Title

Genome-Wide Analysis of Phytophthora cactorum Resistance and Fruit Quality Traits in Cultivated Strawberry (Fragaria x ananassa)

Author

Jimenez, Nicolas Pandeli

Publication Date

2024

Peer reviewed|Thesis/dissertation

Genome-Wide Analysis of *Phytophthora cactorum* Resistance
and Fruit Quality Traits in Cultivated Strawberry (*Fragaria* × *ananassa*)

By

NICOLÁS JIMÉNEZ

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Horticulture and Agronomy

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Steven J. Knapp, Chair

Christine H. Diepenbrock

Mitchell J. Feldmann

Committee in Charge

2024

To my biological mom, Afrodita,
my adoptive mom, Evelyn,
my sister, Francina,
my partner, Miguel,
and my friends
who have been by my side at every step I have taken throughout my life.

Contents

List of Figures

List of Tables

Abstract

This thesis explores advanced strategies to enhance strawberry breeding by applying genomic prediction, harnessing underutilized genetic diversity, and investigating key genetic mutations affecting fruit quality traits. The first chapter focuses on enhancing strawberry resistance to Phytophthora crown rot (PhCR), a soil-borne disease that thrives in warm, wet conditions and severely impacts yields. It addresses the limited genetic gains in breeding for PhCR resistance and the discovery of a significant locus, RPc2, which explains a substantial portion of genetic variance. Incorporating underutilized gene bank resources doubled additive genetic variance and improved genomic prediction accuracy, highlighting the potential of genomic selection for enhancing PhCR resistance.

The second chapter further solidifies the success of these strategies, evaluating the use of useful heterosis and genomic prediction to improve fruit quality traits and shelf life in strawberries. It examines the introgression of favorable alleles from exotic genetic resources into elite breeding pools, resulting in significant improvements in total soluble solids and titratable acidity. The study not only highlights the genetic correlations among fruit traits and the impact of genotype combinations on hybrid performance but also underscores the importance of genomic prediction in efficiently selecting superior genotypes and optimizing breeding programs.

The third chapter investigates the domestication of strawberry for improved fruit firmness and shelf life. It identifies a loss-of-function mutation in the polygalacturonase gene *PG1-6A1*, significantly increasing fruit firmness and reducing perishability. This mutation and several structural variants have been positively correlated with fruit firmness and negatively with gene expression. Developing high-throughput genotyping assays for these mutations

facilitates marker-assisted selection, enabling the breeding of firmer and longer-lasting strawberry cultivars.

Collectively, these studies provide valuable insights into the genetic basis of key traits in strawberries and showcase the potential of genomic technologies to accelerate breeding efforts and improve fruit quality and disease resistance.

Acknowledgments

I want to express my deepest gratitude to everyone who supported me throughout my research journey. First and foremost, my heartfelt thanks go to the members of the Strawberry Breeding Program for their invaluable assistance. I am especially grateful to Randi Famula for her consistent help with my experiments over the years and to Marta Bjornson for her insightful guidance that helped me achieve my research objectives. My sincerest appreciation goes to my mentors, Steve Knapp and Mitchell Feldman, whose unwavering support and mentorship were instrumental in my academic and personal growth at the University of California, Davis. Thank you all for making this journey possible.

CHAPTER 1

Harnessing Underutilized Gene Bank Diversity and Genomic Prediction of Cross Usefulness to Enhance Resistance to *Phytophthora cactorum* in Strawberry

1.1. Abstract

The development of strawberry (*Fragaria* × *ananassa*) cultivars resistant to Phytophthora crown rot (PhCR), a devastating disease caused by the soil-borne pathogen *Phytophthora cactorum*, has been challenging, partly because the resistance phenotypes are quantitative and only moderately heritable. To develop deeper insights into the genetics of resistance and build the foundation for applying genomic selection, a genetically diverse training population was screened for resistance to California isolates of the pathogen. Here we show that genetic gains in breeding for resistance to PhCR have been negligible (3% of the cultivars tested were highly resistant and none surpassed early twentieth century cultivars). Narrow-sense genomic heritability for PhCR resistance ranged from 0.41-0.75 among training population individuals. Using multivariate GWAS, we identified a large-effect locus (predicted to be RPc2) that explained 43.6-51.6% of the genetic variance, was necessary but not sufficient for resistance, and was associated with calcium channel and other candidate genes with known plant defense functions. The addition of underutilized gene bank resources to our training population doubled additive genetic variance, increased the accuracy of genomic selection, and enabled the discovery of individuals carrying favorable alleles that are either rare or not present in modern cultivars. The incorporation of an RPc2-associated SNP as a fixed effect increased genomic prediction accuracy from 0.40 to 0.55. Finally, we show that parent selection using genomic-estimated breeding values, genetic variances, and cross-usefulness holds promise for enhancing resistance to PhCR in strawberry.

1.2. Introduction

Strawberry (*Fragaria* × *ananassa* Duchesne ex Rozier) plant health and production are adversely affected by a broad spectrum of diseases caused by soil-borne pathogens, including *Phytophthora cactorum* (Lebert & Cohn) J. Schröt (Erwin and Ribeiro, 1996; Paulus, 1990). This pathogen has a wide host range, a wide geographic distribution, and causes Phytophthora crown rot (PhCR) of strawberry, a disease that was not reported on strawberry until the middle of the twentieth century in Germany (Deutschmann, 1954) and thrives under warm and wet growing conditions (Erwin and Ribeiro, 1996; Wilcox, 1989). *P. cactorum* produces zoospores from oospores that can persist in soil or infected plant material for many years (Verdecchia et al., 2021), a factor that limits the utility of crop rotation for reducing the incidence of the disease (Rosskopf et al., 2005; Schneider et al., 2003). Soil fumigation with methyl bromide, an ozone-depleting gaseous chemical introduced in the 1960s to control *P. cactorum* and other soil-borne pathogens, greatly decreased the incidence of the diseases they cause, increased yields, decreased production risk, and enabled a phenomenal expansion of stawberry production (Rosskopf et al., 2005; Schneider et al., 2003; Yagi et al., 1993). The application of methyl bromide as a soil fumigant was banned by a global treaty established in 2005 to protect the ozone layer (`https://www.epa.gov/ods-phaseout/methyl-bromide`). The soil fumigants that replaced methyl bromide appear to be less effective for controlling *P. cactorum* and other soil-borne pathogens (Duniway, 2002; Rosskopf et al., 2005; Schneider et al., 2003). Pincot et al. (2022) postulated that the widespread adoption of soil fumigation with methyl bromide decreased both natural selection pressure and the imperative of breeding for resistance to Fusarium wilt, Verticillium wilt, and other diseases caused by soil-borne pathogens in strawberry. They showed that a significant percentage of the cultivars worldwide are susceptible to one or both of these wilt diseases and speculated that the frequency of susceptible cultivars increased from the 1960s onward because they could be commercially produced without disease-associated yield losses in methyl bromide-reliant production systems (Pincot et al., 2022).

The apparent scarcity of PhCR resistant cultivars reported in earlier studies (Eikemo et al., 2000; Mangandi et al., 2017; Shaw et al., 2006) suggests that, similar to the history of breeding for resistance to Verticillium wilt and Fusarium wilt in strawberry (Pincot et al., 2020), breeding for resistance to Phytophthora crown rot has either not been systematic or widespread, not produced significant genetic gains, not substantially increased the frequency of resistant cultivars, or a combination thereof. This was difficult to assess from previous reports because of the diversity of screening protocols, phenotyping methods, pathogen isolates, experiment designs, environments, population composition, and outcomes, in addition to a dearth of common resistant and susceptible checks across studies (Denoyes-Rothan et al., 2004; Eikemo et al., 2003; Mangandi et al., 2017; Parikka, 1998; Pitrat and Risser, 1977; Seemüller, 1977; Shaw et al., 2006,0; van Rijbroek et al., 1997).

The genetic variation previously uncovered for resistance to PhCR in strawberry has been quantitative with broad-sense heritability estimates in the 0.40-0.66 range and narrow-sense heritability estimates in the 0.26-0.39 range; hence, a significant fraction of the phenotypic variation previously observed for resistance to PhCR has been non-genetic (Denoyes-Rothan et al., 2004; Mangandi et al., 2017; Nellist et al., 2019; Shaw et al., 2006). Although several sources of resistance to this disease have been reported, a high percentage of the previously tested cultivars and other genetic resources worldwide appear to be susceptible, gene-for-gene resistance has not been discovered, and breeding for resistance to this pathogen has been challenging (Mangandi et al., 2017; Marin et al., 2022). Here we explore the feasibility of increasing resistance to PhCR through the application of genome-informed breeding approaches, particularly genomic selection and parent selection using genomic-estimated breeding values, genetic variances, and usefulness criteria (Allier et al., 2019; Goddard et al., 2010; Heffner et al., 2009; Lehermeier et al., 2017).

Historically, breeding for PhCR resistance has depended on phenotypic selection, and more recently on phenotypic selection combined with marker-assisted selection (MAS) targeting RPc2, a large-effect quantitative trait locus (QTL) identified by Mangandi et al. (2017).

Genomic selection could perhaps complement phenotypic selection, accelerate genetic gains, and increase the frequency of highly PhCR resistant cultivars in strawberry (Bernardo and Thompson, 2016; Goddard and Hayes, 2007; Habier et al., 2007; Heffner et al., 2009; Meuwissen et al., 2001; Poland and Rutkoski, 2016). RPc2 was discovered by QTL mapping and explained 13.7-25.3% of the phenotypic variation in a University of Florida population (Mangandi et al., 2017).

Earlier studies of the genetics of resistance to PhCR (Denoyes-Rothan et al., 2004; Eikemo et al., 2003; Mangandi et al., 2017; Nellist et al., 2019) pre-dated the development of octoploid reference genomes (Edger et al., 2019; Hardigan et al., 2021b), genotyping platforms populated with single nucleotide polymorphisms (SNPs) physically anchored to the octoploid genome, and genome-wide alignment and mapping of next-generation DNA sequences to the octoploid genome (Hardigan et al., 2020). These advances have enabled the straightforward application of octoploid genome-informed genome-wide association study (GWAS) approaches and high-resolution sub-genome specific genetic mapping of DNA variants (Hardigan et al., 2020; Pincot et al., 2022,2).

Our working hypothesis was that the Phytophthora crown rot resistance breeding problem cannot be fully or effectively solved by targeting individual QTL (Bernardo, 2008), but rather by applying genomic selection with or without the inclusion of RPc2 or other large-effect QTL as fixed effects, an approach that frequently improves genomic prediction accuracy (Bernardo, 2014; Rice and Lipka, 2019; Rutkoski et al., 2014). Our study included a multivariate genome-wide association study (GWAS) search for large-effect loci in a genetically diverse genomic selection training population (George and Cavanagh, 2015; Segura et al., 2012; Tibbs Cortes et al., 2021; Zhang et al., 2010). Lastly, we undertook this study to shed light on historic genetic gains, the impact of past breeding on genetic variation for resistance to PhCR, and the prevalence of resistance to *P. cactorum* among elite and exotic genetic resources, in addition to assessing the feasibility of enhancing resistance to PhCR through genome-informed prediction of breeding values, genetic variances, and usefulness

criteria (Crossa et al., 2017; Heslot et al., 2012; Labroo et al., 2021; Lehermeier et al., 2017; Mohammadi et al., 2015b).

1.3. Materials and Methods

1.3.1. Plant material. Collectively, 435 *F.* × *ananassa*, 18 *F. chiloensis*, and 22 *F. virginiana* individuals (asexually propagated genetic resources) were phenotyped for resistance to PhCR in our study. The *F.* × *ananassa* individuals included 64 cultivars developed at UC Davis (UCD), 282 UCD hybrids (offspring from crosses between non-inbred parents), 75 non-UCD cultivars, and 14 non-UCD hybrids. Thirty-eight individuals were phenotyped in one year, whereas 437 individuals were phenotyped in both years of our studies. The latter were used as the 'training' population for genomic prediction of breeding values and genetic variances. The training population included 321 UCD and 116 non-UCD individuals. Of the latter, 40 were wild ecotypes. The *F. chiloensis* and *F. virginiana* ecotypes were originally collected from habitats across their natural ranges in North and South America (Staudt, 1999). Other than a single *F. virginiana* subsp. *platypetala* ecotype (15X001P001) collected near the Trout Creek Campground, California (41.5°W, -121.9°W), the non-UCD genetic resources were originally acquired as single mother plants from the United States Department of Agriculture (USDA) National Plant Germplasm System (NPGS) National Clonal Germplasm Repository, Corvallis, OR (https://www.ars-grin.gov/). These individuals were multiplied from stolons in a low-elevation field (41 m) in Winters, CA, and maintained in the UC Davis Strawberry Germplasm Collection. To produce bare-root clones ('daughter' plants) for replicated testing, bare-root 'mother' plants were harvested from the low-elevation nursery in January, temporarily stored in the dark at $-3.5°C$, and transplanted to a high-elevation (1,294 m) nursery in April 2017 and 2018 (Cedar Point Nursery, Dorris, CA). Clones (bare-root plants) of each individual were harvested in mid-October of each year and stored in the dark at $3.5°C$ for two to three weeks before pathogen inoculation and planting.

1.3.2. Pathogen isolates and artificial inoculation protocols. Five isolates of *Phytophthora cactorum* (Ph9, Ph10, Ph23, Ph24, and Ph25) cultured from infected plants in coastal California were acquired from Dr. Kelly Ivors (Cal Poly, San Luis Obispo). The *P. cactorum* spore samples we used for artificial inoculation of bare-root plants were produced by California Seed and Plant (CSP) Labs (https://csplabs.com/, Pleasant Grove, CA) from a mixture of these isolates. Samples were prepared by releasing zoospores in water and creating 1×10^4 spores/mL spore solutions the day of planting. The roots of each bare-root plant (four/individual/year) were submerged in 40 ml of inoculum solution for approximately 5 minutes before being transplanted to the field. One month after transplanting, 11 g of *P. cactorum* infected oats were spread around the base of every plant in the field. This inoculum was produced by CSP using using the protocol described by Ivors (2015). The field was periodically sprinkler irrigated over the following week to promote the spread of spores and infection.

1.3.3. Field experiments. Our study populations were phenotyped for resistance to PhCR in field experiments at the UCD Plant Pathology Research Farm, Davis, CA in both years of our study (the soil is classified as a Yolo loam; https://websoilsurvey.sc.egov.usda.gov/). Field preparation and agronomic practices were identical to those previously described for Verticillum wilt resistance screening studies done side-by-side with the present study (Pincot et al., 2020). Our planting sites were pre-plant flat-fumigated by TriCal (https://trical.com/) with a chloropicrin-based fumigant (Pic-Clor 60, Cardinal Professional Products, Woodland, CA; 560 kg/ha) and sealed with a totally impermeable film tarp for one-week post-fumigation. Once the tarps were removed, fields were prepared for planting by creating 15.3 cm high raised beds with 76.2 cm of spacing between beds center-to-center and installing drip irrigation before covering the bed with with black plastic mulch. Subsurface drip irrigation was applied as needed to maintain adequate soil moisture throughout the growing season. Fertilization was done via injection through the drip system with 10-34-0 solution in fall and early spring and 32-0-0 solution in late spring and summer of each year.

Approximately 198 kg/ha of nitrogen was applied over the 2017-18 and 2018-19 growing seasons.

The study populations were planted on 10-23-2017 and 10-10-2018. The roots of each bare-root plant was submerged in the *P. cactorum* inoculum suspension (1×10^4 spores/mL) for 5 min. before transplanting into a single row in the middle of each planting bed. The between-plant spacing within rows was 30.5 cm. The individuals were arranged in 22 $\times$ 22 square lattice experiment designs with 22 individuals/incomplete block $\times$ 22 incomplete blocks (484 individuals) and four single-plant replications/individual arranged in complete blocks (replications) (Hinkelmann and Kempthorne, 2007). The R package *agricolae* (De Mendiburu and Simon, 2015) was used to assign individuals to incomplete blocks and randomize individuals within incomplete blocks.

1.3.4. Disease resistance phenotyping. The training population was visually scored for resistance to *P. cactorum* on eight different dates each year using an ordinal disease rating scale from the onset of symptoms until the resistance scores of check cultivars plateaued. We used the time-series progression of symptoms, phenotypes of check cultivars, and shapes of the phenotypic distributions to guide our phenotyping schedule. The 437 individuals selected for inclusion in the training population were phenotyped both years. We used a combination of stunting, wilting, chlorosis, and die back to score plants, where 1 = symptomless, 2 = mild stunting and wilting of outer leaves, 3 = stunted growth, wilting of outer leaves, and mild chlorosis, 4 = wilting and chlorosis throughout the plant canopy, and 5 = complete die back. The entire population was phenotyped once every 10 to 14 days from 4-13 to 6-28 in 2018 and 4-1 to 7-18 in 2019. We collected 24,600 data points and calculated the area under the disease progression stairs (AUDPS) (Simko and Piepho, 2012) from the eight time points within each year using the R package *agricolae* (De Mendiburu and Simon, 2015).

1.3.5. DNA isolation and SNP genotyping. DNA was extracted from 0.2 g of dried young leaf tissue with the E-Z 96 Plant DNA Kit (Omega Bio-Tek, Norcross, GA) according to manufacturer's instructions. To increase DNA quality and yield, Proteinase K was added

to the lysis buffer to a final concentration of 0.2 mg/ml and lysis incubation was extended to 45 min. at 65°C. Training population individuals were genotyped using these DNA samples with an Axiom^{TM} Strawberry 50K SNP array (Hardigan et al., 2020) by ThemoFisher Scientific Axiom Genotyping Services (Palo Alto, CA; https://www.thermofisher.com/). The raw genotypic data files returned by ThermoFisher Scientific were analyzed using the Axiom Analysis Suite Software v4.0.3.3 (https://www.thermofisher.com/us/en/home/life-science/microarray-analysis/microarray-analysis-instruments-software-services/microarray-analysis-software/axiom-analysis-suite.html). SNP calls for the 50K loci were filtered to identify and only include markers with well separated co-dominant genotypic clusters and identify and eliminate SNPs with minor allele frequencies < 0.05. The filtering process yielded $40,334$ SNPs.

The physical positions of the Axiom^{TM} SNPs in the octoploid genome were originally ascertained by Hardigan et al. (2020) by aligning DNA sequences for array probes to the 'Camarosa' V1 reference (FaCA1; https://phytozome-next.jgi.doe.gov/info/Fxananassa_v1_0_a1) described by Edger et al. (2019). They have since been aligned to the haplotype-phased 'Royal Royce' V1 reference (FaRR1; https://phytozome-next.jgi.doe.gov/info/FxananassaRoyalRoyce_v1_0) described by Hardigan et al. (2021b). The 'Royal Royce' physical addresses and chromosome nomenclature described by Hardigan et al. (2020) were used for GWAS analyses in the present study.

1.3.6. Statistical analyses. The resistance scores observed at the eighth time point and AUDPS metrics were analyzed using linear mixed model (LMM) functions in the R package *lme4::lmer()* (https://cran.r-project.org/web/packages/lme4/index.html; Bates et al. (2015)). The raw phenotypic data were initially analyzed using LMMs for square lattice experiment designs with a completely random effects model. Because the square lattice was found to be no more efficient than a randomized complete blocks experiment design, statistics are reported for analyses of LMMs for the latter only (Hinkelmann and Kempthorne, 2007). Estimated marginal means (EMMs) for training population individuals were estimated using

the R package *emmeans* (https://cran.r-project.org/web/packages/emmeans/; Lenth (2021)). The LMM for individual year analyses was

$$y_{ij} = b_i + G_j + e_{ij} \tag{1.1}$$

where y_{ij} is the observed phenotype for the jth genotype (individual) in the ith block, b_i is the random effect of the ith complete block, G_j is the random effect of the jth individual, e_{ij} is the ijth residual effect, $i = 1, 2, ..., 4$, $j = 1, 2, ..., n$, and n is the number of individuals. The LMM for the across-years analysis was

$$y_{ijk} = b_i + G_j + Y_k + GY_{jk} + e_{ijk} \tag{1.2}$$

where y_{ijk} is the observed phenotype for the jth genotype (individual) in the ith complete block in the kth year, Y_k is random effect of the kth year, GY_{jk} is the random effect of the interaction between the jth genotype and kth year, e_{ijk} is the ijkth residual effect, and k indexes years.

The variance components in these analyses were estimated using the restricted maximum likelihood (REML) method (Bates et al., 2015). The broad-sense heritability on a clone-mean basis was estimated by $\hat{H}^2 = \hat{\sigma}^2_G / \hat{\sigma}^2_{\overline{P}}$, where $\hat{\sigma}^2_G$ is a REML estimate of the among individuals (genotypes) variance component, $\hat{\sigma}^2_{G\times Y}$ is a REML estimate of the genotype $\times$ year interaction variance component, and $\hat{\sigma}_e$ is a REML estimate of the residual variance component, $\hat{\sigma}^2_{\overline{P}} = \hat{\sigma}^2_G + \hat{\sigma}^2_{G\times Y}/y + \hat{\sigma}^2_e/ry$ is a REML estimate of the phenotypic variance on a clone-mean basis, y is the number of years, and r is the harmonic mean of the number of replications/individual ($r = 2.23$ in 2018, 3.60 in 2019, and 5.72 across years). The harmonic mean number of replications was less than four (the number of replicates/individual originally planted) because of the loss of plants to factors other than disease. Narrow-sense genomic heritability was estimated by $\hat{h}^2 = \hat{\sigma}^2_A / \hat{\sigma}^2_{\overline{P}}$, where $\hat{\sigma}^2_A$ is a REML estimate of the genomic additive genetic variance (Endelman, 2011; Mathew et al., 2018). The coefficient of additive genetic variance was estimated by $CV_A = 100 \times \hat{\sigma}_A / \hat{\mu}$, where $\hat{\mu}$ is population mean.

1.3.7. Multivariate GWAS. The EMMs for individuals for each year were analyzed as separate dependant variables in a multivariate GWAS (MV-GWAS) using GEMMA 0.98.1 (Zhou and Stephens, 2012,1). The independent variables for these analyses were the genotypes for 39,195 Axiom^{TM} array SNP markers with alleles coded 0, 1, and 2 (Hardigan et al., 2020). The genomic relationship matrix (K) was estimated from the coded SNP marker genotypes as described by Pincot et al. (2020). The K matrix was used in MV-GWAS analyses to correct for genetic relationships among individuals (Zhou and Stephens, 2012,1). We used a 5% false-discovery rate (FDR)-corrected significance threshold to test the null hypothesis of no SNP effect (Benjamini and Hochberg, 1995). The single most significant SNPs in each cluster of one or more tightly linked SNPs were used as independent variables in multilocus genetic models to estimate the percentage of the genetic variance explained (GVE = $\hat{\sigma}^2_M/\hat{\sigma}^2_G$) and the percentage of the phenotypic variance explained (PVE = $\hat{\sigma}^2_M/\hat{\sigma}^2_{\overline{P}}$) by each locus individually and collectively corrected for other loci in the genetic model, where $\hat{\sigma}^2_M$ is a bias-corrected average semi-variance (ASV) REML estimate of the genetic variance explained by one or more SNP marker loci, $\hat{\sigma}^2_G$ is the genetic variance among clonally replicated individuals, and $\hat{\sigma}^2_{\overline{P}}$ is the phenotypic variance on a clone-mean basis (Feldmann et al., 2021). The genotype (individual) effect (G_j) was partitioned into the effects of SNP marker loci and individuals nested in SNP marker loci (the residual genetic variance not explained by SNP marker loci). The initial GWAS identified two significant SNPs for resistance score and eight significant SNPs for AUDPS, which were further analyzed. The SNP marker loci included in multilocus genetic models were AX-184879834, AX-184292487, AX-184338462, AX-184055612, AX-184127382, AX-184211829, AX-184673648, and AX-184109190 for resistance score and AX-184211684 and AX-184109190 for AUDPS. REML estimates of the variance components for these loci were used to calculate average semi-variance estimates of GVEs and PVEs for each locus (Feldmann et al., 2021). The SNP with the largest effect (AX-184109190) was shared between both traits. We subsequently repeated MV-GWAS analyses of both traits with AX-184109190 as a covariate, which eliminated the other statistically

significant signals; hence, GVEs, PVEs, and other statistics were subsequently estimated for AX-184109190 alone. The additive and dominance effects of AX-184109190 were estimated by $\hat{a} = \hat{\mu}_{AA} - \hat{\mu}_{GG}$ and $\hat{d} = (\hat{\mu}_{AG} - \hat{\mu}_{AA} + \hat{\mu}_{GG})/2$, respectively, where $\hat{\mu}_{AA}$, $\hat{\mu}_{AG}$, and $\hat{\mu}_{GG}$ are the EMMs for individuals with AA, AG, and GG genotypes for the AX-184109190 SNP locus (Falconer and Mackay, 1996; Walsh, 2001). The degree of dominance of the AX-184109190 locus was estimated by $|\hat{d}/\hat{a}|$ (Falconer and Mackay, 1996; Walsh, 2001).

1.3.8. Genomic prediction of breeding values. The statistical methods used to estimate genetic parameters and genomic-estimated breeding values (GEBVs) in the present study were previously described by Pincot et al. (2020) in a parallel study of the genetics of resistance to Verticillum wilt (Pincot et al., 2020). Narrow-sense genomic heritability (h^2) was estimated as described by Mathew et al. (2018). GEBVs for resistance score and AUDPS were estimated using three whole-genome regression methods implemented in the R package *BGLR* (`https://cran.r-project.org/web/packages/BGLR/index.html`; Pérez and de los Campos (2014)) with and without the inclusion of the RPc2-associated SNP marker AX-184109190 as a fixed effect; specifically, genomic-BLUP, the Bayesian Lasso, and Reproducing Kernel Hilbert Spaces (de Los Campos et al., 2010; de los Campos et al., 2013; Gianola and Van Kaam, 2008; Habier et al., 2013). Across-year EMMs for training population individuals were analyzed in *BGLR* by assuming a normal distribution for both traits. We used Monte Carlo cross-validation (MCCV) with 1,000 iterations to estimate the accuracy of genomic predictions of GEBVs by randomly sampling 80% of the original individuals without replacement and predicting GEBVs for the other 20% of the individuals. The estimates from individual year and across-year analyses were used to predict GEBVs for individual years and across years. Genomic prediction accuracy was estimated from estimates of the Pearson's correlation coefficient between EMMs and GEBVs (Dekkers, 2007; Dekkers et al., 2021; Van den Berg et al., 2019).

1.3.9. Genomic prediction of genetic variances and cross usefulness criteria. We used the approach described by Mohammadi et al. (2015b) to estimate genomic-estimated

genetic variances for simulated segregating populations (200 individuals/population) arising from 190,532 crosses (a factorial mating design with reciprocals) among 437 individuals in the training population (prospective parents). The across-year EMMs for resistance score, Axiom^{TM} SNP array genotypes, and a reference genetic map developed for the cultivar 'Camarosa' (Hardigan et al., 2020) were used as input in the R package *PopVar* to estimate population means ($\hat{\mu}$) and genetic variances ($\hat{\sigma}^2_G$) for each of the 190,532 simulated populations (https://cran.r-project.org/web/packages/PopVar/index.html; Tiede and Neyhart (2021)). To compare estimates of these genetic parameters for different subsets of prospective parents under phenotypic and genomic selection scenarios, we selected the 32 most resistant individuals in the training population using EMMs (phenotypic selection) and GEBVs (genomic selection). The EMM cutoff was 2.0, whereas the GEBV cutoff was 2.6. Fifteen of the selected (resistant) parents were common to both subsets, whereas 17 were unique to each subset. The cross usefulness criterion (U) was estimated by $\hat{\mu} + \hat{\sigma}_G$ (Lehermeier et al., 2017).

1.3.10. Data Availability. The data associated with this study are publicly available in a Dryad repository (https://doi.org/10.25338/B86D3M).

1.4. Results and Discussion

1.4.1. Heritability of Resistance to Phytophthora Crown Rot. The variation we observed among training population individuals for Phytophthora crown rot resistance score and area-under-the-disease-pressure-stairs (AUDPS)—a time-series metric that estimates the progression of disease symptoms (Simko and Piepho, 2012)—was continuous and approximately normally distributed, with phenotypes spanning the entire range in both years of our study (Figure 1.1). The additive genetic correlation between resistance score and AUDPS was $\hat{r}_A = 0.82$ ($p \leq 2.2e^{-16}$). The phenotypic variation was noisier in 2017-18 than 2018-19—additive genetic variance and narrow- and broad-sense heritability estimates were lower and the progression of symptoms over time was more erratic in 2017-18 than 2018-19

(Figure 1.1; Table 1.1). The inclusion of a small, albeit highly diverse collection of exotic individuals in the training population doubled or tripled genetic variation for resistance score and AUDPS within and between years. The between-year rank correlations were moderately positive and identical for both traits ($\hat{r} = 0.39$, $p \leq 2.2e^{-16}$), and individual × year interactions were non-significant—REML estimates of $\hat{\sigma}^2_{G\times E}/\hat{\sigma}^2_{\overline{P}}$ were 0.05 for resistance score and 0.09 for AUDPS (Figure 1.1; Table 1.1). Hereafter, we report across-year statistics unless noted otherwise.

TABLE 1.1. REML estimates of narrow-sense genomic heritability ($\hat{h}^2$), broad-sense heritability on a clone-mean basis ($\hat{H}^2$), the proportion of the phenotypic variance explained by the genotype × year interaction variance ($\hat{\sigma}^2_{G\times E}/\hat{\sigma}^2_P$), additive standard deviation ($\hat{\sigma}_A$), coefficient of additive genetic variance ($CV_A = 100 \times \hat{\sigma}_A/\hat{\mu}$) for resistance score and area under the disease pressure stairs (AUDPS) among $n = 437$ training population individuals phenotyped for resistance to *P. cactorum* in 2017-18 and 2018-2019 field experiments in Davis, CA, where $\hat{\mu}$ is the population mean. Statistics are shown for the training population as a whole and for UCD ($n = 321$) and non-UCD ($n = 116$) subsets of individuals in the training population.

Trait	**Population**	**Year**	$\hat{H}^2$	$\hat{\sigma}^2_{G\times E}/\hat{\sigma}^2_P$	$\hat{h}^2$	$\hat{\sigma}_A$	CV_A
Score	Training	2017-18	0.49		0.41	0.21	0.16
		2018-19	0.59		0.41	0.26	0.19
		Combined	0.67	0.05	0.57	0.26	0.18
	UCD	2017-18	0.22		0.16	0.06	0.08
		2018-19	0.44		0.29	0.14	0.14
		Combined	0.51	0.03	0.38	0.11	0.12
	Non-UCD	2017-18	0.75		0.75	0.66	0.27
		2018-19	0.80		0.56	0.56	0.28
		Combined	0.81	0.06	0.75	0.60	0.27
AUDPS	Training	2017-18	0.55		0.37	12.52	0.14
		2018-19	0.74		0.56	37.16	0.18
		Combined	0.71	0.09	0.61	24.82	0.20
	UCD	2017-18	0.37		0.36	8.01	0.11
		2018-19	0.57		0.38	15.26	0.12
		Combined	0.59	0.07	0.40	9.27	0.12
	Non-UCD	2017-18	0.69		0.52	30.56	0.21
		2018-19	0.86		0.63	78.79	0.27
		Combined	0.77	0.12	0.77	61.56	0.31

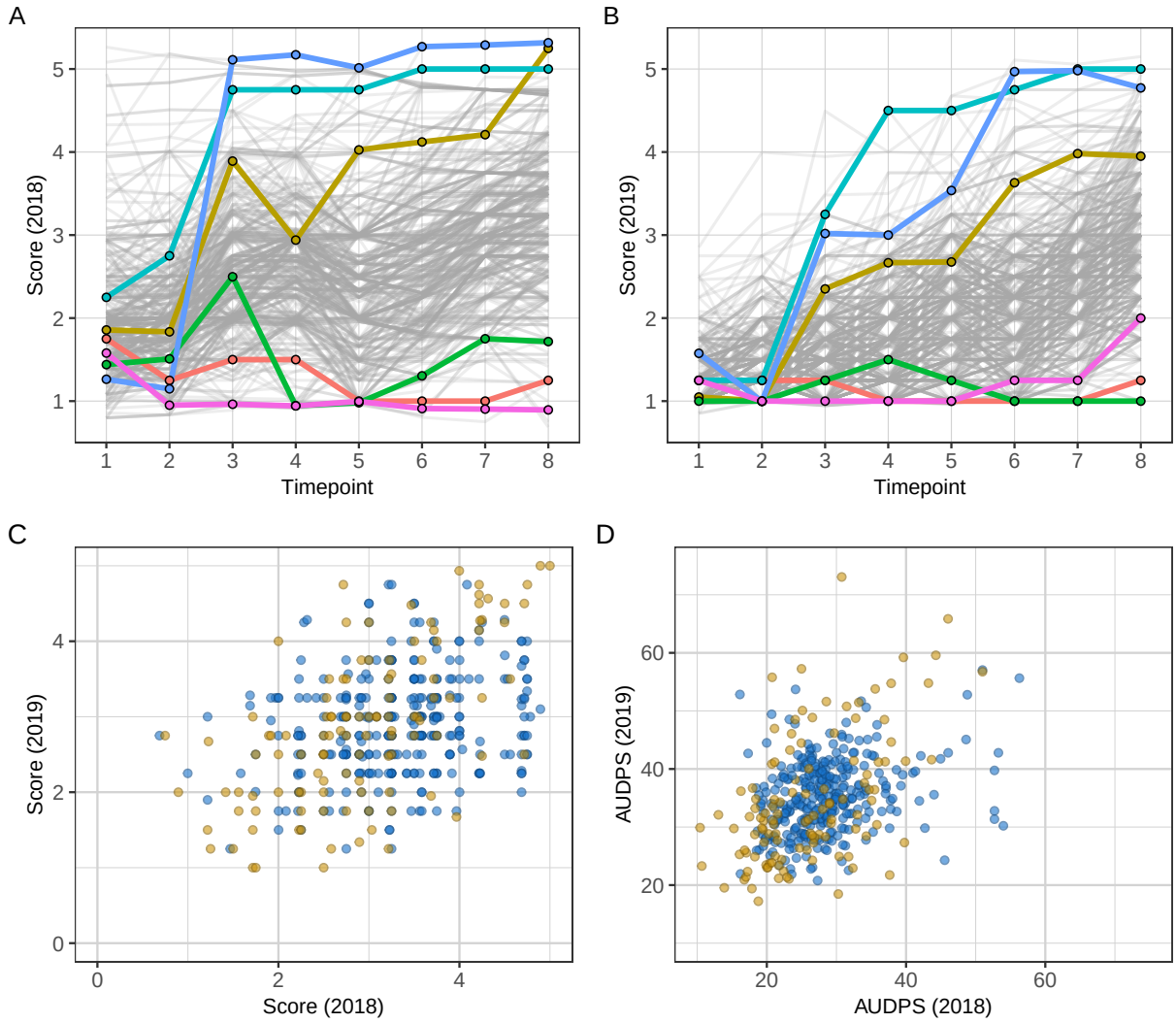

FIGURE 1.1. Phenotypic variation for resistance to PhCR in strawberry. Estimated marginal means (EMMs) were estimated for 475 individuals from three to four single plant replicates (clones)/individual/year and eight time points/year in 2017-18 and 2018-19 field experiments in Davis, CA (24,600 phenotypic observations are displayed). (A-B) EMMs for resistance score are shown for each timepoint in both years. The bold colored lines highlight resistant and susceptible individuals: blue = 'Strawberry Mountain' (PI616601), gold = 'Tamella' (PI551411), teal = 'Sitka D × Red Rich' (PI551472), green = 'Cyclone' (PI551412), red = 'Senga Sengana' (PI264680), and pink = 'N2' (PI616675). (C-D) PhCR resistance score and AUDPS EMMs for UCD individuals (blue) and non-UCD individuals (gold) in the training population are shown for both years. The between-year phenotypic correlations were $r = 0.39$ ($p \leq 0.001$) for both resistance score and AUDPS.

1.4.2. Prospective Donors of Favorable Alleles for Enhancing Resistance to Phytophthora Crown Rot. We found that 93.2% of the genetic resources (clonally propagated individuals preserved in public germplasm collections) in our study population were moderately to highly susceptible to PhCR (Figure 1.2). This conclusion was reached by screening artificially inoculated plants (stolon-propagated clones) of cultivars and other hybrid individuals from the UCD population ($n = 346$), a geographically and historically diverse collection of non-UCD heirloom and modern cultivars ($n = 89$), and a phylogenetically and geographically diverse collection of 18 *F. chiloensis* and 22 *F. virginiana* ecotypes. From previous analyses of biodiversity (Hardigan et al., 2020,2; Pincot et al., 2021), these individuals were predicted to broadly sample global diversity and included several check cultivars previously reported to be resistant or susceptible to PhCR (Bell et al., 1997; Denoyes-Rothan et al., 2004; Eikemo et al., 2003; Nellist et al., 2019; Parikka, 1998; Pérez-Jiménez et al., 2012; Pitrat and Risser, 1977; Schafleitner et al., 2013; Seemüller, 1977; van Rijbroek et al., 1997). We compiled a database of previously reported PhCR resistance phenotypes for reference and comparison. The PhCR resistance classifications for the resistant check 'Senga Sengana' (PI264680; developed in 1954) and the susceptible check 'Tamella' (PI551411; developed in 1964) have been highly consistent across studies and environments.

Of the 475 individuals screened for resistance to PhCR in the present study , 1.3% lacked symptoms and were classified as highly resistant (had across-year EMMs in the $1.0 \leq \overline{y} \leq 1.5$ range), 5.5% had mild symptoms and were classified as resistant ($1.5 < \overline{y} \leq 2.0$), and 93.2% developed moderate to severe symptoms and were classified as moderately to highly susceptible ($2.0 < \overline{y} \leq 5.0$;). The most resistant individual in our study was 'Senga Sengana' ($\overline{y} = 1.25$), an heirloom cultivar widely reported to be highly resistant (Denoyes-Rothan et al., 2004; Eikemo et al., 2003; Parikka, 1998; Pitrat and Risser, 1977; Seemüller, 1977; van Rijbroek et al., 1997). The other individuals in the highly resistant group were 'Cyclone' ($\overline{y} = 1.30$; developed in 1950), 'Addie' ($\overline{y} = 1.38$; 1982), 'MD683' ($\overline{y} = 1.38$; 1955), 12C071P602 ($\overline{y} = 1.47$; 2012), and 'Massey' ($\overline{y} = 1.50$; 1934). Thirty-two individuals had

symptom scores in the resistant to highly resistant range ($1.0 \leq \overline{y} \leq 2.0$), of which 20 were cultivars developed 40 to 99 years before present. Approximately 80% of these cultivars were developed before methyl bromide fumigation was introduced and widely adopted in the 1960s (Wilhelm and Paulus, 1980; Wilhelm et al., 1961).

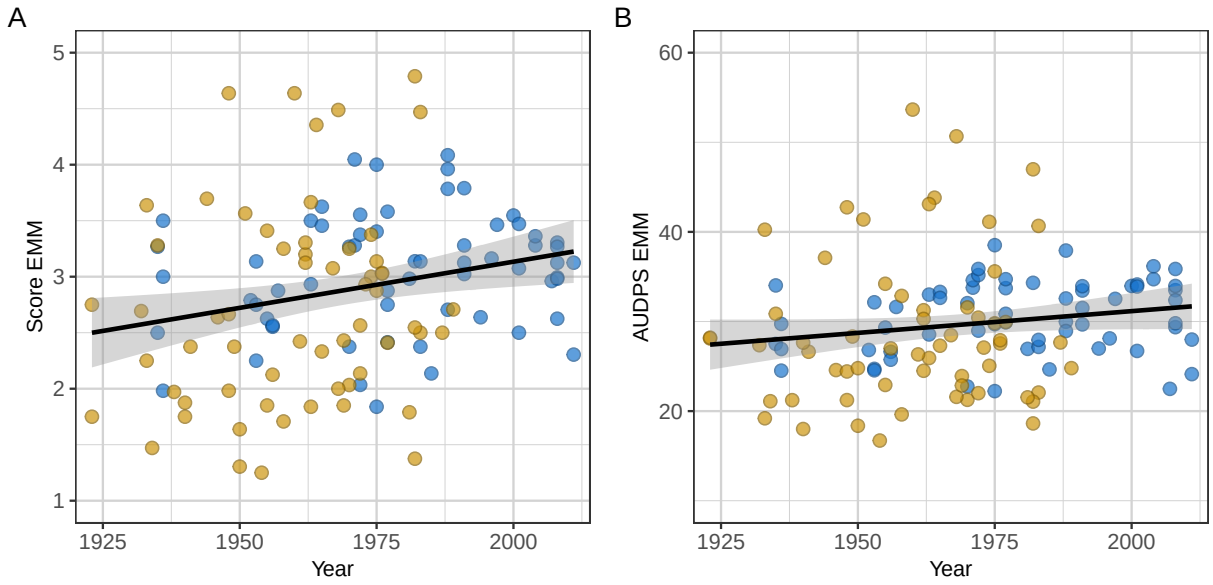

FIGURE 1.2. Estimated-marginal means for PhCR resistance score and AUDPS plotted against the year or origin for 60 UCD (blue points) and 62 non-UCD (gold points) cultivars released since 1923. Slopes for linear regressions (solid black lines) with 95% confidence intervals (gray bands) were significant for resistance score ($R^2 = 0.050$; $p \leq 0.01$) and non-significant for AUDPS ($R^2 = 0.017$; $p = 0.08$).

Of the 64 UCD cultivars screened for resistance to PhCR in the present study, only three were found to be resistant, the heirloom cultivars 'Mrak' ($\overline{y} = 1.84$; developed in 1975), 'Tahoe' ($\overline{y} = 1.98$; 1936), and 'Douglas' ($\overline{y} = 2.03$; 1972). The three modern UCD cultivars reported by Shaw et al. (2006) to be PhCR resistant ('Camino Real', 'Albion', and 'Aromas') were found to be moderately to highly susceptible in our study. The across-year EMMs for these cultivars were $\overline{y} = 2.64$ ('Camino Real'; 1994), $\overline{y} = 3.46$ ('Albion'; 1997), $\overline{y} = 3.79$ ('Aromas'; 1991). Two modern UCD cultivars reported by Shaw et al. (2006) to be PhCR susceptible were found to be susceptible: 'Ventana' ($\overline{y} = 3.16$; 1996) and 'Diamante'

($\overline{y} = 3.02$; 1991). Of the other 282 UCD individuals screened for resistance to PhCR, which represent a broad cross-section of diversity in the circa 2015 UCD population, only 2.5% were classified as resistant. As discussed below, these findings have profound implications for improving resistance to PhCR in the historically and commercially important UCD population and other elite populations that have undergone intense selection and experienced population bottlenecks (Hardigan et al., 2021b).

Finally, we found that PhCR resistance was rare among ecotypes of the octoploid progenitors of cultivated strawberry. Of the 40 *F. chiloensis* and *F. virginiana* ecotypes screened for resistance to PhCR in the present study, only one was found to be resistant ($\overline{y} = 1.56$), whereas the other 39 were found to be moderately to highly susceptible ($2.2 \leq \overline{y} \leq 5.0$). The resistant ecotype was 'N2' (PI616675), an *F. virginiana* subsp. *virginiana* individual collected from a riparian habitat along the Sainte-Anne River, a tributary of the Saint-Lawrence River in Quebec, Canada (49.1°N, -66.5°W).

1.4.3. Genetic Gains in Breeding for Resistance to Phytophthora Crown Rot Have Been Negligible Over the Last Century. Using the assemblage of 122 cultivars developed since 1923 as a barometer (Figure 1.1), genetic gains in breeding for resistance to PhCR appear to have declined over the last century (Figure 1.2). On the ordinal disease symptom rating scale we used, resistance to PhCR decreased as resistance score increased; hence, slopes from linear regressions of resistance score or AUDPS on cultivar year of origin were positive, albeit with weak coefficients of determination because phenotypic variation spanned the entire range over several decades (Figure 1.2).

On closer inspection, we found that the PhCR resistance phenotypes of cultivars developed since 1980 have drifted towards the population mean ($\overline{y} = 3.1$), with the range dropping by approximately one unit on the ordinal disease rating scale in both the upper (more susceptible) and lower (more resistant) tails of the phenotypic distribution (Figure 1.2A;). The pattern for AUDPS was even more pronounced with UCD cultivars falling in a narrower range than non-UCD cultivars (Figure 1.2B). The narrowing of these phenotypic

ranges coincided with a population bottleneck in the UCD breeding program (Hardigan et al., 2021b; Pincot et al., 2021) that appears to have fortuitously decreased genetic variation for resistance to PhCR (Table 1.1; Figure 1.1). The AUDPS range increased in both directions (towards more susceptible and more resistant) among non-UCD individuals in the training population, which is precisely what you would expect in a *random* sample of individuals from a gene bank collection.

Although we sampled a broad cross-section of heirloom and modern cultivars and common ancestors of modern cultivars to build the training population (Pincot et al., 2021), a shortcoming of our study was that plants of non-UCD cultivars developed later than 1989 were either unavailable or could not be acquired for testing from public or proprietary breeding programs in North America and Europe. Every UCD cultivar developed and released between 1935 and 2019, however, was tested . The year of origin for the 75 non-UCD cultivars tested in our study ranged from 1854 to 1989 (the median year of origin was 1965); hence, our insights into PhCR resistance were constrained by the spectrum of non-UCD cultivars tested, which were biased towards older off-patent cultivars (Figure 1.2). The frequency of PhCR resistant cultivars worldwide could obviously be greater than what we are reporting here. Although that seems improbable (Mangandi et al., 2017; Marin et al., 2022; Nellist et al., 2019), thorough sampling and screening of non-UCD modern cultivars is needed to address this question (Figure 1.2).

Our data show that a high frequency of unfavorable alleles persist in the broad cross-section of cultivars and elite and exotic genetic resources sampled, and importantly that the apparent wellspring of favorable alleles found in gene banks have either not been discovered and utilized or have simply been left behind over the last century of breeding (Figures 1.1-1.2). We suspect that the increase in the susceptibility of cultivars over time was caused by the widespread reliance on methyl bromide and other soil fumigants to suppress soilborne pathogens in many parts of the world (Rosskopf et al., 2005; Schneider et al., 2003), a consequent decrease in natural selection pressure, inattention to breeding for resistance to

PhCR and other soil-borne pathogens, and absence of initiatives to pyramid favorable alleles from multiple sources of resistance, most of which are exotic, while simultaneously preserving genetic gains for other horticulturally important traits. The latter poses a significant long-term technical challenge because of the genetic complexity of resistance to this pathogen (as discussed below) and exoticness of the most of the promising donors of favorable alleles for improving resistance to PhCR (Table 1.1). The findings reported here for PhCR resistance are consistent with our findings for Verticillium wilt, another disease caused by a soil-borne pathogen (*Verticillium dahliae* Klebahn) where the genetics of resistance is quantitative and complex, and where genetic gains in breeding for resistance appear to have been declined over the last century (Pincot et al., 2020).

1.4.4. Multivariate GWAS Uncovered the Segregation of a Large-Effect Locus. Our initial GWAS search identified nine statistically significant signals for resistance score and two statistically significant signals for AUDPS in the training population (Table 1.2; Figure 1.3A). Only one was significant for both traits, a SNP (AX-184109190) associated with a QTL on chromosome 7B that was predicted to be RPc2, a large-effect locus originally discovered by Mangandi et al. (2017) in Florida populations and environments. The RPc2-associated SNP AX-184109190 explained 28.7 to 39.7% of the genetic variance for resistance score and 30.7 to 55.6% of the genetic variance for AUDPS (Table 1.2). These percentages were estimated using bias-corrected methods and multilocus genetic models (Feldmann et al., 2021). Several of the GWAS signals observed for resistance score were weak and hypothesized be false positives. When AX-184109190 was used as a covariate in a multivariate GWAS (George and Cavanagh, 2015; Segura et al., 2012; Tibbs Cortes et al., 2021; Zhang et al., 2010), the other eight statistically significant signals decreased or disappeared altogether, and none exceeded the statistical significance threshold (Figure 1.3B). Hence, we only found statistical support for a single large-effect QTL predicted to be RPc2, and concluded that RPc2 was the only locus that merited modeling as a fixed effect in our genomic prediction study and the only locus that clearly warrants direct targeting by MAS.

TABLE 1.2. Multivariate GWAS statistics for SNP loci associated with Phytophthora crown rot (PhCR) resistance phenotypes observed in a population ($n = 437$) genotyped with a 50K Axiom^{TM} SNP array. Artificially inoculated plants (clones) of each individual were phenotyped in Davis, CA field studies in 2017-18 and 2018-19. GWAS statistics are shown for resistance score and area under the disease pressure stairs (AUDPS) within and across years. The across years statistics were estimated using the phenotypes for individual years as separate independent variables in MV-GWAS. Statisitcs for multilocus MV-GWAS analyses were estimated using the full complements of SNP marker loci shown for each trait.

						2017-18		**2018-19**		**Across Years**	
Trait	**SNP Marker**	**Chr**[a]	**Position (bp)**[b]	**MAF**[c]	***p*-value**[d]	PVE (%)[e]	GVE (%)[f]	PVE (%)	GVE (%)	PVE (%)	GVE (%)
					Multiple Loci						
Score	AX-184879834	1D	4,028,756	0.08	$2.07e^{-6}$	3.7	7.0	0.1	0.1	0.4	0.5
	AX-184292487	2B	21,279,975	0.13	$1.19e^{-7}$	6.0	11.3	3.7	5.5	4.9	6.4
	AX-184055612	3C	6,126,473	0.21	$2.75e^{-6}$	1.5	2.9	2.9	4.3	2.9	4.0
	AX-184338462	3D	4,982,835	0.15	$2.21e^{-6}$	1.1	2.1	1.6	2.4	1.9	2.4
	AX-184127382	4C	22,046,504	0.16	$6.82e^{-7}$	0.0	0.0	1.5	2.3	0.7	1.0
	AX-184211829	6B	12,105,164	0.20	$2.47e^{-6}$	1.7	3.2	4.1	6.1	3.8	5.0
	AX-184109190	7B	22,332,550	0.14	$4.60e^{-7}$	15.3	29.0	26.7	39.9	29.0	39.9
	AX-184673648	7D	16,344,212	0.14	$8.84e^{-7}$	1.8	3.5	0.6	0.9	1.1	1.5
AUDPS	AX-184211684	6B	25,289,143	0.10	$1.24e^{-6}$	4.1	7.1	0.5	0.6	1.2	1.5
	AX-184109190	7B	22,332,550	0.14	$6.34e^{-7}$	17.7	30.7	44.9	55.6	42.6	50.9
					Single Locus						
Score	AX-184109190	7B	22,332,550	0.14	$4.60e^{-7}$	22.2	38.5	29.0	43.2	33.6	43.6
AUDPS	AX-184109190	7B	22,332,550	0.14	$6.34e^{-7}$	18.1	30.9	45.4	56.2	43.4	51.6

[a]Chromosome number.
[b]SNP physical positions were ascertained in the FaRR1 haplotype-phased 'Royal Royce' reference genome.
[c]Minor allele frequency (MAF).
[d]FDR-corrected p-value.
[e]Percentage of the phenotypic variance on a clone-mean basis (PVE) explained by a locus.
[f]Percentage of the genetic variance (GVE) explained by a locus.

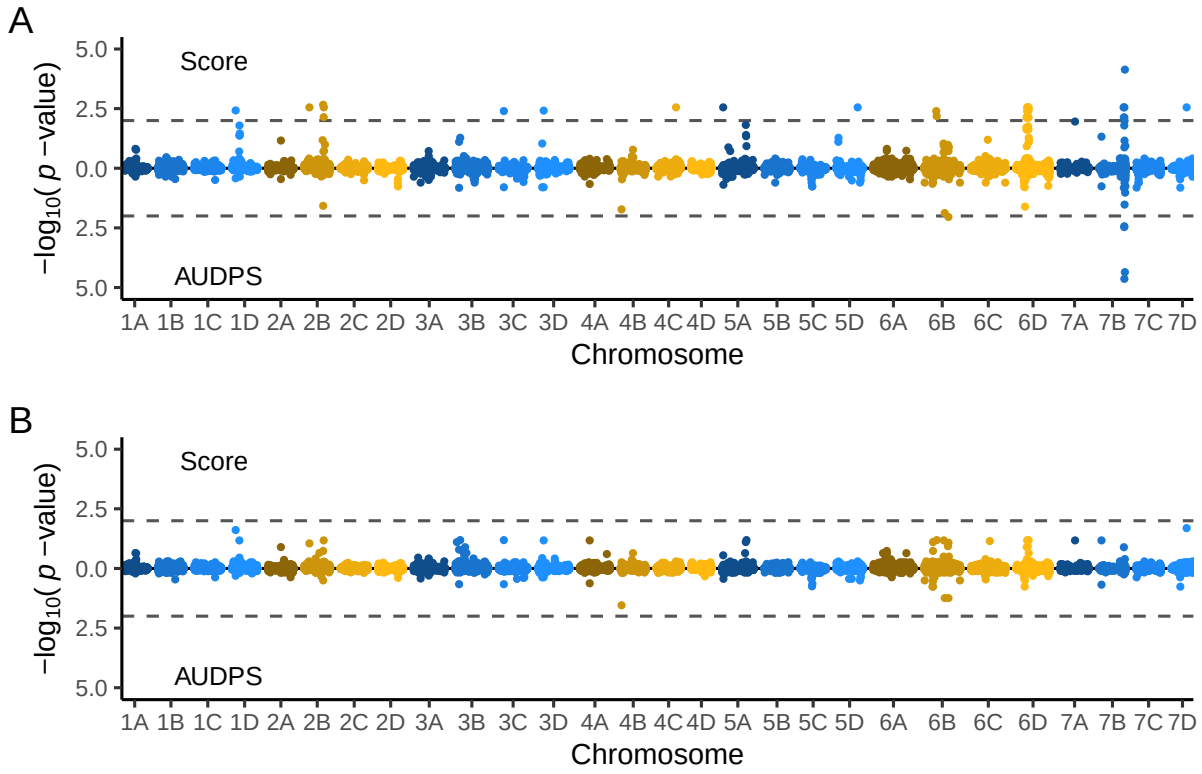

FIGURE 1.3. Genome-wide search for loci affecting resistance to Phytophthora crown rot in a population of 321 UCD and 116 non-UCD individuals genotyped with an Axiom™ 50K SNP array. Estimated marginal means for resistance score and area-under-the-disease-pressure-stairs (AUDPS) were estimated from multiple replicates within each year of the study (2017-18 and 2018-19). The within year EMMs were analyzed as separate dependent variables using multivariate GWAS. False discovery rate (FDR)-adjusted p-values are shown (the horizontal dashed lines depict 0.01 p-value thresholds). The SNP physical positions were ascertained in the haplotype-phased 'Royal Royce' reference genome (FaRR1). (A) Manhattan plot for the initial genome-wide search without covariates. (B) Manhattan plot for a genome-wide search using the RPc2-associated SNP AX-184109190 as a covariate.

1.4.5. The Dominant RPc2 Allele is Necessary But Not Sufficient for Resistance to PhCR. The SNP most strongly associated with the RPc2 locus was AX-184109190, an A/G variant (Figure 1.3; Table 1.2). We used AX-184109190 as a predictor of RPc2-associated phenotypes in the training population. The favorable SNP allele (A) was completely dominant ($|\hat{d}/\hat{a}| = 1.00$) for resistance score, nearly completely dominant for AUDPS ($|\hat{d}/\hat{a}| = 0.91$), and highly frequent among UCD ($f_A = 0.92$) and non-UCD

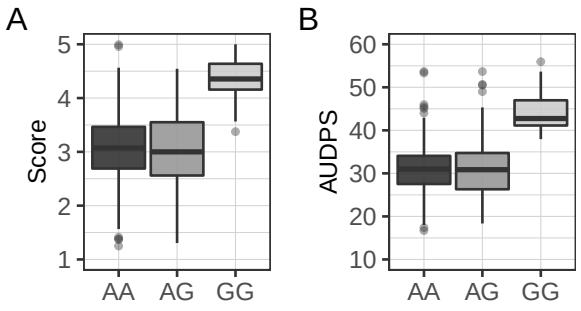

FIGURE 1.4. Estimated marginal means (EMMs) for Phytophthora crown rot resistance score and AUDPS among 437 training population individuals segregating for the RPc2-associated A/G SNP marker AX-184109190, where A is the favorable and G is the unfavorable allele. The minimum (phenotype for the most resistant individual), median, Q1-Q3 inter-quartile range, and maximum (phenotype for the most susceptible individual) are plotted for each AX-184109190 SNP genotype for both traits.

($f_A = 0.67$) individuals, where f_A is the frequency of the A allele. The additive effect of the AX-184109190 locus was $\hat{a} = -0.63$ ($p = 1.76e^{-8}$) for resistance score and $\hat{a} = -6.83$ ($p = 2.23e^{-11}$) for AUDPS. Similarly, the dominance deviation of the AX-184109190 locus was $\hat{d} = 0.63$ ($p = 9.33e^{-07}$) for resistance score and $\hat{d} = 6.24$ ($p = 8.42e^{-08}$) for AUDPS. Hence, the favorable allele increased resistance (decreased the ordinal resistance score) and slowed the progression of disease from the onset of symptoms to the point where symptoms plateaued, which spanned 10-15 weeks across years.

We found that 100% of the UCD individuals in the training population ($n = 321$) were either homozygous (A/A) or heterozygous (A/G) for the favorable SNP allele (Figure 1.4). The frequency of the favorable allele (f_A) was 0.92 among UCD and 0.67 among non-UCD individuals. The high frequency of the favorable allele in the UCD population was not expected because resistance score EMMs spanned the entire range from highly resistant ($\overline{y} = 1.41$) to highly susceptible ($\overline{y} = 4.54$) among A/A homozygotes and heterozygotes. This result

suggests that the favorable RPc2 allele is necessary but not sufficient for resistance, which was consistent with our finding that AX-184109190 explained 43.6% of the genetic variance for resistance score and 51.6% of the genetic variance for AUDPS across years (Table 1.2). The wide range of resistance phenotypes observed within AX-184109190 genotypic classes (Figure 1.4) can be attributed to several factors: non-genetic variation, genetic variation (effects of other QTL), incomplete penetrance of RPc2 alleles, and historic recombination between the SNP and the causal gene underlying RPc2 (MacLeod et al., 2014). One caveat here is that the causal mutation underlying RPc2 is as yet unknown and thus almost certainly not in complete LD with AX-184109190. Even though AX-184109190 appears to be in strong LD with the QTL, a certain percentage of resistant individuals (RPc2/RPc2 and RPc2/rpc2) carry the G allele and vice versa for susceptible individuals (rpc2/rpc2).

The high-density SNP arrays we developed for strawberry have facilitated octoploid genome-informed GWAS and the discovery of agriculturally important loci (Hardigan et al., 2020; Petrasch et al., 2022; Pincot et al., 2022); however, they only capture and interrogate a fraction of the nucleotide variants found throughout the genome (Hardigan et al., 2021b). SNP array genotyping has been incredibly powerful and important in octoploid strawberry genetic studies and facilitated the accumulation and integration of genotypic data across populations and studies with minimal genotyping errors or missing data; however, the discoveries and inferences enabled by SNP array genotypes are limited by LD between the assayed variants and causal mutations (Druet et al., 2014). With the emergence of high quality haplotype-phased genomes, strawberry is positioned to pivot to using whole-genome sequence-facilitated approaches for genomic prediction. Hardigan et al. (2021b) showed that homologous and homoeologous DNA variation can be disentangled in the octoploid using next-generation sequencing and whole-genome shotgun genotyping-by-sequencing, e.g., $>$ 80% of short DNA sequences (150 bp) can be unambiguously physically mapped (aligned) in octoploid reference genomes; hence, using whole-genome sequence data for GWAS and

genomic prediction is feasible and should accelerate the discovery of causal variants and improve the accuracy of genomic prediction across-populations in octoploid strawberry (Druet et al., 2014; Iheshiulor et al., 2016; Meuwissen et al., 2021; Raymond et al., 2018).

1.4.6. Several Immunity-Related Genes are Associated with RPc2. We used historic recombination (GWAS) in the training population to narrow the location of RPc2 down to a short DNA segment on chromosome 7B and searched the octoploid genome for annotated genes with predicted or demonstrated plant defense functions that were in strong linkage disequilibrium with RPc2 (Figure 1.3). Using a $p = 0.001$ statistical significance threshold, AX-184109190 and 10 additional flanking SNPs on the 50K Axiom^{TM} array were found to be strongly associated with PhCR resistance phenotypes. These SNPs spanned Mb 21.73-22.99 in the 'Royal Royce' reference genome (`https://phytozome-next.jgi.doe.gov/info/FxananassaRoyalRoyce_v1_0`). This 1.26 Mb segment harbors 233 annotated genes, of which 18 have immunity-related annotations: seven with homology to intracellular LRR-type immune receptors (Bonardi and Dangl, 2012), one with homology to an immunity-associated WRKY transcription factor (Pandey and Somssich, 2009), two with homology to membrane-localized immune receptors (Boutrot and Zipfel, 2017), one with homology to a Ca^{2+}-dependent protein kinase, and seven with homology to cyclic-nucleotide-gated calcium channels (CNGCs; Seybold et al. (2014)). Loss-of-function CNGC *DEFENSE, NO DEATH1* and *2* mutations have been shown to disrupt broad spectrum disease resistance and inhibit hypersensitive response cell death in Arabidopsis (Clough et al., 2000; Jurkowski et al., 2004). The presence of several CNGCs in close physical proximity to RPc2 makes them intriguing candidates for the causal gene. The gene most proximal to AX-184109190 is homologous to WAK1, another intriguing candidate for RPc2. WAK1 encodes a wall-associated receptor kinase galacturonan-binding protein, senses the presence of cell wall fragments produced during fungal or oomycete attack, and activates plant immune responses in Arabidopsis (Brutus et al., 2010).

1.4.7. Genomic Prediction Accuracy Was Strongly Affected by Population Composition and RPc2 Allele Frequency. Our genomic prediction analyses were informed by population-specific differences in additive genetic variance and heritability and the discovery that RPc2 was segregating in the training population as a whole (Table 1.1; Figure 1.3). GEBVs were separately estimated for UCD and non-UCD individuals and the complete training population with and without the RPc2-associated SNP AX-184109190 as a fixed effect using three whole-genome regression methods: genomic-best linear unbiased prediction (G-BLUP), reproducing kernel Hilbert space (RKHS), and the Bayesian Lasso (BL) (Table 1.3). Substantive differences in prediction accuracy were not observed among methods; hence, findings and conclusions are henceforth only presented for G-BLUP unless otherwise noted.

We went into this study without knowing if the RPc2 locus would segregate or have a significant effect in the training population (Table 1.2; Figure 1.3). As shown earlier, the favorable RPc2-associated SNP allele was nearly completely fixed (non-informative) in the UCD population; hence, the inclusion of the RPc2-associated SNP marker as a fixed effect decreased the accuracy of genomic predictions among UCD individuals for resistance score. Conversely, including AX-184109190 as a fixed effect increased the accuracy of genomic predictions among non-UCD and training population individuals as a whole because the exotic (non-UCD) individuals segregated for RPc2 (Table 1.2). These findings suggest that a fraction of the increase in additive genetic variation associated with the inclusion of exotic (non-UCD) individuals in the training population (Table 1.1) was caused by the introduction of RPc2 alleles (we are allowing here for the presence of multiple favorable and unfavorable alleles). What this shows, in addition, is that genetic variation associated with loci other than RPc2 was insufficient to drive genetic gains for resistance to PhCR in the UCD population without the introduction of novel favorable alleles from non-UCD sources. The prediction accuracy was only in the 0.19 to 0.24 range for resistance score and 0.28 to 0.32 for AUDPS among UCD individuals in the training population (Table 1.3).

TABLE 1.3. Accuracy of genomic prediction of Phytophthora crown rot resistance score and AUDPS breeding values in a training population of 321 UCD and 116 non-UCD individuals phenotyped in 2017-18 and 2018-19 field experiments in Davis, CA. Across-year genomic-estimated breeding value (GEBV) means ± one standard deviation were estimated from 1,000 iterations of 80/20 cross validation using genomic-best linear unbiased prediction (G-BLUP), reproducing kernel Hilbert space (RKHS), and Bayesian Lasso (BL) whole-genome regression methods applied with and without incorporating the RPc2-associated SNP AX-184109190 as a fixed effect (GEBV and GEBV + RPc2, respectively).

		G-BLUP		**RKHS**		**BL**	
Trait	**Population**	GEBV	GEBV+RPc2	GEBV	GEBV+RPc2	GEBV	GEBV+RPc2
Score	Training	0.40 ± 0.10	0.50 ± 0.08	0.41 ± 0.10	0.50 ± 0.08	0.39 ± 0.10	0.49 ± 0.08
	UCD	0.22 ± 0.10	0.19 ± 0.16	0.24 ± 0.10	0.20 ± 0.17	0.23 ± 0.10	0.19 ± 0.16
	Non-UCD	0.55 ± 0.14	0.67 ± 0.11	0.55 ± 0.14	0.67 ± 0.11	0.55 ± 0.14	0.65 ± 0.12
AUDPS	Training	0.33 ± 0.10	0.49 ± 0.08	0.34 ± 0.09	0.49 ± 0.08	0.32 ± 0.10	0.48 ± 0.08
	UCD	0.31 ± 0.10	0.30 ± 0.15	0.32 ± 0.09	0.31 ± 0.15	0.28 ± 0.10	0.30 ± 0.15
	Non-UCD	0.36 ± 0.17	0.59 ± 0.13	0.37 ± 0.15	0.60 ± 0.13	0.35 ± 0.17	0.59 ± 0.13

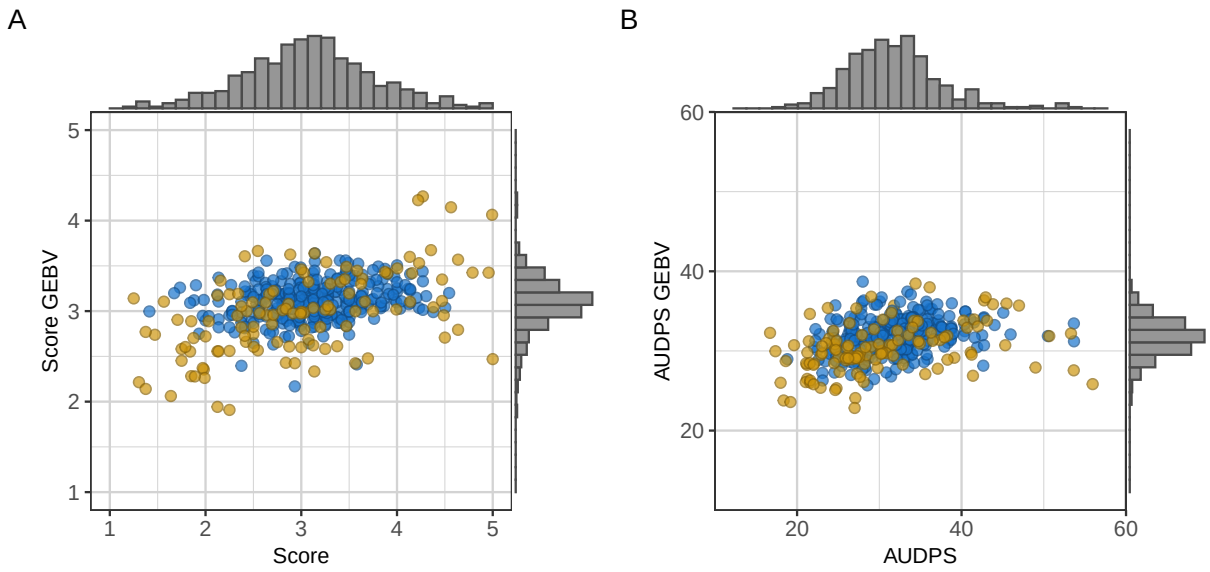

FIGURE 1.5. Phenotypic and genomic-estimated breeding value (GEBV) distributions for Phytophthora crown rot resistance score and AUDPS among 437 individuals phenotyped in 2017-18 and 2018-19 field experiments in Davis, CA. Across-year phenotypes (estimated marginal means) are shown on the x-axis. GEBV means estimated from 1,000 iterations of 80/20 cross-validation using G-BLUP are shown on the y-axis. UCD and non-UCD individuals are shown in blue and gold points, respectively.

Several noteworthy differences were observed in the EMM × GEBV distributions and genomic selection accuracy estimates between UCD and non-UCD individuals in the training population (Table 1.3; Figure 1.5). First, breeding values for UCD individuals were more substantially shrunk towards the population mean than breeding values for non-UCD individuals (Figure 1.5). GEBVs for UCD individuals clustered in a vertically narrow band centered on the population mean (3.1), whereas GEBVs for non-UCD individuals were more diffuse, spanned a much wider range (1.9-4.3), and substantially increased prediction accuracy. The prediction accuracy estimates from cross-validations were 0.22 for G-BLUP without the RPc2-associated SNP as a fixed effect among UCD individuals and 0.67 for G-BLUP with the RPc2-associated SNP as a fixed effect among non-UCD individuals. Even those UCD individuals with EMMs in the resistant range (1.0-2.0) had GEBV estimates close to the population mean (3.1) and were consequently predicted to be moderately susceptible

(Figure 1.4). Excluding one or two outliers in the lower and upper tails of the distribution, the GEBV range among UCD individuals (2.6-3.6) was exceptionally narrow (Figure 1.5). Second, the phenotypic and breeding value distributions were thin tailed (Figure 1.5). Using GEBVs as a selection metric, 19 of the 20-most resistant individuals from the lower tail of the distribution were non-UCD *F.* × *ananassa* cultivars and genetic resources (Figure 1.5). The GEBV range for those individuals was narrow (1.9-2.5). The genomic-estimated breeding value for the most resistant UCD individual (the cultivar 'Tufts') was 2.2. Using the phenotypic mean as a selection metric, the two most resistant UCD individuals were 'Mrak' ($\overline{y} = 1.84$) and 'Tahoe' ($\overline{y} = 1.98$); however, the GEBV estimates for both of those cultivars were 3.1 and thus identical to the population mean.

Third, 10 of the 12-most susceptible individuals from the thin upper tail of the GEBV distribution were from the non-UCD population, disconnected from other individuals, and had GEBVs in the 3.57-4.27 range (Figure 1.5). Seven of the 12 were *F. chiloensis* ecotypes and one was an *F. virginiana* ecotype. These findings further highlight the prediction that unfavorable alleles appear to be substantially more common in the exotic genetic resources we sampled than in the UCD population, which has a lower frequency of highly susceptible individuals. The *F.* × *ananassa* individuals from the highly susceptible tail of the GEBV distribution were the cultivars 'Tamella' (3.67) and 'Jersey Belle' (3.57) and the UCD hybrid '94C016P001' (3.64). 'Tamella' has been widely reported to be highly susceptible and was clearly susceptible in our study (Figure 1.2). Hence, the inclusion of exotic genetic resources in the training population widened the phenotypic and breeding value ranges in both directions, which increased genetic variation and genomic prediction accuracy (Tables 1.1 and 1.3; Figure 1.5). However, further investigations need to focus on whether genomic prediction accuracies using diverse populations will be durable in elite ×elite crosses. A broader range of individuals (e.g., unfavorable genotypes) in the training set may help maintain realistic prediction accuracy.

1.4.8. Revisiting the Selection of Prospective Donors of Favorable Alleles for Enhancing Resistance to PhCR. Among the more compelling story lines to emerge from our study were the scarcity of highly resistant individuals and consequential differences in genetic variation between UCD and non-UCD individuals in the training population, particularly for resistance score in 2017-18 (Table 1.1). The inclusion of a small, albeit highly diverse collection of exotic individuals ($n = 116$) in the training population doubled or tripled genetic variation for resistance score and AUDPS. Nevertheless, our estimates of the additive coefficient of genetic variation (CV_A) or "evolvability" for the complete training population (0.14-0.20) and elite UCD individuals only (0.08-0.14) highlight the challenge of breeding for resistance to PhCR in strawberry (Table 1.1).

Several of the PhCR-resistant heirloom cultivars identified in strawberry have roots in early twentieth century breeding for resistance to red stele or Lanarkshire disease in Scotland where that disease was initially discovered, heavy clay and cool wet soils are common, and the causal pathogen *P. fragariae* thrives (Adams et al., 2020; Wardlaw, 1927). Weaving back through the breeding history (Eikemo et al., 2003,0; Pincot et al., 2021; Van de Weg, 1997) we discovered that many of the cultivars found to be resistant to PhCR have common ancestors previously shown to be resistant to red stele, most notably 'Frith' and 'MD-683'. Frith was a parent or more distant ancestor of several red stele resistant cultivars that were found to be resistant to PhCR in the present study, e.g., 'Climax' (2.38), 'Red Gauntlet' (2.64), and others with Auchincruive ancestry. 'Fairfax' (1.75) and 'Scotland BK-4', a descendant of 'Frith', are parents of 'MD-683' (1.38), a highly resistant parent found in the ancestry of several cultivars shown to be PhCR resistant, e.g., 'Addie' (1.38), 'Stelemaster' (1.64), 'Delite' (1.71), 'Tribute' (1.79), and 'MDUS-5097' (1.89) (Pincot et al., 2021). Several other descendants of 'Fairfax' were found to be resistant to PhCR in the present study, notably 'Empire' (1.75), 'Tribute' (1.79), 'Bounty' (1.84), 'Hood' (1.85), 'Jewel' (1.85), 'Fairland' (1.97), 'Cavalier' (1.98), and 'Red Giant' (2.00) . The interconnections here are intriguing because the development of the cultivars resistant to *P. fragariae* in the early twentieth

century pre-dated the emergence of *P. cactorum*-caused strawberry diseases in the middle of the century (Deutschmann, 1954; Wardlaw, 1927). The correlation seems more than coincidental, however, the genetic mechanisms underlying resistance to these pathogens are different: gene-for-gene resistance to red stele has been rigorously documented and shown to be widespread (Van de Weg, 1997), whereas resistance to PhCR appears to be quantitative, albeit strongly affected by the segregation of the large-effect QTL RPc2 (Table 1.1; Figure 1.4; Mangandi et al. (2017)). One or more 'black box' QTL underlying quantitative resistance to both of these pathogens could be shared but undiscoverable or simply not yet discovered. Moreover, favorable QTL alleles underlying quantitative resistance to red stele could be masked by the dominance of race-specific resistance genes (Van de Weg, 1997).

The breeding value estimates for cultivars trended upward (towards greater susceptibility) from 1960 onward, with few cultivars released after 1960 falling in the resistant range (1.0, 2.0) for either EMMs or GEBVs (Figure 1.2). The eight most highly resistant cultivars ($1.25 \leq \overline{y} \leq 1.75$) in our study, with one exception ('Addie', a cultivar developed in 1981), originated between 1923 and 1958. Among UCD individuals in the training population, the cultivars 'Tufts' ($\overline{y} = 2.93$; GEBV = 2.17; 1963), 'Capitola' ($\overline{y} = 2.38$; GEBV = 2.40; 1983), and 'Santana' (GEBV = 3.58; 1977) and the hybrid 65C065P601 ($\overline{y} = 2.71$; GEBV = 2.63; 1965) had the lowest GEBVs and were consequently predicted to be the most resistant. Using GEBVs as a selection metric, the three most resistant individuals in our study were the non-UCD cultivars 'Sparkle' ($\overline{y} = 2.25$; GEBV = 1.91), 'Earlimiss' ($\overline{y} = 2.13$; GEBV = 1.94), and 'Stelemaster' ($\overline{y} = 1.64$; GEBV = 2.06). Conversely, using phenotypic EMMs as a selection metric, the three most resistant individuals in our study were the non-UCD cultivars 'Senga Sengana' ($\overline{y} = 1.25$; GEBV = 3.14), 'Cyclone' ($\overline{y} = 1.30$; GEBV = 2.21), and 'Addie' ($\overline{y} = 1.38$; GEBV = 2.77). While the divergence between observed phenotypes and GEBVs was pronounced for some of the individuals documented here (e.g., 'Senga Sengana'), the degree of shrinkage observed towards the population mean was expected. 'Senga Sengana' and 'Cyclone' are especially illustrative examples of validated resistant

genetic resources where the divergence between phenotypes and GEBVs markedly differed. The outcome for these two cultivars clearly highlights the merits of selecting prospective donors of favorable alleles using phenotypic means and GEBVs when applying selection to individuals in the training population, versus the more challenging problem of applying selection to non-training population individuals that have not been phenotyped (Müller et al., 2015). Selection on GEBV alone here would exclude 'Senga Sengana', the most important benchmark of resistance identified in the present and previous studies. This sort of hedging seems prudent when selecting parents or prospective donors of favorable alleles because of the uncertainty associated with both phenotypic and breeding value estimates (Tables 1.1 and 1.3).

The phenotypic and breeding value differences observed between UCD and non-UCD individuals in the training population (Figures 1.1 and 1.5) were aligned with insights gained from previous genome-wide studies of nucleotide diversity. Using diverse genetic resources, (Hardigan et al., 2021b) showed that strong directional selection, breeding bottlenecks, and selective sweeps had progressively decreased genetic variation in the UCD population. The decrease was hypothesized to have been driven by significant genetic gains for agriculturally important traits (e.g., fruit yield, size, and firmness) over nearly 70 years of selection in the UCD population, combined with the fixation and loss of alleles through random genetic drift and hitchhiking in selective sweeps (Hardigan et al., 2021b). Our data suggests that resistance to PhCR and Verticillium wilt has declined over the last half century in the UCD population through 2012, the oldest generation analyzed in our studies (Figure 1.1; Pincot et al. (2020)). The non-UCD individuals phenotyped in the present study appear to harbor favorable alleles (lower tail of the GEBV distribution) and unfavorable alleles (upper tail of the GEBV distribution) that are not present in the UCD population and are rare among genetic resources preserved in gene banks (Figure 1.5).

1.4.9. Prospects for Improving Parent Selection and Increasing Genetic Gains Through Genomic Prediction of Genetic Variances and Cross Usefulness Criteria. Although the gene bank diversity we sampled (primarily non-UCD heirloom cultivars and octoploid ecotypes) substantially increased genetic variation, novel favorable alleles appear to be scared and, therefore, highly dispersed (Figure 1.1). Their introduction and accumulation in elite populations will require several generations of hybridization and recombination, which could be accelerated by applying genomic selection (Crossa et al., 2010; Labroo et al., 2021; Poland and Rutkoski, 2016). Our evolvability (CV_A) estimates (Table 1.1) predict that segregating populations developed with parents chosen at *random* from the training population have a low probability of "producing phenotypic variation that is both heritable and adaptive" (Payne and Wagner, 2019; Pigliucci, 2008). This conclusion is consistent with our estimates of historic genetic gains for resistance to PhCR (Figure 1.5) and estimates of narrow-sense heritability in elite California and Florida populations (Table 1.1; Mangandi et al. (2017)). While parents would obviously not be randomly chosen in practice, the low frequency of outstanding parents for PhCR resistance (individuals that have accumulated favorable alleles for multiple QTL underlying resistance to PhCR) drastically limits the choice of prospective parents, a preponderance of which have been selected for traits *other* than PhCR resistance, e.g., out of 95,266 possible crosses among 437 individuals (excluding reciprocal crosses) in the training population, only 15 or 0.0157% would be between pairs of prospective parents predicted to be highly resistant (using phenotypic means as the selection criteria). The percentage predicted from GEBV estimates was virtually identical. To take the parent selection problem one step further, additive genetic variances and usefulness criteria (Allier et al., 2019; Lehermeier et al., 2017) were estimated for phenotypic and genomic selection schemes by simulating segregating populations (full-sib families of 200 individuals each) for all possible crosses (190,532 including reciprocals) among 437 prospective parents (individuals) in the training population (Figures 1.6-1.7). The latter were split into resistant (R) and susceptible (S) groups using across-year phenotypic EMM and GEBV

estimates as selection criteria . The truncation selection cutoffs were EMM < 2.0 for phenotypic selection and GEBV < 2.6 for genomic selection. There were 32 selected parents in both resistant groups (the selected fraction was $32/437 = 0.073$) with an intersection of 15 parents between resistant groups (49 different parents were selected between methods). These analyses yielded classic isosceles triangle-shaped $\hat{\mu} \times \hat{\sigma}_A^2$ distributions similar to the those observed for many agriculturally important traits under directional selection in other domesticated plants (Lado et al., 2017; Mohammadi et al., 2015b). As predicted, crosses between highly resistant parents (R $\times$ R) or highly susceptible parents (S $\times$ S) had the smallest predicted genetic variances, whereas crosses between highly resistant and highly susceptible parents (R $\times$ S) had the largest predicted genetic variances (Figure 1.6). Naturally, as selection intensity increased among individuals with EMM < 2.0, the genetic variance range (altitude of the triangle) and maximum (apex of the triangle) decreased, where EMM here was predicted population mean estimated from 200 individuals/simulated full-sib family.

The subsets of crosses shown in the $\hat{\mu} \times \hat{\sigma}_A^2$ distribution were chosen to study the outcomes of different phenotypic or genomic selection scenarios where resistant UCD and non-UCD individuals were selected and crossed to every individual in the training population or subsets of resistant individuals. The breeding scenarios we explored produced several insights. First, broadly speaking, crosses between resistant non-UCD parents and all other parents unlocked more genetic variation and were predicted to be more resistant (had lower population means) than crosses between UCD parents and all other parents (Figure 1.6A-B). This was fully expected and predicted by the differences observed in the whole-genome regression (WGR) shrinkage of breeding values discussed earlier (Figure 1.5). Second, the genetic parameter estimates for crosses with resistant UCD parents (blue points) were shifted upwards (increased genetic variance) and leftwards (decreased population mean) for genomic selection (Figure 1.6B) compared to phenotypic selection (Figure 1.6A). This pattern suggests that genomic selection on cross usefulness should produce greater gains than phenotypic selection (Figures 1.6-1.7). Third, we observed a distinct band of prospective crosses to the right of

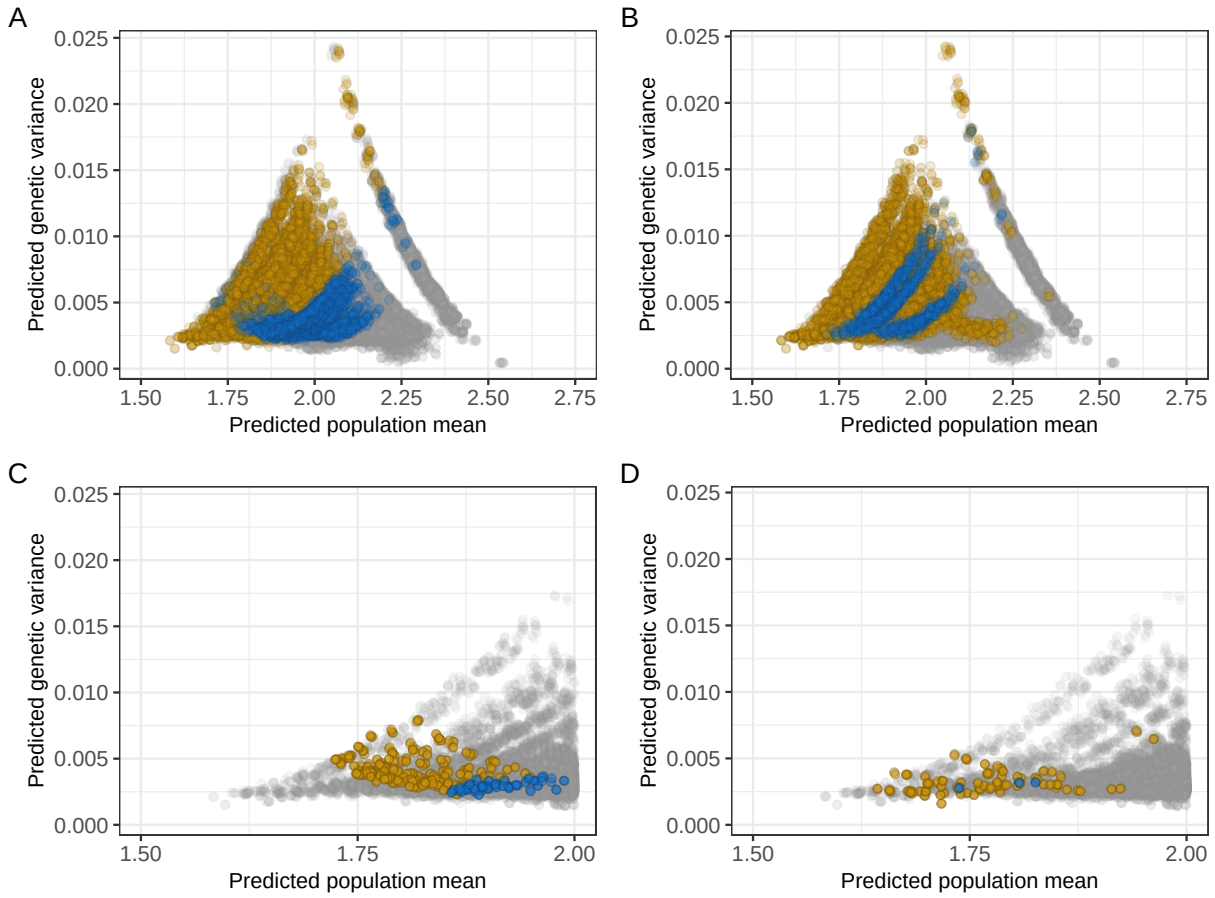

FIGURE 1.6. Genomic-estimated genetic variances and population EMMs for Phytophthora crown rot resistance score are shown for 190,532 simulated segregating populations ($n = 200$ full-sib individuals/population) developed from crosses (with reciprocals) among 437 individuals (prospective parents) in the training population. The prospective parents were classified as resistant or susceptible using resistance score EMMs or GEBVs as a selection criteria to model the outcomes of phenotypic and genomic selection, respectively. The truncation selection cutoffs were EMM < 2.0 for phenotypic selection and GEBV < 2.6 for genomic selection. There were 32 parents in the resistant groups ($32/437 = 0.073$) for phenotypic and genomic selection with an intersection of 15 parents between resistant groups. (A) and (C) show estimates for phenotypic selection, whereas (B) and (D) show estimates for genomic selection. (A) and (B) show statistics for all possible crosses between resistant non-UCD parents and all other training population individuals (gold points) and all possible crosses between resistant UCD parents and all other training population individuals (blue points). (C) and (D) display the lower tails of the EMM distribution (population EMMs < 2.0) and highlight crosses between resistant parents. The gold points identify UCD $\times$ non-UCD crosses, whereas the blue points identify UCD $\times$ UCD crosses. The grey points identify non-UCD $\times$ non-UCD crosses.

the primary distribution that traced to the least resistant parents and most exotic genetic resources in the training population, e.g., *F. chiloensis* and *F. virginiana* ecotypes. These crosses were predicted to unleash the most genetic variation for resistance to PhCR albeit with lower population means (Figure 1.6A-B). There were a small number of crosses with non-UCD resistant parents near the apex in the right-shifted 'exotic' band that warrant further exploration and inclusion in a long-term breeding strategy. These had maximum genetic variance with population means that were only slightly greater than 2.0.

Shifting to the lower (more resistant) tail of the $\hat{\mu} \times \hat{\sigma}^2_A$ and $\hat{\mu} \times$ U distributions and focusing only on crosses between resistant parents, we found that genomic and phenotypic selection scenarios identified different complements of crosses (Figures 1.6C-D and 1.7B-C). Moreover, the complements of UCD $\times$ UCD crosses (blue points) and UCD $\times$ non-UCD crosses (gold points) differed between genomic and phenotypic selection (Figures 1.6C-D and 1.7B-C). The gray points identify non-UCD $\times$ non-UCD crosses. The UCD $\times$ non-UCD crosses depict possible outcomes for the exact breeding scenario that motivated our study, the identification and introduction of possibly novel favorable alleles from highly resistant sources (found to be heirloom cultivars) into modern UCD cultivars. We observed the predicted dichotomy between UCD $\times$ UCD and UCD $\times$ non-UCD crosses for the phenotypic selection scenario with lower genetic variances and population means for the former (Figure 1.6C). The pattern was quite different for the genomic selection scenario where only three UCD $\times$ UCD crosses (0.003% of the crosses simulated) were found in the tail and the UCD $\times$ non-UCD crosses were shifted leftwards toward greater resistance (had lower population means and lower genetic variances) compared to the pattern observed for phenotypic selection (Figure 1.6C-D).

These analyses clearly identified the most promising crosses for pyramiding favorable alleles from different sources and introducing novel favorable alleles from non-UCD sources into elite UCD sources. The triangular $\hat{\mu} \times \hat{\sigma}^2_A$ and ellipsoidal $\hat{\mu} \times$ U distributions predict that useful genetic variation can be unlocked by crosses among the resistant parents identified in our

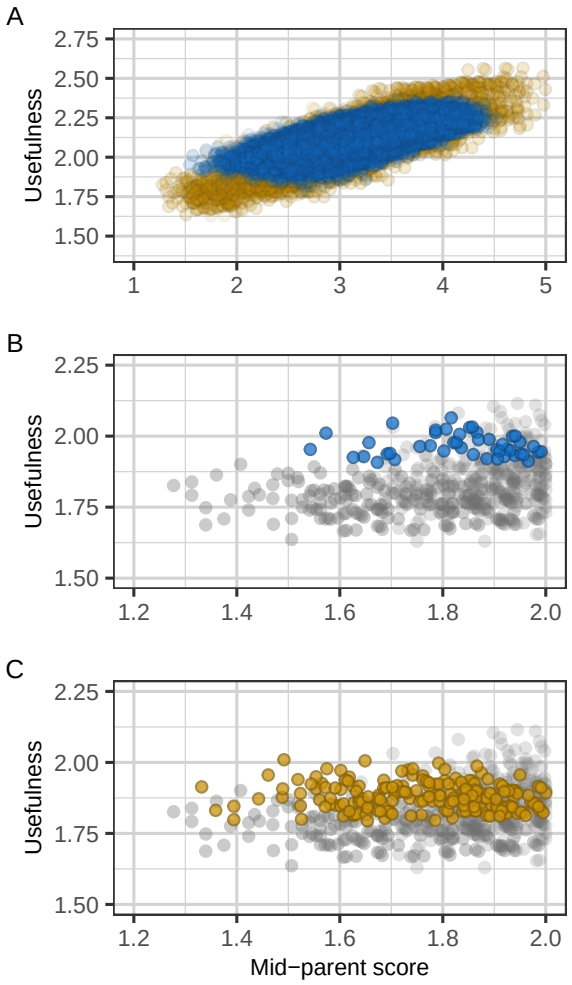

FIGURE 1.7. Cross usefulness criteria (U) and population EMMs for Phytophthora crown rot resistance score are shown for 190,532 simulated segregating populations ($n = 200$ full-sib individuals/population) developed from crosses (with reciprocals) among 437 individuals (prospective parents) in the training population. (A) Statistics are shown for all possible UCD × UCD crosses (blue points) and all possible non-UCD × non-UCD crosses (gold points). (B) Statistics are shown for the lower tail of the EMM distribution (population EMM < 2.0). Crosses between resistant UCD parents only (UCD × UCD crosses) are shown in blue, whereas other crosses between resistant parents (UCD × non-UCD and non-UCD × non-UCD crosses) are shown in gray. (C) Statistics are shown for the lower tail of the EMM distribution (population EMM < 2.0). Crosses between resistant UCD and non-UCD parents (UCD × non-UCD crosses) are shown in gold, whereas other crosses between resistant parents (UCD × UCD and non-UCD × non-UCD crosses) are shown in gray.

study, both exotic and elite (Figures 1.6-1.7). However, the altitudes and apices of the $\hat{\mu} \times \hat{\sigma}^2_A$ distributions (Figure 1.6C-D) clearly show that the small fraction of crosses between pairs of highly resistant parents unlock three to five-fold less genetic variance than crosses with maximum genetic variance at the apex of the entire distribution of 190,532 crosses (Figure 1.6A-B). The latter, of course have intermediate population means. These results suggest that the resistant individuals either share a preponderance of favorable alleles in common, as would be expected from shared ancestry (Hardigan et al., 2020; Pincot et al., 2021), that the cumulative effects of independent QTL combined in thousands of cross combinations are not additive, or both. The prediction accuracy could, in addition, have been partly caused by population structure rather than LD between SNP markers and QTL (Daetwyler et al., 2012), which would further explain the patterns we observed.

1.5. Conclusions

Our study highlighted one of the most pressing challenges ahead for strawberry breeders: stacking resistance to *P. cactorum* and other soil-borne pathogens without eroding genetic gains for yield and other agriculturally important traits that have enabled the phenomenal growth of the strawberry industry since the 1950s. The difficulty of that challenge was illustrated by separately analyzing UCD (90% modern era elite) individuals in the training population—100% of those individuals were predicted to be homozygous or heterozygous for the favorable RPc2 allele. Our analysis of the UCD population mimicked a real world situation where MAS could be applied to a population to fix the favorable RPc2 allele, thereby removing the segregation of RPc2 from the equation, and leaving selection to operate on the 'black box' residual quantitative genetic variation. The specific elite population we targeted (UCD) had significantly less additive genetic variance and appeared to be devoid of many if not most of the favorable alleles for PhCR resistance found in highly resistant heirloom cultivars. We concluded that those alleles had simply been left behind as a consequence of the germplasm conservation, breeding priority, and selection decisions made since the inception of the UCD breeding program in the 1920s. Those decisions profoundly affected the

spectrum and frequencies of alleles preserved among elite individuals that passed through breeding bottlenecks and were preserved in the 2015 rendition of the UCD population we inherited and broadly sampled to assemble the training population for the present study. Our findings suggest that an influx of *novel* favorable alleles from exotic genetic resources into our elite albeit bottlenecked population, and perhaps many others, is necessary to replicate the highly resistant phenotypes of heirloom cultivars.

We undertook this study without any knowledge of the strength or spectrum of resistance to *P. cactorum* in the UCD population that has been the source of commercially important and groundbreaking cultivars for nearly a century. When paired with insights gained from earlier population genomic and forward genetic studies, the genome-informed approaches applied here further pulled back the shroud of mystery that has long surrounded the UCD population, and genetic resources worldwide for that matter, in addition to enabling data-driven decisions to unlock genetic variation for resistance to PhCR and other diseases stowed away in phenotypically anonymous strawberry gene bank collections. We concluded that the genetic complexity of resistance to *P. cactorum*, while important, has perhaps been less of a factor in the scarcity of highly resistant modern cultivars and negative genetic gains than widespread inattention to breeding for resistance. This conclusion was reached because a preponderance of the highly resistant cultivars we discovered were developed in the early twentieth century decades before methyl bromide fumigation emerged and modern genome-informed breeding approaches were invented. The latter have actually come somewhat late to strawberry breeding and cultivar development. While the highly resistant heirloom cultivars documented in the present study are still the benchmark for resistance today, the application of increasingly powerful predictive approaches should enable breeders to replicate the PhCR resistance of early twentieth century heirloom cultivars in high yielding modern cultivars.

1.6. Abbreviations

AUDPS, area under the disease pressure stairs; ASV, average semi-variance; BL, Bayesian Lasso; CNGC, cyclic-nucleotide-gated channel; EMM, estimated marginal mean; FDR, false

discovery rate; G-BLUP, genomic-best linear unbiased prediction; GEBV, genomic-estimated breeding value; GVE, percentage of the genetic variance explained by markers; U, genomic-estimated cross usefulness criteria; GWAS, genome-wide association study; LMM, linear mixed model; LD, linkage disequilibrium, MAS, marker-assisted selection; MAF, minor allele frequency; MCCV, Monte Carlo cross-validation; MV-GWAS, multivariate genome-wide association study; PhCR, Phytophthora crown rot; PVE, percentage of the phenotypic variance explained by markers; QTL, quantitative trait locus; REML, restricted maximum likelihood; RKHS, reproducing kernel Hilbert space; SNP, single nucleotide polymorhism; WGR, whole-genome regression.

CHAPTER 2

Useful Heterosis and Genomic Prediction Strategies for Fruit Quality Traits and Shelf Life in Strawberry

2.1. Abstract

The global expansion of strawberry production has been primarily influenced by the selection of high-yielding long shelf life genotypes. The direct selection of traits for mass production in strawberries, such as fruit size and firmness, has led to a trade-off with fruit quality traits that are important to consumers. Exotic genetic resources might harbor favorable alleles to improve the overall liking of strawberry fruit but may also carry undesired alleles, making breeding efforts challenging. We analyzed the introgression of favorable alleles for fruit quality traits from nine soft-fruited exotic parents into a firm-fruited elite cultivar. We observed a wide range of phenotypic values for firmness ($\overline{x}$ = 0.12-0.45 kg/cm^2), total soluble solids (TSS; $\overline{x}$ = 9.17-11.3%), titratable acidity (TA; $\overline{x}$ = 0.89-1.12%), and anthocyanin content (ANC; $\overline{x}$ = 60.17-96.55 μg/mL) within full-sib families in the elite×exotic population. Significant heterosis was detected in a few hybrids for firmness (0.0-7.7%) and ANC (0.0-14.8%), while a greater number of hybrids showed substantial improvements in TSS (3.9-37.6%) and TA (9.2-55.3%). Our results indicate that genotype combinations notably influence hybrid performance. Genetic correlations among fruit traits indicate that firmness is positively correlated with TA ($\overline{r}$ = 0.40) and negatively with TSS ($\overline{r}$ = -0.50). Furthermore, the lack of genotype×timepoint interaction suggests that measurements taken at harvest might be a reliable predictor of changes in fruit quality traits during storage. Genomic prediction may help overcome challenges associated with introducing exotic resources into elite breeding pools. This study showed that genomic prediction models incorporating half-sib individuals in the

training set can achieve higher predictive abilities for fruit quality traits (0.44-0.91) compared to models with only 20% of individuals from the target family (0.06-0.39) or unrelated germplasm (0.01-0.50). While these results are promising, further investigation is needed to optimize genomic prediction schemes for fruit quality traits, particularly in addressing the impact of genetic distance.

2.2. Introduction

Cultivated strawberry (*Fragaria* × *ananassa*) is one of the most economically valuable and extensively grown small fruits worldwide. The global expansion of strawberries has been driven by the selection of high-yielding long shelf life genotypes over the past century (Feldmann et al., 2024a). While numerous commercially successful modern cultivars show high yields, firmness, and extended shelf life, a prevailing concern among consumers is a perceived compromise in taste quality (Predieri et al., 2021).

The historical genetic gain in the strawberry breeding program at the University of California, Davis suggested that the direct selection of mass-production traits decreased sweetness, acidity, and anthocyanin content (Feldmann et al., 2024a). Thus, mass-production traits are hypothesized to negatively affect fruit quality traits associated with consumer preferences. However, the disparity among studies on the phenotypic and genotypic correlations between fruit quality traits in strawberry (Ghoochani et al., 2015; Hernanz et al., 2008; Lerceteau-Köhler et al., 2012; Masny et al., 2016; Mishra et al., 2015; Shaw, 1988; Singh et al., 2018; Ukalska et al., 2006; Whitaker et al., 2012; Zareei et al., 2022; Zorrilla-Fontanesi et al., 2011) might reflect that the pleiotropic effects among these traits are not fully understood. For instance, some studies reported a positive correlation between total soluble solids and titratable acidity (Lerceteau-Köhler et al., 2012; Shaw, 1988; Whitaker et al., 2012), while others documented negative correlations (Mishra et al., 2015; Singh et al., 2018). These findings underscore the complexity of the genetic interactions underlying fruit quality traits and highlight the need for further studies to improve the overall liking of strawberry fruit without compromising traits for long-scale production.

Genetic resources such as wild relatives and heirloom cultivars might carry novel favorable alleles to expand the genetic base of elite pools and facilitate short- and long-term genetic improvements (Allier et al., 2020). However, the performance gap between exotic resources and modern cultivars often discourages breeders from incorporating them into crosses with elite germplasm in breeding programs. For example, heirloom strawberries are perceived as sweet with superior flavor; yet, their soft texture, susceptibility to damage, and limited shelf life make them unsuitable for modern production environments and long-distance shipping (Collins and Perkins-Veazie, 1993; Pelayo et al., 2003). To address these challenges, genome-wide methods can be used to enhance the efficiency and accuracy of selecting desired genotypes derived from crosses between elite×exotic parents.

The genetic architecture of fruit quality traits in strawberries is complex and involves several genetic factors (Alarfaj et al., 2021; Castro and Lewers, 2016; Lee et al., 2021; Lerceteau-Köhler et al., 2012; Natarajan et al., 2020; Rey-Serra et al., 2021; Verma et al., 2017; Zorrilla-Fontanesi et al., 2011). Identifying reliable and stable loci associated with these traits is crucial for developing DNA markers that allow the selection of desired genotypes. On the other hand, genomic prediction has shown great promise in accurately predicting complex traits in several crops (Ferrão et al., 2021; Gezan et al., 2017; Hong et al., 2020; Medina et al., 2021; Merrick and Carter, 2021), enabling breeders to efficiently identify and select superior genotypes earlier in the breeding process. However, genomic prediction accuracy is influenced by several factors, including the number of markers used for calculating genomic estimated breeding values (Spindel et al., 2015; Tayeh et al., 2015), trait heritability (Kaler et al., 2022; Zhang et al., 2017), training set size (Norman et al., 2018; Tayeh et al., 2015; Zhang et al., 2017), statistical models (Heslot et al., 2012; Spindel et al., 2015), linkage disequilibrium (Habier et al., 2007; Schopp et al., 2017a), the relationship between training and validation sets (Albrecht et al., 2011; Clark et al., 2012), and population structure (De Roos et al., 2009; Norman et al., 2018; Windhausen et al., 2012).

We developed elite×exotic crosses derived from a firm-fruited elite cultivar and nine soft-fruited exotic parents to evaluate the introgression of favorable alleles from exotic genetic resources to enhance fruit quality traits in strawberries. Hybrid performance was evaluated within full-sib families by contrasting each hybrid with its elite parent. Additionally, five full-sib families were studied during post-harvest to evaluate their performance and estimate the trade-off between fruit quality traits. We estimated phenotypic and genetic variance explained by each locus detected for fruit quality traits and their interactions among and within full-sib families. Finally, we considered nine genomic prediction strategies and seven training set sizes to evaluate the effectiveness of genomic prediction for fruit quality traits and the impact of the relatedness between the training and validation sets.

2.3. Materials and Methods

2.3.1. Plant Material. Ten full-sib families were evaluated for fruit quality traits, including an elite×elite backcross between two California breeding lines and elite×exotic crosses derived from the firm-fruited cultivar 'Royal Royce' and nine soft-fruited exotic parents ('Kaoling', 'Morioka 17', 'Primella', 'Madame Moutot', 'Titan', 'MDUS 5130', 'Tangi', 'EarliMiss', and 'Linn'). Exotic parents were initially obtained as bare-root 'mother' plants from the United States Department of Agriculture (USDA) National Plant Germplasm System (NPGS) National Clonal Germplasm Repository in Corvallis, Oregon, USA. Exotic parents were subsequently maintained and propagated asexually at the University of California, Davis (UCD) facilities. Crosses were conducted during the winter of 2017-2018, with seeds sown in the following spring and grown in pots. 795 seedlings (one per plant) and their parents (ten plants per individual) were planted in a non-randomized row design at the Wolfskill Experimental Orchard in Winters, California, in the fall of 2018. The plants were cultivated throughout the 2019 field season.

2.3.2. DNA Extraction and SNP Genotyping. Leaf tissue was lyophilized and ground using a 1600 MiniG® - Automated Tissue Homogenizer and Cell Lyser (SPEX Sample Prep, Metuchen, NJ). 12 mg of ground leaf tissue was used for DNA extraction through an E-Z 96 Plant DNA Kit (Omega Bio-Tek, Norcross, GA) following a modified protocol described by Hardigan et al. (2018). A 50K genotyping array was utilized to identify Single Nucleotide Polymorphisms (SNPs) as described by Hardigan et al. (2020). After filtering for a minor allele frequency ≤ 0.05 and missing data $\geq 10\%$, followed by imputation through the `A.mat` function from the rrBLUP R package (Endelman, 2011), 30,269 SNPs were retained for subsequent analysis.

2.3.3. Population Structure Analysis. A principal component analysis (PCA) was performed to evaluate the relatedness between the ten full-sib families and historical records studied by Feldmann et al. (2024a). SNPs with variance = 0 were excluded from the analysis. The PCA was carried out using the `prcomp()` function with the scale parameter set to TRUE from the R package (R Core Team, 2021).

2.3.4. Fruit Quality Phenotyping. Fruit quality measurements were conducted at full ripeness on May 15 and May 31. Firmness was evaluated from three fruits per individual through a TA.XT plus Texture Analyzer (Stable Micro Systems Ltd., Goldaming, United Kingdoms) equipped with a TA-53 3 mm puncture probe. Subsequently, the fruits were stored at -20°C in Whirl-Pak® Homogenizer Blender Filter Bags (Nasco, Fort Atkinson, WI) until titratable acidity (TA) measurements. TA was determined using the Metrohm Robotic Titrosampler System (Metrohm AG, Herisau, Switzerland) with 1-5 ml of defrosted and homogenized fruit juice. Total soluble solids (TSS) were measured from 200 µl of fruit juice with an RX-5000α-Bev refractometer (ATAGO Co. Ltd., Tokyo, Japan). Anthocyanin content (ANC) was determined by mixing 25 µl of fruit juice in a 200 µl solution of 1% HCl in methanol. Absorption (Abs) was measured at 520 nm using a Synergy HTX Multi-Mode Microplate Reader (BioTek Instruments, Inc.) with Gen5 software. A dilution series of

Pelargonidin (Sigma Aldrich) from 0 - 300 µg/ml was used to calculate a standard curve of the following formula:

$$y = sx + i \tag{2.1}$$

The concentration of anthocyanins in the juice was calculated with the following formula:

$$C(\mu g/ml) = Abs - is \tag{2.2}$$

Post-Harvest Evaluation of Fruit Related-Quality Traits. Fruit quality traits were evaluated for 12 days post-harvest in the elite×elite backcross and four elite×exotic crosses ('Primella', 'Madame Moutot', 'Tangi', and 'EarliMiss'). Images of the fruit were captured every two days through a Sony α 6000 camera equipped with an E PZ 16–50 mm F3.5–5.6 OSS lens (SONY, Tokyo, Japan). Fruit showing signs of fungal infection was discarded and not included in subsequent analyses. Image processing was conducted using a custom macro in Fiji to determine RGB color values (Rueden et al., 2017; Schindelin et al., 2012). These values were then converted into the Lab color space through the `convertColor()` function in R (R Core Team, 2021). Texture, TSS, TA, and ANC were measured every four days.

Genetic correlations among fruit quality traits at 0, 4, and 8 days post-harvest were estimated using Genomic Best Linear Unbiased Predictions (G-BLUP) in the `sommer` R package (Giovanny, 2016). A multitrait model for genetic correlations with all traits at once was not feasible due to computational challenges. The analysis involved estimating 4704 variances and covariances. Thus, pairs of traits were used to calculate genetic correlations.

2.3.5. Statistical Data Analysis. Phenotypic values were analyzed using linear mixed model (LMM) functions from the R package lme4::lmer() (`https://cran.r-project.org/web/packages/lme4/index.html`). Estimated marginal means (EMMs) were calculated

through the `emmeans` R package (Lenth, 2021). The LMM was expressed as:

$$y_{ij} = b_i + G_j + e_{ij} \tag{2.3}$$

where y_{ij} indicates the observed phenotype for the jth genotype (individual) in the ith harvest, b_i is the random effect of the ith harvest, G_j is the random effect of the jth individual, and e_{ij} indicates the ijth residual effect.

Variance components in these analyses were estimated using the restricted maximum likelihood (REML) method (Bates et al., 2015). Narrow-sense genomic heritability ($\hat{h}^2$) was estimated using the formula $\hat{h}^2 = \hat{\sigma}^2_A/\hat{\sigma}^2_P$, where $\hat{\sigma}^2_A$ is the REML estimate of the genomic additive genetic variance and $\hat{\sigma}^2_P$ is the REML estimate of the phenotypic variance (Endelman, 2011; Mathew et al., 2018). Additionally, Genotype, timepoint, and interaction effects were examined for post-harvest assessments using Type I Analysis of Variance (ANOVA) in R.

2.3.6. Useful Heterosis. F1 hybrid performance was evaluated over their check elite parents through useful heterosis (UH) proposed by Meredith Jr. and Bridge (1972). UH was analyzed via the following formula:

$$\text{UH} = \frac{H_i - BP}{BP} \tag{2.4}$$

Here, H_i indicates the EMM of the H_i hybrid, and BP is the EMM of the check elite parent within a specific family. The elite parent selected for the elite×elite backcross was 05C197P002, while for the elite×exotic families was the elite cultivar 'Royal Royce' . Contrasts between hybrid and BP EMMs were estimated for every parent-hybrid combination using the `emmeans` R package (Lenth, 2021).

2.3.7. Genome Wide Association Study. EMMs for firmness, TSS, TA, TSS-TA ratio, and ANC were analyzed through Genome-Wide Association Studies (GWAS). The

Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway (BLINK) method implemented by GAPIT version 3 (Wang and Zhang, 2021) was chosen for its ability to increase statistical power by not assuming an even distribution of trait-associated genes across the genome (Huang et al., 2019). Population structure was corrected by including the genomic relationship matrix (Zhou and Stephens, 2012,1) from the `A.mat()` function in the rrBLUP R package (Endelman, 2011). The physical positions of SNPs were mapped to the haplotype-phased reference genome 'Royal Royce' v1.0 (Hardigan et al., 2021b).

The most significant SNPs identified for each fruit quality trait, along with their interactions, were fitted into a linear model. Phenotypic (PVE = $\hat{\sigma}^2_M/\hat{\sigma}^2_P$) and genetic (GVE = $\hat{\sigma}^2_M/\hat{\sigma}^2_G$) variance explained by each locus and its interactions were calculated. The $\hat{\sigma}^2_M$ indicates a bias-corrected average semivariance REML estimate for the marker among individuals, $\hat{\sigma}^2_P$ is the phenotypic variance and $\hat{\sigma}^2_G$ is the genetic variance (Feldmann et al., 2021).

2.3.8. Genomic Prediction Across and Within Families. Nine genomic prediction strategies and seven training set (TS) sizes were evaluated to estimate progeny and family Genomic Estimated Breeding Values (GEBVs) for fruit quality traits (Figure 2.1). GEBVs were calculated using the `sommer` R package (Giovanny, 2016). Monte Carlo Markov chain simulations with 1,000 iterations were employed to select individuals for the TS using random sampling without replacement. Predictive ability (PA = $\frac{r_{y,\hat{g}}}{\sqrt{\hat{h}}}$) was estimated for each genomic prediction strategy (Daetwyler et al., 2013). Pearson's correlation coefficient was calculated between EMMs and GEBVs ($r_{y,\hat{g}}$) (Dekkers, 2007; Dekkers et al., 2021; Van den Berg et al., 2019) for family mean. Normalized root-mean-square error (nRMSE) was calculated to evaluate the performance of the genomic prediction strategies and TS sizes. nRMSE was estimated through the following formulas:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{2.5}$$

$$\text{nRMSE} = \frac{\sqrt{\text{MSE}}}{y_{\max} - y_{\min}} \tag{2.6}$$

Where y_i is the observed value, $\hat{y}_i$ is the predicted value, n indicates the total number of data points, $y_{\max}$ is the maximum value, and $y_{\min}$ is the minimum value of the measured data.

2.4. Results and Discussion

2.4.1. Genetic Diversity and Phenotypic Variation in Elite×Exotic Families. The genetic distance analysis of ten full-sib families, including an elite×elite backcross population from two California breeding lines and nine crossbred populations derived from the elite cultivar 'Royal Royce' and exotic parents ('Kaoling', 'Morioka 17', 'Primella', 'Madame Moutot', 'Titan', 'MDUS 5130', 'Tangi', 'EarliMiss', and 'Linn'), showed a clear separation between these two populations (Figure 2.2A). However, the analysis of population structure in the elite×exotic crosses displayed isolated genetic clusters involving the exotic parents 'Primella', 'Madame Moutot', and 'Kaoling', while specific full-sib families derived from exotic parents such as 'Morioka 17', 'Titan', 'MDUS 5130', and 'EarliMiss' showed overlapping (Figure 2.2B). Additionally, the elite×elite and elite×exotic populations were analyzed against an extensive compilation of breeding records (Feldmann et al., 2024a) that considered the domestication range from wild ancestors to modern cultivars (Figure 2.3). The elite×elite backcross population and its parents were closely clustered with the California population. In contrast, the elite×exotic families showed divergence from the California pool. The exotic parents were located with Cosmopolitan sources, whereas 'Royal Royce' remained closely related to the California germplasm. Thus, the genetic differentiation introduced by the exotic parents resulted in new subpopulations from the California population, as previously observed by Hardigan et al. (2021b) and Pincot et al. (2021).

A wide range of phenotypic values was observed within and across populations for firmness, TSS, TA, TSS-TA ratio, and ANC (Figure 2.4). The elite×elite population exhibited higher

firmness ($\overline{x}$ = 0.34 kg/cm^2; $p = 3.81 \times 10^{-11}$), TSS-TA ratio ($\overline{x}$ = 13.4; $p = 2.11 \times 10^{-8}$), and ANC ($\overline{x}$ = 92.3 μg/mL; $p = 2.71 \times 10^{-6}$). In contrast, the elite×elite population showed lower TSS ($\overline{x}$ = 9.8%; $p = 1.03 \times 10^{-3}$) and TA ($\overline{x}$ = 0.75%; $p < 2.2 \times 10^{-16}$) compared to the elite×exotic population ($\overline{x}$ = 0.23 kg/cm^2, 10.5, 0.99%, 11.5%, and 75.3 μg/mL, respectively). However, some full-sib families within the elite×exotic population exhibited similar or higher phenotypic values for firmness ($\overline{x}$ = 0.12-0.45 kg/cm^2), TSS ($\overline{x}$ = 9.17-11.3%), TA ($\overline{x}$ = 0.89-1.12%), TSS-TA ratio ($\overline{x}$ = 8.48-13.35), and ANC ($\overline{x}$ = 60.17-96.55 μg/mL) compared to the elite×elite population (Figure 2.4). Our findings align with previous research emphasizing the importance of using diverse germplasm to enhance horticultural and commercial traits in strawberries (Hummer et al., 2023; Mathey et al., 2013). Mathey et al. (2013) analyzed several fruit quality traits from a diverse strawberry collection, including accessions from breeding initiatives such as USDA, Michigan, Florida, New Hampshire, and European programs. Their study revealed a wide range of phenotypic values for TSS ($\overline{x}$ = 3.1-19.5%), TA ($\overline{x}$ = 0.2-2.2%), TSS:TA ratio ($\overline{x}$ = 3.0-35.1), and ANC ($\overline{x}$ = 5.3-1109.3 μg/mL). Thus, harnessing genetic diversity from gene banks might enhance fruit quality traits in strawberry breeding programs by broadening the genetic base and overcoming the limitations of elite germplasm with narrow genetic diversity.

The narrow-sense heritability ($\hat{h}^2$) estimates spanned from low to moderate values for the elite×elite and elite×exotic populations (Table 2.1). We estimated low $\hat{h^2}$ for TSS (0.17-0.18) and TSS-TA ratio (0.03-0.18), whereas moderate $\hat{h^2}$ was observed for firmness (0.35-0.41), TA (0.34-0.38), and ANC (0.23-0.37). In the elite×exotic population, we noticed a wide range of $\hat{h^2}$ estimates across full-sib families (Table 2.1). Several full-sib families from the elite×exotic population showed higher additive genetic effects than the elite×elite population. For instance, Tangi's family exhibited higher $\hat{h^2}$ for firmness (0.60) and ANC (0.43), while Madame Moutot's family for TSS (0.33) and TA (0.55). $\hat{h^2}$ estimates from independent studies (Cockerton et al., 2021; Coman and Popescu, 1997; Gezan et al., 2017; Ghoochani et al., 2015; Mishra et al., 2015; Murti et al., 2012; Ukalska et al., 2006) underscore

TABLE 2.1. **Narrow-sense heritability ($\hat{h}^2$) via restricted maximum likelihood estimates for firmness, total soluble solids (TSS), titratable acidity (TA), TSS-TA ratio, and anthocyanin content (ANC) across diverse strawberry families.** $\hat{h}^2$ was estimated for an elite×elite backcross population between two California breeding lines and elite×exotic population derived from crosses between the elite cultivar 'Royal Royce' and nine exotic parents ('Kaoling', 'Morioka 17', 'Primella', 'Madame Moutot', 'Titan', 'MDUS 5130', 'Tangi', 'EarliMiss', and 'Linn').

Population	Firmness (kg/cm^2)	TSS (%)	TA (%)	TSS:TA Ratio	ANC (μg/mL)
Elite×Elite	0.35	0.18	0.34	0.04	0.21
Elite×Exotic	0.41	0.17	0.38	0.18	0.23
Kaoling	0.38	0.00	0.51	0.04	0.27
Morioka 17	0.39	0.20	0.58	0.11	0.00
Primella	0.00	0.28	0.41	0.26	0.39
Madame Moutot	0.40	0.33	0.55	0.34	0.43
Titan	0.52	0.00	0.45	0.04	0.00
MDUS 5130	0.24	0.00	0.36	0.22	0.25
Tangi	0.60	0.18	0.21	0.12	0.43
EarliMiss	0.21	0.12	0.37	0.13	0.36
Linn	0.24	0.29	0.40	0.26	0.10

the role of genetic backgrounds behind complex traits. The discrepancy of $\hat{h}^2$ estimates across studies might be attributed to differences in population-specific parameters, such as additive and non-additive genetic effects, environmental factors, and statistical methodologies used to estimate $\hat{h}^2$ (Visscher et al., 2008). These findings highlight the relevance of selecting parents to increase genetic variation and genetic gain for fruit quality traits in strawberry breeding programs.

2.4.2. Perspectives on Genotype Combination for Fruit Quality Traits in Elite×Exotic Crosses. We estimated useful heterosis (UH) to evaluate the performance of elite×elite and elite×exotic hybrids (Figure 2.5). The elite×exotic hybrids showed superior performance for TSS (85.9%), TA (85.2%), TSS:TA (45.1%), and ANC (47.3%) and only 3% of them for firmness. By contrast, the elite×elite population showed improvements in firmness (48.2%), TSS ratio (40%), and ANC (61.2%), with few hybrids performing well in TSS (9.4%) and TA (3.5%). The low number of hybrids with high firmness in the elite×exotic population may be attributed to the introgression of deleterious alleles from

exotic parents, which potentially disrupted the favorable alleles accumulated in the elite cultivar 'Royal Royce' (Feldmann et al., 2023; Li et al., 2022). However, a small percentage of the elite×elite and elite×exotic hybrids was significantly greater than their elite parents (Table 2.2). The elite×elite population showed significant improvement over its elite parent for firmness (8.2%) and ANC (7.1%), while the elite×exotic population between all evaluated fruit quality traits, particularly in TSS (16.2%) and TA (28.6%). We noticed that the likelihood of obtaining hybrids with superior performance for fruit quality traits varied among the elite×exotic families (Table 2.2). Previous studies indicated heterosis varies from one hybrid to another for fruit quality traits in strawberry (Kaczmarska et al., 2016; Murti et al., 2012). Several elite×exotic families produced hybrids with higher improvement for TSS (3.9-37.6%) and TA (9.2-55.3%) but showed less superior hybrids for TSS:TA (0.0-7.3%), ANC (0.0-14.8%), and firmness (0.0-7.7%). Thus, the use of exotic genetic resources must be carefully considered to enhance elite genetic pools.

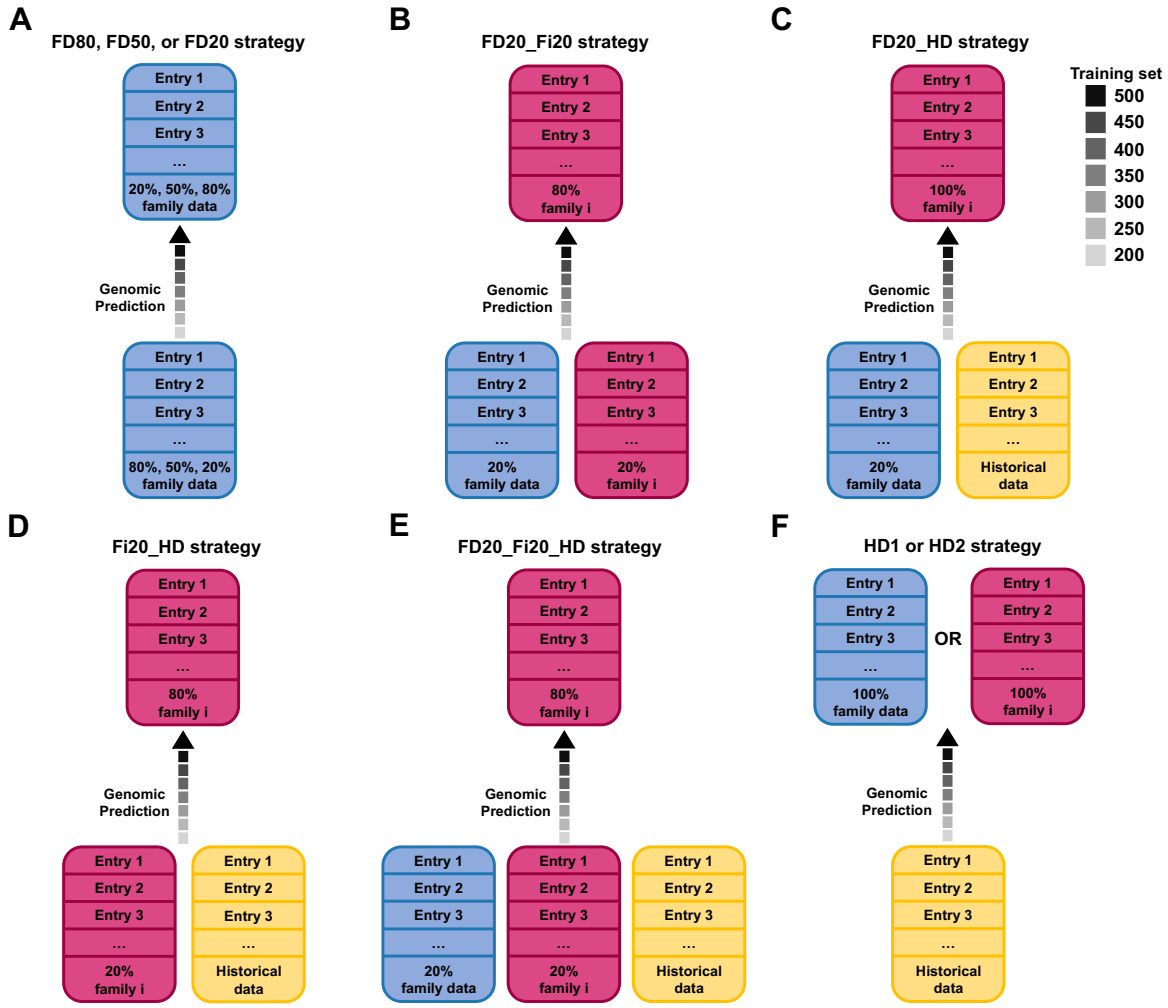

FIGURE 2.1. **Genomic prediction strategies across and within families for fruit quality traits in strawberry.** Nine genomic prediction schemes and seven training set (TS) sizes were evaluated to calculate Genomic Estimated Breeding Values (GEBVs) using Genomic Best Linear Unbiased Predictions (G-BLUP) for both progeny and family mean. Monte Carlo Markov chain simulations with 1,000 iterations were employed to select individuals for the TS using random sampling without replacement. GEBVs across families were calculated by selecting individuals as the TS from A) 80%, 50%, or 20% sampling across families (FD80, FD50, and FD20, respectively) and F) historical records (HD1). For within families, the TS was built from B) 20% sampling across families and 20% from the targeted family (FD20_Fi20), C) 20% sampling across families and historical data (FD20_HD), D) 20% sampling from the targeted family and historical data (Fi20_HD), E) 20% sampling across families, 20% from the targeted family, and historical data (FD20_Fi20_HD), and F) historical records (HD2). Historical data includes breeding records from wild ancestors to modern cultivars (Feldmann et al., 2024a). The grayscale arrow indicates the target population, highlighting the TS sizes from 200 to 500 individuals.

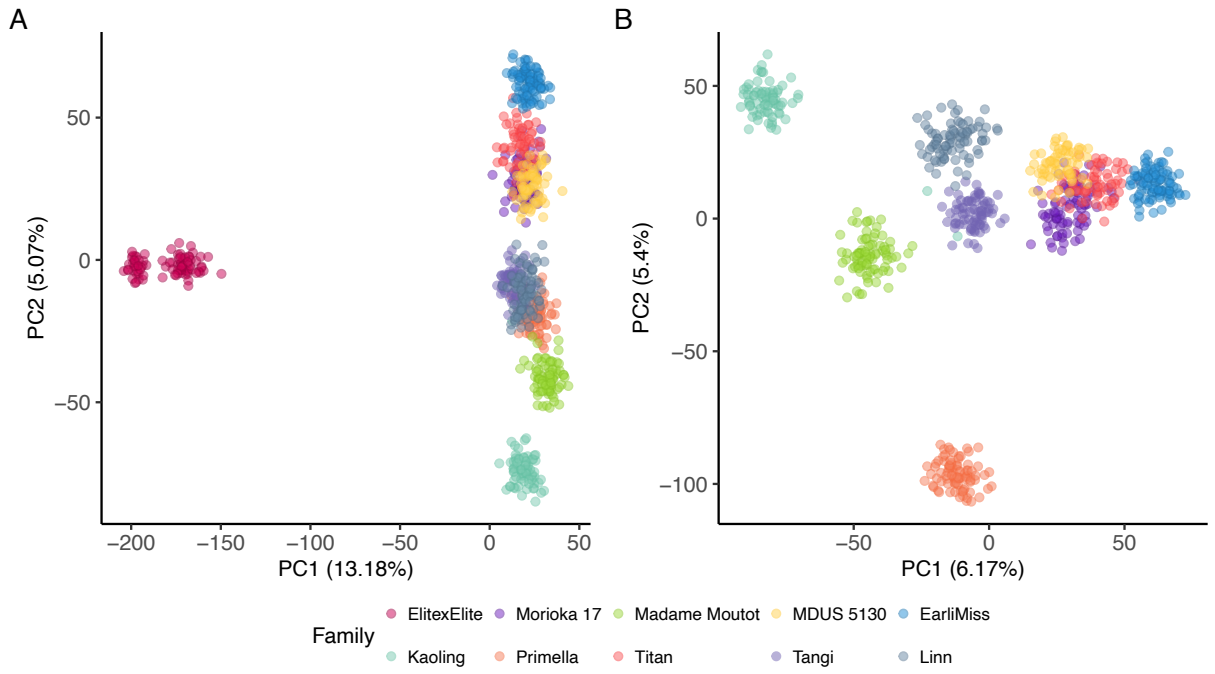

FIGURE 2.2. **Genetic distance unveils distinct subpopulations within the studied families.** The first two principal components, derived from 30,092 Single Nucleotide Polymorphisms (SNPs), show genetic distance A) between an elite×elite backcross and nine elite×exotic crosses and B) only elite×exotic crosses. The elite×elite backcross was developed between two California breeding lines (16C108P605 and 05C197P002) and the elite×exotic population was derived from the elite cultivar 'Royal Royce' and nine exotic parents ('Kaoling', 'Morioka 17', 'Primella', 'Madame Moutot', 'Titan', 'MDUS 5130', 'Tangi', 'EarliMiss', and 'Linn').

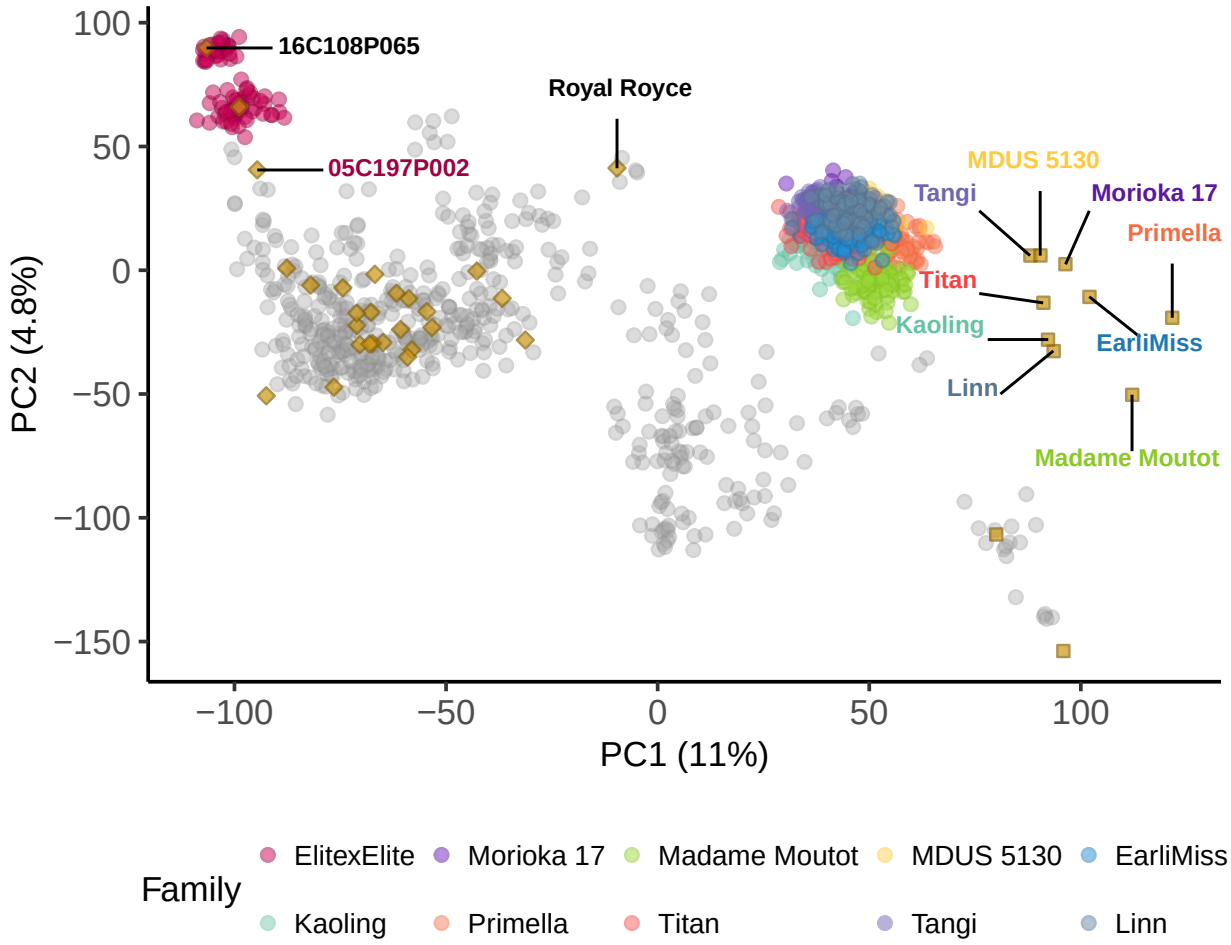

FIGURE 2.3. **Genetic relatedness of elite×exotic families with a California population.** The first two principal components were obtained from 30,092 Single Nucleotide Polymorphisms (SNPs). The California backcross population (E×E) was developed by crossing the California breeding lines 16C108P605 and 05C197P002. The crossbred families were derived from the crosses between the elite cultivar 'Royal Royce' and nine exotic parents ('Kaoling', 'Morioka 17', 'Primella', 'Madame Moutot', 'Titan', 'MDUS 5130', 'Tangi', 'EarliMiss', and 'Linn'). Grey points indicate individuals from a California genetic pool described by Feldmann et al. (2024a). Parents from the California population are represented by gold diamond-shaped points, and exotic parents by gold square-shaped points.

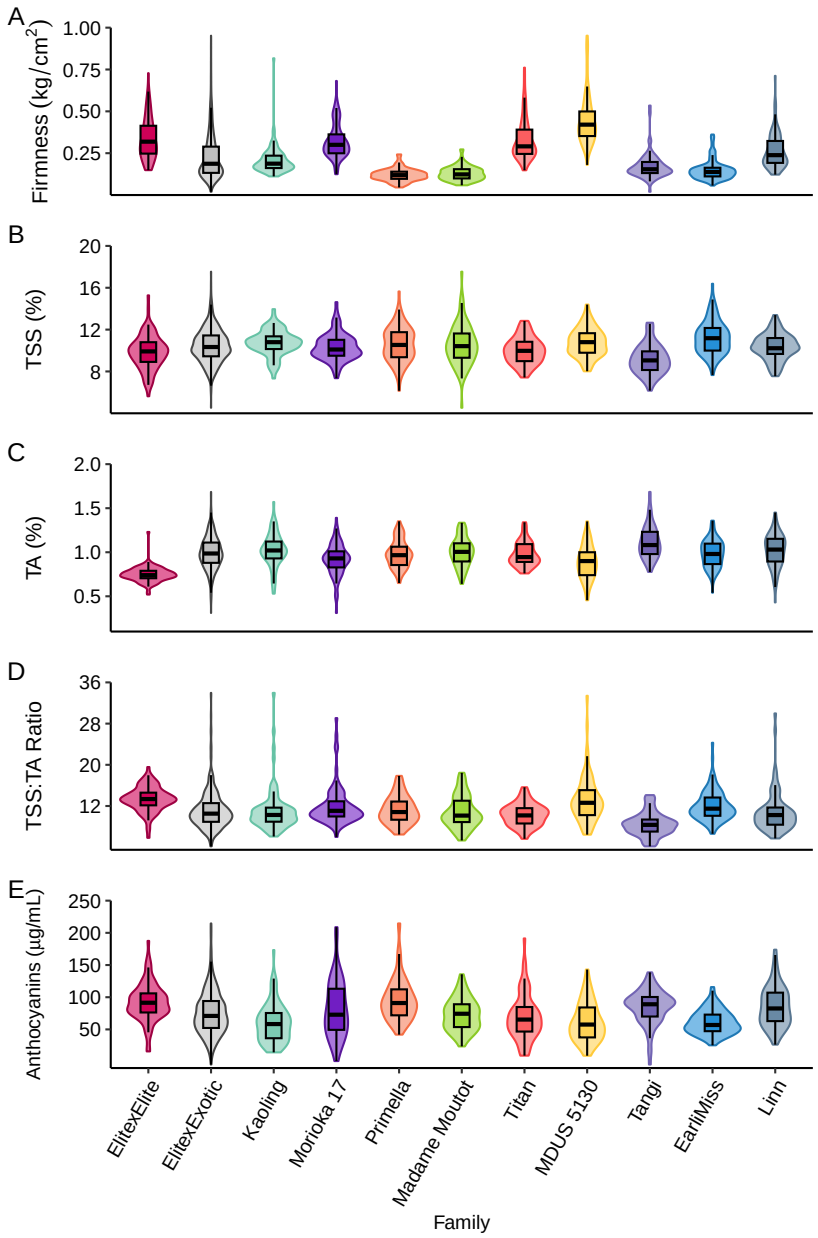

FIGURE 2.4. **Phenotypic variation of fruit quality traits within and across strawberry families.** Phenotypic distribution of A) firmness, B) total soluble solids (TSS), C) titratable acidity (TA), D) TSS:TA ratio, and E) anthocyanin content in an elite×elite backcross and nine elite×exotic crossbred families. Estimated marginal means (EMMs) were obtained from a single plant per individual at two harvests in 2018-2019. Full-sib families include an elite×elite backcross population from two California breeding lines (16C108P065 and 05C197P002) and nine elite×exotic populations derived from the crosses between the elite cultivar 'Royal Royce' and exotic parents ('Kaoling', 'Morioka 17', 'Primella', 'Madame Moutot', 'Titan', 'MDUS 5130', 'Tangi', 'EarliMiss', and 'Linn').

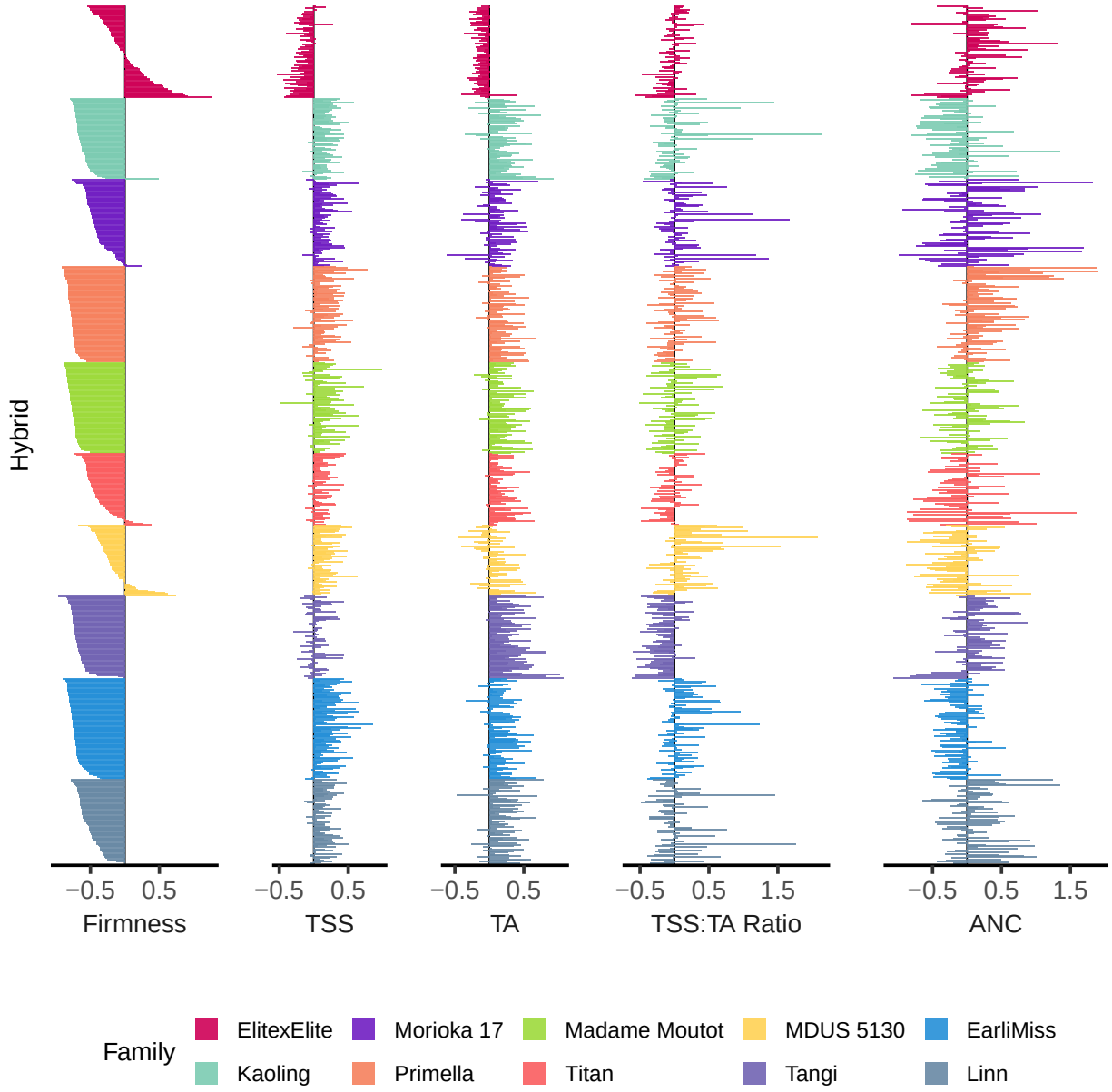

FIGURE 2.5. **Hybrids from the elite×exotic population show superior performance for several fruit quality traits.** Useful heterosis (UH) was evaluated for firmness, total soluble solids (TSS), titratable acidity (TA), TSS-TA ratio, and anthocyanin content (ANC) in elite×elite and elite×exotic populations. The elite×elite family was derived from the backcross between two California breeding lines (16C108P065 and 05C197P002), while the crossbred families between the elite cultivar 'Royal Royce' and nine exotic parents (Kaoling, Morioka 17, Primella, Madame Moutot, Titan, MDUS 5130, Tangi, EarliMiss, and Linn). The California breeding material 05C197P002 and the elite cultivar 'Royal Royce' were used as elite parental sources to estimate UH. Hybrids were organized using firmness as a reference.

TABLE 2.2. **Significant differences between hybrids and their elite parent for fruit quality traits in elite×exotic families.** The percentage of elite×elite and elite×exotic hybrids with useful heterosis (UH) significantly greater than zero (UH > 0), significantly less than zero (UH < 0), and not significantly different from zero (UH = 0) was estimated. We calculated UH between hybrids and the elite parent (UH = $\hat{y}_{f1}$ - $\hat{y}_{BP}$), where $\hat{y}_{f1}$ is the estimated marginal mean (EMM) for the hybrid and $\hat{y}_{BP}$ is the best parent. EMMs were derived from phenotypes of parents and hybrids across single plants and two harvest times for firmness, total soluble solids (TSS), titratable acidity (TA), and anthocyanin content (ANC). The best parent for the elite×elite hybrids (n = 85) was 05C197P002, while the best parent for the elite×exotic hybrids (n = 710) was the elite cultivar 'Royal Royce' .

		Firmness (%)			**TSS (%)**			**TA (%)**			**TSS:TA (%)**			**ANC (%)**		
Population	n	< 0.0	= 0.0	> 0.0	< 0.0	= 0.0	> 0.0	< 0.0	= 0.0	> 0.0	< 0.0	= 0.0	> 0.0	< 0.0	= 0.0	> 0.0
Elite×Elite	85	0.0	91.8	8.2	15.3	84.7	0.0	2.4	97.6	0.0	0.0	100	0.0	0.0	92.9	7.1
Elite×Exotic	710	89.0	9.7	1.3	0.1	83.7	16.2	1.4	70.0	28.6	0.0	96.8	3.2	1.3	92.8	5.9
Kaoling	75	98.7	0.0	1.3	0.0	86.7	13.3	1.3	72.0	26.7	0.0	96.0	4.0	2.7	96.0	1.3
Morioka	82	80.5	18.3	1.2	0.0	92.7	7.3	6.1	76.8	17.1	0.0	92.7	7.3	4.9	82.9	12.2
Primella	88	100	0.0	0.0	0.0	78.4	21.6	0.0	77.3	22.7	0.0	100	0.0	0.0	85.2	14.8
Madame Moutot	84	100	0.0	0.0	1.2	75.0	23.8	0.0	67.9	32.1	0.0	97.6	2.4	1.2	95.2	3.6
Titan	66	77.3	19.7	3.0	0.0	95.5	4.5	0.0	81.8	18.2	0.0	100	0.0	0.0	93.9	6.1
MDUS 5130	65	46.2	46.1	7.7	0.0	83.1	16.9	3.1	87.7	9.2	0.0	93.8	6.2	0.0	96.9	3.1
Tangi	76	96.0	4.0	0.0	0.0	96.1	3.9	0.0	44.7	55.3	0.0	100	0.0	1.3	93.4	5.3
EarliMiss	93	100	0.0	0.0	0.0	62.4	37.6	0.0	63.4	36.6	0.0	95.7	4.3	1.1	98.9	0.0
Linn	81	90.1	9.9	0.0	0.0	90.1	9.9	2.5	63.0	34.6	0.0	95.1	4.9	0.0	93.8	6.2

2.4.3. Post-harvest Dynamics of Fruit Quality Traits. Strawberry fruit is highly perishable and vulnerable to mechanical damage, dehydration, and physiological irregularities during storage (Vu et al., 2011). Although several technologies have helped to mitigate these challenges (Castelló et al., 2010; Joshi et al., 2019; Maraei and Elsawy, 2017; Muley and Singhal, 2020), cultivar selection is critical for extending post-harvest quality in strawberry (Prange and DeEll, 1997). We evaluated fruit quality-related traits in the elite×elite and four elite×exotic full-sib families at post-harvest. We did not observe visual changes during the storage period (Figure 2.6A), following common patterns of variation for fruit quality traits across families (Figure 2.6B). We noticed post-harvest changes such as increasing TSS, TA, and ANC accumulation that have been previously reported in strawberry fruit (Cordenunsi et al., 2003; Lee et al., 2022; Nunes et al., 2006). Our results indicated an increase in strawberry firmness as well as Chandra et al. (2015), whereas Collins and Perkins-Veazie (1993) and Ali et al. (2011) reported firmness decay during storage. Thus, fruit quality traits might be influenced by genetic backgrounds and storage conditions. On the other hand, our analysis of variance indicated no significant genotype × timepoint interaction for these observed changes, except in the cases of elasticity, ANC, and color components (Table 2.3). The lack of genotype×timepoint interaction for most of the evaluated traits suggests that assessing fruit quality traits at harvest might be a reliable predictor of post-harvest changes.

TABLE 2.3. **Prediction of post-harvest performance for strawberry fruit quality traits using harvest metrics.** Analysis of variance for texture (firmness, skin strength, and elasticity), taste (total soluble solids (TSS), titratable acidity (TA), TSS:TA ratio, and pH), pigmentation (anthocyanin content (ANC), a, b, and L values), and fruit size. An elite×elite and four elite×exotic full-sib families were evaluated at 0, 4, 8, and 12 post-harvest days.

Firmness	**Df**	**Sum Sq**	**Mean Sq**	**F value**	**Pr(>F)**	**Significance**
Genotype	430	27.62	0.06	8.78	$<2 \times 10^{-16}$	***
Timepoint	3	0.99	0.33	45.03	$<2 \times 10^{-16}$	***
Genotype:Timepoint	1044	8.27	0.01	1.08	0.0756	

Residuals	1684	12.32	0.01			
Skin Strength						
Genotype	430	10.37	0.02	5.27	$<2 \times 10^{-16}$	***
Timepoint	3	0.64	0.21	46.33	$<2 \times 10^{-16}$	***
Genotype:Timepoint	1044	4.49	0.00	0.94	0.8650	
Residuals	1684	7.71	0.00			
Elasticity						
Genotype	430	755.80	1.76	3.53	$<2 \times 10^{-16}$	***
Timepoint	3	2205.18	735.06	1475.71	$<2 \times 10^{-16}$	***
Genotype:Timepoint	1044	637.45	0.61	1.23	0.0001	***
Residuals	1684	838.81	0.50			
TSS						
Genotype	430	8727.39	20.30	2.46	$<2 \times 10^{-16}$	***
Timepoint	3	316.54	105.51	12.81	2.83×10^{-8}	***
Genotype:Timepoint	1044	5923.98	5.67	0.69	1	
Residuals	1682	13859.54	8.24			
TA						
Genotype	430	107.58	0.25	5.67	$<2 \times 10^{-16}$	***
Timepoint	3	31.19	10.40	235.70	$<2 \times 10^{-16}$	***
Genotype:Timepoint	1038	43.23	0.04	0.94	0.8468	
Residuals	1655	73.01	0.04			
TSS:TA Ratio						
Genotype	430	16854.80	39.20	2.55	$<2 \times 10^{-16}$	***
Timepoint	3	1295.24	431.75	28.04	$<2 \times 10^{-16}$	***

Genotype:Timepoint	1038	14420.35	13.89	0.90	0.9660	
Residuals	1655	25482.57	15.40			
pH						
Genotype	430	85.37	0.20	4.90	$<2 \times 10^{-16}$	***
Timepoint	3	0.22	0.07	1.82	0.1418	
Genotype:Timepoint	1038	34.46	0.03	0.82	0.9998	
Residuals	1655	67.10	0.04			
ANC						
Genotype	430	381.32	0.89	4.84	$<2 \times 10^{-16}$	***
Timepoint	3	20.83	6.94	37.91	$<2 \times 10^{-16}$	***
Genotype:Timepoint	1044	228.71	0.22	1.20	0.0006	***
Residuals	1682	308.11	0.18			
a						
Genotype	425	77090.62	181.39	8.07	$<2 \times 10^{-16}$	***
Timepoint	3	17051.73	5683.91	253.02	$<2 \times 10^{-16}$	***
Genotype:Timepoint	1026	25567.48	24.92	1.11	0.0328	*
Residuals	1587	35651.33	22.46			
b						
Genotype	425	192301.61	452.47	10.34	$<2 \times 10^{-16}$	***
Timepoint	3	36717.16	12239.05	279.61	$<2 \times 10^{-16}$	***
Genotype:Timepoint	1026	49946.92	48.68	1.11	0.0296	*
Residuals	1587	69465.09	43.77			
L						
Genotype	425	120938.78	284.56	11.92	$<2.2 \times 10^{-16}$	***

Timepoint	3	584.29	194.76	8.16	$<2.2\times10^{-16}$	***
Genotype:Timepoint	1026	27076.38	26.39	1.11	0.0372	*
Residuals	1587	37876.08	23.87			
Fruit Size						
Genotype	430	39132.00	91.00	3.22	$<2\times10^{-16}$	***
Timepoint	3	16442.53	5480.84	193.93	$<2\times10^{-16}$	***
Genotype:Timepoint	1044	20200.42	19.35	0.68	1	
Residuals	1684	47593.23	28.26			

Significance levels: *** 0.001 and * 0.05

Genetic correlations among the evaluated fruit quality traits during post-harvest exhibited different relationship patterns when the elite×elite population was included (Figure 2.7). The first two principal components for both the elite×elite and elite×exotic populations explained 51.3% of the total additive genetic variance (Figure 2.7A). Firmness was positively associated with fruit size ($\overline{r}$ = 0.62), skin strength ($\overline{r}$ = 0.74) and TSS:TA ($\overline{r}$ = 0.42), but negatively with TSS ($\overline{r}$ = -0.51) and TA ($\overline{r}$ = -0.74). The elite×elite population was closely associated with firmness and fruit size, while the elite×exotic population clustered with color components, TSS, and TA (Figure 2.7B). On the other hand, the first two principal components for the elite×exotic population explained 40.2% of the total additive genetic variance (Figure 2.7C). Firmness was positively associated with skin strength ($\overline{r}$ = 0.85), TA ($\overline{r}$ = 0.40) and negatively correlated with TSS ($\overline{r}$ = -0.50), TSS:TA ($\overline{r}$ = -0.62) and ANC ($\overline{r}$ = -0.38). We observed the four elite×exotic families were differentially pooled among the studied fruit quality traits (Figure 2.7D). The 'Tangi' full-sib family was closely associated with firmness, size, and TA; the 'Primella' full-sib family with ANC; and both the 'Madame Moutot' and 'EarliMiss' full-sib families with color components, elasticity, and TSS. Similar discrepancies have been observed in previous studies for phenotypic and genotypic correlations of fruit quality traits in strawberries (Adams et al., 2020; Ghoochani et al., 2015;

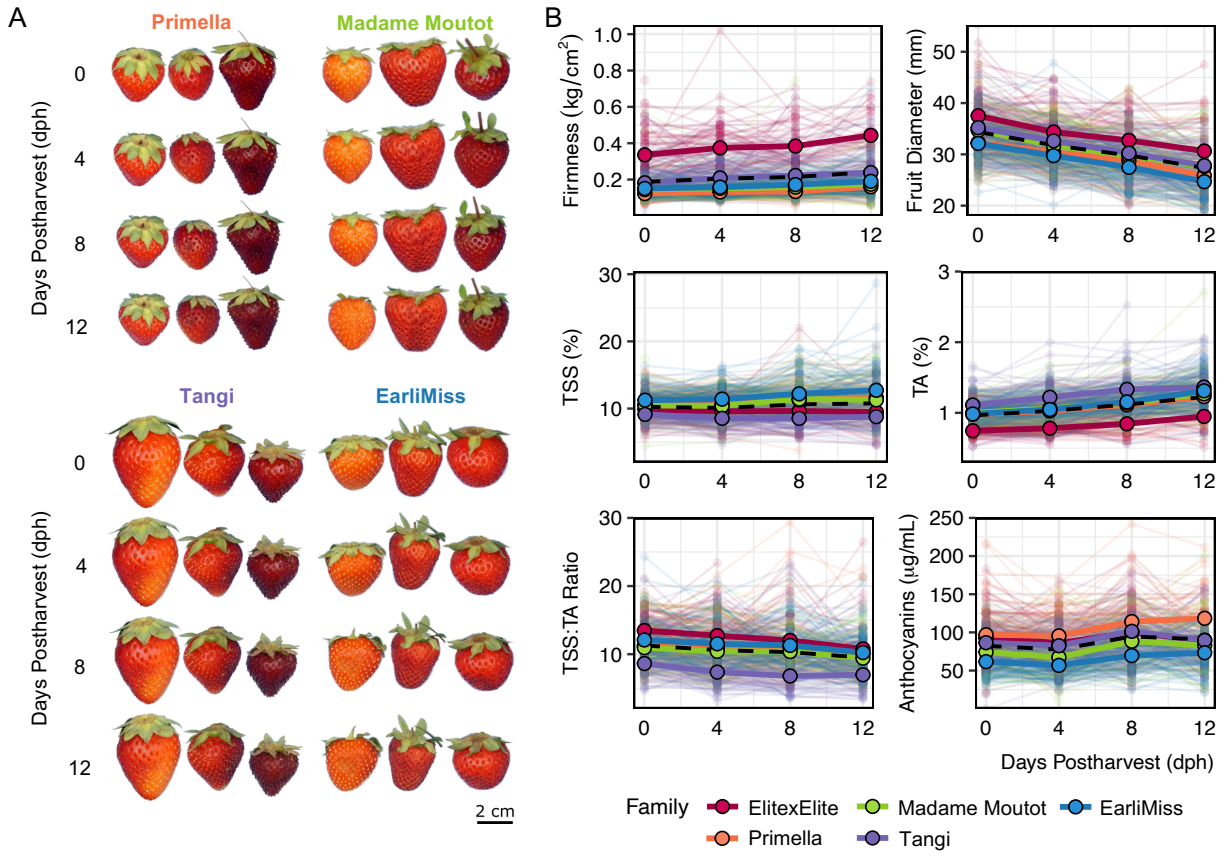

FIGURE 2.6. **Absence of genotype×environment interaction for fruit quality traits in strawberry during post-harvest storage.** Progeny from four elite×exotic crossbred families derived from the elite cultivar 'Royal Royce' and exotic parents ('Primella', 'Madame Moutot', 'Tangi', and 'EarliMiss'), along with an elite×elite backcross family between two California breeding lines (16C108P065 and 05C197P002), were evaluated over a 12-day post-harvest period (dph). Phenotypic changes of A) strawberry fruit in elite×exotic full-sib families at post-harvest, including B) firmness, fruit diameter, total soluble solids (TSS), titratable acidity (TA), TSS-TA ratio, and anthocyanin content. Solid and thick lines indicate the average value of the full-sib family, while dashed and black lines are the average value of the elite×elite and elite×exotic populations.

Lerceteau-Köhler et al., 2012; Masny et al., 2016; Mishra et al., 2015; Shaw, 1988; Singh et al., 2018; Ukalska et al., 2006; Whitaker et al., 2012; Zareei et al., 2022; Zorrilla-Fontanesi et al., 2011). Genetic correlations can vary among populations from changes in additive genetic variances influenced by different environmental conditions (Holloway et al., 1990; Sgrò and Hoffmann, 2004) and inbreeding (Rose, 1984; Whitlock and Fowler, 1999). Therefore, our

findings may be attributed to the population structure differences between the elite×elite and elite×exotic populations (Figure 2.3). The elite×elite family was closely related to the California population, which is highly inbred and has a narrow genetic base. In contrast, the elite×exotic families represent new subpopulations created through the introgression of novel alleles from exotic parents. Additionally, the population size in the elite×elite family (n = 85) may have affected the sampling additive genetic variance of the genetic correlations among fruit quality traits. The small sample size from the elite×elite family may have introduced bias and weighted more for the genetic correlations among fruit quality traits (Visscher, 1998). Although some genetic correlations showed opposite directions when the elite×elite family was included in the analysis, our study suggests a trade-off between firmness and consumer preference-linked traits (e.g., sugar content). The identification of loci without pleiotropic effects may help break down these negative correlations and improve the overall liking of strawberry fruit.

2.4.4. Large-Effect Genetic Sources Within Families Influence Fruit Quality Traits. We explored the genetic basis of fruit quality traits using the assumption of non-uniform gene distribution across the genome through the Bayesian Information and Linkage Disequilibrium Iteratively Nested Keyway (BLINK) method (Huang et al., 2019). Our analysis revealed several statistically significant signals distributed throughout the genome (Figure 2.8; Table 2.4). We identified 11 significant associations for firmness ($\lambda = 1.18$; Figure 2.8A), four for TSS ($\lambda = 1.10$; Figure 2.8B), five for TA ($\lambda = 1.10$; Figure 2.8C), one for TSS:TA ($\lambda = 1.06$; Figure 2.8D), and four for ANC ($\lambda = 1.01$; Figure 2.8E). Interestingly, two loci were detected in the same chromosome at different physical positions for firmness (Figure 2.8A) and TA (Figure 2.8C). Moreover, our analysis did not reveal co-localized loci among fruit quality traits. These findings, along with previous studies (Alarfaj et al., 2021; Castro and Lewers, 2016; Lee et al., 2021; Lerceteau-Köhler et al., 2012; Natarajan et al., 2020; Rey-Serra et al., 2021; Verma et al., 2017; Zorrilla-Fontanesi et al., 2011), suggest that fruit quality traits are controlled by a complex genetic architecture in strawberry.

TABLE 2.4. **Genome-Wide Association Study (GWAS) statistics for firmness, total soluble solids (TSS), titratable acidity (TA), TSS-TA ratio, and anthocyanin content (ANC) using 795 hybrids genotyped with a 50K Axiom^{TM} SNP array.**

Trait	**Marker**	**Chr**[a]	**Position (bp)**[b]	***p*-value**[c]	**Effect**[d]
Firmness	AX-184140198	1A	18891080	1.26×10^{-3}	-0.03
	AX-184605081	2C	11102222	1.83×10^{-3}	-0.03
	AX-184579310	3A	14173665	2.49×10^{-3}	0.02
	AX-166512777	3B	24485373	2.89×10^{-3}	-0.01
	AX-166512704	3D	8534698	1.40×10^{-4}	0.03
	AX-184863957	4B	28803270	1.54×10^{-7}	0.03
	AX-184162213	5C	10072899	1.26×10^{-4}	-0.05
	AX-184608879	6A	7765714	3.79×10^{-11}	-0.07
	AX-184089164	6A	25964115	1.74×10^{-7}	-0.01
	AX-166527345	6C	31891573	4.47×10^{-6}	-0.05
	AX-184244674	7A	17402423	1.26×10^{-4}	0.04
TSS	AX-184312863	1B	15148429	4.74×10^{-3}	-0.64
	AX-184148450	2C	10789397	1.45×10^{-4}	-0.72
	AX-184416160	3B	9710031	1.18×10^{-4}	0.55
	AX-184643646	6A	17383194	1.70×10^{-7}	0.57
TA	AX-184495179	2A	12662897	1.31×10^{-5}	-0.13
	AX-184166518	2A	21657588	3.12×10^{-3}	0.10
	AX-166504739	3D	29585699	6.89×10^{-3}	-0.05
	AX-184251393	4B	21918607	6.89×10^{-3}	-0.11
	AX-184419884	5C	5024907	7.43×10^{-4}	-0.04
TSS:TA Ratio	AX-184868046	7D	8136080	9.07×10^{-4}	1.45
ANC	AX-184673952	1B	7893376	4.16×10^{-6}	-11.08
	AX-184439966	5A	6936484	8.70×10^{-3}	-9.89
	AX-123363930	5B	8770250	1.29×10^{-4}	8.34
	AX-184367856	7A	20878974	2.62×10^{-3}	11.73

[a] Chromosome number.

[b] Physical position mapped in the haplotype-phased 'Royal Royce' reference genome (FaRR V1).

[c] FDR-corrected *p*-value.

[d] Estimated effect size of a genetic variant on the trait.

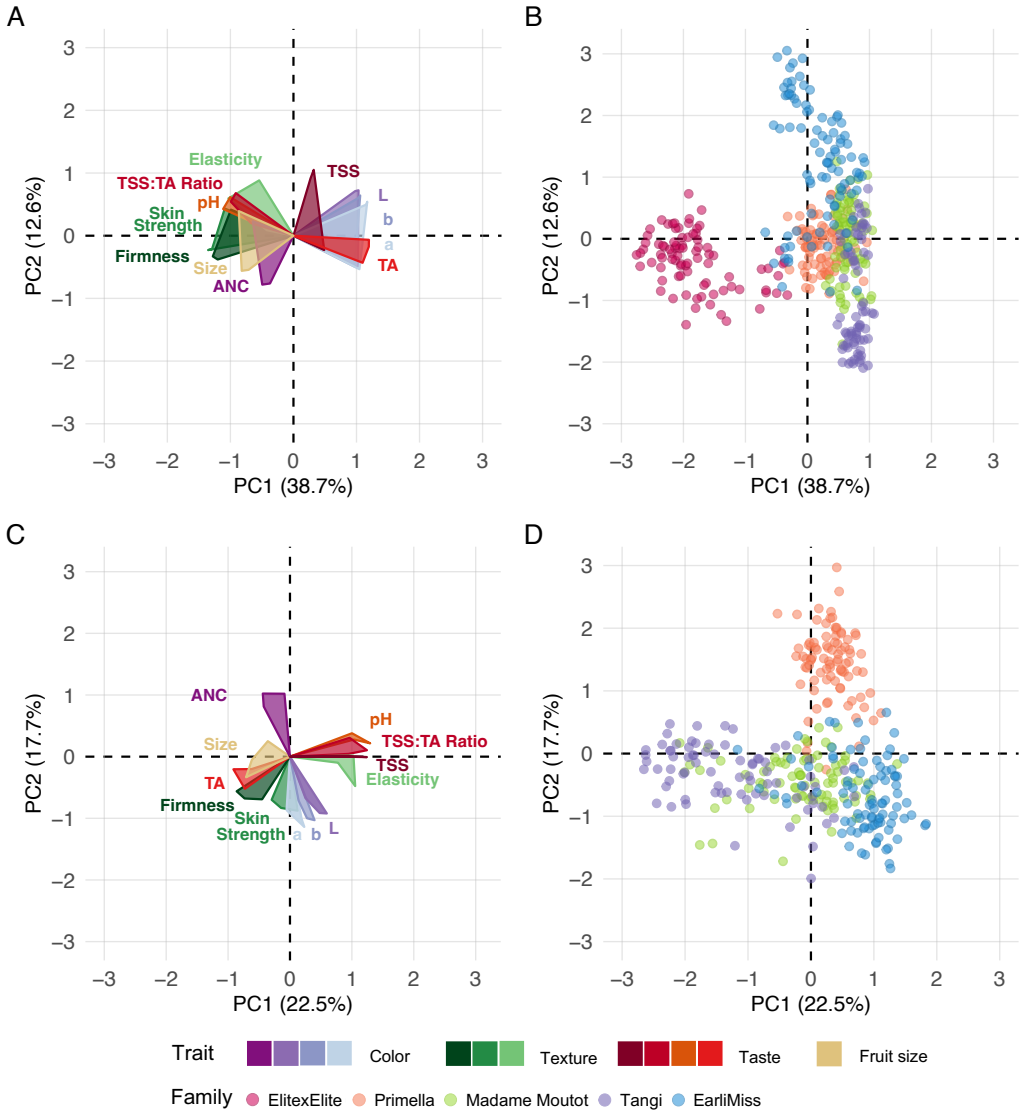

FIGURE 2.7. **Genetic correlations among fruit quality traits in strawberry.** Genomic estimated breeding values were calculated for texture (firmness, skin strength, and elasticity), taste (total soluble solids (TSS), titratable acidity (TA), TSS:TA ratio, and pH), fruit size, anthocyanin content (ANC), and color components (L, a, and b). The two principal components were obtained from an elite×elite and four elite×exotic full-sib families ('Primella', 'Madame Moutot', 'Tangi', and 'EarliMiss') evaluated at 0, 4, and 8 days post-harvest. Genetic correlations for fruit quality traits were analyzed A) and B) by including both populations, while C) and D) considering exclusively the elite×exotic full-sib families. Polygons represent the distribution patterns of fruit quality traits during post-harvest in the principal component analysis.

The most significant single nucleotide polymorphisms (SNPs) associated with fruit quality traits (Table 2.4), along with the interaction between the identified loci as well as those with two loci on the same chromosome, were fitted into a linear model. Our analysis indicated that the genetic variance explained (GVE) and phenotypic variance explained (PVE) by all firmness-linked loci were 76.53% and 49.05%, respectively (Table 2.5). We observed one large-effect locus on chromosome 1A (GVE = 21.41% and PVE = 13.75%) for firmness. However, three large-effect loci were identified within full-sib families from the elite×exotic population (Table 2.5). These loci are located on chromosome 3D in Morioka 17 (GVE = 26% and PVE = 16.08%), 4B in Titan (GVE = 35.52% and PVE = 11.78%), and 6A in both 'Primella' (GVE = 37.81% and PVE = 17.58%) and 'EarliMiss' (GVE = 29.87% and PVE = 15.25%). The large-effect locus associated with firmness on chromosome 6A has been previously identified (Cockerton et al., 2021; Hardigan et al., 2018; Lee et al., 2021). This genomic region contains three tandemly arranged polygalacturonase (PG) genes reported by Hardigan et al. (2021b). PGs are enzymes involved in pectin degradation, which influences cell wall integrity and fruit softening in strawberry (Ramos et al., 2018). One of these tandem PG genes, named *FaPG1*, has been extensively studied for its role in strawberry firmness (Julio-González et al., 2018; López-Casado et al., 2023; Paniagua et al., 2020,2; Pose et al., 2015; Posé et al., 2013; Quesada et al., 2009; Ramos et al., 2018; Redondo-Nevado et al., 2001; Villarreal et al., 2009,0). Strikingly, the knockout of *PG1* did not significantly affect other fruit quality-related traits of strawberries, such as sweetness, color, and nutrient content (Julio-González et al., 2018; Pose et al., 2015; Quesada et al., 2009). The unlinked effect of *PG1* on other fruit quality traits suggests that the general appeal of strawberries can be improved with minimal trade-offs by identifying loci without pleiotropic effects. On the other hand, a small fraction of the genetic and phenotypic variation was explained by all the identified loci for TSS (GVE = 32.04% and PVE = 11.85%; Table 2.6), TA (GVE = 13.08% and PVE = 7.09%; Table 2.7), TSS:TA (GVE = 13.02% and PVE = 4.75%; Table 2.8), and ANC (GVE = 30.72% and PVE = 11.23%; Table 2.9). We detected one

large-effect locus on chromosome 5C for TA in Madame Moutot (GVE = 20.32% and PVE = 11.28%), while the loci on chromosomes 1B and 5B had a major effect on ANC in the elite×elite population (GVE = 35.16% and PVE = 10.30%) and in Kaoling (GVE = 88.46% and PVE = 71.03%), respectively. Additionally, the interaction between loci did not explain much of the phenotypic and genetic variation observed for fruit quality traits. We observed variable effects across and within families for the studied traits. For instance, epistatic effects were observed for firmness both across and within two specific families (Table 2.5). In contrast, no epistatic interaction between loci for TSS was detected across families. However, epistatic effects were present in five of the six families with more than one locus (Table 2.6). Therefore, we found that the additive genetic component of each locus explained a high proportion of the phenotypic and genotypic variance for fruit quality traits. Monnahan and Kelly (2015) used a similar approach to estimate pairwise epistasis for several floral traits in *Mimulus guttatus*. Although they identified substantial epistatic interactions for most traits and showed variable effects across different genetic backgrounds, the genetic variance was predominantly explained by the additive effects of individual loci.

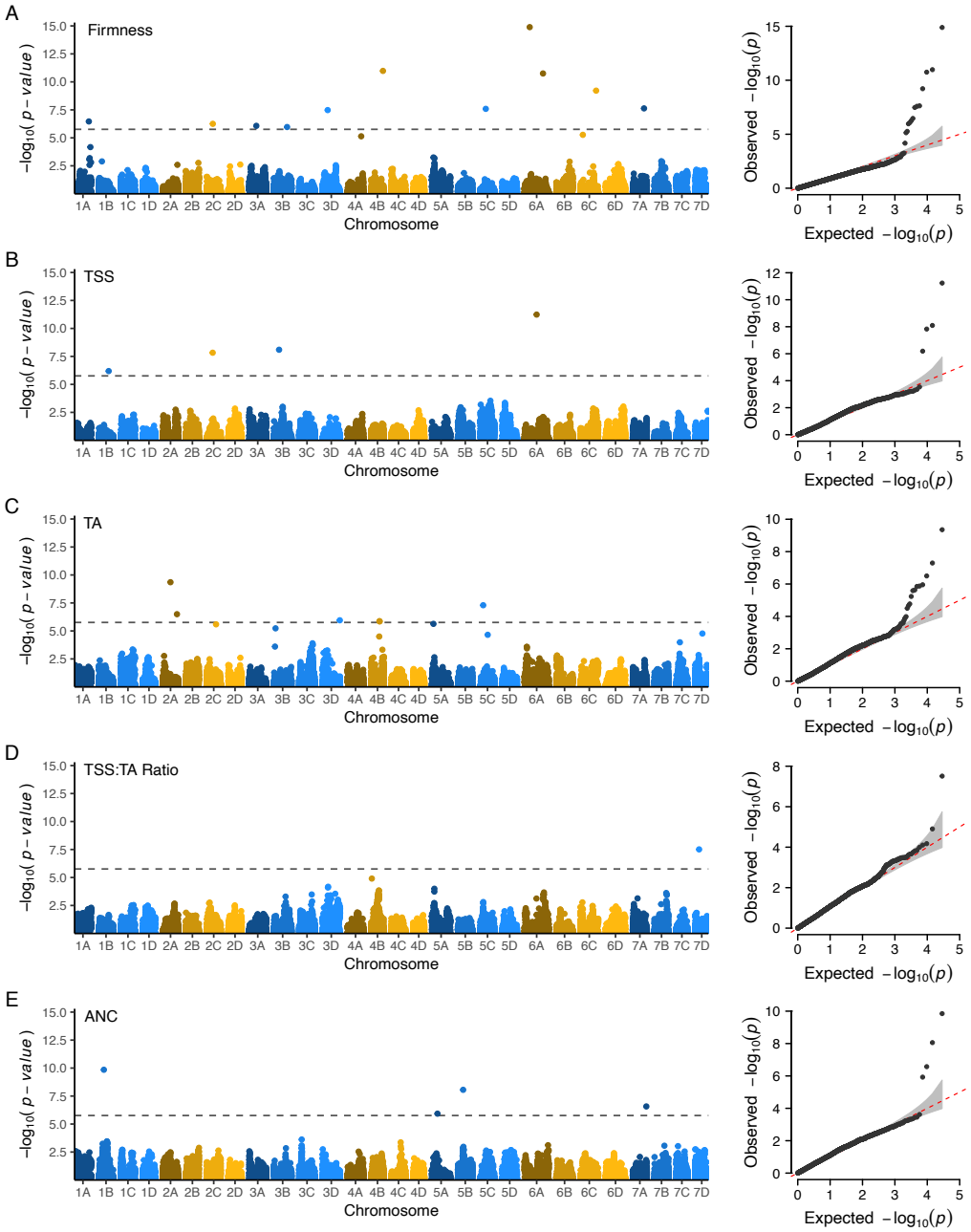

FIGURE 2.8. **Genome-wide association analysis uncovers loci involved in fruit quality traits in strawberry.** Estimated marginal means (EMMs) were calculated from two harvest times in 2019. Through a genome-wide association study, EMMs of 795 individuals were statistically associated with 31,269 single nucleotide polymorphisms (SNPs). Manhattan and Q-Q plots for A) firmness, B) total soluble solids (TSS), C) titratable acidity (TA), and anthocyanin content (ANC). Dashed lines indicate the Bonferroni threshold. SNP physical positions were mapped to the haplotype-phased genome of 'Royal Royce' (FaRR V1).

TABLE 2.5. **Contribution of single- and multi-locus effects to phenotypic and genetic variation for firmness in strawberry.** The percentage of phenotypic (PVE) and genetic variance explained (GVE) by all identified single nucleotide polymorphisms (SNPs) for firmness using a genome-wide association study was estimated by fitting single SNPs and their interactions in a linear model. PVE and GVE were calculated within the studied population and across the ten full-sib families. Dashes indicate loci removed from the analysis due to either a minor allele frequency ≤ 0.05 or the observation of only one genotypic class.

Chr[a]		**Population**	**elite×elite**	**Kaoling**	**Morioka 17**	**Primella**	**Madame Moutot**	**Titan**	**MDUS 5130**	**Tangi**	**EarliMiss**	**Linn**
1A	PVE (%)	13.75	5.13	-	-	0.00	-	1.83	0.00	-	-	0.00
	GVE (%)	21.45	10.88	-	-	0.00	-	5.05	0.00	-	-	0.00
2C	PVE (%)	2.42	-	1.46	-	0.00	0.00	-	-	0.00	-	0.00
	GVE (%)	3.78	-	3.98	-	0.00	0.00	-	-	0.00	-	0.00
3A	PVE (%)	1.59	1.34	0.00	0.00	0.00	4.84	0.00	0.00	4.62	0.27	0.00
	GVE (%)	2.48	2.84	0.00	0.00	0.00	7.06	0.00	0.00	12.21	0.53	0.00
3B	PVE (%)	2.19	-	1.33	0.00	0.00	2.16	1.98	4.40	0.48	0.00	0.00
	GVE (%)	3.42	-	3.62	0.00	0.00	3.15	5.48	10.21	1.26	0.00	0.00
3D	PVE (%)	1.79	-	0.00	16.08	-	0.00	-	0.58	0.00	-	-
	GVE (%)	2.79	-	0.00	26.00	-	0.00	-	1.35	0.00	-	-
4B	PVE (%)	6.46	4.39	-	-	-	-	11.78	-	-	-	-
	GVE (%)	10.08	9.30	-	-	-	-	32.52	-	-	-	-
5C	PVE (%)	4.39	0.00	2.21	0.00	0.00	0.00	-	-	-	-	2.88
	GVE (%)	6.85	0.00	6.01	0.00	0.00	0.00	-	-	-	-	11.45
6A-1	PVE (%)	6.61	-	-	0.00	-	-	-	-	-	-	-
	GVE (%)	10.31	-	-	0.00	-	-	-	-	-	-	-
6A-2	PVE (%)	3.26	-	0.00	5.95	17.58	8.99	-	1.55	3.66	15.25	1.95
	GVE (%)	5.09	-	0.00	9.62	37.81	13.10	-	3.60	9.67	29.87	7.75

Table 2.5 continued from previous page

6C	PVE (%)	3.92	-	3.80	5.55	-	-	-	-	-	-	0.00
	GVE (%)	6.11	-	10.33	8.97	-	-	-	-	-	-	0.00
7A	PVE (%)	2.67	-	11.09	-	-	0.75	-	15.87	0.00	-	0.84
	GVE (%)	4.17	-	30.19	-	-	1.09	-	36.80	0.00	-	3.33
6A-1:2[b]	PVE (%)	0.00	-	-	5.88	-	-	-	-	-	-	-
	GVE (%)	0.00	-	-	9.50	-	-	-	-	-	-	-
Interaction[c]	PVE (%)	3.42	0.00	0.00	0.00	0.00	0.00	2.48	1.81	0.00	0.00	0.00
	GVE (%)	5.33	0.00	0.00	0.00	0.00	0.00	6.84	4.19	0.00	0.00	0.00

[a]Chromosome for the most significant SNPs identified. Chr6A-1 and Chr6A-2 indicate the first and second peaks detected on chromosome 6A, respectively.

[b]Interaction term between 6A-1 and 6A-2.

[c]Interaction term between the most significant SNPs detected for firmness.

TABLE 2.6. **Contribution of single- and multi-locus effects to phenotypic and genetic variation for total soluble solids in strawberry.** The percentage of phenotypic (PVE) and genetic variance explained (GVE) by all identified single nucleotide polymorphisms (SNPs) for total soluble solids using a genome-wide association study was estimated by fitting single SNPs and their interactions in a linear model. PVE and GVE were calculated within the studied population and across the ten full-sib families. Dashes indicate loci removed from the analysis due to either a minor allele frequency ≤ 0.05 or the observation of only one genotypic class.

Chr[a]		**Population**	**elite×elite**	**Kaoling**	**Morioka 17**	**Primella**	**Madame Moutot**	**Titan**	**MDUS 5130**	**Tangi**	**EarliMiss**	**Linn**
1B	PVE (%)	2.37	-	-	-	-	-	0.00	-	0.13		-
	GVE (%)	6.41	-	-	-	-	-	0.00	-	0.80		-
2C	PVE (%)	1.09	-	-	-	2.55	-	-	-	-	0.00	-
	GVE (%)	2.94	-	-	-	5.21	-	-	-	-	0.00	-
3B	PVE (%)	3.36	0.64	0.00	-	-	0.00	0.00	-	-	0.62	-
	GVE (%)	9.09	1.35	0.00	-	-	0.00	0.00	-	-	2.37	-
6A	PVE (%)	5.03	1.26	0.00	-	8.24	0.00	-	-	-	0.20	2.34
	GVE (%)	13.60	2.69	0.00	-	16.93	0.00	-	-	-	0.76	6.40
Interaction[b]	PVE (%)	0.00	7.96	5.39	-	3.51	11.78	1.88	-	-	0.00	-
	GVE (%)	0.00	16.91	33.01	-	7.22	29.79	10.98	-	-	0.00	-

[a]Chromosome for the most significant SNPs identified. Chr6A-1 and Chr6A-2 indicate the first and second peaks detected on chromosome 6A, respectively.

[b]Interaction term between the most significant SNPs detected for firmness.

TABLE 2.7. **Contribution of single- and multi-locus effects to phenotypic and genetic variation for titratable acidity in strawberry.** The percentage of phenotypic (PVE) and genetic variance explained (GVE) by all identified single nucleotide polymorphisms (SNPs) for firmness using a genome-wide association study was estimated by fitting single SNPs and their interactions in a linear model. PVE and GVE were calculated within the studied population and across the ten full-sib families. Dashes indicate loci removed from the analysis due to either a minor allele frequency ≤ 0.05 or the observation of only one genotypic class.

Chr[a]		**Population**	**elite×elite**	**Kaoling**	**Morioka 17**	**Primella**	**Madame Moutot**	**Titan**	**MDUS 5130**	**Tangi**	**EarliMiss**	**Linn**
2A-1	PVE(%)	2.08	-	-	-	-	-	-	-	3.19	-	-
	GVE(%)	3.84	-	-	-	-	-	-	-	8.27	-	-
2A-2	PVE(%)	0.00	-	5.34	-	-	-	-	-	-	-	-
	GVE(%)	0.00	-	10.00	-	-	-	-	-	-	-	-
3D	PVE(%)	1.18	5.85	0.00	0.00	2.83	0.06	5.29	0.00	0.58	0.00	0.56
	GVE(%)	2.17	11.22	0.00	0.00	5.88	0.12	7.09	0.00	1.50	0.00	0.87
4B	PVE(%)	0.00	7.87	-	-	0.00	0.00	-	0.00	2.79	0.31	-
	GVE(%)	0.00	15.09	-	-	0.00	0.00	-	0.00	7.22	0.48	-
5C	PVE(%)	3.83	2.02	1.18	7.39	7.46	11.28	0.00	3.93	0.89	0.00	0.00
	GVE(%)	7.07	3.87	2.21	14.10	5.50	20.32	0.00	6.11	2.31	0.00	0.00
2A-1:2[b]	PVE(%)	0.10	-	-	-	-	-	-	-	-	-	-
	GVE(%)	0.19	-	-	-	-	-	-	-	-	-	-
Interation[c]	PVE(%)	4.65	0.00	4.18	0.00	3.19	0.00	0.00	0.94	0.00	2.13	0.00
	GVE(%)	8.59	0.00	7.84	0.00	6.63	0.00	0.00	1.46	0.00	3.30	0.00

[a]Chromosome for the most significant SNPs identified. 2A-1 and 2A-2 indicate the first and second peaks detected on chromosome 2A, respectively.

[b]Interaction term between 2A-1 and 2A-2.

[c]Interaction term between the most significant SNPs detected for firmness.

TABLE 2.8. **Contribution of single-locus effect to phenotypic and genetic variation for sugar:acid ratio in strawberry.** The percentage of phenotypic (PVE) and genetic variance explained (GVE) by the identified single nucleotide polymorphisms (SNP) for the ratio between total soluble solids and titratable acidity using a genome-wide association study was estimated. PVE and GVE were calculated within the studied population and across the ten full-sib families. Dashes indicate locus removed from the analysis due to either a minor allele frequency ≤ 0.05 or the observation of only one genotypic class.

Chr[a]		**Population**	**elitexelite**	**Kaoling**	**Morioka 17**	**Primella**	**Madame Moutot**	**Titan**	**MDUS 5130**	**Tangi**	**EarliMiss**	**Linn**
7D	PVE (%)	4.75	-	-	-	2.38	-	-	-	-	1.27	5.14
	GVE (%)	13.20	-	-	-	4.82	-	-	-	-	7.00	8.62

[a]Chromosome for the most significant SNPs identified.

TABLE 2.9. **Contribution of single- and multi-locus effects to phenotypic and genetic variation for anthocyanin content in strawberry.** The percentage of phenotypic (PVE) and genetic variance explained (GVE) by all identified single nucleotide polymorphisms (SNPs) for anthocyanin content using a genome-wide association study was estimated by fitting single SNPs and their interactions in a linear model. PVE and GVE were calculated within the studied population and across the ten full-sib families. Dashes indicate loci removed from the analysis due to either a minor allele frequency ≤ 0.05 or the observation of only one genotypic class.

Chr[a]		**Population**	**elitexelite**	**Kaoling**	**Morioka 17**	**Primella**	**Madame Moutot**	**Titan**	**MDUS 5130**	**Tangi**	**EarliMiss**	**Linn**
1B	PVE(%)b	6.23	10.30	0.00	0.00	2.02	0.21	4.96	0.90	3.01	-	-
	GVE(%)	17.04	35.16	0.00	0.00	3.58	0.39	10.78	4.29	5.55	-	-
5A	PVE(%)	1.35	0.00	-	-	-	0.00	0.53	0.24	-	0.89	-
	GVE(%)	3.70	0.00	-	-	-	0.00	1.14	1.15	-	1.51	-
5B	PVE(%)	2.23	-	71.03	1.15	0.49	3.43	0.00	0.82	0.51	4.90	0.00
	GVE(%)	6.10	-	88.46	5.75	0.86	6.22	0.00	3.90	0.93	8.34	0.00
7A	PVE(%)	1.42	-	0.00	0.00	-	0.00	-	-	-	-	0.00
	GVE(%)	3.88	-	0.00	0.00	-	0.00	-	-	-	-	0.00
Interaction[b]	PVE(%)	0.00	1.68	0.06	0.00	0.00	0.00	6.53	0.00	0.00	0.92	0.00
	GVE(%)	0.00	5.76	0.08	0.00	0.00	0.00	14.18	0.00	0.00	1.57	0.00

[a] Chromosome for the most significant SNPs identified.

[b] Interaction term between the most significant SNPs detected for anthocyanin content.

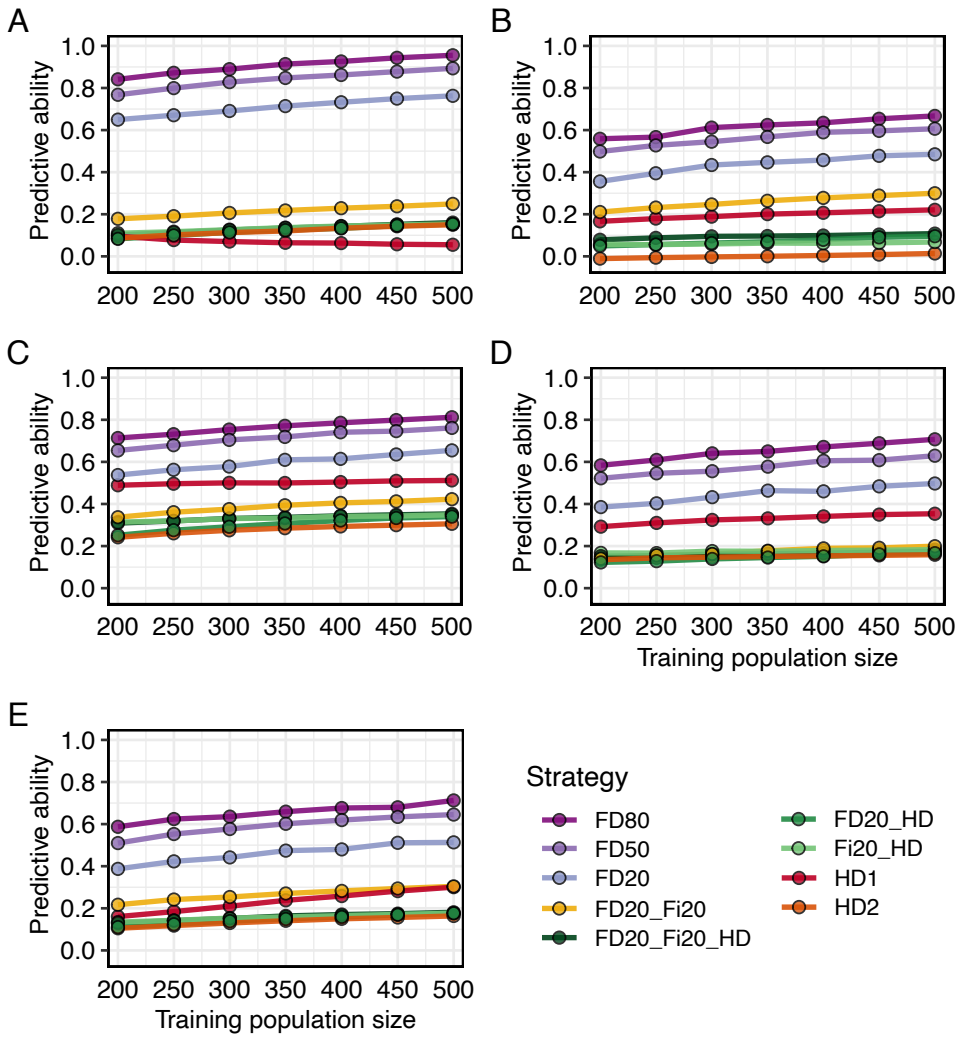

FIGURE 2.9. **Training set optimization to improve genomic prediction for fruit quality traits in strawberry.** Genomic-estimated breeding values (GEBVs) were obtained from estimated marginal means (EMMs) through Genomic Best Linear Unbiased Predictions (G-BLUP). EMMs were calculated for 795 individuals from a single plant per individual at two harvests in 2018-2019. Individuals were genotyped with an Axiom^{TM} 50K SNP array. We used Monte Carlo Markov Chain simulations with 1,000 iterations to randomly select individuals for the training set using non-replacement sampling to estimate GEBVs for the test set. The simulations were executed under nine scenarios described in Figure 2.1. The predictive ability was calculated as the correlation between true breeding values and GEBVs divided by the square root of the narrow-sense heritability for A) firmness, B) total soluble solids (TSS), C) titratable acidity (TA), D) TSS-TA ratio, and anthocyanin content (ANC).

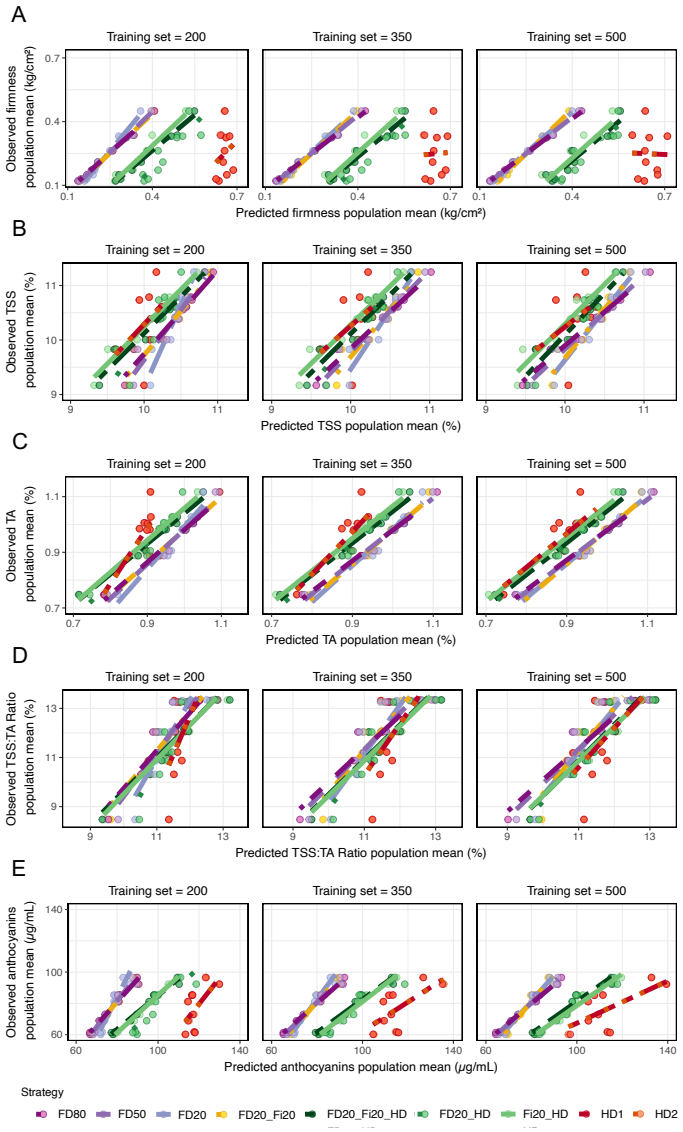

FIGURE 2.10. **Accuracy of population mean estimates using genomic prediction strategies for fruit quality traits in strawberry.** Pearson's correlation between the observed and predicted family mean for A) firmness, B) total soluble solids (TSS), C) titratable acidity (TA), D) TSS-TA ratio, and E) anthocyanin content. Genomic-estimated breeding value (GEBV) mean for each family was calculated from the average of progeny GEBVs. Monte Carlo Markov Chain simulations with 1,000 iterations were employed for random non-replacement sampling. Nine scenarios were evaluated to select individuals for the training set (Figure 2.1) through seven different training population sizes (from 200 to 500 individuals). Genomic prediction accuracy is shown for the nine breeding strategies using a training set of 200, 350, and 500 individuals.

2.4.5. Harnessing Genomic Prediction for Targeted Improvement of Fruit Quality in Diverse Strawberry Families. Plant breeding programs typically structure their populations using biparental families that rarely exceed 100 individuals each (Lehermeier et al., 2014). These programs face challenges with small populations for population-specific genomic prediction analysis, leading to the inclusion of unrelated populations (De Roos et al., 2009; Lehermeier et al., 2014). The balance of incorporating information from both related and unrelated populations into training sets (TS) for genomic prediction remains uncertain. Several genomic prediction strategies were analyzed to improve genomic prediction for fruit quality traits using the ten full-sib families studied here (Figure 2.1). We calculated progeny and family genomic-estimated breeding values (GEBVs) by targeting individuals across or within full-sib families. Genomic prediction was evaluated across full-sib families using individuals for the TS from either i) 80%, 50%, and 20% of the total population (FD80, FD50, and FD20, respectively) or ii) historical records analyzed by Feldmann et al. (2024a) (HD1). In the case of within families, individuals were selected from iii) 20% of the target family and 20% across full-sib families (FD20_Fi20), iv) 20% of the target family, 20% across full-sib families, and historical records (FD20_Fi20_HD), v) 20% of the target family and historical records (Fi20_HD), vi) 20% of the population and historical records (FD20_HD), or vii) historical records (HD2). These genomic prediction strategies were implemented through seven different TS sizes.

Predictive ability (PA) of progeny GEBVs showed substantial differences among genomic prediction strategies for fruit quality traits (Figure 2.9). We observed higher PA values when progeny GEBVs were estimated across full-sib families than within families. For instance, the average PA of progeny GEBVs for firmness was higher using the FD80 ($\overline{PA} = 0.91$) strategy, following by FD50 ($\overline{PA} = 0.84$), FD20 ($\overline{PA} = 0.71$), FD20_Fi20 ($\overline{PA} = 0.22$), FD20_Fi20_HD ($\overline{PA} = 0.13$), FD20_HD ($\overline{PA} = 0.13$), Fi20_HD, FD20_HD ($\overline{PA} = 0.12$), HD2 ($\overline{PA} = 0.12$), and HD1 ($\overline{PA} = 0.07$). The low PA observed for fruit quality traits within families may be influenced by using only 20% of the target family in the TS, which is likely insufficient for

accurately predicting the Mendelian sampling term within the family (Werner et al., 2020). Moreover, PA values were generally lower when the TS included individuals from historical records. Previous studies have shown that TS consisting of individuals related to the target population provide higher PA than those with unrelated individuals (Brauner et al., 2020; Clark et al., 2012; Herter et al., 2019; Riedelsheimer et al., 2013; Stewart-Brown et al., 2019). Interestingly, our forward validation using historical records to estimate progeny GEBVs across full-sib families (HD1) exhibited comparable PA to the FD20_Fi20 strategy for TSS (Figure 2.9B) and ANC (Figure 2.9E). Additionally, the HD1 strategy showed higher PA for TA (Figure 2.9C) and TSS:TA ratio (Figure 2.9D) than within family strategies. Individuals from the elite×exotic population were not completely distant from historical records (Figure 2.3), likely sharing ancestral linkage disequilibrium (LD) between markers and quantitative trait loci (QTL). In general, our results may be attributed to differences in LD (De Roos et al., 2009; Schopp et al., 2017a), QTL×genetic background interaction (Lorenz and Cohen, 2012; Melchinger et al., 2008; Mohammadi et al., 2015a), and the presence of population-specific QTL alleles (Lorenz et al., 2012; Schulz-Streeck et al., 2012).

The correlation between observed and predicted family means varied across genomic prediction strategies (Figure 2.10). Our results showed strong correlations across genomic prediction strategies that included individuals from the elite×elite and elite×exotic full-sib families ($r \geq 0.85$). The HD1 and HD2 strategies exhibited weak correlations for firmness ($\overline{r} = 0.08$) and moderate to high for TSS ($\overline{r} = 0.45$), TSS-TA ratio ($\overline{r} = 0.56$), ANC ($\overline{r} = 0.63$), and TA ($\overline{r} = 0.77$). Moreover, normalized root-mean-square error (nRMSE) indicated genomic prediction accuracy is highly affected by unrelated strawberry germplasm (Table 2.10). For instance, genomic prediction strategies for firmness showed lower nRMSE when the TS was exclusively built from individuals in the elite×elite and elite×exotic full-sib families (FD20, FD50, FD80, FD20_Fi20; nRMSE = 0.03-0.13) compared to either combined data (FD20_HD, Fi20_HD, FD20_Fi20_HD; nRMSE = 0.38-0.63) or historical records alone

(HD1 and HD2; nRMSE = 1.25-1.27). The high nRMSE values in genomic prediction strategies that included individuals from historical records may be due to missing relatedness and different sets of QTL segregating to the target population (Lehermeier et al., 2014; Schopp et al., 2017b). These findings underscore the importance of optimizing the TS in genomic prediction for complex traits to achieve desired breeding outcomes in strawberry breeding programs.

The TS size did not show great differences in genomic prediction accuracy for fruit quality traits (Figure 2.9 and 2.10). On average, we observed an increase of up to 0.01 in PA with the addition of 50 individuals to the training population. Several studies have shown that increasing the TS size improves PA but with diminishing returns at higher TS sizes (Herter et al., 2019; Jarquín et al., 2014; Osorio et al., 2021; Stewart-Brown et al., 2019; Windhausen et al., 2012; Zhang et al., 2017). We observed a decrease in PA for firmness when larger TS sizes for the HD1 strategy were used to estimate progeny GEBVs (Figure 2.9A). The correlation between the observed and predicted family mean dropped from 0.23 (TS = 200) to -0.02 (TS = 500), which may be explained by opposite linkage phases with key QTLs between the TS and the target population for firmness (Riedelsheimer et al., 2013). Future efforts should focus on optimizing relatedness in TS design to enhance short-term PA without compromising long-term genetic gains. Osorio et al. (2021) suggested using average predictions from multiple cycles in training genomic prediction models, particularly up to four cycles when common individuals are included across trials in strawberries. Consequently, frequent re-estimation of SNP effects is essential to incorporate the most recent phenotypic data from relatives (Habier et al., 2010). Our results advocate for combining several half-sib families when available to form the TS. If insufficient individuals from such families are available, other highly related genetic sources should be considered.

TABLE 2.10. **Genomic prediction performance of family mean estimates for fruit quality traits in strawberry.** Normalized root-mean-square error was calculated across genomic prediction strategies and training population sizes (Figure 2.1) for firmness, total soluble solids (TSS), titratable acidity (TA), TSS-TA ratio, and anthocyanin content (ANC).

Trait	Training size	FD80	FD50	FD20	FD20_Fi20	FD20_HD	Fi20_HD	FD20_Fi20_HD	HD1	HD2
Firmness	200	0.06	0.08	0.13	0.06	0.63	0.38	0.42	1.27	1.27
	250	0.05	0.06	0.11	0.07	0.63	0.42	0.46	1.27	1.27
	300	0.04	0.06	0.10	0.08	0.62	0.44	0.48	1.27	1.27
	350	0.04	0.05	0.09	0.08	0.61	0.46	0.50	1.26	1.26
	400	0.04	0.05	0.08	0.09	0.61	0.49	0.52	1.26	1.26
	450	0.03	0.04	0.08	0.09	0.61	0.51	0.54	1.26	1.26
	500	0.03	0.04	0.07	0.09	0.61	0.53	0.55	1.25	1.25
TSS	200	0.11	0.14	0.19	0.12	0.20	0.16	0.12	0.28	0.28
	250	0.10	0.13	0.18	0.12	0.19	0.17	0.13	0.28	0.28
	300	0.08	0.11	0.17	0.13	0.18	0.17	0.13	0.28	0.28
	350	0.08	0.10	0.16	0.13	0.18	0.18	0.14	0.27	0.27
	400	0.07	0.10	0.15	0.13	0.17	0.18	0.14	0.27	0.27
	450	0.06	0.09	0.15	0.13	0.17	0.19	0.15	0.27	0.27
	500	0.06	0.08	0.14	0.13	0.16	0.19	0.15	0.27	0.27
TA	200	0.08	0.09	0.12	0.08	0.15	0.13	0.10	0.28	0.28
	250	0.07	0.08	0.11	0.08	0.14	0.14	0.11	0.27	0.27
	300	0.07	0.08	0.10	0.09	0.13	0.15	0.11	0.26	0.26
	350	0.07	0.08	0.10	0.09	0.13	0.15	0.11	0.26	0.26
	400	0.07	0.07	0.09	0.09	0.13	0.15	0.11	0.26	0.25
	450	0.07	0.07	0.09	0.09	0.13	0.16	0.11	0.25	0.25

Table 2.10 continued from previous page

	500	0.06	0.07	0.09	0.09	0.12	0.16	0.11	0.25	0.25
TSS:TA Ratio	200	0.17	0.19	0.22	0.17	0.20	0.14	0.14	0.27	0.27
	250	0.16	0.18	0.22	0.18	0.19	0.14	0.14	0.27	0.27
	300	0.16	0.18	0.21	0.18	0.18	0.14	0.14	0.27	0.27
	350	0.16	0.17	0.21	0.19	0.18	0.14	0.14	0.26	0.26
	400	0.15	0.16	0.20	0.19	0.17	0.14	0.14	0.26	0.26
	450	0.15	0.16	0.19	0.19	0.17	0.14	0.14	0.26	0.26
	500	0.15	0.16	0.19	0.19	0.16	0.14	0.14	0.26	0.26
ANC	200	0.12	0.15	0.23	0.13	0.63	0.45	0.43	1.18	1.18
	250	0.11	0.14	0.21	0.14	0.61	0.49	0.46	1.15	1.15
	300	0.10	0.13	0.19	0.15	0.59	0.51	0.48	1.13	1.13
	350	0.09	0.12	0.18	0.15	0.58	0.53	0.49	1.10	1.10
	400	0.08	0.11	0.17	0.16	0.58	0.55	0.51	1.08	1.08
	450	0.07	0.10	0.17	0.16	0.57	0.57	0.51	1.07	1.07
	500	0.07	0.09	0.15	0.16	0.57	0.59	0.52	1.06	1.06

2.5. Conclusions

Our study shows that incorporating exotic genetic resources into elite strawberry breeding programs can significantly enhance fruit quality traits. We evaluated the introgression of favorable alleles for fruit quality from nine soft-fruited exotic parents into a firm-fruited elite cultivar, observing significant phenotypic variation within the resulting hybrid population. These hybrids exhibited notable improvements in traits such as TSS and TA, indicating the potential for significant genetic gain through strategic hybridization. The genetic correlations among fruit quality traits revealed a trade-off between firmness and consumer-preferred attributes like sweetness and acidity, suggesting the complexity of improving multiple traits simultaneously. Moreover, the lack of genotype×timepoint interaction for most traits indicated that fruit quality evaluation at harvest could reliably predict post-harvest performance. We identified several large-effect loci associated with fruit quality harbored within specific full-sib families. Notably, a significant locus on chromosome 6A linked to firmness was consistent with previous studies, suggesting its potential as a target for marker-assisted selection. The genetic architecture of these traits, predominantly additive in nature, underscores the importance of selecting parents with favorable alleles to maximize genetic gains. Genomic prediction showed promise in predicting complex traits, with predictive abilities varying across different TS strategies. TS designs that included closely related individuals, particularly half-sib families, yielded higher predictive accuracies compared to those including unrelated individuals. This highlights the need to optimize TS composition to enhance the effectiveness of genomic selection in breeding programs. In summary, our findings suggest that exotic resources can lead to substantial improvements in fruit quality traits. Genomic-informed breeding strategies can further refine breeding efforts to select superior genotypes. Future research should focus on validating identified loci and optimizing TS designs to maximize the benefits of genomic selection in strawberry breeding programs.

2.6. Abbreviations

LMM, linear mixed model; EMM, estimated marginal mean; TSS, total soluble solids; TA, titratable acidity; ANC, anthocyanin content; UH, useful heterosis; REML, restricted maximum likelihood; PVE, percentage of the phenotypic variance explained by markers; GVE, percentage of the genetic variance explained by markers; GWAS, genome-wide association study; QTL, quantitative trait loci; LD, linkage disequilibrium; SNP, single nucleotide polymorphism; TS, training set; G-BLUP, genomic-best linear unbiased prediction; GEBV, genomic-estimated breeding value; PA, predictive ability; nRMSE, normalized root-mean-square error.

CHAPTER 3

Loss-of-Function Mutations in the Fruit Softening Gene *POLYGALACTURONASE1* Doubled Fruit Firmness in Strawberry

3.1. Abstract

Wildtype fruit of cultivated strawberry (*Fragaria* × *ananassa*) are typically soft and highly perishable when fully ripe. The development of firm-fruited cultivars by phenotypic selection has greatly increased shelf life, decreased post-harvest perishability, and driven the expansion of strawberry production worldwide. Hypotheses for the firm-fruited phenotype include mutations affecting the expression of genes encoding polygalacturonases that soften fruit by degrading cell wall pectins. Here we show that loss-of-function mutations in the fruit softening gene *POLYGALACTURONASE1* (*PG1-6A1*) double fruit firmness in strawberry. *PG1-6A1* was one of three tandemly duplicated polygalacturonase genes found to be in linkage disequilibrium with a quantitative trait locus affecting fruit firmness on chromosome 6A. *PG1-6A1* was strongly expressed in soft-fruited (wildtype) homozygotes and weakly expressed in firm-fruited (mutant) homozygotes. Genome-wide association, quantitative trait transcript, DNA sequence, and expression-QTL analyses identified genetic variants in linkage disequilibrium with *PG1-6A1* that were positively correlated with fruit firmness and negatively correlated with *PG1-6A1* expression. An *Enhancer/Suppressor-mutator* (*En/Spm*) transposable element insertion was discovered upstream of *PG1-6A1* in mutant homozygotes that we hypothesize transcriptionally downegulates the expression of *PG1-6A1.* The *PG1-6A1* locus was incompletely dominant and explained 26-76% of the genetic variance for fruit firmness among phenotypically diverse individuals. Additional loci are hypothesized to underlie the missing heritability. Highly accurate, codominant genotyping assays were

developed for modifying fruit firmness by marker-assisted selection of the *En/Spm* insertion and SNPs associated with the *PG1-6A1* locus.

3.2. Introduction

The development of cultivars with increased fruit firmness and decreased perishability has been one of the most important domestication changes in cultivated strawberry (*Fragaria* × *ananassa*; Darrow et al., 1966; Feldmann et al., 2024a; Fletcher, 1917; Lawrence et al., 1990; Shaw and Larson, 2008). Wildtypes are generally, but not universally, exceptionally soft and highly perishable when fully ripe, characteristics that enhance seed dispersal in nature (Darrow et al., 1966; Hancock et al., 2001a,0). Significant phenotypic variation for fruit firmness has been observed in natural and domesticated strawberry populations, which has facilitated the development of cultivars with a wide range of fruit firmness and shelf life characteristics (Hancock et al., 2001a,0,1; Shaw and Larson, 2008).

The improvement of fruit firmness has a long history in strawberry breeding with references to cultivars developed in the late 1800s and early 1900s that were firmer than the wildtype (Darrow et al., 1966; Fletcher, 1917). 'Blakemore', developed in 1931, was one of the earliest cultivars reported to "set the standard for firmness" necessary for shipping long distances without significant post-harvest losses (Lawrence et al., 1990). The firmness of that cultivar and other early firm-fruited cultivars, however, has been surpassed by the firmness of 'Royal Royce', 'Eclipse', and other modern long shelf life (LSL) cultivars developed since the 1970s in California (Darrow et al., 1966; Feldmann et al., 2024a; Knapp et al., 2023; Shaw and Larson, 2008). Using wild relatives and heirloom cultivars as benchmarks, Feldmann et al. (2024a) showed that the genetic gains from breeding for increased fruit firmness have been in the 240-770% range since the 1950s. Royce S. Bringhurst has been credited with identifying sthisces of favorable alleles for increased fruit firmness and developing the extremely firm-fruited, modern LSL cultivars that revolutionized strawberry production in California, several of which were analyzed in this study (Darrow et al., 1966; Feldmann et al., 2024a; Shaw and Larson, 2008).

Strawberry is a non-climacteric fruit that ripens gradually and lacks a well-defined climacteric peak, the point in the ripening process when ethylene production and respiration rate sharply increase in climacteric fruits (Given et al., 1988; Gu et al., 2019; Symons et al., 2012). The non-climacteric development of the fruit means that they do not ripen further once harvested and that flavor and texture are determined at harvest. Their shelf life is determined by the speed at which they deteriorate in storage. This post-harvest deterioration of the fruit includes desiccation, dissipation of volatile aromatic compounds, diminished texture and flavor, a decrease in glossiness, and rotting caused by gray mold (*Botrytis cinerea*) and other fungal diseases (Gu et al., 2019; Jouki and Khazaei, 2014; Krivorot and Dris, 2000; Matar et al., 2018; Petrasch et al., 2019,2; Shehata et al., 2020; Symons et al., 2012). The speed of deterioration is partially determined by handling (harvest date, postharvest storage) but is widely known to be strongly correlated with the softness of the fruit at harvest (Darrow et al., 1966; Given et al., 1988; Petrasch et al., 2022).

The phenotypic variation observed for fruit firmness in cultivated strawberry has been shown to be highly correlated with the expression of *POLYGALACTURONASE1* (*FaPG1*), a fruit softening gene expressed in the fleshy receptacle (López-Casado et al., 2023; Paniagua et al., 2020; Quesada et al., 2009). *POLYGALACTURONASE1* (hereafter abbreviated *PG1*) plays an important role in pectin disassembly and fruit softening in the wildtype (Quesada et al., 2009). Hardigan et al. (2021b) discovered that three tandemly duplicated paralogs of *PG1* were contained within a large-effect quantitative trait locus (QTL) for fruit firmness on chromosome 6A in a phenotypically diverse genome-wide association study (GWAS) population. They hypothesized that one of these tandemly duplicated genes was the causal gene underlying the QTL. We targeted these genes in this study of genetic variation for fruit firmness in strawberry, in addition to searching for additional genotype-to-phenotype associations using GWAS, transcript-to-phenotype associations using quantitative trait transcript (QTT) analysis, and genotype-to-transcript associations using expression-quantitative

trait locus (eQTL) analysis (Albert and Kruglyak, 2015; Gilad et al., 2008; Hill et al., 2021; Kliebenstein, 2009; Lappalainen et al., 2013; Zhang et al., 2015).

The fruit firmness QTL on chromosome 6A described by Hardigan et al. (2021b) appears to be proximal to the fruit firmness QTL identified by Cockerton et al. (2021), Lee et al. (2021), Munoz et al. (2024), and Prohaska et al. (2024) in independent segregating populations. Lee et al. (2021) did not identify *PG1* as a candidate gene for the QTL but proposed others. *PG1*, however, has shown to be important in fruit softening (Paniagua et al., 2020,1; Pose et al., 2015; Posé et al., 2013; Quesada et al., 2009). The downregulation of *PG1* by antisense RNA silencing in transgenic plants and CRISPR/Cas9 insertion-mutations of *PG1* in gene-edited plants either greatly decreased or eliminated the expression of *PG1* and increased fruit firmness (López-Casado et al., 2023; Paniagua et al., 2020,2,1; Pose et al., 2015; Posé et al., 2013; Quesada et al., 2009).

By aligning the DNA sequence for the *PG1* gene (`https://www.ncbi.nlm.nih.gov/nuccore/AF380299`; Paniagua et al., 2020; Quesada et al., 2009) targeted in those reverse genetic studies to the Camarosa and Royal Royce reference genomes and cross-referencing positions of physically-anchored array-genotyped SNPs in the present study (Edger et al., 2019; Hardigan et al., 2021a; `https://phytozome-next.jgi.doe.gov/info/FxananassaRoyalRoyce_v1_0`), we hypothesized that the QTL identified in independent forward genetic studies of phenotypic variation for fruit firmness was likely caused by mutations affecting one of the three tandemly duplicated polygalacturonase genes (*PG1-6A1*, *-6A2*, and *-6A3*) found to be in linkage disequilibrium (LD) with the QTL on chromosome 6A. Here we show that a loss-of-function mutation in one of the three genes (*PG1-6A1*) doubled fruit firmness in strawberry, that the *PG1-6A1* gene is strongly expressed in ripe fruit of wildtype homozygotes and weakly expressed in ripe fruit of mutant homozygotes, that the favorable (mutant) allele ($PG1\text{-}6A1^{+}$) explains a significant fraction of the genetic variance for fruit firmness, and that the most firmly fruited cultivars are homozygous for the incompletely dominant mutant allele ($PG1\text{-}6A1^{+}/PG1\text{-}6A1^{+}$). Although the causal mutation remains uncertain, we show

that the mutant phenotype is associated with an *Enhancer/Suppressor-mutator* (*En/Spm*) transposable element insertion upstream of *PG1-6A1* and hypothesize that this transposable element disrupted the transcription of *PG1-6A1* in firm-fruited mutants (Feschotte, 2008). Lastly, we describe highly predictive, high-throughput, codominant genotyping assays for marker-assisted selection (MAS) of genetic variants in linkage disequilibrium with the *PG1-6A1* locus, and discuss the utility of native *PG1-6A1* alleles for modifying fruit firmness in strawberry.

3.3. Materials and Methods

3.3.1. Plant material and phenotyping. The fruit firmness (kg-force) phenotypes reported in this chapter were recorded from samples of ripe fruit harvested from 178 individuals grown in field experiments in Oxnard, Santa Maria, Davis, and Winters, California, and a greenhouse experiment in Davis, California. The number and composition of individuals differed across these experiments. They included 69 *F.* × *ananassa* cultivars developed between 1854 and 2017, 91 clonally propagated hybrid individuals with diverse *F.* × *ananassa* parentage developed at UC Davis, four *F. chiloensis* ecotypes, 13 *F. virginiana* ecotypes, and one *F. vesca* ecotype. Of the 91 hybrid individuals, 12 were developed between 1953 and 2012, and 79 were developed between 2016 and 2020. Of the latter, 43 originated from crosses between elite, firm-fruited UC parents and soft-fruited non-UC heirloom cultivars, whereas 36 originated from crosses between elite, firm-fruited UC parents.

The firmness of ripe fruit (kg-force) was measured using a FR-5120 Digital Fruit Firmness Tester (penetrometer) fitted with a 3 mm puncture probe (Stable Micro Systems Ltd., Goldalming, United Kingdom). The penetrometer was held in a stand with the probe pointed upward, and fruit was lowered horizontally onto the instrument so that the probe targeted receptacle tissue near the latitudinal center.

3.3.2. Discovery population. The phenotypes for the genome-wide association study (GWAS), quantitative trait transcript (QTT), and expression-QTL analyses were collected

from 85 *F.* × *ananassa* individuals grown in Santa Maria and Oxnard, California field experiments in 2020-21 and 2021-22 (hereafter the 'discovery' population). They included 13 cultivars and 72 other hybrid individuals from the UC strawberry breeding program. They were arranged in randomized complete blocks experiment designs and grown on commercial farms using standard management practices. The experimental units were 20-plant plots with plants equally spaced in raised beds planted with diagonal staggering of plants in three rows to a density of 54,362 plants/hectare. The bare-root plants (clones) for these experiments were produced in a commercial nursery (Cedar Point, Dorris, CA) between April and October of each planting year. They were harvested and placed in cold storage (4°C) for less than one week before being planted at the Oxnard and Santa Maria field sites. Three ripe fruit were harvested from each plot in March and April of each year, placed in cold storage (4°C) for less than 24 hours, and phenotyped as described above. Across-year phenotypic means were estimated from 24 observations/individual (2 replications/individual × 3 samples/replication × 2 harvests/year × 2 years).

The transcriptomes of the 85 discovery population individuals were analyzed using RNA samples isolated from ripe fruit harvested in March, 2021 from three 20-plant plot replications/individual in both locations. Those data were previously analyzed and described by Fan et al. (2022) and are available in the NCBI Short-Read Archive under Bioproject #PRJNA787565.

3.3.3. Diversity population. We assembled and phenotyped a genetically diverse collection of clonally propagated hybrid individuals ($n = 92$) for the identification of genetic variants associated with a fruit firmness QTL on chromosome 6A, the identification of genetic variants associated with gene expression differences in different stages of fruit development, the development and validation of high-throughput genotyping assays for SNPs and INDELs, specifically Kompetitive allele-specific PCR (KASP) assays (Semagn et al., 2014), and analysis of the domestication history of long shelf life cultivars. These individuals were chosen

to sample the widest possible range of fruit firmness phenotypes and for testing hypotheses in different studies.

We used data for every individual and different subsets of individuals for specific analyses. The subsets included eight individuals used for a quantitative RT-PCR study (see below), 43 individuals used for an analysis of genetic variants among long-read DNA sequences (see below), and 101 individuals for a meta-analysis of fruit firmness phenotypes across experiments. Those included phenotypes of 85 individuals collected for the discovery population (see above), phenotypes collected from individual field grown plants of 87 individuals preserved in the UC Davis Strawberry Germplasm Collection at the Wolfskill Experiment Orchard (WEO), Winters, California in 2020-21 and 2021-22, and phenotypes of three clones/individual of 43 long-read DNA sequenced individuals grown in a UC Davis greenhouse in 2023. We phenotyped three fruit/harvest on six harvest dates in April and May 2021 and four harvest dates in April and May 2022 at WEO and three fruit/clone on five harvest dates in the greenhouse experiment. The data collected for the 87 individuals at WEO were used for KASP marker testing and validation.

3.3.4. Full-sib families. We selected four full-sib families from 178 crosses within the 2023-24 breeding pipeline that showed three genotypic classes using K-676, K-SPM, K-732, and K-253. We harvested all ripe fruits from one-year-old, seed-propagated plants grown at WEO ($n = 152$) across two harvest dates in May, with an average collection of two and four ripe fruits per individual, respectively. Fruit firmness was measured using the previously described protocol.

Estimated marginal means (EMMs) were calculated for one-year-old, seed-propagated plants through the *sommer* R package (Covarrubias-Pazaran, 2016) and `emmeans` R packages

$$y_{hg} = H_h + G_g + e_{hg} \tag{3.1}$$

where y_{hg} represents the observed phenotype for the g^{th} genotype in the h^{th} harvest. G_g is the effect of the g^{th} genotype and H_h is the random effect of the h^{th} harvest time.

3.3.5. GWAS population. Fruit firmness (kg-force) was measured on 460 wild and domesticated individuals in 2018 from replicated 6-plant plots grown in Ventura, CA under commercial field conditions. The data from the first harvest were first reported by Hardigan et al. (2021b). Fruit firmness was measured on six randomly selected berries from each plot from two harvests in March and April of 2018 as the maximum force with a QA Supplies FT2 handheld penetrometer equipped with a 3mm probe (QA Supplies, Norfolk, VA, USA) EMMs were calculated for this population using the *lme4*

$$y_{hgbf} = H_h + B_b + G_g + GB_{gb} + e_{hgbf} \tag{3.2}$$

where y_{hgbf} represents the f^{th} fruit phenotype for the g^{th} genotype in the b^{th} block during the h^{th} harvest. In this model, H_h denotes the fixed effect of the h^{th} harvest, G_g is the fixed effect of the g^{th} genotype. B_b is the random effect of the b^{th} block, GB_{gb} represents the random effect of the interaction between the g^{th} genotype and the b^{th} block (i.e., plot and experimental unit), and e_{hgbf} is the residual effect of $hgbf^{th}$ fruit (i.e., subsample).

3.3.6. Statistical analyses. Initial analysis of the raw phenotypic data was conducted using a LMM through the *sommer* R package (Covarrubias-Pazaran, 2016). EMMs were calculated for the UCD breeding collection via the *emmeans* R package (Lenth, 2021). The LMM across years was formulated as:

$$y_{lgyb} = L_l + G_g + Y_y + LY_{ly} + GY_{gy} + LYB_{lyb} + e_{lgyb} \tag{3.3}$$

where y_{ijkl} is the observed phenotype for the g^{th} genotype in the l^{th} location during the y^{th} year at the b^{th} complete block. L_l indicates the effect of the l^{th} location, G_g is the effect of

the g^{th} genotype. Y_y is the random effect of the y^{th} year, LY_{ly} is the random effect of the interaction between the l^{th} location and the y^{th} year, GY_{gy} is the random effect of the g^{th} genotype in the y^{th} year, LYB_{lyb} indicates the random effect of the interaction among the l^{th} location, y^{th} year, and b^{th} complete block, and e_{lgyb} is the residual effect.

Variance components were calculated for the random effects in LMM using the restricted maximum likelihood (REML) method. Broad-sense heritability on a clone-mean basis was estimated by $\hat{H}^2 = \hat{\sigma}^2_G/\hat{\sigma}^2_P$, where $\hat{\sigma}^2_G$ is the genetic variance among clonally propagated individuals (genotypes), $\hat{\sigma}^2_P = \hat{\sigma}^2_G + \hat{\sigma}^2_{GxY}/y + \hat{\sigma}^2_e/ry$ is the phenotypic variance on a clone-mean basis, $\hat{\sigma}^2_{GxY}$ is the genotype × year interaction variance, $\hat{\sigma}^2_e$ is the residual variance, y is the number of years, and $r = 11.23$ is the harmonic mean of the number of replications per genotype across years. Narrow-sense heritability was estimated by $\hat{h}^2 = \hat{\sigma}^2_A/\hat{\sigma}^2_P$, where $\hat{\sigma}^2_A$ is an estimate of the additive genetic variance from an RR-BLUP analysis (Endelman, 2011; Mathew et al., 2018) and $\hat{\sigma}^2_P$ is the phenotypic variance on a clone-mean basis.

3.3.7. Genome-wide association study. A single locus analysis for the across-year EMMs was performed using GEMMA v0.98.1 (Zhou and Stephens, 2012,1). This analysis incorporated 49,330 Axiom® array SNP markers mapped in the 'Royal Royce' reference genome (Hardigan et al., 2021b). The genomic relationship matrix (**K**), derived from SNP marker genotypes (Pincot et al., 2020), was used to adjust for genetic relationships among individuals (Zhou and Stephens, 2012,1). Percentage of genetic variance explained (GVE = $\hat{\sigma}^2_M/\hat{\sigma}^2_G$) and phenotypic variance explained (PVE = $\hat{\sigma}^2_M/\hat{\sigma}^2_P$) by the most significant SNPs were calculated, where $\hat{\sigma}^2_M$ is the bias-adjusted REML semi-variance estimate for the SNP marker, and $\hat{\sigma}^2_G$ and $\hat{\sigma}^2_P$ represent respective clone-mean-based genetic and phenotypic variance (Feldmann et al., 2021).

3.3.8. DNA sequence analyses. We analyzed the structure and function of the Fxa6A g103973 gene model in the Royal Royce genome using the Integrative Genomics Viewer (IGV) v2.16.2 (Thorvaldsdóttir et al., 2013) and the InterPro database (`https://www.ebi.ac.uk/interpro/`; Paysan-Lafosse et al., 2023), respectively. We used BLASTN 2.12.0+

(Zhang et al., 2000), implemented by the Genome Database for Rosaceae (GDR, `http://www.rosaceae.org`; Jung et al., 2019), to identify homologous sequences for the first and last four exons, including the large intron between them in the Fxa6Ag103973 gene model. The queried sequences were aligned against the reference genomes 'Camarosa' v1.0.a1 (Edger et al., 2019), 'Camarosa' v1.0.a2 (Liu et al., 2021) and *Fragaria vesca* v4.0.a1 (Edger et al., 2018). Microsynteny for the curated Fxa6Ag103973 gene model was visualized with JCVI v1.3.6 (Tang et al., 2008).

3.3.9. Phylogenomic analyses of polygalacturonase genes. Polygalacturonase (PG) genes were identified in the 'Royal Royce' reference genome using the gene ontology terms for PG activity (GO:0004650). PG genes exhibiting high sequence similarity to those within the firmness locus on chromosome 6A were detected via BLAST v2.15 (Altschul et al., 1990). Local synteny analysis was performed on homoeologous *PG1* genes. Alignments of amino acid sequences for *PG1* paralogs and homoeologs identified in this study were performed with Clustal Omega (`https://www.ebi.ac.uk/Tools/msa/clustalo/`; Madeira et al., 2022). the Interactive Tree Of Life v6 (`https://itol.embl.de`) was used to visualize the evolutionary relationships among these PGs.

3.3.10. Quantitative trait transcript analysis. Transcriptomic data from the discovery collection was aligned against the Royal Royce reference genome through STAR v2.7.10a (Dobin et al., 2013). Normalization, filtration, and conversion to $\log_2$-scaled counts per million ($\log_2$ CPM) were performed using `calcNormFactors()`, `filterByExpr()`, and `cpm()` functions from the edgeR package in R (Robinson et al., 2010), respectively. We further adjusted the data by including the **K** matrix in the `polygenic` function from the GenABEL package in R (Aulchenko et al., 2007). A Pearson's correlation analysis of gene expression profiles (60,685 transcripts) and EMMs of firmness values was conducted using the `cor.test` function in R. We applied a threshold of 0.005 to identify transcripts as quantitative trait transcripts (QTTs) when this stringent criterion was met.

3.3.11. Gene expression analyses. Expression profiles of the curated Fxa6Ag103973 gene model were analyzed in the firm-fruited cultivar 'Royal Royce' and the soft-fruited cultivar 'Mara des Bois' across the unripe, white, and ripe fruit development stages. Total RNA was isolated from pooled three ripe fruits after flash-freezing in liquid nitrogen using the Quick-RNA Miniprep Kit from Zymo Research. Sequencing on the Illumina NovoSeq platform yielded an average of 17.6 Gb per sample. We aligned the sequencing reads to the updated gene annotation of the 'Royal Royce' reference genome. Subsequently, transcript abundances were normalized to $\log_2$ CPM and visualized with the `pheatmap` package in R (Kolde and Kolde, 2015).

3.3.12. Real-time quantitative RT-PCR analyses. We quantified the expression levels of *PG1-6A1*, *PG1-6A2*, and *PG1-6D1* to evaluate their role in fruit firmness during ripening. Eight accessions were selected based on their firmness phenotypes and known genetic information concerning the *En/Spm* element upstream of the *PG1-6A1* gene. The accessions were grown at a UCD greenhouse facility in 2023 and represented by three biological replicates in a completely randomized design. We harvested three fruits for each biological replicate and pooled them for RNA isolation. After flash-freezing the samples in liquid nitrogen, the RNA extraction process was carried out using the Quick-RNA Miniprep Kit from Zymo Research. First-strand cDNA from total RNA extraction was synthesized through the RevertAid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific, Inc.). We established a DNA dilution of 1:125 for qPCR after analyzing the amplification efficiency for each primer set through a series of dilution assays (Figure 3.1). Details of the primers designed for each of the three PG genes with Primer3 v0.4.0 (Untergasser et al., 2012) are provided in Table 3.1. We prepared the DNA template mixtures to a final volume of 10 μL, which included 4 μL DNA template, 1 μL of 10 μM primer mix, and 5 μL PowerUp^{TM} SYBR^{TM} Green Master Mix (Applied Biosystem^{TM}). Amplifications were carried out on a QuanStudio^{TM} 3 Real-Time PCR System (Applied Biosystem^{TM}) under the following thermal profile: initial denaturation at 95 °C for 10 minutes, 40 amplification cycles at 95 °C for

15 seconds and 60 °C for 1 minute, and followed by a dissociation curve analysis ramping from 60 °C to 95 °C at 1.6°C/second. We considered three technical replicates per sample and included water as a no-template control in every run. The mRNA levels were normalized to the expression of *Fragaria* × *ananassa* DNA binding protein (*DBP*; Galli et al., 2015). To elucidate the relationship between each PG gene and fruit firmness at different developmental stages, we calculated Pearson's correlations in R.

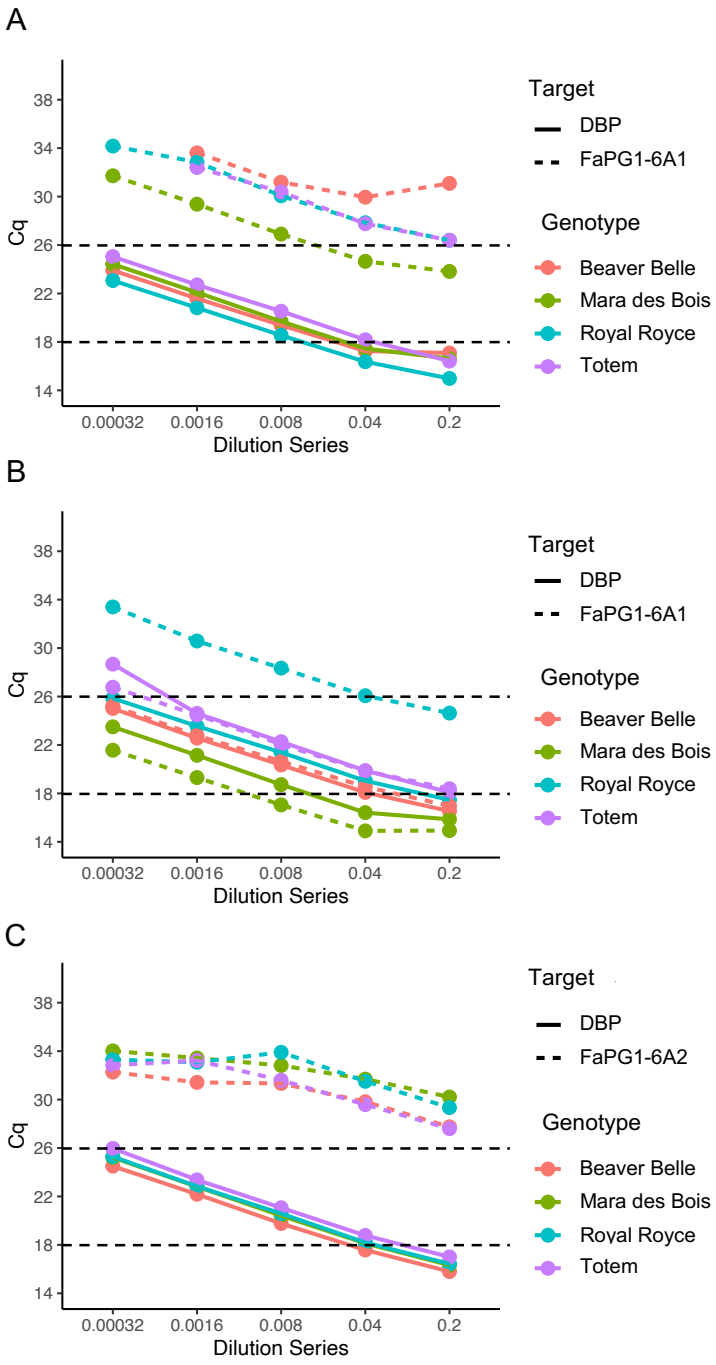

FIGURE 3.1. **Reaction efficiency for quantitative polymerase chain reaction (qPCR) analysis.** Dilution series in a fivefold decrease from 0.2 to 0.00032 for qPCR efficiency of DNA binding protein (DBP) and *PG1-6A1* at A) unripe and B) white fruit stages, and C) DBP and *PG1-6A2* at ripe fruit stage. Black dashed lines represent the threshold for optimal qPCR dilution (0.008).

TABLE 3.1. **List of primers for evaluating the gene expression of *PG1-6A1*, *PG1-6A2*, and *PG1-6D1*, as well as for genotyping the PG1 locus using KASP markers.**

Primer Name	**Sequence (5' ->3')**	**Description**
DBP qPCR-F	TTGGCAGCGGGACTTTACC	Forward qPCR primer targeting the DNA-binding protein housekeeping gene described by Galli et al. (2015).
DBP qPCR-R	CGGTTGTGTGACGCTGTCAT	Reverse qPCR primer targeting the DNA-binding protein housekeeping gene described by Galli et al. (2015).
PG1-6A1 qPCR-F	CCCATGGGGTCAGTGTAATAA	Forward qPCR primer targeting the PG1-6A1 gene (Fxa6Ag103973a)
PG1-6A1 qPCR-R	CGCAGTTGAAGTTGTCCCTA	Reverse qPCR primer targeting the PG1-6A1 gene (Fxa6Ag103973a)

Table 3.1 continued from previous page

Primer Name	Sequence (5' ->3')	Description
PG1-6A2 qPCR-F	CAAAGGCTTGGAAAGCAAAT	Forward qPCR primer targeting the PG1-6A2 gene (Fxa6Ag103973c)
PG1-6A2 qPCR-R	TTCTTCATCTGGTGGAACCTG	Reverse qPCR primer targeting the PG1-6A2 gene (Fxa6Ag103973c)
PG1-6D1 qPCR-F	GTTCAAAACCCAATCGTGCT	Forward qPCR primer targeting the PG1-6D1 gene (Fxa6Dg103429)
PG1-6D1 qPCR-R	GCAATCTTCACGCCTAATGC	Reverse qPCR primer targeting the PG1-6D1 gene (Fxa6Dg103429)
EnSpm1 KASP-FCOM	CTTTACGTATGAGTGCTAAGTCTTTGATTT	Forward KASP common primer targeting the EnSpm-1 transposable element upstream of the PG1-6A1 gene promoter

Table 3.1 continued from previous page

Primer Name	**Sequence (5' ->3')**	**Description**
EnSpm1 KASP-RHEX	gaaggtcggagtcaacggattAGCTTTGAGAAATTATTTTG TTACTCGC	Reverse KASP-specific A HEX primer targeting the deletion of EnSpm-1 transposable element upstream of the PG1-6A1 gene promoter

Table 3.1 continued from previous page

Primer Name	Sequence (5' ->3')	Description
EnSpm1 KASP-RFAM	gaaggtgaccaagttcatgctCTAGCTTTGAGAAATTATTTTGTTACTCGT	Reverse KASP specific ACGACACTACCA-GAAATTTCATTTTGGGC-GAC-GAAAAAAAATCGTCGC-CCNGTCGCCCGG TTTTTTTTGTTTCGT CGCCCAAGACTCATTTT CTGGTAGTG FAM primer targeting the EnSpm-1 transposable element upstream of the PG1-6A1 gene
PG1-GWAS KASP-FCOM	GTCTTGGCGTGGAATTCCAAACCTT	Forward KASP common primer targeting the GWAS peak SNP (AX-184210676)

Table 3.1 continued from previous page

Primer Name	Sequence (5' ->3')	Description
PG1-GWAS KASP-RHEX	gaaggtcggagtcaacggattGACAAGAATACCTTATTGGAGGAATG	Reverse KASP-specific C HEX primer targeting the GWAS peak SNP (AX-184210676)
PG1-GWAS KASP-RFAM	gaaggtgaccaagttcatgctTTTGACAAGAATACCTTATTGGAGGAATT	Reverse KASP-specific A FAM primer targeting the GWAS peak SNP (AX-184210676)
PG1-5UTR732 KASP-RCOM	CGAGAGCTAAGCAGTTTATAGTGTGATTT	Reverse KASP common primer targeting the 5' untranslated region (5'UTR) of the PG1-6A1 gene (physical position: 27,751,732)

Table 3.1 continued from previous page

Primer Name	Sequence (5' ->3')	Description
PG1-5UTR732 KASP-FHEX	gaaggtcggagtcaacggattGTTGCAGGATGTCTTAGTGCCGT	Forward KASP specific T HEX primer targeting the 5' untranslated region (5'UTR) of the PG1-6A1 gene (physical position: 27,751,732)
PG1-5UTR732 KASP-FFAM	gaaggtgaccaagttcatgctTGCAGGATGTCTTAGTGCCGG	Forward KASP specific G FAM primer targeting the 5' untranslated region (5'UTR) of the PG1-6A1 gene (physical position: 27,751,732)
PG1-6A1-eQTL KASP-FCOM	GACGGCAGATATTGCCACAAATGCAA	Forward KASP common primer targeting the eQTL SNP peak (AX-184242253)

Table 3.1 continued from previous page

Primer Name	Sequence (5' ->3')	Description
PG1-6A1-eQTL KASP-RHEX	gaaggtcggagtcaacggattGTAACTCACACATCTTCTAGTGCC	Reverse KASP-specific G HEX primer targeting the eQTL SNP peak (AX-184242253)
PG1-6A1-eQTL KASP-RFAM	gaaggtgaccaagttcatgctTGTAACTCACACATCTTCTAGTGCT	Reverse KASP specific A FAM primer targeting the eQTL SNP peak (AX-184242253)

3.3.13. Long-read DNA sequencing and genetic variant analyses. We used highly accurate long sequencing reads (HiFi reads) from a diverse assemblage of 43 strawberry accessions, including wilds, heirlooms, advanced selections, and modern cultivars, to identify genomic variants in the *PG1* locus. HiFi reads were aligned to the 'Royal Royce' reference genome by using the `pbmm2` tool (Li, 2018). We delimited the analysis of structural variations within the *PG1* locus via `samtools view` from Samtools v1.13 (Danecek et al., 2021). Subsequently, the identification and characterization of structural variation signatures were executed through the `pbsv discover` and `pbsv call` functions, available in pbsv v2.8.0 (https://github.com/PacificBiosciences/pbsv).

The insertion-deletion (INDELs) found upstream of the *PG1-6A1* gene's promoter were quantitatively associated with firmness metrics. The fruit firmness of the diverse strawberry panel grown at a UCD greenhouse facility was measured in 2023. We computed EMMs from five biological replicates and three ripe fruits per accession using a completely random experiment design across five harvest times. Pearson correlation coefficients and the corresponding p-values were calculated using R statistical software. In addition, we examined the identity of the INDEL located at -3,926 bp from the ATG codon of the *PG1-6A1* gene using the CENSOR web server (https://www.girinst.org/; Kohany et al., 2006).

We identified SNPs within the *PG1-6A1* gene using DeepVariant v1.4 (Poplin et al., 2018). SNPs were filtered using BCFTools v1.19 (Danecek et al., 2021) with the `view` command, applying a minor allele frequency threshold of ≤ 0.05 and a quality score threshold of ≥ 30. SNPs identified in the *PG1-6A1* gene were classified using SnpEff v5.2c (Cingolani et al., 2012b). We used SnpSift v5.2c (Cingolani et al., 2012a) to identify SNPs significantly associated with fruit firmness by coding soft-fruited accessions as 1 (wildtype) and firm-fruited accessions as 2 (mutant). Additionally, cis-eQTLs were identified within the discovery population using the MatrixEQTL package in R (Shabalin, 2012), following the approach described by Fan et al. (2022). For further analysis, we focused on the most significant cis-eQTL associated with the *PG1-6A1* gene.

3.3.14. DNA marker development. Four Kompetitive allele-specific PCR (KASP) assays were designed to target the *PG1* locus in a diverse strawberry panel ($n = 92$) and one-year-old, seed-propagated plants grown in 2023-24 ($n = 152$). The design of KASP markers was undertaken using the services provided by LGC Biosearch™ Technologies. We aimed to target the GWAS peak SNP, the most significant eQTL peak SNP, the *En/Spm-1* transposable element, and an intragenic SNP variant in the *PG1-6A1* coding sequence (Table 3.1).

We established a statistical relationship between firmness EMMs and KASP markers by conducting Pearson's correlation in the R platform. The additive and dominance effects were estimated by $\hat{a} = \hat{\mu}_{PG1\text{-}6A1^-/PG1\text{-}6A1^-} - \hat{\mu}_{PG1\text{-}6A1^+/PG1\text{-}6A1^+}$ and $\hat{d} = \hat{\mu}_{PG1\text{-}6A1^+/PG1\text{-}6A1^-} - (\hat{\mu}_{PG1\text{-}6A1^-/PG1\text{-}6A1^-} + \hat{\mu}_{PG1\text{-}6A1^+/PG1\text{-}6A1^+})/2$, where $\hat{\mu}_{PG1\text{-}6A1^-/PG1\text{-}6A1^-}$, $\hat{\mu}_{PG1\text{-}6A1^+/PG1\text{-}6A1^-}$, $\hat{\mu}_{PG1\text{-}6A1^+/PG1\text{-}6A1^+}$ are the EMMs for individuals with *PG1-6A1*$^-$/*PG1-6A1*$^-$, *PG1-6A1*$^+$/*PG1-6A1*$^-$, and *PG1-6A1*$^+$/*PG1-6A1*$^+$ genotypes for each KASP marker. *PG1-6A1*$^-$ is the unfavorable allele (wildtype), while *PG1-6A1*$^+$ is the favorable allele (mutant). The degree of dominance of each KASP marker was estimated by $|\hat{d}/\hat{a}|$ (Falconer and Mackay, 1996; Walsh, 2001). We calculated the accuracy of KASP markers based on SNPs and the *En/Spm* element using data from the 50K array (n = 90) and HiFi data (n = 37), respectively.

3.4. Results and Discussion

3.4.1. A genome-wide association study confirmed the segregation of a large-effect QTL for fruit firmness on chromosome 6A. To identify genetic variants associated with fruit firmness, we analyzed a population of 85 phenotyped, 50K Axiom® SNP array genotyped, and ripe-fruit transcriptome profiled hybrid individuals developed at the University of California, Davis. Their phenotypic means ($\overline{y}$) were estimated from multiple biological replicates (clones of hybrid individuals) and multiple harvests of ripe fruit over two growing seasons (24 phenotypic observations/individual), ranged from extremely soft (0.13 kg-force) to extremely firm (0.45 kg-force), and were approximately normally distributed. This population was dominated by firm-fruited individuals: 74 had phenotypic means in the

0.30 to 0.49 kg-force range (the firmest individual was 17C140P012). This group included the cultivar 'Royal Royce' ($\overline{y} = 0.48$ kg-force), 11 other UC cultivars, and 62 other UC hybrid individuals. The soft-fruited group included the UC cultivar 'Chandler' ($\overline{y} = 0.21$) and 10 other UC hybrid individuals. Approximately half of the phenotypic variation observed for fruit firmness in this discovery population was genetic. The REML estimates of narrow- and broad-sense heritability were $h^2 = 0.46$ and $H^2 = 0.50$, respectively; hence, 92% of the genetic variance for fruit firmness was estimated to be additive in the discovery population. Using genetically validated physical positions of 49,330 array-genotyped SNPs for GWAS in the discovery population, four SNPs on chromosome 6A (AX-184953741: Mb 27,253,734, AX-184023221: Mb 27,593,175, AX-184726882: Mb 27,671,092, and AX-184210676: Mb 27,676,285) were found to be strongly associated with fruit firmness variation (Figure 3.2A-B). We noticed that the fruit firmness QTL identified on chromosome 6A was discovered in previous genome-wide association and genetic mapping studies (Cockerton et al., 2021; Feldmann et al., 2024b; Hardigan et al., 2021b; Lee et al., 2021; Munoz et al., 2024; Prohaska et al., 2024). We did not observe statistically significant SNPs elsewhere in the genome, nor did Hardigan et al. (2021b) using a larger and more genetically diverse population ($n = 466$). We substantiated this by repeating the GWAS analysis of that population using the original phenotypic data and previously unpublished phenotypic data (10.9 fruit/individual $\times$ 460 individuals = 5,014 phenotypic observations). The broad-sense heritability on a clone-mean basis was 0.81 for fruit firmness in that study. The two most significant SNPs in that analysis were AX-184107692 (-$\log_{10}$ p-value = 7.1; Mb 27,812,192) and AX-184242253 (-$\log_{10}$ p-value = 6.4; Mb 27,888,596; Figure 3.3), both of which were in strong LD with SNPs identified in the discovery population ($n = 85$).

The four statistically significant SNPs identified by GWAS in the discovery population ($n = 85$) were in complete linkage disequilibrium (LD) with one another ($\hat{r}^2 = 1.0$), spanned a 0.6 Mb haploblock on chromosome 6A in the FaRR V1 reference genome (Mb 27.3-27.9 in FaRR V1), and exceeded a 1% false discovery rate (FDR)-corrected statistical significance

threshold (Figure 3.2). They were linked with array-genotyped SNPs previously shown to be associated with a fruit firmness QTL and three tandemly duplicated polygalacturonase-encoding genes (Mb 7.046-7.064) on chromosome 6A (Hardigan et al., 2021b). The latter were annotated as FxaC_6-1g13880, FxaC_6-1g13900, and FxaC_6-1g13910 in the 'Camarosa' reference genome (FaCA V1; Edger et al., 2019).

3.4.2. Three tandemly duplicated polygalacturonase genes are in strong linkage disequilibrium with the fruit firmness QTL on chromosome 6A. Transcriptomic and BLAST analyses confirmed that the three polygalacturonase-encoding genes in LD with the QTL on chromosome 6A are *FaPG1* (*PG1*) paralogs (Figure 3.4A; López-Casado et al., 2023; Paniagua et al., 2020; Quesada et al., 2009); however, we discovered errors in their annotations of the 'Camarosa' genome (FaCA V1; Edger et al., 2019), a reannotation of the 'Camarosa' genome (FaCA V1R; Liu et al., 2021), and annotations of the 'Royal Royce' genome (FaRR V1; https://phytozome-next.jgi.doe.gov/info/FxananassaRoyalRoyce_v1_0). Specifically and briefly, the first *PG1* paralog was merged with an *Enhancer/Suppressor-mutator* (*En/Spm*) transposon and other non-genic DNA sequences in the original annotation of the 'Camarosa' genome (Edger et al., 2019), the first *PG1* paralog was missing in the reannotation of the 'Camarosa' genome by Liu et al. (2021), and the second and third *PG1* paralogs were merged with a gene of unknown function (Fxa6Ag103973) in the original annotation of the 'Royal Royce' genome (https://phytozome-next.jgi.doe.gov/info/FxananassaRoyalRoyce_v1_0). These annotation errors were manually corrected by mapping transcripts isolated from unripe and ripe fruit to the 'Royal Royce' reference genome (FaRR V1). We identified three tandemly duplicated *PG1* paralogs, which appear to be arranged as shown in the uppermost chromosome displayed in Figure 3.4A. Fxa6Ag103973 in the original FaRR V1 annotation was split into two *PG1* paralogs (Fxa6Ag103973a and Fxa6Ag103973c) and a gene of unknown function (Fxa6Ag103973b) in the corrected annotation of FaRR V1.

The imperfections inherent in genome annotations (Salzberg, 2019), and the need to distinguish paralogs and homoeologs (Koonin, 2005) underscore the complications that commonly arise when cross-referencing genes, loci, and alleles in genetic studies of octoploid strawberry, a species where community-wide guidelines have not yet been adopted for naming chromosomes, linkage groups, homologous and homoeologous genes, haplotypes, and alleles (Figure 3.4A-Figure 3.5). To address this in the present study, we created and adopted gene names of chromosome numbers (1-7) and genome letters (A-D) as suffixes followed by consecutive numbers $(1, 2, ..., n)$ to identify paralogs where necessary. The chromosome identifiers (1A, 1B, 1C, 1D, ..., 7A, 7B, 7C, 7D) follow the nomenclature proposed by Hardigan et al. (2020). Chromosome numbers (1, 2, ..., 7) in the ancestral genomes (A, B, C, and D) match those originally assigned in *F. vesca*, the diploid ancestor of the A-genome (Edger et al., 2019,1; Session and Rokhsar, 2023; Shulaev et al., 2011).

Using these naming guidelines and the corrected FaRR V1 gene annotations, the gene identifiers for the three tandemly duplicated *PG1* paralogs in LD with the fruit firmness QTL on chromosome 6A are *PG1-6A3* (Fxa6Ag103971), *PG1-6A1* (Fxa6Ag10397a), and *PG1-6A2* (Fxa6Ag10397c), as shown in Figure 3.4A. Their arrangement and annotations match those found in the 'Hawaii 4' reference genome for *F. vesca* (FvH4 V1; Edger et al., 2018), as shown in the lowermost chromosome displayed in Figure 3.4A. The sequences of these genes in FaRR V1 are 99.5% identical to their homologs in the FvH4 V1 diploid and 100.0% identical to their homologs in the FaCA V1 octoploid reference genomes. *PG1-6A1* is 73.9% identical to *PG1-6A2* (62.1% similarity at the amino acid level) and 61.5% identical to *PG1-6A3* (49.9% similarity at the amino acid level; $E \leq 1.0 \times 10^{-43}$). The *PG1-6A1* sequence in the UC cultivar 'Royal Royce' is 100% identical to the *FaPG1* sequence previously isolated from the UC cultivar 'Chandler' (`https://www.ncbi.nlm.nih.gov/nuccore/AF380299`; Paniagua et al., 2020; Quesada et al., 2009).

3.4.3. A single paralog of the fruit softening gene *POLYGALACTURONASE1* underlies the fruit firmness QTL on chromosome 6A. We hypothesized that one of

the three tandemly duplicated *PG1* paralogs was the causal gene underlying the fruit firmness QTL on chromosome 6A (Figure 3.4). To test this, quantitative trait transcript (QTT) analysis was applied to short-read sequences of ripe-fruit RNAs isolated from the discovery population (Figure 3.2B; https://www.ncbi.nlm.nih.gov/bioproject/PRJNA787565). Transcripts were observed in ripe fruit for 1,081 of the 1,761 annotated genes found in the 2.4 Mb genomic segment flanking the QTL. Three of these genes (*PG1-6A1* = Fxa6Ag103973a, Fxa6Ag104099, and Fxa6Ag104340) were found to be significantly associated with the fruit firmness QTL ($p \leq 0.005$; Figure 3.2B). *PG1-6A1* was in complete LD with the four peak SNPs identified by GWAS in the discovery population (Figure 3.2D). Moreover, the expression of *PG1-6A1* was negatively genetically correlated with the firmness of ripe fruit ($r = -0.47$; $p = 1 \times 10^{-6}$) in the discovery population (Figure 3.2). The function of Fxa6Ag104099 remains unknown, while the other QTT-significant gene (Fxa6Ag104340) encodes an S-adenosyl-L-Met decarboxylase gene (*FaSAMDC*) known to play an important role in the regulation of strawberry fruit ripening (Guo et al., 2018). *FaSAMDC* is located approximately one Mb downstream of *PG1-6A2*; hence, we concluded that the fruit firmness variation associated with the QTL on chromosome 6A was likely caused by the segregation of mutant and wildtype *PG1-6A1* alleles (Figure 3.2). Throughout this chapter, alleles that increase fruit firmness are described as favorable and identified by a plus sign superscript (*PG1-6A1*$^+$), whereas alleles that decrease fruit firmness are described as unfavorable and identified by a minus sign superscript (*PG1-6A1*$^-$). These designations are market class dependant, e.g., mutant alleles that knockdown or knockout the expression of *PG1-6A1* are favorable when breeding firm-fruited cultivars and unfavorable when breeding soft-fruited cultivars.

BLAST analyses suggest that approximately 227 polygalacturonase-encoding genes have survived polyploid evolution in the octoploid. Apart from the three *PG1* paralogs on chromosome 6A, 15 *PG1* homoeologs were identified on chromosomes 6B, C, and D (Figure 3.5). Of these, only *PG1-6A1* was strongly expressed in ripe fruit (Figure 3.4). *PG1-6A2*,

PG1-6A3, and Fxa6Dg103429 (*PG1-6D1*), a *PG1* homoeolog on chromosome 6D, were very weakly expressed in unripe-green fruit and unexpressed in ripe fruit (Figure 3.4B). The other 14 *PG1* B-, C-, and D-genome homoeologs were unexpressed in fruit (Figure 3.6). *POLYGALACTURONASE2* (*PG2*) is another polygalacturonase gene previously studied that has been associated with fruit firmness in strawberry (Paniagua et al., 2020). Paniagua et al. (2020) showed that fruit firmness could be increased by transgenic anti-sense silencing of *PG2* in combination with *PG1-6A1*, and that *PG2* was more weakly expressed than *PG1-6A1* in ripe fruit, but expressed nonetheless. Consistent with our findings, Sánchez-Sevilla et al. (2017) did not observe the expression of *PG2* in ripe fruit of the firm-fruited cultivar 'Camarosa', whereas *PG1-6A1* was strongly expressed.

The transcriptomes of the soft-fruited cultivar 'Mara des Bois' ($\overline{y}$ = 0.10 kg-force) and firm-fruited cultivar 'Royal Royce' ($\overline{y}$ = 0.35 kg-force) were compared by RNA-seq across different stages of fruit development, where $\overline{y}$ is the estimated marginal mean (EMM) for fruit firmness (Figure3.4B). As predicted by the QTT analysis, *PG1-6A1* was more strongly expressed in ripe fruit of the soft-fruited wildtype ('Mara des Bois'; 2,033.1 CPM) than the firm-fruited mutant ('Royal Royce'; 163.2 CPM), a 12.5-fold difference (Figure 3.4B). The expression of *PG1-6A1* increased from 57.0 CPM in unripe to 2,033.1 CPM in ripe fruit of the wildtype, a 35.7-fold increase. By contrast, a 544-fold increase was observed in the mutant from unripe (0.30 CPM) to ripe fruit (163.2 CPM). The weak *PG1-6A1* expression in ripe fruit of the firm-fruited cultivar 'Royal Royce' was consistent with the observation in reverse genetic studies that *PG1* knockdown and knockout mutations result in increased fruit firmness (López-Casado et al., 2023; Paniagua et al., 2020; Quesada et al., 2009). The fruit firmness increases reported by Paniagua et al. (2020) from anti-sense silencing of *PG1* ranged from 0 to 140% (approximated by us from phenotypic means of *PG1* transgenic plants displayed in their bar chart). Similarly, López-Casado et al. (2023) reported increases in fruit firmness in the 37 to 70% range among transgenic plants carrying knockdown edited

PG1 genes. These forward and reverse genetic studies clearly substantiate the importance of the *PG1-6A1* gene.

To further substantiate the association between *PG1-6A1* expression and the QTL, we developed paralog-specific quantitative reverse transcription polymerase chain reaction (qRT-PCR) assays for *PG1-6A1*, *PG1-6A2*, and *PG1-6D1* and screened RNAs isolated from unripe, green, and ripe fruit of three firm-fruited and five soft-fruited cultivars (Figure 3.7). *PG1-6A1*, *PG1-6A2*, and *PG1-6D1* are the only *PG1* paralogs that were observed to be weakly to strongly expressed in developing fruit. The expression of *PG1-6A1* was weak in unripe-green fruit of every cultivar, progressively increased in the five soft-fruited cultivars as they ripened, was 13.0-fold greater in ripe fruit of soft- than firm-fruited cultivars, and was negatively correlated with fruit firmness in ripe fruit ($\hat{r} = -0.82$; $p = 1.4 \times 10^{-2}$; Figure 3.7 A). These analyses substantiated that *PG1-6A2* and *PG1-6D1* were very weakly expressed in developing fruit and that their expression was uncorrelated with fruit firmness variation (Figure 3.7 B-C).

3.4.4. An *En/Spm* transposon insertion-deletion and single nucleotide polymorphisms associated with differentially expressed *PG1-6A1* alleles. To search for prospective causal mutations and identify predictive genetic variants within and proximal to the *PG1-6A1* gene, we developed and analyzed high fidelity (long-read) DNA sequences for 43 individuals, including 37 *F.* × *ananassa* cultivars, two *F. chiloensis* ecotypes, and four *F. virginiana* ecotypes differing in fruit firmness (phenotypic means for these individuals ranged from 0.06 to 0.35 kg-force in a greenhouse study and 0.07 to 0.41 kg-force in field studies). DNA sequences for those individuals were aligned to the FaRR V1 reference genome to identify structural variants and SNPs associated with the *PG1-6A1* locus (Table 3.2; Figure 3.8 and 3.9). This analysis of the 0.6 Mb segment flanking the QTL (Mb 27.3-27.9) identified 7,221 SNPs (one SNP every 83.1 bp) among the 43 individuals. The Mb 27.5-27.8 segment harboring the QTL and tandemly duplicated polygalacturonase genes (*PG1-6A1*, *PG1-6A2*,

and *PG1-6A3*) was homozygous in 'Royal Royce' and the other nine firm-fruited UC cultivars we sequenced. Analyses of pedigree records suggested that the favorable *PG1-6A1* alleles found in these cultivars are identical-by-descent (Pincot et al., 2021).

TABLE 3.2. **Statistics[a] for an Enhancer/Suppressor-mutator (*En/Spm*) insertion-deletion (INDEL) and single nucleotide polymorphisms (SNPs) associated with a polygalacturonase gene (*PG1-6A1*) and fruit firmness QTL on chromosome 6A in a population of 43 soft- to firm-fruited individuals.**

		Position						$\hat{a}$		$\hat{d}$		
Marker (M)[b]	Variant[c]	(bp)	FAF[d]	$\hat{r}_{\overline{y},M}$	$\hat{r}_{RE,M}$	PVE	GVE	(kg-force)	$Pr > \mathrm{F}$	(kg-force)	$Pr > \mathrm{F}$	$\lvert\hat{d}/\hat{a}\rvert$
AX-184953741	G/T	27,253,734	0.49	0.77	-0.86	0.68	0.76	0.08	< 0.0001	0.04	0.07	0.48
AX-184210676	C/A	27,676,285	0.55	0.68	-0.91	0.22	0.26	0.07	< 0.0001	0.06	0.01	0.82
En/Spm INDEL	+/-	27,743,085	0.55	0.76	-0.79	0.49	0.55	0.08	< 0.0001	0.05	0.01	0.59
5'-UTR	G/T	27,751,732	0.45	0.81	-0.91	0.53	0.61	0.08	< 0.0001	0.03	0.24	0.33
AX-184242253	G/A	27,888,596	0.49	0.78	-0.75	0.54	0.61	0.08	< 0.0001	0.04	0.05	0.50

[a] [a]Correlation between the phenotypic mean for fruit firmness ($\overline{y}$) and marker locus genotypes ($\hat{r}_{\overline{y},M}$), correlation between the relative expression (RE) of *PG1-6A1* in ripe fruit and marker locus genotypes ($\hat{r}_{RE,M}$), estimates of the additive ($\hat{a}$) and dominance ($\hat{d}$) effects of the marker locus on fruit firmness, degree-of-dominance ($|\hat{d}/\hat{a}|$), fraction of the phenotypic variance explained by the marker locus (PVE), and fraction of the genetic variance explained by the marker locus (GVE). Correlation coefficient estimates were significantly greater than zero for every genetic variant ($p \leq 0.0001$ for $\hat{r}_{\overline{y},M}$ and $0.020 \leq p \leq 0.008$ for $\hat{r}_{RE,M}$). Additive and dominance effects were estimated by linear contrasts among genotypic means ($\overline{y}_{+/+}$, $\overline{y}_{+/-}$, and $\overline{y}_{-/-}$). Significance levels ($Pr > \mathrm{F}$) are shown in the columns to the right of additive and dominance effect estimates for tests of the null hypothesis that the linear contrast was not significantly different from zero, where $\hat{a} = (\overline{y}_{+/+} - \overline{y}_{-/-})/2$ and $\hat{d} = (\overline{y}_{+/+} + \overline{y}_{-/-})/2 - \overline{y}_{+/-}$.

[b] [b]Genetic variants were genotyped by genotyping-by-sequencing (GBS). SNP and INDEL genotypes were called by aligning long-read DNA sequences for 43 individuals to the 'Royal Royce' reference genome (FaRR1). AX-184953741 is one of four 50K array SNP markers identified by GWAS that was most strongly associated with the QTL on chromosome 6A. AX-184210676 is the 50K array SNP marker identified by GWAS that was most tightly linked to *PG1-6A1* among the four that were strongly associated with the QTL on chromosome 6A and in complete LD within the discovery population. The 4,948-bp *En/Spm* INDEL is located 3,926 bp upstream of *PG1-6A1*. The G/T SNP in the 5'-UTR is one of three that were in complete LD in the 5'-UTR of *PG1-6A1*. Statistics were identical for the three highly predictive 5'-UTR SNPs identified by sequence and QTL analyses among the 43 DNA sequenced individuals and are only shown for the G/T SNP (bp 27,751,732). The other two 5'-UTR SNPS were A/G (bp 27,751,041) and T/C (bp 27,751,106). AX-184242253 is the 50K array SNP in LD with *PG1-6A1* that was identified by expression-QTL analysis.

[c] [c]The first nucleotide shown for each SNP is the favorable allele (the SNP associated with the *PG1-6A1* allele predicted to increase fruit firmness). The 4,498-bp insertion (+) is the favorable allele for the INDEL.

[d] [d]FAF is the frequency of the SNP or *En/Spm* INDEL allele associated with the favorable *PG1-6A1* allele.

We did not identify a definitive causal mutation in *PG1-6A1*; however, several genetic variants were found to be highly predictive of *PG1-6A1* expression differences and fruit firmness variation (Table 3.2; Figure 3.8). SIFT (Sorting Intolerant From Tolerant) statistics estimated with SNPeff (Cingolani et al., 2012b) and SNPsift (Cingolani et al., 2012a) were used to search for SNPs predicted to affect the function of the polygalacturonase encoded by *PG1-6A1*, especially non-synonymous SNPs that could cause amino acid substitutions predicted to be deleterious to protein function. SIFT analyses were done by comparing *PG1-6A1* alleles between soft-fruited ($\overline{y} < 0.30$ kg-force) and firm-fruited ($\overline{y} \geq 0.30$ kg-force) individuals.

Thirty-seven SNPs were identified among *PG1-6A1* alleles, 19 in the 5'-UTR, 12 in exons, four in introns, and two in the 3'-UTR. Of the 12 SNPs found in exons, four caused synonymous and eight caused non-synonymous amino acid changes, none of which were predicted to be deleterious.

We used SIFT statistics and single-marker QTL analyses to identify genetic variants associated with fruit firmness phenotypes among the 43 long-read sequenced individuals (Table 3.2; Figure 3.8). Three genotyping-by-sequencing (GBS) called SNPs in complete LD with one another in the 5'-UTR (bp 27,751,041, 27,751,106, and 27,751,732; MAF = 0.45) were found to be positively correlated with fruit firmness ($\hat{r} = 0.81$; $p = 1.73 \times 10^{-8}$) and negatively correlated with the expression of *PG1-6A1* in ripe fruit ($\hat{r} = -0.91$; $p = 0.002$; data are only shown for one of the three in Table 3.2). A single exonic SNP (bp 27,752,865; MAF = 0.44) was similarly predictive of fruit firmness variation ($\hat{r} = 0.76$; $p = 8.63 \times 10^{-7}$). This SNP caused a conservative, non-synonymous amino acid change (M220T) and was predicted by SNPsift to have a moderate effect on protein function, if any.

To complement these analyses, we used expression-QTL (eQTL) analysis to search the genome for associations between 50K array genotyped SNPs and transcript abundance (count per million) in ripe fruit of the 85 discovery population individuals. That analysis identified 10,780 eQTL associations in 5,159 differentially expressed genes in ripe fruit transcriptomes. We identified a SNP (AX-184242253; $p = 4.69 \times 10^{-10}$) slightly downstream of *PG1-6A1* that was significant in the GWAS analysis, tightly linked with *PG1-6A1* and the four GWAS peak SNPs, positively correlated with variation in fruit firmness, and negatively correlated with *PG1-6A1* expression differences (Table 3.2; Figure 3.2).

The analyses of GBS-called variants uncovered a 4,948-bp insertion-deletion 3,926 bp upstream of *PG1-6A1* that was positively associated with fruit firmness variation ($\hat{r} = 0.76$; $p < 0.0001$) and negatively correlated with the expression of *PG1-6A1* in ripe fruit ($\hat{r} = -0.79$; p-value = 0.02) (Figure 3.7-3.8C-3.9; Table 3.2). The *En/Spm* insertion was only

found in firm-fruited individuals, was homozygous in 'Royal Royce' and the other nine modern, firm-fruited UC cultivars that were long-read sequenced, and was the only structural variant proximal to *PG1-6A1* that was strongly associated with variation in fruit firmness and expression of *PG1-6A1* in wildtype and mutant individuals (Table 3.8; Figure 3.7A). The 4,948-bp insertion was discovered to be a CACTA family class II *En/Spm* transposable element (TE) with sequence similarity to *F. vesca*, *F.* × *ananassa*, and other *En/Spm* TEs in plants (Bennetzen, 2000; Bennetzen and Wang, 2014; McClintock, 1950,5). This TE carries the characteristic 12-bp terminal inverted repeats (5′-CACTACCAGAAA-3′) and proximal subterminal direct repeats found in CACTA family class II TEs (Bennetzen, 2000; Feschotte et al., 2002; Wessler, 1988). *En/Spm* elements are important mediators of phenotypic diversity in plants (Bennetzen, 2000; Bennetzen and Wang, 2014; Feschotte et al., 2002; Hirsch and Springer, 2017; Wessler, 1988), and often create novel phenotypes by disrupting the function or modifying the expression of genes through insertional mutagenesis in coding or promoter sequences. Although numerous TE insertional mutations have been described in plants, and their effects are often predictable, the mutant phenotype observed in this study was not caused by an insertional mutation in *PG1-6A1 per se* (Feschotte, 2008; Feschotte et al., 2002; Hirsch and Springer, 2017). Castillejo et al. (2020) discovered an *En/Spm*-2 TE insertion in the promoter of *MYB10-2*, an R2R2 MYB transcription factor that regulates anthocyanin biosynthesis in strawberry. They showed that the mutation (TE insertion) enhanced the expression of *MYB10-2* and increased the accumulation of anthocyanins in ripe fruit. This appears to be the only TE insertional mutant reported to date in strawberry. Apart from insertional mutagenesis, TEs create novel phenotypes by altering the regulation of nearby genes (reviewed by

3.4.5. Wildtype *PG1-6A1* alleles are incompletely dominant. Table 3.2 displays the additive and dominance effects and other statistics for the five GBS-called genetic variants that were most highly predictive of *PG1-6A1* expression differences and fruit firmness variation in diverse germplasm. These include array-genotyped SNPs identified by GWAS

(AX-184953741 and AX-184210676) or expression-QTL analysis (AX-184242253), one of three SNPs in the 5'-UTR of *PG1-6A1* identified by SIFT, sequence, and QTL analyses, and the *En/Spm* INDEL identified by sequence and QTL analyses (Table 3.2; Figure 3.8). The three SNPs in the 5'-UTR of *PG1-6A1* were in complete LD among the 43 long-read sequenced individuals; hence, the statistics shown for the G/T SNP (bp 27,751,732) were identical for the other two SNPs among those 43 individuals (Table 3.2).

The favorable allele frequencies, correlations between marker genotypes and phenotypic means ($\hat{r}_{\overline{y},M}$), correlations between marker genotypes and relative expression levels of *PG1-6A1* ($\hat{r}_{\overline{y},M}$), and additive ($\hat{a}$) and dominance ($\hat{d}$) effect estimates were similar across genetic variants (Table 3.2; Figure 3.8). The additive effects of these *PG1-6A1*-associated genetic variants were highly significant ($p \leq 0.0001$) and greater than their dominance effects, which were significant for three of the five genetic variants ($p \leq 0.05$). The wildtype *PG1-6A1*$^{-}$ allele was incompletely dominant to nearly dominant. The ($|\hat{d}/\hat{a}|$ estimates ranged from 0.33 to 0.82 for the five genetic variants (Table 3.2; Figure 3.8). Four of the genetic variants explained 49-68% of the phenotypic variance and 55-76% of the genetic variance for fruit firmness among the 43 long-read sequenced individuals. The estimate of the genetic variance explained by the AX-184210676 SNP (26%) in the population of 43 long-read-sequenced individuals was markedly lower than estimates for the other genetic variants, even though the *PG1-6A1* genotypes predicted by that SNP were strongly correlated with differences in *PG1-6A1* expression and phenotypic variation for firm firmness in the discovery population (Table 3.2).

3.4.6. Quantitative trait transcript and co-expression analyses identified several differentially expressed genes known or predicted to affect fruit development and ripening. We knew from GWAS analyses that approximately 24-45% of the heritability for fruit firmness was likely missing and undoubtedly caused by the segregation of loci other than *PG1-6A1* in the discovery population (Table 3.2). Although we only observed one statistically significant genotype-to-phenotype QTL in the genome-wide association study,

multiple QTL have been identified in other forward genetic studies (Cockerton et al., 2021; Feldmann et al., 2024b; Lee et al., 2021; Munoz et al., 2024; Prohaska et al., 2024), and several statistically significant transcript-to-phenotype QTL were observed in the QTT analysis (Figure 3.2). To shed light on the biological functions of the latter, we conducted a co-expression analysis of the 39 genes identified by QTT analysis (Figure 3.10; Table 3.3). This revealed three co-expression modules: one comprised of genes that were moderately to strongly expressed in ripe fruit and negatively correlated with fruit firmness (module 1, upper clade); one comprised of genes with low expression in ripe fruit that were negatively correlated with fruit firmness (module 2, middle clade); and one comprised of genes with that were moderately to strongly expressed in ripe fruit and positively correlated with fruit firmness (module 3, lower clade), as shown in Figure 3.10. While some of these could be false positives, future studies are bound to identify and validate additional QTL with predictable effects across environments and genetic backgrounds.

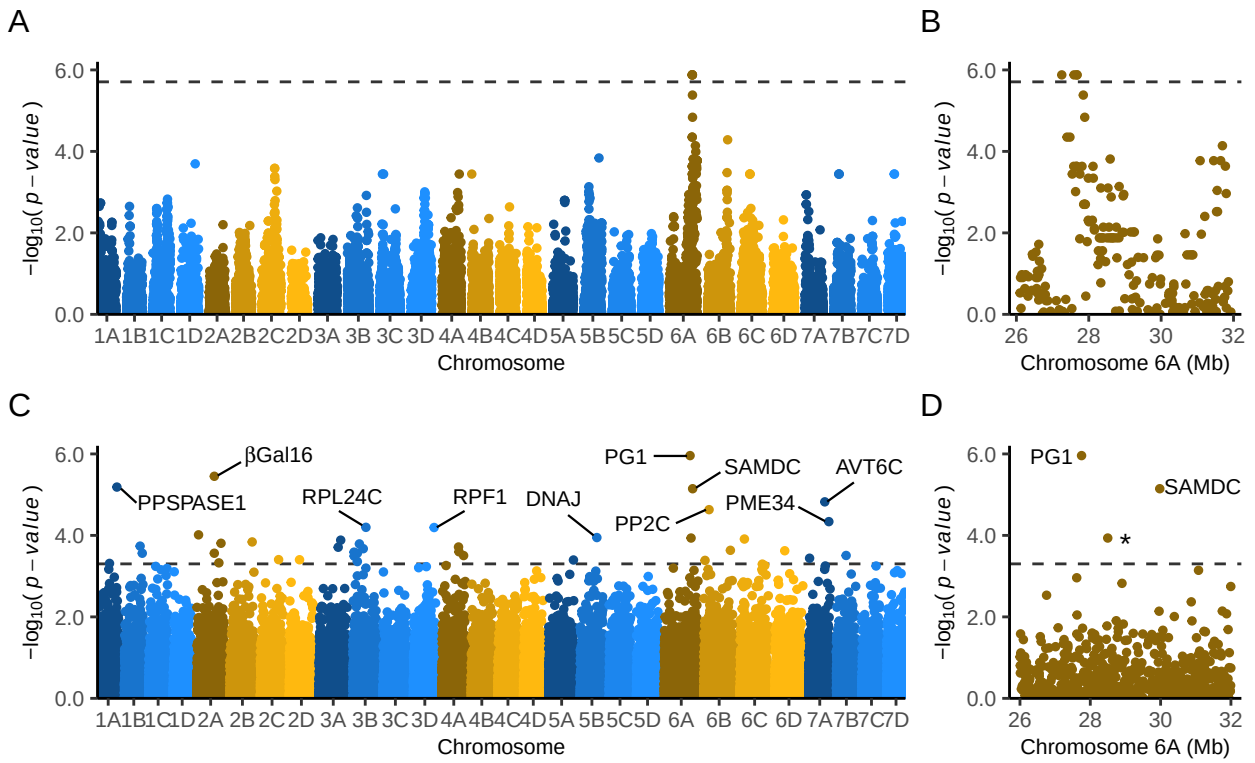

FIGURE 3.2. **Genome-wide association study (GWAS) and quantitative trait transcript (QTT) analyses identify genetic variants associated with phenotypic variation for fruit firmness and transcripts associated with differentially expressed genes among soft- and firm-fruited individuals ($n = 85$).** Study population individuals were genotyped for 49,330 single nucleotide polymorphisms (SNPs) physically anchored to the FaRR1 reference genome. (A) The GWAS Manhattan plot illustrates SNPs associated with fruit firmness across the strawberry genome (physical positions of array-genotyped SNP on the x-axis are shown in the FaRR1 reference genome). GWAS was applied to phenotypic means estimated from 24 observations/individual using a Bonferroni-corrected significance threshold of 5.7 (depicted by the horizontal dashed line). (B) The physical positions of SNPs associated with phenotypic variation for fruit firmness are shown for Mb 26-32 on chromosome 6A. (C) The QTT Manhattan plot was constructed from analyses of 59,126 transcripts mapped in the FaRR1 reference genome using mRNAs isolated from ripe fruit of soft- or firm-fruited individuals in the study population ($n = 85$). QTT was applied to transcript counts estimated from short-read mRNA sequences using a Bonferroni-corrected significance threshold of 3.3 (depicted by the horizontal dashed line). The differentially expressed genes labeled in the QTT Manhattan plot are pyrophosphate-specific phosphatase1 (PPSPASE1), β-galactosidase 16 (βGal16), ribosomal protein L24C (RPL24C), RNA processing factor 1 (RPF1), chaperone DNAj-domain (DNAJ), polygalacturonase 1 (PG1), S-adenosylmethionine decarboxylase (SAMDC), protein phosphatase 2c (PP2C), pectin methylesterase 34 (PME34), and amino acid transporter avt6c (AVT6C). (D) The physical positions of differentially expressed genes are shown for Mb 26-32 on chromosome 6A. Asterisk indicates a gene with unknown function on chromosome 6A.

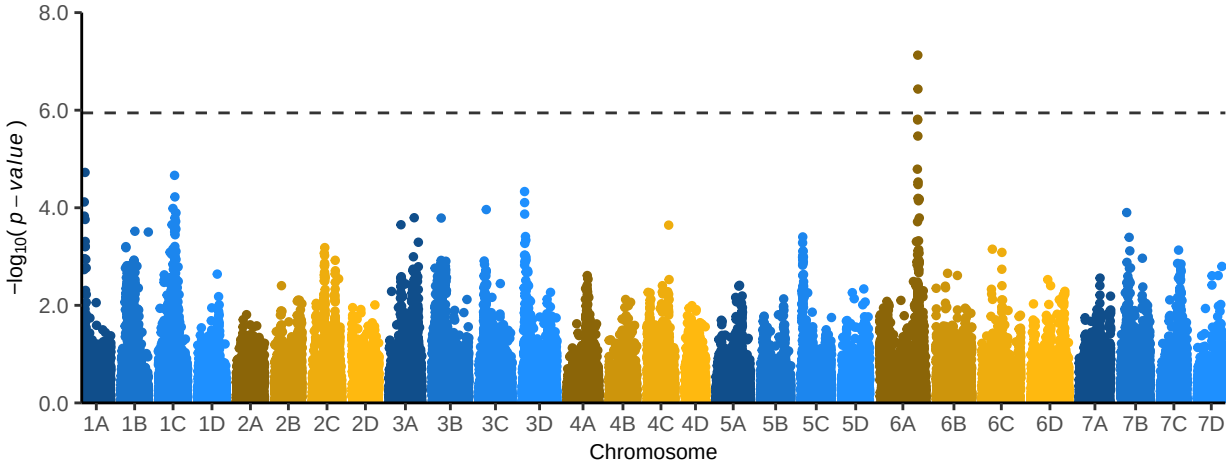

FIGURE 3.3. Single nucleotide polymorphisms (SNPs) associated with phenotypic variation for fruit firmness among a genetically diverse collection of soft- to firm-fruited individuals ($n = 460$) phenotyped as described by Hardigan et al. (2021b). These individuals were genotyped with a 50K Axiom® SNP array. The physical positions of SNPs on the x-axis of the Manhattan plot are coordinates in the Royal Royce reference genome (FaRR1). The genome-wide association study analysis was applied to phenotypic means estimated from 11 observations/individual using a Bonferroni-corrected significance threshold of 5.9 (depicted by the horizontal dashed line).

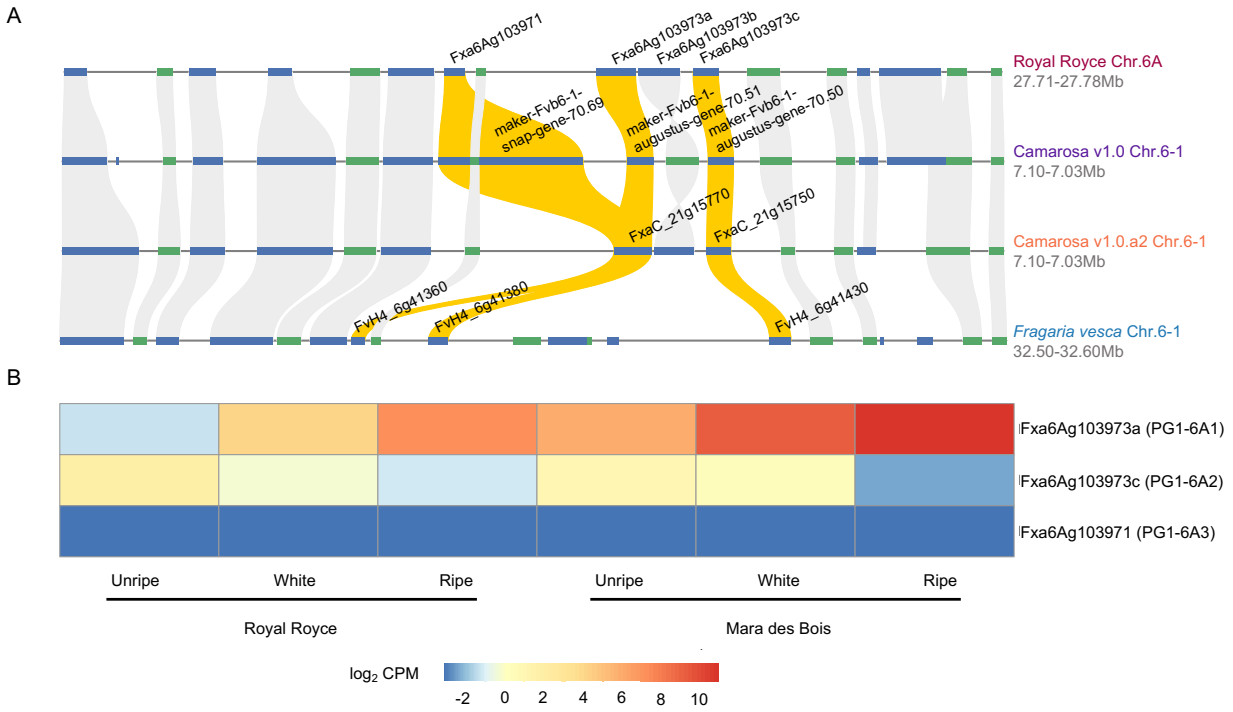

FIGURE 3.4. **Annotations and physical positions of polygalacturonase genes in linkage disequilibrium with a fruit firmness QTL on chromosome 6A in octoploid strawberry.** (A) Organization and synteny of three tandemly duplicated polygalacturonase-encoding genes on chromosome 6A in the 'Royal Royce' and 'Camarosa' genomes and chromosome 6 in the 'Hawaii 4' *F. vesca* genome. (B) Transcript counts per million (CPM) for four polygalacturonase-encoding genes observed in the soft-fruited cultivar 'Mara des Bois' and firm-fruited cultivar 'Royal Royce'. CPMs were estimated from short-read RNA sequences normalized for sequencing depth. Gold lines depict the synteny of *PG1* paralogs across genomes.

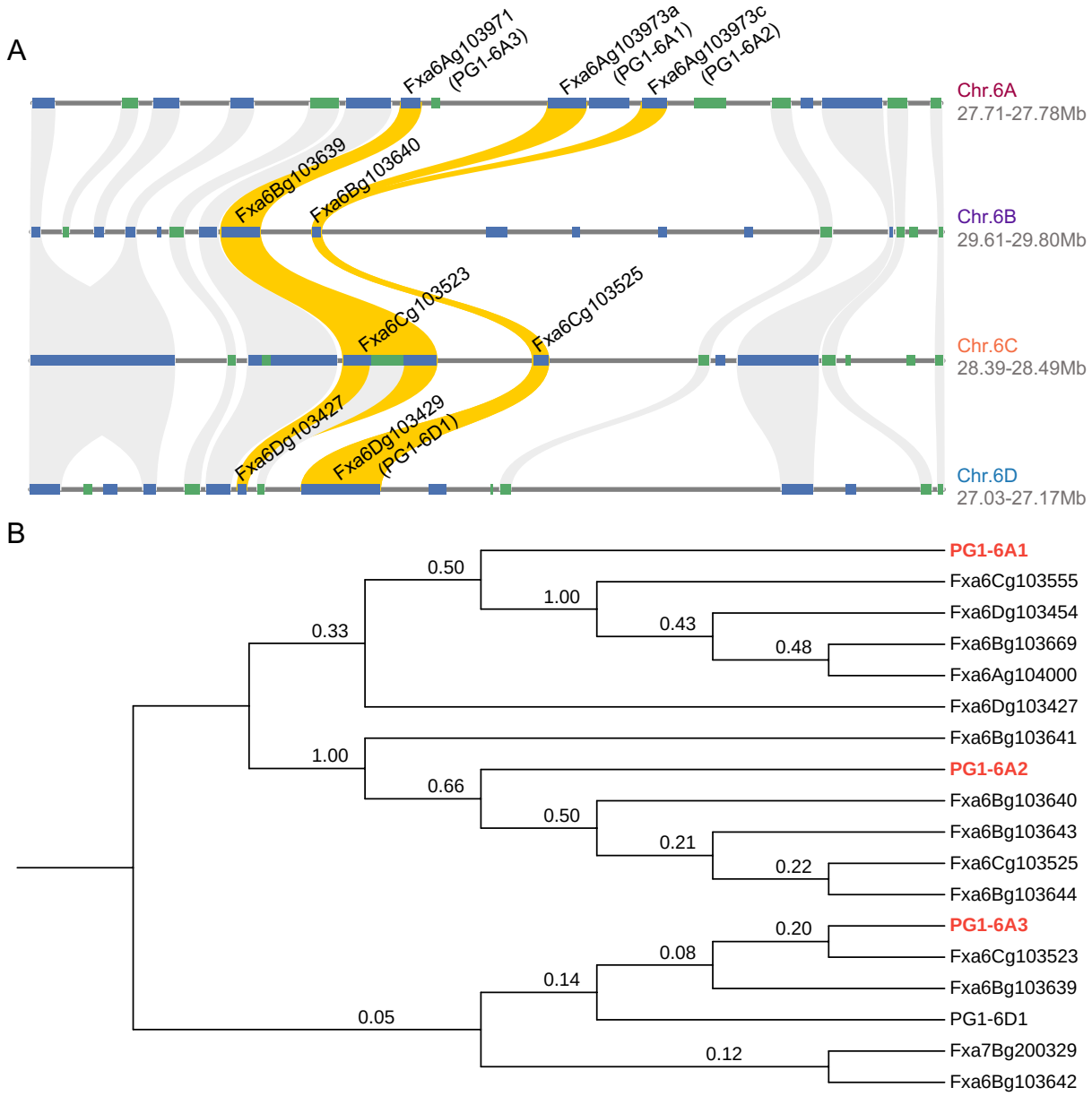

FIGURE 3.5. **Local synteny and phylogenetic tree.** A) Synteny analysis of homoeologous *PG1* genes across strawberry subgenomes using the 'Royal Royce' reference genome. The synteny relationships for the *PG1* genes across different subgenomes are indicated by gold lines. B) Evolutionary relationships among homoeologous *PG1* genes from different subgenomes, highlighting tandemly arranged *PG1* genes on chromosome 6A in red. Phylogenetic tree constructed using amino acid sequences of all homoeologous *PG1* genes identified in this study. Bootstrap values are shown at the nodes, representing the confidence level for each branch based on 1,000 bootstrap replicates.

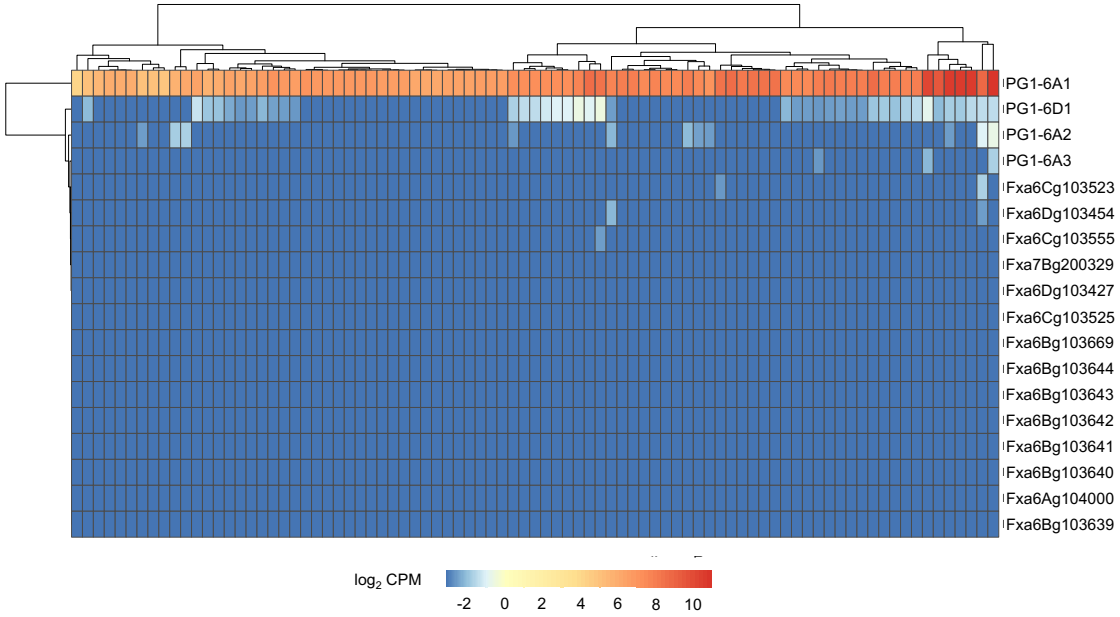

FIGURE 3.6. **The expression profiles of 18 polygalacturonase-encoding genes in ripe fruit of 85 individuals.** The 18 genes include three tandemly duplicated paralogs associated with a fruit firmness QTL on chromosome 6A (*PG1-6A1*, *PG1-6A2*, and *PG1-6A3*), a homoeolog on chromosome 6D (*PG1-6D1*), and 14 additional homoeologs on chromosomes 6B, 6C, and 6D. The heat map color indexes the $\log_2$-transformed transcript count per million estimated from short-read RNA sequences isolated from ripe fruit.

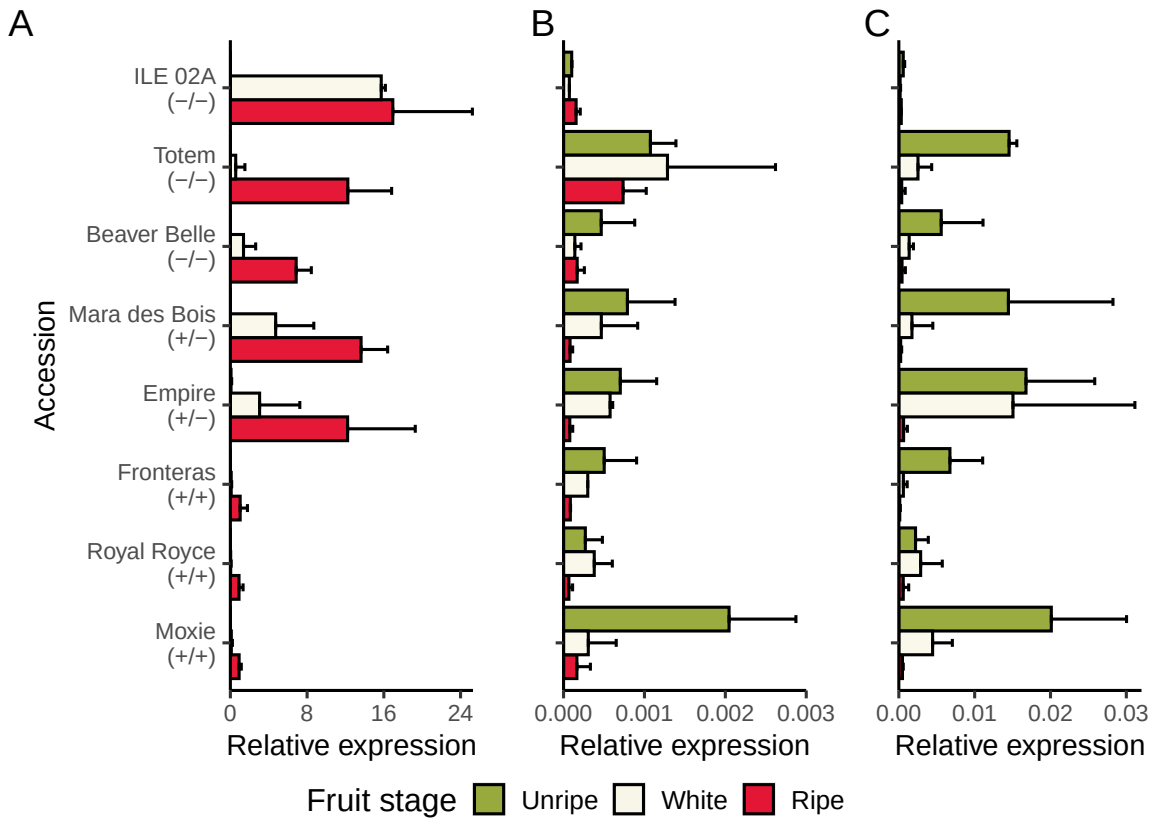

FIGURE 3.7. **Relative gene expression profiles of polygalacturonase genes across strawberry fruit ripening stages.** Quantitative analysis of mRNA abundance for A) *PG1-6A1*, B) *PG1-6A2*, and C) *PG1-6D1* genes in strawberry accessions with distinct fruit firmness profiles at unripe, white, and ripe stages. Expression was quantified over three biological and technical replicates, using a DNA-binding protein gene as the normalization standard. The favorable allele for firmness is indicated as +.

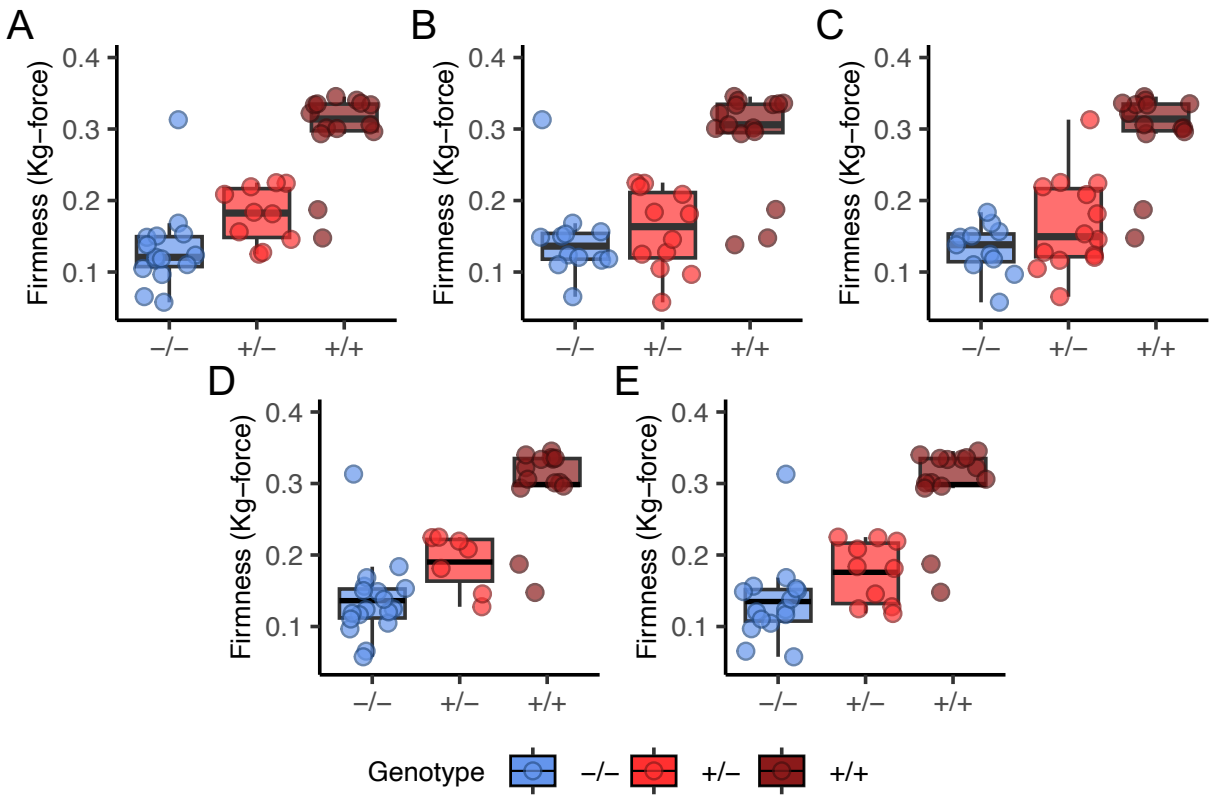

FIGURE 3.8. **Fruit firmness variation among 43 soft- to firm-fruited individuals genotyped for an Enhancer/Suppressor-mutator (*En/Spm*) insertion-deletion (INDEL) and single nucleotide polymorphisms (SNPs) associated with the *PG1-6A1* locus on chromosome 6A.** Genetic variants were genotyped using genotyping-by-sequencing. The points display phenotypic means (estimated marginal means) estimated from five biological replicates (clones)/individual, five harvests, and three subsamples/replicate/harvest among greenhouse grown plants of the DNA sequenced individuals (11 observations/individual). The box displays the genotypic median and interquartile range within each genotypic class, where -/- are unfavorable allele homozygotes, +/- are heterozygotes, and +/+ are favorable allele homozygotes. (A) SNP interrogated by AX-184953741, one of four Axiom® 50K array SNP markers identified by GWAS found upstream of *PG1-6A1* and in complete LD with one another. (B) SNP interrogated by AX-184210676, one of four Axiom® 50K array SNP markers identified by GWAS found upstream of *PG1-6A1* and in complete LD with one another. (C) A 4,948-bp *En/Spm* INDEL 3,926 bp upstream of *PG1-6A1*. (D) A G/T SNP in the 5'-UTR of *PG1-6A1*. (E) SNP interrogated by AX-184242253, an Axiom® 50K array SNP marker identified by expression-QTL analysis found downstream of *PG1-6A1*.

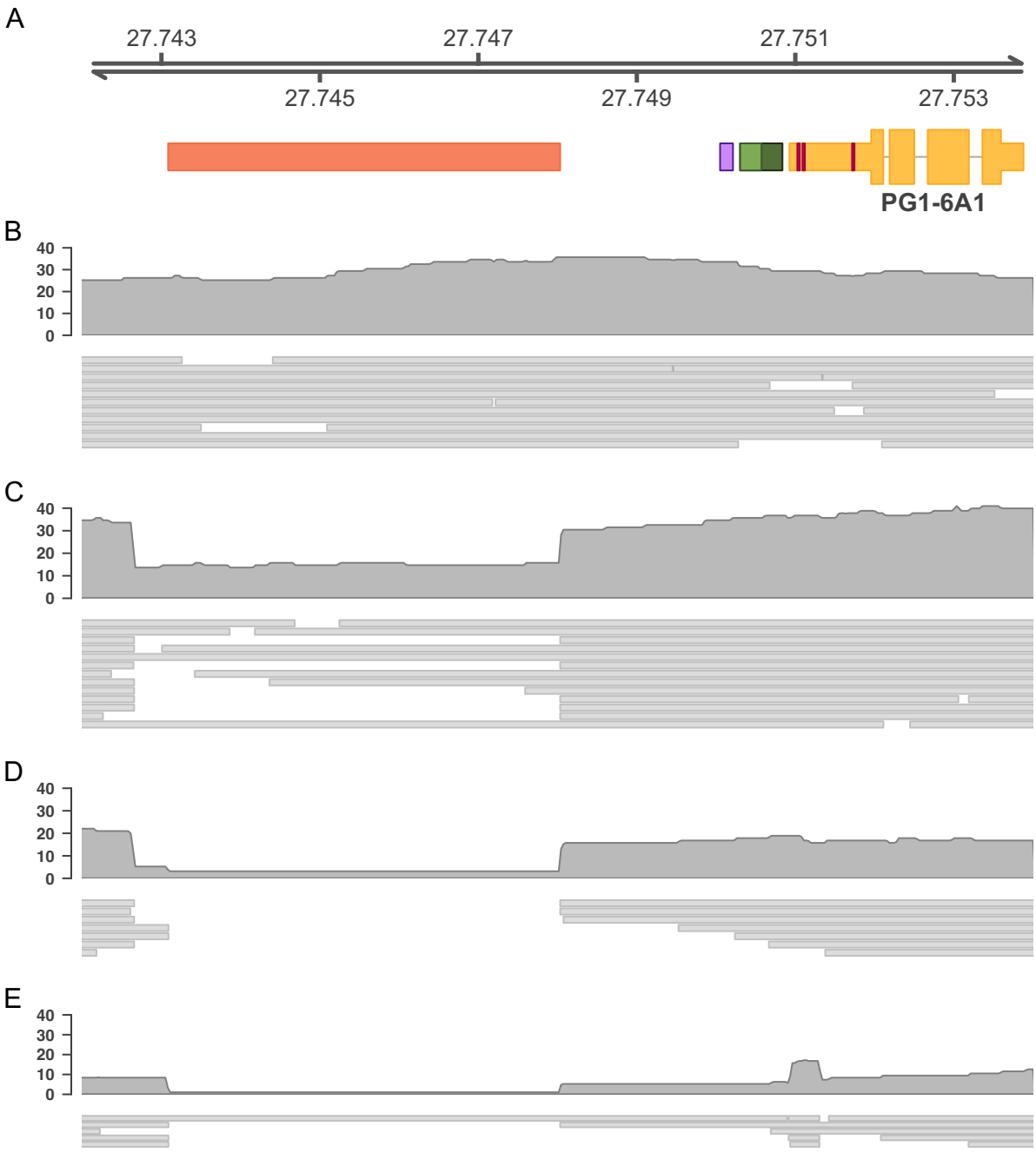

FIGURE 3.9. **Gene model and sequence alignment for *PG1-6A1* in firm- and soft-fruited strawberry accessions.**(A) Gene model for *PG1-6A1* showing single nucleotide polymorphisms (SNPs) detected in the 5'UTR as well as insertion-deletion (INDEL) events located at 1121, 1385, 1745, and 3926 bp upstream from the ATG codon. Alignment of high fidelity long-read DNA sequences was shown for the firm-fruited cultivar (B) 'Royal Royce' (0.35 kg-force), and the soft-fruited exotics (C) 'Mara des Bois' (0.10 kg-force), (D) 'Beaver Belle' (0.09 kg-force), and (E) 'ILE 02' (0.13 kg-force).

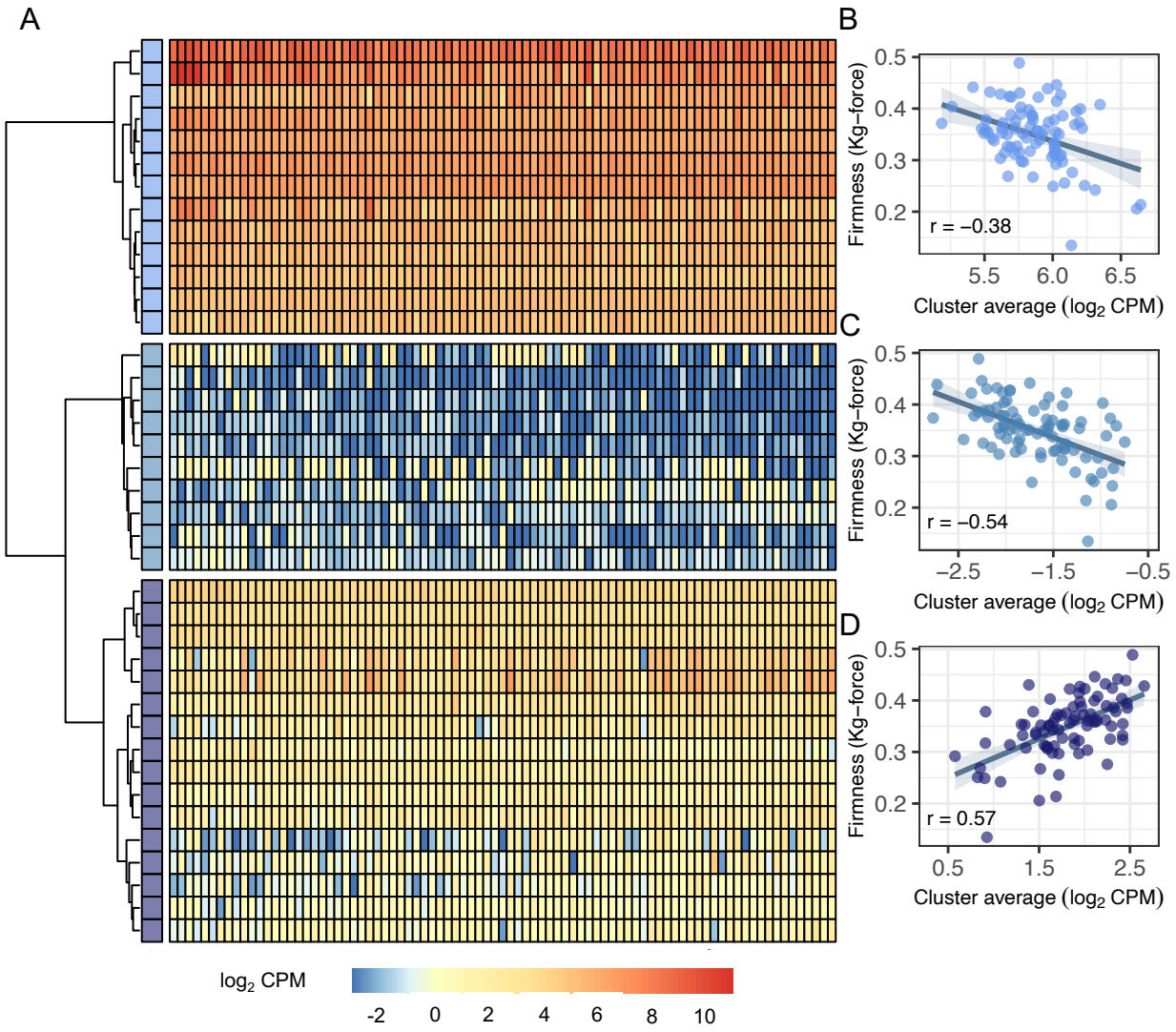

FIGURE 3.10. **Co-expression network analysis of transcripts in ripe fruit of 85 discovery population individuals.** (A) Transcript abundance heat map for genes in the *PG1-6A1* polygalacturonase co-expression network (upper panel) and two other networks identified by co-expression analysis (middle and lower panels) using hierarchical cluster analysis. (B) The correlation between transcript abundance and fruit firmness for genes in the co-expression network is shown in the upper panel of A. (B) The correlation between transcript abundance and fruit firmness for genes (nodes) in the co-expression network is shown in the middle panel of A. (B) The correlation between transcript abundance and fruit firmness for genes (nodes) in the co-expression network is shown in the lower panel of A.

TABLE 3.3. **List of genes significantly associated with strawberry firmness.** Correlation between transcript abundance and fruit firmness in 85 individuals was analyzed through a quantitative trait transcript (QTT) analysis. Descriptions of QTTs significantly associated with fruit firmness ($p \leq 0.005$) include the gene ID, chromosome location, physical position in bp, predicted function, and homology to *Arabidopsis thaliana*.

Gene ID	**Chr**	**Position (bp)**	**Annotation**	**AGI homolog**
Fxa1Ag101585	1A	9326707-9329306	Potential Queuosine, Q, salvage protein family	-
Fxa1Ag102738	1A	18599667-18601666	Putative Phosphatase	AT1G73010
Fxa1Bg202718	1B	21523590-21526505	C2 domain	AT1G74720
Fxa1Bg202317	1B	16467571-16474035	PHD finger protein male meiocyte death 1-related	AT1G33420
Fxa2Ag103645	2A	25479922-25483823	Zinc carboxypeptidase	AT5G42320
Fxa2Ag103376	2A	23950123-23952228	Nucleotide-diphospho-sugar transferases	AT4G33330
Fxa2Ag102751	2A	20233768-20239168	Beta-galactosidase 16	AT1G77410
Fxa2Ag102750	2A	20224814-20228497	Proteasome subunit alpha/beta	AT1G21720
Fxa2Ag100287	2A	1913880-1915521	Glucosyl/Glucuronosyl transferases	AT1G05675
Fxa2Bg203454	2B	25810859-25813703	ATP-sulfurylase	AT1G19920
Fxa2Cg202753	2C	21517798-21520068	Tetratricopeptide-like helical domain superfamily	AT5G48850
Fxa2Dg201701	2D	15643431-15645723	Tetratricopeptide-like helical domain superfamily	AT5G48850
Fxa3Ag103343	3A	24740383-24743537	CRAL/TRIO domain	AT3G46450
Fxa3Ag102927	3A	21001962-21004626	Nucleic acid-binding, OB-fold	AT5G08020
Fxa3Bg200155	3B	726016-729670	Aspartic proteinase-like	AT1G11910
Fxa3Bg200138	3B	646897-648129	Chaperonin-like RbcX superfamily	AT4G04330
Fxa3Bg200956	3B	5741483-5744448	Rx N-terminal domain	AT3G07040
Fxa3Bg200627	3B	3696159-3701280	Acetolactate synthase small subunit 1, chloroplastic	AT5G16290
Fxa3Bg200626	3B	3688029-3691371	Trehalose-phosphatase	AT1G06410
Fxa3Bg202102	3B	15275341-15276575	Ribosomal protein L24e, conserved site	AT2G44860

Table 3.3 continued from previous page

Fxa3Bg201592	3B	10628038-10639237	L-type lectin-domain containing receptor kinase IX.1-like	AT5G10530
Fxa3Dg203357	3D	27620912-27624542	PentatricoPeptide Repeat (ppr) superfamily protein	AT1G12700
Fxa4Ag103318	4A	25110615-25112551	EamA-like transporter family	AT5G40240
Fxa4Ag102677	4A	21393662-21398206	HEAT repeat	AT3G25800
Fxa4Ag102621	4A	21064572-21065514	-	AT5G15880
Fxa5Ag203613	5A	25050026-25053600	Methyltransferase	AT1G78140
Fxa5Bg102468	5B	18678005-18680164	DnaJ domain	AT3G12170
Fxa6Ag104340	6A	29976026-29976574	S-adenosylmethionine decarboxylase, core	AT3G02470
Fxa6Ag104099	6A	28496345-28497132	-	-
Fxa6Ag103973	6A	27751958-27753554	Pectin lyase fold/virulence factor	AT3G07820
Fxa6Bg100154	6B	954903-957130	-	AT5G51410
Fxa6Bg100769	6B	4567122-4575719	ATP synthase (E/31 kDa) subunit	AT2G38600
Fxa6Bg104123	6B	32806865-32808033	OS03G0209700 Protein	AT3G29280
Fxa6Cg100650	6C	4042067-4042554	PentatricoPeptide Repeat (ppr) superfamily protein	AT1G63330
Fxa6Dg101561	6D	10880821-10881339	Zinc finger AN1-type	AT2G36320
Fxa7Ag203101	7A	21166929-21170332	Pectinesterase	AT3G49220
Fxa7Ag202485	7A	17801960-17803847	Amino acid transporter AVT6C	AT3G56200
Fxa7Ag200163	7A	1310646-1315814	S-adenosyl-L-methionine-dependent methyltransferase	AT5G13710
Fxa7Bg201670	7B	14335172-14342432	PHD-zinc-finger like domain	AT3G61740

TABLE 3.4. **The physical locations and putative functions of genes with statistically significant quantitative trait transcript analysis (QTT) analysis signals that belong to the *PG1-6A1* co-expression network.** The correlation (r) between transcript abundance (count per million) and fruit firmness was estimated from short-read RNA sequences of 85 individuals. Chromosome numbers (CNs), physical positions, and gene IDs are from annotations of genes in the Royal Royce reference genome. The predicted gene functions and Arabidopsis Genome Initiative (AGI) locus IDs are from annotations of genes in The Arabidopsis Information Resource (TAIR).

Gene ID	CN	Position (bp)	Predicted Function	Locus ID	r	p-value
Fxa2Ag102750	2A	20,224,814-20,228,497	proteasome β subunit C1	AT1G21720	-0.39	2.7×10^{-4}
Fxa2Ag102751	2A	20,233,768-20,239,168	β-galactosidase 16	AT1G77410	-0.48	3.5×10^{-6}
Fxa3Ag103343	3A	24,740,383-24,743,537	SEC14 phosphoglyceride transfer protein	AT3G46450	-0.40	1.3×10^{-4}
Fxa3Bg200155	3B	726,016-729,670	aspartic proteinase A1	AT1G11910	-0.38	3.2×10^{-4}
Fxa3Bg200626	3B	3,688,029-3,691,371	trehalose-6-phosphate synthase/phosphatase	AT1G06410	-0.37	4.4×10^{-4}
Fxa3Bg200627	3B	3,696,159-3,701,280	valine-tolerant 1	AT5G16290	-0.39	2.7×10^{-4}
Fxa4Ag102677	4A	21,393,662-21,398,206	protein phosphatase 2A subunit A2	AT3G25800	-0.39	2.5×10^{-4}
Fxa6Ag103973a	6A	27,750,926-27,751,957	polygalacturonase 1	AT3G07820	-0.50	1.1×10^{-6}
Fxa6Ag104340	6A	29,976,026-29,976,574	S-adenosylmethionine decarboxylase	AT3G02470	-0.47	7.2×10^{-6}
Fxa6Dg101561	6D	10,880,821-10,881,339	A20/AN1-like zinc finger protein	AT2G36320	-0.39	2.4×10^{-4}
Fxa7Ag200163	7A	1,310,646-1,315,814	sterol methyltransferase 1	AT5G13710	-0.38	3.6×10^{-4}
Fxa7Ag203101	7A	21,166,929-21,170,332	pectin methylesterase 34	AT3G49220	-0.43	4.6×10^{-5}
Fxa7Bg201670	7B	14,335,172-14,342,432	SET domain protein 14	AT3G61740	-0.38	3.1×10^{-4}

Notably, module 1 (Table 3.4) contains *PG1-6A1*, *FaSAMDC* (Guo et al., 2018), and other genes involved in cell wall remodeling, e.g., β-galactosidase 16 (βGAL16);Paniagua et al., 2016,2) and pectin methylesterase 34 (Ponce et al., 2023). Although the genes in module 2 were less strongly expressed than those in module 1, several were identified to have functions associated with abscisic acid signaling through phosphate metabolism (Jung et al., 2020; Zhou et al., 2012) and fruit softening through carbon metabolism and methylation reactions e.g., a glycogenin-like starch initiation protein (Ahn et al., 2019) and a SAM-dependent methyltransferase (Ahn et al., 2019; Soares et al., 2021; Xiong et al., 2024). Interestingly, module 3 contains several genes with DNA/RNA binding activity, which may reflect the alternate metabolic profile of firm fruit with reduced cell wall breakdown. For example, the restorer of fertility-like pentatricopeptide repeat protein is essential for efficiently processing a mitochondrial NAD4 mRNA (Hölzle et al., 2011). Incorrect splicing of the mitochondrial NAD4 transcript might disrupt carbon metabolism and ATP production due to abnormal

NADH metabolism (Nakagawa and Sakurai, 2006). Further study of these genes, some of which have not been studied or well characterized, could shed additional light on the processes underlying fruit softening in strawberry.

3.4.7. KASP assays for marker-assisted selection of *PG1-6A1*-associated genetic variants are specific and accurate. To facilitate marker-assisted selection of the *PG1-6A1* locus, Kompetitive Allele-Specific PCR (KASP) assays were developed and tested for one of the peak SNPs identified by GWAS (AX-184210676), three SNPs in the 5'-UTR identified by SIFT and QTL analyses, a SNP downstream of *PG1-6A1* identified by e-QTL analysis (AX-184242253), and the *En/Spm* INDEL (Table 3.5; Figure 3.11). To assess their specificity, accuracy, and predictive values, 92 individuals spanning the domestication and fruit firmness spectra in strawberry (hereafter the 'diversity' population) were selected and genotyped. They included 29 heirloom cultivars (1854-1991), 27 UC cultivars (1935-2017), 23 additional UC hybrid individuals (1961-2020), four *F. chiloensis* ecotypes, eight *F. virginiana* ecotypes, and a single *F. vesca* ecotype. Their fruit firmness phenotypes ranged from 0.07 to 0.45 kg-force.

The KASP assays we developed for *PG1-6A1*-associated genetic variants appear to be paralog-specific (only amplified *PG1-6A1* alleles), but differed in genotyping accuracy and genotype call rates (Table 3.5; Figure 3.11). Assays designed from the 50K array SNPs (K-676 for the AX-184210676 SNP and K-253 for the AX-19424253 SNP) were highly accurate with distinct, codominant genotypic clusters and 98.9-100.0% genotype call rates (Table 3.5; Figure 3.2 and 3.11). The clusters were more well-separated and equidistant for K-253 than K-676 (Figure 3.12). Although both performed well, the genotype-to-phenotype correlation was stronger for K-253 ($\hat{r}_{\overline{y},M} = 0.83$) than K-676 ($\hat{r}_{\overline{y},M} = 0.67$). This is in keeping with the reduced variation explained by AX-184210676 in the 43 long-read sequenced individuals, indicating that this SNP may lose value in more diverse populations.

Of the three KASP assays designed for SNPs in the 5'-UTR of *PG1-6A1*, one was highly accurate (K-732), one failed, and the other was slightly less accurate and did not perform

TABLE 3.5. **Statistics[a] for Kompetitive Allele-Specific PCR (KASP) markers[b] developed for genotyping an Enhancer/Suppressor-mutator (*En/Spm*) insertion-deletion (INDEL) and single nucleotide polymorphisms (SNPs) associated with the *PG1-6A1* locus on chromosome 6A.**

Population[c]	Reference Variant	KASP Marker (M)	FAF[d]	Accuracy (%)[e]	Call Rate (%)[f]	$\hat{r}_{\hat{y},M}$	$\hat{a}$ (kg-force)	$Pr > F$	$\hat{d}$ (kg-force)	$Pr > F$	$\lvert\hat{d}/\hat{a}\rvert$
Diversity	AX-184210676	K-676	0.55	97.8	98.9	0.67	0.09	< 0.0001	0.04	0.05	0.44
	En/Spm INDEL	K-SPM	0.65	81.0	86.0	0.64	0.09	< 0.0001	0.00	0.95	0.02
	5'-UTR	K-732	0.55	100.0	82.8	0.82	0.10	< 0.0001	0.01	0.38	0.14
	AX-184242253	K-253	0.50	95.6	100.0	0.83	0.10	< 0.0001	0.02	0.22	0.19
Full-sibs	AX-184210676	K-676	0.54	NA	99.4	0.35	0.04	< 0.0001	0.01	0.28	0.33
	En/Spm INDEL	K-SPM	0.62	NA	97.0	0.45	0.04	< 0.0001	0.03	0.01	0.74
	5'-UTR	K-732	0.40	NA	84.9	0.48	0.05	< 0.0001	0.03	0.01	0.56
	AX-184242253	K-253	0.50	NA	95.8	0.51	0.05	< 0.0001	0.02	0.03	0.41

[a] [a]$r_{\overline{y},M}$ is the correlation between the phenotypic mean ($\overline{y}$) for fruit firmness (kg-force) and marker genotypes, $\hat{a}$ is the additive and ($\hat{d}$) is the dominance effect, and $|\hat{d}/\hat{a}|$ is the degree-of-dominance of the KASP marker locus. The correlation coefficeint estimates ($r_{\overline{y},M}$) for every marker locus were highly significant ($p \leq 0.0001$) in both populations. Additive and dominance effects were estimated by linear contrasts among genotypic means ($\overline{y}_{+/+}$, $\overline{y}_{+/-}$, and $\overline{y}_{-/-}$). Significance levels ($Pr > F$) are shown in columns to the right of additive and dominance effect estimates for tests of the null hypothesis that the linear contrast was not significantly different from zero, where $\hat{a} = (\overline{y}_{+/+} - \overline{y}_{-/-})/2$ and $\hat{d} = (\overline{y}_{+/+} + \overline{y}_{-/-})/2 - \overline{y}_{+/-}$.
[b] [b]KASP markers were designed for the *En/Spm* INDEL and SNPs shown in Table 3.2.
[c] [c]The diversity population consisted of 92 soft- to firm-fruited individuals. The full-sib population consisted of 152 one-year-old individuals within four full-sib families segregating for mutant and wildtype *PG1-6A1* alleles.
[d] [d]FAF is the frequency of the KASP-SNP or KASP-INDEL marker allele associated with the favorable *PG1-6A1* allele.
[e] [e]Accuracy (%) was estimated for the K-SPM and K-732 marker by comparing genotypes called by GBS among long-read DNA sequences of 43 individuals with genotypes called by KASP. Accuracy was estimated for the K-676 and K-253 markers by comparing genotypes called among 92 Axiom® 50K array-genotyped individuals with genotypes called by KASP. KASP marker accuracy could not be estimated among full-sib individuals because they were only genotyped with the KASP markers and were not sequenced or genotyped with the 50K SNP array.
[f] [f]The call rate is the percentage of individuals where KASP genotypes were successfully called.

as well as K-732. K-732 genotypes were strongly correlated with fruit firmness phenotypes ($\hat{r}_{\overline{y},M} = 0.82$). The genotype call rate for the K-732 assay (82.8%) was similar to that for K-SPM assay (86.0%), both of which were markedly lower than the the genotype call rates observed for the K-676 and K-253 assays (Table 3.11; Figure 3.11 and 3.12).

The KASP assay designed for the *En/Spm* INDEL (K-SPM) was less accurate than the K-732 and K-253 assays (Table 3.5; Figure 3.11 and 3.12). The lower performance of that assay was presumably caused by the technical difficulty of amplifying DNA sequences bordering the 4,948-bp insertion (Figure 3.9). Using long-read DNA sequences as a reference, genotyping errors for the K-SPM assay appear to have been primarily caused by mistyping insertion homozygotes as heterozygotes, which caused significant segregation distortion among the full-sib progeny phenotyped and genotyped in this study (Figure 3.11).

The observed segregation ratios for the K-676, K732, and K-253 assays were not significantly different from the expected segregation ratio among the full-sib progeny genotyped in the KASP marker validation study (1 +/+ : 2 +/- : 1 -/-): K-676 ($\chi^2 = 2.88; p = 0.24$), K-732 ($\chi^2 = 0.44$; $p = 0.81$), and K-253 ($\chi^2 = 0.25; p = 0.88$). The observed segregation ratio for the K-SPM assay (69 +/+ : 51 +/- : 32 -/-), by comparison, was significantly distorted with an excess of favorable allele homozygotes and shortage of heterozygotes ($\chi^2 = 34.36; p \leq 0.0001$). Although the *En/Spm* was homozygous and appears to be predictive of the favorable *PG1-6A1* allele found in modern, firm-fruited UC cultivars, the *En/Spm* insertion-deletion proved to be more difficult to accurately genotype by KASP than the SNPs we targeted.

3.4.8. The favorable *PG1-6A1* allele appears to be nearly fixed in a population with a long history of selection for increased fruit firmness. Using GBS- and KASP-called genetic variants, we discovered that the favorable *PG1-6A1* allele was nearly fixed in modern, firm-fruited UC cultivars and associated genetic resources, hereafter identified as the California population (Figure 3.13 and 3.14). We discovered this in part by screening the parents of 178 full-sib families with three KASP markers: K-676, K-SPM, and K-253. Using those KASP markers, the frequency of the favorable *PG1-6A1* allele was estimated to range from 0.93 to 0.97 among 10,650 full-sib individuals sampled from the California population in the 2023-24 cycle of selection. Of the 178 full-sib families, only four originated from crosses where both parents were heterozygous for KASP marker alleles and predicted to segregate 1 $PG1\text{-}6A1^{+}/PG1\text{-}6A1^{+}$: 2 $PG1\text{-}6A1^{+}/PG1\text{-}6A1^{-}$: 1 $PG1\text{-}6A1^{-}/PG1\text{-}6A1^{-}$. Importantly, those were the only families in the survey of California population progeny in 2023-24 where the additive and dominance effects of the *PG1-6A1* locus could be estimated (where both parents were heterozygous for *PG1-6A1* alleles). These findings further substantiate that phenotypic selection alone has strongly swept the favorable *PG1-6A1* allele within the California population (Feldmann et al., 2023,2).

Figure 3.11 displays the phenotypes of individuals observed within KASP marker genotypic classes in the diversity and full-sib populations. The additive effect estimates for the *PG1-6A1* locus were nearly identical assay-to-assay and highly significant in the diversity population (0.07-0.08 kg-force), whereas the dominance effects were smaller, mostly non-significant, and slightly more variable (0.0-0.04 kg-force; Table 3.5; Figure 3.11). The same pattern was observed in the full-sib population; however, the additive effect estimates were 50% smaller, and the KASP markers were 30-49% less predictive. Wildtype *PG1-6A1* homozygotes were firmer, and the dominance of the wildtype *PG1-6A1* allele was greater in the full-sib than the diversity population (Table 3.5; Figure 3.11). The degree-of-dominance ($|d/a|$) estimates for KASP markers associated with the *PG1-6A1* locus ranged from 0.02 (nearly additive) to 0.44 in the diversity population and 0.33 to 0.74 in the full-sib population, and increased as the dominance of the wildtype allele increased.

The assay-to-assay variation observed within populations was attributed to differences in KASP assay accuracy, identity-by-state variation, and stochastic variation (Table 3.5; Figure 3.11). We attributed the dampened effect of the *PG1-6A1* locus in the full-sib population to genetic background effects (Table 3.5; Figure 3.11). Although the analysis was limited to four full-sib families because of the high frequency of the favorable *PG1-6A1* allele, phenotypes observed within *PG1-6A1* genotypic classes suggest that favorable alleles for multiple QTL have been targeted by phenotypic selection and accumulated in the California population (Figure 3.8 and 3.11; Tables 3.2 and 3.5). That conclusion is consistent with the finding that long-term selection appears to have virtually eliminated genetic variation for fruit firmness within the California population (Feldmann et al., 2024a,2), which we have shown was substantially caused by the fixation of the favorable *PG1-6A1* allele (Figure 3.13 and 3.14).

3.4.9. *PG1-6A1* loss-of-function mutations more than double fruit firmness in strawberry. The ecotypes and cultivars screened in this study spanned the domestication spectrum from extremely soft-fruited wild relatives to extremely firm-fruited cultivars (Table

3.5; Figure 3.11 and 3.14). Figure 3.15 displays the phenotypic means for 381 additional cultivars and other hybrid individuals not shown in Figure 3.13. The firmest individual in this study (17C140P012), a favorable allele homozygote (*PG1-6A1*$^{+}$/*PG1-6A1*$^{+}$), withstood 0.45 kg-force of pressure (Figure 3.13). The UC cultivars 'Royal Royce' (0.41-0.45 kg-force) and 'Surfline' (0.41-0.44 kg-force) were similarly firm. At the other extreme were three *F. virginiana* ecotypes with extremely soft, fragile, and easily bruised fruit (0.07 to 0.08 kg-force). They were predicted to be homozygous for wildtype (unfavorable) alleles.

The phenotypes of several *F. chiloensis* and *F. virginiana* ecotypes are shown to the left of *F.* × *ananassa* cultivars in Figure 3.13 and 3.15. The analysis shows that ecotypes of the wild relatives are extremely to moderately soft-fruited (0.07 to 0.20 kg-force; $\tilde{y} = 0.15$). The fruit firmness medians ($\tilde{y}$) and maximums ($\overline{y}_{MAX}$) were lower for *F. virginiana* ecotypes ($\tilde{y} = 0.15$ and $\overline{y}_{MAX} = 0.18$ kg-force) and *F. chiloensis* ecotypes ($\tilde{y} = 0.16$ and $\overline{y}_{MAX} = 0.20$ kg-force) than *F.* × *ananassa* cultivars ($\tilde{y} = 0.24$ and $\overline{y}_{MAX} = 0.45$ kg-force) (Table 3.5; Figure 3.11 and 3.13). Statistics for the latter were estimated from the phenotypes of cultivars spanning the fruit firmness range, from 'Sitka' (PI616777; $\overline{y} = 0.09$ kg-force) and 'Jucunda' (PI551623; $\overline{y} = 0.10$ kg-force) at the lower extreme to 'UC Surfline' ($\overline{y} = 0.41$ kg-force) and 'UC Royal Royce' ($\overline{y} = 0.44$ kg-force) at the upper extreme. The fruit firmness median for modern UC cultivars (1988-present; $\tilde{y} = 0.35$ kg-force) was double that of heirloom cultivars ($\tilde{y} = 0.18$ kg-force) and slightly more than double that of wild species ecotypes (Figure 3.13).

We suspect that the phenotypic minimums for *F. chiloensis* and *F. virginiana* ecotypes were overestimated (biased upward) because their fruit were often too soft to phenotype or even disintegrated when fully ripe. Nevertheless, using the phenotypic extremes observed in the diversity population, we estimated that domestication has increased fruit firmness by 628% from the softest wild species ecotypes (0.07 kg-force) to the firmest cultivars (0.45 kg-force). That estimate is remarkably close to the estimate of 768% reported for fruit firmness in a study of historical genetic gains in strawberry (Feldmann et al., 2024a). That genetic gain was substantially, but not solely driven by *PG1-6A1* mutations, which appear to double fruit

firmness in most genetic backgrounds, e.g., the mean fruit firmness predicted by the KASP assay for the *En/Spm* INDEL was $\overline{y}_{+/+} = 0.35$ kg-force for the favorable allele homozygote and $\overline{y}_{-/-} = 0.14$ kg-force for the unfavorable allele homozygote (Table 3.5; Figure 3.11). The *PG1-6A1* genotypes predicted by the AX-184242253 SNP are depicted in a visual reconstruction of the history of breeding for increased fruit firmness in strawberry (Figure 3.13 and 3.14). The phenotypic means ($\overline{y}$) displayed in Figure 3.13 were estimated from a meta-analysis of phenotypic observations collected from field experiments over the course of these studies (Figure 3.15). The hockey stick (exponential growth) curve for fruit firmness uncovered by this analysis traces the increase in the frequency of the favorable *PG1-6A1* allele and favorable allele homozygotes (*PG1-6A1*$^{+}$/*PG1-6A1*$^{+}$) from the early 1950s onwards when moderately firm-fruited cultivars (*PG1-6A1* heterozygotes) began emerging and became catalysts for the expansion of strawberry production in California (Feldmann et al., 2024a). The change in fruit firmness was negligible from 1850 to 1950, apart from notable outliers, e.g., heirloom cultivars 'Aberdeen' and 'Titan' (Figure 3.13).

3.4.10. Outliers and the prediction accuracy of identical-by-state genetic variants. This study could not unambiguously predict *PG1-6A1* genotypes across diverse germpl asm because the SNPs targeted for marker development are identical-by-state (IBS) (Table 3.2 and 3.5). They are highly predictive proxies for the causal mutations in elite UC genetics but are less than 100% predictive across diverse germplasm. Using five array-genotyped SNPs (AX-184953741, AX-184726882, AX-184210676, AX-184275052, and AX-184242253) as proxies for *PG1-6A1* alleles, the frequency of the wildtype *PG1-6A1* allele was estimated to range from 0.61 (AX-184275052) to 0.94 (AX-184242253) and 0.97 (AX-184953741) among 18 ecotypes of the wild relatives screened in this study. This highlights the breakdown in prediction accuracy of non-causal (IBS) genetic variants in strawberry, a highly polymorphic and heterozygous species, e.g., using the AX-18424253 SNP, the *F. chiloensis* ecotype ILE 02A (PI552038; 0.13 kg-force) and *F. virginiana* ecotype PI612320 (0.17 kg-force) were predicted to be heterozygous for *PG1-6A1* alleles. Their soft phenotypes, however, suggest

that they are either homozygous or heterozygous for wildtype alleles. Also, the *PG1-6A1* gene was more strongly expressed in ripe fruit of ILE 02A than any of the $F \times$ *ananassa* cultivars screened by RT-PCR in this study (Figure 3.7). These examples highlight the ambiguities associated with predictions using IBS genetic variants. Although only five of the 18 array-genotyped ecotypes were long-read sequenced, they were found to be homozygous for the *En/Spm* deletion. The *En/Spm* INDEL appears to be highly predictive of wildtype and mutant *PG1-6A1* alleles that are nevertheless riddled with IBS genetic variants.

We suspect that the phenotype observed for 'Aberdeen' in this study (1910; PI551630; $\overline{y} = 0.35$ kg-force) could be erroneous. 'Aberdeen' was predicted to be homozygous for a wildtype *PG1-6A1* allele (Figure 3.13), was described by Darrow et al. (1966) as soft-fruited and "too soft to ship", and was soft-fruited (0.10 kg-force) in the Hardigan et al. (2021b) study. Aberdeen could conceivably carry a favorable *PG1-6A1* allele different from the one found in modern UC cultivars or be a source of favorable alleles for loci other than *PG1-6A1*, but that seems unlikely.

Titan (PI551398; $\overline{y} = 0.33$ kg-force), another firm-fruited heirloom cultivar predicted to be heterozygous for *PG1-6A1* alleles, was as firm as many of the firm-fruited UC cultivars known to be homozygous for the favorable allele (Figure 3.13). This cultivar could be heterozygous for SNPs in LD with different favorable *PG1-6A1* alleles or could carry favorable alleles at loci other than *PG1-6A1*. Although we systematically sampled genetic diversity in the UC and USDA genetic resource collections (https://www.ars-grin.gov/), deeper sampling in other collections might identify additional outliers and sources of novel favorable alleles and shed additional light on the domestication history of strawberry (Figure 3.13 and 3.14).

3.4.11. *PG1-6A1*, a single gene of pivotal importance to strawberry domestication. These analyses show that the mutant *PG1-6A1* allele was swept up by phenotypic selection over two decades (1953-1973) in the UC breeding program before mutant homozygotes and long shelf life cultivars emerged that transformed strawberry production (Figure

3.13, 3.14, and 3.15; Feldmann et al., 2024a). The fruit firmness transitions were exceedingly fast compared to typical Neolithic plant domestication time scales (Doebley et al., 2006; Meyer and Purugganan, 2013); however, strawberry domestication primarily occurred in parallel with the Industrial Revolution and has been anything but typical (Darrow et al., 1966; Feldmann et al., 2024a; Hardigan et al., 2021b). However, the central importance of *PG1* mutations to strawberry domestication had not been previously established, despite early clues from differential gene expression analyses (Salentijn et al., 2003). The phenotypic range for early cultivars was found to be virtually identical to that for wild species ecotypes in our study, most of which were predicted to be homozygous for the wildtype *PG1-6A1* allele using KASP markers (blue points in Figure 3.13 and 3.15). 'Lassen' (1935; 36C003P001; not phenotyped), a descendant of 'Blakemore' (1929; PI551421; $\overline{y} = 0.18$ kg-force), was the oldest UC cultivar predicted to be heterozygous for wildtype and mutant *PG1-6A1* alleles (Figure 3.14). GBS- and KASP-called genotypes suggest that the favorable *PG1-6A1* allele was present but infrequent in early cultivars, was present in UC genetics from inception, and was inherited from 'Blakemore' (Figure 3.14). We found that selection for increased fruit firmness in the early 1950s began exposing the effect of the incompletely dominant favorable *PG1-6A1* allele transmitted by 'Tioga' (1953; 53C009P002; $\overline{y} = 0.25$ kg-force) and 'Tufts' (1963; 63C120P011; $\overline{y} = 0.25$ kg-force), both of which are descendants of 'Lassen' and are the second oldest UC cultivars predicted to be heterozygous for *PG1-6A1* alleles (Figure 3.13 and 3.14).

These early heterozygotes were superseded in the 1970s by the emergence of super firm-fruited *PG1-6A1*$^+$ homozygotes with greatly increased shelf life and decreased perishability, starting with the cultivars 'Douglas' (1972; 72C266P604; $\overline{y} = 0.13$ kg-force) and 'Selva' (1975; 75C071P107; $\overline{y} = 0.30$ kg-force). Digging deeper into the breeding history, the favorable *PG1-6A1* allele found in modern UC cultivars was targeted by phenotypic selection in segregating populations as early as 1953 and increased in frequency until favorable *PG1-6A1* allele homozygotes began emerging approximately two decades later (Figure 3.13 and

3.14). There are thousands of descendants of 'Tioga', 'Tufts', 'Douglas', and 'Selva' in UC pedigree records, which include every UC cultivar developed since 1970, all of which appear to be homozygous for the favorable *PG1-6A1* allele (Figure 3.13 and 3.14). The abbreviated family tree developed for UC cultivars, from 'Lassen' and 'Tioga' to 'Golden Gate' and 'Surfline', illustrates the transition from soft- to firm-fruited phenotypes, the emergence and flow of the favorable *PG1-6A1* allele in UC parents and progeny, and the critical importance of the favorable *PG1-6A1* allele to strawberry domestication (Figure 3.14). The speed with which *PG1-6A1* mutant homozygotes emerged in the twentieth century seems exceedingly slow from our modern, genome-informed breeding vantage point: marker-assisted selection and CRISPR/Cas9 editing of the *PG1-6A1* gene could conceivably collapse the half century domestication process into mere two years (López-Casado et al., 2023; Rodríguez-Leal et al., 2017; Zsögön et al., 2018).

3.5. Conclusions

These results showed that *PG1-6A1* is a major player in determining fruit firmness. The expression of *PG1-6A1* was significantly higher in soft-fruited genotypes compared to firm-fruited ones, suggesting a negative correlation between *PG1-6A1* expression and fruit firmness. The *PG1* locus was significantly associated with several structural variants, which were highly predictive of firmness phenotypes. Furthermore, KASP markers for the *PG1-6A1*-associated SNPs and the *En/Spm* INDEL might efficiently facilitate the selection of firmer genotypes to improve fruit firmness and shelf life in strawberry breeding programs. Additionally, this study identified additional genes associated significantly with fruit firmness, highlighting the co-expression between *PG1-6A1*and genes such as *FaSAMDC* and βGAL16. Overall, these findings provide valuable insights into the genetic architecture of fruit firmness in strawberries and offer practical tools for its improvement through targeted breeding efforts.

3.6. Abbreviations

EMM, estimated marginal means; GWAS, genome-wide association study; QTT, quantitative trait transcript; MAS, marker-assisted selection; SNP, single nucleotide polymorphism; INDEL, insertion-deletion; LMM, linear mixed model; REML, restricted maximum likelihood; PVE, phenotypic variance explained; GVE, genetic variance explained; qPCR, quantitative polymerase chain reaction; KASP, Kompetitive allele-specific PCR; LD, linkage disequilibrium; TE, transposable element; RNA-seq, RNA sequencing.

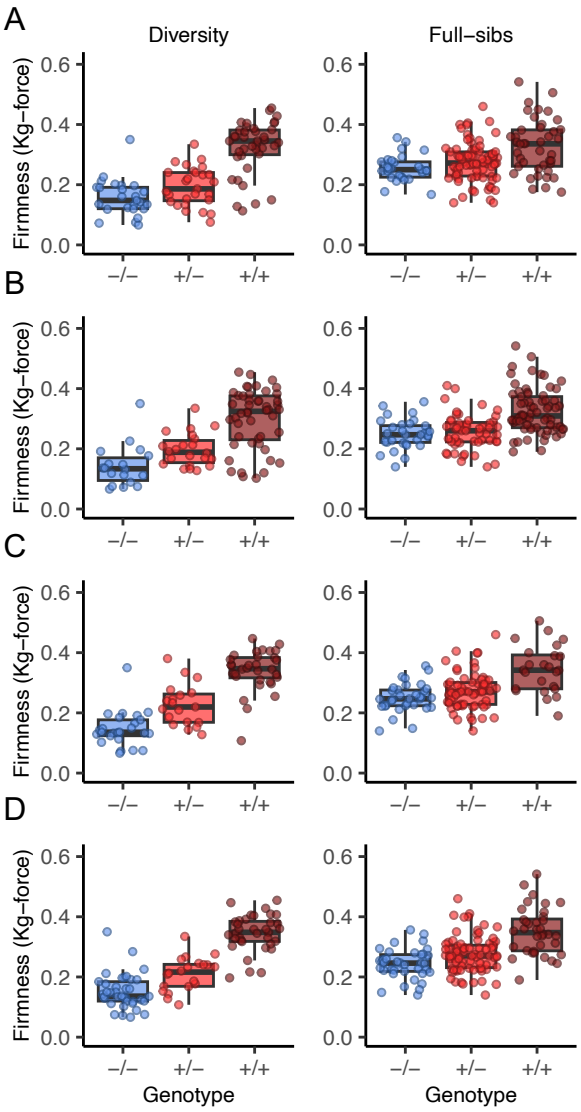

FIGURE 3.11. **Fruit firmness variation among 92 soft- to firm-fruited individuals (the diversity population; left column) and 152 full-sib progeny (the full-sib population; right column) genotyped with KASP markers developed for an Enhancer/Suppressor-mutator (*En/Spm*) insertion-deletion (INDEL) and single nucleotide polymorphisms (SNPs) associated with the *PG1-6A1*) locus.** The points display phenotypic means (estimated marginal means) for 92 individuals in the diversity population (four observations/individual) and 152 individuals in the full-sib population (six observations/individual). The box displays the genotypic median and interquartile range within each genotypic class for each KASP marker, where -/- are unfavorable allele homozygotes, +/- are heterozygotes, and +/+ are favorable allele homozygotes. Genotypes and phenotypes are shown for four KASP markers associated with the *PG1-6A1 locus*: (A) K-676 (Mb 27,676,285); K-SPM (Mb 27,743,085); K-732 (Mb 27,751,732); and K-253 (Mb 27,888,596).

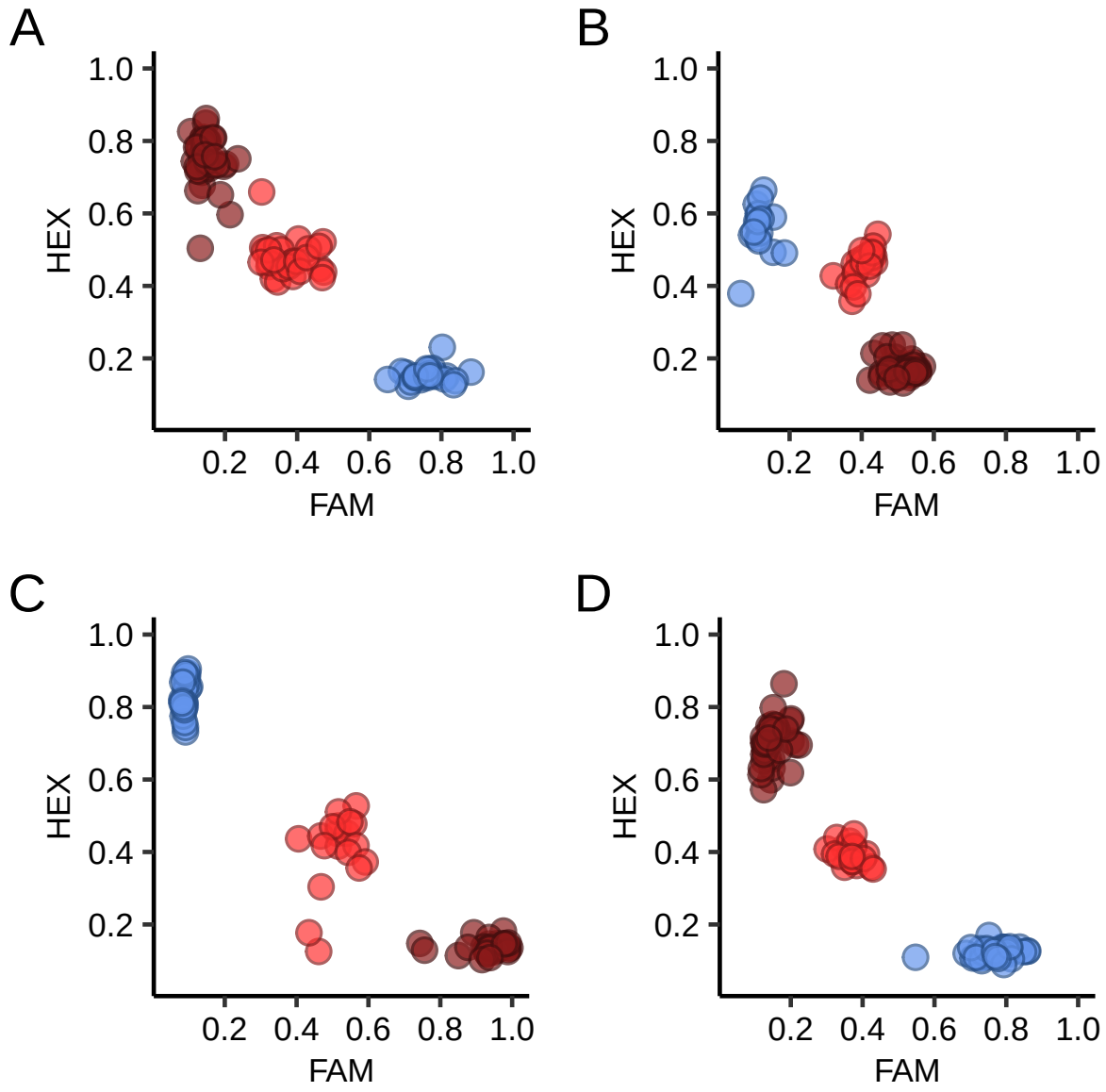

FIGURE 3.12. Allele discrimination scatter plot displaying the fluorescence intensity of FAM signal on the x-axis and HEX signal on the y-axis for 92 octoploid strawberry individuals (the diversity population) genotyped with four KASP markers: (A) K-676; (B) K-SPM; (C) K-732; and (D) K-253.

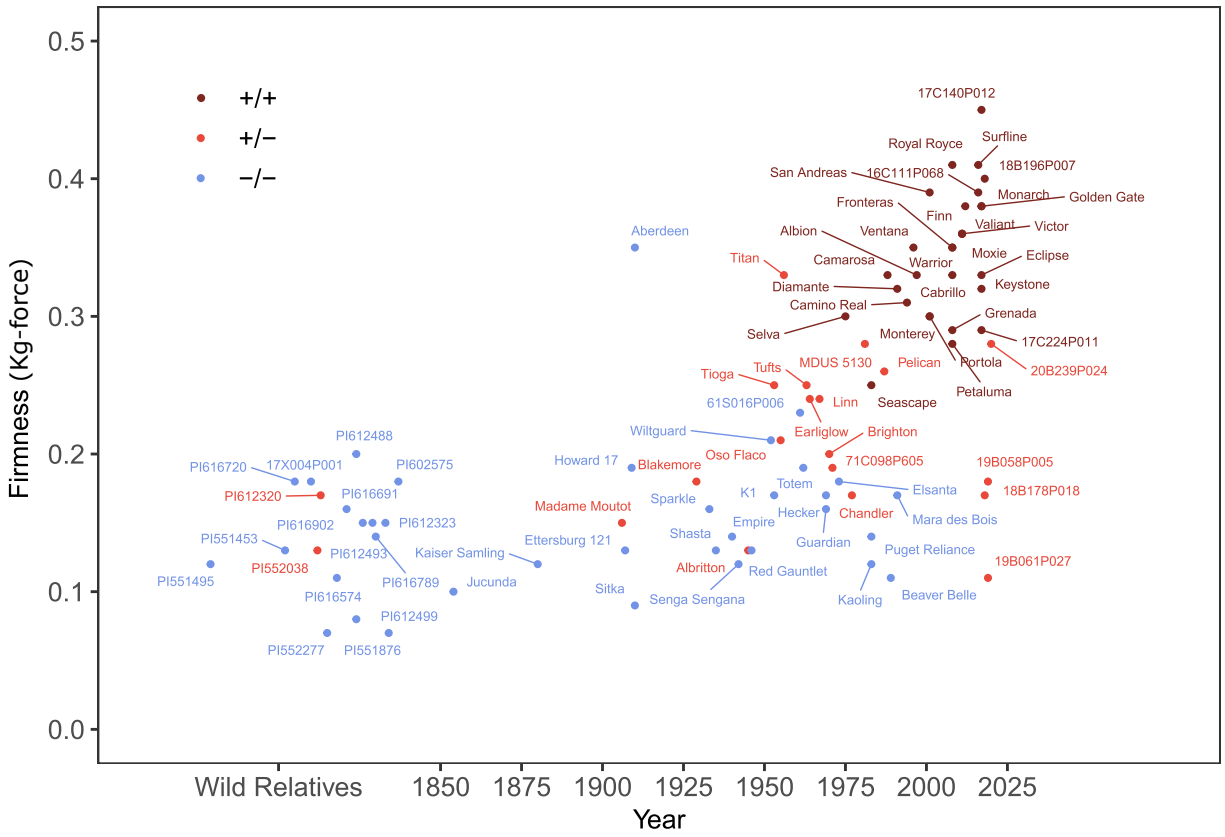

FIGURE 3.13. **Fruit firmness phenotypes of wild relatives, cultivars, and other genetic resources of cultivated strawberry originating 1850 to present.** The birth years of *F.* × *ananassa* cultivars are plotted on the x-axis. The phenotypes of several *F. chiloensis* and *F. virginiana* ecotypes are shown in random order to the left of 1850 on the x-axis. Genotypes of the AX-184242253 SNP were used to predict to *PG1-6A1* unfavorable allele homozygotes (-/-; blue points), heterozygotes (+/-; red points), and favorable allele homozygotes (+/+; brown points), where the favorable (mutant) allele ($PG1\text{-}6A2^{+}$) increases fruit firmness. See Figure 3.15 for a version of this figure showing additional cultivars.

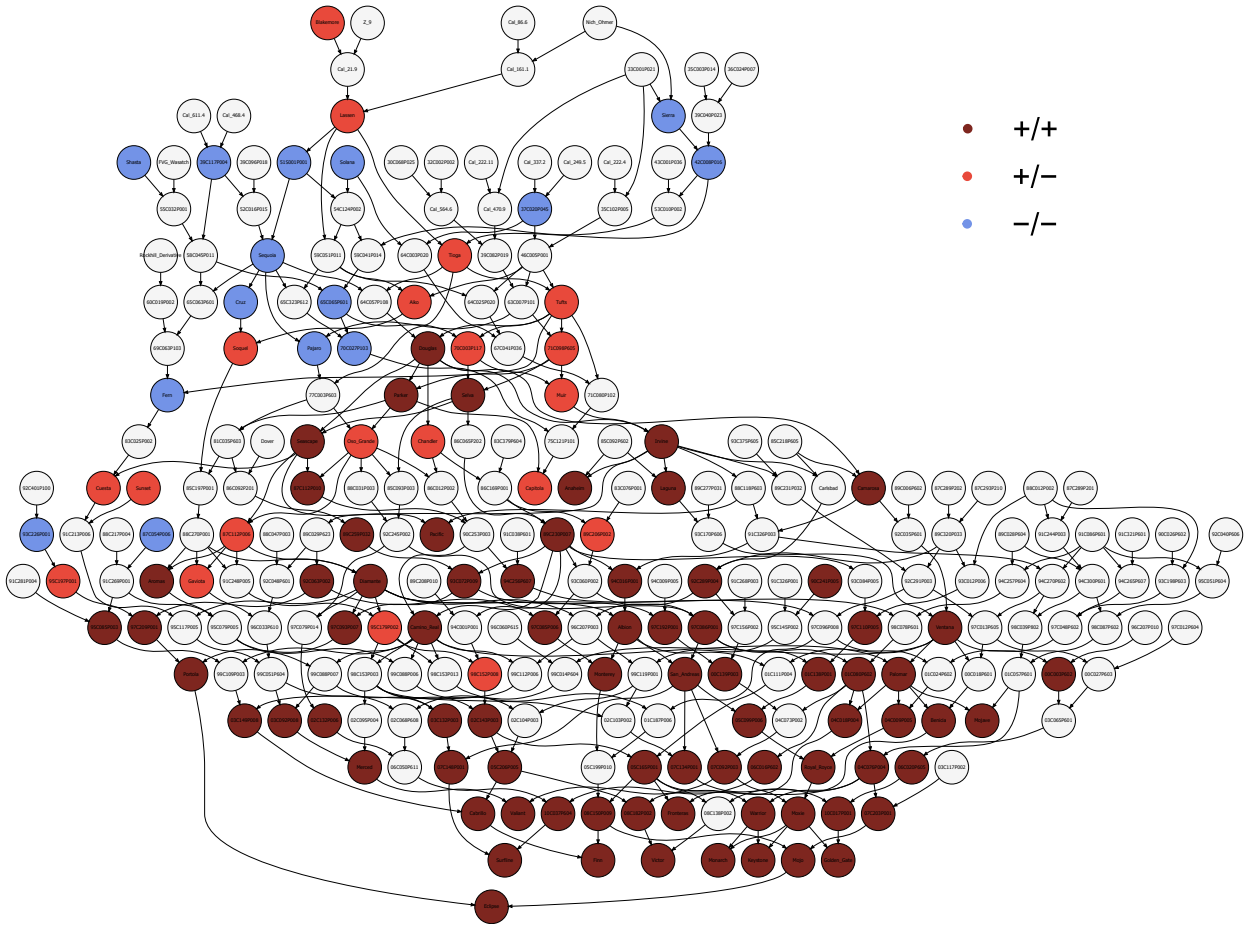

FIGURE 3.14. **Tracing the ancestry of the favorable (mutant) *PG1-6A1* allele found in firm-fruited UC cultivars.** The family tree illustrates a small fraction of the thousands of descendants of 'Tioga' and 'Tufts' in the pedigree records of firm-fruited progeny developed at UC, including every UC cultivar developed since 1970. Genotypes of the AX-184242253 SNP were used to predict *PG1-6A1* unfavorable allele homozygotes (-/-; blue points), heterozygotes (+/-; red points), and favorable allele homozygotes (+/+; brown points), where the favorable (mutant) allele (*PG1-6A2*$^{+}$) increases fruit firmness. Gray nodes identify individuals that were not genotyped or phenotyped, many of which are extinct.

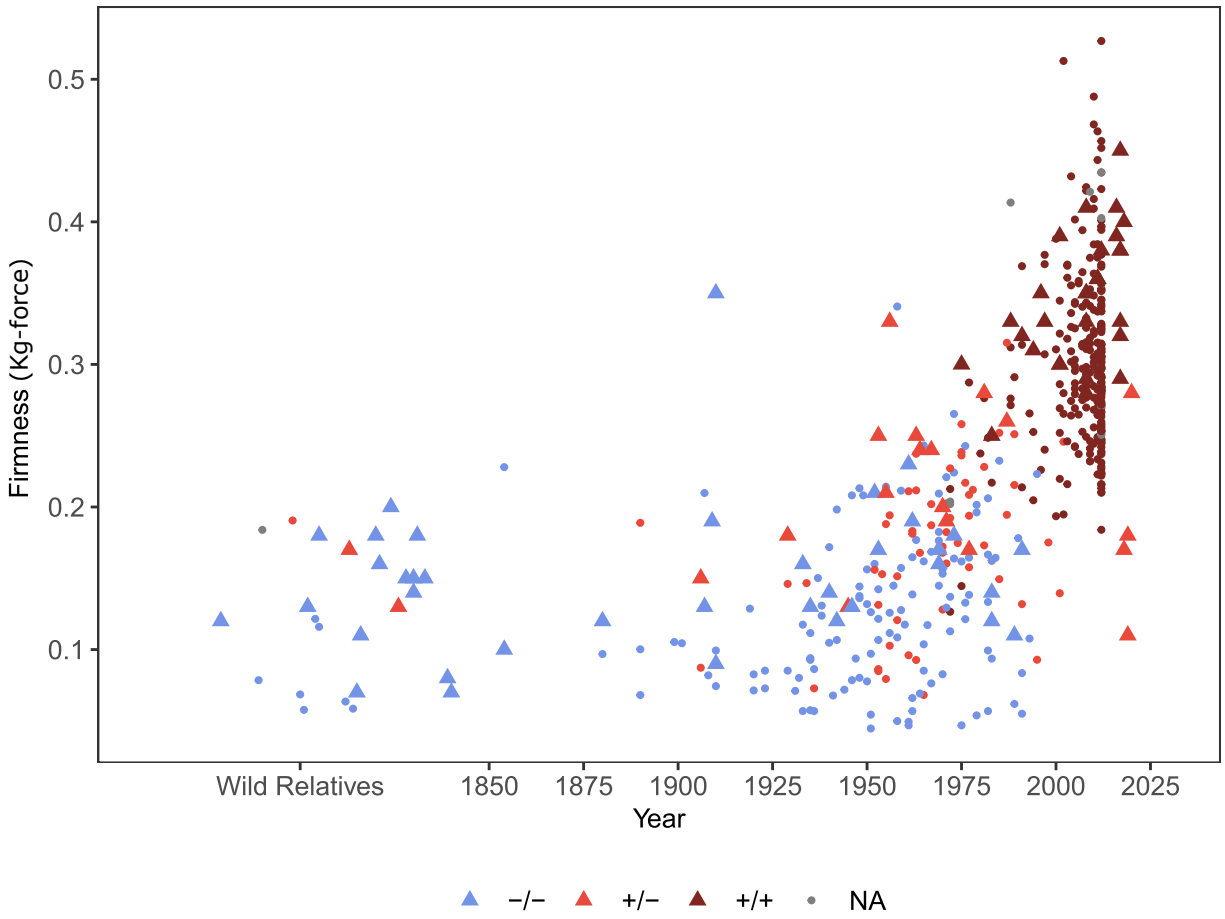

FIGURE 3.15. **An expanded version of Figure 3.13 showing the fruit firmness phenotypes of 500 wild relatives, cultivars, and other genetic resources of cultivated strawberry originating 1850 to present.** The birth years of *F.* × *ananassa* cultivars and other hybrid individuals are plotted on the x-axis. The phenotypes of several *F. chiloensis* and *F. virginiana* ecotypes are shown in random order to the left of 1850 on the x-axis. Genotypes of the AX-184242253 SNP were used to predict to *PG1-6A1* unfavorable allele homozygotes (-/-; blue points), heterozygotes (+/-; red points), and favorable allele homozygotes (+/+; brown points), where the favorable (mutant) allele (*PG1-6A1*$^{+}$) increases fruit

References

Adams, T. M., Armitage, A. D., Sobczyk, M. K., Bates, H. J., Tabima, J. F., Kronmiller, B. A., Tyler, B. M., Grünwald, N. J., Dunwell, J. M., Nellist, C. F., et al. (2020). Genomic investigation of the strawberry pathogen *Phytophthora fragariae* indicates pathogenicity is associated with transcriptional variation in three key races. *Front. Microbiol.*, 11:490. https://doi.org/10.3389/fmicb.2020.00490.

Ahn, S. Y., Kim, S. A., and Yun, H. K. (2019). Differentially expressed genes during berry ripening in de novo rna assembly of vitis flexuosa fruits. *Horticulture, Environment, and Biotechnology*, 60:531–553.

Alarfaj, R., El-Soda, M., Antanaviciute, L., Vickerstaff, R., Hand, P., Harrison, R. J., and Wagstaff, C. (2021). Mapping qtl underlying fruit quality traits in an f1 strawberry population. *J. Hortic. Sci. Biotechnol.*, 96(5):634–645.

Albert, F. W. and Kruglyak, L. (2015). The role of regulatory variation in complex traits and disease. *Nat. Rev. Genet.*, 16(4):197–212.

Albrecht, T., Wimmer, V., Auinger, H.-J., Erbe, M., Knaak, C., Ouzunova, M., Simianer, H., and Schön, C.-C. (2011). Genome-based prediction of testcross values in maize. *Theor. Appl. Genet.*, 123:339.

Ali, A., Abrar, M., Sultan, M., Din, A., Niaz, B., et al. (2011). Post-harvest physicochemical changes in full ripe strawberries during cold storage. *J Anim Plant Sci*, 21(1):38–41.

Allier, A., Moreau, L., Charcosset, A., Teyssèdre, S., and Lehermeier, C. (2019). Usefulness criterion and post-selection parental contributions in multi-parental crosses: application to polygenic trait introgression. *G3: Genes Genomes Genet.*, 9:1469–1479. https://doi.org/10.1534/g3.119.400129.

Allier, A., Teyssèdre, S., Lehermeier, C., Moreau, L., and Charcosset, A. (2020). Optimized breeding strategies to harness genetic resources with different performance levels. *BMC Genom.*, 21:1–16.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.*, 215(3):403–410.

Aulchenko, Y. S., Ripke, S., Isaacs, A., and Van Duijn, C. M. (2007). Genabel: an r library for genome-wide association analysis. *Bioinformatics*, 23(10):1294–1296.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.*, 67(1):1–48. https://doi.org/10.18637/jss.v067.i01.

Bell, J., Simpson, D., and Harris, D. (1997). Development of a method for screening strawberry germplasm for resistance to *Phytophthora cactorum*. In *Acta Hortic.*, volume 439, pages 175–180. https://doi.org/10.17660/ActaHortic.1997.439.23.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc., Ser. B, Methodol.*, 57:289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Bennetzen, J. L. (2000). Transposable element contributions to plant gene and genome evolution. *Plant Mol. Biol.*, 42:251–269.

Bennetzen, J. L. and Wang, H. (2014). The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annu Rev Plant Biol.*, 65:505–530.

Bernardo, R. (2008). Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop Sci.*, 48:1649–1664. https://doi.org/10.2135/cropsci2008.03.0131.

Bernardo, R. (2014). Genomewide selection when major genes are known. *Crop Sci.*, 54(1):68–75. https://doi.org/10.2135/cropsci2013.05.0315.

Bernardo, R. and Thompson, A. M. (2016). Germplasm architecture revealed through chromosomal effects for quantitative traits in maize. *Plant Genome*, 9. https://doi.org/10.3835/plantgenome2016.03.0028.

Bonardi, V. and Dangl, J. L. (2012). How complex are intracellular immune receptor signaling complexes? *Front. Plant Sci.*, 3:237. https://doi.org/10.3389/fpls.2012.00237.

Boutrot, F. and Zipfel, C. (2017). Function, discovery, and exploitation of plant pattern recognition receptors for broad-spectrum disease resistance. *Annu. Rev. Phytopathol.*, 55:257–286.

Brauner, P. C., Müller, D., Molenaar, W. S., and Melchinger, A. E. (2020). Genomic prediction with multiple biparental families. *Theor. Appl. Genet.*, 133(1):133–147.

Brutus, A., Sicilia, F., Macone, A., Cervone, F., and Lorenzo, G. D. (2010). A domain swap approach reveals a role of the plant wall-associated kinase 1 (WAK1) as a receptor of oligogalacturonides. *Proc. Natl. Acad. Sci. U.S.A.*, 107(20):9452–9457. https://doi.org/10.1073/pnas.1000675107.

Castelló, M., Fito, P., and Chiralt, A. (2010). Changes in respiration rate and physical properties of strawberries due to osmotic dehydration and storage. *J. Food Eng.*, 97(1):64–71.

Castillejo, C., Waurich, V., Wagner, H., Ramos, R., Oiza, N., Muñoz, P., Triviño, J. C., Caruana, J., Liu, Z., Cobo, N., et al. (2020). Allelic variation of myb10 is the major force controlling natural variation in skin and flesh color in strawberry (fragaria spp.) fruit. *Plant Cell*, 32(12):3723–3749.

Castro, P. and Lewers, K. S. (2016). Identification of quantitative trait loci (qtl) for fruit-quality traits and number of weeks of flowering in the cultivated strawberry. *Mol. Breed.*, 36(10):138.

Chandra, D., Choi, A. J., Lee, J. S., Lee, J., Kim, J. G., et al. (2015). Changes in physico-chemical and sensory qualities of "Goha" strawberries treated with different conditions of carbon dioxide. *Agric. Sci.*, 6(03):325.

Cingolani, P., Patel, V., Coon, M., Nguyen, T., Land, S., Ruden, D., and Lu, X. (2012a). Using drosophila melanogaster as a model for genotoxic chemical mutational studies with a new program, snpsift. *Front. Genet.*, 3.

Cingolani, P., Platts, A., Coon, M., Nguyen, T., Wang, L., Land, S., Lu, X., and Ruden, D. (2012b). A program for annotating and predicting the effects of single nucleotide polymorphisms, snpeff: Snps in the genome of drosophila melanogaster strain w1118; iso-2; iso-3. *Fly*, 6(2):80–92.

Clark, S. A., Hickey, J. M., Daetwyler, H. D., and van der Werf, J. H. (2012). The importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. *Genet. Sel. Evol.*, 44(1):1–9.

Clough, S. J., Fengler, K. A., ching Yu, I., Lippok, B., Smith, R. K., and Bent, A. F. (2000). The Arabidopsis dnd1 "defense, no death" gene encodes a mutated cyclic nucleotide-gated ion channel. *Proc. Natl. Acad. Sci. U.S.A.*, 97(16):9323–9328. https://doi.org/10.1073/pnas.150005697.

Cockerton, H. M., Karlström, A., Johnson, A. W., Li, B., Stavridou, E., Hopson, K. J., Whitehouse, A. B., and Harrison, R. J. (2021). Genomic informed breeding strategies for strawberry yield and fruit quality traits. *Front. Plant Sci.*, 12.

Collins, J. and Perkins-Veazie, P. (1993). Postharvest changes in strawberry fruit stored under simulated retail display conditions 1. *J. Food Qual.*, 16(2):133–143.

Coman, M. and Popescu, A. (1997). Inheritance of some strawberry quantitative traits.

Cordenunsi, B., Nascimento, J. d., and Lajolo, F. (2003). Physico-chemical changes related to quality of five strawberry fruit cultivars during cool-storage. *Food Chem.*, 83(2):167–173.

Covarrubias-Pazaran, G. (2016). Genome-assisted prediction of quantitative traits using the R package sommer. *PloS One*, 11:e0156744.

Crossa, J., de los Campos, G., Pérez, P., Gianola, D., Burgueño, J., Araus, J. L., Makumbi, D., Singh, R. P., Dreisigacker, S., Yan, J., et al. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics*, 186(2):713–724. https://doi.org/10.1534/genetics.110.118521.

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., Burgueño, J., González-Camacho, J. M., Pérez-Elizalde, S., Beyene, Y., et al. (2017). Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci.*, 22:961–975. https://doi.org/10.1016/j.tplants.2017.08.011.

Daetwyler, H., Kemper, K., Van Der Werf, J., and Hayes, B. J. (2012). Components of the accuracy of genomic prediction in a multi-breed sheep population. *J. Anim. Sci.*, 90:3375–3384. https://doi.org/10.2527/jas.2011-4557.

Daetwyler, H. D., Calus, M. P., Pong-Wong, R., de los Campos, G., and Hickey, J. M. (2013). Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. *Genetics*, 193(2):347–365.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., et al. (2021). Twelve years of samtools and bcftools. *Gigascience*, 10(2):giab008.

Darrow, G. M. et al. (1966). The strawberry. history, breeding and physiology. *The strawberry. History, breeding and physiology.*

de Los Campos, G., Gianola, D., Rosa, G. J., Weigel, K. A., and Crossa, J. (2010). Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. *Genet. Res.*, 92:295–308. https://doi.org/10.1017/S0016672310000285.

de los Campos, G., Hickey, J. M., Pong-Wong, R., Daetwyler, H. D., and Calus, M. P. (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics*, 193(2):327–345. https://doi.org/10.1534/genetics.112.143313.

De Mendiburu, F. and Simon, R. (2015). Agricolae-ten years of an open source statistical tool for experiments in breeding, agriculture and biology. Technical report, PeerJ PrePrints. https://doi.org/10.7287/peerj.preprints.1404v1.

De Roos, A., Hayes, B., and Goddard, M. (2009). Reliability of genomic predictions across multiple populations. *Genetics*, 183(4):1545–1553.

Dekkers, J. (2007). Prediction of response to marker-assisted and genomic selection using selection index theory. *J. Anim. Breed. Genet.*, 124(6):331–341. https://doi.org/10.1111/j.1439-0388.2007.00701.x.

Dekkers, J., Su, H., and Cheng, J. (2021). Predicting the accuracy of genomic predictions. *Genet. Sel. Evol.*, 53:1–23. https://doi.org/10.1186/s12711-021-00647-w.

Denoyes-Rothan, B., Lerceteau-Kohler, E., Guérin, G., Bosseur, S., Bariac, J., Martin, E., and Roudeillac, P. (2004). QTL analysis for resistances to *Colletotrichum acutatum* and *Phytophthora cactorum* in octoploid strawberry (*Fragaria* × *ananassa*). *Acta Hortic.*, 663:147–152. https://doi.org/10.17660/ActaHortic.2004.663.19.

Deutschmann, V. (1954). Eine wurzelfäule an erdbeeren, hervorgerufen durch *Phytophthora cactorum* (leb. et cohn) schröt. *Nachrichtenblatt des Deutschen Pflanzenschutzdienstes*, 6:7–9.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013). Star: ultrafast universal rna-seq aligner. *Bioinformatics*, 29(1):15–21.

Doebley, J. F., Gaut, B. S., and Smith, B. D. (2006). The molecular genetics of crop domestication. *Cell*, 127(7):1309–1321.

Druet, T., Macleod, I., and Hayes, B. (2014). Toward genomic prediction from whole-genome sequence data: impact of sequencing design on genotype imputation and accuracy of predictions. *Heredity*, 112:39–47. https://doi.org/10.1038/hdy.2013.13.

Duniway, J. M. (2002). Status of chemical alternatives to methyl bromide for pre-plant fumigation of soil. *Phytopathology*, 92:1337–1343. https://doi.org/10.1094/PHYTO.2002.92.12.1337.

Edger, P. P., Poorten, T. J., VanBuren, R., Hardigan, M. A., Colle, M., McKain, M. R., Smith, R. D., Teresi, S. J., Nelson, A. D., Wai, C. M., et al. (2019). Origin and evolution of the octoploid strawberry genome. *Nat. Genet.*, 51:541–547. https://doi.org/10.1038/s41588-019-0356-4.

Edger, P. P., VanBuren, R., Colle, M., Poorten, T. J., Wai, C. M., Niederhuth, C. E., Alger, E. I., Ou, S., Acharya, C. B., Wang, J., et al. (2018). Single-molecule sequencing and optical mapping yields an improved genome of woodland strawberry (fragaria vesca) with chromosome-scale contiguity. *Gigascience*, 7(2):gix124.

Eikemo, H., Stensvand, A., Davik, J., and Tronsmo, A. (2003). Resistance to crown rot (*Phytophthora cactorum*) in strawberry cultivars and in offspring from crosses between cultivars differing in susceptibility to the disease. *Ann. Appl. Biol.*, 142(1):83–89. https://doi.org/10.1111/j.1744-7348.2003.tb00232.x.

Eikemo, H., Stensvand, A., and Tronsmo, A. (2000). Evaluation of methods of screening strawberry cultivars for resistance to crown rot caused by *Phytophthora cactorum*. *Ann. Appl. Biol.*, 137:237–244. https://doi.org/10.1111/j.1744-7348.2000.tb00064.x.

Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with r package rrblup. *Plant Genome*, 4(3).

Erwin, D. C. and Ribeiro, O. K. (1996). *Phytophthora diseases worldwide*. The American Phytopathological Society, St. Paul, Minnesota, US.

Falconer, D. and Mackay, T. (1996). *Introduction to Quantitative Genetics*. Harlow, Essex, UK. Longmans Green.

Fan, Z., Tieman, D. M., Knapp, S. J., Zerbe, P., Famula, R., Barbey, C. R., Folta, K. M., Amadeu, R. R., Lee, M., Oh, Y., et al. (2022). A multi-omics framework reveals strawberry flavor genes and their regulatory elements. *New Phytol.*, 236(3):1089–1107.

Feldmann, M. J., Piepho, H.-P., Bridges, W. C., and Knapp, S. J. (2021). Average semivariance yields accurate estimates of the fraction of marker-associated genetic variance and heritability in complex trait analyses. *PLoS Genet.*, 17(8):1–27. https://doi.org/10.1371/journal.pgen.1009762.

Feldmann, M. J., Pincot, D. D., Hardigan, M. A., Seymour, D. K., Famula, R. A., López, C. M., Bjornson, M., Jiménez, N. P., Petrasch, S., Blanco-Ulate, B., Cole, G. S.,

and Knapp, S. J. (2023). Wide hybrids uncover useful heterosis good inbreeding and outbreeding depression in strawberry. *Genome Genetics*, 2023:1–XX.

Feldmann, M. J., Pincot, D. D. A., Cole, G. S., and Knapp, S. J. (2024a). Genetic gains underpinning a little-known strawberry green revolution. *Nat. Commun.*, 15(1):2468.

Feldmann, M. J., Pincot, D. D. A., Hardigan, M. A., Seymour, D. K., Famula, R. A., Jiménez, N. P., Lopez, C. M., Cole, G. S., and Knapp, S. J. (2024b). A dominance hypothesis argument for historical genetic gains, inbreeding-associated genetic erosion, and the disappearance of heterosis in octoploid strawberry. *Genetics*, Submitted(TBD):TBD.

Ferrão, L. F. V., Amadeu, R. R., Benevenuto, J., de Bem Oliveira, I., and Munoz, P. R. (2021). Genomic selection in an outcrossing autotetraploid fruit crop: lessons from blueberry breeding. *Front. Plant Sci.*, 12:676326.

Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. *Nature Reviews Genetics*, 9(5):397–405.

Feschotte, C., Jiang, N., and Wessler, S. R. (2002). Plant transposable elements: where genetics meets genomics. *Nat. Rev. Genet.*, 3(5):329–341.

Fletcher, S. W. (1917). *Strawberry-growing*. Macmillan.

Galli, V., Borowski, J. M., Perin, E. C., da Silva Messias, R., Labonde, J., dos Santos Pereira, I., dos Anjos Silva, S. D., and Rombaldi, C. V. (2015). Validation of reference genes for accurate normalization of gene expression for real time-quantitative pcr in strawberry fruits using different cultivars and osmotic stresses. *Gene*, 554(2):205–214.

George, A. W. and Cavanagh, C. (2015). Genome-wide association mapping in plants. *Theor. Appl. Genet.*, 128(6):1163–1174. https://doi.org/10.1007/s00122-015-2497-x.

Gezan, S. A., Osorio, L. F., Verma, S., and Whitaker, V. M. (2017). An experimental validation of genomic selection in octoploid strawberry. *Hort. Res.*, 4:16070.

Ghoochani, R., Vosough, A., and Karami, F. (2015). Heritability, genetic variability and relationship among morphological and chemical parameters of strawberry cultivars. In

Biological Forum, volume 7, page 218. Research Trend.

Gianola, D. and Van Kaam, J. B. (2008). Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics*, 178:2289–2303. https://doi.org/10.1534/genetics.107.084285.

Gilad, Y., Rifkin, S. A., and Pritchard, J. K. (2008). Revealing the architecture of gene regulation: the promise of eqtl studies. *Trends Genet.*, 24(8):408–415.

Giovanny, C.-P. (2016). Genome assisted prediction of quantitative traits using the r package sommer. *PLoS One*, 11:1–15.

Given, N., Venis, M., and Gierson, D. (1988). Hormonal regulation of ripening in the strawberry, a non-climacteric fruit. *Planta*, 174:402–406.

Goddard, M. and Hayes, B. (2007). Genomic selection. *J. Anim. Breed. Genet.*, 124(6):323–330. https://doi.org/10.1111/j.1439-0388.2007.00702.x.

Goddard, M. E., Hayes, B. J., and Meuwissen, T. H. E. (2010). Genomic selection in livestock populations. *Genet. Res.*, 92:413–421. https://doi.org/10.1017/S0016672310000613.

Gu, T., Jia, S., Huang, X., Wang, L., Fu, W., Huo, G., Gan, L., Ding, J., and Li, Y. (2019). Transcriptome and hormone analyses provide insights into hormonal regulation in strawberry ripening. *Planta*, 250:145–162.

Guo, J., Wang, S., Yu, X., Dong, R., Li, Y., Mei, X., and Shen, Y. (2018). Polyamines regulate strawberry fruit ripening by abscisic acid, auxin, and ethylene. *Plant Physiol.*, 177(1):339–351.

Habier, D., Fernando, R. L., and Dekkers, J. C. (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics*, 177:2389–2397. https://doi.org/10.1534/genetics.107.081190.

Habier, D., Fernando, R. L., and Garrick, D. J. (2013). Genomic BLUP decoded: a look into the black box of genomic prediction. *Genetics*, 194:597–607. https://doi.org/10.1534/genetics.113.152207.

Habier, D., Tetens, J., Seefried, F.-R., Lichtner, P., and Thaller, G. (2010). The impact of genetic relationship information on genomic breeding values in german holstein cattle. *Genet. Sel. Evol.*, 42:1–12.

Hancock, J., Callow, P., Dale, A., Luby, J., Finn, C., Hokanson, S., and Hummer, K. E. (2001a). From the andes to the rockies: Native strawberry collection and utilization. *HortScience*, 36(2):221–225.

Hancock, J., Finn, C., Hokanson, S., Luby, J., Goulart, B., Demchak, K., Callow, P., Serce, S., Schilder, A., and Hummer, K. E. (2001b). A multistate comparison of native octoploid strawberries from north and south america. *J. Am. Soc. Hortic. Sci.*, 126(5):579–586.

Hancock, J. F., Finn, C. E., Luby, J. J., Dale, A., Callow, P. W., and Serçe, S. (2010). Reconstruction of the strawberry, fragaria× ananassa, using genotypes of f. virginiana and f. chiloensis. *HortScience*, 45(7):1006–1013.

Hardigan, M. A., Feldmann, M. J., Lorant, A., Bird, K. A., Famula, R., Acharya, C., Cole, G., Edger, P. P., and Knapp, S. J. (2020). Genome synteny has been conserved among the octoploid progenitors of cultivated strawberry over millions of years of evolution. *Front. Plant Sci.*, 10:1789. https://doi.org/10.3389/fpls.2019.01789.

Hardigan, M. A., Feldmann, M. J., Pincot, D. D., Famula, R. A., Vachev, M. V., Madera, M. A., Zerbe, P., Mars, K., Peluso, P., Rank, D., Ou, S., Saski, C. A., Acharya, C. B., Cole, G. S., Yocca, A. E., Platts, A. E., Edger, P. P., and Knapp, S. J. (2021a). Blueprint for phasing and assembling the genomes of heterozygous polyploids: Application to the octoploid genome of strawberry. *bioRxiv*. https://doi.org/10.1101/2021.11.03.467115.

Hardigan, M. A., Lorant, A., Pincot, D. D. A., Feldmann, M. J., Famula, R. A., Acharya, C. B., Lee, S., Verma, S., Whitaker, V. M., Bassil, N., Zurn, J., Cole, G. S., Bird, K., Edger, P. P., and Knapp, S. J. (2021b). Unraveling the Complex Hybrid Ancestry and Domestication History of Cultivated Strawberry. *Mol. Biol. Evol.*, 38(6):2285–2305.

Hardigan, M. A., Poorten, T. J., Acharya, C. B., Cole, G. S., Hummer, K. E., Bassil, N., Edger, P. P., and Knapp, S. J. (2018). Domestication of temperate and coastal hybrids with distinct ancestral gene selection in octoploid strawberry. *Plant Genome*, 11.

Heffner, E. L., Sorrells, M. E., and Jannink, J.-L. (2009). Genomic selection for crop improvement. *Crop Sci.*, 49(1):1–12. https://doi.org/10.2135/cropsci2008.08.0512.

Hernanz, D., Recamales, Á. F., Meléndez-Martínez, A. J., González-Miret, M. L., and Heredia, F. J. (2008). Multivariate statistical analysis of the color- anthocyanin relationships in different soilless-grown strawberry genotypes. *J. Agric. Food Chem.*, 56(8):2735–2741.

Herter, C. P., Ebmeyer, E., Kollers, S., Korzun, V., Würschum, T., and Miedaner, T. (2019). Accuracy of within-and among-family genomic prediction for fusarium head blight and septoria tritici blotch in winter wheat. *Theor. Appl. Genet.*, 132:1121–1135.

Heslot, N., Yang, H.-P., Sorrells, M. E., and Jannink, J.-L. (2012). Genomic selection in plant breeding: a comparison of models. *Crop Sci.*, 52(1):146–160.

Hill, M. S., Vande Zande, P., and Wittkopp, P. J. (2021). Molecular and evolutionary processes generating variation in gene expression. *Nature Reviews Genetics*, 22(4):203–215.

Hinkelmann, K. and Kempthorne, O. (2007). *Design and analysis of experiments: Introduction to experimental design*. John Wiley & Sons, Hoboken, New Jersey, US.

Hirsch, C. D. and Springer, N. M. (2017). Transposable element influences on gene expression in plants. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1860(1):157–165.

Holloway, G. J., Povey, S. R., and Sibly, R. M. (1990). The effect of new environment on adapted genetic architecture. *Heredity*, 64(3):323–330.

Hölzle, A., Jonietz, C., Törjek, O., Altmann, T., Binder, S., and Forner, J. (2011). A restorer of fertility-like ppr gene is required for 5-end processing of the nad4 mrna in mitochondria of arabidopsis thaliana. *The Plant Journal*, 65(5):737–744.

Hong, J.-P., Ro, N., Lee, H.-Y., Kim, G. W., Kwon, J.-K., Yamamoto, E., and Kang, B.-C. (2020). Genomic selection for prediction of fruit-related traits in pepper (capsicum spp.). *Front. Plant Sci.*, 11:570871.

Huang, M., Liu, X., Zhou, Y., Summers, R. M., and Zhang, Z. (2019). Blink: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience*, 8(2):giy154.

Hummer, K. E., Bassil, N. V., Zurn, J. D., and Amyotte, B. (2023). Phenotypic characterization of a strawberry (fragaria ×ananassa duchesne ex rosier) diversity collection. *Plants people planet*, 5(2):209–224.

Iheshiulor, O. O., Woolliams, J. A., Yu, X., Wellmann, R., and Meuwissen, T. H. (2016). Within-and across-breed genomic prediction using whole-genome sequence and single nucleotide polymorphism panels. *Genet. Sel. Evol.*, 48:1–15. https://doi.org/10.1186/s12711-016-0193-1.

Ivors, K., editor (2015). *PROTOCOL 02-12.1: Production of Phytophthora inoculum on oat grain*. The American Phytopathological Society.

Jarquín, D., Kocak, K., Posadas, L., Hyma, K., Jedlicka, J., Graef, G., and Lorenz, A. (2014). Genotyping by sequencing for genomic prediction in a soybean breeding population. *BMC Genom.*, 15(1):1–10.

Joshi, K., Tiwari, B., Cullen, P. J., and Frias, J. M. (2019). Predicting quality attributes of strawberry packed under modified atmosphere throughout the cold chain. *Food Packag. Shelf Life.*, 21:100354.

Jouki, M. and Khazaei, N. (2014). Effect of low-dose gamma radiation and active equilibrium modified atmosphere packaging on shelf life extension of fresh strawberry fruits. *Food Packag. Shelf Life.*, 1(1):49–55.

Julio-González, L. C., Matas, A. J., and Mercado, J. A. (2018). Characterization of quality traits in transgenic strawberry fruits with genes encoding pectinolytic enzymes down-regulated. *Rev. Colomb. Biotecnol.*, 20(1):42–50.

Jung, C., Nguyen, N. H., and Cheong, J.-J. (2020). Transcriptional regulation of protein phosphatase 2c genes to modulate abscisic acid signaling. *International journal of molecular sciences*, 21(24):9517.

Jung, S., Lee, T., Cheng, C.-H., Buble, K., Zheng, P., Yu, J., Humann, J., Ficklin, S. P., Gasic, K., Scott, K., et al. (2019). 15 years of gdr: New data and functionality in the genome database for rosaceae. *Nucleic Acids Res.*, 47(D1):D1137–D1145.

Jurkowski, G. I., Smith Jr, R. K., Yu, I.-c., Ham, J. H., Sharma, S. B., Klessig, D. F., Fengler, K. A., and Bent, A. F. (2004). Arabidopsis DND2, a second cyclic nucleotide-gated ion channel gene for which mutation causes the "defense, no death" phenotype. *Mol. Plant Microbe Interact.*, 17:511–520. https://doi.org/10.1094/MPMI.2004.17.5.511.

Kaczmarska, E., Gawroński, J., Jabłońska-Ryś, E., Zalewska-Korona, M., Radzki, W., and Sławińska, A. (2016). Hybrid performance and heterosis in strawberry (fragaria× ananassa duchesne), regarding acidity, soluble solids and dry matter content in fruits. *Plant Breed.*, 135(2):232–238.

Kaler, A. S., Purcell, L. C., Beissinger, T., and Gillman, J. D. (2022). Genomic prediction models for traits differing in heritability for soybean, rice, and maize. *BMC Plant Biol.*, 22(1):1–11.

Kliebenstein, D. (2009). Quantitative genomics: analyzing intraspecific variation using global gene expression polymorphisms or eqtls. *Annual review of plant biology*, 60(1):93–114.

Knapp, S. J., Cole, G. S., Pincot, D. D., Lòpez, C. M., Gonzalez-Benitez, O. A., and Famula, R. A. (2023). 'uc eclipse', a summer plant-adapted photoperiod-insensitive strawberry cultivar. *HortScience*, 58(12):1568–1572.

Kohany, O., Gentles, A. J., Hankus, L., and Jurka, J. (2006). Annotation, submission and screening of repetitive elements in repbase: Repbasesubmitter and censor. *BMC Bioinform.*, 7(1):1–7.

Kolde, R. and Kolde, M. R. (2015). Package 'pheatmap'. *R package*, 1(7):790.

Koonin, E. V. (2005). Orthologs, paralogs, and evolutionary genomics. *Annu. Rev. Genet.*, 39(1):309–338.

Krivorot, A. and Dris, R. (2000). Shelf life and quality changes of strawberry cultivars. In *IV International Strawberry Symposium 567*, pages 755–758.

Labroo, M. R., Studer, A. J., and Rutkoski, J. E. (2021). Heterosis and hybrid crop breeding: a multidisciplinary review. *Front. Genet.*, 12:643761. https://doi.org/10.3389/fgene.2021.643761.

Lado, B., Battenfield, S., Guzmán, C., Quincke, M., Singh, R. P., Dreisigacker, S., Peña, R. J., Fritz, A., Silva, P., Poland, J., et al. (2017). Strategies for selecting crosses using genomic prediction in two wheat breeding programs. *Plant Genome*, 10:1–12. https://doi.org/10.3835/plantgenome2016.12.0128.

Lappalainen, T., Sammeth, M., Friedländer, M. R., 't Hoen, P. A., Monlong, J., Rivas, M. A., Gonzalez-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P. G., et al. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*, 501(7468):506–511.

Lawrence, F., Galletta, G., and Scott, D. (1990). Strawberry breeding work of the us department of agriculture. *HortScience*, 25(8):895–896.

Lee, C., Lee, J., and Lee, J. (2022). Relationship of fruit color and anthocyanin content with related gene expression differ in strawberry cultivars during shelf life. *Sci. Hortic.*, 301:111109.

Lee, H.-E., Manivannan, A., Lee, S. Y., Han, K., Yeum, J.-G., Jo, J., Kim, J., Rho, I. R., Lee, Y.-R., Lee, E. S., Kang, B.-C., and Kim, D.-S. (2021). Chromosome level assembly of homozygous inbred line 'wongyo 3115' facilitates the construction of a high-density linkage map and identification of qtls associated with fruit firmness in octoploid strawberry (fragaria × ananassa). *Front. Plant Sci.*, 12.

Lehermeier, C., De los Campos, G., Wimmer, V., and Schön, C.-C. (2017). Genomic variance estimates: With or without disequilibrium covariances? *J. Anim. Breed. Genet.*,

134(3):232–241. https://doi.org/10.1111/jbg.12268.

Lehermeier, C., Krämer, N., Bauer, E., Bauland, C., Camisan, C., Campo, L., Flament, P., Melchinger, A. E., Menz, M., Meyer, N., et al. (2014). Usefulness of multiparental populations of maize (zea mays l.) for genome-based prediction. *Genetics*, 198(1):3–16.

Lenth, R. V. (2021). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.5.5-1.

Lerceteau-Köhler, E., Moing, A., Guérin, G., Renaud, C., Petit, A., Rothan, C., and Denoyes, B. (2012). Genetic dissection of fruit quality traits in the octoploid cultivated strawberry highlights the role of homoeo-qtl in their control. *Theor. Appl. Genet.*, 124(6):1059–1077.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18):3094–3100.

Li, Y., Shi, F., Lin, Z., Robinson, H., Moody, D., Rattey, A., Godoy, J., Mullan, D., Keeble-Gagnere, G., Hayden, M. J., et al. (2022). Benefit of introgression depends on level of genetic trait variation in cereal breeding programmes. *Front. Plant Sci.*, 13:786452.

Liu, T., Li, M., Liu, Z., Ai, X., and Li, Y. (2021). Reannotation of the cultivated strawberry genome and establishment of a strawberry genome database. *Hort. Res.*, 8.

López-Casado, G., Sánchez-Raya, C., Ric-Varas, P. D., Paniagua, C., Blanco-Portales, R., Muñoz-Blanco, J., Pose, S., Matas, A. J., and Mercado, J. A. (2023). Crispr/cas9 editing of the polygalacturonase fapg1 gene improves strawberry fruit firmness. *Hort. Res.*, 10(3):uhad011.

Lorenz, A., Smith, K., and Jannink, J.-L. (2012). Potential and optimization of genomic selection for fusarium head blight resistance in six-row barley. *Crop Sci.*, 52(4):1609–1621.

Lorenz, K. and Cohen, B. A. (2012). Small-and large-effect quantitative trait locus interactions underlie variation in yeast sporulation efficiency. *Genetics*, 192(3):1123–1132.

MacLeod, I. M., Hayes, B. J., and Goddard, M. E. (2014). The effects of demography and long-term selection on the accuracy of genomic prediction with sequence data. *Genetics*, 198:1671–1684. https://doi.org/10.1534/genetics.114.168344.

Madeira, F., Pearce, M., Tivey, A. R., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A., and Lopez, R. (2022). Search and sequence analysis tools services from embl-ebi in 2022. *Nucleic Acids Res.*, 50(W1):W276–W279.

Mangandi, J., Verma, S., Osorio, L., Peres, N. A., van de Weg, E., and Whitaker, V. M. (2017). Pedigree-based analysis in a multiparental population of octoploid strawberry reveals QTL alleles conferring resistance to *Phytophthora cactorum*. *G3: Genes Genomes Genet.*, 7:1707–1719. https://doi.org/10.1534/g3.117.042119.

Maraei, R. W. and Elsawy, K. M. (2017). Chemical quality and nutrient composition of strawberry fruits treated by γ-irradiation. *J. Radiat. Res. Appl. Sci.*, 10(1):80–87.

Marin, M. V., Seijo, T. E., Baggio, J. S., Whitaker, V. M., and Peres, N. A. (2022). Resistance of strawberry cultivars and the effects of plant ontogenesis on *Phytophthora cactorum* and *P. nicotianae* causing crown rot. *Plant Dis.* https://doi.org/10.1094/PDIS-01-22-0203-RE.

Masny, A., Pruski, K., Żurawicz, E., and Mkadry, W. (2016). Breeding value of selected dessert strawberry (fragaria× ananassa duch.) cultivars for ripening time, fruit yield and quality. *Euphytica*, 207:225–243.

Matar, C., Gaucel, S., Gontard, N., Guilbert, S., and Guillard, V. (2018). Predicting shelf life gain of fresh strawberries 'charlotte cv'in modified atmosphere packaging. *Postharvest Biol. Technol.*, 142:28–38.

Mathew, B., Léon, J., and Sillanpää, M. J. (2018). A novel linkage-disequilibrium corrected genomic relationship matrix for snp-heritability estimation and genomic prediction. *Heredity*, 120(4):356–368.

Mathey, M. M., Finn, C. E., Mookerjee, S., Gündüz, K., Hancock, J. F., Iezzoni, A. F., Mahoney, L. L., Davis, T. M., Bassil, N. V., Hummer, K. E., Stewart, P. J., Whitaker,

V. M., Sargent, D. J., Denoyes-Rothan, B., Amaya, I., and van de Weg, W. E. (2013). Large-scale standardized phenotyping of strawberry in rosbreed. *J. Am. Pomol. Soc.*, 67(4):205–217.

McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. U.S.A.*, 36(6):344–355.

McClintock, B. (1956). Controlling elements and the gene. In *Cold Spring Harbor symposia on quantitative biology*, volume 21, pages 197–216. Cold Spring Harbor Laboratory Press.

Medina, C. A., Kaur, H., Ray, I., and Yu, L.-X. (2021). Strategies to increase prediction accuracy in genomic selection of complex traits in alfalfa (medicago sativa l.). *Cells*, 10(12):3372.

Melchinger, A., Utz, H., and Schon, C. (2008). Genetic expectations of quantitative trait loci main and interaction effects obtained with the triple testcross design and their relevance for the analysis of heterosis. *Genetics*, 178(4):2265–2274.

Meredith Jr., W. R. and Bridge, R. R. (1972). Heterosis and gene action in cotton, gossypium hirsutum l.1. *Crop Sci.*, 12(3):cropsci1972.0011183X001200030015x.

Merrick, L. F. and Carter, A. H. (2021). Comparison of genomic selection models for exploring predictive ability of complex traits in breeding programs. *Plant Genome*, 14(3):e20158.

Meuwissen, T., Hayes, B., and Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4):1819–1829. https://doi.org/10.1093/genetics/157.4.1819.

Meuwissen, T., van den Berg, I., and Goddard, M. (2021). On the use of whole-genome sequence data for across-breed genomic prediction and fine-scale mapping of QTL. *Genet. Sel. Evol.*, 53:1–15. https://doi.org/10.1186/s12711-021-00607-4.

Meyer, R. S. and Purugganan, M. D. (2013). Evolution of crop species: genetics of domestication and diversification. *Nature reviews genetics*, 14(12):840–852.

Mishra, P. K., Ram, R. B., and Kumar, N. (2015). Genetic variability, heritability, and genetic advance in strawberry (fragaria× ananassa duch.). *Turk. J. Agric. For.*, 39(3):451–458.

Mohammadi, M., Blake, T. K., Budde, A. D., Chao, S., Hayes, P. M., Horsley, R. D., Obert, D. E., Ullrich, S. E., and Smith, K. P. (2015a). A genome-wide association study of malting quality across eight us barley breeding programs. *Theor. Appl. Genet.*, 128:705–721.

Mohammadi, M., Tiede, T., and Smith, K. P. (2015b). PopVar: A genome-wide procedure for predicting genetic variance and correlated response in biparental breeding populations. *Crop Sci.*, 55(5):2068–2077. https://doi.org/10.2135/cropsci2015.01.0030.

Monnahan, P. J. and Kelly, J. K. (2015). Epistasis is a major determinant of the additive genetic variance in mimulus guttatus. *PLoS Genet.*, 11(5):e1005201.

Muley, A. B. and Singhal, R. S. (2020). Extension of postharvest shelf life of strawberries (fragaria ananassa) using a coating of chitosan-whey protein isolate conjugate. *Food Chem.*, 329:127213.

Müller, D., Technow, F., and Melchinger, A. E. (2015). Shrinkage estimation of the genomic relationship matrix can improve genomic estimated breeding values in the training set. *Theor. Appl. Genet.*, 128:693–703. https://doi.org/10.1007/s00122-015-2464-6.

Munoz, P., Roldan-Guerra, F. J., Verma, S., Ruiz-Velazquez, M., Torreblanca, R., Oiza, N., Castillejo, C., Sevilla, J. F. S., and Amaya, I. (2024). Genome-wide association studies in a diverse strawberry collection unveil loci controlling agronomic and fruit quality traits. *bioRxiv*, pages 2024–03.

Murti, R. H., Kim, H. Y., and Yeoung, Y. R. (2012). Heritability of fruit quality in the progenies of day-neutral and short day hybrid strawberry cultivars. *AGRIVITA, J Agric Sci.*, 34(2):105–114.

Nakagawa, N. and Sakurai, N. (2006). A mutation in at-nmat1a, which encodes a nuclear gene having high similarity to group ii intron maturase, causes impaired splicing of

mitochondrial nad4 transcript and altered carbon metabolism in arabidopsis thaliana. *Plant and cell physiology*, 47(6):772–783.

Natarajan, S., Hossain, M. R., Kim, H.-T., Denison, M. I. J., Ferdous, M. J., Jung, H.-J., Park, J.-I., and Nou, I.-S. (2020). ddrad-seq derived genome-wide snps, high density linkage map and qtls for fruit quality traits in strawberry (*Fragaria x ananassa*). *3 Biotech*, 10(8):353. PMID: 32760641.

Nellist, C. F., Vickerstaff, R. J., Sobczyk, M. K., Marina-Montes, C., Wilson, F. M., Simpson, D. W., Whitehouse, A. B., and Harrison, R. J. (2019). Quantitative trait loci controlling *Phytophthora cactorum* resistance in the cultivated octoploid strawberry (*Fragaria* × *ananassa*). *Hort. Res.*, 6:1–14. https://doi.org/10.1038/s41438-019-0136-4.

Norman, A., Taylor, J., Edwards, J., and Kuchel, H. (2018). Optimising genomic selection in wheat: effect of marker density, population size and population structure on prediction accuracy. *G3: Genes Genomes Genet.*, 8(9):2889–2899.

Nunes, M. C. N., Brecht, J. K., Morais, A. M., and Sargent, S. A. (2006). Physicochemical changes during strawberry development in the field compared with those that occur in harvested fruit during storage. *J. Sci. Food Agric.*, 86(2):180–190.

Osorio, L. F., Gezan, S. A., Verma, S., and Whitaker, V. M. (2021). Independent validation of genomic prediction in strawberry over multiple cycles. *Front. Genet.*, 11:1862.

Pandey, S. P. and Somssich, I. E. (2009). The role of wrky transcription factors in plant immunity. *Plant Physiol.*, 150(4):1648–1655.

Paniagua, C., Blanco-Portales, R., Barceló-Muñoz, M., García-Gago, J. A., Waldron, K. W., Quesada, M. A., Muñoz-Blanco, J., and Mercado, J. A. (2016). Antisense down-regulation of the strawberry β-galactosidase gene faβgal4 increases cell wall galactose levels and reduces fruit softening. *Journal of Experimental Botany*, 67(3):619–631.

Paniagua, C., Ric-Varas, P., García-Gago, J. A., López-Casado, G., Blanco-Portales, R., Muñoz-Blanco, J., Schückel, J., Knox, J. P., Matas, A. J., Quesada, M. A., et al. (2020). Elucidating the role of polygalacturonase genes in strawberry fruit softening.

J. Exp. Bot., 71(22):7103–7117.

Paniagua, C., Sánchez-Raya, C., Blanco-Portales, R., Mercado, J. A., Palomo-Ríos, E., and Posé, S. (2021). Silencing of fapg1, a fruit specific polygalacturonase gene, decreased strawberry fruit fungal decay during postharvest. In *Biology and Life Sciences Forum*, volume 11, page 96. MDPI.

Paniagua, C., Santiago-Doménech, N., Kirby, A. R., Gunning, A. P., Morris, V. J., Quesada, M. A., Matas, A. J., and Mercado, J. A. (2017). Structural changes in cell wall pectins during strawberry fruit development. *Plant Physiol. Biochem.*, 118:55–63.

Parikka, P. (1998). Mansikkalajikkeiden tyvimädänkestävyys testattu. *Puutarha&kauppa*, 2:4–5.

Paulus, A. O. (1990). Fungal diseases of strawberry. *HortScience*, 25:885–889. https://doi.org/10.21273/HORTSCI.25.8.885.

Payne, J. L. and Wagner, A. (2019). The causes of evolvability and their evolution. *Nat. Rev. Genet.*, 20:24–38. https://doi.org/0.1038/s41576-018-0069-z.

Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., Salazar, G. A., Bileschi, M. L., Bork, P., Bridge, A., Colwell, L., et al. (2023). Interpro in 2022. *Nucleic Acids Res.*, 51(D1):D418–D427.

Pelayo, C., Ebeler, S., and Kader, A. (2003). Postharvest life and flavor quality of three strawberry cultivars kept at 5 C in air or air+ 20 kPa CO2. *Postharvest Biol. Technol.*, 27(2):171–183.

Pérez, P. and de los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics*, 198(2):483–495. https://doi.org/10.1534/genetics.114.164442.

Pérez-Jiménez, R., De Cal, A., Melgarejo, P., Cubero, J., Soria, C., Zea-Bonilla, T., and Larena, I. (2012). Resistance of several strawberry cultivars against three different pathogens. *Span. J. Agric. Res.*, 10:502–512. https://doi.org/10.5424/sjar/2012102-345-11.

Petrasch, S., Knapp, S. J., Van Kan, J. A., and Blanco-Ulate, B. (2019). Grey mould of strawberry, a devastating disease caused by the ubiquitous necrotrophic fungal pathogen *Botrytis cinerea*. *Mol Plant. Pathol.*, 20:877–892.

Petrasch, S., Mesquida-Pesci, S. D., Pincot, D. D., Feldmann, M. J., López, C. M., Famula, R., Hardigan, M. A., Cole, G. S., Knapp, S. J., and Blanco-Ulate, B. (2022). Genomic prediction of strawberry resistance to postharvest fruit decay caused by the fungal pathogen botrytis cinerea. *G3: Genes Genomes Genet.*, 12(1):jkab378.

Pigliucci, M. (2008). Is evolvability evolvable? *Nat. Rev. Genet.*, 9:75–82. https://doi.org/10.1038/nrg2278.

Pincot, D. D., Feldmann, M. J., Hardigan, M. A., Vachev, M. V., Henry, P. M., Gordon, T. R., Bjornson, M., Rodriguez, A., Cobo, N., Famula, R. A., et al. (2022). Novel fusarium wilt resistance genes uncovered in natural and cultivated strawberry populations are found on three non-homoeologous chromosomes. *Theor. Appl. Genet.*, 135:2121–2145. https://doi.org/10.1007/s00122-022-04102-2.

Pincot, D. D., Hardigan, M. A., Cole, G. S., Famula, R. A., Henry, P. M., Gordon, T. R., and Knapp, S. J. (2020). Accuracy of genomic selection and long-term genetic gain for resistance to Verticillium wilt in strawberry. *Plant Genome*, 13(3):e20054. https://doi.org/10.1002/tpg2.20054.

Pincot, D. D., Ledda, M., Feldmann, M. J., Hardigan, M. A., Poorten, T. J., Runcie, D. E., Heffelfinger, C., Dellaporta, S. L., Cole, G. S., and Knapp, S. J. (2021). Social network analysis of the genealogy of strawberry: Retracing the wild roots of heirloom and modern cultivars. *G3: Genes Genomes Genet.*, 11:jkab015. https://doi.org/10.1093/g3journal/jkab257.

Pitrat, M. and Risser, G. (1977). Susceptibility of strawberry to *Phytophthora cactorum* in artificial inoculation. In *Annales de l'Amelioration des Plantes*.

Poland, J. and Rutkoski, J. (2016). Advances and challenges in genomic selection for disease resistance. *Annu Rev Phytopathol.*, 54:79–98. https://doi.org/10.1146/annurev

-phyto-080615-100056.

Ponce, E., Núñez-Lillo, G., Bravo, C., Vidal, J., Tapia-Reyes, P., Meneses, C., Pedreschi, R., and Fuentealba, C. (2023). Cell wall disassembly, metabolome and transcriptome analysis in sweet cherry fruit with induced surface pitting. *Postharvest Biology and Technology*, 198:112262.

Poplin, R., Chang, P.-C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., Newburger, D., Dijamco, J., Nguyen, N., Afshar, P. T., et al. (2018). A universal snp and small-indel variant caller using deep neural networks. *Nat. Biotechnol.*, 36(10):983–987.

Pose, S., Kirby, A. R., Paniagua, C., Waldron, K. W., Morris, V. J., Quesada, M. A., and Mercado, J. A. (2015). The nanostructural characterization of strawberry pectins in pectate lyase or polygalacturonase silenced fruits elucidates their role in softening. *Carbohydr. Polym.*, 132:134–145.

Posé, S., Paniagua, C., Cifuentes, M., Blanco-Portales, R., Quesada, M. A., and Mercado, J. A. (2013). Insights into the effects of polygalacturonase fapg1 gene silencing on pectin matrix disassembly, enhanced tissue integrity, and firmness in ripe strawberry fruits. *J. Exp. Bot.*, 64(12):3803–3815.

Prange, R. K. and DeEll, J. R. (1997). Preharvest factors affecting postharvest quality of berry crops. *HortScience*, 32(5):824–830.

Predieri, S., Lippi, N., and Daniele, G. (2021). What can we learn from consumers' perception of strawberry quality? In *Acta Hortic.*, volume 1309 of *ActaHortic.*, pages 987–994. International Society for Horticultural Science (ISHS), Leuven, Belgium. Conference on International Research on Food Security, Natural Resource Management and Rural Development.

Prohaska, A., Rey-Serra, P., Petit, J., Petit, A., Perrotte, J., Rothan, C., and Denoyes, B. (2024). Exploration of a european-centered strawberry diversity panel provides markers and candidate genes for the control of fruit quality traits. *Horticulture Research*, page uhae137.

Quesada, M. A., Blanco-Portales, R., Posé, S., García-Gago, J. A., Jiménez-Bermudez, S., Muñoz-Serrano, A., Caballero, J. L., Pliego-Alfaro, F., Mercado, J. A., and Munoz-Blanco, J. (2009). Antisense down-regulation of the fapg1 gene reveals an unexpected central role for polygalacturonase in strawberry fruit softening. *Plant Physiol.*, 150(2):1022–1032.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramos, P., Parra-Palma, C., Figueroa, C. R., Zuñiga, P. E., Valenzuela-Riffo, F., Gonzalez, J., Gaete-Eastman, C., and Morales-Quintana, L. (2018). Cell wall-related enzymatic activities and transcriptional profiles in four strawberry (fragaria x ananassa) cultivars during fruit development and ripening. *Sci. Hortic.*, 238:325–332.

Raymond, B., Bouwman, A. C., Schrooten, C., Houwing-Duistermaat, J., and Veerkamp, R. F. (2018). Utility of whole-genome sequence data for across-breed genomic prediction. *Genet. Sel. Evol.*, 50:1–12. https://doi.org/10.1186/s12711-018-0396-8.

Redondo-Nevado, J., Moyano, E., Medina-Escobar, N., Caballero, J., and Muñoz-Blanco, J. (2001). A fruit-specific and developmentally regulated endopolygalacturonase gene from strawberry (fragaria x ananassa cv. chandler). *J. Exp. Bot.*, 52(362):1941–1945.

Rey-Serra, P., Mnejja, M., and Monfort, A. (2021). Shape, firmness and fruit quality qtls shared in two non-related strawberry populations. *Plant Sci.*, 311:111010.

Rice, B. and Lipka, A. E. (2019). Evaluation of rr-BLUP genomic selection models that incorporate peak genome-wide association study signals in maize and sorghum. *Plant Genome*, 12(1). https://doi.org/10.3835/plantgenome2018.07.0052.

Riedelsheimer, C., Endelman, J. B., Stange, M., Sorrells, M. E., Jannink, J.-L., and Melchinger, A. E. (2013). Genomic predictability of interconnected biparental maize populations. *Genetics*, 194(2):493–503.

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edger: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*,

26(1):139–140.

Rodríguez-Leal, D., Lemmon, Z. H., Man, J., Bartlett, M. E., and Lippman, Z. B. (2017). Engineering quantitative trait variation for crop improvement by genome editing. *Cell*, 171(2):470–480.

Rose, M. R. (1984). Genetic covariation in drosophila life history: untangling the data. *Am. Nat.*, 123(4):565–569.

Rosskopf, E. N., Chellemi, D. O., Kokalis-Burelle, N., and Church, G. T. (2005). Alternatives to methyl bromide: A Florida perspective. *Plant Health Prog.*, 6:19. https://doi.org/10.1094/PHP-2005-1027-01-RV.

Rueden, C. T., Schindelin, J., Hiner, M. C., DeZonia, B. E., Walter, A. E., Arena, E. T., and Eliceiri, K. W. (2017). ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinform.*, 18:529.

Rutkoski, J. E., Poland, J. A., Singh, R. P., Huerta-Espino, J., Bhavani, S., Barbier, H., Rouse, M. N., Jannink, J.-L., and Sorrells, M. E. (2014). Genomic selection for quantitative adult plant stem rust resistance in wheat. *Plant Genome*, 7:plantgenome2014.02.0006. https://doi.org/10.3835/plantgenome2014.02.0006.

Salentijn, E. M., Aharoni, A., Schaart, J. G., Boone, M. J., and Krens, F. A. (2003). Differential gene expression analysis of strawberry cultivars that differ in fruit-firmness. *Physiologia Plantarum*, 118(4):571–578.

Salzberg, S. L. (2019). Next-generation genome annotation: we still struggle to get it right. *Genome Biol.*, 20(1):92.

Sánchez-Sevilla, J. F., Vallarino, J. G., Osorio, S., Bombarely, A., Posé, D., Merchante, C., Botella, M. A., Amaya, I., and Valpuesta, V. (2017). Gene expression atlas of fruit ripening and transcriptome assembly from rna-seq data in octoploid strawberry (fragaria× ananassa). *Scientific reports*, 7(1):13737.

Schafleitner, S., Bonnet, A., Pedeprat, N., Rocca, D., Chartier, P., and Denoyes, B. (2013). Genetic variation of resistance of the cultivated strawberry to crown rot caused by *Phytophthora cactorum*. *J. Berry Res.*, 3:79–91. https://doi.org/10.3233/JBR-130052.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9:676–682.

Schneider, S. M., Rosskopf, E. N., Leesch, J. G., Chellemi, D. O., Bull, C. T., and Mazzola, M. (2003). United states department of agriculture—agricultural research service research on alternatives to methyl bromide: pre-plant and post-harvest. *Pest management science: formerly Pesticide Science*, 59(6-7):814–826. https://doi.org/10.1002/ps.728.

Schopp, P., Müller, D., Technow, F., and Melchinger, A. E. (2017a). Accuracy of genomic prediction in synthetic populations depending on the number of parents, relatedness, and ancestral linkage disequilibrium. *Genetics*, 205(1):441–454.

Schopp, P., Müller, D., Wientjes, Y. C., and Melchinger, A. E. (2017b). Genomic prediction within and across biparental families: means and variances of prediction accuracy and usefulness of deterministic equations. *G3: Genes Genomes Genet.*, 7(11):3571–3586.

Schulz-Streeck, T., Ogutu, J., Karaman, Z., Knaak, C., and Piepho, H. (2012). Genomic selection using multiple populations. *Crop Sci.*, 52(6):2453–2461.

Seemüller, E. (1977). Resistenzverhalten von erdbeersorten gegen den erreger der rhizomfäule, *Phytophthora cactorum*. *Nachrichtenblatt des Deutschen Pflanzenschutzdienst (Braunschweig)*, 29:124–126.

Segura, V., Vilhjálmsson, B. J., Platt, A., Korte, A., Seren, Ü., Long, Q., and Nordborg, M. (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat. Genet.*, 44(7):825–830. https://doi.org/10.1038/ng.2314.

Semagn, K., Babu, R., Hearne, S., and Olsen, M. (2014). Single nucleotide polymorphism genotyping using kompetitive allele specific pcr (kasp): overview of the technology and its application in crop improvement. *Mol. Breed.*, 33:1–14.

Session, A. M. and Rokhsar, D. S. (2023). Transposon signatures of allopolyploid genome evolution. *Nat. Commun.*, 14(1):3180.

Seybold, H., Trempel, F., Ranf, S., Scheel, D., Romeis, T., and Lee, J. (2014). Ca2+ signalling in plant immune response: from pattern recognition receptors to ca2+ decoding mechanisms. *New Phytol.*, 204(4):782–790.

Sgrò, C. M. and Hoffmann, A. A. (2004). Genetic correlations, tradeoffs and environmental variation. *Heredity*, 93(3):241–248.

Shabalin, A. A. (2012). Matrix eqtl: ultra fast eqtl analysis via large matrix operations. *Bioinformatics*, 28(10):1353–1358.

Shaw, D. V. (1988). Genotypic variation and genotypic correlations for sugars and organic acids of strawberries. *J. Am. Soc. Hortic. Sci.*, 113(5):770–774.

Shaw, D. V., Hansen, J., and Browne, G. T. (2006). Genotypic variation for resistance to *Phytophthora cactorum* in a California strawberry breeding population. *J. Am. Soc. Hortic. Sci.*, 131(5):687–690. https://doi.org/10.21273/JASHS.131.5.687.

Shaw, D. V., Hansen, J., Browne, G. T., and Shaw, S. M. (2008). Components of genetic variation for resistance of strawberry to *Phytophthora cactorum* estimated using segregating seedling populations and their parent genotypes. *Plant Pathol.*, 57(2):210–215. https://doi.org/10.1111/j.1365-3059.2007.01773.x.

Shaw, D. V. and Larson, K. D. (2008). Performance of early-generation and modern strawberry cultivars from the university of california breeding programme in growing systems simulating traditional and modern horticulture. *J. Hortic. Sci. Biotechnol.*, 83(5):648–652.

Shehata, S. A., Abdeldaym, E. A., Ali, M. R., Mohamed, R. M., Bob, R. I., and Abdelgawad, K. F. (2020). Effect of some citrus essential oils on post-harvest shelf life and

physicochemical quality of strawberries during cold storage. *Agronomy*, 10(10):1466.

Shulaev, V., Sargent, D. J., Crowhurst, R. N., Mockler, T. C., Folkerts, O., Delcher, A. L., Jaiswal, P., Mockaitis, K., Liston, A., Mane, S. P., et al. (2011). The genome of woodland strawberry (fragaria vesca). *Nat. Genet.*, 43(2):109–116.

Simko, I. and Piepho, H.-P. (2012). The area under the disease progress stairs: calculation, advantage, and application. *Phytopathology*, 102:381–389. https://doi.org/10.1094/PHYTO-07-11-0216.

Singh, G., Kachwaya, D. S., Kumar, R., Vikas, G., and Singh, L. (2018). Genetic variability and association analysis in strawberry (fragaria x ananassa duch). *Electron. J. Plant Breed.*, 9(1):169–182.

Soares, C. G., do Prado, S. B. R., Andrade, S. C., and Fabi, J. P. (2021). Systems biology applied to the study of papaya fruit ripening: the influence of ethylene on pulp softening. *Cells*, 10(9):2339.

Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redoña, E., Atlin, G., Jannink, J.-L., and McCouch, S. R. (2015). Genomic selection and association mapping in rice (oryza sativa): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLoS Genet.*, 11(2):e1004982.

Staudt, G. (1999). *Systematics and geographic distribution of the American strawberry species: Taxonomic studies in the genus Fragaria (Rosaceae: Potentilleae)*, volume 81. University of California Press.

Stewart-Brown, B. B., Song, Q., Vaughn, J. N., and Li, Z. (2019). Genomic selection for yield and seed composition traits within an applied soybean breeding program. *G3: Genes Genomes Genet.*, 9(7):2253–2265.

Symons, G., Chua, Y.-J., Ross, J., Quittenden, L., Davies, N., and Reid, J. (2012). Hormonal changes during non-climacteric ripening in strawberry. *J. Exp. Bot.*, 63(13):4741–4750.

Tang, H., Bowers, J. E., Wang, X., Ming, R., Alam, M., and Paterson, A. H. (2008). Synteny and collinearity in plant genomes. *Science*, 320(5875):486–488.

Tayeh, N., Klein, A., Le Paslier, M.-C., Jacquin, F., Houtin, H., Rond, C., Chabert-Martinello, M., Magnin-Robert, J.-B., Marget, P., Aubert, G., et al. (2015). Genomic prediction in pea: effect of marker density and training population size and composition on prediction accuracy. *Front. Plant Sci.*, 6:941.

Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P. (2013). Integrative genomics viewer (igv): high-performance genomics data visualization and exploration. *Brief. Bioinform.*, 14(2):178–192.

Tibbs Cortes, L., Zhang, Z., and Yu, J. (2021). Status and prospects of genome-wide association studies in plants. *Plant Genome*, 14(1):e20077. https://doi.org/10.1002/tpg2.20077.

Tiede, T. and Neyhart, J. (2021). *PopVar: Genomic Breeding Tools: Genetic Variance Prediction and Cross-Validation*. R package version 1.3.0.

Ukalska, J., Madry, W., Ukalski, K., Masny, A., and Zurawicz, E. (2006). Patterns of variation and correlation among traits in a strawberry germplasm collection (Fragaria × ananassa duch.). *J. Fruit Ornam. Plant Res.*, 14:5.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., and Rozen, S. G. (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Res.*, 40(15):e115–e115.

Van de Weg, W. (1997). A gene-for-gene model to explain interactions between cultivars of strawberry and races of *Phytophthora fragariae* var. *fragariae*. *Theor. Appl. Genet.*, 94:445–451. https://doi.org/10.1007/s001220050435.

Van den Berg, I., Meuwissen, T., MacLeod, I., and Goddard, M. (2019). Predicting the effect of reference population on the accuracy of within, across, and multibreed genomic prediction. *J. Dairy Sci.*, 102:3155–3174. https://doi.org/10.3168/jds.2018-15231.

van Rijbroek, P., Meulenbroek, E., and van de Lindeloof, C. (1997). Development of a screening method for resistance to *Phytophthora cactorum*. *Acta Hortic.*, 439(1):181–184. https://doi.org/10.1002/tpg2.20077.

Verdecchia, E., Ceustermans, A., Baets, D., Ferreira, J., Bonants, P., Melis, P., Van Hemelrijck, W., Bylemans, D., Rediers, H., and Lievens, B. (2021). Quantitative PCR for detection and quantification of *Phytophthora cactorum* in the cultivation of strawberry. *Eur. J. Plant Pathol.*, 160:867–882. https://doi.org/10.1007/s10658-021-02290-z.

Verma, S., Zurn, J. D., Salinas, N., Mathey, M. M., Denoyes, B., Hancock, J. F., Finn, C. E., Bassil, N. V., and Whitaker, V. M. (2017). Clarifying sub-genomic positions of QTLs for flowering habit and fruit quality in U.S. strawberry (Fragaria×ananassa) breeding populations using pedigree-based QTL analysis. *Hort. Res.*, 4:17062.

Villarreal, N. M., Martínez, G. A., and Civello, P. M. (2009). Influence of plant growth regulators on polygalacturonase expression in strawberry fruit. *Plant Sci.*, 176(6):749–757.

Villarreal, N. M., Rosli, H. G., Martínez, G. A., and Civello, P. M. (2008). Polygalacturonase activity and expression of related genes during ripening of strawberry cultivars with contrasting fruit firmness. *Postharvest Biol. Technol.*, 47(2):141–150.

Visscher, P. M. (1998). On the sampling variance of intraclass correlations and genetic correlations. *Genetics*, 149(3):1605–1614.

Visscher, P. M., Hill, W. G., and Wray, N. R. (2008). Heritability in the genomics era—concepts and misconceptions. *Nat. Rev. Genet.*, 9(4):255–266.

Vu, K., Hollingsworth, R., Leroux, E., Salmieri, S., and Lacroix, M. (2011). Development of edible bioactive coating based on modified chitosan for increasing the shelf life of strawberries. *Food Res. Int.*, 44(1):198–203.

Walsh, B. (2001). Quantitative genetics in the age of genomics. *Theor Popul Biol.*, 59(3):175–184. https://doi.org/10.1006/tpbi.2001.1512.

Wang, J. and Zhang, Z. (2021). Gapit version 3: Boosting power and accuracy for genomic association and prediction. *Genomics, Proteomics Bioinformatics*, 19(4):629–640. Bioinformatics Commons.

Wardlaw, C. (1927). The strawberry disease in Lanarkshire. *Ann. Appl. Biol.*, 14:197–201. https://doi.org/10.1111/j.1744-7348.1927.tb07075.x.

Werner, C. R., Gaynor, R. C., Gorjanc, G., Hickey, J. M., Kox, T., Abbadi, A., Leckband, G., Snowdon, R. J., and Stahl, A. (2020). How population structure impacts genomic selection accuracy in cross-validation: implications for practical breeding. *Front. Plant Sci.*, 11:592977.

Wessler, S. R. (1988). Phenotypic diversity mediated by the maize transposable elements ac and spm. *Science*, 242(4877):399–405.

Whitaker, V. M., Osorio, L. F., Hasing, T., and Gezan, S. (2012). Estimation of genetic parameters for 12 fruit and vegetative traits in the university of florida strawberry breeding population. *J. Am. Soc. Hortic. Sci.*, 137(5):316–324.

Whitlock, M. C. and Fowler, K. (1999). The changes in genetic and environmental variance with inbreeding in drosophila melanogaster. *Genetics*, 152(1):345–353.

Wilcox, W. (1989). First report of crown rot (*Phytophthora cactorum*) of strawberry in eastern North America. *Plant Dis.*, 73:183. https://doi.org/10.1094/PD-73-0183E.

Wilhelm, S. and Paulus, A. O. (1980). How soil fumigation benefits the california strawberry industry. *Plant Dis.*, 64:264–270.

Wilhelm, S., Storkan, R. C., and Sagen, J. E. (1961). Verticillium wilt of strawberry controlled by fumigation of soil with chloropicrin and chloropicrin-methyl bromide mixtures. *Phytopathology*, 51:744–748.

Windhausen, V. S., Atlin, G. N., Hickey, J. M., Crossa, J., Jannink, J.-L., Sorrells, M. E., Raman, B., Cairns, J. E., Tarekegne, A., Semagn, K., et al. (2012). Effectiveness of genomic prediction of maize hybrid performance in different breeding populations and environments. *G3: Genes Genomes Genet.*, 2(11):1427–1436.

Xiong, J., Liu, Y., Wu, P., Bian, Z., Li, B., Zhang, Y., and Zhu, B. (2024). Identification and virus-induced gene silencing (vigs) analysis of methyltransferase affecting tomato (solanum lycopersicum) fruit ripening. *Planta*, 259(5):109.

Yagi, K., Williams, J., Wang, N., and Cicerone, R. (1993). Agricultural soil fumigation as a source of atmospheric methyl bromide. *Proc. Natl. Acad. Sci. U.S.A.*, 90:8420–8423. https://doi.org/10.1073/pnas.90.18.842.

Zareei, E., Karami, F., Aryal, R., and Saed-Moucheshi, A. (2022). Genotypic by phenotypic interaction affects the heritability and relationship among quantity and quality traits of strawberry (fragaria× ananassa). *N. Z. J. Crop Hortic. Sci.*, pages 1–20.

Zhang, A., Wang, H., Beyene, Y., Semagn, K., Liu, Y., Cao, S., Cui, Z., Ruan, Y., Burgueño, J., San Vicente, F., et al. (2017). Effect of trait heritability, training population size and marker density on genomic prediction accuracy estimation in 22 bi-parental tropical maize populations. *Front. Plant Sci.*, 8:1916.

Zhang, F.-T., Zhu, Z.-H., Tong, X.-R., Zhu, Z.-X., Qi, T., and Zhu, J. (2015). Mixed linear model approaches of association mapping for complex traits based on omics variants. *Sci. Rep.*, 5(1):10298.

Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., Bradbury, P. J., Yu, J., Arnett, D. K., Ordovas, J. M., and Buckler, E. S. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.*, 42(4):355–360. https://doi.org/10.1038/ng.546.

Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. (2000). A greedy algorithm for aligning dna sequences. *J. Comput. Biol.*, 7(1-2):203–214.

Zhou, X. and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.*, 44:821. https://doi.org/10.1038/ng.2310.

Zhou, X. and Stephens, M. (2014). Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat. Methods*, 11(4):407–409.

Zhou, X. E., Soon, F.-F., Ng, L.-M., Kovach, A., Suino-Powell, K. M., Li, J., Yong, E.-L., Zhu, J.-K., Xu, H. E., and Melcher, K. (2012). Catalytic mechanism and kinase interactions of aba-signaling pp2c phosphatases. *Plant signaling & behavior*, 7(5):581–588.

Zorrilla-Fontanesi, Y., Cabeza, A., Domínguez, P., Medina, J. J., Valpuesta, V., Denoyes-Rothan, B., Sánchez-Sevilla, J. F., and Amaya, I. (2011). Quantitative trait loci and underlying candidate genes controlling agronomical and fruit quality traits in octoploid strawberry (*Fragaria × ananassa*). *Theor. Appl. Genet.*, 123(5):755–778.

Zsögön, A., Čermák, T., Naves, E. R., Notini, M. M., Edel, K. H., Weinl, S., Freschi, L., Voytas, D. F., Kudla, J., and Peres, L. E. P. (2018). De novo domestication of wild tomato using genome editing. *Nature biotechnology*, 36(12):1211–1216.

Relevant Publications

Jiménez, N. P., Feldmann, M. J., Famula, R. A., Pincot, D. D., Bjornson, M., Cole, G. S., & Knapp, S. J. (2023). Harnessing underutilized gene bank diversity and genomic prediction of cross usefulness to enhance resistance to *Phytophthora cactorum* in strawberry. *Plant Genome*, 16(1), e20275.

Feldmann, M. J., Pincot, D. D. A., Hardigan, M. A., Seymour, D. K., Famula, R. A., Jiménez, N. P., Lopez, C. M., Cole, G. S., and Knapp, S. J. (2024). A dominance hypothesis argument for historical genetic gains, inbreeding-associated genetic erosion, and the disappearance of heterosis in octoploid strawberry. *Genetics* (*manuscript in review*).

Jiménez, N. P., Bjornson, M., Famula, R. A., Pincot, D. D. A., Hardigan, M. A., Madera, M. A., Lopez, C. M., Cole, G. S., Feldmann, M. J., and Knapp, S. J. (2024). Loss-of-Function Mutations in the Fruit Softening Gene *POLYGALACTURONASE1* Doubled Fruit Firmness in Strawberry. *Hortic. Res.* (*manuscript in preparation*).