somers, et al. 2021. "Apple Ripening Is Controlled by a NAC Transcription Factor." *Frontiers in Genetics* 12 (June): 671300. https://doi.org/10.3389/fgene.2021.671300.

Miller, Allison J., and Briana L. Gross. 2011. "From Forest to Field: Perennial Fruit Crop Domestication." *American Journal of Botany* 98 (9): 1389–1414. https://doi.org/10.3732/ajb.1000522.

Money, Daniel, Zoe Migicovsky, Kyle Gardner, and Sean Myles. 2017. "LinkImputeR: User-Guided Genotype Calling and Imputation for Non-Model Organisms." *BMC Genomics* 18 (1): 1—12. https://doi.org/10.1186/s12864-017-3873-5.

Mudge, Elizabeth M., Paula N. Brown, and Susan J. Murch. 2019. "The Terroir of Cannabis: Terpene Metabolomics as a Tool to Understand Cannabis Sativa Selections." *Planta Medica* 85 (9–10): 781—796. https://doi.org/10.1055/a-0915-2550.

Murata, Masatsune, Ikuko Noda, and Seiichi Homma. 1995. "Enzymatic Browning of Apples on the Market: Relationship between Browning, Polyphenol Content, and Polyphenol Oxidase." *Nippon Shokuhin Kagaku Kogaku Kaishi* 42 (10): 820–26. https://doi.org/10.3136/nskkk.42.820.

Musacchi, Stefano, and Sara Serra. 2018. "Apple Fruit Quality: Overview on Pre-Harvest Factors." *Scientia Horticulturae* 234 (April): 409–30. https://doi.org/10.1016/j.scienta.2017.12.057.

Myles, Sean. 2013. "Improving Fruit and Wine: What Does Genomics Have to Offer?" *Trends in Genetics: TIG* 29 (4): 190–96. https://doi.org/10.1016/j.tig.2013.01.006.

Myles, Sean, Jason Peiffer, Patrick J. Brown, Elhan S. Ersoz, Zhiwu Zhang, Denise E. Costich, and Edward S. Buckler. 2009. "Association Mapping: Critical Considerations Shift from Genoty[Ing to Experimental Design." *The Plant Cell* 121: 2194–2202. https://doi.org/10.1105/tpc.109.068437.

Nagle, Michael, Annabelle Déjardin, Gilles Pilate, and Steven H. Strauss. 2018. "Opportunities for Innovation in Genetic Transformation of Forest Trees." *Frontiers in Plant Science* 9 (October): 1443. https://doi.org/10.3389/fpls.2018.01443.

Naim, Fatima, Benjamin Dugdale, Jennifer Kleidon, Anthony Brinin, Kylie Shand, Peter Waterhouse, and James Dale. 2018. "Gene Editing the Phytoene Desaturase Alleles of Cavendish Banana Using CRISPR/Cas9." *Transgenic Research* 27 (5): 451–60. https://doi.org/10.1007/s11248-018-0083-0.

Nasti, Ryan A., and Daniel F. Voytas. 2021. "Attaining the Promise of Plant Gene Editing at Scale." *Proceedings of the National Academy of Sciences of the United States of America* 118 (22). https://doi.org/10.1073/pnas.2004846117.

Naville, Sebastien. 2019. "$340 Billion: The Global Cannabis Market." Geneva Business News. August 9, 2019. https://www.gbnews.ch/340-billion-the-global-cannabis-market/.

Nieuwenhuizen, Niels J., Xiuyin Chen, Mindy Y. Wang, Adam J. Matich, Ramon Lopez Perez, Andrew C. Allan, Sol A. Green, and Ross G. Atkinson. 2015. "Natural Variation in Monoterpene Synthesis in Kiwifruit: Transcriptional Regulation of Terpene Synthases by NAC and ETHYLENE-INSENSITIVE3-like Transcription Factors." *Plant Physiology* 167 (4): 1243–58. https://doi.org/10.1104/pp.114.254367.

Nishitani, Chikako, Narumi Hirai, Sadao Komori, Masato Wada, Kazuma Okada, Keishi Osakabe, Toshiya Yamamoto, and Yuriko Osakabe. 2016. "Efficient Genome Editing in Apple Using a CRISPR/Cas9 System." *Scientific Reports* 6 (1): 31481. https://doi.org/10.1038/srep31481.

Noiton, Dominique A. M., and Peter A. Alspach. 1996. "Founding Clones, Inbreeding, Coancestry, and Status Number of Modern Apple Cultivars." *Journal of the American Society for Horticultural Science. American Society for Horticultural Science* 121 (5): 773–82. https://doi.org/10.21273/jashs.121.5.773.

Nordborg, Magnus, and Detlef Weigel. 2008. "Next-Generation Genetics in Plants." *Nature* 456 (7223): 720–23. https://doi.org/10.1038/nature07629.

Noronha, Henrique, Angélica Silva, Zhanwu Dai, Philippe Gallusci, Adamo D. Rombolà, Serge Delrot, and Hernâni Gerós. 2018. "A Molecular Perspective on Starch Metabolism in Woody Tissues." *Planta* 248 (3): 559–68. https://doi.org/10.1007/s00425-018-2954-2.

Novembre, John, Toby Johnson, Katarzyna Bryc, Zoltán Kutalik, Adam R. Boyko, Adam Auton, Amit Indap, et al. 2008. "Genes Mirror Geography within Europe." *Nature* 456 (7218): 98–101. https://doi.org/10.1038/nature07331.

Nybom, Hilde, Masoud Ahmadi-Afzadi, Jasna Sehic, and Maarten Hertog. 2013. "DNA Marker-Assisted Evaluation of Fruit Firmness at Harvest and Post-Harvest Fruit Softening in a Diverse Apple Germplasm." *Tree Genetics & Genomes* 9 (1): 279–90. https://doi.org/10.1007/s11295-012-0554-z.

Ogura, Takehiko, and Wolfgang Busch. 2015. “From Phenotypes to Causal Sequences: Using Genome Wide Association Studies to Dissect the Sequence Basis for Variation of Plant Development.” *Current Opinion in Plant Biology* 23 (February): 98–108. https://doi.org/10.1016/j.pbi.2014.11.008.

Oraguzie, N. C., H. Iwanami, J. Soejima, T. Harada, and A. Hall. 2004. “Inheritance of the Md-ACS1 Gene and Its Relationship to Fruit Softening in Apple ( Malus x Domestica Borkh.).” *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik* 108 (8): 1526–33. https://doi.org/10.1007/s00122-003-1574-8.

Osakabe, Yuriko, Zhenchang Liang, Chong Ren, Chikako Nishitani, Keishi Osakabe, Masato Wada, Sadao Komori, et al. 2018. “CRISPR–Cas9-Mediated Genome Editing in Apple and Grapevine.” *Nature Protocols* 13 (12): 2844–63. https://doi.org/10.1038/s41596-018-0067-9.

Peace, Cameron P., Luca Bianco, Michela Troggio, Eric van de Weg, Nicholas P. Howard, Amandine Cornille, Charles-Eric Durel, et al. 2019. “Apple Whole Genome Sequences: Recent Advances and New Prospects.” *Horticulture Research* 6 (April): 59. https://doi.org/10.1038/s41438-019-0141-7.

Pearce, Daniel D., Katherine Mitsouras, and Kristopher J. Irizarry. 2014. “Discriminating the Effects of Cannabis Sativa and Cannabis Indica: A Web Survey of Medical Cannabis Users.” *The Journal of Alternative and Complementary Medicine* 20 (10): 787—791. https://doi.org/10.1089/acm.2013.0190.

Pech, Jean-Claude, Alain Latché, and Mondher Bouzayen. 2010. “Ethylene Biosynthesis.” In *Plant Hormones: Biosynthesis, Signal Transduction, Action!*, edited by Peter J. Davies, 115–36. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-1-4020-2686-7_6.

Peil, A., F. Dunemann, K. Richter, M. Hoefer, I. Király, H. Flachowsky, and M-V Hanke. 2008. “Resistance Breeding in Apple at Dresden-Pillnitz.” In *Ecofruit - 13th International Conference on Cultivation Technique and Phytopathological Problems in Organic Fruit-Growing: Proceedings to the Conference from 18th February to 20th February 2008 at Weinsberg/Germany*, 220–25. http://orgprints.org/13698/.

Petit, Jordi, Elma M. J. Salentijn, Maria-João Paulo, Christel Denneboom, Eibertus N. van Loo, and Luisa M. Trindade. 2020. “Elucidating the Genetic Architecture of Fiber Quality in Hemp (Cannabis Sativa L.) Using a Genome-Wide Association Study.” *Frontiers in Genetics* 11 (September): 566314. https://doi.org/10.3389/fgene.2020.566314.

Petit, Jordi, Elma M. J. Salentijn, Maria-João Paulo, Christel Denneboom, and Luisa M. Trindade. 2020. “Genetic Architecture of Flowering Time and Sex Determination in Hemp (Cannabis Sativa L.): A Genome-Wide Association Study.” *Frontiers in Plant Science* 11: 1704. https://doi.org/10.3389/fpls.2020.569958.

Pirona, Raul, Iban Eduardo, Igor Pacheco, Cassia Da Silva Linge, Mara Miculan, Ignazio Verde, Stefano Tartarini, et al. 2013. "Fine Mapping and Identification of a Candidate Gene for a Major Locus Controlling Maturity Date in Peach." *BMC Plant Biology* 13 (October): 166. https://doi.org/10.1186/1471-2229-13-166.

Pompili, Valerio, Lorenza Dalla Costa, Stefano Piazza, Massimo Pindo, and Mickael Malnoy. 2020. "Reduced Fire Blight Susceptibility in Apple Cultivars Using a High-Efficiency CRISPR/Cas9-FLP/FRT-Based Gene Editing System." *Plant Biotechnology Journal* 18 (3): 845–58. https://doi.org/10.1111/pbi.13253.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, et al. 2007. "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses." *American Journal of Human Genetics* 81 (3): 559—575. https://doi.org/10.1086/519795.

Qi, Xiliang, Yuanxin Dong, Congli Liu, Lulu Song, Lei Chen, and Ming Li. 2022. "The PavNAC56 Transcription Factor Positively Regulates Fruit Ripening and Softening in Sweet Cherry (Prunus Avium)." *Physiologia Plantarum* 174 (6): e13834. https://doi.org/10.1111/ppl.13834.

R Core Team. 2018. "R: A Language and Environment for Statistical Computing." Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Ren, Chong, Xianju Liu, Zhan Zhang, Yi Wang, Wei Duan, Shaohua Li, and Zhenchang Liang. 2016. "CRISPR/Cas9-Mediated Efficient Targeted Mutagenesis in Chardonnay (Vitis Vinifera L.)." *Scientific Reports* 6 (August): 32289. https://doi.org/10.1038/srep32289.

Ren, Guangpeng, Xu Zhang, Ying Li, Kate Ridout, Martha L. Serrano-Serrano, Yongzhi Yang, Ai Liu, et al. 2021. "Large-Scale Whole-Genome Resequencing Unravels the Domestication History of *Cannabis Sativa*." *Science Advances* 7 (29). https://doi.org/10.1126/sciadv.abg2286.

Ren, Meng, Zihua Tang, Xinhua Wu, Robert Spengler, Hongen Jiang, Yimin Yang, and Nicole Boivin. 2019. "The Origins of Cannabis Smoking: Chemical Residue Evidence from the First Millennium BCE in the Pamirs." *Science Advances* 5 (6): eaaw1391. https://doi.org/10.1126/sciadv.aaw1391.

Ren, Qiurong, Simon Sretenovic, Shishi Liu, Xu Tang, Lan Huang, Yao He, Li Liu, et al. 2021. "PAM-Less Plant Genome Editing Using a CRISPR-SpRY Toolbox." *Nature Plants* 7 (1): 25–33. https://doi.org/10.1038/s41477-020-00827-4.

Rodríguez-Leal, Daniel, Zachary H. Lemmon, Jarrett Man, Madelaine E. Bartlett, and Zachary B. Lippman. 2017. "Engineering Quantitative Trait Variation for Crop Improvement by Genome Editing." *Cell* 171 (2): 470-480.e8. https://doi.org/10.1016/j.cell.2017.08.030.

Rouf, Reyazul, Mathew Reynolds, Francisco Pinto, and Mohd Anwar. 2019. “Plant Science High-Throughput Phenotyping for Crop Improvement in the Genomics Era.” *Plant Science*, no. May 2018: 0–1. https://doi.org/10.1016/j.plantsci.2019.01.007.

Rousset, François, and Rotem Sorek. 2021. “A Treasure Trove of Molecular Scissors.” *Science*. https://doi.org/10.1126/science.abm2239.

Ru, Sushan, Dorrie Main, Kate Evans, and Cameron Peace. 2015. “Current Applications, Challenges, and Perspectives of Marker-Assisted Seedling Selection in Rosaceae Tree Fruit Breeding.” *Tree Genetics & Genomes* 11 (1): 8. https://doi.org/10.1007/s11295-015-0834-5.

Rupasinghe, H. P. Vasantha, Laixin Wang, Gwendolyn M. Huber, and Nancy L. Pitts. 2008. “Effect of Baking on Dietary Fibre and Phenolics of Muffins Incorporated with Apple Skin Powder.” *Food Chemistry* 107 (3): 1217–24. https://doi.org/10.1016/j.foodchem.2007.09.057.

Russo, Ethan B. 2011. “Taming THC: Potential Cannabis Synergy and Phytocannabinoid-terpenoid Entourage Effects.” *British Journal of Pharmacology* 163 (7): 1344–64. https://doi.org/10.1111/j.1476-5381.2011.01238.x.

Russo, Ethan B., and Jahan Marcu. 2017. “Chapter Three - Cannabis Pharmacology: The Usual Suspects and a Few Promising Leads.” In *Advances in Pharmacology*, edited by David Kendall and Stephen P. H. Alexander, 80:67–134. Academic Press. https://doi.org/10.1016/bs.apha.2017.03.004.

Sakamoto, Koichi, Yukio Akiyama, Kiichi Fukui, Hiroshi Kamada, and Shinobu Satoh. 1998. “Characterization; Genome Sizes and Morphology of Sex Chromosomes in Hemp (Cannabis Sativa L.).” *Cytologia* 63 (4): 459–64. https://doi.org/10.1508/cytologia.63.459.

Samara, Mohamed, Ahmed Nasser, and Uri Mingelgrin. 2022. “Critical Examination of the Suitability of the Folin-Ciocalteu Reagent Assay for Quantitative Analysis of Polyphenols—the Case of Olive-Mill Wastewater.” *American Journal of Analytical Chemistry* 13 (11): 476–93. https://doi.org/10.4236/ajac.2022.1311032.

Sanoner, Philippe, Sylvain Guyot, Nathalie Marnet, Daniel Molle, and J-F Drilleau. 1999. “Polyphenol Profiles of French Cider Apple Varieties ( Malus Domestica Sp.).” *Journal of Agricultural and Food Chemistry* 47 (12): 4847–53. https://doi.org/10.1021/jf990563y.

Sawler, Jason, Jake M. Stout, Kyle M. Gardner, Darryl Hudson, John Vidmar, Laura Butler, Jonathan E. Page, and Sean Myles. 2015. “The Genetic Structure of Marijuana and Hemp.” *PloS One* 10 (8): 1—9. https://doi.org/10.1371/journal.pone.0133292.

Saxena, Rachit K., David Edwards, and Rajeev K. Varshney. 2014. “Structural Variations in Plant Genomes.” *Briefings in Functional Genomics* 13 (4): 296–307. https://doi.org/10.1093/bfgp/elu016.

Schaid, Daniel J., Wenan Chen, and Nicholas B. Larson. 2018. “From Genome-Wide Associations to Candidate Causal Variants by Statistical Fine-Mapping.” *Nature Reviews. Genetics* 19 (8): 491–504. https://doi.org/10.1038/s41576-018-0016-z.

Schlathölter, Ina, Giovanni A. L. Broggini, Sebastian Streb, Bruno Studer, and Andrea Patocchi. 2023. “Field Study of the Fire-Blight-Resistant Cisgenic Apple Line C44.4.146.” *The Plant Journal: For Cell and Molecular Biology* 113 (6): 1160–75. https://doi.org/10.1111/tpj.16083.

Schwabe, Anna L., and Mitchell E. McGlaughlin. 2019. “Genetic Tools Weed out Misconceptions of Strain Reliability in Cannabis Sativa: Implications for a Budding Industry.” *Journal of Cannabis Research* 1 (3). https://doi.org/10.1186/s42238-019-0001-1.

Seeley, S. D., and J. L. Anderson. 2003. “Apple-Orchard Freeze Protection.” In *Apples: Botany, Production and Uses.*, edited by D. C. Ferree and I. J. Warrington, 521–38. CABI Publishing.

Segura, Vincent, Bjarni J. Vilhjálmsson, Alexander Platt, Arthur Korte, Ümit Seren, Quan Long, and Magnus Nordborg. 2012. “An Efficient Multi-Locus Mixed-Model Approach for Genome-Wide Association Studies in Structured Populations.” *Nature Genetics* 44 (7): 825–30. https://doi.org/10.1038/ng.2314.

Singleton, Vernon L., Rudolf Orthofer, and Rosa M. Lamuela-Raventós. 1999. “Analysis of Total Phenols and Other Oxidation Substrates and Antioxidants by Means of Folin-Ciocalteu Reagent.” In *Methods in Enzymology*, 299:152–78. Academic Press. https://doi.org/10.1016/S0076-6879(99)99017-1.

Small, E., and H. D. Beckstead. 1973. “Letter: Cannabinoid Phenotypes in Cannabis Sativa.” *Nature* 245 (5421): 147–48. https://doi.org/10.1038/245147a0.

Small, Ernest. 2015. “Evolution and Classification of Cannabis Sativa (Marijuana, Hemp) in Relation to Human Utilization.” *The Botanical Review; Interpreting Botanical Progress* 81 (3): 189—294. https://doi.org/10.1007/s12229-015-9157-3.

Small, Ernest, H. D. Beckstead, and Allan Chan. 1975. “The Evolution of Cannabinoid Phenotypes in Cannabis.” *Economic Botany* 29 (3): 219–32. http://www.jstor.org/stable/4253607.

Smith, Christiana J., Daniela Vergara, Brian Keegan, and Nick Jikomes. 2021. “The Phytochemical Diversity of Commercial Cannabis in the United States.” *BioRxiv*. https://doi.org/10.1101/2021.07.05.451212.

Sun, Xuepeng, Chen Jiao, Heidi Schwaninger, C. Thomas Chao, Yumin Ma, Naibin Duan, Awais Khan, et al. 2020. "Phased Diploid Genome Assemblies and Pan-Genomes Provide Insights into the Genetic History of Apple Domestication." *Nature Genetics*, November. https://doi.org/10.1038/s41588-020-00723-9.

Sun, Yanqing, Lianguang Shang, Qian-Hao Zhu, Longjiang Fan, and Longbiao Guo. 2021. "Twenty Years of Plant Genome Sequencing: Achievements and Challenges." *Trends in Plant Science*, November. https://doi.org/10.1016/j.tplants.2021.10.006.

Szczesniak, Alina Surmacka. 2002. "Texture Is a Sensory Property." *Food Quality and Preference* 13 (4): 215–25. https://doi.org/10.1016/S0950-3293(01)00039-8.

Tan, Qiuping, Sen Li, Yuzheng Zhang, Min Chen, Binbin Wen, Shan Jiang, Xiude Chen, et al. 2021. "Chromosome-Level Genome Assemblies of Five Prunus Species and Genome-Wide Association Studies for Key Agronomic Traits in Peach." *Horticulture Research* 8 (1): 213. https://doi.org/10.1038/s41438-021-00648-2.

Temple, Leslie Mendoza, and Jerrold B. Leikin. 2020. "Tetrahydrocannabinol - Friend or Foe? - Debate." *Clinical Toxicology* 58 (2): 75–81. https://doi.org/10.1080/15563650.2019.1610567.

Thompson-Witrick, Katherine A., Katheryn M. Goodrich, Andrew P. Neilson, E. Kenneth Hurley, Gregory M. Peck, and Amanda C. Stewart. 2014. "Characterization of the Polyphenol Composition of 20 Cultivars of Cider, Processing, and Dessert Apples ( Malus × Domestica Borkh.) Grown in Virginia." *Journal of Agricultural and Food Chemistry* 62 (41): 10181–91. https://doi.org/10.1021/jf503379t.

Tian, Zhixi, Jia-Wei Wang, Jiayang Li, and Bin Han. 2021. "Designing Future Crops: Challenges and Strategies for Sustainable Agriculture." *The Plant Journal: For Cell and Molecular Biology* 105 (5): 1165–78. https://doi.org/10.1111/tpj.15107.

Tibbs Cortes, Laura, Zhiwu Zhang, and Jianming Yu. 2021. "Status and Prospects of Genome-Wide Association Studies in Plants." *The Plant Genome*, January, e20077. https://doi.org/10.1002/tpg2.20077.

Toivonen, Peter M. A. 2006. "Fresh-Cut Apples: Challenges and Opportunities for Multi-Disciplinary Research." *Canadian Journal of Plant Science. Revue Canadienne de Phytotechnie* 86 (Special Issue): 1361–68. https://doi.org/10.4141/p06-147.

Troggio, M., A. Gleave, S. Salvi, D. Chagné, A. Cestaro, S. Kumar, R. N. Crowhurst, and S. E. Gardiner. 2012. "Apple, from Genome to Breeding." *Tree Genetics & Genomes* 8 (3): 509–29. https://doi.org/10.1007/s11295-012-0492-9.

Turnbull, Crystal, Morten Lillemo, and Trine A. K. Hvoslef-Eide. 2021. "Global Regulation of Genetically Modified Crops Amid the Gene Edited Crop Boom - A Review." *Frontiers in Plant Science* 12 (February): 630396. https://doi.org/10.3389/fpls.2021.630396.

Urrestarazu, Jorge, Hélène Muranty, Caroline Denancé, Diane Leforestier, Elisa Ravon, Arnaud Guyader, Rémi Guisnel, et al. 2017. "Genome-Wide Association Mapping of Flowering and Ripening Periods in Apple." *Frontiers in Plant Science* 8 (November): 1923. https://doi.org/10.3389/fpls.2017.01923.

"US Apple and Pear Forecast." n.d. The World Apple and Pear Association (WAPA). Accessed October 26, 2020. http://www.wapa-association.org/asp/page_1.asp?doc_id=447.

Vallebona, Chiara, Alberto Mantino, and Enrico Bonari. 2016. "Exploring the Potential of Perennial Crops in Reducing Soil Erosion: A GIS-Based Scenario Analysis in Southern Tuscany, Italy." *Applied Geography* 66 (January): 119–31. https://doi.org/10.1016/j.apgeog.2015.11.015.

Valois, Sarah, Ian A. Merwin, and Olga I. Padilla-Zakour. 2006. "Characterization of Fermented Cider Apple Cultivars Grown in Upstate New York." *Journal-American Pomological Society* 60 (3): 113–28. http://dx.doi.org/.

Varkonyi-Gasic, Erika, Tianchi Wang, Charlotte Voogd, Subin Jeon, Revel S. M. Drummond, Andrew P. Gleave, and Andrew C. Allan. 2019. "Mutagenesis of Kiwifruit CENTRORADIALIS-like Genes Transforms a Climbing Woody Perennial with Long Juvenility and Axillary Flowering into a Compact Plant with Rapid Terminal Flowering." *Plant Biotechnology Journal* 17 (5): 869–80. https://doi.org/10.1111/pbi.13021.

Varshney, Rajeev K., Abhishek Bohra, Jianming Yu, Andreas Graner, Qifa Zhang, and Mark E. Sorrells. 2021. "Designing Future Crops: Genomics-Assisted Breeding Comes of Age." *Trends in Plant Science* 0 (0). https://doi.org/10.1016/j.tplants.2021.03.010.

Velasco, Riccardo, Andrey Zharkikh, Jason Affourtit, Amit Dhingra, Alessandro Cestaro, Ananth Kalyanaraman, Paolo Fontana, et al. 2010. "The Genome of the Domesticated Apple (Malus × Domestica Borkh.)." *Nature Genetics* 42 (10): 833–39. https://doi.org/10.1038/ng.654.

Vergara, Daniela, Ezra L. Huscher, Kyle G. Keepers, Robert M. Givens, Christian G. Cizek, Anthony Torres, Reggie Gaudino, and Nolan C. Kane. 2019. "Gene Copy Number Is Associated with Phytochemistry in Cannabis Sativa." *BioRxiv*, 736181. https://doi.org/10.1101/736181.

Verma, S., K. Evans, Y. Guan, J. J. Luby, U. R. Rosyara, N. P. Howard, N. Bassil, M. C. A. M. Bink, W. E. van de Weg, and C. P. Peace. 2019. "Two Large-Effect QTLs, Ma and Ma3, Determine Genetic Potential for Acidity in Apple Fruit: Breeding Insights from a Multi-Family Study." *Tree Genetics & Genomes* 15 (2): 18. https://doi.org/10.1007/s11295-019-1324-y.

Vinson, Joe A., Xuehui Su, Ligia Zubik, and Pratima Bose. 2001. "Phenol Antioxidant Quantity and Quality in Foods: Fruits." *Journal of Agricultural and Food Chemistry* 49 (11): 5315–21. https://doi.org/10.1021/jf0009293.

Volk, Gayle M., C. Thomas Chao, Jay Norelli, Susan K. Brown, Gennaro Fazio, Cameron Peace, Jim McFerson, Gan-Yuan Zhong, and Peter Bretting. 2015. "The Vulnerability of US Apple (Malus) Genetic Resources." *Genetic Resources and Crop Evolution* 62 (5): 765–94. https://doi.org/10.1007/s10722-014-0194-2.

Walsh, Bruce, and Jean-Luc Jannink. n.d. "Association Mapping in Plant Populations." In *Quantitative Genetics, Genomics and Plant Breeding*, edited by M. S. Kang. Cambridge: CAB International.

Wang, Nan, Shenghui Jiang, Zongying Zhang, Hongcheng Fang, Haifeng Xu, Yicheng Wang, and Xuesen Chen. 2018. "Malus Sieversii: The Origin, Flavonoid Synthesis Mechanism, and Breeding of Red-Skinned and Red-Fleshed Apples." *Horticulture Research* 5 (1): 70. https://doi.org/10.1038/s41438-018-0084-4.

Ward, G. M. 1978. *A History of the Research Station in Harrow, Ontario*. Vol. 10. Historical Series. Research Branch, Canada Department of Agriculture.

Watkins, C. B. 2003. "Principles and Practices of Postharvest Handling and Stress." In *Apples: Botany, Production and Uses*, edited by D. C. Ferree and I. J. Warrington, 585–614. Cambridge: CABI Publishing.

Weigel, Detlef, and Magnus Nordborg. 2005. "Natural Variation in Arabidopsis. How Do We Find the Causal Genes?" *Plant Physiology* 138 (2): 567–68. https://doi.org/10.1104/pp.104.900157.

Winkler, J. A., A. B. Cinderich, S. D. Ddumba, D. Doubler, J. Nikolic, Perdinan, A. M. Pollyea, D. R. Young, and C. Zavalloni. 2013. "Understanding the Impacts of Climate on Perennial Crops." In *Climate Vulnerability*, edited by Roger A. Pielke, 37–49. Oxford: Academic Press. https://doi.org/10.1016/B978-0-12-384703-4.00208-2.

Wittkopp, Patricia J., and Gizem Kalay. 2011. "Cis-Regulatory Elements: Molecular Mechanisms and Evolutionary Processes Underlying Divergence." *Nature Reviews. Genetics* 13 (1): 59–69. https://doi.org/10.1038/nrg3095.

Xie, Xianrong, Xingliang Ma, Qinlong Zhu, Dongchang Zeng, Gousi Li, and Yao-Guang Liu. 2017. “CRISPR-GE: A Convenient Software Toolkit for CRISPR-Based Genome Editing.” *Molecular Plant* 10 (9): 1246–49. https://doi.org/10.1016/j.molp.2017.06.004.

Xu, Kenong, Aide Wang, and Susan Brown. 2012. “Genetic Characterization of the Ma Locus with PH and Titratable Acidity in Apple.” *Molecular Breeding: New Strategies in Plant Improvement* 30 (2): 899–912. https://doi.org/10.1007/s11032-011-9674-7.

Yang, Jian, Noah A. Zaitlen, Michael E. Goddard, Peter M. Visscher, and Alkes L. Price. 2014. “Advantages and Pitfalls in the Application of Mixed-Model Association Methods.” *Nature Genetics* 46 (2): 100–106. https://doi.org/10.1038/ng.2876.

Yu, Cindy H. J., Zoë Migicovsky, Jun Song, and H. P. Vasantha Rupasinghe. 2023. “(Poly)Phenols of Apples Contribute to in Vitro Antidiabetic Properties: Assessment of Canada’s Apple Biodiversity Collection.” *Plants, People, Planet* 5 (2): 225–40. https://doi.org/10.1002/ppp3.10315.

Yu, Jianming, Gael Pressoir, William H. Briggs, Irie Vroh Bi, Masanori Yamasaki, John F. Doebley, Michael D. McMullen, et al. 2006. “A Unified Mixed-Model Method for Association Mapping That Accounts for Multiple Levels of Relatedness.” *Nature Genetics* 38 (2): 203–8. https://doi.org/10.1038/ng1702.

Zager, Jordan J., Iris Lange, Narayanan Srividya, Anthony Smith, and B. Markus Lange. 2019. “Gene Networks Underlying Cannabinoid and Terpenoid Accumulation in Cannabis 1.” https://doi.org/10.1104/pp.18.01506.

Zaidi, Syed Shan-E-Ali, Hervé Vanderschuren, Matin Qaim, Magdy M. Mahfouz, Ajay Kohli, Shahid Mansoor, and Mark Tester. 2019. “New Plant Breeding Technologies for Food Security.” *Science* 363 (6434): 1390–91. https://doi.org/10.1126/science.aav6316.

Zhang, Hui, Jinshan Zhang, Zhaobo Lang, José Ramón Botella, and Jian-Kang Zhu. 2017. “Genome Editing—Principles and Applications for Functional Genomics Research and Crop Improvement.” *Critical Reviews in Plant Sciences* 36 (4): 291–309. https://doi.org/10.1080/07352689.2017.1402989.

Zhang, Liyi, Jiang Hu, Xiaolei Han, Jingjing Li, Yuan Gao, Christopher M. Richards, Caixia Zhang, et al. 2019. “A High-Quality Apple Genome Assembly Reveals the Association of a Retrotransposon and Red Fruit Colour.” *Nature Communications* 10 (1): 1–13. https://doi.org/10.1038/s41467-019-09518-x.

Zhang, Xiaoyu, Gencheng Xu, Chaohua Cheng, Lei Lei, Jian Sun, Ying Xu, Canhui Deng, et al. 2021. “Establishment of an Agrobacterium-Mediated Genetic Transformation and CRISPR/Cas9-Mediated Targeted Mutagenesis in Hemp (Cannabis Sativa L.).” *Plant Biotechnology Journal* 19 (10): 1979–87. https://doi.org/10.1111/pbi.13611.

Zhang, Yanzi, Pengmin Li, and Lailiang Cheng. 2010. “Developmental Changes of Carbohydrates, Organic Acids, Amino Acids, and Phenolic Compounds in ‘Honeycrisp’ Apple Flesh.” *Food Chemistry* 123 (4): 1013–18. https://doi.org/10.1016/j.foodchem.2010.05.053.

Zhao, Keyan, Chih-Wei Tung, Georgia C. Eizenga, Mark H. Wright, M. Liakat Ali, Adam H. Price, Gareth J. Norton, et al. 2011. “Genome-Wide Association Mapping Reveals a Rich Genetic Architecture of Complex Traits in Oryza Sativa.” *Nature Communications* 2 (September): 467. https://doi.org/10.1038/ncomms1467.

Zhu, Chengsong, Michael Gore, Edward S. Buckler, and Jianming Yu. 2008. “Status and Prospects of Association Mapping in Plants.” *The Plant Genome Journal* 1 (1): 5–20. https://doi.org/10.3835/plantgenome2008.02.0089.

Zhu, Haocheng, Chao Li, and Caixia Gao. 2020. “Applications of CRISPR-Cas in Agriculture and Plant Biotechnology.” *Nature Reviews. Molecular Cell Biology* 21 (11): 661–77. https://doi.org/10.1038/s41580-020-00288-9.

Zohary, D., M. Hopf, and E. Weiss. 2012. *Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in Southwest Asia, Europe, and the Mediterranean Basin*. Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199549061.001.0001.

## Appendix I: Quantifying apple diversity: a phenomic characterization of Canada's Apple Biodiversity Collection (Chapter 2)

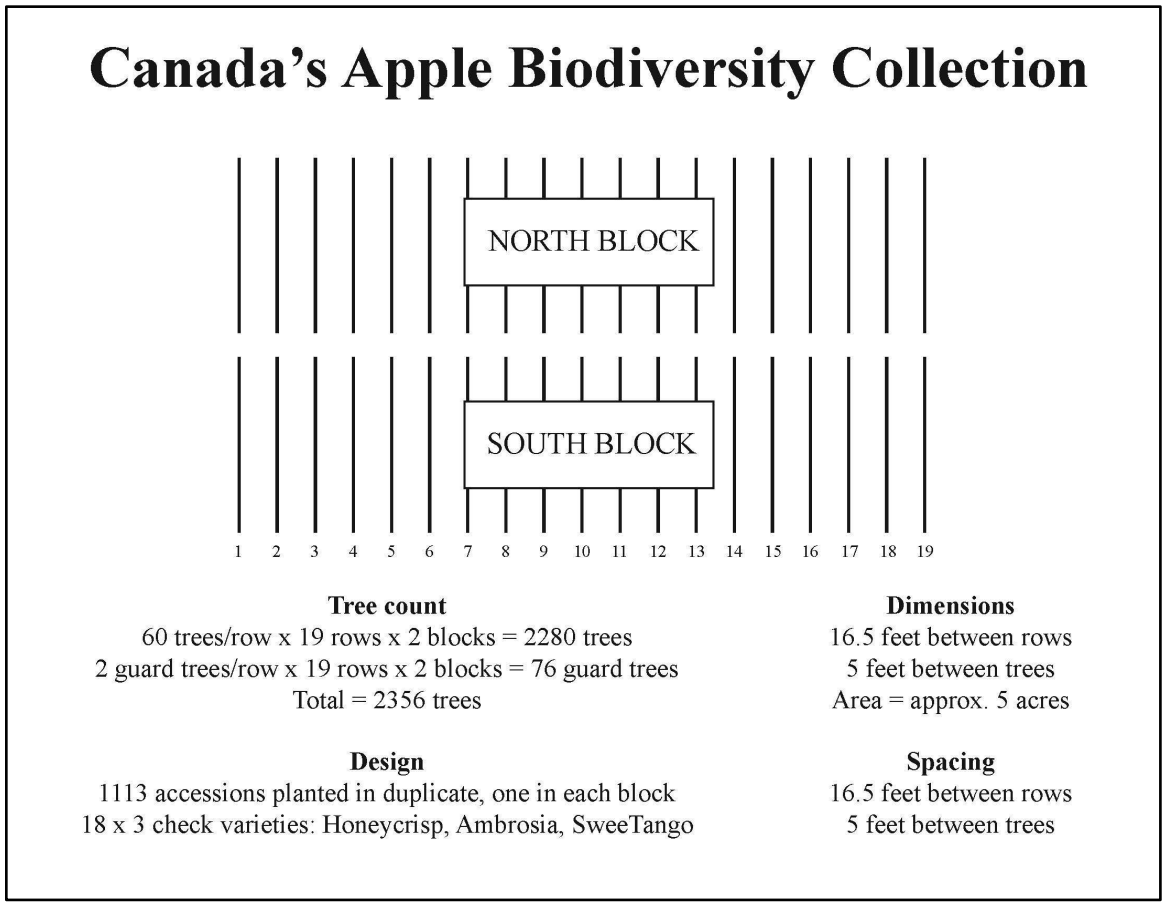


**Figure I-I.** The orchard design of Canada's Apple Biodiversity Collection.

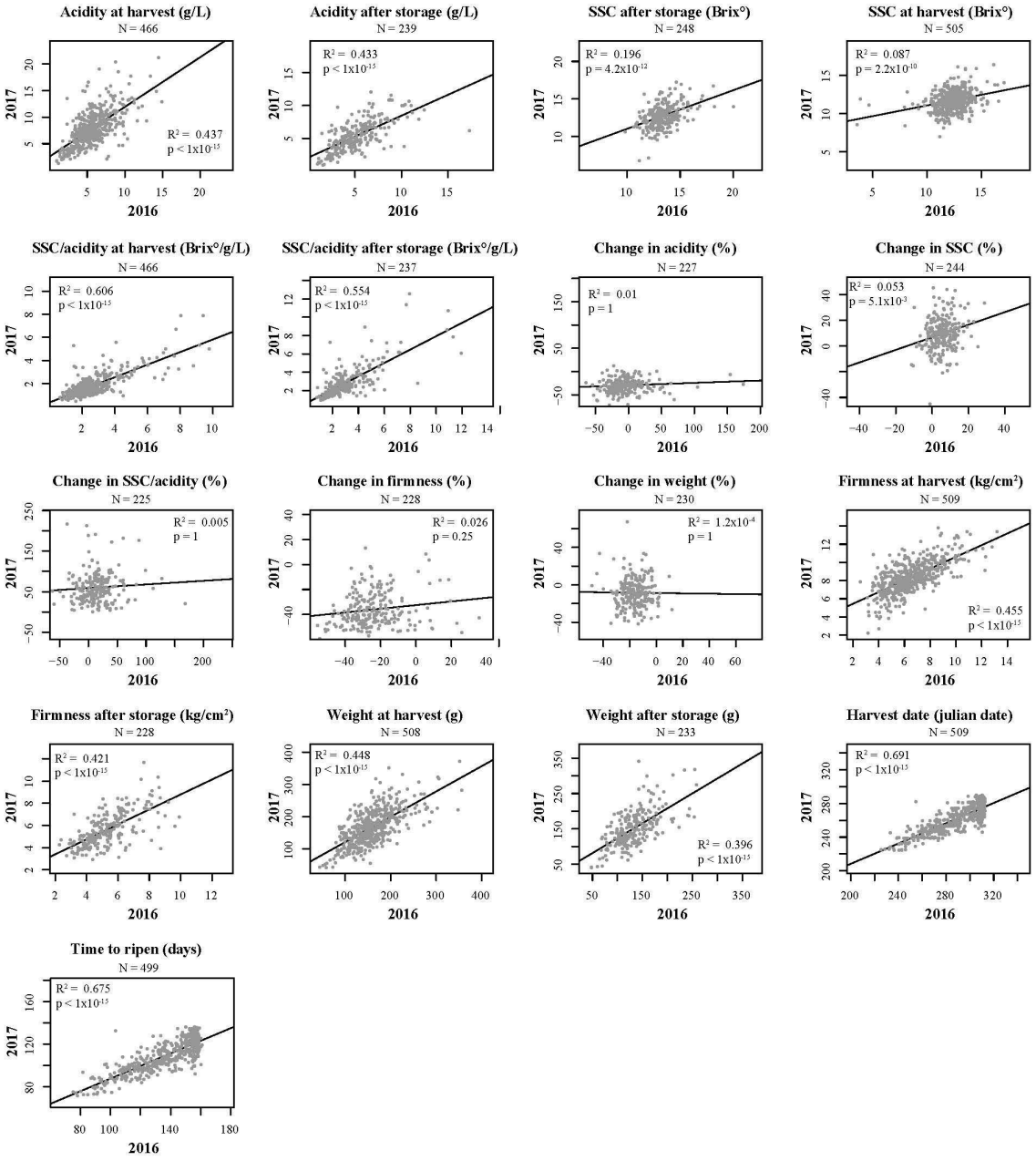


**Figure I-II.** Correlation plots of fruit quality and phenology traits between 2016 and 2017.

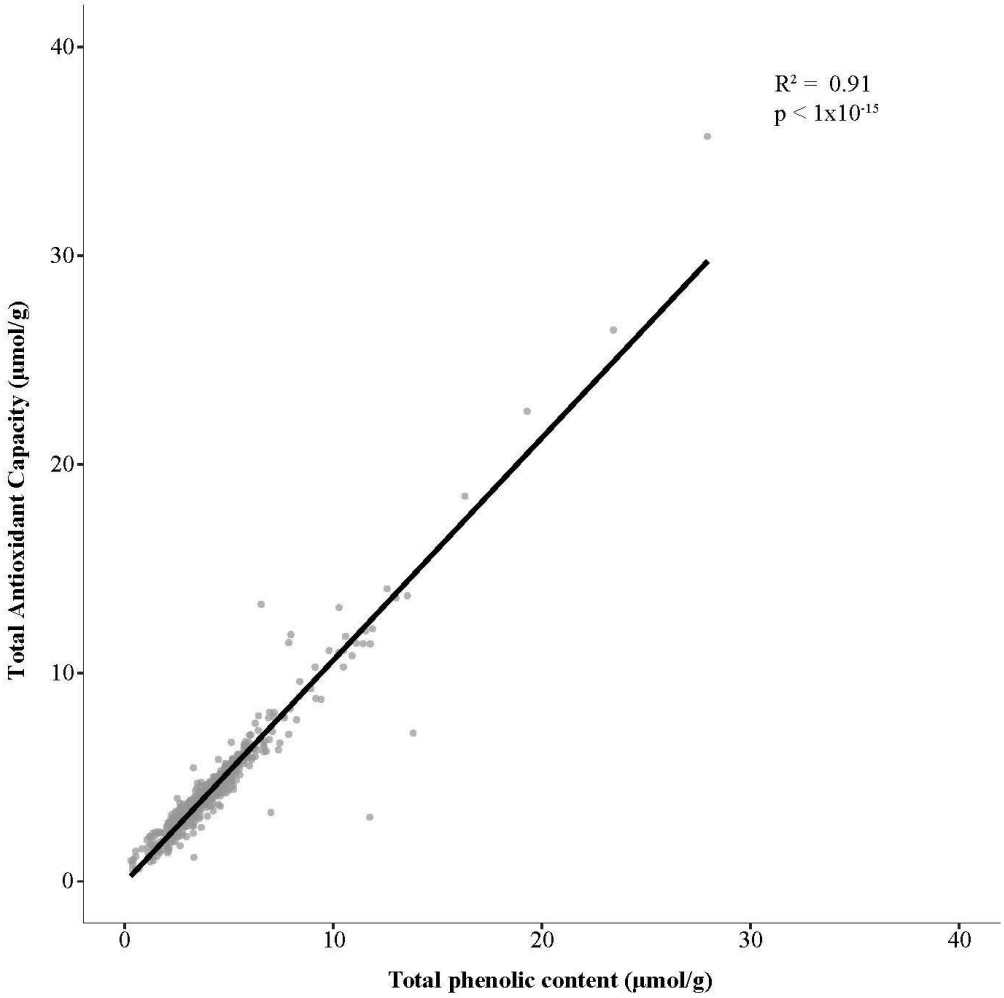


**Figure I-III.** Correlation of total phenolic content and total antioxidant capacity**.**

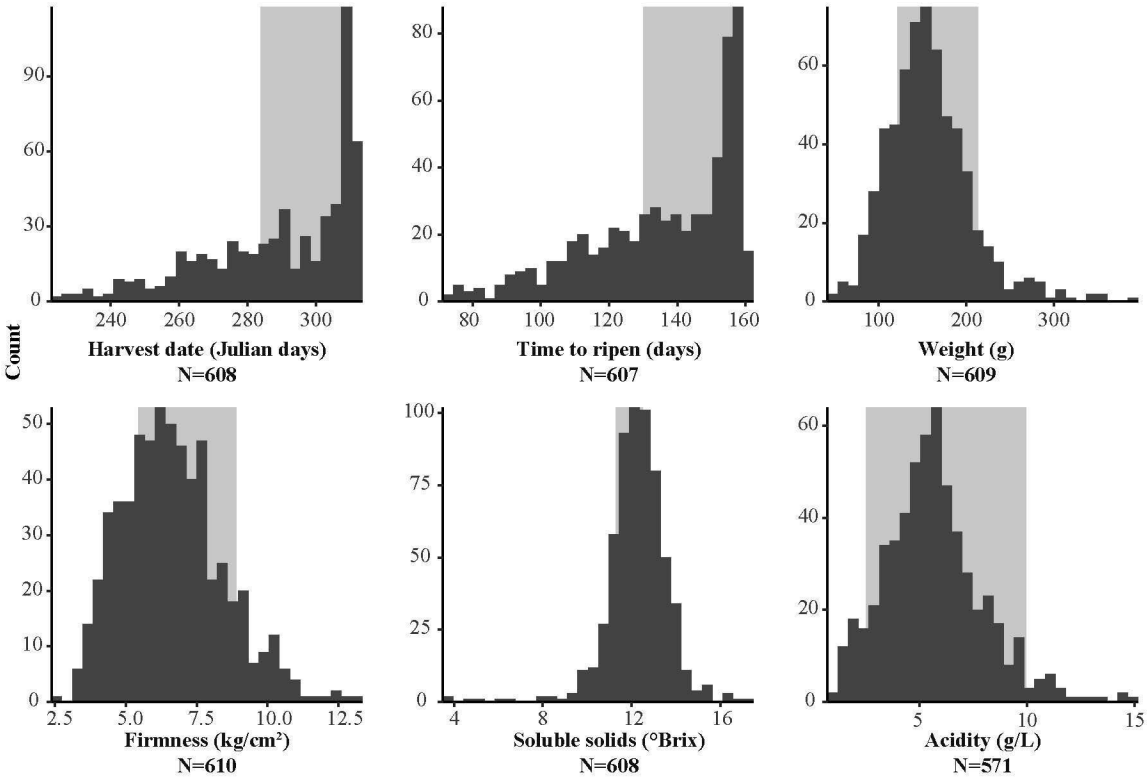


**Figure I-IV:** Distribution of apple phenotypes from 2016. Shaded grey areas represent the range of values observed across the top 9 cultivars grown in the USA.

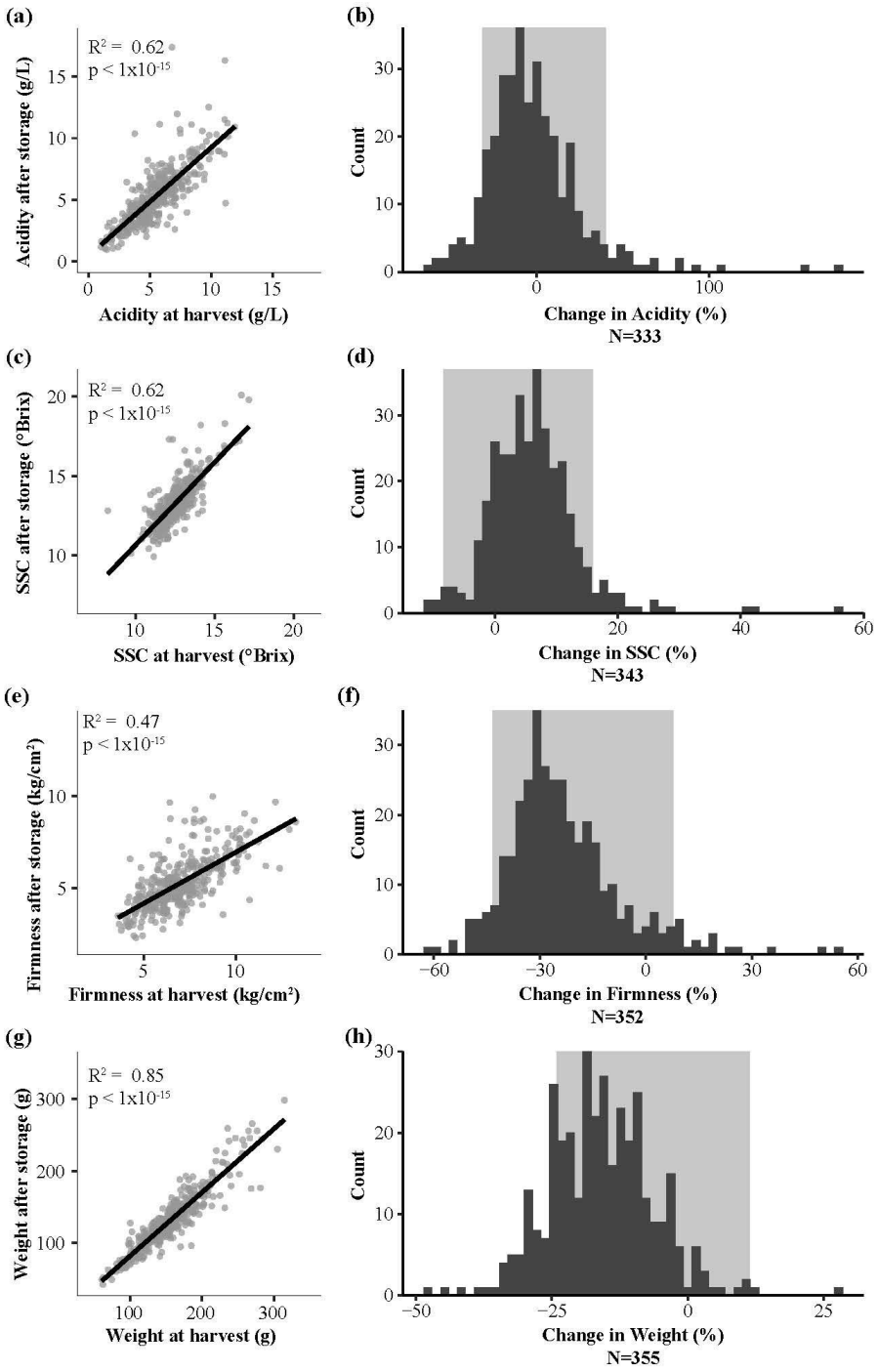


**Figure I-V.** Fruit quality measurements before and after 3 months of cold storage in 2016. Correlations of fruit quality measures taken before and after storage are found in the left column (a, c, e, g). Distribution of the percent change of each trait during storage

are shown in the right column (b, d, f, h). The shaded grey areas represent the range occupied by the "top 9" cultivars.

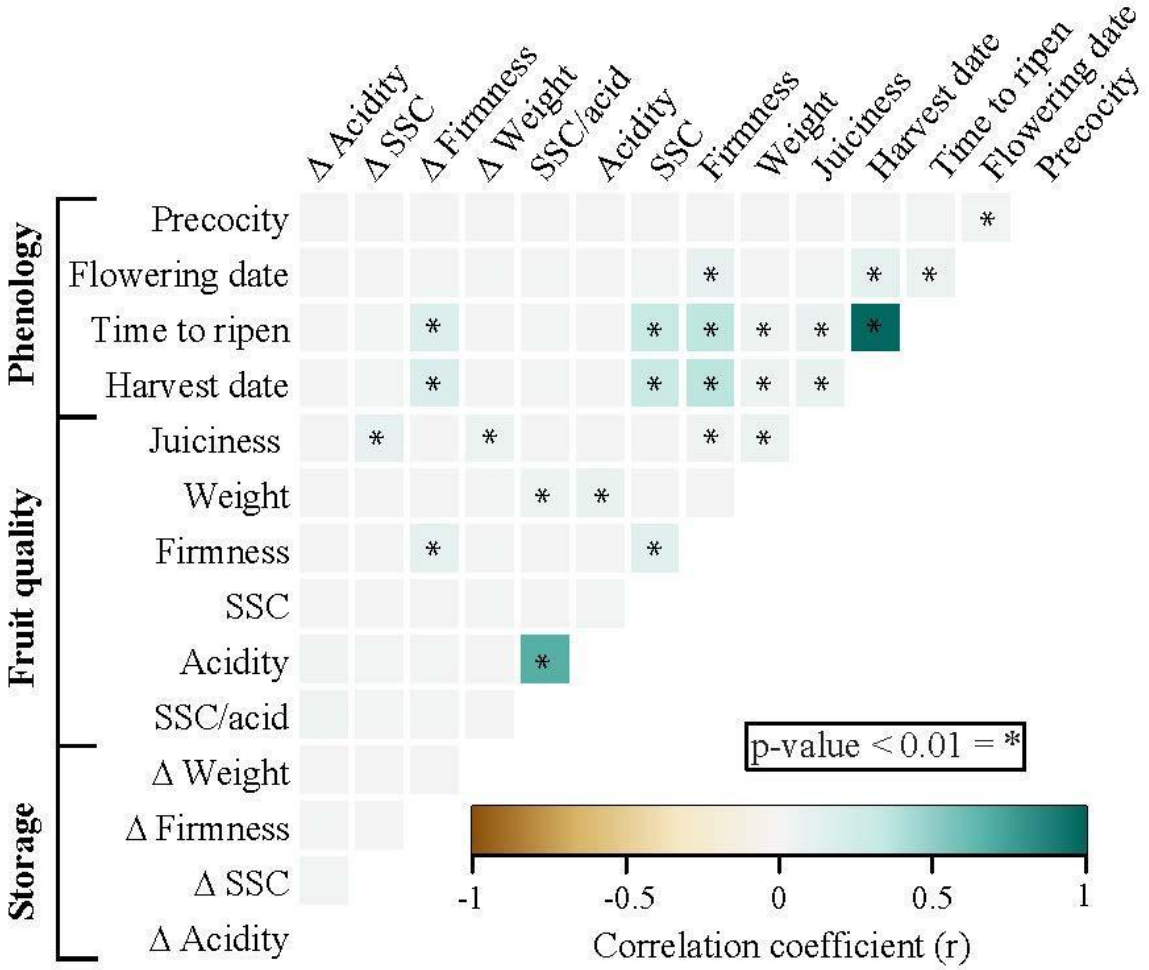


**Figure I-VI.** Heat map of pairwise correlations between phenology, fruit quality and storage traits from 2016. Significant correlations ($p < 0.01$) after correcting for multiple comparisons are indicated with an asterisk.

**Table I-I.** Phenotype data for 1119 accessions of apple from Canada's Apple Biodiversity Collection. [ELECTRONIC SUPPLEMENT]

**Table I-II.** Summary statistics for the distribution of phenotype measurements from 2017. [ELECTRONIC SUPPLEMENT]

**Table I-III.** Correlation coefficients and p-values from Pearson correlations between phenotypic trait measurements from 2017. [ELECTRONIC SUPPLEMENT]

**Table I-IV.** Correlation coefficients and p-values from Pearson correlations between phenotypic trait measurements from 2016. [ELECTRONIC SUPPLEMENT]

**Table I-V.** Median, standard error and Wilcoxon p-values for comparisons of phenotypic measurements between geographic origin, species and end use. [ELECTRONIC SUPPLEMENT]

**Table I-VI.** Correlation of release year with phenotypic traits. [ELECTRONIC SUPPLEMENT]

## Appendix II: Large-scale GWAS in apple reveals NAC18.1 as a master regulator of ripening traits (Chapter 3)

**Figure II-I.** Manhattan and QQ plots from GWAS for phenology and fruit quality traits. [ELECTRONIC SUPPLEMENT]

**Figure II-II.** Correlations between phenotypes with significant GWAS hits at NAC18.1. [ELECTRONIC SUPPLEMENT]

**Figure II-III.** Manhattan and QQ plots from GWAS with harvest date as a covariate. [ELECTRONIC SUPPLEMENT]

**Table II-I.** R-square and p-values from the correlations between PC1 values and fruit quality and phenology traits. [ELECTRONIC SUPPLEMENT]

**Table II-II.** R-square and p-values from the correlations between PC2 values and fruit quality and phenology traits. [ELECTRONIC SUPPLEMENT]

**Table II-III.** Phenotypic variance explained by the top SNPs from the GWAS. [ELECTRONIC SUPPLEMENT]

**Table II-IV.** Ripening model statistics. [ELECTRONIC SUPPLEMENT]

**Table II-V.** List of gene models within 50 kilobases of either side of the top GWAS SNPs. [ELECTRONIC SUPPLEMENT]

## Appendix III: Cannabis labelling is associated with genetic variation in terpene synthase genes (Chapter 4)

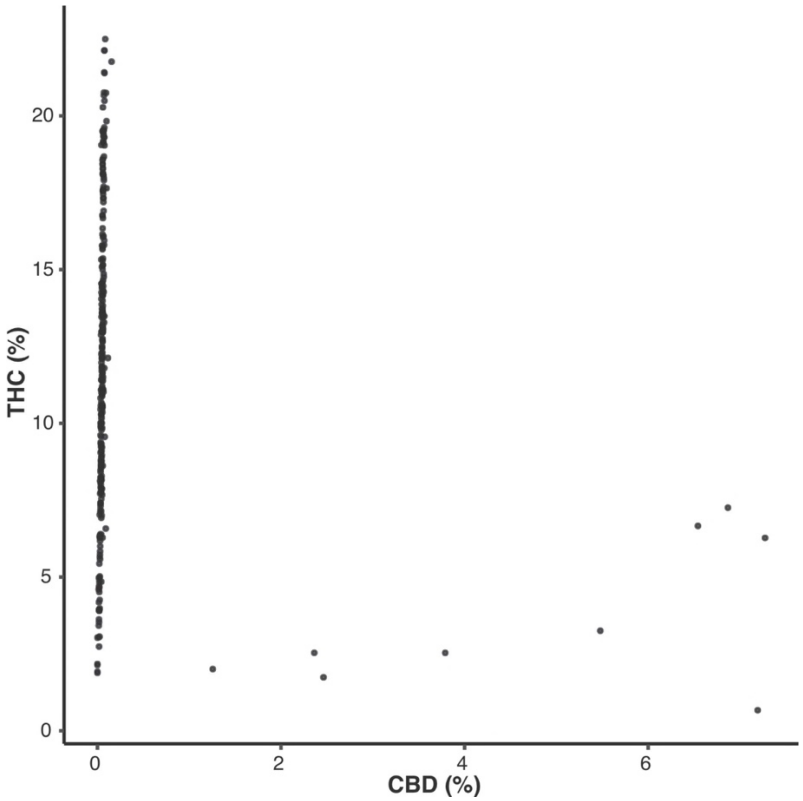


**Figure IV-I.** Plot of percent CBD versus percent THC content.

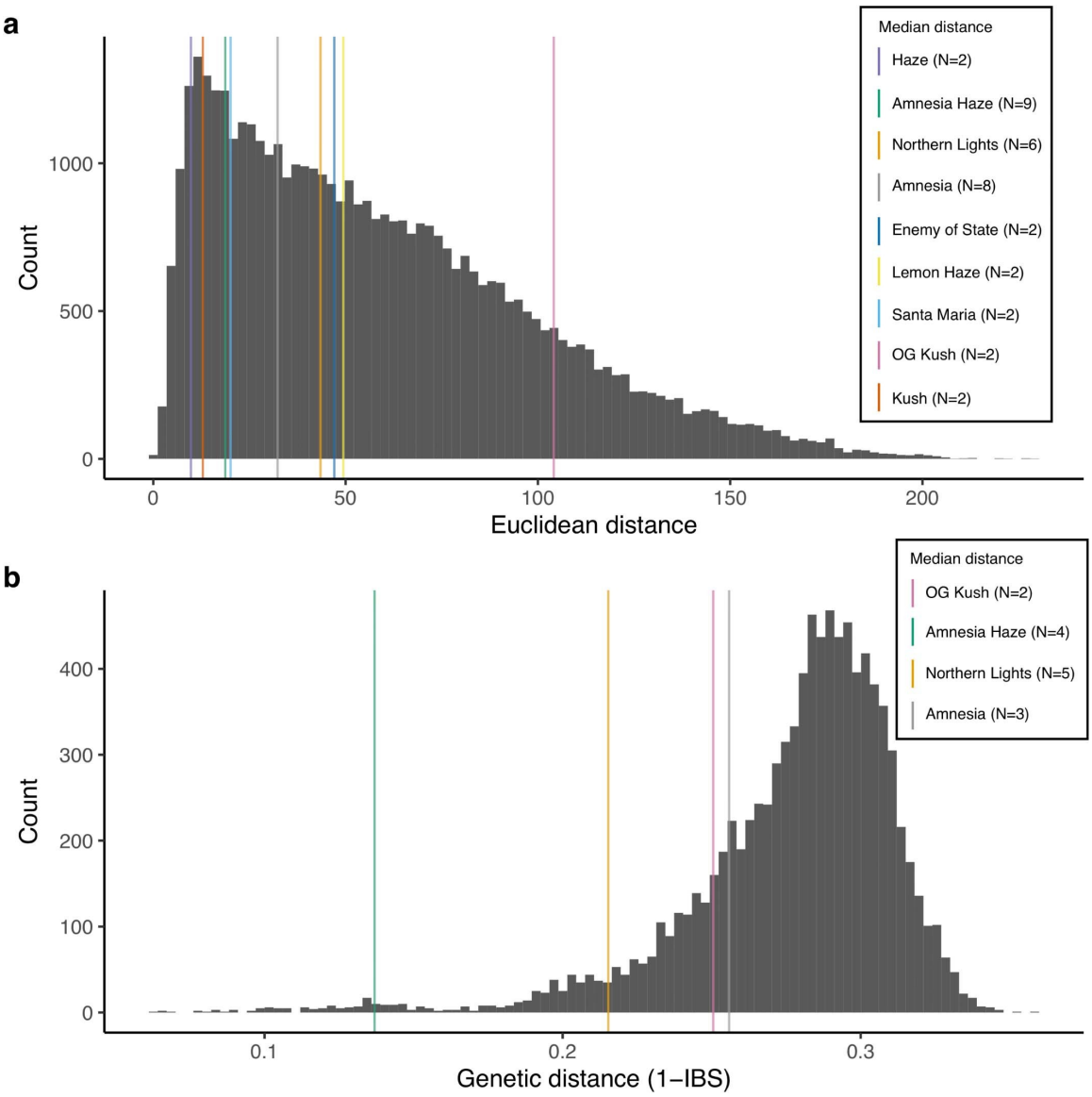


**Figure IV-II.** Histograms of **a**) pairwise chemical distances and **b**) pairwise genetic distances among all pairs of samples. Vertical lines indicate the median distance between pairs of samples with the same name.

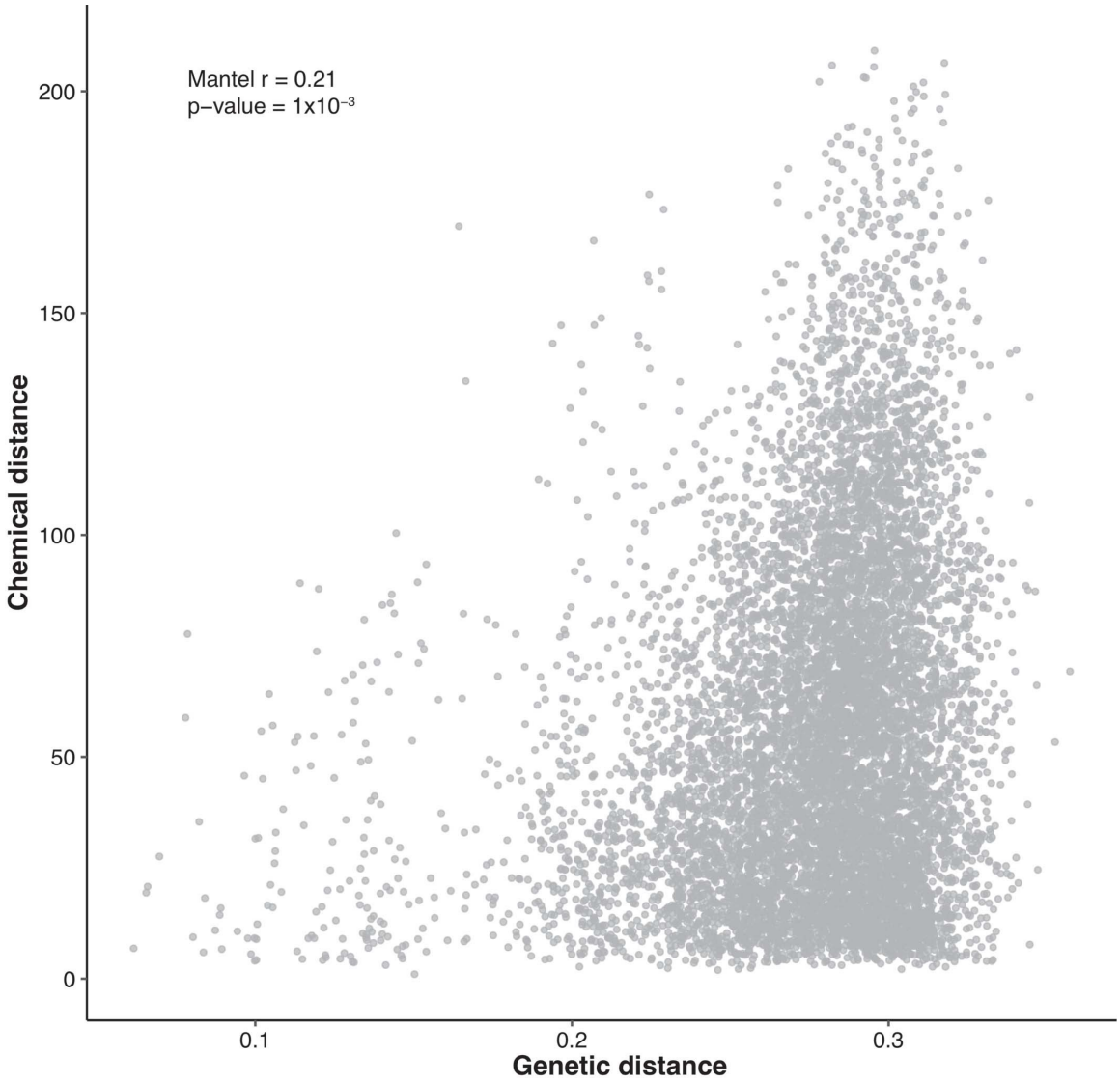


**Figure IV-III.** Plot of genetic distance versus chemical distance between pairs of samples. The Mantel r statistic and p-value are reported.

**Figure IV-IV.** Boxplots of chemical concentrations across the Sativa-Indica scale, n= 68 'Sativa', n= 27 'Sativa-Hybrid', n= 115 'Hybrid', n= 25 'Indica-Hybrid' and n= 62 'Indica' samples. The lower and upper bounds of the boxplots correspond to the 25% and 75% quantiles, the minima and maxima of whiskers extend to the lowest and largest value, respectively, within 1.5 times the inter-quartile range, and the centre of the boxes

represent the median value. Outliers beyond the whiskers are plotted individually. Y-axis is expressed as milligrams per gram. Asterisks denote chemicals with tentative identifications. [ELECTRONIC SUPPLEMENT]

**Figure IV-V.** Manhattan plots from the standard mixed linear model (MLM) and the multi-locus mixed-linear model (MLMM) GWAS and QQ plots from the standard MLM GWAS. Asterisks denote chemicals with tentative identifications. [ELECTRONIC SUPPLEMENT]

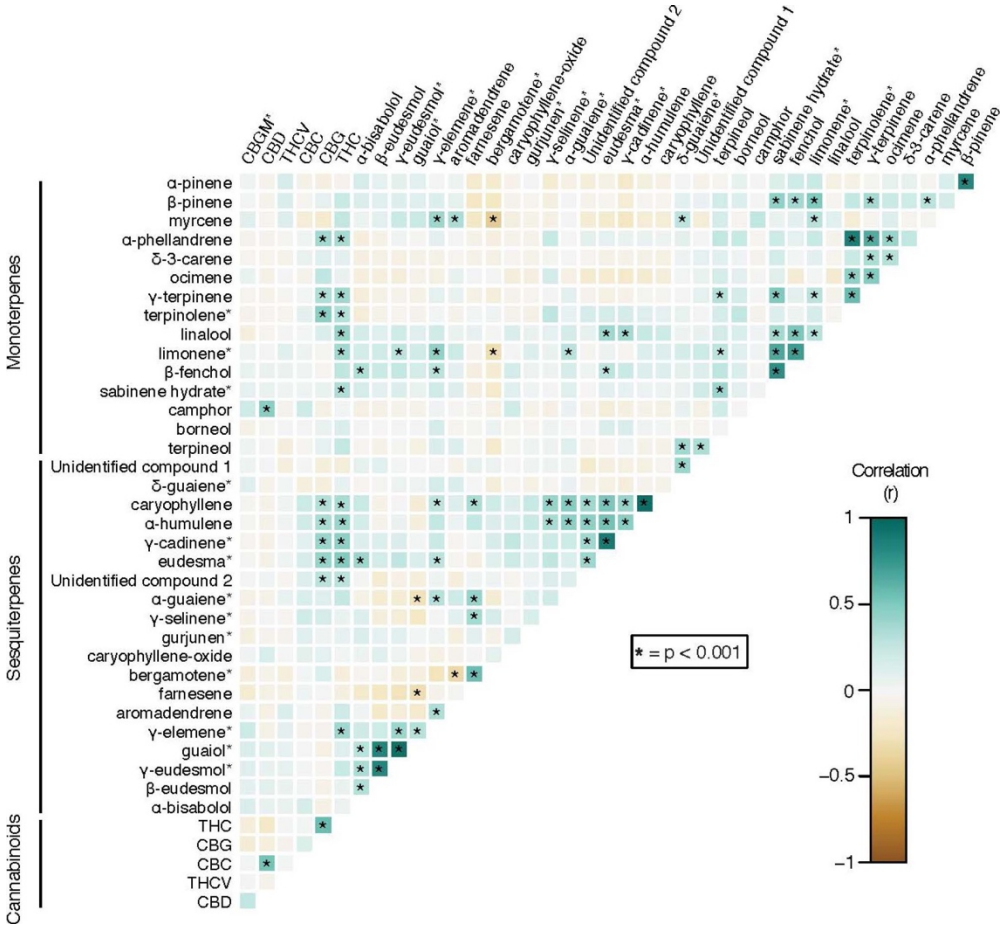


**Figure IV-VI.** Heatmap displaying the Pearson correlation between the concentrations of the 40 terpenes and cannabinoids.

**Table IV-I.** Chemical concentrations and labels across 297 Cannabis samples. The asterisks denote chemicals with tentative identifications. [ELECTRONIC SUPPLEMENT]

**Table IV-II.** A list of compound names identified by Hazekamp et al. (2016), a list of authentic standards used for compound identification and a list of compound names based on re-analysis of the methods used in Hazekamp et al. (2016). [ELECTRONIC SUPPLEMENT]

**Table IV-III.** Significant SNPs from the MLMM GWAS for myrcene and three sesquiterpenes. Only SNPs identified as significantly ($P < 6.69 \times 10^{-7}$) associated with a trait according to the MLMM GWAS are shown. The genomic coordinates and annotations, $P$ value, $R^2$ value and nearby candidate genes are shown. [ELECTRONIC SUPPLEMENT]

## Appendix IV: Copyright release

### Chapter 2: Quantifying apple diversity: A phenomic characterization of Canada's Apple Biodiversity Collection

Watts, S., Migicovsky, Z., McClure, K.A., Cindy, H.J., Amyotte, B., Baker, T., Bowlby, D., Burgher-MacLellan, K., Butler, L., Donald, R., Fan, L., Fillmore, S., Flewelling, J., Gardner, K., Hodges, M., Hughes, T., Jagadeesan, V., Lewis, N., MacDonell, E., MacVicar, L., McElroy, M., Money, D., O'Hara, M., Ong, Q., Campbell Palmer, L., Sawler, J., Vinqvist-Tymchuk, M., Rupasinghe, HP V., DeLong, J.M., Forney, C.F., Song, J., Myles, S. (2021). Quantifying Apple Diversity: Phenomic Characterization of Canada's Apple Biodiversity Collection. Plants, People, Planet, 3, 6. https://doi.org/10.1002/ppp3.10211



### Chapter 4: Cannabis labelling is associated with genetic variation in terpene synthase genes.

Watts, S., McElroy, M., Migicovsky, Z., Maassen, H., van Velzen, R., Myles, S. (2021). Cannabis labelling is associated with genetic variation in terpene synthase genes. Nature Plants, 10, 7. https://doi.org/10.1038/s41477-021-01003-y