

UC Davis

UC Davis Electronic Theses and Dissertations

Title

Developing Potential Biomarkers for Non-Infectious Bud Failure in Almond (P. Dulcis) And Methods on the Genetic Improvement of Zinnias (Z. Elegans)

Permalink

<https://escholarship.org/uc/item/3rg4z7jf>

ISBN

9798193264538

Author

Hazzard, William R

Publication Date

2026-09-08

Peer reviewed|Thesis/dissertation

Developing Potential Biomarkers for Non-Infectious Bud Failure in Almond (*P. Dulcis*)

And

Methods on the Genetic Improvement of Zinnias (*Z. Elegans*)

By

WILLIAM R. HAZZARD
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:

Thomas Gradziel, Chair

Patrick J. Brwon

Luca Comai

Committee in Charge

2026

Acknowledgements and Funding

Chapter 1: This project was funded by grant #063021 through the California Department of Agriculture (CDFA) Fruit Tree, Tree Nut and Grapevine Improvement Advisory Board (IAB), and Project HORT35 from the Almond Board of California. Funding for this project was also provided by the University of California, Davis Department of Plant Sciences Jastro-Shields Graduate Research Award. Contributing collaborators were Jonathan Fresnedo-Ramirez, Elizabeth Anderson, Thomas Gradziel and Patrick J. Brown. WRH, JFR and TMG were responsible for conceptual contributions to this paper. WRH, JFR and ESA were responsible for practical and methodological contributions to this paper. WRH and PJB were responsible for empirical contributions to this paper. Manuscript was prepared by WRH and PJB. Special thanks to UC Agriculture and Natural Resources Almond Variety trials for providing plant material for this experiment.

Chapter 2: This project was funded by grant #24-0831-000-SA through the California Department of Agriculture (CDFA) Fruit Tree, Tree Nut and Grapevine Improvement Advisory Board (IAB). Funding for this project was also provided by the University of California, Davis Department of Plant Sciences Jastro-Shields Graduate Research Award. Contributing collaborators were Matthew Davis, Chaehee Lee, John Grey Monroe, Pedro J. Martínez-García, Thomas Gradziel and Patrick J. Brown. WRH, MWD, CL, PJMG and TMG were responsible for the practical and methodological contributions to this paper. WRH, MWD and CL were responsible for the empirical contributions to this paper. Manuscript was prepared by WRH,

MWD, CL, PJMG and PJB. Special thanks to Sierra Gold Nurseries for contributing plant material and technical advice for this project.

Chapter 3: This project was funded by USDA-NIFA Organic Research and Extension Initiative #2023-51300-40963- Extending the Scope of SCOPE: Sharing tools and expanding partnerships for organic cultivar development and USDA-NIFA Organic Research and Extension Initiative #2020-51300-32275-SCOPE 2.0: Refining organic breeding pipelines to produce improved varieties and workforce. Contributing collaborators were Laura Roser, Antonia Palkovic, Charlie Brummer, Allen Van Deynze and Cree King. WRH, LCR and AP were responsible for practical and methodological contributions to this paper. WRH, LCR, AP and CK were responsible for empirical contributions to this paper. Manuscript was prepared by WRH, LCR, AP and ECB. Interns from the Student Collaborative Organic Plant Breeding and Education program contributed significant labor and intellectual contributions to this project.

This dissertation is dedicated to my wife, Ada Waleder, without who's love and support, none of this work would be possible.

Dissertation Abstract

Many cultivars of almond (*Prunus dulcis* Mill. D. A. Webb) suffer from a disorder known as non-infectious bud failure in which preformed meristematic tissue becomes necrotic during dormancy, then fails to emerge in the following growing season. Non-infectious bud failure is a noteworthy disorder in perennial crops due to its mitotic inheritance through asexual propagation and variance in severity between genetic clones. While the cause of the disorder is unknown, previous research has shown possible connections between DNA methylation and symptoms of the disorder. Non-infectious bud failure has been mitigated through propagation from basal epicormic buds, which show reduced non-infectious bud failure in clonal progeny relative to the axillary canopy buds typically used for propagation. In this dissertation, we explored several possibilities of detecting differences in DNA methylation and expression to develop potential biomarkers for non-infectious bud failure in almonds.

First, we explored variation in DNA methylation between epicormic and axillary shoots that may provide insight into the spread and origins of non-infectious bud failure. We hypothesized that shoots derived from different tissues (epicormic and axillary buds) and growing environments would have distinct genome-wide patterns of DNA methylation. We sampled leaf tissue from epicormic and axillary canopy shoots from two almond varieties at four locations, then used whole-genome enzymatic methyl-sequencing to determine global methylation levels and differentially methylated regions between tissues. We found no significant differences in genome-wide methylation levels between different shoot types, locations, and varieties. Epicormic and axillary shoots displayed 23 differentially-methylated regions in one variety and 45 in the other, with no overlap in these regions between varieties. In addition, we identified

4546 regions that appeared differentially methylated between varieties; however, DNA polymorphism may account for some of these differences. In pairwise comparisons between locations, we identified 312 regions, with overlapping regions between some, but not all, locations. The remarkable stability of DNA methylation patterns across tissues and locations motivates further investigation into possible perturbations of these patterns by non-infectious bud failure and related genetic disorders in perennial plants.

Next, we explored differences in DNA methylation and RNA expression between trees displaying different severities of non-infectious bud failure. To do so, we developed a new haplotype-phased reference assembly for the variety “Carmel”. One parent of Carmel, “Texas” has no history of bud failure symptoms, while its other parent, “Nonpareil” does. We then scored Carmel trees in a commercial orchard for non-infectious bud failure severity, then sampled leaf tissue from those trees. We used enzymatic methyl-seq to measure whole-genome DNA methylation and perform differential methylation analysis and we used 3' TAG RNA sequencing to measure gene expression between the levels. Our results show 20 regions of differential methylation between the two haplotypes. Additionally, 2083 genes were differentially expressed between the two haplotypes, with the expression of certain genes highly correlated with bud failure level. Three of these DMRs fall within 5000 base pairs of the differentially expressed genes. These areas of variation in methylation and expression data may indicate candidate genes for non-infectious bud failure, and may aid in the development of future diagnostic tools.

Finally, we describe a successful program for zinnia breeding. Zinnias (*Z. Elegans*) have emerged as an important crop for many California organic growers. Due to their cold intolerance,

zinnias cannot be refrigerated for extended periods of time. Therefore, production cannot be easily moved to locations that require long transportation times, which has resulted in localized production. However, most zinnia seed currently in production has been bred for conventional systems. Additionally, very few resources exist on the methods and strategies for improving zinnias for cut flower production. After evaluating available germplasm, we made 463 crosses, then selected 85 to evaluate in a single field season. Using a modified synthetic breeding strategy, we performed selection on color, petal doubling, stem length and flower size for five years. We introduce two new cultivars, ZP.23.225 and ZC.23.227, that have been bred specifically for organic systems and addressing new and emerging market demands. ZP.23.225 is a zinnia with a light pink color and a small flower head but maintains agronomic vigor. ZC.23.227 is an orange “cactus” type zinnia with good field performance and an exceptionally large flower head. These varieties were developed through the Student Organic Plant Breeding and Education program at the University of California, Davis.

Chapter 1: DNA methylation is conserved between vegetative propagules in almond (*Prunus dulcis* Mill D.A. Webb)

Abstract

Many cultivars of almond (*Prunus dulcis* Mill. D. A. Webb) suffer from a disorder known as non-infectious bud failure in which preformed meristematic tissue becomes necrotic during dormancy, then fails to emerge in the following growing season. While the cause of the disorder is unknown, previous research has shown possible connections between DNA methylation and symptoms of the disorder. Additionally, non-infectious bud failure has been mitigated through propagation from basal epicormic buds, which show reduced non-infectious bud failure in clonal progeny relative to the axillary canopy buds typically used for propagation. In this study, we explored variation in DNA methylation between epicormic and axillary shoots that may provide insight into the spread and origins of non-infectious bud failure. We hypothesized that shoots derived from different tissues (epicormic and axillary buds) and growing environments would have distinct genome-wide patterns of DNA methylation. We sampled leaf tissue from epicormic and axillary canopy shoots from two almond varieties at four locations, then used whole-genome enzymatic methyl-sequencing to determine global methylation levels and differentially methylated regions between tissues. We found no significant differences in genome-wide methylation levels between different shoot types, locations, and varieties. Epicormic and axillary shoots displayed 23 differentially-methylated regions in one variety and 45 in the other, with no

overlap in these regions between varieties. In addition, we identified 4546 regions that appeared differentially methylated between varieties; however, DNA polymorphism may account for some of these differences. In pairwise comparisons between locations, we identified 312 regions, with overlapping regions between some, but not all, locations. The remarkable stability of DNA methylation patterns across tissues and locations motivates further investigation into possible perturbations of these patterns by non-infectious bud failure and related genetic disorders in perennial plants.

Significance of the Study

1. What is known about this subject?

Non-infectious bud failure (NBF) of almond is an economically significant production problem of unknown cause. Epicormic propagation has been shown to reduce NBF, and previous studies demonstrated a link between DNA methylation and NBF.

2. What are the new findings?

Global levels of DNA methylation are similar across tissues, locations, and varieties. We identified a small number of differentially methylated regions between epicormic and axillary shoots in each of the two varieties tested, but there was no overlap in these regions between varieties.

3. What is the expected impact on horticulture?

There is still no clear path towards developing a bioassay for NBF in almond. However, genotype-specific differentially methylated sites between epicormic and canopy tissue may be used by propagation specialists to ensure that their propagation material is from epicormic buds, thereby reducing the potential spread of NBF.

Introduction

Heritable epigenetic disorders are an emerging issue in clonally propagated crops. One of these disorders is non-infectious bud failure (NBF) in almond (*Prunus dulcis* [Mill] D.A. Webb), which can be identified by the transition of healthy preformed buds into necrotic tissue that fails to emerge as shoots during bud break in the spring (Kester 2004). After a bud fails, other buds along the same branch will push, resulting in the characteristic “witch’s broom” or “crazy top” pattern. NBF was first described in the variety “Jordanolo” in the 1930s. In 1964 the disorder was reported in “Nonpareil” (Micke 1996), which currently comprises 40% of California almond acreage (USDA 2020). “Carmel” and other varieties developed using Nonpareil as a parent also subsequently developed NBF symptoms.

NBF is not associated with any viral, fungal, or bacterial pathogen, and fails to meet Koch’s postulates (Fenton 1988). Instead, NBF is primarily spread through asexual propagation. While the exact cause of NBF remains uncertain, the inheritance of the disorder can be modeled by the level of bud-failure potential accumulated in an individual propagule. Once a threshold of bud-failure potential has been surpassed from repeated generations of propagation, symptoms of the disorder will manifest. The severity and percentage of affected scaffolds are proportional to the level of hypothetical bud-failure potential accumulated. Additionally, the onset of symptoms is expedited by high mean summer temperatures, resulting in environmental variance in onset and severity of symptoms for trees propagated from the same nursery (Kester 1978). Propensity for NBF is also passed on through sexual reproduction, where the bud failure potential of one or

both parents is inherited by the progeny, sometimes resulting in the abandonment of varieties during their development (Gradziel 2017).

The severity of bud failure has a proportional effect on yield, making NBF an existential threat to almond production. The appearance of bud failure in some varieties has led to a drastic reduction in their production acreage. However, the effects of NBF are largely mitigated through the unique horticultural practice of propagation using basal epicormic branches from “relic” trees. Epicormic shoots are adventitious shoots arising from latent trace meristems (Meier 2010). These trace meristems serve as systems for regeneration and rejuvenation in woody plants, lying dormant in parenchyma cells below the bark layer (Lanner 2002) until a wounding event, when auxin signaling results in the elongation of the bud primordia (Burrows 2003), thus resulting in shoot growth. Compared to other meristematic tissue or neoformed growth, epicormic shoots have a distinct branching pattern and less bud differentiation than canopy axillary growth (Negron 2014) and are epigenetically younger than canopy buds, having gone through fewer cell divisions. Despite the importance of epicormic shoots in preserving key almond varieties, it remains unclear why their use reduces NBF symptoms.

Due to its asexual inheritance and correlation with specific tissue types and environments, NBF is hypothesized to be an epigenetic disorder, or phenotypic changes resulting from a genetic modification other than the underlying genetic sequence. To explore this hypothesis, DNA methylation has been examined for its possible role in the variance of NBF severity and expression between different varieties and clones (Fresnedo-Ramirez et al. 2017). Differentially methylated regions (DMRs) are areas with unequal levels of cytosine methylation between

groups (Chen 2016), which can be identified using either predefined or undefined regions of the genome (Zhang 2014). Identification of DMRs associated with NBF could potentially allow for the development of predictive markers for the disorder (D’Amico-Willman 2022b). Evidence from other plant species suggests a relationship between differences in plant development, epigenetic state, and DNA methylation. For instance, juvenile shoots have reduced levels of DNA methylation in chestnut (*Castanea dentata*; Valledor 2007), while in apple, the origin of grafted material can affect the epiallelic state of the clonal offspring (Fortes 2017). Whole methylomes can even be inherited through grafting (Perrin 2020). However, methylome differences between clones, tissues, and environments have been little explored in almond or other woody perennial crops.

We hypothesized that almond shoots derived from epicormic and axillary buds possess distinct DNA methylation patterns that can be quantified and deployed in bioassays for prediction across a range of almond varieties. By exploring variation in DNA methylation between tissue types, locations, and varieties, we aimed to elucidate the possible connections between epicormic buds, DNA methylation, and NBF.

Methods

Tissue Sampling

Samples were collected from the University of California Agriculture and Resources Regional Almond Variety Trials (RAVT). Two varieties within the trial were selected for testing: UCD 8-160 and UCD 8-201. These varieties were selected due to their known pedigree relationships with both Nonpareil and Jordano, two cultivars with a history of NBF. The RAVT trials were

established in 2014 in Chico, Salida, Chowchilla, and Billings, California, and samples were collected in Spring of 2022. Trees were arranged in alternating rows of experimental varieties and Nonpareil, with 11 trees per row and four replicated rows of each experimental genotype per trial.

Leaves were sampled from current season growth originating from two locations: the upper canopy and epicormic growth on the basal trunk of the tree above the graft line. From each of the replicated rows, one tree was selected at random. Epicormic shoots were identified and selected based on the lateral position of the emerging shoot relative to a central bud location on the basal area of the trunk and the presence of some prior wounding event relative to the location of the shoot. Multiple leaves were collected and pooled for both upper canopy and epicormic samples within each replicated row ($n=4$) of each genotype ($n=2$) at each location ($n=4$), for a total of 64 samples (Figure 1). Samples were flash frozen in liquid nitrogen, then stored in paper envelopes on dry ice during transport and moved to a -80°C freezer within 12 hours of sampling.

DNA Extraction

Leaf samples were frozen in liquid nitrogen, then ground into a fine powder using a mortar and pestle. Twenty grams of pulverized tissue were used for DNA extraction using the SILEX isolation method (Vilanova et al. 2020). Quality and concentration were measured by fluorometry using Qubit 4 and Qubit 1X dsDNA HS Kit. To prepare the library samples for EM-seq, 100 μg of DNA was mixed in 48 μl of TE buffer with 1 μl each of CpG unmethylated Lambda and CpG methylated pUC19 control DNA provided by the NEBNext© Enzymatic Methyl-seq kit. Samples were then sonicated using a Covaris ® S220 focused-ultrasonicator in

microtube AFA Fiber Pre-Slit Snap-Cap 6 mm x 16 mm tubes. Then, the libraries were prepared using the NEBNext© Enzymatic Methyl-seq kit following the standard protocol.

MethylSeq Library Preparation and Sequencing

Library concentration and quality was assessed using the Qubit 4 and Qubit 1X dsDNA HS Kit and by electrophoresis using a TapeStation. Libraries were pooled equimolarly, then cleaned using an equal volume of NEBNext© Sample Purification Beads. The pool was then eluted in 25 µl TE buffer and then concentration and DNA quality were evaluated using the above methods. The library pool was sequenced on two lanes of Illumina® HiSeq 4000 platform to generate 150-bp paired-end reads.

Alignment of MethylSeq Libraries

Read quality of each sample was assessed using FastQC v 0.12.0. Reads were then trimmed using Trimmomatic version 0.33 using the default settings, then quality was reassessed. Forward and reverse strands were then combined into a single fastq file. Reads were then aligned to the Nonpareil almond genome 2.0 (D'Amico Wilman 2022a). Samples were deduplicated, then methylation reads were called using Bismark version 0.22.3 (Krueger 2011) using default settings in paired-end mode. Nonpareil genome was selected as a reference due to both genotypes sharing “Nonpareil” in their lineage. To test conversion efficiency, reads were also aligned to the Lambda and pUC19 nucleotide sequence fasta files provided by NEB. All analyses were performed using resources from the Ohio Supercomputer Center (Ohio Supercomputer Center. 1987).

Whole Genome Methylation and Differential Methylation Analysis

Global methylation read counts were extracted from the Bismark methylation call report. To calculate total methylation, the total number of methylated cytosines was divided by the total number of reads for cytosines of a given sample. For each context (CpG, CHG, and CHH), percent methylation was calculated by the number of methylated cytosines at that context divided by the total numbers of cytosines at that context. Results were loaded into R statistical software v. 4.2.1. Data was tested for normality using Shapiro-Wilks test and F-test was performed to test homogeneity of variance. To analyze differences between varieties and tissue types, a Student's T-test was performed, and an ANOVA between the locations. Pairwise comparisons between locations were performed using the `emmeans()` functions from the R package `emmeans` v1.11.0.

Differentially methylated regions were called using the R package `DMRcaller` version 1.41.0 (Catoni 2018). CX report files generated by Bismark were loaded into R and used to generate pooled Granges objects. To analyze pooled samples, `readBismarkPool()` was used to pool biological replicates into a single object, while individual sample reports were loaded using the `readBismark()` function. Lists were then generated based on genotype, assigning samples to different treatment groups for “shoot type” and “location.” Separate lists for each genotype were created to avoid confounding polymorphisms between genotypes as differences in methylation. For the locations, pairwise DMR calling was performed between each of the locations. Methylation profiles for each sample were made using the `computeMethylationProfile()` function, then profiles were compiled into a single list. DMRs were called using the `filterDMRs()` function for each chromosome, based on annotation from the GTF

file for the Nonpareil almond loaded into R using read.gtf(). Each included DMR had a minimum coverage of three, occurred over at least four cytosines, and had an adjusted p-value < 0.01 with multiple testing correction. DMRs were also called using the “bins” method in the computeDMRs() function, selected under the same parameters as the region-called DMRs.

Protein Analysis

Locus tags identified in the DMR calls stored in a text file, then uploaded to UniProt (The UniProt Consortium 2025). Locus tags were used as the Gene Name input parameter, then mapped to the Uniport database using Prunus Dulcis as the selected taxonomy. Proteins were categorized by protein family, then grouped by gain and loss of methylation relative to the axillary level.

Results

Whole Genome Methylation

In this study, we sampled tissue derived from epicormic and axillary canopy shoots of two almond genotypes at four locations across California to determine how DNA methylation varies between

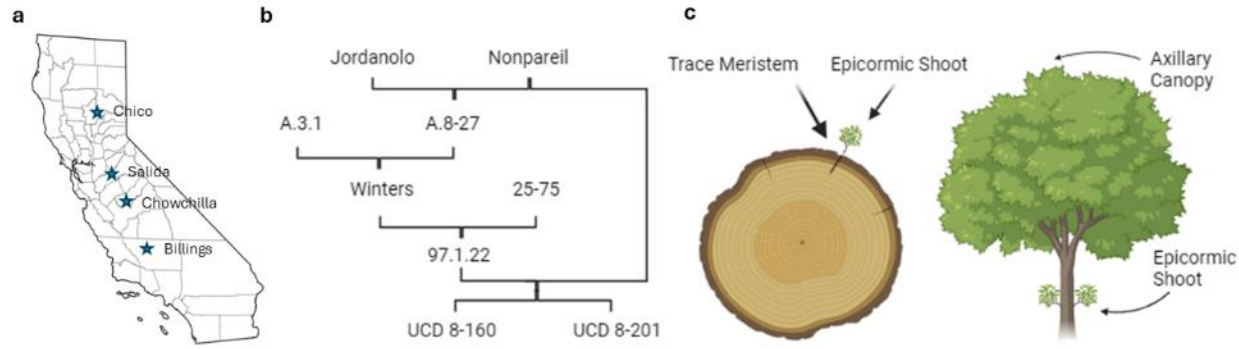


Figure 1.1. Experimental design. a) Four orchards were sampled in the Central Valley of California. b) Two genotypes (8-160 and 8-201) were sampled, both with the same pedigree that includes two NBF-afflicted cultivars, Nonpareil and Jordanolo. c) Two tissue types, epicormic shoots and canopy shoots, were sampled. Four replicates of each tissue type were sampled for each genotype at each location, for a total of 64 samples.

tissues in mature individuals of a genetic clone. We collected samples from multiple locations and cultivars to examine how factors other than tissue type affect methylation (Figure 1.1). This experimental design, along with the inclusion of biological replicates, was intended to filter noise caused by externalities and increase the confidence of differences in methylation between epicormic and axillary canopy-derived shoots.

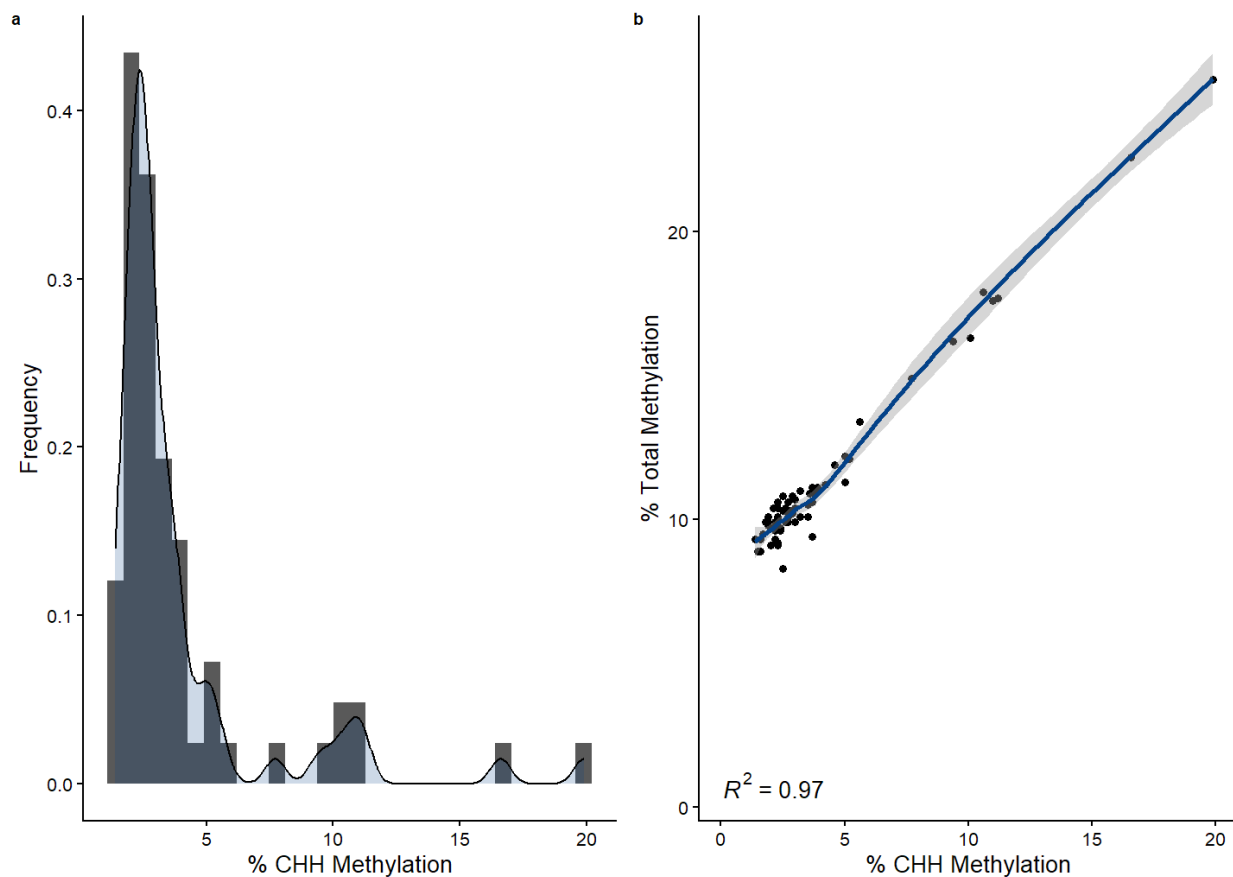


Figure 1.2. a) Distribution of CHH global methylation across all almond samples. b) Correlation between total methylation and CHH methylation across all samples. Outliers with high total and CHH methylation are presumed to result from incomplete digestion during EM-Seq library preparation.

Preliminary analysis of total methylation data revealed outliers: nine samples had levels of CHH methylation and total methylation more than four standard deviations above the mean (Figure 1.2). These outliers were not associated with any of our treatment groups or any other biological factors between the replicates. Because of this high level of variance, we concluded that these outliers may have resulted from incomplete digestion during enzymatic conversion, whereby the catalytic polypeptide-like 3A did not completely deaminate cytosine in the library preparation step, (Morrison 2021), and thus were removed from subsequent downstream analyses.

After removing outlier samples, mean total global methylation in all three contexts (CpG, CHG, and CHH) did not differ significantly between genotypes, tissues, or locations (Figure 1.3). In a two-sample *t*-test, there were no methylation differences between epicormic and axillary canopy shoots in the CpG context ($p=0.09$), the CHG context ($p=0.18$), or the CHH context ($p=0.87$). There were no differences in total methylation either ($p=0.63$). Similar results were obtained when two-sample *t*-tests were performed between UCD 8-160 and UCD 8-201, with no differences in the CpG context ($p=0.98$), the CHG context ($p=0.42$), or the CHH context ($p=0.59$) between the varieties. When a one-way ANOVA was performed on location, there were no significant differences among the locations for total methylation ($p\text{-value}=0.62$) or in the three contexts ($p\text{-value}=0.91$, $p\text{-value}=0.94$, $p\text{-value}=0.49$, for CpG, CHG, and CHH, respectively). The interaction of variety and location does appear to be significant for the CpG ($p=0.03$) and CHG ($p=0.04$) contexts, but not for total methylation or the CHH context. The interactions between tissue and variety, tissue and location, and all three treatment groups were not significant.

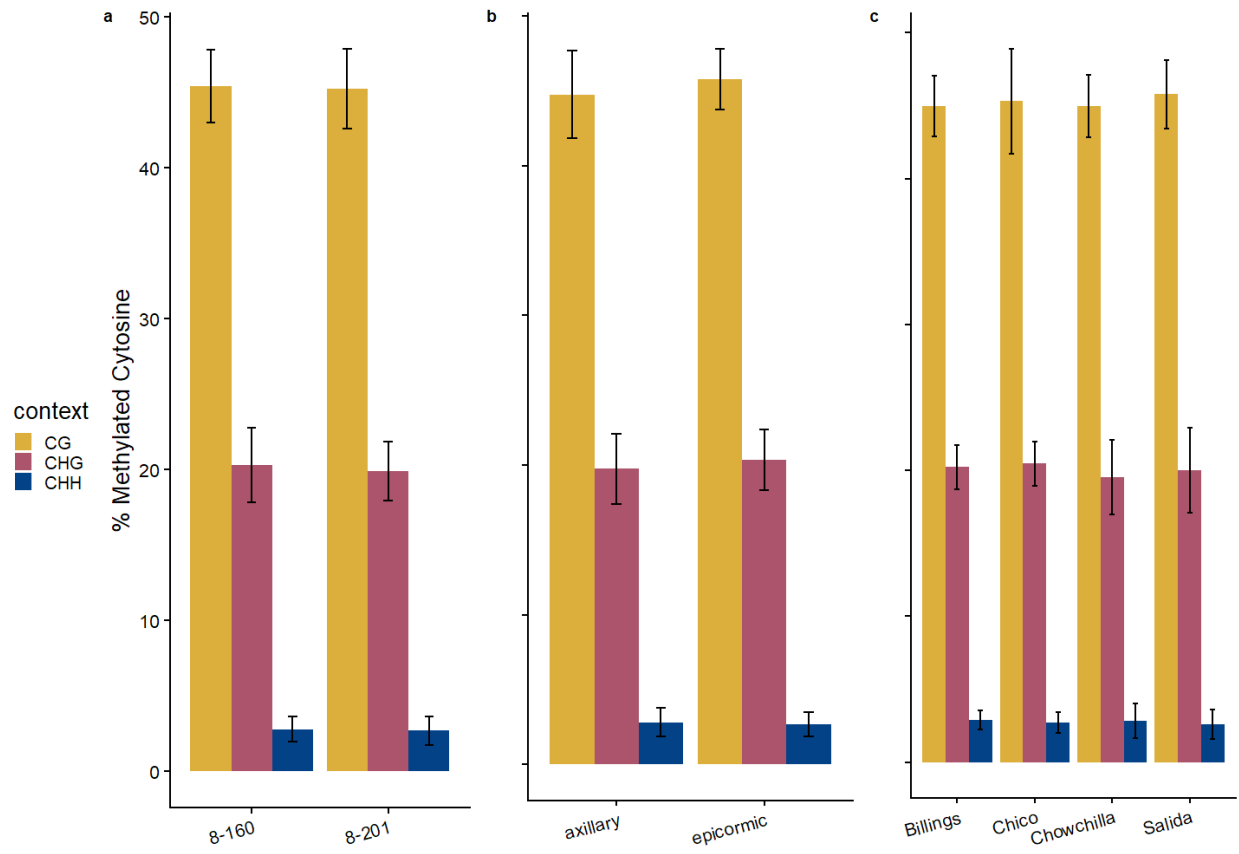


Figure 1.3. Global methylation in all three contexts (CG, CHG, and CHH) across genotypes (a), tissue types (b), and locations (c).

Relationships between Samples

To explore robust differences in methylation patterns among genotypes, tissues, and locations, we pooled biological replicates and calculated the proportion of methylated cytosines in each context (CpG, CHG, and CHH) in 1,000, 10,000, and 100,000 base-pair windows across the genome. For each window, we also calculated gene counts and gene density, or the proportion of the interval made up of genes. Principal component analysis on the resulting dataset showed that samples cluster by genotype, regardless of the window size or methylation context (Figure 1.4). For each of the PCA plots at the CpG context, samples clustered by genotype along the first

principal component, but the variation decreased proportionally with window size. These trends were repeated in the CHG context. However, the relationships were not as pronounced in the CHH context, and genotypes separated along the second principal component.

We then constructed a distance dendrogram based on each window. Samples were aligned into two major clades based on the two varieties. This trend was maintained for each of the three window sizes. The smaller clades within each genotype clade were dispersed more randomly. Some samples were closely related to each other by tissue type, while others were more closely related by location. Association between tissue types increases somewhat with the smaller window sizes; however, no discernable pattern emerges.

Differentially Methylated Regions

We used two methods to determine define regions when calling DMRs: independent windows of various intervals and genomic positions from the annotated Nonpareil almond genome (D'Amico-Willman et al. 2022a). Then, using these two region types, we compared methylation between sample groups.

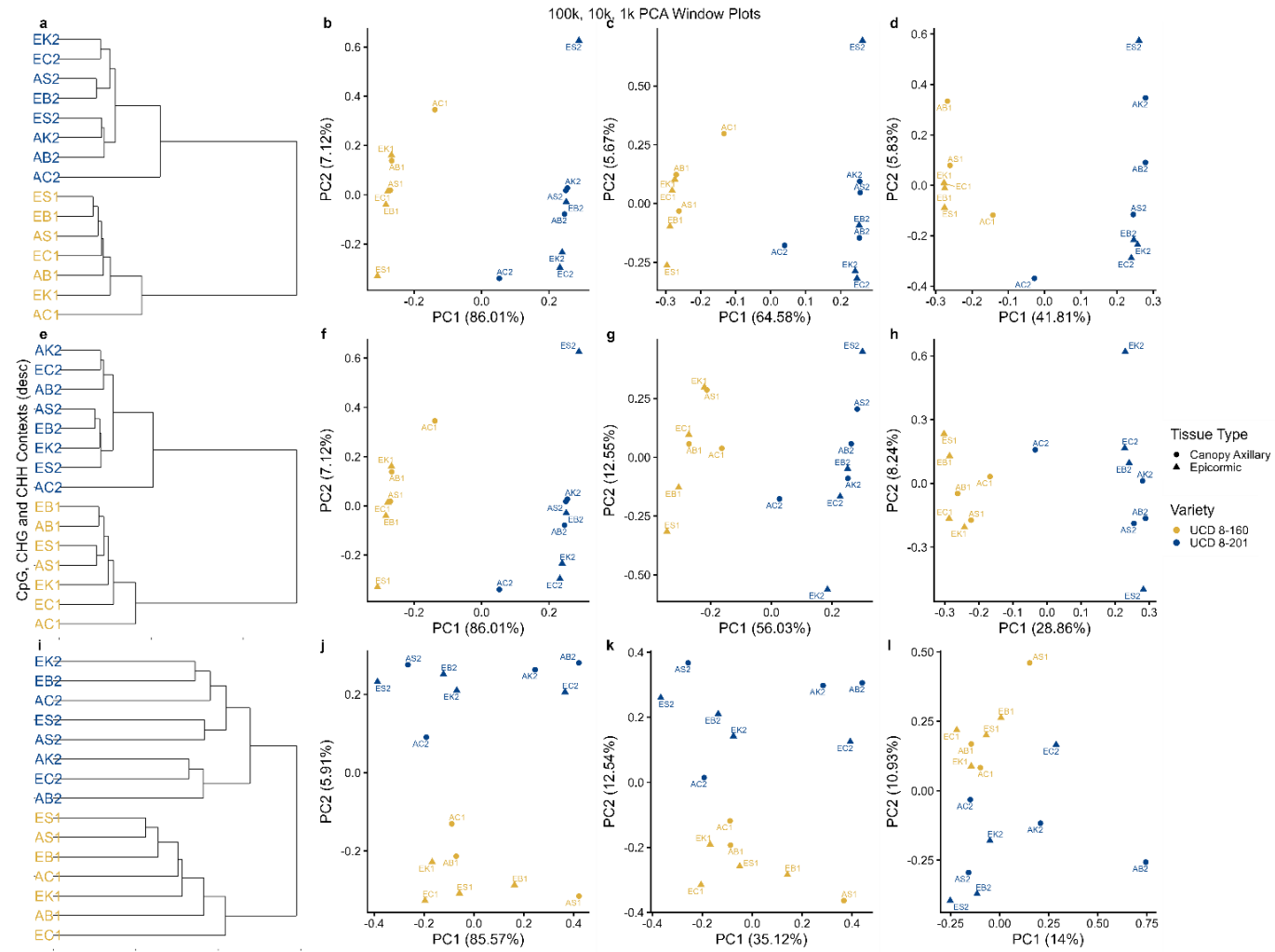


Figure 1.4. Variation in methylation patterns between samples in the CG context (b, c, d), CHG context (f, g, h), and CHH context (j, k, l). Methylation frequencies in each context were summarized over 100-kb (b, f, j), 10-kb (c, g, k) and 1-kb (d, h, l) windows. Dendrograms were constructed for using data from 100-kb (a), 10-kb (e) and 1-kb (i) windows at the CpG context. Three-letter codes for each sample indicate the tissue type (E=epicormic; A=axillary canopy), location (K=Chowchilla, C=Chico, S=Salida, B=Billings), and genotype (1=8_160; 2=8_201). Axillary samples are represented as circles, and

We separated the DMR calling to different tissue types within the varieties, then called DMRs between varieties. Within each variety, we also performed pairwise comparisons between each of the locations. First, we called DMRs using 100,000, 10,000, and 1,000 base pair windows in each context. In UCD 8-160, there were no significant DMRs in any of the window sizes or in any context. In UCD 8-201, there was a single significant DMR on chromosome 5 at 1,000 base-pair windows in the CpG context. There were no significant DMRs in any of the other contexts at any of the other window sizes. However, we did identify 68 DMRs when analyzing differences in methylation in annotated regions (Figure 1.6). In genotype UCD 8-160, we found a total of 23 DMRs between tissues: 18 in the CpG context, 4 in the CHG context, and 1 in the CHH context. For genotype UCD 8-201, we found a total of 46 DMRs between tissues: 27 in the CpG context, 14 in the CHG context, and 4 in the CHH context. There were no overlapping regions between the two varieties for DMRs associated with tissue type. There were 4546 total DMRs between varieties: 4264 and 284 in the CpG and CHG contexts respectively, and no DMRs in the CHH context. For the locations, there were 312 total DMRs for each comparison of the in the pairwise comparison between the locations across all three contexts (Table 1.1) Of these 312 DMRs, 96 have a shared position with at least one other comparison, but none shared between all six comparisons.

Table 1.1 Pairwise DMR Counts between Location

<i>Comparison</i>	<i>DMR Counts</i>		
	CpG	CHG	CHH
<i>Billings - Chico</i>	31	9	3
<i>Billings - Chowchilla</i>	22	9	2
<i>Billings - Salida</i>	28	17	2
<i>Chico – Chowchilla</i>	47	13	2
<i>Chico – Salida</i>	39	19	1
<i>Chowchilla – Salida</i>	32	35	1

Associated Proteins with Differentially Methylated Regions

To elucidate some of the functional properties of the differential methylated regions, we examined the known protein functions of the gene bodies of the DMRs. We extracted the locus tags, or gene identifiers of entire gene models, associated with the DMRs, then found known proteins associated with those. Of the 68 DMRs identified between epicormic and axillary tissue, 48 had an unidentified or unknown protein associated with them (Figure 1.5). With the remaining 20 DMRs, we identified 17 distinct families of proteins. Of those distinct families, only TRAFAC class myosin-kinesin ATPase superfamily, Kiensin and PP2C families had more than one DMR within that distinct family. only TRAFAC class myosin -kinesin ATPase superfamily

proteins showed an overall gain in methylation, while the PP2C class of proteins showed a split between hyper – and hypo methylation.

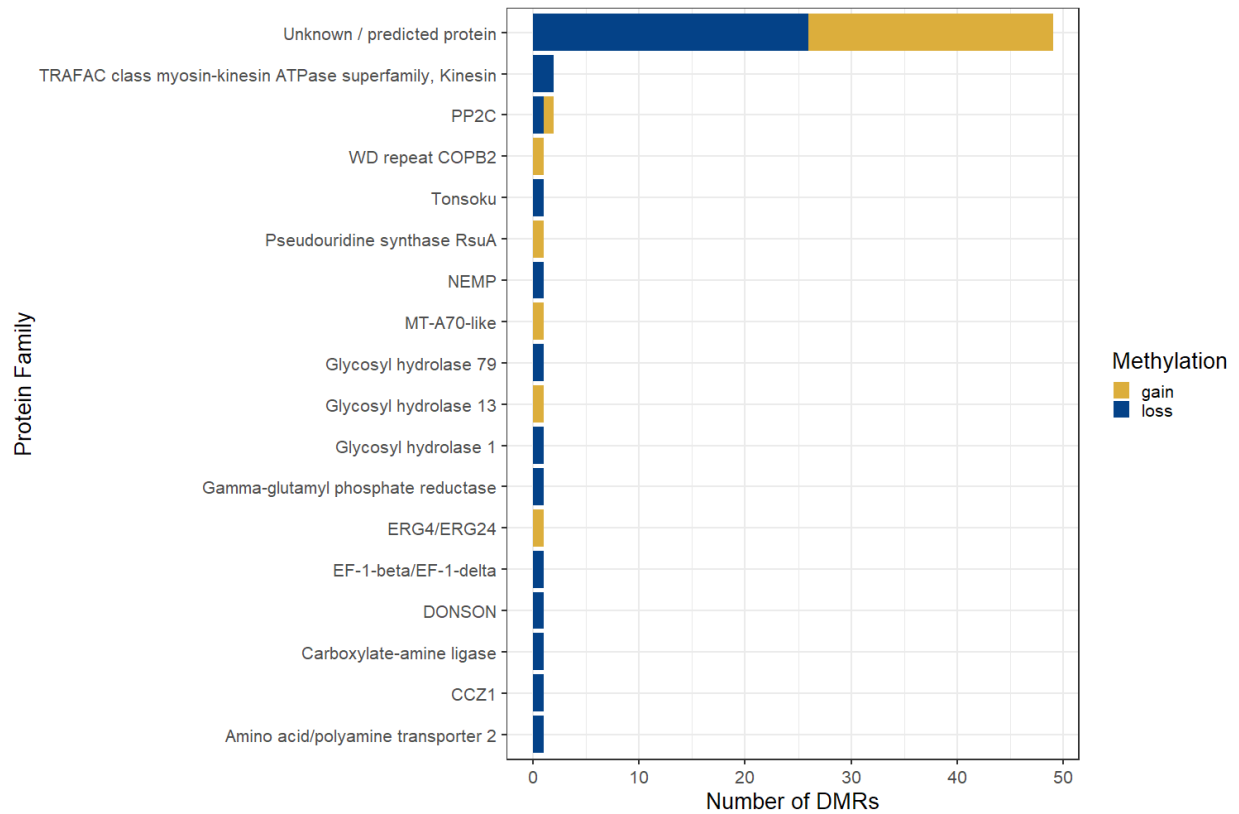


Figure 1.5. Protein families associated with DMR between epicormic and axillary derived tissue. Color indicates whether the DMR showed a gain or loss of methylation relative to the axillary tissue.

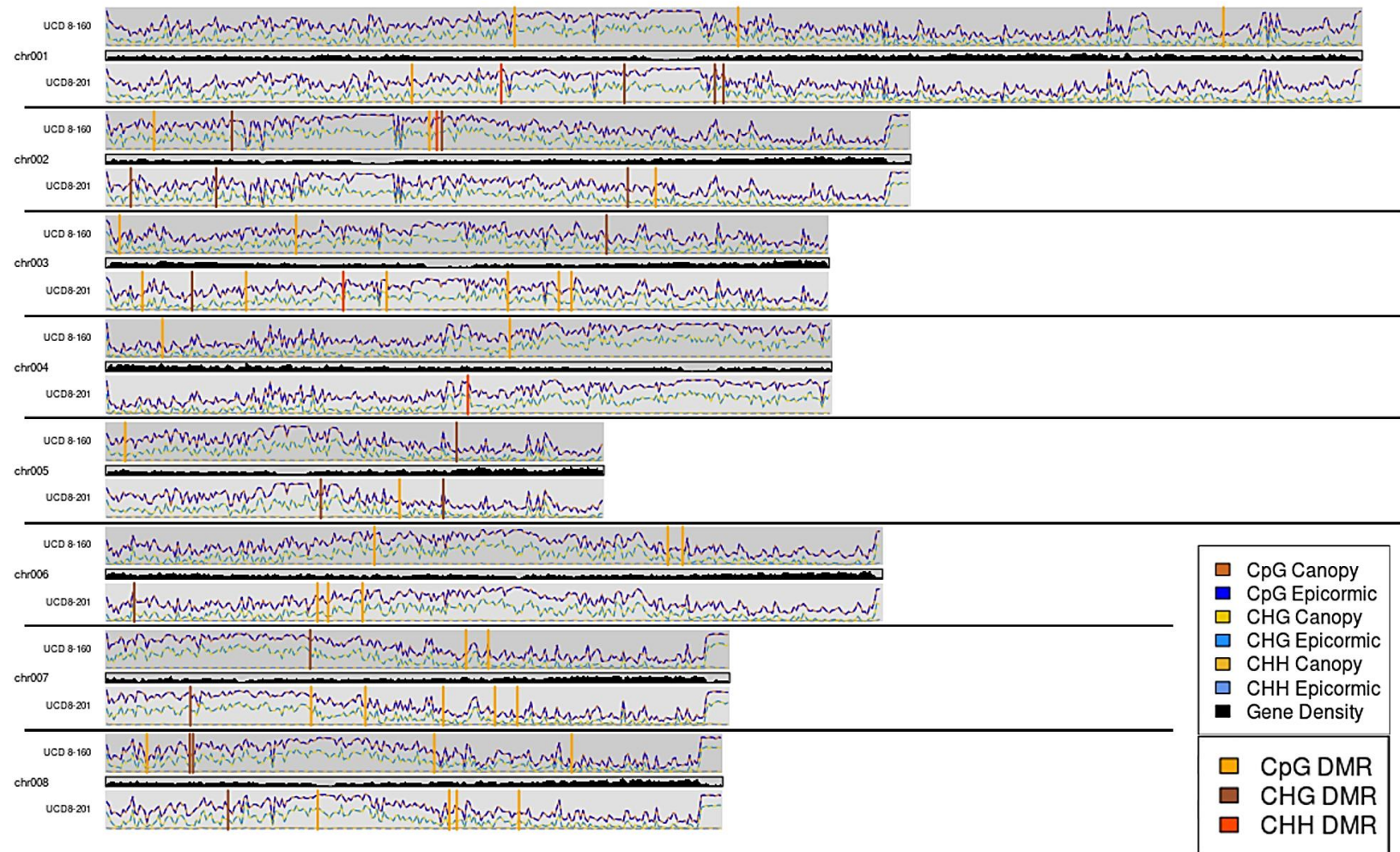


Figure 1.6. Identification of differentially methylated regions (DMRs) between epicormic and axillary shoots. The y-axis of each chromosome represents proportion of methylated cytosines at 100kb windows. Lines showing DNA methylation profiles of the two, with the dash lines used to represent their overlapping nature. The black bar plots show gene density at that location. DMRs (thick vertical lines) were identified in all three methylation contexts for each genotype separately and to show their approximate position on the genome.

Discussion

Global Methylation is Similar between Epicormic and Canopy Axillary Tissue

Our epigenetic analysis of samples displaying different propensity to NBF does not support the hypothesis that differential DNA methylation is involved. Additionally, there were no significant differences in global methylation between either locations or genotypes. In contrast, previous studies have found heritable changes in DNA methylation during in-vitro propagation (Ghosh 2021). For example, in rice, methylation was lost after plants were regenerated from callus (Stroud 2014) and in soybeans, long periods of continuous tissue culture were shown to increase overall DNA methylation, particularly in the CHH context (Ji 2019). Environmental factors including moisture and elevation have also been associated with phenotypic plasticity and DNA methylation of potato clones in the Andes (Ibanez 2021). However, our results in almond suggest genome-wide DNA methylation is stable for a given cultivar across multiple environments and tissue sources. Some apparent methylation differences between genotypes in this study may result from single nucleotide polymorphisms (Noshay 2020). Similar patterns of stable global methylation with distinct changes in regional methylation levels have also been observed in maize (Stepflugh 2014). Notably, and in contrast to many of the studies cited above, our study used whole-genome enzymatic sequencing rather than bisulfite sequencing and had a relatively large sample size with biological replicates.

Differential Methylation between Epicormic and Axillary Tissue is Genotype-Specific

This study was based on the rationale that understanding methylation differences between epicormic and axillary buds could help elucidate the mechanisms underlying NBF. NBF could be

an aging-related disorder (Gradziel 2019), and a previous study in almond found significant differences in methylation between cohorts of two-, seven- and 11-year-old trees (D'Amico-Wilman 2024). Because epicormic buds originate as trace meristematic tissue under the bark, often remaining dormant for years to decades, epicormic tissue is theorized to be ontogenetically younger, and may be exposed to fewer environmental stressors than the canopy. Epicormic buds contribute to the longevity of individual trees (Borges 2009) and DNA methylation has been used as an age predictor in perennial species (Gardener 2023). In addition, apical meristems have been shown to have variable methylation in stress response genes during first emergence after dormancy (Le Gac 2018). Epicormic growth itself is often associated with stress (Zverevs 2001) and exposure to abiotic stresses can result in methylation changes in genes of stress response pathways. Accumulated methylation changes, therefore, could function as a stress memory (Sun 2021). Given these premises, we hypothesized that both tissue type (epicormic versus canopy axillary shoot) and location would be associated with variation in DNA methylation. Additionally, PP2C family proteins have been associated with response to abiotic stress factors in the Rosaceae (Wang 2021). Although we observed significant differences in DNA methylation between epicormic and canopy shoots at specific genomic regions, these regions were not shared between the two genotypes, only partially supporting our hypothesis. Surprisingly, no significant differences in DNA methylation were found among locations. However, the time scale of this experiment may be too short, and future studies could be conducted on older trees to determine whether methylation differences eventually accumulate between tissues and locations.

Practical Applications

Applied research on epicormic growth has aimed to develop an assay to screen for NBF before the onset of symptoms. To determine the viability of such a diagnostic assay, it will be necessary to conduct whole-genome enzymatic methyl-sequencing experiments on trees showing different levels of bud failure. Results from our study suggest that assays could be developed for epicormic-derived shoots, which should be at lower risk for NBF. However, such assays would likely be genotype-specific. In the future, assays could also be used to select NBF-free breeding material, or help develop tools to select epigenetically superior clones (Shaik 2022).

Conclusion

Global DNA methylation states are highly conserved across tissues and locations in clonal propagules of almond. DMRs between epicormic and axillary shoots were detected for both genotypes, but there was no overlap of DMRs between genotypes. DNA polymorphism was the largest driver of differences in methylation. The remarkable stability of DNA methylation patterns within almond genotypes suggests that perturbation of these patterns by NBF or other disorders would be detectable with modest sample sizes.

References

1. Borges, R. M., (2009) Phenotypic plasticity and longevity in plants and animals: cause and effect? *J. Biosci* 34(4)
2. Burrows, G. E., Offord, C. A., Meagher, P. F., & Ashton, K. (n.d.). *Axillary Meristems and the Development of Epicormic Buds in Wollemi Pine (Wollemia nobilis)*. <https://doi.org/10.1093/aob/mcg207>
3. Catoni, M., Tsang, J. M. F., Greco, A. P., & Zabet, N. R. (2018). DMRcaller: A versatile R/Bioconductor package for detection and visualization of differentially methylated

regions in CpG and non-CpG contexts. *Nucleic Acids Research*, 46(19).

<https://doi.org/10.1093/nar/gky602>

4. Chen, D. P., Lin, Y. C., & Fann, C. S. J. (2016). Methods for identifying differentially methylated regions for sequence- and array-based data. *Briefings in Functional Genomics*, 15(6), 485–490. <https://doi.org/10.1093/bfgp/elw018>
5. D'Amico-Willman, K. M. (a), Ouma, W. Z., Meulia, T., Sideli, G. M., Gradziel, T. M., & Fresnedo-Ramírez, J. (2022). Whole-genome sequence and methylome profiling of the almond [*Prunus dulcis* (Mill.) D.A. Webb] cultivar 'Nonpareil.' G3: *Genes|Genomes|Genetics*, 12(5). <https://doi.org/10.1093/G3JOURNAL/JKAC065>
6. D'Amico-Willman, K. M. (b), Sideli, G. M., Allen, B. J., Anderson, E. S., Gradziel, T. M., & Fresnedo-Ramírez, J. (2022). Identification of Putative Markers of Non-infectious Bud Failure in Almond [*Prunus dulcis* (Mill.) D.A. Webb] Through Genome Wide DNA Methylation Profiling and Gene Expression Analysis in an Almond × Peach Hybrid Population. *Frontiers in Plant Science*, 13(February). <https://doi.org/10.3389/fpls.2022.804145>
7. D'Amico-Willman, K. M., Niederhuth, C. E., Sovic, M. G., Anderson, E. S., Gradziel, T. M., & Fresnedo-Ramírez, J. (2024). Hypermethylation and small RNA expression are associated with increased age in almond (*Prunus dulcis* [Mill.] D.A. Webb) accessions. *Plant Science*, 338, 111918. <https://doi.org/10.1016/j.plantsci.2023.111918>
8. Feng, S., Cokus, S. J., Zhang, X., Chen, P. Y., Bostick, M., Goll, M. G., Hetzel, J., Jain, J., Strauss, S. H., Halpern, M. E., Ukomadu, C., Sadler, K. C., Pradhan, S., Pellegrini, M., & Jacobsen, S. E. (2010). Conservation and divergence of methylation patterning in plants and animals. *Proceedings of the National Academy of Sciences of the United States of America*, 107(19), 8689–8694. <https://doi.org/10.1073/pnas.1002720107>
9. Fenton, C. A. L., Kester, D. E., & Kuniyuki, A. H. (1988). Models for noninfectious bud failure in almond. *Phytopathology*, 78, 139. <https://doi.org/10.1094/Phyto-78-139>
10. Fortes, A. M., & Gallusci, P. (2017). Plant stress responses and phenotypic plasticity in the epigenomics era: Perspectives on the grapevine scenario, a model for perennial crop plants. *Frontiers in Plant Science*, 8(FEBRUARY), 1–9. <https://doi.org/10.3389/fpls.2017.00082>
11. Fresnedo-Ramírez, J., Chan, H. M., Parfitt, D. E., Crisosto, C. H., & Gradziel, T. M. (2017). Genome-wide DNA-(de)methylation is associated with Noninfectious Bud-failure exhibition in Almond (*Prunus dulcis* [Mill.] D.A. Webb). *Scientific Reports* 2017 7:1, 7(1), 1–12. <https://doi.org/10.1038/srep42686>
12. Gardner, S. T., Bertucci, E. M., Sutton, R., Horcher, A., Aubrey, D., & Parrott, B. B. (2023). Development of DNA methylation-based epigenetic age predictors in loblolly pine (*Pinus taeda*). *Molecular Ecology Resources*, 23(1), 131–144. <https://doi.org/10.1111/1755-0998.13698>
13. Ghosh, A., Igamberdiev, A. U., & Debnath, S. C. (2021). Tissue culture-induced DNA methylation in crop plants: a review. In *Molecular Biology Reports* (Vol. 48, Issue 1, pp. 823–841). Springer Science and Business Media B.V. <https://doi.org/10.1007/s11033-020-06062-6>
14. Gradziel, T. M., & Shackel, K. A. (2021). Propagation of an epigenetic age-related disorder in almond is governed by vegetative bud ontogeny rather than chimera-type cell lineage. *Horticulturae*, 7(7). <https://doi.org/10.3390/HORTICULTURAE7070190>

15. Gradziel, T., Lampinen, B., & Preece, J. E. (2019). Propagation from basal epicormic meristems remediates an aging-related disorder in almond clones. *Horticulturae*, 5(2), 1–9. <https://doi.org/10.3390/horticulturae5020028>
16. Gradziel, T., & Fresnedo, J. (2019). Noninfectious Bud-failure As a Model For Studying Age Related Genetic Disorders in Long-Lived Perennial Plants. *Journal of the American Pomological Society*, 73(4), 240–253.
17. Ibañez, V. N., Masuelli, R. W., & Marfil, C. F. (2021). Environmentally induced phenotypic plasticity and DNA methylation changes in a wild potato growing in two contrasting Andean experimental gardens. *Heredity*, 126(1), 50–62. <https://doi.org/10.1038/s41437-020-00355-z>
18. Ji, L., Mathioni, S. M., Johnson, S., Tucker, D., Bewick, A. J., Kim, K. do, Daron, J., Slotkin, R. K., Jackson, S. A., Parrott, W. A., Meyers, B. C., & Schmitz, R. J. (n.d.). *Genome-Wide Reinforcement of DNA Methylation Occurs during Somatic Embryogenesis in Soybean*. <https://doi.org/10.1105/tpc.19.00255>
19. Kester, D. E., Shackel, K. A., Micke, W. C., Viveros, M., & Gradziel, T. M. (2004). Noninfectious Bud Failure in “Carmel” Almond: I. Pattern of Development in Vegetative Progeny Trees. *Journal of the American Society for Horticultural Science*, 129(2), 244–249. <https://doi.org/10.21273/jashs.129.2.0244>
20. Kester, D. E., Asay R. N. (1978). Variability in Noninfectious Bud-failure of “Nonpareil” Almond. I. Location and Environment1. In *J. Amer. Soc. Hort. Sci* (Vol. 103, Issue 3).
21. Krueger, F., & Andrews, S. R. (2011). Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, 27(11), 1571–1572. <https://doi.org/10.1093/bioinformatics/btr167>
22. Lanner, R. M. (2002). Why do trees live so long? *Ageing Research Reviews*, 1(4), 653–671. [https://doi.org/10.1016/S1568-1637\(02\)00025-9](https://doi.org/10.1016/S1568-1637(02)00025-9)
23. Lawrence, M., Gentleman, R., & Carey, V. (2009). rtracklayer: An R package for interfacing with genome browsers. *Bioinformatics*, 25(14), 1841–1842. <https://doi.org/10.1093/bioinformatics/btp328>
24. Le Gac A-L, Lafon-Placette C, Delaunay A & Maury S (2019) Developmental, genetic and environmental variations of global DNA methylation in the first leaves emerging from the shoot apical meristem in poplar trees. *Plant Signaling & Behavior*, 14(6). <https://doi.org/10.1080/15592324.2019.1596717>
25. Lin, W., Xiao, X., Sun, W., Liu, S., Wu, Q., Yao, Y., Zhang, H., & Zhang, X. (2022). Genome-Wide Identification and Expression Analysis of Cytosine DNA Methyltransferase Genes Related to Somaclonal Variation in Pineapple (*Ananas comosus* L.). *Agronomy*, 12(5). <https://doi.org/10.3390/agronomy12051039>
26. Meier, A. R., Saunders, M. R., & Michler, C. H. (2012). Epicormic buds in trees: a review of bud establishment, development and dormancy release. *Tree Physiology*, 32, 565–584. <https://doi.org/10.1093/treephys/tps040>
27. Micke, W.C. (1996) Almond Production Manual. Oakland: University of California (System). Division of Agriculture and Natural Resources. UCANR Publications 3364.
28. Morrison, J., Koeman, J. M., Johnson, B. K., Foy, K. K., Beddows, I., Zhou, W., Chesla, D. W., Rossell, L. L., Siegwald, E. J., Adams, M., & Shen, H. (2021). Evaluation of whole-genome DNA methylation sequencing library preparation protocols. *Epigenetics and Chromatin*, 14(1). <https://doi.org/10.1186/s13072-021-00401-y>

29. Negrón, C., Contador, L., Lampinen, B. D., Metcalf, S. G., Guédon, Y., Costes, E., & Dejong, T. M. (2014). Differences in proleptic and epicormic shoot structures in relation to water deficit and growth rate in almond trees (*Prunus dulcis*). *Annals of Botany*, 113(3), 545–554. <https://doi.org/10.1093/AOB/MCT282>
30. Noshay, J. M., & Springer, N. M. (2020). *Stories that can't be told by SNPs; DNA methylation variation in plant populations*.
31. Ohio Supercomputer Center. 1987. Ohio Supercomputer Center. Columbus OH: Ohio Supercomputer Center. <http://osc.edu/ark:/19495/f5s1ph73>.
32. Perrin, A., Daccord, N., Roquis, D., Celton, J. M., Vergne, E., & Bucher, E. (2020). Divergent DNA methylation signatures of juvenile seedlings, grafts and adult apple trees. *Epigenomes*, 4(1). <https://doi.org/10.3390/epigenomes4010004>
33. Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
34. Shaikh, A. A., Chachar, S., Chachar, M., Ahmed, N., Guan, C., & Zhang, P. (2022). Recent Advances in DNA Methylation and Their Potential Breeding Applications in Plants. *Horticulturae*, 8(7), 562. <https://doi.org/10.3390/horticulturae8070562>
35. Stelpflug, S. C., Eichten, S. R., Hermanson, P. J., Springer, N. M., & Kaeppler, S. M. (2014). Consistent and heritable alterations of DNA methylation are induced by tissue culture in maize. *Genetics*, 198(1), 209–218. <https://doi.org/10.1534/GENETICS.114.165480>
36. Stroud, H., Ding, B., Simon, S. A., Feng, S., Bellizzi, M., Pellegrini, M., Wang, G. L., Meyers, B. C., & Jacobsen, S. E. (2013). Plants regenerated from tissue culture contain stable epigenome changes in rice. *ELife*, 2013(2). <https://doi.org/10.7554/ELIFE.00354>
37. Sun, R. Z., Liu, J., Wang, Y. Y., & Deng, X. (2021). DNA methylation-mediated modulation of rapid desiccation tolerance acquisition and dehydration stress memory in the resurrection plant *Boea hygrometrica*. *PLoS Genetics*, 17(4). <https://doi.org/10.1371/journal.pgen.1009549>
38. The UniProt Consortium , UniProt: the Universal Protein Knowledgebase in 2025, *Nucleic Acids Research*, Volume 53, Issue D1, 6 January 2025, Pages D609–D617, <https://doi.org/10.1093/nar/gkae1010>
39. United States Department of Agriculture National Agricultural Statistics Service, (2020) *2020 California Almond Objective Measurement Report*, https://www.nass.usda.gov/Statistics_by_State/California/Publications/Specialty_and_Other_Releases/Almond/Objective-Measurement/202007almom.pdf
40. Valledor, L., Hasbún, R., Meijón, M., Rodríguez, J. L., Santamaría, E., Viejo, M., Berdasco, M., Feito, I., Fraga, M. F., Cañal, M. J., & Rodríguez, R. (2007). Involvement of DNA methylation in tree development and micropropagation. *Plant Cell, Tissue and Organ Culture*, 91(2), 75–86. <https://doi.org/10.1007/s11240-007-9262-z>
41. Wang, G., Sun, X., Guo, Z., Joldersma, D., Guo, L., Qiao, X., Qi, K., Gu, C., & Zhang, S. (2021). Genome-wide Identification and Evolution of the PP2C Gene Family in Eight Rosaceae Species and Expression Analysis Under Stress in *Pyrus bretschneideri*. *Frontiers in Genetics*, 12. <https://doi.org/10.3389/fgene.2021.770014>
42. Zhang, H., Tang, K., Wang, B., Duan, C. G., Lang, Z., & Zhu, J. K. (2014). Protocol: A beginner's guide to the analysis of RNA-directed DNA methylation in plants. *Plant Methods*, 10(1), 1–9. <https://doi.org/10.1186/1746-4811-10-18>

43. Zvereva, E. L., & Kozlov, M. v. (2001). Effects of pollution-induced habitat disturbance on the response of willows to simulated herbivory. *Journal of Ecology*, 89(1), 21–30.
<https://doi.org/10.1046/J.1365-2745.2001.00509.X>

Chapter 2: Severity of Non-Infectious Bud Failure in Almond (*Prunus Dulcis* Mill DA Webb) is Correlated with Differential Gene Expression and Methylation

Abstract

Non-infectious bud failure is a disorder in almond characterized by the transformation of bud tissue from healthy to necrotic during the dormancy period. The death of meristematic tissue in the terminal shoot results in the distinct “bushy top” pattern and ensuing heavy crop loss. Non-infectious bud failure is a noteworthy disorder in perennial crops due to its mitotic inheritance through asexual propagation and variance in severity between genetic clones. Additionally, the propensity for non-infectious bud failure is also sexually inherited, leading to variance between cultivars to show propensity for the disorder. In this study, we developed a new haplotype-phased reference assembly for the variety “Carmel”. One parent of Carmel, “Texas” has no history of bud failure symptoms, while its other parent, “Nonpareil” does. We then scored Carmel trees in a commercial orchard for non-infectious bud failure severity, then sampled leaf tissue from those trees. We used enzymatic methyl-seq to measure whole-genome DNA methylation and perform differential methylation analysis and we used 3' TAG RNA sequencing to measure gene expression between the levels. Our results show 20 regions of differential methylation between the two haplotypes. Additionally, 2083 genes were differentially expressed between the two haplotypes, with the expression of certain genes highly correlated with bud failure level. Three of these DMRs fall within 5000 base pairs of the differentially expressed genes. These areas of variation in methylation and expression data may indicate candidate genes for non-infectious bud failure and may aid in the development of future diagnostic tools.

Introduction

Non-infectious bud failure (NBF) has been a major threat to almond production for the last century, yet a definitive cause of the disorder has been elusive. NBF is a disorder in almond characterized by the abscission of vegetative buds in the spring after the meristematic tissue became necrotic in a previous growing season (Hellali 1978). This failure of buds in the terminal shoot leads to the characteristic “mule tail” or “bushy top” where subsequent axillary buds emerge rapidly to produce vegetative groups, with new apical meristems then failing to create an iterative pattern of extensive vegetative branches. This rapid onset of new vegetative tissue reduces the potential for floral bud development, which results in severe crop loss on affected scaffolds. NBF was first described in the cultivar “Jordanolo” in the early 20th century. In 1964, it was reported in the variety “Nonpareil” (Micke 1996). Currently, Nonpareil composes 40% of all almond acreage in California. The distribution and onset of symptoms of NBF in the variety “Carmel” have been particularly dramatic (Kester 1998), leading to the reduction of its total acreage from 36% in 1994 to only 4% today (USDA 2020).

Despite its significant effect on almond production, the underlying cause of the disorder has remained a mystery. NBF is noteworthy due to the variation in expression and severity of symptoms between clonally propagated plants. Potential pathogens associated with NBF fail to meet Koch’s postulates, and the spread of the disorder has not been linked to any viral, fungal or bacterial pathogen (Fenton 1988). Instead, spread of the disorder has been linked to clonal propagation. The inheritance is modeled as the accumulation of “bud-failure potential”, whereby each individual propagule of almond has the ability to show symptoms of NBF, and potential is accumulated after repeated propagation, leading to a threshold where it expresses symptoms (Kester 1978). Bud-failure potential can also be inherited during sexual propagation, leading to trees showing symptoms without any previous propagation (Gradziel 2021). Additionally, the severity and speed of onset is associated with increased mean summer temperatures (Kester 1978). There is no known cure for the disease. The spread of the disorder, however, has been

mitigated by the propagation nursery stock using basal epicormic buds of near original seedling clonal-source trees (Gradziel 2019).

Because the NBF phenotype varies among clonally propagated trees, identifying the root cause of the disorder has been difficult. Because of variation between genetic clones and the correlation between severity and the environment, the underlying cause of NBF is believed to be the result of some epigenetic mark (Fresnedo-Ramirez 2017). Between cultivars, NBF has been associated with variation in DNA methylation, a common and heritable epigenetic mark. DNA methylation plays a major role in transcriptional regulation (Poestch 2011), but is complex and varied. DNA methylation in the promoter region often inhibits gene transcription (Zhang 2018). At the same time, it has been shown to activate certain genes during fruit ripening in tomato (Lang 2017). In pear buds, genes with methylated promoters showed lower expression than non-methylated promoters (Li 2024a), however methylation upstream of related genes in rice correlated with higher expression, while gene body methylation was correlated with lower gene expression (Zhang 2016). Changes in promoter methylation and subsequent downstream effects on gene expression have been shown to be induced by abiotic and abiotic stress factors, such as in wild strawberry (Lopez 2022). The context of the differential methylation may play a role in variations in gene transcription as well. In maize, CpG methylation was positively correlated with gene expression in embryos, while CHG methylation was negatively correlated with gene expression (Lu 2015). Exploring the relationship between DNA methylation and its effect on transcription could provide valuable insights into the potential causes of NBF.

Long-read sequencing technologies also create new avenues to explore NBF. Certain cultivars are more prone to NBF than others, with some varieties having no history of symptoms (Gradziel 2017). Inferred haplotype phasing identifies the contribution of alleles from parental chromosomes (Sun 2022).

Determining the source of alleles in hybrid cultivars of almond derived from parents with divergent histories of NBF may help pinpoint a genetic region of interest. Additionally, new advances allow for the

direct observation of base pair modifications from PacBio HiFi (Ni 2022) and Oxford Nanopore (Liu 2021) sequencing without the need for additional laboratory procedures in the library preparation step. These advances allow for new methods of exploration and validation of DNA methylation data.

Previous research studies have not tried to connect observable differences of the NBF phenotype with measurable differences in genomic predictors. Differentially methylated regions associated with NBF and their downstream effects on RNA expression have been characterized in peach-almond hybrids (D’Amico-Willman 2022), and may be a potential source of putative markers predicting bud failure. In this study, we seek to directly observe differences in methylation between trees showing varying degrees of bud failure in a commercial setting using robust statistical methods and high-throughput sequencing. First, we developed a new haplotyped-phased reference assembly for the “Carmel” variety of almond using long-read sequencing technologies. Using haplotype-resolved reference assemblies have proven fruitful in detecting allele specific methylation and expression in pear (Li 2024) and allele specific methylation in tetraploid potato (Sun 2022). Because of the divergent histories of NBF for each of the Carmel parents, understanding methylation on a haplotype level may be illuminating. We then evaluated almond trees of the Carmel variety in commercial settings for NBF symptoms. We sampled leaf tissue from these trees and performed enzymatic methyl sequencing to evaluate whole genome methylation, and 3’ TAG RNA sequencing to evaluate gene expression. We hypothesized that if the underlying cause of NBF is epigenetically controlled as previous research has suggested, there will be significant differences in the methylation and expression of certain genes that will be correlated with the observed differences in the phenotype.

Methods

Tree Sampling

Almond trees from the variety “Carmel” grown in commercial conditions were scored for NBF severity. Scores were evaluated on a scale of 1-4: 1 or “low” were trees where 0-25% of scaffolds showed symptoms of NBF; 2 or “medium-low” were trees where 25-50% of scaffolds showed symptoms of NBF with some crop loss; 3 or “medium-high” were trees where 50-75% of scaffolds showed symptoms of NBF with substantial crop loss; 4 or “high” 75-100% of scaffolds showed symptoms with severe crops loss. Criteria for these scores were developed using commercial nursery standards for NBF severity. From the evaluated trees, 32 trees were randomly selected, with eight replicates from each of the four levels of bud failure. Approximately 50-100 gms of leaves were sampled from each tree sampled from around the canopy to capture microvariations. After abscission from the tree, samples were immediately submerged in liquid nitrogen, then placed in 100 ml falcon tubes placed on dry ice. Samples were then transferred to a -80°C freezer for storage.

For the reference assembly, samples were taken from a divergent population of NBF positive and negative trees propagated specifically to maximize differences in NBF severity. Leaves samples were taken from a positive tree and one from a negative tree. Approximately 50-100 g of leaves were sampled from each tree, then immediately frozen with liquid nitrogen and placed on dry ice. Samples were stored in a -80°C freezer before high molecular weight DNA extraction.

DNA Extraction

Leaf samples were frozen in liquid nitrogen, then ground into a fine powder using a mortar and pestle. 20 g of pulverized tissue were used for DNA extraction using the SILEX isolation method (Vilanova et al. 2020). Quality and concentration were measured by fluorometry using Qubit 4 and Qubit 1X dsDNA HS Kit. To prepare the library samples for EM-seq, 100 µg of DNA was mixed in 48 µl of TE buffer with 1 µl each of CpG unmethylated Lambda and CpG methylated pUC19 control DNA provided by the NEBNext® Enzymatic Methyl-seq kit. Samples were then sonicated using a Covaris® S220 focused-

ultrasonicator in microtube AFA Fiber Pre-Slit Snap-Cap 6 mm x 16 mm tubes. Then, the libraries were prepared using the NEBNext® Enzymatic Methyl-seq kit following the standard protocol.

MethylSeq Library Preparation and Sequencing

Library concentration and quality was assessed using the Qubit 4 and Qubit 1X dsDNA HS Kit and by electrophoresis using a TapeStation. Libraries were pooled into equimolar ratios then cleaned using an equal volume of NEBNext® Sample Purification Beads. The pool was then eluted in 25 µl TE buffer and then concentration and DNA quality were reevaluated using the same methods as the original samples. The library pool was sequenced on two lanes of Illumina® HiSeq 4000 platform to generate 150bp paired-end reads.

RNA Extraction

Samples were refrozen using liquid nitrogen, then ground into a fine powder using a mortar and pestle. 50 mg of the pulverized tissue were used for the Thermo Fisher Scientific® MagMax Kit and Plant RNA Isolation aid using the standard protocol. RNA samples were eluted in 50 µl of elution buffer in 2 ml tubes. 5 µl of RNA were removed from each samples and its quality was tested using the Agilent® Bioanalyzer RNA Nano 6000 assay and the Agilent® Bioanalyzer 2100 using the standard protocols.

High Molecular Weight Extraction

While frozen, the leaf tissue was fractured into small pieces and approximately one gram was used to isolate nuclei according to the PacBio® ‘Isolating nuclei from plant tissue using TissueRuptor® disruption procedure’ and its connected checklist. High molecular weight (HMW) DNA was isolated using the PacBio® Nanobind Plant Nuclei Big DNA kit, following the included protocol.

PacBio Library Preparation and Sequencing

For PacBio sequencing, the HiFi SMRTbell library was constructed using the SMRTbell prep kit 3.0 (Pacific Biosciences, Menlo Park, CA; Cat. #102-182-700) according to the manufacturer's instructions. HMW gDNA was sheared to a target DNA mode size of ~20Kb using Diagenode's Megaruptor 3 system (Diagenode, Belgium; cat. B06010003). The sheared gDNA was concentrated using 1X of SMRTbell cleanup beads provided in the SMRTbell prep kit 3.0 for the repair and a-tailing incubation at 37 °C for 60 minutes and 65 °C for 5 minutes, followed by ligation of overhang adapters at 20 °C for 30 minutes, clean-up using 1X SMRTbell cleanup beads, and nuclease treatment at 37 °C for 15 minutes. Size selection was performed on the SMRTbell library using the LightBench CF system from Yourgene Health for the removal of 9Kb and smaller fragments (Yourgene Health, Manchester, UK; Cat. #CG-12500-03). The resulting 20.5Kb average HiFi SMRTbell library was sequenced at the UC Davis DNA Technologies Core (Davis, CA) using one 25 M SMRT cell (Pacific Biosciences, Menlo Park, CA; Cat #102-202-200) on the Revio system with SPRQ sequencing chemistry and a 30-hour movie time.

Oxford Nanopore Library Preparation and Sequencing

Genomic DNA was cleaned using a high-salt-phenol-chloroform method. DNA fragments shorter than 10 Kb were removed using LightBench Discover (Yourgene Health). Sequencing libraries were prepared from 1.5 µg of genomic DNA using a PCR free native barcoding kit SQK-NBD114.24 (Oxford Nanopore Technologies, Oxford, UK). Libraries were made following the instructions of the manufacturer with the exception of extended incubation times for DNA damage repair, end repair, ligation and bead elutions. 900 ng of the final library was loaded on the PromethION R10.4.1 flow cell (Oxford Nanopore Technologies, Oxford, UK). Sequencing runs were performed using MinKNOW software version PromethION 24.11.11. Data was basecalled during the run with ont-dorado-server-for-promethion 7.6.8 using super-accurate basecalling model v4.3.0, 400 bps and all-context 5mc-5hmc modifications.

Genome Assembly

Prior to genome assembly, the Oxford Nanopore Technologies (ONT) adapters were trimmed using porechop v0.2.4 (<https://github.com/rrwick/porechop>). To assemble the haplotype-phased reference genome of ‘Carmel’, the PacBio High-Fidelity Circular Consensus Sequencing (HiFi CCS) and trimmed ONT long-reads from the plant with no bud-failure were used. Additionally, whole genome sequencing data from the parent *Prunus dulcis* cultivars ‘Nonpareil’ (D’Amico-Wilman 2022) and ‘Texas’/‘Mission’ (Alito 2020) were leveraged for triobinning, and contigs were assembled using hifiasm v0.25.0-r726 (Cheng 2021, 2022, 2024). Prior to assembly, parentage was confirmed using Merqury /Meryl v1.3 (Rhie 2020). The phased contigs were then scaffolded using RagTag v2.1.0 (Alonge 2022) to the Nonpareil assembly.

The assemblies were assessed for telomeric sequences using custom R(R Core Team) scripts modified from the ggenomics package (<https://github.com/matthewwdavis/ggenomics>). Using SyRI v1.5.4 (2019) and plotSR v0.5.4 (Goel 2022), haplotypes were assessed for large scale synteny. Small contigs were appended to the ends of chromosomes after the telomeric sequences during scaffolding on several chromosomes. In all cases, these contigs were asyntenic to the corresponding regions on the complementary haplotype. The lack of synteny, small size, and location of these contigs indicated they were instances of incorrect scaffolding. These were corrected by manually editing the assembly placement file, followed by using this edited file to inform an additional round of contig scaffolding with RagTag. The assembled haplotypes were concatenated, resulting in six final assemblies: a ‘Nonpareil’ haplotype assembly with unanchored contigs, a chromosome only ‘Nonpareil’ haplotype assembly, a ‘Texas’/‘Mission’ haplotype assembly with unanchored contigs, a chromosome only ‘Texas’/‘Mission’ haplotype assembly, a concatenated haplotype assembly with unanchored contigs, and a chromosome only concatenated haplotype assembly.

Genome Assembly Assessment

The sequencing reads were initially assessed for 21-mer content using jellyfish v2.2.10 (Marcais 2012), and GenomeScope v2.0 (Ranallo-Benavidez 2020) was used to predict genome size and heterozygosity. Assembly statistics were evaluated on all assemblies with Seqkit v2.10.0 (Shen 2016) and completeness was identified with BUSCO v5.8.3 (Tegenfeldt 2025) using the eudicotyledons_odb12 database. Telomere presence and scaffold locations were identified using custom R scripts modified from the ggenomics package. Assembly comparisons. The assembled ‘Nonpareil’ and ‘Texas’/‘Mission’ haplotypes were aligned using Minimap2 v2.30-r1287. Large scale structural variation and synteny were visualized using a dot plot generated by DGENIES(Cabanettes 2018). SyRI was used to generate calls of structural variation and synteny, which were visualized using plotSR.

3' TAG RNA Library Preparation and Sequencing

Barcoded 3' Tag-Seq libraries were prepared using the QuantSeq FWD kit for multiplexed sequencing according to the recommendations of the manufacturer. The fragment size distribution of the libraries was verified via micro-capillary gel electrophoresis on an Agilent® Bioanalyzer 2100. The libraries were quantified by fluorometry on a Qubit instrument and pooled in equimolar ratios. 32 libraries were sequenced on one lane of an Aviti® sequencer with single-end 100bp reads. The sequencing generated more than 4 million reads per library.

Gene, Transposable Element, and Tandem Repeat Annotations

Gene models were lifted from the Nonpareil reference genome to each haplotype assembly using Liftoff v1.6.3 (Shumate and Salzberg 2021), and independently predicted gene structures on each assembly with Helixer v0.3.6 (Stiehler et al. 2020; Holst et al. 2026) using the land_plant model. The two annotations were then merged with agat_sp_complement_annotations.pl from AGAT (Another GTF/GFF Analysis Toolkit) v1.7.0 (Dainat et al. 2026), which adds a Liftoff gene model only when it does not overlap a Helixer prediction. This preserves the Helixer gene structures, which were structurally more accurate and complete, while recovering loci missed by de novo prediction. After merging, we removed genes with

internal or missing stop codons. The completeness of the final gene models was assessed using BUSCO v5.8.0 (Manni et al. 2021) with the eudicotyledons dataset (eudicotyledons_odb12). The final gene models were functionally annotated with InterProScan 5 (Jones et al. 2014) and eggNOG-mapper (Cantalapiedra et al. 2021), using the protein sequence of the longest isoform per gene. Transposable elements were annotated with the Extensive de novo TE Annotator (EDTA) v2.2.2 (Ou et al. 2019), run on the eight chromosomes with default parameters. The genome-wide distribution of tandem repeats was characterized with TRASH (Tandem Repeat Annotation and Structural Hierarchy; Wlodzimierz, Hong, and Henderson 2023).

Alignment of MethylSeq Libraries

Read quality of each sample was assessed using FastQC v 0.12.0. Reads were then trimmed using Trimmomatic version 0.33 using the default settings, then quality was reassessed. Forward and reverse strands were then combined into a single fastq file. Reads were then aligned to the concatenated “Carmel” reference genome. Samples were deduplicated, then methylation reads were called using Bismark version 0.22.3 (Krueger 2011) using default settings in paired-end mode. To test conversion efficiency, reads were also aligned to the Lambda and pUC19 nucleotide sequence fasta files provided by NEB. All analyses were performed using resources from the Ohio Supercomputer Center (Ohio Supercomputer Center. 1987).

Alignment of RNA Libraries

Read quality was assessed using FastQC v.0.12.0. Reads were trimmed using BBMap v35.85 (Bushnell 2014) using a minimum length of 20 base-pairs. Read quality was then reassessed using the same procedure. Gff3 files for each of the haplotypes were concatenated, then converted into a GTF file. An index for the Carmel genome was produced using STAR 2.7.11b (Dobin 2012) utilizing the concatenated haplotype Carmel genome fasta and associated GTF file, then reads were aligned using STAR quant mode

to generate gene counts. All analyses were performed using resources from the Ohio Supercomputer Center (Ohio Supercomputer Center. 1987).

RNA Seq Analysis

After alignment, unnormalized count files were compiled into a single matrix by loading the individual STAR output into R statistical computing software. Samples were filtered by genes that had expression counts equal to zero, removing unexpressed genes. Differential gene analysis was conducted using the R package edgeR v 3.23 (Chen 2025). Samples were normalized by read depth, and dispersion between samples were calculated. A generalized linear model was then fit using a design matrix of the experimental design (1 treatment, 4 levels, 32 total samples), then the model was fit relative to the “treatmentLow” factor, or samples that displayed little to no bud failure. Differentially expressed genes were then joined to the concatenated gff3 file to determine genomic positions. Genes were then filtered into individual haplotypes, then cross-referenced by locus tag, or shared gene identifiers based off the Liftoff annotation, to determine allelic position of the genes, and where genes overlapped between haplotypes.

Hierarchical clustering was performed on samples using R package gg dendro v 0.20 (de Vries 2025) with all samples, then samples were filtered by coefficient of variation to determine 500 most variable genes, then clustering was performed again. Samples were then bootstrapped using R package pvclust v 2.20 (Suzuki 2019) using the “ward.D” method with n=100 to validate relationships between samples. Genes were also clustered using k-means clustering using r package cluster v 2.1.8.3 (Maechler 2026). The gap statistic was calculated, then k was set to k=3. Mean expression across chromosomes was also calculated to test chromosome level differences in expression. Rolling means of each known expressed gene was calculated using the R package zoo v 1.9(Zeileis 2005), then smoothed every 50 genes. A Wilcox signed-rank test was performed between the treatments at each chromosome.

Whole Genome Methylation and Differential Methylation Analysis

Global methylation read counts were extracted from the Bismark methylation call report. To calculate total methylation, the total number of methylated cytosines was divided by the total number of reads for cytosines of a given sample. For each context (CpG, CHG, and CHH), percent methylation was calculated by the number of methylated cytosines at that context divided by the total numbers of cytosines at that context. Results were loaded into R statistical software v. 4.2.1. Data was tested for normality using Shapiro-Wilks test and F-test was performed to test homogeneity of variance. A linear model was built using the `lm()` function in base R, then the `emmeans` of the total methylation was calculated using the R package `emmeans()` v2.04 (Lenth 2026) To analyze differences in total methylation between treatment levels, and ANOVA was performed between treatment levels at each of the contexts.

Differentially methylated regions were called using the R package DMRcaller version 1.41.0 (Catoni 2018). CX report files generated by Bismark were loaded into R and used to generate pooled Granges objects. To analyze pooled samples, `readBismarkPool()` was used to pool biological replicates into a single object, while individual sample reports were loaded using the `readBismark()` function. DMRs were called using the `filterDMRs()` function for each chromosome, based on annotation from the concatenated GFF3 file for Carmel loaded into R using `read.gtf()` from the package `genomation` v 1.44(Akalin 2014). DMRs were also called using the `computeDMRs()` function using the neighborhood method with a window size of 100 base pairs. Each included DMR had a minimum coverage of three, occurred over at least four cytosines, and had an adjusted p-value < 0.01 with multiple testing correction.

RNA Seq Analysis

After alignment, unnormalized count files were compiled into a single matrix by loading the individual STAR output into R statistical computing software. Samples were filtered by genes that had expression counts equal to zero, removing unexpressed genes. Differential gene analysis was conducted using the R

package edgeR v 3.23(Chen 2025). Samples were normalized by read depth, and dispersion between samples were calculated. A generalized linear model was then fit using a design matrix of the experimental design(One treatment, four levels, 32 total samples), then the model was fit relative to the “treatmentLow” factor, or samples that displayed little to no bud failure. Differentially expressed genes were then joined to the concatenated gff3 file to determine genomic positions. Genes were then filtered into individual haplotypes, then cross-referenced by locus tag to determine allelic position of the genes, and where genes overlapped between haplotypes.

Hierarchical clustering was performed on samples using R package gg dendro v 0.20(de Vries 2025) with all samples, then samples were filtered by coefficient of variation to determine 500 most variable genes, then clustering was performed again. Samples were then bootstrapped using R package pvclust v 2.20 (Suzuki 2019) using the “ward.D” method with n=100 to validate relationships between samples. Genes were also clustered using k-means clustering using r package cluster v 2.1.8.3 (Maechler 2026). The gap statistic was calculated, then k was set to k=3. Mean expression across chromosomes was also calculated to test chromosome level differences in expression. Rolling means of each known expressed gene was calculated using the R package zoo v 1.9(Zeileis 2005), then smoothed every 50 genes. A Wilcox signed-rank test was performed between the treatments at each chromosome.

Whole Genome Methylation and Differential Methylation Analysis

Global methylation read counts were extracted from the Bismark methylation call report. To calculate total methylation, the total number of methylated cytosines was divided by the total number of reads for cytosines of a given sample. For each context (CpG, CHG, and CHH), percent methylation was calculated by the number of methylated cytosines at that context divided by the total numbers of cytosines at that context. Results were loaded into R statistical software v. 4.2.1. Data were tested for normality using Shapiro-Wilks test and F-test was performed to test homogeneity of variance. A linear model was built using the lm() function in base R, then the emmeans of the total methylation was calculated using the R

package `emmeans()` v2.04 (Lenth 2026) To analyze differences in total methylation between treatment levels, and ANOVA was performed between treatment levels at each of the contexts.

Differentially methylated regions were called using the R package `DMRcaller` version 1.41.0 (Catoni 2018). CX report files generated by Bismark were loaded into R and used to generate pooled Granges objects. To analyze pooled samples, `readBismarkPool()` was used to pool biological replicates into a single object, while individual sample reports were loaded using the `readBismark()` function. DMRs were called using the `filterDMRs()` function for each chromosome, based on annotation from the concatenated GFF3 file for Carmel loaded into R using `read.gtf()` from the package `genomation` v 1.44 (Akalin 2014). DMRs were also called using the `computeDMRs()` function using the neighborhood method with a window size of 100 base pairs. Each included DMR had a minimum coverage of three, occurred over at least four cytosines, and had an adjusted p-value < 0.01 with multiple testing correction.

Results

Carmel Haplotype Phased Assembly and Annotation

PacBio HiFi reads and Oxford Nanopore reads were used in combination with trio-binning, which uses short reads to resolve the parental haplotypes of the parents, to create a new assembly for Carmel (Figure 2.1). 151 contigs were assembled for the “Texas” haplotype totaling 280Mb and 267 contigs totaling 296Mb for the “Nonpareil” haplotype. Chromosome level assembly and phasing was achieved by trio-binning using the ultra-long reads from the Nanopore sequencing to verify parental haplotypes. This resulted in eight chromosomes totaling 265Mb in length for the Texas haplotype and 268Mb in length for the Nonpareil haplotype. The haplotypes each scored a high BUSCO completeness of 97.8 percent with 2 contigs gaps for the Texas haplotype and 98.7 with 0 contig gaps for the Nonpareil haplotype. Using

Helixer and Liftoff from the Nonpareil annotation, 38,376 genes were identified in the Texas haplotype and 43,429 were identified in the Nonpareil haplotype.

3' TAG Seq Expression

Table 2.1. RNA Reads and Haplotype Mappability

<i>Chromosome</i>	<i>Hap1 Only Reads</i>	<i>Hap2 Only Reads</i>	<i>Both Haplotypes</i>	<i>Total Reads</i>
			<i>Read</i>	
<i>Chr1</i>	11.5	12.2	76.3	37224787
<i>Chr2</i>	8.2	55.5	36.3	24341470
<i>Chr3</i>	8.6	8.4	83.0	18534740
<i>Chr4</i>	54.2	7.1	38.2	26787922
<i>Chr5</i>	17.6	19.9	62.5	10267053
<i>Chr6</i>	14.6	14.0	71.4	21424671
<i>Chr7</i>	14.3	11.6	74.4	13862614
<i>Chr8</i>	7.2	15.8	77.0	24725568

Reads were mapped to the concatenated genome across both haplotypes (Table 2.1) Most reads mapped across both haplotypes, with an average of 64.2% mapping to both haplotypes, with an average of 17.6% of reads mapping to Texas haplotype and 18.1% of reads mapping uniquely to the Nonpareil haplotype. However, Chromosome 2 and Chromosome 4 demonstrate a bias to one of the two haplotypes. Chromosome 2 has a higher percentage of unique reads mapped to the Nonpareil haplotype (55.5%), while Chromosome 4 has a higher percentage (54.2%) of reads mapped uniquely to the Texas haplotype.

To test for differential gene expression, we tested the different levels of bud failure against the “low” treatment, which showed little or no symptoms of NBF (Figure 2.3). 2093 differentially

expressed genes relative to the low NBF treatment were found, out of 11494 total expressed genes in the samples. 981 of these genes are downregulated and 1102 were up-regulated. Of the differentially expressed genes, 971 of those genes were aligned to the Texas haplotype, with 448 down-regulated and 523 up-regulated, and 1102 were associated with the Nonpareil haplotype, with 497 down regulated and 543 up regulated. The top differentially expressed genes based on p value and false discovery rate ($\alpha=0.05$) tended to have high expression in the “low” samples and declined in gene expression as the severity of bud failure increased. To examine how genes aligned to

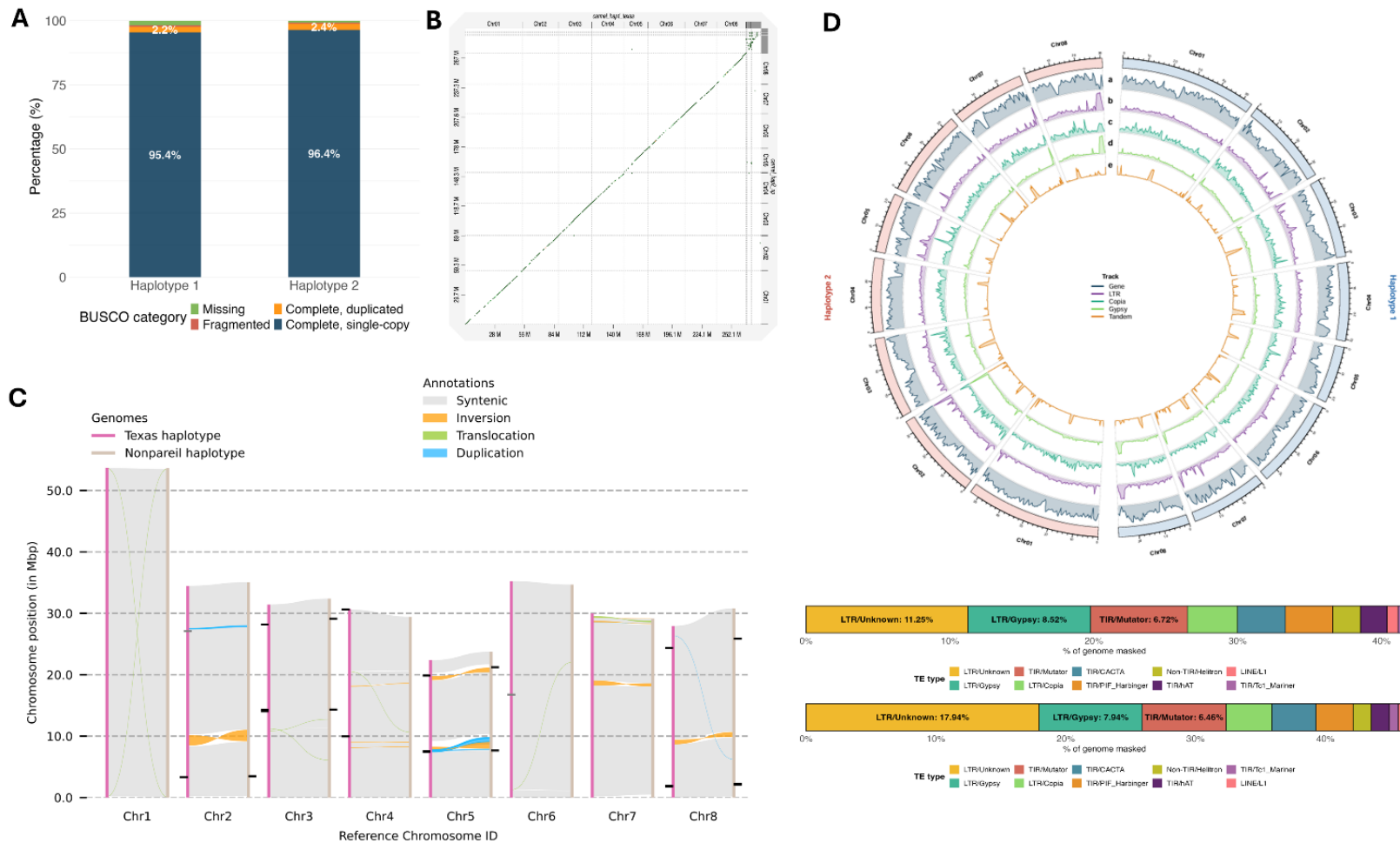


Figure 2.1. Assembly Statistics of the Carmel genome a) BUSCO scores of the Texas Haplotype (Haplotype1) and Nonpareil Haplotype(Haplotype2). B)Contig alignment between the two haplotypes. C) Syntenic areas between the two haplotypes. Orange areas represent inversion area between two haplotypes, blue represents duplication, and green is translocation. D) Transposable elements (TE) distribution and composition throughout the genome.

individual haplotypes compared to one another, associated locus tags, or non-haplotype specific gene identifiers from the liftoff annotation, were cross referenced between genes associated with each haplotype. In total, 557 genes shared locus tags between the two haplotypes, while 312 genes had unique locus tags for the Texas haplotype and 367 genes associated with the Nonpareil haplotype had unique locus tags. Using the total differentially expressed genes, genes were organized by the coefficient of variation, then the 500 most variable genes were subset to test their association. Hierarchical clustering was used to test dissimilarities between samples. Samples were broken into primary clades with “low” and “medium low” samples occupying one clade, and “medium high” and “high” samples occupying the other clade. Subclades had groups of samples within the same treatment group clustering together. These dissimilarities were validated by bootstrapping the samples, and then hierarchical clustering was performed again.

Whole Genome Methylation

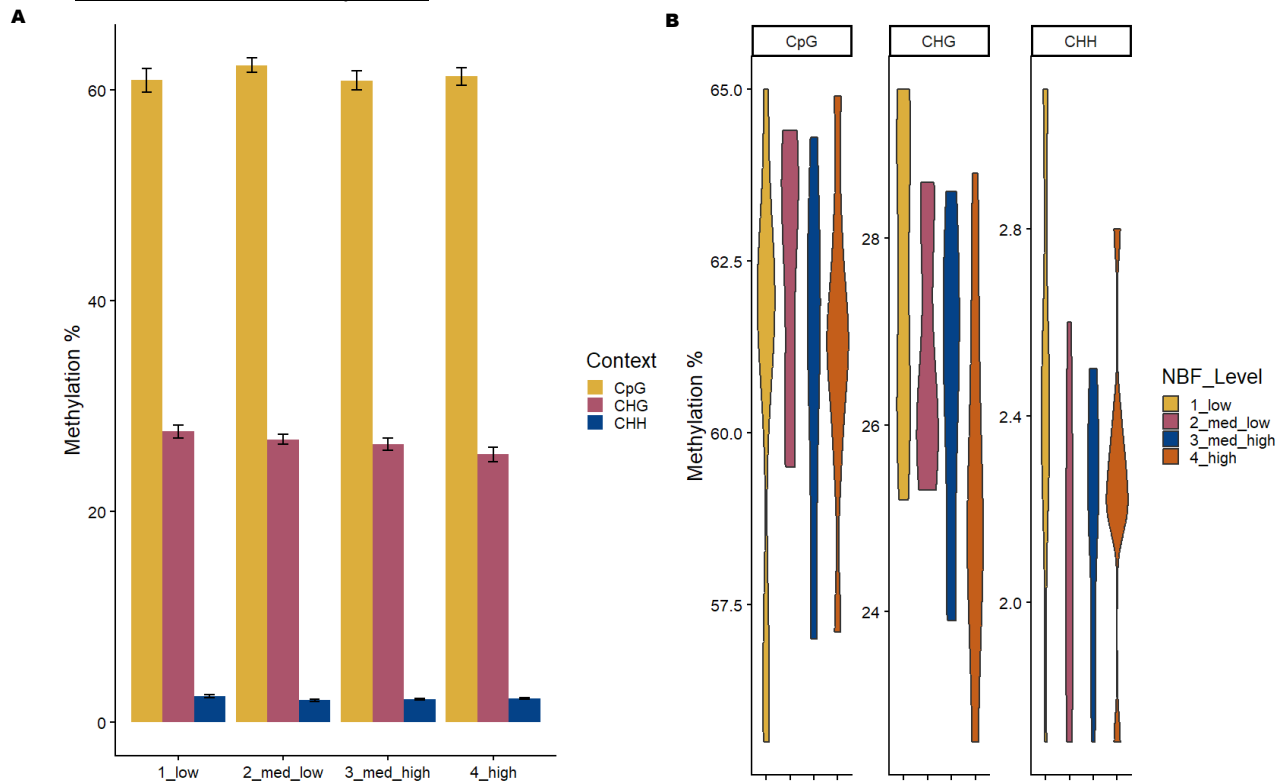


Figure 2.2. Genome-wide DNA methylation a) Mean Genome-wide DNA methylation across all three contexts for each level of NBF. B) Violin plots for genome-wide DNA methylation for each level of NBF across all three contexts.

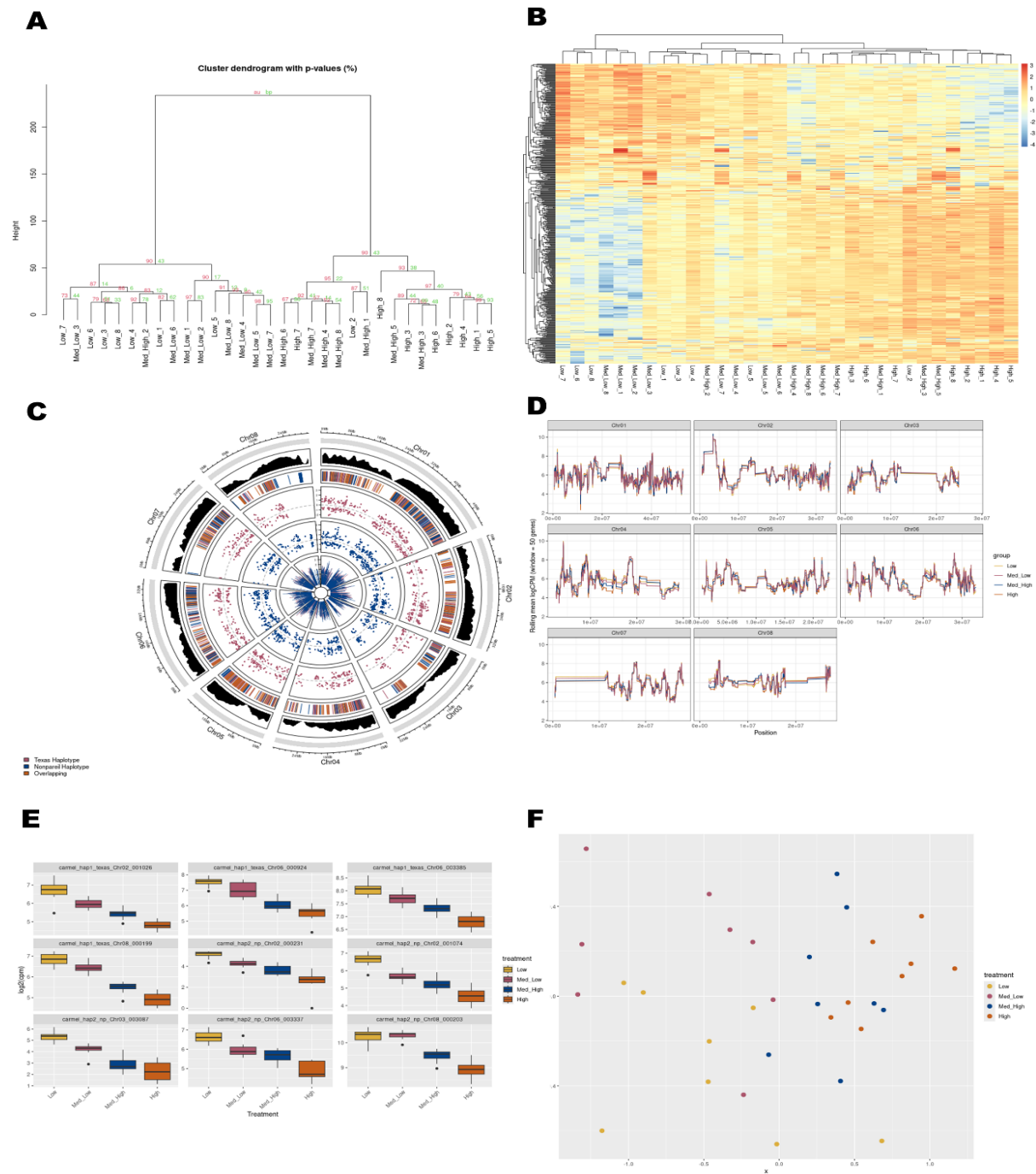


Figure 2.3: Gene Expression in leaf tissue of Carmel showing four different levels of bud failure a) Bootstrapped dendrogram between the RNA seq samples, with the majority broken down between two clades. Green values are the bootstrap percentage values and red values are the approximate unbiased values. b) Heatmap showing over and under expression relative to the “low” treatment of 500 most variable genes. c) Circos plot showing gene density, then distribution of differentially expressed genes and whether they aligned to specific haplotypes or both. d) Rolling means for the expression of every 50 genes showing chromosome-level differences in expression. e) Nine most significant differentially expressed genes of all differentially expressed genes f) MDS plot of all differentially expressed genes

contexts (Figure 2.2). CpG had a mean methylation value of 60.9% for the low treatment, 62.3% for the medium low, 60.9% for the medium high and 61.2% for high. Anova had a pvalue of 0.63. CHG had a mean methylation value of 27.6% for the low treatment, 26%.8 for the medium low, 26.4% for the medium high and 25.4% for high. The anova had a pvalue of 0.09. CHH had a mean methylation value of 2.5% for the low treatment, 2.1% for the medium low, 2.2% for the medium high and 2.3% for high. The anova had a pvalue of 0.18.

Differential Methylation

Based on the signals from the 3'TAG RNA seq data, samples were grouped into a “positive” group consisting of medium-high and high samples and a “negative” group consisting of the medium-low and low samples. Samples were pooled and comparisons were made between the two groups relative to the negative group. Regions were determined by known regions in the concatenated annotation between the two haplotypes. Additionally, DMRs were called using 100bp windows under the same parameters.

There were 742 DMRs called from the 100bp bins: 393 in the CpG context, 348 in the CHG context and 1 in the CHH context. 399 of these DMRs showed a gain of methylation in the positive cohort, while 343 showed a loss of methylation. There were 20 total unique DMRs between the two groups using the genes as regions: 15 in the CpG context, four in the CHG context and one in the CHH context. Of these DMRs, 13 showed hypermethylation and seven displayed hypomethylation in the positive group. 15 of these genes mapped to exons, 11 in the CpG context and four in the CHG context. The remaining five DMRs mapped to the 5' UTRs: four in the CpG context and one in the CHH context. These DMRs mapped uniquely to each haplotype, with 11 on the Texas haplotype and 9 on the Nonpareil haplotype. DMRs were found on both gene-body exons and the 5' UTR regions of certain genes, with 15 DMRs called on exons and 5 called on 5' UTR.

To infer a relationship between the DMRs and differentially expressed genes, overlaps between the two datasets were examined (Figure 2.4). DMR positions were compared to the genomic position of differentially expressed genes. First, direct overlaps between DMRs and differentially expressed gene bodies were tested. Next, proximity between DMRs and genes were tested by identifying DMRs that fell within 5000 base pairs up or downstream of a differentially expressed gene.

Of the 20 DMRs identified from the annotation, none overlapped directly with any corresponding differentially expressed genes. However, 3 DMRs were in proximity to differentially expressed genes. On the Texas haplotype, gene Chr01_004284 had loss methylation at a DMR 3175bp upstream of the start in the positive cohort, which corresponds to an up regulation of the gene in the “high” treatment. Gene Chr04_002016 or L3X38_023430 had a gain in methylation 1602bp upstream at the end of the gene. This gene was upregulated in the “high” treatments, and it likely encodes a peptidyl-prolyl cis-trans isomerase CYP37chloroplastic [*Prunus dulcis*], which facilitates protein folding during transcription and peptide buildup. In the Nonpareil Haplotype, Gene Chr05_000303 or L3X38_027468 also displayed a gain in methylation downstream from the gene body. However, this gene showed a decrease in expression in the high treatment. It encoded an uncharacterized protein.

When we examined the DMRs from the 100bp bins, there were a total of seven DMRs that directly overlap with genes, and 32 that were proximal to differentially expressed genes. 5 of these occurred in the CpG context, while only two occurred in the CHG context. Of the 32 DMRs within proximity of DE genes, 11 occurred upstream of differentially expressed genes and the remaining 21 occurred downstream. 27 of these DMRs occurred in the CpG context, 5 occurred in CHG and none occurred in CHH. Between the DMRs called with the gene annotation and the 100bp bins, there were overlapping hits on Chr04_002026 and Chr05_000303.

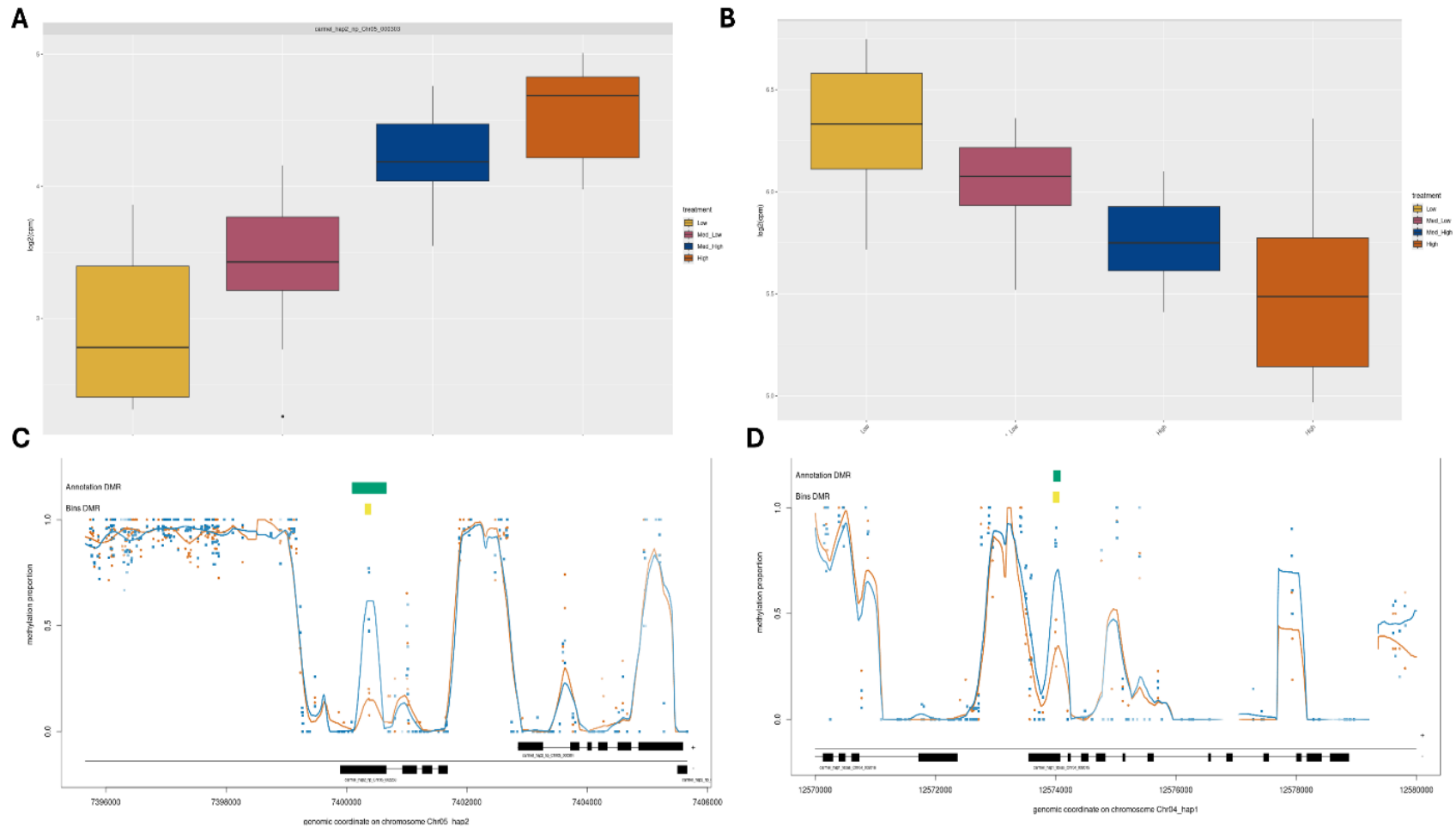


Figure 2.4. Overlap of DNA methylation and differential gene expression at two genes with significant differences between NBF levels in both. (a,b) Box plots showing gene counts of DE genes between the different levels of bud failure. (c,d) Methylation profiles of the same genes over 10000 bp, where the Orange line is NBF-negative and the blue line is NBF-positive, and the boxes show the DMRs called, where the green box is a DMR called using the annotation, and the yellow box is a DMR called using a 100bp bin.

Function of Differentially Expressed Genes

To test the functional enrichment and gain some insight into the differentially expressed genes between the treatment groups, we performed gene ontology (GO) analysis to test enrichment of differentially expressed genes against the universe of all expressed genes (Figure 2.5). In total, there were 43 significant GO terms between the differentially expressed genes. Translation composed the largest section of genes, 162 significantly differentially expressed genes of the 392 in the total list of expressed Ribosomal activity and assembly also composed a large set of differentially expressed genes, with 10 differentially expressed genes of the 12 of all expressed. Regulation of malate metabolic processes composed the next largest subunit, with 6 of the 8 genes associated with the processes were found to be significant.

Discussion

This study directly observed phenotypic differences in NBF severity between genetic clones, then used sequencing data to explore a possible underlying genetic cause. A statistically robust sample dataset was collected, and confounding factors were reduced by using a single commercial orchard and focusing on a single variety that commonly show NBF, strengthening the analysis relative to previous studies. To improve the fidelity and accuracy of these analyses, the most complete almond genome to date using a combination of long-read sequencing technologies.

Differential expression is correlated with NBF severity

The results show differences in DNA methylation and differential gene expression that correlated to differences in phenotype. Among top differentially expressed genes, there was a clear trend of decreased gene expression as the severity of bud failure increased. Generally, these genes tend to cluster together based on the treatment level, with two broad trends of a “negative” group, or the low and medium-low

samples, and the “positive” group of high and medium-high treatment levels. Narrowing these down to the most variable genes, the subgroups cluster by the replicates of each group, with some overlap contained within those larger macro groups. These clusters remain in this order when the samples are bootstrapped, providing some validation to the relationships between the treatment levels. The result suggests there is a relationship between gene expression and severity of bud failure, both on an individual level and on a larger gene scale as well. There were two noteworthy exceptions to these general trends: low2 and medium-high5. These two samples clustered with the opposite macrogroup than other samples of the same cohort. These outliers could be the result of incorrect scoring of the phenotype or mislabeling. However, these outliers could also be some sort of phase transition whereby differences in gene expression may preclude an observable difference in the phenotype. GO analysis suggests large disruptions to translation between the treatment levels, as well as disruptions to malate metabolic processes. Disruptions to ribosomal activity have been associated with stress response in several plant species, such as ryegrass (Zha 2025) and *Nicotiana benthamiana* (Fakih 2023). In *Arabidopsis*, disruption of pre-rRNA processing has occurred during heat stress (Darriere 2022). Malate metabolic processes regulate many processes in horticultural plants like apple, and play important roles in fruit quality and stress response (Ting-Zhao 2025). NBF severity has been correlated with mean summer temperatures and could be part of some complex or disrupted stress response mechanism in the plant.

Figure 2.5. Gene ontology results from the differentially expressed genes. A) A scatterplot of gene ontology terms distributed by the first two dimensions of multi-dimensional scaling, and point size representing significance. Colors differentiate difference GO terms and groups. B) Treemap showing nested rectangles, where the color represent similar GO terms and size represent significance.

Genes are differentially methylated between positive and negative cohorts

There was no significant differences in global methylation between the different levels of NBF. There was a small trend of lower methylation as NBF increased but failed to meet a statistically significant threshold. Further studies may want to increase sample size even further to explore these differences in global methylation. When grouping the samples into positive and negative categories, there were significant differences in methylation. These differences were observed using known genes as predefined regions, but also in smaller windows irrespective of known gene bodies as well. To determine whether there was a relationship between the differences in methylation and gene expression, the positions of DMRs were referenced against the position of known differentially expressed genes. When examining the DMRs taken from using known genes, there was no direct overlap between regions. However, taking proximity into account, three genes emerge as being within 5000bp of differentially expressed genes. Even more DMRs overlap within 5000bp of differentially expressed genes when DMRs were called using 100bp bins. However, the biological relationship between these data is unclear. While some DMRs with a loss of methylation in the positive group corresponded with an increase of expression in samples with higher levels of NBF and vice versa, an expected relationship, there was a roughly equal number of DMRs that had the opposite relationship to gene expression. While this atypical relationship between gene expression and methylation has been reported (Lu 2015), it muddies any sort of definitive conclusion between the relationship between NBF, gene expression and methylation. These regions could be the area for further study, including the development of assays for NBF or targeted methylation of alleles or methylation state.

Differential methylation is an additional symptom of NBF

Because of the variation between clones, sexual and asexual heritability, the influence of the environment on the phenotype, a heritable epigenetic mark such as DNA methylation seemed like a likely culprit.

Evidence in other tree species, such as Lombardy poplar, has shown that significant changes in the epigenome can result in observable differences in the phenotype (Broeck 2018). However, our results have significantly fewer differentially methylated regions than previous studies on NBF. These studies also didn't observe differences in methylation directly and instead used a proxy phenotype between peach almond hybrids (D'Amico Willman 2022). Fresnedo-Ramirez looked at genome methylation differences between varieties with different levels of bud failure, as well as different varieties with histories of bud failure. However, methylation can be affected by genotype and can have a stronger effect than environmental components (Yeo 2026). Even a few single nucleotide polymorphisms can affect the accrual of epimutations relative to other lines in the same lineage (Becker 2011). Therefore, it is plausible that the observed differences in methylation and bud failure between the varieties was a result of difference in the varieties, rather than an underlying epigenetic mechanism of NBF affecting the methylome. Differences in DNA methylation could have also been the result of insufficient data, such as low coverage of sites leading to noise from whole-genome bisulfite sequencing or the lack of biological reps leading to greater differences in DNA methylation (Ziller 2013). Often this level of replication can be difficult due to the relatively high cost whole-genome bisulfite sequencing (Agius 2023). As replication increases and other confounding factors are controlled for though, the relative difference between samples is small and are largely driven by genotype (Hazzard 2026). Additionally, the biological connection between the changes in methylation, variation in expression and the resulting phenotypic differences remain unclear. Further results need to establish a stronger connection between phenotypic differences and differences found through sequencing, including improving methods for phenotyping NBF with a higher degree of accuracy.

Alternative hypotheses of NBF

The significant differences in DNA methylation between the different phenotypes and gene expression may not be the cause of bud failure but could potentially be another symptom. NBF symptoms could be the result of some previously undetected structural variation. Structural variation is an unexplored source

of plant phenotypic diversity, but new sequencing technologies have opened the possibility of exploration (Gabur 2019). In several studies in humans, a connection has been established between DNA methylation, differential expression and structural variants. Blood samples in over 7000 individuals revealed sequence variants correlated with CpG rates and downstream mRNA expression (Stefansson 2024). In brain cell tumors, structural variants were found to be within proximity of CpG islands showing differential methylation and often showed unexpected patterns of downstream changes in gene expression (Chen 2025). In plants, structural variants had bi-directional effects on DNA methylation regulating transcription binding factors motifs in *Brassica oleracea*, serving to modulate influences in different morphotypes (Li 2024). In cotton, methylation changes associated with structural variants had a stronger influence on phenotype than methylation changes alone (Shao 2025). Therefore, based on the differences in gene expression and methylation between clones, it seems plausible that NBF may be related to some undetected structural variant. Another possibility could lie in tandem repeat disorders, a group of genetic disorders resulting from changes in the number of microsatellite DNA repeat units that disrupt transcription (Moon 2024). NBF has similar characteristics to Huntington's disease, which is caused by expansion of a trinucleotide repeat that is inversely correlated with age and shows heritability (Hong 2024). A triplet repeat expansion was shown to cause a detrimental phenotype in *Arabidopsis* (Sureshkumar 2009) causing epigenetic silencing (Sureshkumar 2025). Further exploration of these tandem repeats in almond will be necessary to explore this plausible alternative hypothesis to methylation.

Conclusion

Gene expression data is correlated with changes in the severity of NBF. In the most significant differentially expressed genes, there is a trend of decreasing expression as the severity of NBF increases, and samples cluster together into “positive and negative groups”. While there are no significant

differences in genome-wide methylation between treatment levels, there are significant areas of differential methylation specific to each haplotype. Some of the differential gene expression may be related to the areas of differential methylation, however more experiments will be needed to confirm the biological connection and resulting effect on phenotype. These could serve as potential sites for future assays of NBF. Methylation itself may not be the direct cause of NBF and be the result of other genomic or epigenomic differences, such as structural variants or expansion of tandem repeats, that result in these observable differences in gene expression, DNA methylation, and phenotype.

References

1. Alioto, T., Alexiou, K. G., Bardil, A., Barteri, F., Castanera, R., Cruz, F., Dhingra, A., Duval, H., Fernández i Martí, Á., Frias, L., Galán, B., García, J. L., Howad, W., Gómez-Garrido, J., Gut, M., Julca, I., Morata, J., Puigdomènech, P., Ribeca, P., ... Arús, P. (2020). Transposons played a major role in the diversification between the closely related almond and peach genomes: Results from the almond genome sequence. *The Plant Journal*, 101(2), 455–472. <https://doi.org/10.1111/tpj.14538>
2. Alonge, M., Lebeigle, L., Kirsche, M., Jenike, K., Ou, S., Aganezov, S., Wang, X., Lippman, Z. B., Schatz, M. C., & Soyk, S. (2022). Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biology*, 23, 258. <https://doi.org/10.1186/s13059-022-02823-7>
3. Ariel Lou Fenton, C., Kester, D. E., & Liu, L. (1988). The temperature response of cultured almond cells derived from tree sources with and without non-infectious bud failure. I. Specific growth rate. *Plant Science*, 56(3), 239–243. [https://doi.org/10.1016/0168-9452\(88\)90103-3](https://doi.org/10.1016/0168-9452(88)90103-3)
4. Becker, C., Hagmann, J., Müller, J., Koenig, D., Stegle, O., Borgwardt, K., & Weigel, D. (2011). Spontaneous epigenetic variation in the Arabidopsis thaliana methylome. *Nature*, 480(7376), 245–249. <https://doi.org/10.1038/nature10555>
5. Cabanettes, F., & Klopp, C. (2018). D-GENIES: Dot plot large genomes in an interactive, efficient and simple way. *PeerJ*, 6, e4958. <https://doi.org/10.7717/peerj.4958>
6. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, 38(12), 5825–5829. <https://doi.org/10.1093/molbev/msab293>
7. Chen, F., Zhang, Y., Li, W., Sedlazeck, F. J., Shen, L., & Creighton, C. J. (2025). Global DNA methylation differences involving germline structural variation impact gene expression in pediatric brain tumors. *Nature Communications*, 16(1), 4713. <https://doi.org/10.1038/s41467-025-60110-y>
8. Cheng, H., Asri, M., Lucas, J., Koren, S., & Li, H. (2024). Scalable telomere-to-telomere assembly for diploid and polyploid genomes with double graph. *Nature Methods*, 21(6), 967–970. <https://doi.org/10.1038/s41592-024-02269-8>
9. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., & Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods*, 18(2), 170–175. <https://doi.org/10.1038/s41592-020-01056-5>

10. Cheng, H., Jarvis, E. D., Fedrigo, O., Koepfli, K.-P., Urban, L., Gemmell, N. J., & Li, H. (2022). Haplotype-resolved assembly of diploid genomes without parental data. *Nature Biotechnology*, 10.1038/s41587-022-01261-x. <https://doi.org/10.1038/s41587-022-01261-x>
11. D'Amico-Willman, K. M., Ouma, W. Z., Meulia, T., Sideli, G. M., Gradziel, T. M., & Fresnedo-Ramírez, J. (2022). Whole-genome sequence and methylome profiling of the almond [*Prunus dulcis* (Mill.) D.A. Webb] cultivar 'Nonpareil.' *G3: Genes|Genomes|Genetics*, 12(5). <https://doi.org/10.1093/G3JOURNAL/JKAC065>
12. D'Amico-Willman, K. M., Sideli, G. M., Allen, B. J., Anderson, E. S., Gradziel, T. M., & Fresnedo-Ramírez, J. (2022). Identification of Putative Markers of Non-infectious Bud Failure in Almond [*Prunus dulcis* (Mill.) D.A. Webb] Through Genome Wide DNA Methylation Profiling and Gene Expression Analysis in an Almond × Peach Hybrid Population. *Frontiers in Plant Science*, 13(February). <https://doi.org/10.3389/fpls.2022.804145>
13. Darriere, T., Jobet, E., Zavala, D., Escande, M. L., Durut, N., de Bures, A., Blanco-Herrera, F., Vidal, E. A., Rompais, M., Carapito, C., Gourbiere, S., & Sáez-Vásquez, J. (n.d.). Upon heat stress processing of ribosomal RNA precursors into mature rRNAs is compromised after cleavage at primary P site in *Arabidopsis thaliana*. *RNA Biology*, 19(1), 719–734. <https://doi.org/10.1080/15476286.2022.2071517>
14. Fakh, Z., Plourde, M. B., & Germain, H. (2023). Differential Participation of Plant Ribosomal Proteins from the Small Ribosomal Subunit in Protein Translation under Stress. *Biomolecules*, 13(7), 1160. <https://doi.org/10.3390/biom13071160>
15. Fresnedo-Ramírez, J., Chan, H. M., Parfitt, D. E., Crisosto, C. H., & Gradziel, T. M. (2017). Genome-wide DNA-(de)methylation is associated with Noninfectious Bud-failure exhibition in Almond (*Prunus dulcis* [Mill.] D.A. Webb). *Scientific Reports* 2017 7:1, 7(1), 1–12. <https://doi.org/10.1038/srep42686>
16. Gabur, I., Chawla, H. S., Snowdon, R. J., & Parkin, I. A. P. (2019). Connecting genome structural variation with complex traits in crop plants. *Theoretical and Applied Genetics*, 132(3), 733–750. <https://doi.org/10.1007/s00122-018-3233-0>
17. Goel, M., & Schneeberger, K. (2022). plotsr: Visualizing structural similarities and rearrangements between multiple genomes. *Bioinformatics*, 38(10), 2922–2926. <https://doi.org/10.1093/bioinformatics/btac196>
18. Goel, M., Sun, H., Jiao, W.-B., & Schneeberger, K. (2019). SyRI: Finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biology*, 20, 277. <https://doi.org/10.1186/s13059-019-1911-0>
19. Gradziel, T., & Fresnedo, J. (2019). Noninfectious Bud-failure As a Model For Studying Age Related Genetic Disorders in Long-Lived Perennial Plants. *Journal of the American Pomological Society*, 73(4), 240–253.
20. Gradziel, T., Lampinen, B., & Preece, J. E. (2019). Propagation from basal epicormic meristems remediates an aging-related disorder in almond clones. *Horticulturae*, 5(2), 1–9. <https://doi.org/10.3390/horticulturae5020028>
21. Hazzard, W., Fresnedo-Ramírez, J., Anderson, E., Brown, P. J., & Gradziel, T. (2026). DNA methylation is conserved between vegetative propagules in almond (*Prunus dulcis* Mill D.A. Webb). *European Journal of Horticultural Science*, 91(1). <https://doi.org/10.1079/ejhs.2026.0007>
22. Hellali, R., Lin, J., & Kester, D. E. (1978). Morphology of Noninfectious Bud Failure Symptoms in Vegetative Buds of Almond (*Prunus amygdalus* Batsch)1. *Journal of the American Society for Horticultural Science*, 103(4), 459–464. <https://doi.org/10.21273/JASHS.103.4.459>
23. Hong, E. P., Ramos, E. M., Aziz, N. A., Massey, T. H., McAllister, B., Lobanov, S., Jones, L., Holmans, P., Kwak, S., Orth, M., Ciosi, M., Lomeikaite, V., Monckton, D. G., Long, J. D., Lucente, D., Wheeler, V. C., Gillis, T., MacDonald, M. E., Sequeiros, J., ... Lee, J.-M. (2024). Modification of Huntington's disease by short tandem repeats. *Brain Communications*, 6(2), fcae016. <https://doi.org/10.1093/braincomms/fcae016>

24. Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., & Hunter, S. (2014a). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, 30(9), 1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>
25. Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., & Hunter, S. (2014b). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, 30(9), 1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>
26. Kester, D. E. (n.d.). *NONINFECTIOUS BUD FAILURE (1) The nature and origin*.
27. Kester, D. E., & Say2, R. N. A. (1978). Variability in Noninfectious Bud-failure of “Nonpareil” Almond. I. Location and Environment1. In *J. Amer. Soc. Hort. Sci* (Vol. 103, Issue 3, pp. 377–382).
28. Kester, D. E., Shackel, K. A., Micke, W. C., Viveros, M., & Gradziel, T. M. (2004). Noninfectious Bud Failure in “Carmel” Almond: I. Pattern of Development in Vegetative Progeny Trees. *Journal of the American Society for Horticultural Science*, 129(2), 244–249. <https://doi.org/10.21273/jashs.129.2.0244>
29. Kester2, E., & Jones3, R. W. (n.d.). *Noninfectious Bud-Failure from Breeding Programs of Almond (Prunus amygdalus Batsch)1*.
30. Kuznetsov, D., Tegenfeldt, F., Manni, M., Seppey, M., Berkeley, M., Kriventseva, E. V., & Zdobnov, E. M. (2022). OrthoDB v11: Annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Research*, 51(D1), D445–D451. <https://doi.org/10.1093/nar/gkac998>
31. Lang, Z., Wang, Y., Tang, K., Tang, D., Datsenka, T., Cheng, J., Zhang, Y., Handa, A. K., & Zhu, J. K. (2017). Critical roles of DNA demethylation in the activation of ripening-induced genes and inhibition of ripening-repressed genes in tomato fruit. *Proceedings of the National Academy of Sciences of the United States of America*, 114(22), E4511–E4519. <https://doi.org/10.1073/PNAS.1705233114;WGROU:STRING:PUBLICATION>
32. Li, H. L., Wang, Y., Guo, D., Zhu, J. H., Wang, Y., Dai, H. F., & Peng, S. Q. (2024). Reprogramming of DNA methylation and changes of gene expression in grafted Hevea brasiliensis. *Frontiers in Plant Science*, 15, 1407700. <https://doi.org/10.3389/FPLS.2024.1407700/BIBTEX>
33. Li, Q., Qiao, X., Li, L., Gu, C., Yin, H., Qi, K., Xie, Z., Yang, S., Zhao, Q., Wang, Z., Yang, Y., Pan, J., Li, H., Wang, J., Wang, C., Rieseberg, L. H., Zhang, S., & Tao, S. (2024). Haplotype-resolved T2T genome assemblies and pangenome graph of pear reveal diverse patterns of allele-specific expression and the genomic basis of fruit quality traits. *Plant Communications*, 5(10). <https://doi.org/10.1016/j.xplc.2024.101000>
34. Li, X., Wang, Y., Cai, C., Ji, J., Han, F., Zhang, L., Chen, S., Zhang, L., Yang, Y., Tang, Q., Bucher, J., Wang, X., Yang, L., Zhuang, M., Zhang, K., Lv, H., Bonnema, G., Zhang, Y., & Cheng, F. (2024). Large-scale gene expression alterations introduced by structural variation drive morphotype diversification in Brassica oleracea. *Nature Genetics*, 56(3), 517–529. <https://doi.org/10.1038/s41588-024-01655-4>
35. Liu, Y., Rosikiewicz, W., Pan, Z., Jillette, N., Wang, P., Taghbalout, A., Foox, J., Mason, C., Carroll, M., Cheng, A., & Li, S. (2021). DNA methylation-calling tools for Oxford Nanopore sequencing: A survey and human epigenome-wide evaluation. *Genome Biology*, 22(1), 295. <https://doi.org/10.1186/s13059-021-02510-z>
36. López, M.-E., Roquis, D., Becker, C., Denoyes, B., & Bucher, E. (n.d.). *DNA methylation dynamics during stress response in woodland strawberry (Fragaria vesca)*. <https://doi.org/10.1093/hr/uhac174>
37. Lu, X., Wang, W., Ren, W., Chai, Z., Guo, W., Chen, R., Wang, L., Zhao, J., Lang, Z., Fan, Y., Zhao, J., & Zhang, C. (2015). *Genome-Wide Epigenetic Regulation of Gene Transcription in Maize Seeds*. <https://doi.org/10.1371/journal.pone.0139582>

38. Manni, M., Berkeley, M. R., Seppey, M., & Zdobnov, E. M. (2021). BUSCO: Assessing Genomic Data Quality and Beyond. *Current Protocols*, 1(12), e323. <https://doi.org/10.1002/cpz1.323>
39. Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, 27(6), 764–770. <https://doi.org/10.1093/bioinformatics/btr011>
40. Moon, J. (2025). Tandem repeat disorders: From diagnosis to emerging therapeutic strategies. *Encephalitis*, 5(2), 27–35. <https://doi.org/10.47936/encephalitis.2024.00122>
41. Ni, P., Nie, F., Zhong, Z., Xu, J., Huang, N., Zhang, J., Zhao, H., Zou, Y., Huang, Y., Li, J., Xiao, C.-L., Luo, F., & Wang, J. (n.d.). DNA 5-methylcytosine detection and methylation phasing using PacBio circular consensus sequencing. <https://doi.org/10.1038/s41467-023-39784-9>
42. Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J. R. A., Hellings, A. J., Lugo, C. S. B., Elliott, T. A., Ware, D., Peterson, T., Jiang, N., Hirsch, C. N., & Hufford, M. B. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology*, 20, 275. <https://doi.org/10.1186/s13059-019-1905-y>
43. Poetsch, A. R., & Plass, C. (2011). *Transcriptional regulation by DNA methylation*. <https://doi.org/10.1016/j.ctrv.2011.04.010>
44. R Core Team (2024) *R A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. - References—Scientific Research Publishing. (n.d.). Retrieved August 24, 2026, from <https://www.scirp.org/reference/referencespapers?referenceid=3887298>
45. Ranallo-Benavidez, T. R., Jaron, K. S., & Schatz, M. C. (2020). GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications*, 11, 1432. <https://doi.org/10.1038/s41467-020-14998-3>
46. Rhie, A., Walenz, B. P., Koren, S., & Phillippy, A. M. (2020). Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology*, 21, 245. <https://doi.org/10.1186/s13059-020-02134-9>
47. Shao, L., Jin, S., Jiang, H., Pan, T., Rui, Z., Shi, X., Zhao, T., Si, Z., Guan, X., Hu, Y., Zhang, T., & Fang, L. (2026). Structural Variation and 3D Genome-Driven DNA/RNA Methylation Divergence Contributing to Cotton Fiber Domestication. *Advanced Science*, 13(9), e14381. <https://doi.org/10.1002/advs.202514381>
48. Shen, W., Le, S., Li, Y., & Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLOS ONE*, 11(10), e0163962. <https://doi.org/10.1371/journal.pone.0163962>
49. Shumate, A., & Salzberg, S. L. (2021). Liftoff: Accurate mapping of gene annotations. *Bioinformatics*, 37(12), 1639–1643. <https://doi.org/10.1093/bioinformatics/btaa1016>
50. Stefansson, O. A., Sigurpalsdottir, B. D., Rognvaldsson, S., Halldorsson, G. H., Juliusson, K., Sveinbjornsson, G., Gunnarsson, B., Beyter, D., Jonsson, H., Gudjonsson, S. A., Olafsdottir, T. A., Saevarsdottir, S., Magnusson, M. K., Lund, S. H., Tragante, V., Oddsson, A., Hardarson, M. T., Eggertsson, H. P., Gudmundsson, R. L., ... Stefansson, K. (2024). The correlation between CpG methylation and gene expression is driven by sequence variants. *Nature Genetics*, 56(8), 1624–1631. <https://doi.org/10.1038/s41588-024-01851-2>
51. Sun, H., Jiao, W.-B., Krause, K., Campoy, J. A., Goel, M., Folz-Donahue, K., Kukat, C., Huettel, B., & Schneeberger, K. (2022). Chromosome-scale and haplotype-resolved genome assembly of a tetraploid potato cultivar. *Nature Genetics*, 54(3), 342–348. <https://doi.org/10.1038/s41588-022-01015-0>
52. Sureshkumar, S., Bandaranayake, C., Lv, J., Dent, C. I., Bhagat, P. K., Mukherjee, S., Sarwade, R., Atri, C., York, H. M., Tamizhselvan, P., Shamaya, N., Folini, G., Bergey, B. G., Yadav, A. S., Kumar, S., Grummisch, O. S., Saini, P., Yadav, R. K., Arumugam, S., ... Balasubramanian, S. (2024). SUMO protease FUG1, histone reader AL3 and chromodomain protein LHP1 are integral to repeat expansion-induced gene silencing in *Arabidopsis thaliana*. *Nature Plants*, 10(5), 749–759. <https://doi.org/10.1038/s41477-024-01672-5>

53. Sureshkumar, S., Todesco, M., Schneeberger, K., Harilal, R., Balasubramanian, S., & Weigel, D. (2009). A Genetic Defect Caused by a Triplet Repeat Expansion in *Arabidopsis thaliana*. *Science*, 323(5917), 1060–1063. <https://doi.org/10.1126/science.1164014>
54. Vanden Broeck, A., Cox, K., Brys, R., Castiglione, S., Ciccattelli, A., Guarino, F., Heinze, B., Steenackers, M., & Vander Mijnsbrugge, K. (2018). Variability in DNA Methylation and Generational Plasticity in the Lombardy Poplar, a Single Genotype Worldwide Distributed Since the Eighteenth Century. *Frontiers in Plant Science*, 9. <https://doi.org/10.3389/fpls.2018.01635>
55. Vilanova, S., Alonso, D., Gramazio, P., Plazas, M., García-Forte, E., Ferrante, P., Schmidt, M., Díez, M. J., Usadel, B., Giuliano, G., & Prohens, J. (2020). SILEX: a fast and inexpensive high-quality DNA extraction method suitable for multiple sequencing platforms and recalcitrant plant species. *Plant Methods*, 16(1). <https://doi.org/10.1186/S13007-020-00652-Y>
56. Yeo, D. S. G., Eydam, J., Bruckmüller, J., Kauder, F., Lübeck, J., Kaier, A., Sonnewald, S., & Sonnewald, U. (2026). Genotype-dependent DNA methylation patterns are negatively associated with allelic variation rather than heat-induced gene expression in two contrasting potato genotypes. *The Plant Journal*, 125(2), e70690. <https://doi.org/10.1111/tpj.70690>
57. Zha, W., Zeng, L., Zhang, Z., Zhang, J., & Xu, B. (2025). Comparative transcriptomic analysis revealed specific modules and hub heat-responsive genes in perennial ryegrass. *Grass Research*, 5(1), 0–0. <https://doi.org/10.48130/grares-0025-0021>
58. Zhang, H., Lang, Z., & Zhu, J. K. (2018). Dynamics and function of DNA methylation in plants. *Nature Reviews Molecular Cell Biology*, 19(8), 489–506. <https://doi.org/10.1038/s41580-018-0016-z>
59. Zhang, H. Y., Zhao, H. X., Wu, S. H., Huang, F., Wu, K. T., Zeng, X. F., Chen, X. Q., Xu, P. Z., & Wu, X. J. (2016). Global methylation patterns and their relationship with gene expression and small RNA in rice lines with different ploidy. *Frontiers in Plant Science*, 7(JULY2016), 205963. <https://doi.org/10.3389/FPLS.2016.01002/BIBTEX>
60. Zhao, T.-T., Du, L.-D., Wang, C.-K., Wei, M.-M., & Hu, D.-G. (2025). Malate metabolism in horticultural crops: Mechanistic insights and agricultural practices for quality improvement. *Molecular Horticulture*, 5, 58. <https://doi.org/10.1186/s43897-025-00196-6>
61. Ziller, M. J., Hansen, K. D., Meissner, A., & Aryee, M. J. (2015). Coverage recommendations for methylation analysis by whole genome bisulfite sequencing. *Nature Methods*, 12(3), 230–232. <https://doi.org/10.1038/nmeth.3152>

Chapter 3: On the Genetic Improvement of Zinnias for Organic Production: ZC.23.227 and ZP.23.225 Breeding and Methods

Abstract

Zinnias (*Z. Elegans*) have emerged as an important crop for many California organic growers. Due to their cold intolerance, zinnias cannot be refrigerated for extended periods of time. Therefore, production cannot be easily moved to locations that require long transportation times, which has resulted in localized production. However, most zinnia seed currently in production have been bred for conventional systems. Additionally, very few resources exist on the methods and strategies for improving zinnias for cut flower production. After evaluating available germplasm, we made 463 crosses, then selected 85 to evaluate in a single field season. Using a modified synthetic breeding strategy, we performed selection on color, petal doubling, stem length and flower size for five years. We introduce two new cultivars, ZP.23.225 and ZC.23.227, that have been bred specifically for organic systems and addressing new and emerging market demands. ZP.23.225 is a zinnia with a light pink color and a small flower head but maintains agronomic vigor. ZC.23.227 is an orange “cactus” type zinnia with good field performance and an exceptionally large flower head. These varieties were developed through the Student Organic Plant Breeding and Education program at the University of California, Davis.

Introduction

Cultivated zinnia (*Z. elegans*) is a flower in the *Asteraceae* family endemic to the central United States and Mexico. It is a diploid organism with 24 chromosomes, however some variation may exist, including some tetraploid varieties (Saifudin 2021). Zinnia was first introduced to Europe in 1829, and the first documented case of selected breeding was reported in 1858 with the release of the variety “Flore Peno” and subsequent “Nana Flore Peno” in Germany (Stitmart 2007). Early selection focused on flower shape

and color, followed by production of larger varieties that focused on superior stem length for cut flower production, such as the “Giants of California” series, were released in 1926.

Many key traits in zinnias have been characterized genetically (Song 2025). Flower hue is primarily controlled by two genes: a dominant expression gene for anthocyanins (red) and a dominant suppression gene for carotenoids (yellow) (Boyle 1988). The interaction of these two pigment genes leads to common color phenotypes, such as orange and white flowers. However, it is unclear how color phenotypes such as greens, pinks and multicolor flowers are genetically controlled. Petal shade is likely controlled by variation in petal pH (Stimart 2007), while presence of chlorophyll in the petals may alter the underlying flower tone (Boyle 1988). Flower doubling is a polygenic trait that is largely controlled by three Mendelian traits that control rows of petals (Gotoh 1954). Differences in flower heads also are controlled by environmental conditions, with heads reverting to the single type in late summer and autumn.

Few differences exist between different cultivars of zinnias in terms of disease resistance for pathogens such as powdery mildew and *Alternaria*, however, it is unclear whether selection for resistance was ever carried out on common cultivars (Gombert 2001). *Xanthomonas* in particular poses a major problem to seed production and quality (Myung 2012). Some disease resistance has been conferred through hybridization of *Z. elegans* and *Z. angustifolia* (Boyle 1982), as well as some sterile hybrids of *Z. elegans* and *Z. peruviana* (Shanin 1971).

Preferences in the cut flower market are changing. There is a growing demand from consumers to understand where and how the flowers are produced (Etheridge 2024a). Consumers are showing increased preferences in flowers that are produced using sustainable practices, including organic production. Additionally, consumers place the highest value on flowers that are locally sourced. Florists are responding to these trends and adjusting their sourcing practices in kind. Economic performance of flower

shops is significantly improved when flowers are sourced and marketed in a way that promotes environmentalism, sustainability and localism (Etheridge 2024b). To meet these growing trends in consumer and florist preferences, breeders need to adapt to meet market demands.

Here, we implemented a breeding program for *Z. elegans* for organic production that meet emerging market demands. We evaluated and selected germplasm, performed crosses to generate new hybrids under a modified synthetic system, then performed selections over subsequent generations to produce new varieties. The varieties were evaluated and tested in multiple locations and different environments to assess performance and gene by environment interaction.

Methods

Breeding Goals

The primary goal of the breeding program was to develop new varieties of zinnia that are adapted to organic growing conditions. Organic growing conditions are defined by guidelines from the USDA National Organic Program established by the Organic Food Production Act of 1990. These practices include, but are not limited to, reduced chemical fertilizer input, cultural weed and pest control, and planned crop rotations. Varieties needed to respond to these management practices versus conventional farming practices. Specific goals of the program included improving quantitative and qualitative traits that appealed to primary and end users of zinnias. Primary users we defined as growers managing certified organic farmers and end users as consumers, florists and wholesalers. We conducted an informal survey from these users to determine which traits should be the focus of the breeding program that would provide the maximum value. We selected traits that applied to all aspects of the supply chain.

First, novel colors were the primary target trait. From growers and florists, we received comments from the survey that stated that many current zinnia cultivars did not meet current trends in flower arrangements. Many zinnias available at the time of the survey were dominated by bright, saturated hues, and flower heads only presented a single-color. Florists wanted “blending” and “blush” tones, referring to colors with reduced color saturation. Additionally, the survey identified flower heads with multiple colors as highly desirable. Related to flower color is the petal shape and number of petals. Growers and florists find “single petal” flowers, or flower heads with a single row of petals, to be very undesirable. Next, there were three quantitative traits that were important to both growers and end users: stem length, flower size and vase life. By increasing stem size, flowers become more easily harvestable for the grower and increase the number of marketable flowers able to be sold to wholesalers or directly to florists. For florists, longer stems make bouquet arrangement and flower preparation easier, as there is more stem to work with and cut down to appropriate sizes. With flower size, there are conflicting desires between end users. Growers enjoy larger flowers because of their charismatic appeal. However, many florists want smaller flowers because they take up less room in a bouquet. Increasing the post harvest shelf life of individual flowers was also a highly important trait among constituent groups. Other traits that were considered important were disease resistance, particularly to powdery mildew and the unique petal phenotype referred to as “cactus”, which consist of rolled petals.

To manage the various breeding goals, we divided them into four ideotypes. These ideotypes represented a phenotypic goal that the breeding program worked toward that encompasses several of the target breeding traits. First, there was “Pastel Intermediate”. This ideotype focused primarily on color selection and selecting on hue. Second, there was “Cactus Bicolor”. The goal of this ideotype was to express the multicolor phenotype on a flower with the rolled cactus petals. This would create a unique effect that was very novel relative to what was commercially available in 2020. The third ideotype was “Improved Queen Lime”. One of the popular multicolor varieties available is the “Queen Lime” series of flowers. These

flowers have petals that are green and some other color, the most common being pink, orange and red. However, these Queen Lime flowers have very poor field performance due to their low vigor, short stems, and high disease susceptibility. Because of these problems in the field, Queen Lime was not commonly grown in commercial settings. Therefore, we set out to make substantial improvements to its agronomic performance while maintaining the unique color phenotype. The final ideotype was “Smaller Benary’s Giant”. “Benary’s Giant” is a common and while grown variety due to its high vigor and strong agronomic performance. However, many florists do not like using it due to its large flower size, which makes it less suitable for floral arrangements. For this ideotype, the goal was to select smaller flowers from Benary’s giant to maintain its field performance but have more desirable flower sizes.

Germplasm Selection

Germplasm selection consisted of collecting material from public germplasm repositories and commercial sources (Table 3.1). For the accessions from the public germplasm repository, we used the United States Department of Agriculture-Agricultural Research Service (USDA-ARS) Germplasm Resources Information Network (GRIN). In total, we acquired 21 different accessions from GRIN. These accessions were selected because they were all available zinnia cultivars at the time of germplasm selection. For the accessions from commercial sources, we sought to create a broad range of germplasm to both understand the current market for zinnias and provide an adequate base for breeding. Additionally with the commercial germplasm, we collected “mixes”, or named varieties that contained several different possible genotypes. These included named varieties with different colors and petal types.

Table 3.1. Germplasm Sources

<i>PI 442425</i>	Single Variety	GRIN
<i>PI 596403</i>	Single Variety	GRIN
<i>PI 597633</i>	Single Variety	GRIN
<i>PI 545697</i>	Single Variety	GRIN
<i>PI 667549</i>	Single Variety	GRIN
<i>PI 586635</i>	Single Variety	GRIN
<i>PI 597634</i>	Single Variety	GRIN
<i>PI 564843</i>	Single Variety	GRIN
<i>Benary's Giant Salmon Rose</i>	Single Variety	Student Farm
<i>PI 545695</i>	Single Variety	GRIN
<i>PI 596404</i>	Single Variety	GRIN
<i>PI 597637</i>	Single Variety	GRIN
<i>PI 278179</i>	Single Variety	GRIN
<i>PI 442438</i>	Single Variety	GRIN
<i>PI 597636</i>	Single Variety	GRIN
<i>PI 442439</i>	Single Variety	GRIN
<i>PI 603136</i>	Single Variety	GRIN
<i>PI 597640</i>	Single Variety	GRIN
<i>PI 597639</i>	Single Variety	GRIN
<i>PI 170349</i>	Single Variety	GRIN
<i>PI 319385</i>	Single Variety	GRIN
<i>Mazurkia</i>	Single Variety	Baker Creek Heirloom Seeds
<i>Zinderella Lilac</i>	Single Variety	Harris Seeds
<i>Queen Lime Orange</i>	Single Variety	Student Farm
<i>Cactus Aztec</i>	Single Variety	Baker Creek Heirloom Seeds
<i>Giant Dahlia Creamy Yellow</i>	Single Variety	Johnny's
<i>Queen Lime Blush</i>	Single Variety	Fruition Seeds
<i>Zinderella Orange</i>	Single Variety	Harris Seeds
<i>Old Mexico</i>	Single Variety	Burpee
<i>Queen Lime Red</i>	Single Variety	Student Farm
<i>Zinderella Peach</i>	Single Variety	Fruition Seeds
<i>Oklahoma Mix</i>	Mix	Johnny's
<i>Candy Cane Mix</i>	Mix	Burpee
<i>Persian Carpet</i>	Single Variety	Fruition Seeds
<i>Dahlia Flowered Mix</i>	Mix	Swallowtail Garden Seeds
<i>Benary's Giant Mix</i>	Mix	Harris Seeds
<i>Cactus Chrysanthemum mix</i>	Mix	Baker Creek Heirloom Seeds
<i>Jazzy Mix</i>	Mix	Johnny's

Crossing Scheme

With the initial germplasm selection completed, we developed crossing methods. Crossing zinnias is performed by removing the floret around the central core of the flower, then removing the anthers from the floret. If the pollen is mature, it will dehisce and easily fall from the anthers. If the pollen is harvested before dehiscence, the anthers can be agitated using tweezers or some other sharp object to stimulate the release of pollen. Once the pollen is available, crossing is performed by tapping the pollen onto the stigma, located on the outer petals of the plant.

In order to advance our breeding goals, we used a quasi-synthetic breeding scheme to develop a phenotypically consistent cultivar but maintained some level of genotypic diversity (Figure 3.1). With the initial germplasm selection, we conducted over 154 wide crosses to test the viability of our crossing program. We then collected seeds from each of these crosses, and counted the number of harvestable seed. 86 viable crosses were then grown out in the first field season. These 86 F1s were selected based on the phenotypes of the parents from the germplasm selection, and how well they may contribute to the four ideotypes.

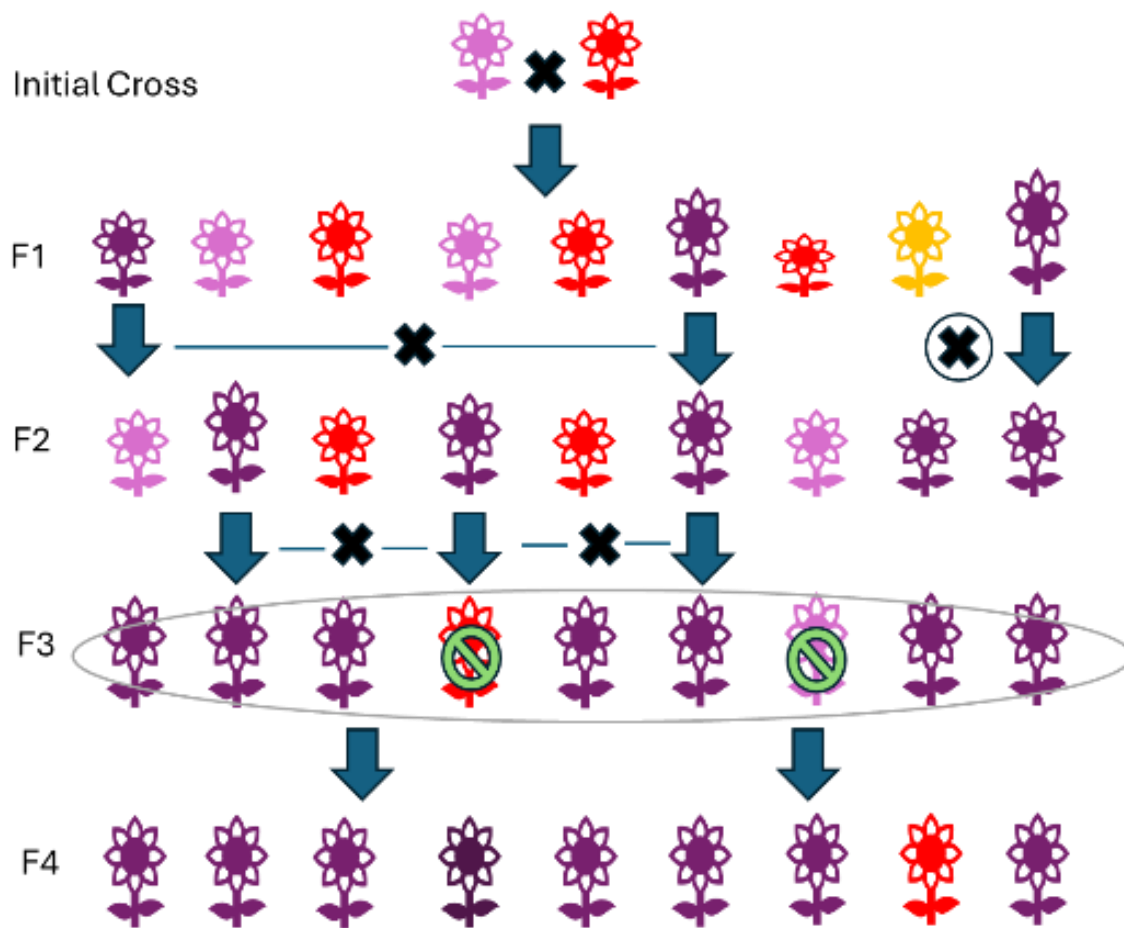


Figure 3.1. Crossing diagram showing a schematic of the general crossing strategy. The initial cross is made, then ideotypes are selected from the progeny. Desired individuals are then selected in subsequent generations while off types are removed from the population. Hand crossing is done in the first two generations, then isolated polycrosses are performed in alter generations.

Field Plan

The breeding program took place over six field seasons, starting in 2020. Seeds were sown in the greenhouse in April and transplanted in May. A randomized complete block design with three replicated plots was planted at one or more of the following locations: the UC Davis Student Farm, Cal Poly Pomona Spadra Farm and the UC Santa Cruz (UCSC) Farm. Plot sizes increased in size in each subsequent year of the program. As populations advanced, plots became larger to better evaluate

uniformity. Individual plots began at up to 12 plants per plot, spaced 12 inches apart. Starting in 2024, the plot size consisted of up to 36 plants per plot spaced 12 inches apart. In 2025, at UC Davis, 162 plant plots were planted with 8 inch spacing in 3 rows per bed, to replicate standard production spacing.

Trials were fertilized, drip-irrigated, and pruned according to standard organic management practices for flowers at each farm. The first buds of each plant were pinched approximately 4 weeks after transplanting to encourage additional branching, and plants were harvested from and pruned regularly throughout the growing season to ensure continued flowering in line with standard production practices.

Phenotyping

Flower diameter and stem length phenotyping typically took place over the course of 2 weeks in mid-to-late July. Stem length was defined as the length between the base (sepals) of a flower at harvesting stage and the next axillary on the branch. This distance from flower to next stem mirrors the standard harvest practice to cut stems directly above an axillary bud whenever the desired stem length is reached. Flower diameter was measured from tip to tip of the opposing flower petals. Both stem length and flower diameter were measured using a ruler. Three flowers were cut and measured from every plant in a plot. In 2025 at UC Davis, 100 random stems were harvested in each plot, instead of three flowers from each plant. Flower color was assessed using a color-scoring chart developed by the breeding program. The chart includes 9 color types (ex. red, yellow, orange, pink) and 5 numerical shade scores within those color types, ranging from most saturated (+2) to least saturated (-2). The chart is referenced to RGB color codes.

Data Analysis

Analysis was performed on individual plants in early generations, then on plant populations in later generations. Results were loaded into R statistical software v. 4.2.1. Data was tested for normality using

Shapiro-Wilks test and F-test was performed to test homogeneity of variance. Data was then plotted to interpret broad trends. A mixed linear model was generated using R package lme4 v. 2.0-6 with plot and block as random effects, then estimated marginal means were calculated for each population using emmeans v.2.0.4. An ANOVA was used to test differences between each of the populations.

To make selections, a logical vector was constructed using selection intensity. In early generations, if individuals with desirable qualitative traits were identified, we advanced if they met the selection intensity of the quantitative traits ($i=0.4$). In later generations when the phenotypic homogeneity of the acceptable, selection based on the quantitative traits between populations was used. From the model, variances components were extracted to calculate an estimate of heritability for stem length and flower diameter. To test differences between checks and varieties, a one-way ANOVA was performed for stem length and diameter. To test differences in doubling, Fisher's exact test was used to compare ratios of single and double flower heads between checks and varieties.

Results

General Description of Final Varieties

The breeding program has resulted in the public release of two completed varieties through the University of California, Davis. These two varieties will hereby be referred to by their field designations and their common names interchangeably.

Shirley or ZC.23.227 is a zinnia bred for organic cut flower production with a mix of peach and light orange flowers with quilled or “cactus” type petals (Figure 3.2). The “cactus” flower phenotype is identifiable by its distinct downward-curved petals, which can vary from moderate frills to near full tubes. The blooms range from 9.9 to 10.4 cm in diameter, on average, the largest



Figure 3.2: Photos of Shirley flowers, 2024-2025. **Top left:** Phenotyping board with flowers chosen to represent all of the colors observed in a single plot in 2025. **Right:** Plot in the 2025 Davis field. **Bottom left:** Marketing image taken in 2024 to highlight the primary flower phenotype of Shirley. (Photo: SCOPE team, UC Davis; bottom left, Amanda Saichaie, UC Davis)

amongst varieties tested in our field trials. Stem length, or distance from the flower head to the next branching node, averages 29.5 cm. The stem length of Shirley has no statistical difference with ‘Benary’s Giant Salmon Rose’, a common pastel zinnia cultivar, and is an average of 3.2 cm longer than ‘Cactus Chrysanthemum’, a common “cactus” type mix.

Estelita or ZP.23.225 is a zinnia bred for organic cut flower production with small, pastel “double” blooms in a complimentary mix of colors (Figure 3.3). Estelita flowers range from light

yellow (RGB R:255, G:217, B:146) to pastel pink (RGB R:255, G:200, B:234). The “double” blooms are 5.7 centimeters in diameter. Stem length, or distance from the flower head to the next branching node, averages at about 24.6 centimeters as compared to an average of 30.5 centimeters for ‘Benary’s Giant Salmon Rose’ and 12.3 centimeters for ‘Oklahoma Ivory’, two common pastel zinnia cultivars suitable for cut flowers. Photos of Estelita, as compared to Benary’s Giant Salmon Rose and Oklahoma Ivory.



Figure 3.3: Photos of Estelita flowers, 2024-2025. **Right:** Estelita plot in the 2025 Davis field **Left:** Marketing image taken in 2024 to highlight the primary pastel pink flower phenotype of Estelita.(Photo: SCOPE team, UC Davis; bottom right, Amanda Saichaie, UC Davis)

Pedigrees

Shirley or ZC.23.227 : In 2021, a single plant of the cultivar ‘Pinca’ was bagged to exclude pollinators and self-pollinated by hand at the UC Davis Student Farm to produce seeds designated ZCF21177. ZCF21177 was planted in replicated plots at the UC Davis Student Farm in 2022. Plants with “cactus” type flowers and unique yellow-orange (RGB values of R:255, G:173, B:90) and peach (RGB values of R:255, G:217, B:146) flower colors were identified, based on a color chart developed by the program, and the best of those were selected using an equally weighted index of longer stem length and larger flower size. Cuttings were taken of six

selected plants, which were moved to the greenhouse, intercrossed by hand during the winter 2022-2023, and seed bulked to form a population called ZC.23.118.

In 2023, ZC.23.118 was planted in replicated plots at the UC Davis Student Farm. Seven plants with the appropriate qualitative criteria of color, cactus-type petals, and “double” flower heads were selected from a single plot at a selection intensity of 0.4 for longer stems and larger flowers. The selected plants were isolated using insect-proof netting and pollinated with blue bottle flies. This population was given the name ZC.23.227 and which we refer to with the tentative name ‘Shirley’ in the remainder of the document.

In 2024, two plots of Shirley (ZC.23.227) were rogued by removing plants with non-cactus petal types, “single” flower heads, and colors that did not match the original color selections. Following rogueing, each plot contained 20-25 uniform, healthy plants with long stems and yellow-orange or peach, cactus-petaled flowers. Both plots were isolated and pollinated using blue bottle flies. The resulting seed from these two plots was bulked together and designated pre-breeder seed.

In 2025, approximately 100-120 uniform plants derived from 2024 seed that had long stems and yellow-orange or peach, cactus-petaled flowers were isolated and pollinated using blue bottle flies. Plants with non-cactus-petals, pink off-type flower color and “single” petal types were rogued. This seed was designated breeder seed.

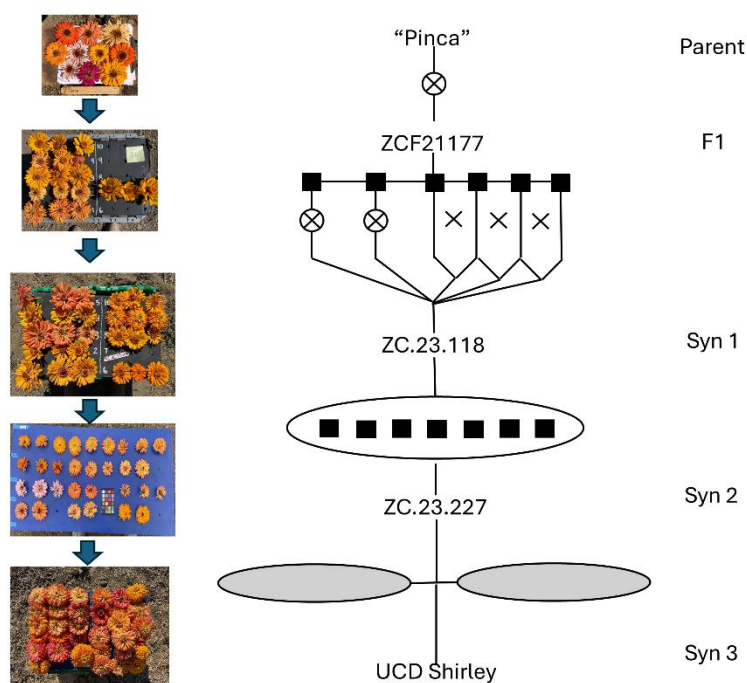


Figure 3.4: Visual pedigree for variety "Shirley".

Estelita or ZP.23.225 : In 2020, a single cross was made in the field at the UC Davis Student Farm between two yellow-flowered individuals from the cultivar 'Oklahoma Mix' and a pink individual from the cultivar 'Benary's Giant Mix'. This cross was labeled ZSG20088.

In 2021 the seeds of cross ZSG20088 were planted in a field plot at the UC Davis Student Farm, and a single cross was made between an individual plant in the ZSG20088 family and a single plant from 'Miniature Double', a new germplasm introduction bred by Floret, a flower seed and breeding company in Mt. Vernon, WA. This cross was named ZAI21115.

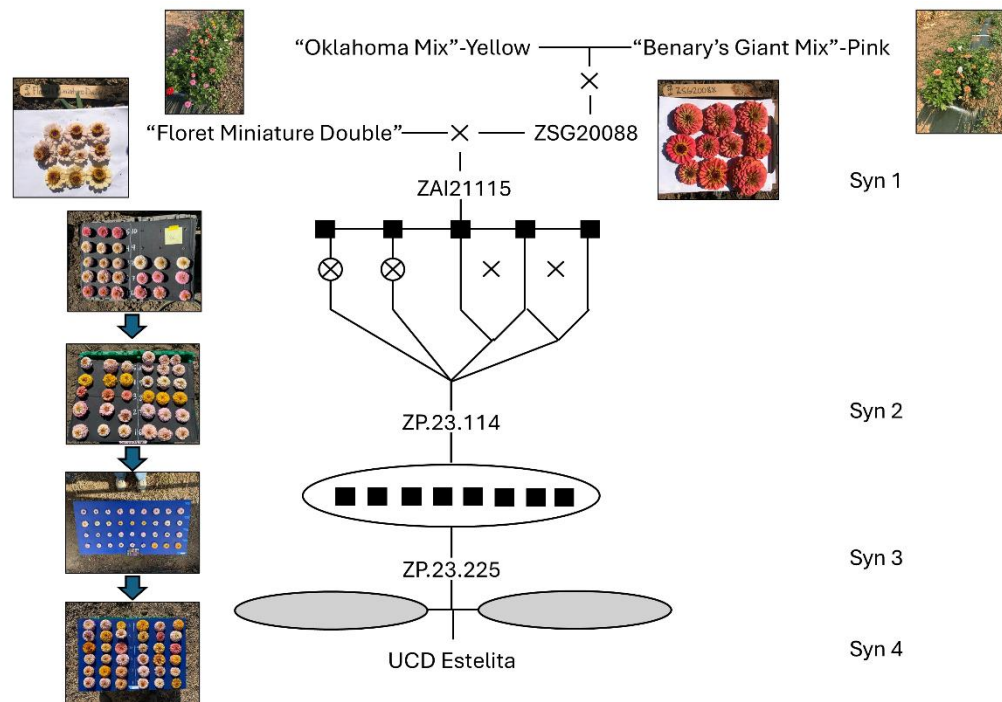


Figure3.5: Visual Pedigree for variety "Estelita".

In 2022, ZAI21115 was planted in replicated plots in the field at the UC Davis Student Farm. Plants with pastel pink color (RGB R:255, G:200, B:234), based on a color chart developed by the program, were selected using an equally weighted index of longer stem length and smaller flower size. Cuttings were taken of the best five plants for a winter greenhouse crossing block. In a winter greenhouse from 2022-2023, the five propagated ZAI21115 individuals were intercrossed by hand and seed was bulked to form a population named ZP.23.114.

In summer 2023, two plots of ZP.23.114 were planted at the UC Davis Student Farm. Off-type plants not matching the pastel pink color targeted for selection were removed. A selection intensity of 0.4 was used, with long stem length and smaller flower size as a selection criteria, in addition to color. Eight uniform plants in one plot of ZP.23.114 were isolated using insect-proof

netting, and pollinated with blue bottle flies. This population was named ZP.23.225, which we tentatively refer to as Estelita in the remainder of the document.

In 2024, two plots of Estelita (ZP.23.225), were selected based on their uniformity, long stems, and small flowers. Each plot contained between 10 and 30 plants with pastel flowers, after off colors, such as bright yellows and pinks, were rogued from the plot. Both plots were isolated using insect-proof netting, and pollinated with blue bottle flies. The resulting seed from these plots were bulked together for 2025 planting. This seed was designated as Estelita pre-breeder seed

In 2025, two plots of Estelita were planted, each containing 100-120 uniform plants with pastel pink, small flowers after any bright yellow and bright pink off-types were rogued. Each plot of Estelita was isolated and pollinated using blue bottle flies. This seed was designated breeder seed.

Color Improvements

Shirley is characterized by its mix of light orange (RGB values of R:255, G:173, B:90) and peach (RGB values of R:255, G:217, B:146) colors. Shirley is novel for its moderated color saturation and “pastel” tones, which are an underrepresented color palette in the current market.

In 2024 and 2025, less than 1% of flowers had a darker pink color that did not match selected colors, however this small variation is tolerable in commercial zinnia varieties.

Estelita is characterized by two primary flower colors: pastel pink (RGB R:255, G:200, B:234) and a light golden yellow (RGB R:255, G:217, B:146), creating a uniform mix of complimentary colors. These pastel colors are both highly desirable for growers and florists, and allow the grower to provide different colors for different markets and meet emerging trends. In 2024, Estelita had approximately 75% pastel pink flowers, 10-25% pale golden yellow flowers, and fewer than 10% off-type flowers (bright pink or bright yellow) based on counting individual plant phenotypes. Similar proportions based on visual estimates were observed in 2025. This percentage of off-types is normal in the zinnia market.

Stem Length

Longer stem length is desirable, but stem length can vary significantly based on environment and management practices. Estelita had longer stems than the most comparable small-flower, pastel cultivar ‘Oklahoma Ivory’ in two of three trials (Table 3.6). ‘Benary’s Giant Salmon Rose’ is another leading pastel cultivar known for its long stems and large flowers. In three of six trials, Estelita was not significantly different from Benary’s Giant Salmon Rose’ but had shorter stems in the other three trials. These comparisons used the estimated marginal means at each trial location, then statistical tests of significance were performed using a one-way ANOVA.

Table 3.2. Stem length per site-year combination, 2024-2025

<i>Location</i>	<i>Year</i>	<i>Estelita</i>	<i>Oklahoma Ivory</i>	<i>Benary's Giant Salmon Rose</i>
		(cm)	(cm)	(cm)
<i>UCD</i>	2024	19.2 ^a	17.3 ^b	20.3 ^{a*}
<i>CPP</i>	2024	10.1 ^a	13.9 ^b	15.8 ^b
<i>UCSC</i>	2024	22.2 ^a	16.8 ^b	23.8 ^a
<i>UCD</i>	2025	24.6 ^a	N/A	30.5 ^b
<i>CPP</i>	2025	10.5 ^a	N/A	13.5 ^b
<i>UCSC</i>	2025	23.2 ^a	N/A	24.0 ^a

*Values within rows followed by a different letter are significantly different at $p < 0.05$.

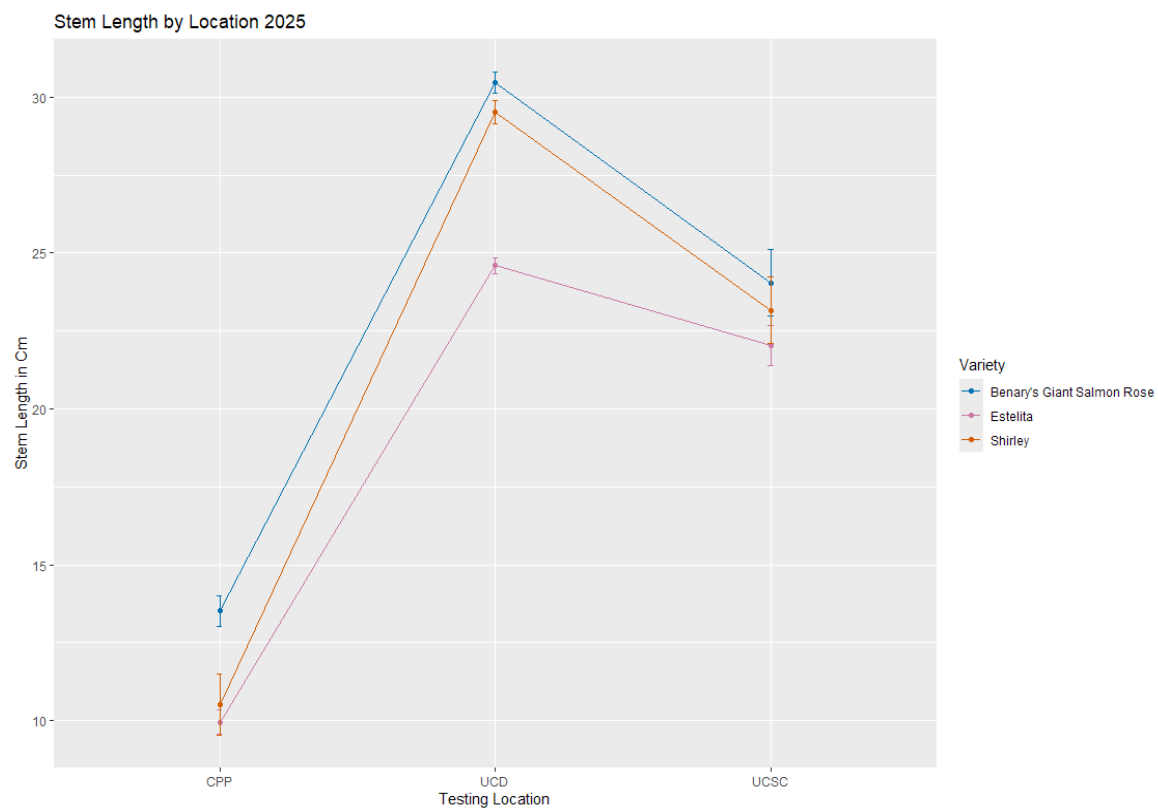
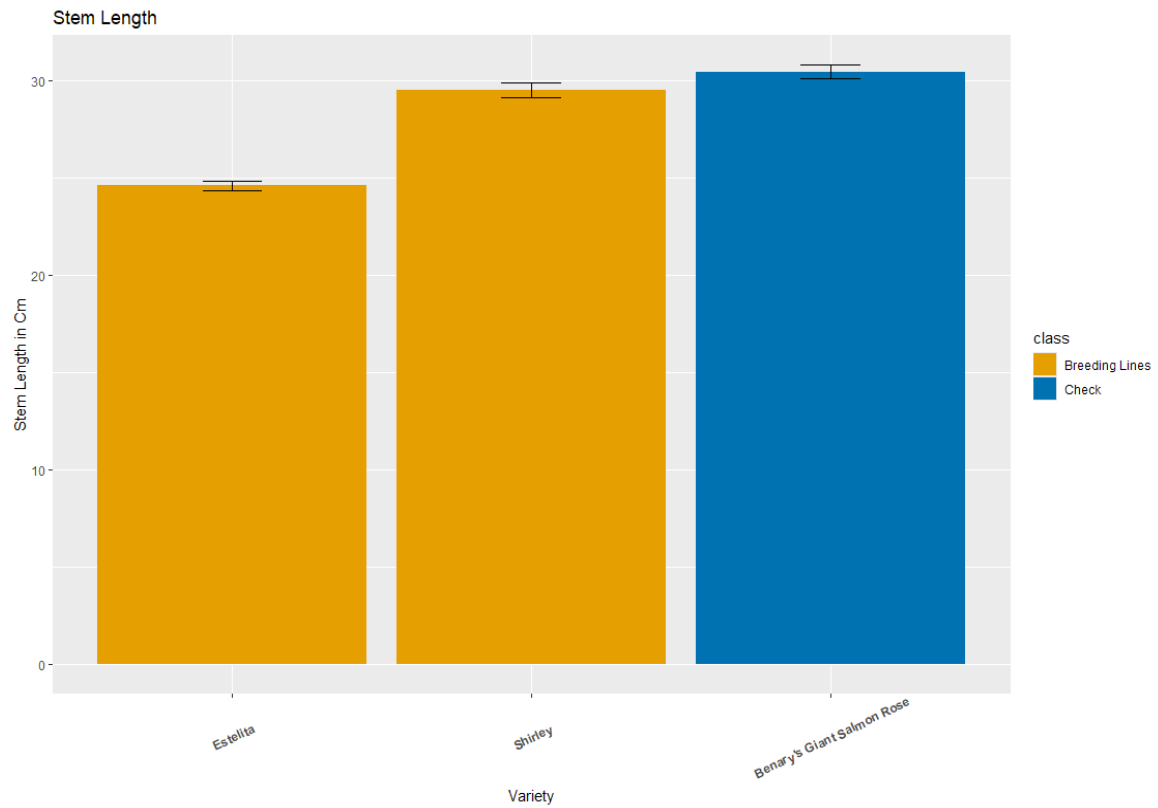


Figure 3.6: Stem length comparisons between release varieties and major check. Multi-location plot showing performance of varieties across locations. CPP=Cal Poly Pamona, UCD= UC Davis, UCSC= .UC Santa Cruz

type zinnias currently on the market (Cactus Chrysanthemum), and is statistically the same as Benary's Giant Salmon Rose, one of the most commonly grown pastel zinnia varieties on the market that is known for its long stems and large flowers. Stem length can vary significantly based on environment and management practices, but we found that Shirley consistently produces stems as long as Benary's Giant Salmon Rose. These comparisons were using the estimated marginal means at each trial location, then statistical tests of significance were performed using a one-way ANOVA.

Table 3.3. Stem length per site-year combination, 2024-2025

<i>Location</i>	<i>Year</i>	<i>Shirley</i>	<i>Cactus Chrysanthemum</i>	<i>Benary's Giant Salmon Rose</i>
		(cm)	(cm)	(cm)
<i>UCD</i>	2024	19.2 ^a	16.5 ^b	20.3 ^{a*}
<i>UCSC</i>	2024	23.8 ^a	19.7 ^b	23.8 ^a
<i>UCD</i>	2025	29.5 ^a	N/A	30.5 ^a
<i>CPP</i>	2025	10.5 ^a	N/A	13.5 ^b
<i>UCSC</i>	2025	23.2 ^a	N/A	24.0 ^a

**Values within rows followed by a different letter are significantly different at $p < 0.05$*

Flower Diameter

Shirley also has exceptionally large flowers compared to other varieties currently available on the market. Flower size has a strong genetic control is heritable and stable between generations, with little influence from the environment or management (Figure 3.7). We found that Shirley had the same or larger flower diameter as compared to Benary's Giant Salmon Rose, one of the most popular large-flowered zinnias on the market, in all five trials where it was tested (Table 3). Analysis was performed using Fischer's exact test to compare combinations of varieties. These

comparisons were using the estimated marginal means at each trial location, then statistical tests of significance were performed using a one-way ANOVA.

Estelita is noteworthy for its small flower size, which is a highly desirable trait for floral arrangements. Flower size is highly heritable and stable between generations, with little influence from the environment or management. Estelita produces flowers that aren't statistically different in size from Oklahoma Ivory but often with longer stems. Benary's Giant Salmon Rose is a large-flowered, pastel zinnia cultivar and Estelita had smaller flowers in all trials, which is desirable. These comparisons were made using the estimated marginal means at each trial location, then statistical tests of significance were performed using a one-way ANOVA.

Table 3.4 Shirley Flower Diameter per site-year combination, 2024-2025

<i>Location</i>	<i>Year</i>	<i>Shirley</i>	<i>Cactus</i>	<i>Benary's Giant Salmon Rose</i>
		(cm)	<i>Chrysanthemum</i> (cm)	(cm)
<i>UCD</i>	2024	9.3 ^a	8.5 ^b	8.4 ^{b*}
<i>UCSC</i>	2024	9.3 ^a	7.7 ^b	9.0 ^a
<i>UCD</i>	2025	10.4 ^a	N/A	8.9 ^b
<i>CPP</i>	2025	9.7 ^a	N/A	9.5 ^a
<i>UCSC</i>	2025	9.5 ^a	N/A	8.9 ^a

*Values within rows followed by a different letter are significantly different at $p < 0.05$

Table 3.5. Estelita Flower Diameter per site-year combination, 2024-2025

<i>Location</i>	<i>Year</i>	<i>Estelita</i>	<i>Oklahoma Ivory</i>	<i>Benary's Giant Salmon Rose</i>
		(cm)	(cm)	(cm)
<i>UCD</i>	2024	5.7 ^a	5.3 ^a	8.4 ^{b*}
<i>CPP</i>	2024	6.5 ^a	6.2 ^a	9.9 ^b
<i>UCSC</i>	2024	6.4 ^a	5.7 ^a	9.0 ^b
<i>UCD</i>	2025	5.7 ^a	N/A	8.9 ^b
<i>CPP</i>	2025	6.2 ^a	N/A	9.5 ^b
<i>UCSC</i>	2025	5.4 ^a	N/A	8.9 ^b

**Values within rows followed by a different letter are significantly different at $p < 0.05$.*

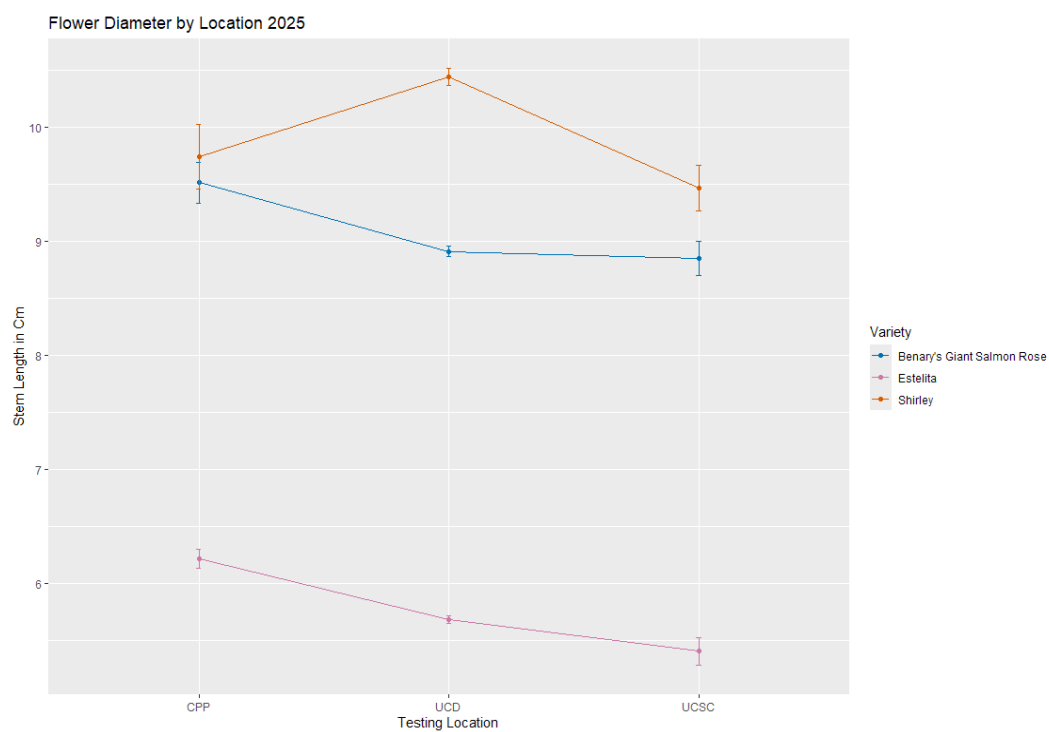
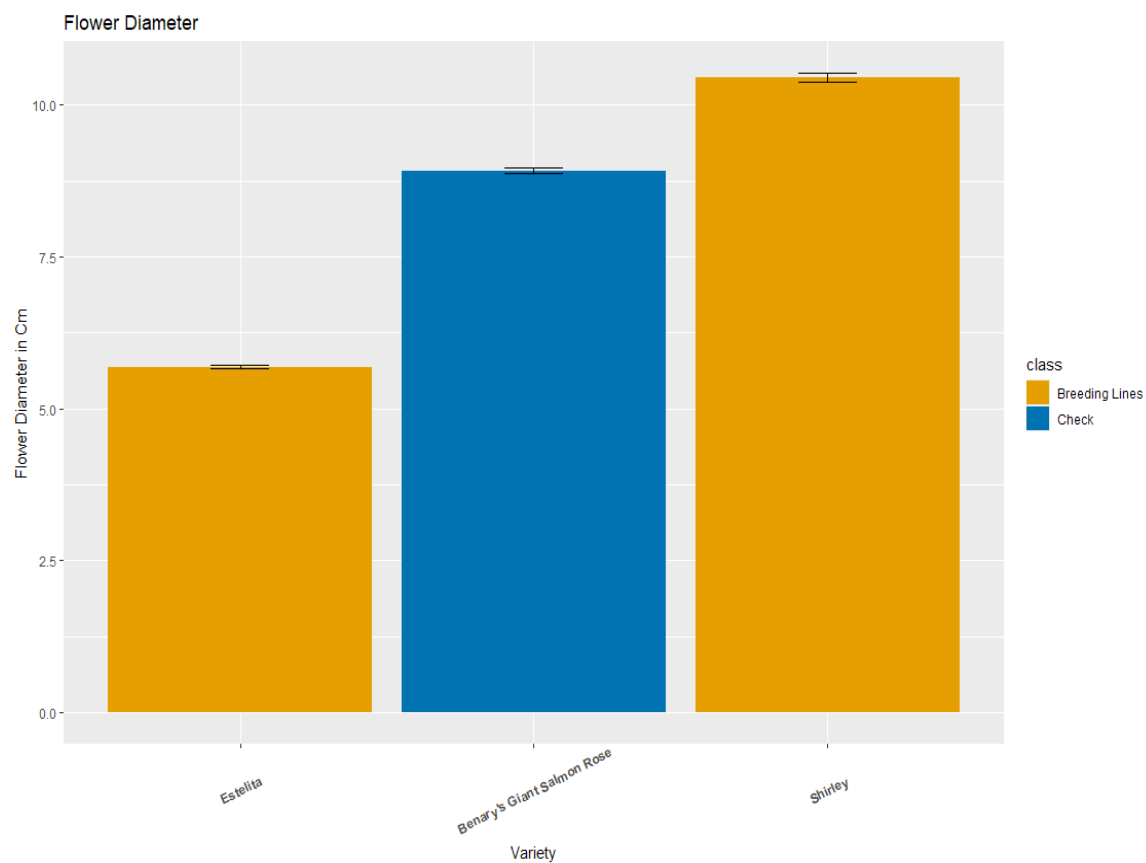


Figure3.7. Stem length comparisons between release varieties and major check. Multi-location plot showing performance of varieties across locations. CPP=Cal Poly Pomona, UCD= UC Davis, UCSC= .UC Santa Cruz

Discussion

This study describes the successful breeding program of an outcrossing seed propagated crop for organic production. Rarely does a breeder have the luxury of building a program from the ground up, with little institutional knowledge behind it. Indeed, often breeding programs are layered with the history of the breeders that preceded them. But, here there was a rare opportunity to design everything, from the crossing structure and methods to the name and manner of the release, from the ground up. With this framework in mind, it would be useful to discuss the success and failures, and the lessons that were learned over the course of this program's design and execution. A breeding program doesn't often go through the entire cycle of germplasm selection to variety release in the span of only six years. However, other comparable zinnias were released in the same time span, so the speed of the release is in line with other programs. This provides even more opportunities to analyze the success and failures of this program because of this condensed period.

Variety Releases

We presented two different varieties that are phenotypically distinct, meet different market needs, and are the result of specific goals. ZP.25.225 is a zinnia with relatively small flower heads, but maintains relatively good vigor compared to high-vigor varieties, but outperforms other varieties with smaller heads. The variety is also most notable for its color, a distinct light pink and yellow mix that is highly desirable in the cut flower market. By contrast, ZC.25.227 is a zinnia with exceptionally large flower heads, larger on average than the comparable check. While the color, a light red-orange, is desirable, it is not as highly sought after as other pastel or blush tones.

However, it makes up in this regard with its strong field performance, high-vigor, and petal shape, which exhibits a high proportion of the cactus petal phenotype, often in greater quantity than comparable checks. Each of these varieties align to specific goals that were established at the inception of the program.

Development for these varieties also included the development of intellectual property protections and plans for commercialization. In other crops within the SCOPE breeding program, fully developed varieties have languished somewhere in the liminal zone between the university and a grower's field. To protect these varieties, we have explored two options for protections of the variety: a commercial trademark and a plant variety protection (PVP), each with their pros and cons. The SCOPE program is committed to the open access of its genetics in the spirit of a public breeding program. At the same time, we want to protect the integrity of the program and those genetics that emerge from it. A trademark offers an interesting avenue to do so by protecting the name of the variety itself, and potentially any variety produced by the SCOPE program, regardless of species. A PVP offers a somewhat different avenue. By pursuing a PVP, the program will be able to have some claim specifically over the genetics themselves, rather than just the name for marketing and distribution. This is a stronger intellectual property claim for the breeder, and also provides a clearer path for future breeders, both within and outside of the program. However, this is a more formal certification process administered through the United States Department of Agriculture. Current PVPs that have been filed for zinnias are for *Z. marylandica* interspecific hybrids sold for the bedding plant market. These zinnias have a much higher genotypic and phenotypic homogeneity than some of the varieties produced through the SCOPE program, so the question of whether a PVP is even possible remains.

Education and Mentorship

There are more intangible successes of this zinnia program, particularly when it comes to education and mentorship. The success is intangible because unlike the variety releases, there are no specific deliverables at the end of the breeding cycle. However, that doesn't make them any less noteworthy. Over the course of the six years of the breeding program, there have been nearly 50 undergraduate students who have participated in the zinnia program in some capacity. Of those 50, nearly a dozen of those have been involved in a more formal and focused capacity as a Lead Student Farmer through the UC Davis Student Farm. Often, these students started out as an intern and chose to involve themselves deeper in the breeding program, sometimes over several years. These students have participated at all levels of the program, from necessary labor tasks such as weeding and crossing, to important intellectual advancements such as selections and phenotypic data analysis. Many of these students have then gone on to careers in plant breeding, both in industry and in academia.

Zinnias have also been used to teach students outside of the SCOPE program as well. For four years, selections from the zinnia program have been used in lab exercises at the UC Davis class "Introduction to Plant Breeding". Exercises have included taking phenotypic data, performing crosses, then tracking the progress of these crosses over successive years of the class. While this is only anecdotal evidence, zinnias have a strong charismatic quality as a teaching tool. Because they do not require emasculation, and the separate male and female flowers are easily distinguished visually, teaching students how to cross is very simple and intuitive, especially when compared to crops such as tomato or common beans. Additionally, since the primary

horticultural characteristic of zinnias are their flower color, rather than something somewhat more abstract such as yield, disease resistance or brix, it is very simple to explain goals and for students to intuit what they need to be doing. It is impossible to determine whether working with zinnias in class “inspired” students to pursue careers in plant breeding, however many students who participated in zinnia related activities in the class did go on to join SCOPE and get internships and Lead Student Farmer positions after having taken the class.

There is another aspect to these successes. Because the program was built essentially from scratch including the overall breeding scheme, there was a considerable amount of trial and error. While much can be learned from the successes of the program, an equal amount, if not more knowledge, can be obtained from its failures or challenges as well.

Inbreeding Depression

One of the major challenges that the program faced was inbreeding and the corresponding loss of promising varieties. Two populations from the program highlight the consequences that inbreeding can have on varietal development. In the first instance, a variety known as “Alba” was abandoned due to its poor seed performance. In the case of Alba, we found an unusual and appealing color in some of our early test crosses. Alba also displayed a high degree of doubling flower heads. After three generations of hand crossing, we had a sizable population of approximately 100 plants that showed good field performance. However, after they were polycrossed, very few seeds were produced: roughly 10% of the seeds of other comparably advanced populations. We hypothesized that the low seed count could come from one of two sources. First, the variety had a high rate of doubling that complicating the identification of the

male flowers, which often were concealed and obstructed by the female flowers. It may be possible that they were so concealed that pollinators in the cages were unable to reach the flowers. Second, the gene pool of the population was smaller than desirable, and that genetic self-incomparability prevented enough seeds from being produced. The second variety that had to be abandoned was colloquially referred to as “Clara”. On paper, Clara was an exceptional variety, with some of the highest mean stem performance of anything in the program and a unique peach-orange color that was popular with trial growers. However, from the start, Clara had unique problems stemming from a series of mutations that were common in the population. Clara seeds would fail to germinate correctly, with too many, too few, or fused cup-like cotyledons that would fail to produce true leaves. Even when the cotyledons would appear healthy, the first true leaves would sometimes fuse, resulting in a single heart-shaped leaf that would result in fewer leaf nodes on subsequent stem growth, if at all. These two issues alone resulted in Clara having a successful germination rate of only 54% in its largest trial, far below the average of other breeding lines and further below the industry germination standard of 95%. Even the plants that did survive displayed a number of undesirable characteristics, such as a lack of gravitropism, lack of branching nodes, and full sterility. Because of the frequency of these undesirable traits, as high as 15-20% of the population in some cases, the decision was made to abandon development of the variety. This decision was also made late in the breeding process, with Clara trial in the advanced stages after five generations, as well as trials in multiple states. The reason that Clara had so many problems is a bit easier to pinpoint. From its cross and one generation of hand crosses, four flowers were polycrossed to make the population that would become Clara. Such a small number of flowers were necessary to isolate the specific color that made Clara so appealing, but the inbreeding was too intense, and most likely gave rise to the

traits that made Clara so problematic. While each of these varieties were ultimately not successful, they did provide useful insights of the effects of heavy inbreeding.

Breeding Strategy Efficiency and Optimization

A question that remains is whether and how the breeding strategy could be improved. Looking at the development of the completed varieties and the effect of selection, a very specific trend begins to emerge. The quantitative traits, stem length and flower size, did not change in significant ways while undergoing active selection. From the visual pedigrees, the resulting varieties have different colors and characteristics than their parents, however it doesn't appear there was a strong response to selection for stem length and flower size. This calls into question the effectiveness of hand crosses and later stages of selection. Improved strategies could mimic clonal selection, where a series of wide crosses are performed, and then populations from those crosses are first selected, then maintained through polycrossing with some refinement, and finally managed like a clonal population. The initial selection would thus define the characteristics of the released population. According to this strategy, the number of new crosses would increase and the need for series of replicated trials over multiple years would decrease. There would be increased emphasis on negative selection of off-color types rather than selecting individual flowers to increase the performance of the population. If this is the case, further efficiencies could be gained and time from cross to release could be reduced.

Phenotyping Automation

In the search for efficient breeding, one of the most challenging was automation of the phenotyping process. The primary approach for this was using image analysis software, such as

ImageJ, to perform flower and stem measurements. By designing the phenotyping boards with the correct color scheme and spacing arrangement, a computer could perform the task of measuring each flower head and stem, as well as taking an RGB score for each flower. There were several reasons for emphasizing automation. First, there was the need for objective color phenotyping. In the initial stages of selection, it was difficult to calibrate color measurement between multiple people. Despite sharing color reference cards, variation in individual color perception caused wildly different color scores for the same flowers. This variability made selection and data collection difficult. The other reason for exploring automation was to save time. Phenotyping had to occur when the flowers were at peak maturity before summer heat intensified and disease pressures set in. Often, this would require flowers to be phenotyped in late July. The process of hand measuring each individual plant multiple times required weeks of work, and each day was limited by the peak heat. If some aspects of this could be automated, reducing the need to be in the sun while also providing stronger objective measurements would increase efficiency. However, several problems prevented implementing automation. The first issue was designing the boards correctly. At first, the flowers were too close together, making segmentation of the individual flowers difficult and preventing the computer from measuring flower heads correctly. The photo boards and flower arrangements were redesigned multiple times to overcome this segmentation issue. However, the new procedures required to arrange the board for taking the picture increased the total time to phenotype. Even after these redesigns enabled successful photo analysis, the results were difficult to interpret. We then tried designing boards to photograph individual plants and their flowers and stems. By taking photos on individuals rather than populations, this design would eliminate segmentation on a board and improve accuracy. However, again, this caused total time of phenotyping to increase and was still

limited by the peak heat of the day. For the stems and flower sizes, there was a strong correlation between the scores obtained through the images and the ground truth scores ($R^2=0.92$). However, it is unclear if these gains justified the effort. There was also the issue of the color scores for the flower heads. Often, petals would have multiple colors on a single head, making a mean RGB score of a flower difficult to parse. Additionally, many flowers are similar in color at the flower center near the disk florets. This similarity between flowers led to flowers with noticeable different petal colors to have similar RGB scores depending on the relative size of the flower head, particularly for small flowers. Ultimately, an objective flower score wasn't necessary to select for a phenotypically homogeneous population, and simple visual selection was sufficient.

Color Inheritance

We attempted to explore the inheritance of color phenotypes. While there was information in the literature on the inheritance of carotenoid and anthocyanin pigments, there was less data on the chlorophyll presence in petals and how it resulted in a green petal color. Therefore, we attempted to perform controlled crosses to create a segregating F2 population to explore this trait. We generated two populations: white (v. "Polar Bear") and green (v. "Benary's Giant Tequila Lime") and "Queen Lime Blush" and green. The resulting F1 populations were white and pink, respectively. We then selfed these F1 flowers to create an F2. In the F2 for the white/green cross, all the plants were white (65 w, 0 g). With the Queen Lime Blush/green population, only four flowers produced viable progeny, and we could not gain any statistically significant insight into the segregation ratios. This result created more questions than answers. Several hypotheses could explain our observations. First, the green color may be maternally inherited. The crosses were

not reciprocal and only performed in the green to white direction. If this was the case, then we may see green flowers in the F3 population. However, no seed was saved from these F2 flowers and no additional selfs were made. Another hypothesis is that there is an environmental component to green, such as soil pH or mean summer temperature. Anecdotally, the program has heard from other breeders in different climates that green (and several other traits) is more common in their regions than Davis. Finally, chlorophyll and the other color pigments could interact in an unknown manner. Green tones are often observed on the underside of petals, or green shades may be present in other colors. In conclusion, green petals may be a much more complex trait that was initially believed.

Vase Life Experiments

Finally, we attempted to phenotype (cut flower) vase life. Because this trait is of significant interest to multiple end users, the program tried to develop an effective assay for vase life, or time until full desiccation after harvest. In the initial years of the program, some evidence suggested a genetic component to this trait. In crosses from 2021 only six genotypes were tested with three replicates. However, as the experiment was scaled up to test over 30 genotypes, the signal became less clear and there were fewer significant differences between the genotypes. Other problems arose as the experiment was expanded. First, there were issues with the testing environment and the level of variability. Originally, the experiment was performed in a greenhouse headhouse in a large open area in the middle of summer. The test was conducted for two to three weeks in the middle of summer around harvest time, and variances in temperature due to doors opening and changes in outside temperature played a role. Eventually, this was overcome by moving the testing conditions to a climate-controlled room, but the original results

contained a lot of uncontrolled variance from the environment. The other, larger issue, was phenotyping. The central question of the experiment was “Is there a relationship between genotype and vase life”? or, phrased differently, “Do some varieties live longer as cut flowers than others?” The question then became, “what is a dead flower and how does one measure this?”. When attempting to phenotype the flowers after they had been cut, there was high variability and different results depending on the tester. We tried to account for this in several ways. First, we tried to set objective standards for flower quality, then trained testers to this standard. We told testers to mark a flower as “dead” if 50% or more of the flower petals were fully desiccated and color had been lost. However, this also proved difficult because death estimation depended on the original flower color. For example, pink flowers looked more alive than orange flowers of the same age and survival state. To overcome this, we attempted to use imaging software to track total change in RGB score over a two-week period. We also took weights of individual flowers to measure the total changes from fresh weight to dry weight over that period. While we were successful at measuring change in weight over time to estimate total desiccation, establishing a “color curve” to measure the change in flower quality turned out to be exceedingly difficult, and the end result data was unsatisfactory. In conclusion, measuring vase life remains a challenge and will require new approaches in order to appropriately phenotype the trait.

Plant Breeding as an Art

It is often remarked how plant breeding is as much as art and science. While countless books and papers have outlined the science of plant breeding, little has been documented on the “art” aspect and what exactly this is trying to articulate. The concept of the “breeders’ eye” has been broadly

defined as an individual breeder's ability to identify winners and losers in the population. This is probably some combination of experience working with a specific species and familiarity with the subtle nuances of a specific population or program. This could be the art that is often referred to because it tries to describe an indescribable quality required when making successful selections. But that is only one aspect of breeding.

Here, we have chronicled a breeding program from its inception and documented the various processes that went into its creation. The program was struck with the unique problem of producing a successful breeding program and strategy when there was little documentation on how to do so. The art of plant breeding lies in this: how through trial and error, and through experiments with marginal outcomes, can one make incremental progress, and define the intangible qualities that lead to success. In this process, successes are as important as failures, and there is still plentiful room for improvement and optimization. Development of our two varieties is a testament on what can be accomplished with limited resources and a relatively small program. Its product is also something beautiful. There are few opportunities where the outcome of a breeding program is to produce something that moves people, even in other ornamental programs. Established programs are focused on making incremental improvements on elite varieties, but having a completely blank slate is a rare thing.

References

1. Boyle, T. H., & Stimart, D. P. (1982). Interspecific hybrids of *Zinnia elegans* Jacq. and *Z. angustifolia* HBK: Embryology, morphology and powdery mildew resistance. *Euphytica*, 31(3), 857–867. <https://doi.org/10.1007/BF00039226>

2. Boyle, T. H., & Stimart, D. P. (1988). Inheritance of Ray Floret Color in *Zinnia angustifolia* HBK and *Z. elegans* Jacq. *Journal of Heredity*, 79(4), 289–293. <https://doi.org/10.1093/oxfordjournals.jhered.a110511>
3. Etheredge, C. L., DelPrince, J., & Waliczek, T. M. (2024). Perceptions of Environmentalism and the Use of Sustainable Floral Design Practices by Professional Florists. *HortTechnology*, 34(6), 812–818. <https://doi.org/10.21273/HORTTECH05546-24>
4. Etheredge, C. L., Waliczek, T. M., & DelPrince, J. (2024). Demographic Differences in United States Consumers' Perceptions and Willingness to Pay for Sustainable Environmental Practices in the Floral Industry. *HortTechnology*, 34(1), 27–51. <https://doi.org/10.21273/HORTTECH05325-23>
5. Gombert, L., Windham, M., & Hamilton, S. (n.d.). *Evaluation of Disease Resistance among 57 Cultivars of Zinnia*.
6. Gotoh, K. (1954). Inheritance of Doubleness in *Zinnia elegans* L. *Japanese Journal of Breeding*, 4(1), 37–40. <https://doi.org/10.1270/jsbbs1951.4.37>
7. Lou, X. Y., Hu, Q. S., Bao, M. Z., & Ye, Y. M. (2010). Analysis of combining ability of two-types of male sterile and four restorer lines of *Zinnia elegans*. *Euphytica*, 174(1), 91–103. <https://doi.org/10.1007/S10681-010-0143-X/FIGURES/2>
8. Myung, I.-S., Lee, J. Y., Yoo, H. L., Wu, J. M., & Shim, H.-S. (2012). Bacterial Leaf Spot of *Zinnia* Caused by *Xanthomonas campestris* pv. *Zinniae*, a New Disease in Korea. *Plant Disease*, 96(7), 1064–1064. <https://doi.org/10.1094/PDIS-03-12-0319-PDN>
9. Saifudin, Shafira, S., Dwiranti, A., & Salamah, A. (2021). Short communication: Analysis of the chromosome numbers of *zinnia elegans* jacq. In single, double, and pom-pom flowers. *Biodiversitas*, 22(7), 2771–2777. <https://doi.org/10.13057/biodiv/d220726>
10. Song, C., Liu, X., Xu, M., Ying, M., Fu, J., Zhang, C., Song, C., Liu, X., Xu, M., Ying, M., Fu, J., & Zhang, C. (2025). Germplasm resource and genetic breeding of *Zinnia*: A review. *Ornamental Plant Research*, 5(1). <https://doi.org/10.48130/opr-0025-0014>
11. Stimart, D., & Boyle, T. (2007). *Chapter 12 ZINNIA Zinnia elegans, Z. angustifolia*.
12. Terry-Lewandowski, V. M., Baughan, G. R., & Stimart, D. P. (2011). Cytology and breeding behavior of interspecific hybrids and induced amphiploids of *Zinnia elegans* and *Zinnia angustifolia*. *https://doi.org/10.1139/G84-007*, 26(1), 40–45. <https://doi.org/10.1139/G84-007>
13. Terry-lewandowski, V. M., & Stimar, and DENNIS P. (n.d.). THE INHERITANCE OF RESISTANCE TO POWDERY MILDEW IN INTERSPECIFIC HYBRIDS AND INDUCED AMPHIPLOIDS OF ZINNIA ELEGANS AND Z. ANGUSTIFOLIA HBK]. *Euphytica*, 34, 483487.
14. Torres, A. M. (n.d.). *TAXONOMY OF ZINNIA 1*.
15. Ye, Y. M., Zhang, J. W., Ning, G. G., & Bao, M. Z. (2008). A comparative analysis of the genetic diversity between inbred lines of *Zinnia elegans* using morphological traits and RAPD and ISSR markers. *Scientia Horticulturae*, 118(1), 1–7. <https://doi.org/10.1016/J.SCIENTA.2008.05.025>