

EXPLORING DEVELOPMENT AND GENETIC VARIATION WITHIN *VITIS*

By

Eleanore Jeanne Ritter

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Plant Biology – Doctor of Philosophy

2024

ABSTRACT

The genus *Vitis* (grapevine) contains about 70 species, including domesticated grapevine (*Vitis vinifera* L.). Since its domestication about 11,000 years ago, humans have generated thousands of domesticated grapevine cultivars exhibiting large phenotypic diversity, likely owing to the use of grapes in wine. Advances in genome sequencing technologies have allowed plant geneticists to disentangle the genetic basis of many of these diverse, complex traits and have led to the production of many *Vitis* genomes to aid in this effort. In this dissertation, I utilized modern genomic resources in *Vitis* to investigate the genetic mechanisms that influence various phenotypes within *Vitis* species (*V. vinifera* and *V. riparia* Michx.) and to generate a new genomic resource. In Chapter 1, I reviewed the history of *Vitis* and domesticated grapevine, as well as the current state of genetic research in *Vitis* that I have built upon in this dissertation. In Chapter 2, I investigated the genetic basis of the Witch's Broom bud sports in domesticated grapevine, which are shoots that randomly arise on otherwise normal grapevine plants—presumably through somatic mutations—and exhibit dwarf phenotypes and reduced fertility. To do so, I sequenced two independent cases of the Witch's Broom bud sport alongside their wild-type counterparts and identified putative causal genetic variants. I also characterized the phenotypes of the two cases, which revealed that these bud sports display developmental defects early on within the developing buds and that these independent cases display distinct phenotypes. In Chapter 3, I investigated the molecular genetic basis of mite domatia (hereafter, "domatia") in *V. riparia*, which are tiny structures the plants form on the undersides of their leaves that mediate a mutualism with beneficial mites by providing them shelter in return for protection. These mites protect against small herbivores and pathogenic fungi, including

significant grapevine pests like powdery mildew. Using transcriptome sequencing of two *V. riparia* genotypes with heritable distinct investments into domatia, I was able to identify key molecular genetic pathways involved in domatia development and intraspecific variation in domatia traits. This work, coupled with comparing leaf shapes between the two genotypes, also demonstrated a strong potential link between domatia traits and overall leaf development in *V. riparia*. In Chapter 4, I assembled and annotated the genome of the Dakapo variety of grapevine, a teinturier (“dyer”) variety that produces pigmented berry flesh (unlike most grapevine varieties). This high-quality genome assembly and the accompanying annotations will support future work on berry flesh color and anthocyanin production, as well as other work genetic research in domesticated grapevine. In Chapter 5, I described future directions for these research projects. Overall, the work described in this dissertation has provided unique insights into grapevine biology and generated new genomic resources, which will greatly facilitate future research in grapevine and other plant systems.

This dissertation is dedicated to my mother, Lisa Ritter. Her boundless love strengthens and defines me every day, even since her passing in 2017.

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the wisdom from many brilliant scientists, the advice from key mentors (both formal and informal), the camaraderie of friends (both near and far), and the love and support of family. I am forever grateful for the people I am surrounded by and could not have written this dissertation without the support and guidance of many.

First and foremost, I would like to thank my advisor, Dr. Chad Niederhuth. I will forever value our discussions, regarding research specifically as well as other topics in science, academia, and life in general. Your unwavering candidness during our discussions was always appreciated. Thank you for supporting all my crazy ideas (except the intractable ones) and helping me grow into the scientist I am today.

I would also like to thank the other members of my committee—Dr. Robin Buell, Dr. Dan Chitwood, and Dr. Emily Josephs. Thank you, Robin, for your expansive expertise in plant genomics and your honesty on all matters. Thank you, Dan, for passing along your knowledge and enthusiasm about grapevines, as well as your overall exuberance for all things science. Finally, thank you, Emily, for your expertise in evolutionary biology and your mentorship during my final year that always kept me encouraged.

I would also like to thank other faculty and staff members in the Department of Plant Biology at Michigan State University (MSU) that greatly supported me during my time at MSU and throughout Chad's departure from MSU to join Corteva. Thank you, Dr. David Lowry, for taking me into your lab (informally) and allowing me to get a head start on learning about the magical monkeyflower. Thank you to all members of the Lowry Lab—David, Leslie, Lauren, Daniel, Madison, Andrew, Otto, Cam, and Paige—for welcoming me

into your lab with open arms and for allowing me to be a part of your exuberant, supportive group. Thank you, Dr. Andrea Case, for supporting me through Chad's departure and going above and beyond to make sure that my final year went smoothly. Thank you, Sara Kraeuter, for your constant support and assistance navigating the bureaucracy that went above and beyond. Thank you, Kelley Rose, Heather Stallone, Krystal Witt, and Trevor Simmons, for answering all my questions and for bringing so much joy to the department. Thank you, Dr. Shin Han Shiu and Dr. Jyothi Kumar, for your dedication to the IMPACTs program.

I am also grateful to have had help from the MSU Genomics Core, particularly Dr. Kevin Childs, who helped me plan some of the sequencing described here.

To the former members of the Niederhuth lab—thank you all for your support. To Dr. Leslie Kollar—thank you for being an amazing mentor and friend. To Hannah Brown—thank you for all your help within the lab and for your contagious excitement for research. To Dr. Sunil Kenchanmane Raju—thank you for all the advice and support, both in the lab and on the tennis courts. Thank you to Aidan Kile and Diego DeSousa, my two undergraduate mentees, for bringing so much enthusiasm to the lab and for being the best mentees I could have asked for.

Thank you to my friends at MSU (past and present)—Dr. Nate Catlin, Maya Wilson-Brown, Erika LaPlante, Caroline Edwards, Michael Foisy, Julia Brose, Riley Pizza, Carolyn Graham, Kate Wynne, Seth Hunt, Madison Plunkert, Andrew Bleich, Otto Kailing, Cam Durant, Sara Hugentobler, Meaghan Clark, Olivia Fitch, Brooke Jeffreys, and many, many others—thank you all for the friendship, the laughter, and all the kindness.

Thank you to the many faculty members of the Skidmore College Biology Department that encouraged me to pursue this degree, including Dr. David Domozych, Dr. Elaine Larsen,

Dr. Patti Steinberger, Dr. Monica Raveret-Richter, Dr. Bernard Possidente, Denise McQuade, Dr. Erika Schielke, and many others. Applying to graduate school and finishing my undergraduate degree right after the passing of my mom would have been impossible without the support of you all. I am eternally grateful for the kindness and support you all showed me, as well as all the knowledge and skills I gained from you all during my time at Skidmore.

Thank you to my family and bonus family (the Martins/Collishaws) for all the love and support. Thank you to my mom, for raising me as a single mother and for instilling in me the importance of love, kindness, and learning. All my successes I owe to the love and dedication of my family.

Finally, thank you to my husband, Bruce, and our dog, Mulberry. You two are my best friends, and words cannot fully express how grateful I am to have you both in my life. I love you both so much. Bruce—thank you for all the love, support, and laughter during this process.

TABLE OF CONTENTS

CHAPTER 1: Introduction.....	1
REFERENCES.....	17
CHAPTER 2: From buds to shoots: insights into grapevine development from the Witch’s Broom bud sport.....	26
REFERENCES.....	62
APPENDIX A: CHAPTER 2 SUPPLEMENT.....	68
CHAPTER 3: Small, but mitey: investigating the molecular genetic basis for mite domatia development and intraspecific variation in <i>Vitis riparia</i> using transcriptomics.....	81
REFERENCES.....	112
APPENDIX A: CHAPTER 3 SUPPLEMENT.....	120
CHAPTER 4: The assembly and annotation of two teinturier grapevine varieties, Dakapo and Rubired.....	131
REFERENCES.....	153
APPENDIX A: CHAPTER 4 SUPPLEMENT.....	161
CHAPTER 5: Concluding remarks.....	162
REFERENCES.....	166

CHAPTER 1: Introduction

OVERVIEW

This dissertation focuses on understanding development and genetic variation in *Vitis* (grapevine) using modern genomic resources. In this introduction, I introduce the genus *Vitis* and detail the current state of genomic resources available in grapevine. I also describe past research on traits of interest in grapevine that I expanded upon within other chapters of this dissertation.

THE GENUS *VITIS*

The genus *Vitis* (grapevine), which contains the domesticated grapevine (*Vitis vinifera* L.), is a member of the angiosperm family Vitaceae and order Vitales, which diverged early on from rosids and Saxifragales about 115-126 million years ago (mya) (Zeng et al., 2017). Vitaceae itself contains about 950 species, distributed across 16 genera (Ma et al., 2021). The family is present on all continents, aside from Antarctica, however most species are present in tropical regions (Wen et al., 2018). A key feature that differentiates Vitaceae from other angiosperm families is the presence of leaf-opposed tendrils (Gerrath et al., 2015), which are present in most species, aside from a few species of *Cyphostemma* (Gerrath & Posluszny, 2007). The presence of tendrils allows the plants to climb as they grow, and as a result, most species are climbing woody vines (also known as “lianas”) (Gerrath et al., 2015).

There are five tribes within Vitaceae, with *Vitis* being in the tribe Viteae, along with *Ampleocissus*, *Nothocissus*, and *Pterisanthes* (Ma et al., 2021). *Vitis* is distinguished from the other three genera by being dioecious and producing a calyptra, or flower cap, that covers

the flowers prior to bloom. It is also the only genus within the tribe including species native to North America (Gerrath et al., 2015). *Vitis* contains about 70 species divided between two recognized subgenera, *Vitis* subg. *Muscadinia*, with only two species, and *Vitis* subg. *Vitis* encompassing the remaining species (Liu et al., 2016). The genus as a whole is thought to have originated in North America in the late Eocene (~39.4 mya) and subsequently spread to Europe and Asia in the late Eocene as well (~37.3 mya) (Liu et al., 2016). Both North America and East Asia are hotspots of *Vitis* diversity, however, the subgenus *Muscadinia* is only present within North America (Liu et al., 2016; Zecca et al., 2012) (Figure 1.1).

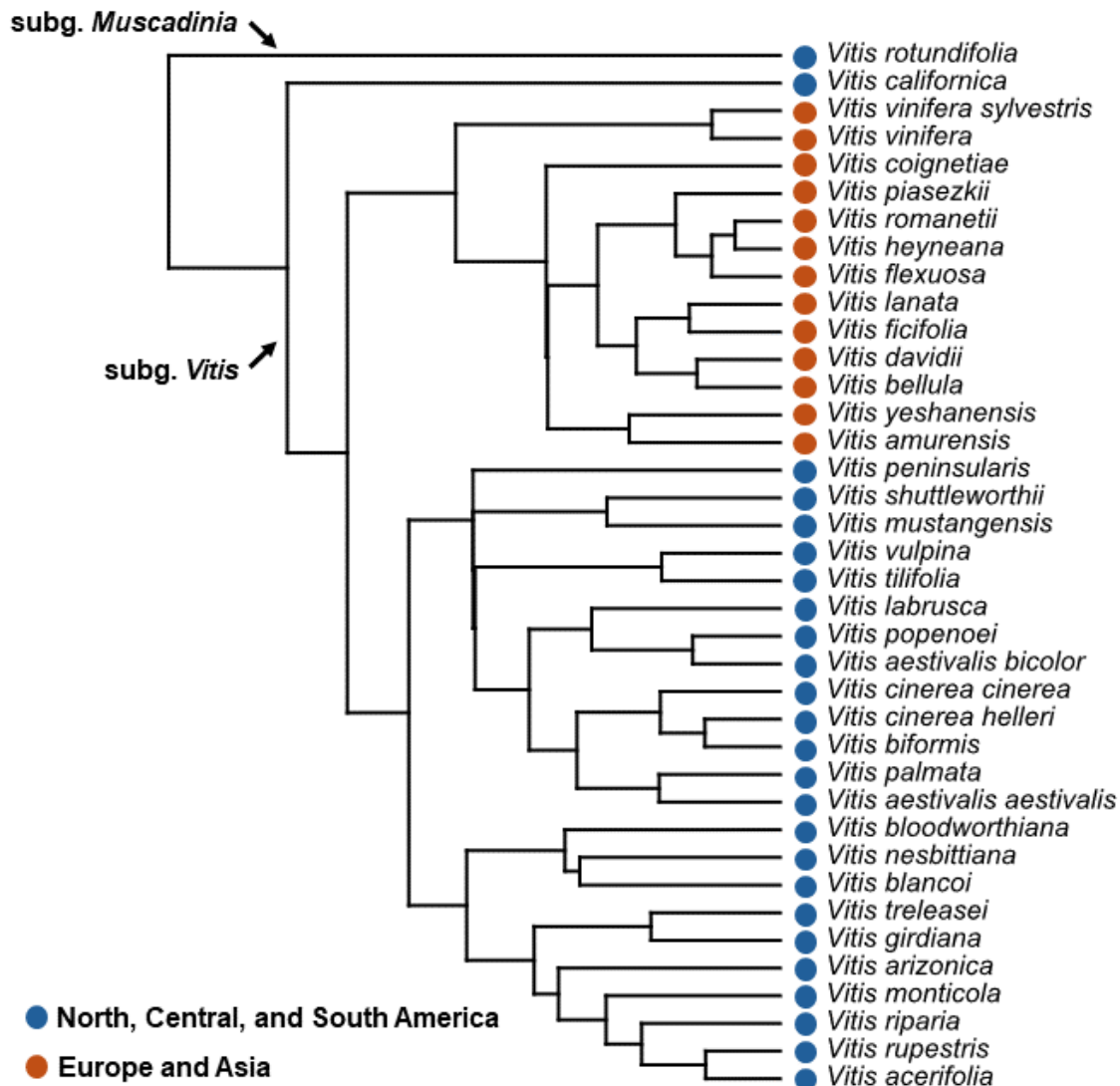


Figure 1.1. A phylogeny of 38 species of *Vitis* (Marjorie G. Weber, unpublished) colored by native range, with North, Central, and South American *Vitis* shown with a blue circle and European and Asian *Vitis* shown with an orange circle. The two subgenera of *Vitis* are noted as well.

Notably, domesticated grapevines (*Vitis vinifera* ssp. *vinifera*, hereafter *V. vinifera*) emerged outside of these hotspots from both Europe and Western Asia through domestication of *Vitis vinifera* ssp. *sylvestris* (hereafter *V. sylvestris*) (Dong et al., 2023).

THE HISTORY OF GRAPEVINE DOMESTICATION

Domesticated grapevine (*V. vinifera*) holds both economic importance, as the fifth

most produced fruit crop (FAO, 2023), and cultural significance, owing to its long history of cultivation by humans (McGovern, 2013). It is estimated to have been domesticated twice ~11,000 years ago in Western Asia and the Caucasus, with two distinct domestication events for both table and wine grapes (Dong et al., 2023). While initially grown for consumption as a food source, evidence of winemaking with grapes dates back as far as ~8,000 years ago (McGovern et al., 2017). The two domestication events seem to have occurred within similar timeframes, but from two distinct Eastern populations of *V. sylvestris* (Dong et al., 2023). Following domestication, the population domesticated in the Caucasus spread north of the Black Sea, likely spreading as far as the Carpathian Basin. However, the population domesticated in Western Asia spread much further. The Western Asia population was likely domesticated within the Fertile Crescent (McGovern, 2013), which is thought to be the “cradle of agriculture” where many other crop species were first domesticated (Lev-Yadun et al., 2000; Riehl et al., 2013). From there, this population then spread in four distinct directions across Eurasia and North Africa, following known human migration routes (Dong et al., 2023). During dispersal into Europe, this population appears to have undergone a stepwise diversification, with the first step being introgression from the Western population of *V. sylvestris* that seems to be ancestral to all European domesticated grapevines. After spreading to Western Europe through the Balkans and Iberian Peninsula, a unique second round of introgression from the Western population of *V. sylvestris* occurred within the Western European population (Dong et al., 2023).

The spread of domesticated grapevine occurred again thousands of years later with European colonization efforts between the 15th-18th centuries. It was first introduced to the New World when Christopher Columbus brought it to Hispaniola in 1493 as a part of the

Colombian exchange (Gade, 2015). It was later introduced to Australia in 1788 through British colonization (Read, 2015). As a result, domesticated grapevines are still cultivated globally (aside from Antarctica). The Columbian exchange was not unidirectional, however, and resulted in the introduction of grape phylloxera (*Daktulosphaira vitifoliae* Fitch) to Europe in the mid-19th century from the New World (Tello et al., 2019). *Vitis vinifera* roots are highly susceptible to phylloxera, unlike their North American counterparts. The introduction of phylloxera led to a rapid decimation of vineyards globally, which is known as the “great wine blight” (Alston & Sambucci, 2019). However, North American *Vitis* species with resistance to phylloxera were used to develop rootstocks that were grafted to *V. vinifera* cuttings, allowing for the survival and reestablishment of vineyards globally (This et al., 2006).

While it is thought that phylloxera decimated genetic diversity within domesticated grapevine (This et al., 2006), between 6-12,000 varieties of domesticated grapevine exist today (International Organisation of Vine and Wine, 2017). The extensive diversification of domesticated grapevine is in large part due to the use of grapes in winemaking.

Domesticated grapevine varieties are diverse in a number of traits, from differences in berry color or sugar content, to differences in yield or responses to biotic stress. Advances in genetic/genomic sequencing and analysis in the 21st century have recently invigorated efforts to understand the genetic basis of these traits in grapevine.

GENOMIC RESOURCES FOR *VITIS*

Modern genome sequencing technologies have allowed for a better understanding of not only the history of grapevine domestication (Dong et al., 2023; Myles et al., 2011; Zhou et al., 2019), but also the genetic basis of many key traits [such as berry skin color (Azuma et

al., 2009, 2008; Kobayashi et al., 2004) or cold stress (Rubio et al., 2019)]. The increase in studies using genetics and genomics to investigate grapevines has in large part been driven by the creation and refinement of new sequencing technologies. This was initially enabled by next generation sequencing (NGS) technologies, such as Illumina sequencing, which allowed for parallelized sequencing of reads. These NGS technologies have been refined over time, further reducing sequencing costs (Slatko et al., 2018). However, NGS technologies produce short read sequences, typically up to 600 base pairs (bp) long (Satam et al., 2023), making it challenging to accurately assemble highly repetitive regions (Tyson et al., 2018). The advent of third-generation sequencing technologies that with increased read lengths, including PacBio and Oxford Nanopore Technologies (ONT) sequencing, have made genome assembly both easier and more accurate due their ability to produce long read sequences that are 10 kilobase pairs (kbp) long or more (Satam et al., 2023; Wang et al., 2021), thus allowing accurate sequencing of repetitive regions and large structural variants (Tyson et al., 2018). The improvement of these sequencing technologies, along with the spread of additional technologies like optical mapping or Hi-C that improve scaffolding, have made it easier and more cost effective to assemble accurate genomes (Pollard et al., 2018). This has allowed for an eruption of available domesticated grapevine genome assemblies, as well as reference genomes for wild *Vitis* species (Figure 1.2).

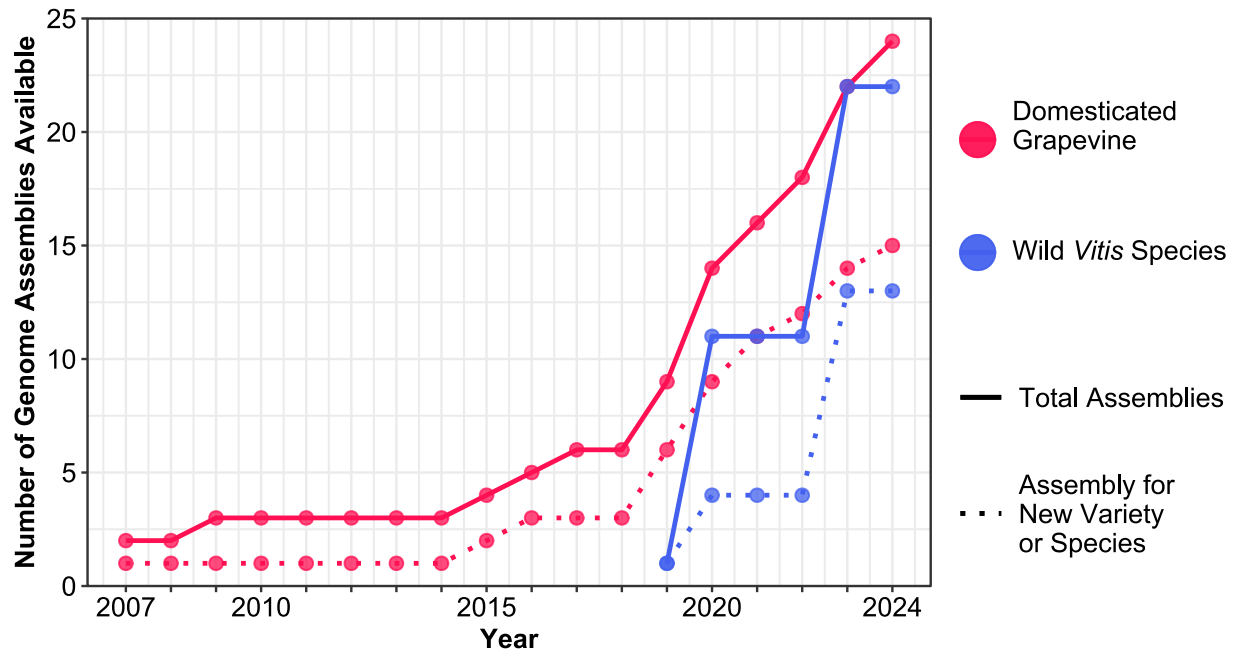


Figure 1.2. A timeline of the release of domesticated grapevine and wild *Vitis* genome assemblies. The total assemblies include assemblies for distinct clones of a variety/species as well as updated assembly releases. The release of a new variety for domesticated grapevine or a new *Vitis* species are denoted with the dotted lines.

The initial genome assembly for domesticated grapevine was released in 2007 and was the first genome assembly for a fruit crop. Domesticated grapevine is highly heterozygous, which would have made genome assembly somewhat challenging as the genome was assembled before new technologies like long-read sequencing and optical mapping were widely accessible and affordable. As a result, the PN40024 Pinot Noir genotype was used for this assembly due to its high homozygosity (~93%) that resulted from repeated selfings. This genome was a shotgun whole-genome assembly, assembled using 6.2 million end-reads sequenced through Sanger sequencing. This approach yielded a somewhat complete, highly fragmented genome with ~19,000 contigs. An annotation for this genome was released as well, with 30,434 genes annotated (Jaillon et al., 2007). The PN40024 reference genome assembly has been updated three times, with substantial

improvements each time. The first update was the release of the 12X.v0 grapevine genome assembly in 2009 which increased the sequencing coverage used in the assembly to 12X, resulting in reduced fragmentation and more complete coverage of the genome (The French-Italian Public Consortium, 2009). New annotations, named CRIBIv1, were later released for this genome independently (Forcato, 2010). The grapevine reference genome assembly was further updated in 2017 with the 12X.v2 assembly produced using parental maps and mate paired sequences that enabled improved scaffolding of the 12X genome. This assembly was accompanied with the release of the VCost.v3 genome annotations as well (Canaguier et al., 2017). The most recent grapevine genome assembly, PN40024.v4, has been the most substantial update thus far. This genome used the 12X.v0 scaffolds, along with 27X coverage Single-Molecule Real-Time (SMRT) sequencing PacBio reads and 15X coverage Illumina reads, to produce two high quality assemblies for the PN40024 reference haplotype and the PN40024 alternate haplotype. These assemblies are highly continuous and complete. The PN40024.v4 genomes were released with new annotations that are higher quality and less fragmented than the VCost.v3 genome annotations which seemed to include many erroneous annotations (Velt et al., 2023). The release of the PN40024.v4 genome assemblies and annotations, which are far more complete and accurate than past releases, will sustain continued genetic research in grapevine and likely improve the ease with which these studies can be conducted.

Beyond the PN40024 genotype used for the grapevine reference genome, reference genomes have been released for at least 14 other grapevine varieties to date, spanning a wide range of phenotypic diversity (Blanco-Ulate et al., 2015; Chin et al., 2016; Maestri et al., 2022; Massonnet et al., 2020; Minio et al., 2019, 2022, 2024a, 2024b; Onetto et al., 2023;

Sichel et al., 2023; Urra et al., 2023; Vondras et al., 2021, 2019; Zhou et al., 2019; Zou et al., 2021). These reference genomes have been used to address a variety of questions, from understanding biotic stress (Blanco-Ulate et al., 2015) to investigating sex determination within grapes (Massonnet et al., 2020). Grapevines exhibit substantial genetic variation among varieties, including large structural variants and differences in gene content (Maestri et al., 2022; Minio et al., 2019; Zhou et al., 2019). As a result, variety-specific reference genomes are vital for grapevine genetic research aimed at understanding the genetic basis of traits. Variety-specific reference genomes reduce the likelihood of erroneous findings owing to differences in gene content between the grapevine being studied and the reference genome being utilized. In addition, the wide availability of assemblies for numerous grapevine varieties has allowed for direct comparisons between assemblies that have provided insight into basic biological questions, including the evolutionary genomics of grapevine (Zhou et al., 2019). Presently, no pangenome is available for domesticated grapevine. The next crucial step for the field of genomics of domesticated grapevine is the creation of a pangenome, which will provide a more holistic understanding of genome dynamics in domesticated grapevine and streamline future genetic studies within domesticated grapevine.

Progress assembling the genomes of wild *Vitis* species has accelerated drastically since the release of the first wild *Vitis* genome for *V. riparia* in 2019 (Girrollet et al., 2019), and assemblies are now available for at least 13 different wild *Vitis* species (Badouin et al., 2020; Cochetel et al., 2023; Girrollet et al., 2019; Holtgräwe et al., 2020; Li & Gschwend, 2023; Li et al., 2024; Massonnet et al., 2020; Patel et al., 2020; Ramos et al., 2020). The super-pangenome for wild *Vitis* species was also released in 2023 and incorporates nine

haplotype-type resolved genomes of wild North American *Vitis* species (Cochetel et al., 2023). The availability of these genome assemblies is not only important for improving our understanding of evolutionary dynamics within *Vitis* but agriculturally relevant due to several wild *Vitis* species (*V. riparia*, *V. rupestris*, and *V. berlandieri*) being commonly used as rootstocks for growing domesticated grapevines. While assemblies have been released for nearly half of the total wild *Vitis* species native to North America, only two wild Eurasian *Vitis* species have genomes available to date. Sequencing additional wild Asian *Vitis* species and incorporating these genomes into a wild *Vitis* super-pangenome will greatly improve our understanding of the evolution of *Vitis*. Many questions remain regarding dispersal and diversification of *Vitis*. The assembly of additional Asian *Vitis* species will allow for more studies comparing the genomes of North American and Asian *Vitis* species to understand key questions that remain regarding *Vitis* evolution, including how *Vitis* spread from North America to Eurasia. Past work has been inconclusive and shown support for *Vitis* spreading through North Atlantic land bridges and/or through intercontinental long-distance dispersal (Liu et al., 2016), but increased genomic resources could clarify which migration route is most likely.

CURRENT STATUS OF GRAPEVINE RESEARCH

The releases of the PN40024 grapevine reference genome and other genomic resources for domesticated grapevine have enabled substantial advances in understanding grapevine biology and elucidating the genetic basis of traits important for viticulture. Below is a summary of key molecular genetic findings in various grapevine traits that this dissertation builds upon.

Development

Due to their ability to produce tendrils and unique growth habit as lianas, grapevine development and growth is distinct from other model plant systems, such as *Arabidopsis* (*Arabidopsis thaliana*). Most work on grapevine development has focused on the development of flowers and berries, due to their agricultural importance, as well as tendrils. The grapevine orthologs to key *Arabidopsis* floral development regulators regulate floral development in grapevine as well, including orthologs to *FLOWERING LOCUS T (FT)*, *TERMINAL FLOWER1 (TFL)*, *LEAFY*, *APETALA1 (AP1)* and *FRUITFULL (FUL)* (Boss et al., 2006; Calonje et al., 2004; Carmona et al., 2007, 2008, 2002; Joly et al., 2004; Sreekantan & Thomas, 2006). Additional work has shown that several of these genes are also involved in tendril development, including *VvFUL-L* and *VvAP1*, orthologous to *FUL* and *AP1* in *Arabidopsis*, respectively (Calonje et al., 2004). As a result of the overlap between floral and tendril development genes, *Vitis* tendrils are thought to be modified inflorescences (Gerrath et al., 2015). Regulation of transitions between tendrils and flowers is strongly controlled by a combination of hormones and environmental conditions, with low temperature/light and gibberellic acid (GA) promoting the initiation of tendrils and high temperature/light and cytokinin promoting the initiation of flowers (Srinivasan & Mullins, 1981).

While several key regulators of grapevine development have been identified, the difficulty and time consuming nature of transformation within grapevine (Campos et al., 2021) has inhibited prolific work investigating the genes involved in grapevine development. However, bud sports in grapevine provide natural mutants for investigating a variety of traits, including developmental features. Bud sports are new growth on plants that display a distinct phenotype from the rest of the plant, generally due to spontaneous

somatic mutations (Foster & Aranzana, 2018). Several bud sports in grapevine exhibit defects in development (Foster & Aranzana, 2018), including the Witch's Broom bud sport, which produces dwarf vegetative growth and no flowers, among other abnormalities. In Chapter 2, I utilized two independent cases of Witch's Broom in grapevine to understand the genetic basis of these developmental defects as well as how they manifest over developmental time.

Defense against powdery mildew and other grapevine pests

Grapevine powdery mildew, caused by the fungus *Erysiphe necator*, is a major pest of grapevine. The fungus can infect most green above-ground organs of grapevine, including flowers and bunches (Gadoury et al., 2012). While infection is typically not fatal for the infected grapevine plant, it can cause substantial yield loss and reductions in berry quality (Calonnec et al., 2004). *Erysiphe necator* was introduced to Europe from North America in the 19th century and has since become a widespread pest (Qiu et al., 2015). While many wild *Vitis* species from North American exhibit resistance to powdery mildew (Dry et al., 2010; Staudt, 2015), most domesticated grapevine varieties are highly susceptible (Gaforio et al., 2015; Staudt, 2015). To reduce the pressure of powdery mildew and other fungal pests on grapevines, a colossal amount of fungicides are applied to vineyards annually. From 2001-2003, 81,000 tons of fungicides were applied to vineyards within the European Union (EU) (~67% of the total fungicides applied in the EU to crops during this period) (Eurostat, 2007). As a result, understanding the genetic mechanisms of powdery mildew defense in grapevine is an important area of research (Qiu et al., 2015). Notably, many *Vitis* species produce mite domatia (hereafter "domatia"), which provide a key mechanism for regulating powdery mildew resistance (Graham et al., 2023).

Domatia are diverse structures that form constitutively on the undersides of their leaves in many plants and facilitate mutualisms with beneficial predatory and fungivorous mites. *Vitis* specifically bear tuft domatia, which are small structures at the vein axils composed of hairs that cover a depression in the leaf surface. These domatia provide shelter to the mites that inhabit them, and in return, the mites will consume pathogenic fungi and/or small, herbivorous arthropods. Domesticated grapevine and at least 30 other species within *Vitis* produce domatia. There is substantial variation in domatia traits within species (English-Loeb & Norton, 2006; English-Loeb et al., 2002), which has been shown to be heritable (English-Loeb et al., 2002). Domatia have been the most extensively studied in *V. riparia*, likely owing to wild *V. riparia* plants generally having very dense populations of mutualistic mites on their leaves (English-Loeb et al., 1999; Norton et al., 2000) and often producing large and/or dense domatia (Graham et al., 2023). In one study, the presence of domatia on *V. riparia* plants led to a 48% reduction in powdery mildew compared to plants with blocked domatia (Norton et al., 2000), making domatia a powerful defense mechanism against powdery mildew. Despite domatia offering considerable protection against powdery mildew, very little is known about the molecular genetic mechanisms that regulate their development. In Chapter 3, I investigated the molecular genetic mechanisms that regulate domatia development and intraspecific variation in *V. riparia* domatia by performing transcriptome sequencing on domatia from two genotypes with distinct domatia phenotypes.

Berry color

Much early research on the genetic underpinnings of grapevine traits focused on berry skin color due to its importance for both table grape consumption and wine grape use

in viticulture. Consumers judge the quality of fruits based on several metrics, one of which is color (Abbott, 1999). Berry color also drives the color of red wines, which has a large influence on the perceived quality of wine (Parpinello et al., 2009; Sáenz-Navajas et al., 2011). There is substantial variation in berry skin color in grapevine, including white, pink/grey, red, and black (Ferreira et al., 2018). Berry flesh color varies as well, with most berries having white flesh but several varieties producing berries with red flesh (Ferreira et al., 2018). Many studies have set out to investigate the genetic basis of berry skin color and flesh color in domesticated grapevine.

Berry skin color is driven by the production of anthocyanins, which are colored flavonoids (Ferreira et al., 2018). Most berry skin color mutants in grapevine are the result of somatic mutations impacting the two anthocyanin genes *VvMybA1* and *VvMybA2* within the berry color locus on chromosome 2 (Ferreira et al., 2018; Walker et al., 2007). Both *VvMybA1* and *VvMybA2* are key regulators of anthocyanin biosynthesis in berry skin (Ferreira et al., 2018; S. Kobayashi et al., 2002). There are multiple known mutations impacting both or one of these genes that cause white-skinned berries in grapevine (Ferreira et al., 2018; Kobayashi et al., 2004; Walker et al., 2007, 2006; Yakushiji et al., 2006), with a common mutation being the insertion of the *Gret1* retrotransposon upstream of *VvMybA1* (Kobayashi et al., 2004; Walker et al., 2006). Further work has elucidated the basis of other less common berry skin colors that arise through somatic mutations, including grey-skinned berries and bronze-skinned berries, both of which are also caused by mutations impacting *VvMybA1* and *VvMybA2* (Walker et al., 2006).

The genetic basis of berry flesh color in grapevine has also been explored in teinturier grapes, which are unique in that they produce berries with red flesh as opposed

to white flesh. The pigmentation in teinturier berry flesh is the result of anthocyanin accumulation. While previous work has shown that a 408 bp repeat within the promoter of *VvMybA1* causes pigmentation within berry flesh of most teinturier varieties (Röckel et al., 2020; Zhang et al., 2023), other work has shown that this repeat is not essential for red berry flesh in grapevine and that red berry flesh can be caused by changes in DNA methylation (Azuma & Kobayashi, 2022). Further, the 408 bp repeat within the promoter of *VvMybA1* is present in different copy numbers (2, 3, and 5) within teinturier grape varieties, and the number of copies of this repeat has a direct impact on the total concentration of anthocyanins within the berry flesh (Röckel et al., 2020). In spite of this, teinturier grapevine varieties vary greatly in the variety of anthocyanin molecules produced and the overall composition of anthocyanins produced, regardless of the number of copies of the 408 bp repeat present upstream of *VvMybA1* (Kőrösi et al., 2022). Recently, the first teinturier grapevine genome and annotations were released for the Yan73 variety (Zhang et al., 2023). While the release of this genome provided additional insight into genes that may drive berry flesh pigmentation in teinturier grapes, without additional teinturier genomes, it is challenging to study anthocyanin variation between teinturier grapes. In Chapter 4, I present the assembly and annotation of two teinturier grape varieties, Dakapo and Rubired, which will allow for the investigation of remaining questions about anthocyanin production within berry flesh, such as the mechanisms regulating the production of specific anthocyanin molecules within the flesh of teinturier berries.

DISSERTATION PROJECTS AND SIGNIFICANCE

The objectives of this dissertation were to use modern genomic resources to (i) investigate the genetic basis of the Witch's Broom bud sport in grapevine and characterize

the developmental defects of the bud sport over developmental time, (ii) identify the key genetic pathways involved in domatia development and intraspecific variation in domatia in *V. riparia* using transcriptome sequencing, and (iii) generate a high-quality genome assembly and annotation for the Dakapo variety of grapevine.

REFERENCES

- Abbott, J. A. (1999). Quality measurement of fruits and vegetables. *Postharvest Biology and Technology*, 15(3), 207–225.
- Alston, J. M., & Sambucci, O. (2019). Grapes in the World Economy. In D. Cantu & M. A. Walker (Eds.), *The Grape Genome* (pp. 1–24). Springer International Publishing.
- Azuma, A., & Kobayashi, S. (2022). Demethylation of the 3' LTR region of retrotransposon in *VvMYBA1^{BEN}* allele enhances anthocyanin biosynthesis in berry skin and flesh in 'Brazil' grape. *Plant Science*, 322, 111341.
- Azuma, A., Kobayashi, S., Goto-Yamamoto, N., Shiraishi, M., Mitani, N., Yakushiji, H., & Koshita, Y. (2009). Color recovery in berries of grape (*Vitis vinifera* L.) 'Benitaka', a bud sport of 'Italia', is caused by a novel allele at the *VvmybA1* locus. *Plant Science*, 176(4), 470–478.
- Azuma, A., Kobayashi, S., Mitani, N., Shiraishi, M., Yamada, M., Ueno, T., Kono, A., Yakushiji, H., & Koshita, Y. (2008). Genomic and genetic analysis of *Myb*-related genes that regulate anthocyanin biosynthesis in grape berry skin. *Theoretical and Applied Genetics*, 117(6), 1009–1019.
- Badouin, H., Velt, A., Gindraud, F., Flutre, T., Dumas, V., Vautrin, S., Marande, W., Corbi, J., Sallet, E., Ganofsky, J., Santoni, S., Guyot, D., Ricciardelli, E., Jepsen, K., Käfer, J., Berges, H., Duchêne, E., Picard, F., Hugueney, P., ... Marais, G. A. B. (2020). The wild grape genome sequence provides insights into the transition from dioecy to hermaphroditism during grape domestication. *Genome Biology*, 21(1), 223.
- Blanco-Ulate, B., Amrine, K. C. H., Collins, T. S., Rivero, R. M., Vicente, A. R., Morales-Cruz, A., Doyle, C. L., Ye, Z., Allen, G., Heymann, H., Ebeler, S. E., & Cantu, D. (2015). Developmental and Metabolic Plasticity of White-Skinned Grape Berries in Response to *Botrytis cinerea* during Noble Rot. *Plant Physiology*, 169(4), 2422–2443.
- Boss, P. K., Sreekantan, L., & Thomas, M. R. (2006). A grapevine *TFL1* homologue can delay flowering and alter floral development when overexpressed in heterologous species. *Functional Plant Biology*, 33(1), 31–41.
- Bouquet, A. (1986). Introduction dans l'espèce *Vitis vinifera* L. d'un caractère de résistance à l'oidium (*Uncinula necator* Schw. Burr.) issu de l'espèce *Muscadinia rotundifolia* (Michx.) small. *Vignevine*, 12,141–146.
- Calonje, M., Cubas, P., Martínez-Zapater, J. M., & Carmona, M. J. (2004). Floral meristem identity genes are expressed during tendril development in grapevine. *Plant Physiology*, 135(3), 1491–1501.
- Calonnec, A., Cartolaro, P., Poupot, C., Dubourdieu, D., & Darriet, P. (2004). Effects of *Uncinula necator* on the yield and quality of grapes (*Vitis vinifera*) and wine. *Plant*

Pathology, 53(4), 434–445.

- Campos, G., Chialva, C., Miras, S., & Lijavetzky, D. (2021). New Technologies and Strategies for Grapevine Breeding Through Genetic Transformation. *Frontiers in Plant Science*, 12, 767522.
- Canaguier, A., Grimplet, J., Di Gaspero, G., Scalabrin, S., Duchêne, E., Choisine, N., Mohellibi, N., Guichard, C., Rombauts, S., Le Clainche, I., Bérard, A., Chauveau, A., Bounon, R., Rustenholz, C., Morgante, M., Le Paslier, M.-C., Brunel, D., & Adam-Blondon, A.-F. (2017). A new version of the grapevine reference genome assembly (12X.v2) and of its annotation (VCost.v3). *Genomics Data*, 14, 56–62.
- Carmona, M. J., Calonje, M., & Martínez-Zapater, J. M. (2007). The *FT/TFL1* gene family in grapevine. *Plant Molecular Biology*, 63(5), 637–650.
- Carmona, M. J., Chaïb, J., Martínez-Zapater, J. M., & Thomas, M. R. (2008). A molecular genetic perspective of reproductive development in grapevine. *Journal of Experimental Botany*, 59(10), 2579–2596.
- Carmona, M. J., Cubas, P., & Martínez-Zapater, J. M. (2002). *VFL*, the grapevine *FLORICAULA/LEAFY* ortholog, is expressed in meristematic regions independently of their fate. *Plant Physiology*, 130(1), 68–77.
- Chin, C.-S., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion, G. T., Clum, A., Dunn, C., O'Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., Cramer, G. R., Delledonne, M., Luo, C., Ecker, J. R., Cantu, D., Rank, D. R., & Schatz, M. C. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. *Nature Methods*, 13(12), 1050–1054.
- Cochetel, N., Minio, A., Guarracino, A., Garcia, J. F., Figueroa-Balderas, R., Massonnet, M., Kasuga, T., Londo, J. P., Garrison, E., Gaut, B. S., & Cantu, D. (2023). A super-pangenome of the North American wild grape species. *Genome Biology*, 24(1), 290.
- Divilov, K., Barba, P., Cadle-Davidson, L., & Reisch, B. I. (2018). Single and multiple phenotype QTL analyses of downy mildew resistance in interspecific grapevines. *Theoretical and Applied Genetics*, 131(5), 1133–1143.
- Dong, Y., Duan, S., Xia, Q., Liang, Z., Dong, X., Margaryan, K., Musayev, M., Goryslavets, S., Zdunić, G., Bert, P.-F., Lacombe, T., Maul, E., Nick, P., Bitskinashvili, K., Bisztray, G. D., Drori, E., De Lorenzis, G., Cunha, J., Popescu, C. F., ... Chen, W. (2023). Dual domestications and origin of traits in grapevine evolution. *Science*, 379(6635), 892–901.
- Dry, I. B., Feechan, A., Anderson, C., Jermakow, A. M., Bouquet, A., Adam-Blondon, A.-F., & Thomas, M. R. (2010). Molecular strategies to enhance the genetic resistance of grapevines to powdery mildew. *Australian Journal of Grape and Wine Research*, 16, 94–105.

- English-Loeb, G., & Norton, A. (2006). Lack of trade-off between direct and indirect defence against grape powdery mildew in riverbank grape. *Ecological Entomology*, *31*(5), 415–422.
- English-Loeb, G., Norton, A. P., Gadoury, D. M., Seem, R. C., & Wilcox, W. F. (1999). Control of Powdery Mildew in Wild and Cultivated Grapes by a Tydeid Mite. *Biological Control*, *14*(2), 97–103.
- English-Loeb, G., Norton, A. P., & Walker, M. A. (2002). Behavioral and population consequences of acarodomatia in grapes on phytoseiid mites (Mesostigmata) and implications for plant breeding. *Entomologia Experimentalis et Applicata*, *104*(2–3), 307–319.
- EUROSTAT EC. (2007). The use of plant protection products in the European Union. Data 1992-2003. Luxembourg: Office for Official Publications of the European Communities.
- Fan, J. J., Wang, P., Xu, X., Liu, K., Ruan, Y. Y., Zhu, Y. S., Cui, Z. H., & Zhang, L. J. (2015). Characterization of a TIR-NBS-LRR gene associated with downy mildew resistance in grape. *Genetics and Molecular Research*, *14*(3), 7964–7975.
- FAO. (2023). *Agricultural production statistics 2000–2022* (FAOSTAT Analytical Briefs No. 79).
- Ferreira, V., Pinto-Carnide, O., Arroyo-García, R., & Castro, I. (2018). Berry color variation in grapevine as a source of diversity. *Plant Physiology and Biochemistry*, *132*, 696–707.
- Forcato, C. (2010). *Gene prediction and functional annotation in the Vitis vinifera genome* [Università degli studi di Padova].
- Foster, T. M., & Aranzana, M. J. (2018). Attention sports fans! The far-reaching contributions of bud sport mutants to horticulture and plant biology. *Horticulture Research*, *5*, 44.
- Gade, D. W. (2015). Particularizing the Columbian exchange: Old World biota to Peru. *Journal of Historical Geography*, *48*, 26–35.
- Gadoury, D. M., Cadle-Davidson, L., Wilcox, W. F., Dry, I. B., Seem, R. C., & Milgroom, M. G. (2012). Grapevine powdery mildew (*Erysiphe necator*): a fascinating system for the study of the biology, ecology and epidemiology of an obligate biotroph. *Molecular Plant Pathology*, *13*(1), 1–16.
- Gaforio, L., García-Muñoz, S., Cabello, F., & Muñoz-Organero, G. (2015). Evaluation of susceptibility to powdery mildew (*Erysiphe necator*) in *Vitis vinifera* varieties. *Vitis*, *50*, 123–126.
- Gerrath, J. M., & Posluszny, U. (2007). Shoot architecture in the Vitaceae. *Canadian Journal of Botany*, *85*(8), 691–700.

- Gerrath, J., Posluszny, U., & Melville, L. (2015). *Taming the Wild Grape*. Springer International Publishing.
- Girollet, N., Rubio, B., Lopez-Roques, C., Valière, S., Ollat, N., & Bert, P.-F. (2019). *De novo* phased assembly of the *Vitis riparia* grape genome. *Scientific Data*, 6(1), 1–8.
- Graham, C. D. K., Forrestel, E. J., Schillmiller, A. L., Zemenick, A. T., & Weber, M. G. (2023). Evolutionary signatures of a trade-off in direct and indirect defenses across the wild grape genus, *Vitis*. *Evolution*, 77(10), 2301–2313.
- Hoffmann, S., Di Gaspero, G., Kovács, L., Howard, S., Kiss, E., Galbács, Z., Testolin, R., & Kozma, P. (2008). Resistance to *Erysiphe necator* in the grapevine ‘Kishmish vatkana’ is controlled by a single locus through restriction of hyphal growth. *Theoretical and Applied Genetics*, 116(3), 427–438.
- Holtgräwe, D., Rosleff Soerensen, T., Hausmann, L., Pucker, B., Viehöver, P., Töpfer, R., & Weisshaar, B. (2020). A Partially Phase-Separated Genome Sequence Assembly of the *Vitis* Rootstock “Börner” (*Vitis riparia* × *Vitis cinerea*) and Its Exploitation for Marker Development and Targeted Mapping. *Frontiers in Plant Science*, 11, 156.
- International Organisation of Vine and Wine. (2017). *Distribution of the world's grapevine varieties*. <https://www.oiv.int/public/medias/5888/en-distribution-of-the-worlds-grapevine-varieties.pdf>.
- Jaillon, O., Aury, J.-M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., Vezzi, A., Legeai, F., Huguene, P., Dasilva, C., Horner, D., Mica, E., Jublot, D., Poulain, J., Bruyère, C., ... Wincker, P. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, 449(7161), 463–467.
- Joly, D., Perrin, M., Gertz, C., Kronenberger, J., Demangeat, G., & Masson, J. E. (2004). Expression analysis of flowering genes from seedling-stage to vineyard life of grapevine cv. Riesling. *Plant Science*, 166(6), 1427–1436.
- Kobayashi, S., Ishimaru, M., Hiraoka, K., & Honda, C. (2002). *Myb*-related genes of the Kyoho grape (*Vitis labruscana*) regulate anthocyanin biosynthesis. *Planta*, 215(6), 924–933.
- Kobayashi, Shozo, Goto-Yamamoto, N., & Hirochika, H. (2004). Retrotransposon-induced mutations in grape skin color. *Science*, 304(5673), 982.
- Kőrösi, L., Molnár, S., Teszlák, P., Dörnyei, Á., Maul, E., Töpfer, R., Marosvölgyi, T., Szabó, É., & Röckel, F. (2022). Comparative Study on Grape Berry Anthocyanins of Various Teinturier Varieties. *Foods*, 11(22).
- Lev-Yadun, S., Gopher, A., & Abbo, S. (2000). Archaeology. The cradle of agriculture. *Science*, 288(5471), 1602–1603.

- Li, B., & Gschwend, A. R. (2023). *Vitis labrusca* genome assembly reveals diversification between wild and cultivated grapevine genomes. *Frontiers in Plant Science*, *14*, 1234130.
- Li, H., Liu, Y., Fan, P., Dai, Z., Hao, J., Duan, W., Liang, Z., & Wang, Y. (2024). The Genome of *Vitis zhejiang-adstricta* Strengthens the Protection and Utilization of the Endangered Ancient Grape Endemic to China. *Plant & Cell Physiology*, *65*(2), 216–227.
- Liu, X.-Q., Ickert-Bond, S. M., Nie, Z.-L., Zhou, Z., Chen, L.-Q., & Wen, J. (2016). Phylogeny of the *Ampelocissus-Vitis* clade in Vitaceae supports the New World origin of the grape genus. *Molecular Phylogenetics and Evolution*, *95*, 217–228.
- Ma, Z.-Y., Nie, Z.-L., Ren, C., Liu, X.-Q., Zimmer, E. A., & Wen, J. (2021). Phylogenomic relationships and character evolution of the grape family (Vitaceae). *Molecular Phylogenetics and Evolution*, *154*, 106948.
- Maestri, S., Gambino, G., Lopatriello, G., Minio, A., Perrone, I., Cosentino, E., Giovannone, B., Marcolungo, L., Alfano, M., Rombauts, S., Cantu, D., Rossato, M., Delledonne, M., & Calderón, L. (2022). “Nebbiolo” genome assembly allows surveying the occurrence and functional implications of genomic structural variations in grapevines (*Vitis vinifera* L.). *BMC Genomics*, *23*(1), 159.
- Massonnet, M., Cochetel, N., Minio, A., Vondras, A. M., Lin, J., Muyle, A., Garcia, J. F., Zhou, Y., Delledonne, M., Riaz, S., Figueroa-Balderas, R., Gaut, B. S., & Cantu, D. (2020). The genetic basis of sex determination in grapes. *Nature Communications*, *11*(1), 2902.
- McGovern, P. E. (2013). *Ancient Wine*. Princeton University Press.
- McGovern, P., Jalabadze, M., Batiuk, S., Callahan, M. P., Smith, K. E., Hall, G. R., Kvavadze, E., Maghradze, D., Rusishvili, N., Bouby, L., Failla, O., Cola, G., Mariani, L., Boaretto, E., Bacilieri, R., This, P., Wales, N., & Lordkipanidze, D. (2017). Early Neolithic wine of Georgia in the South Caucasus. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(48), E10309–E10318.
- Minio, A., Cochetel, N., Figueroa-Balderas, R., & Cantu, D. (2024a). *Grapegenomics.com - Genome release: Vitis vinifera cv. Chardonnay cl. 04 v2.0*. <https://doi.org/10.5281/zenodo.10578344>.
- Minio, A., Cochetel, N., Figueroa-Balderas, R., Cantu, D. (2024b). *Grapegenomics.com - Genome release: Vitis vinifera cv. Muscat of Alexandria cl. FPS02*. <https://doi.org/10.5281/zenodo.10570019>.
- Minio, A., Cochetel, N., Vondras, A. M., Massonnet, M., & Cantu, D. (2022). Assembly of complete diploid-phased chromosomes from draft genome sequences. *G3*, *12*(8).
- Minio, A., Massonnet, M., Figueroa-Balderas, R., Castro, A., & Cantu, D. (2019). Diploid Genome Assembly of the Wine Grape Carménère. *G3*, *9*(5), 1331–1337.

- Myles, S., Boyko, A. R., Owens, C. L., Brown, P. J., Grassi, F., Aradhya, M. K., Prins, B., Reynolds, A., Chia, J.-M., Ware, D., Bustamante, C. D., & Buckler, E. S. (2011). Genetic structure and domestication history of the grape. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(9), 3530–3535.
- Norton, A. P., English-Loeb, G., Gadoury, D., & Seem, R. C. (2000). Mycophagous mites and foliar pathogens: Leaf domatia mediate tritrophic interactions in grapes. *Ecology*, *81*(2), 490–499.
- Onetto, C. A., Ward, C. M., & Borneman, A. R. (2023). The Genome Assembly of *Vitis vinifera* cv. Shiraz. *Australian Journal of Grape and Wine Research*, 2023.
- Parpinello, G. P., Versari, A., Chinnici, F., & Galassi, S. (2009). Relationship among sensory descriptors, consumer preference and color parameters of Italian Novello red wines. *Food Research International*, *42*(10), 1389–1395.
- Patel, S., Robben, M., Fennell, A., Londo, J. P., Alahakoon, D., Villegas-Diaz, R., & Swaminathan, P. (2020). Draft genome of the Native American cold hardy grapevine *Vitis riparia* Michx. “Manitoba 37.” *Horticulture Research*, *7*(1), 92.
- Pauquet, J., Bouquet, A., This, P., & Adam-Blondon, A.-F. (2001). Establishment of a local map of AFLP markers around the powdery mildew resistance gene *Run1* in grapevine and assessment of their usefulness for marker assisted selection. *Theoretical and Applied Genetics*, *103*(8), 1201–1210.
- Pollard, M. O., Gurdasani, D., Mentzer, A. J., Porter, T., & Sandhu, M. S. (2018). Long reads: their purpose and place. *Human Molecular Genetics*, *27*(R2), R234–R241.
- Qiu, W., Feechan, A., & Dry, I. (2015). Current understanding of grapevine defense mechanisms against the biotrophic fungus (*Erysiphe necator*), the causal agent of powdery mildew disease. *Horticulture Research*, *2*, 15020.
- Qu, J., Dry, I., Liu, L., Guo, Z., & Yin, L. (2021). Transcriptional profiling reveals multiple defense responses in downy mildew-resistant transgenic grapevine expressing a TIR-NBS-LRR gene located at the *MrRUN1/MrRPV1* locus. *Horticulture Research*, *8*(1), 161.
- Ramos, M. J. N., Coito, J. L., Faísca-Silva, D., Cunha, J., Costa, M. M. R., Amâncio, S., & Rocheta, M. (2020). Portuguese wild grapevine genome re-sequencing (*Vitis vinifera sylvestris*). *Scientific Reports*, *10*(1), 18993.
- Read, S. (2015). *Early vineyards and viticulture in the Sydney basin*. https://www.gardenhistorysociety.org.au/wp-content/uploads/2023/05/Stuart-Read-Vineyards-of-Sydney_15-10-15-SA-Symposium-paper.pdf.
- Riehl, S., Zeidi, M., & Conard, N. J. (2013). Emergence of agriculture in the foothills of the Zagros Mountains of Iran. *Science*, *341*(6141), 65–67.

- Röckel, F., Moock, C., Braun, U., Schwander, F., Cousins, P., Maul, E., Töpfer, R., & Hausmann, L. (2020). Color Intensity of the Red-Fleshed Berry Phenotype of *Vitis vinifera* Teinturier Grapes Varies Due to a 408 bp Duplication in the Promoter of *VvmybA1*. *Genes*, *11*(8).
- Rubio, S., Noriega, X., & Pérez, F. J. (2019). Abscisic acid (ABA) and low temperatures synergistically increase the expression of CBF/DREB1 transcription factors and cold-hardiness in grapevine dormant buds. *Annals of Botany*, *123*(4), 681–689.
- Sáenz-Navajas, M.-P., Echavarri, F., Ferreira, V., & Fernández-Zurbano, P. (2011). Pigment composition and color parameters of commercial Spanish red wine samples: linkage to quality perception. *European Food Research and Technology*, *232*(5), 877–887.
- Satam, H., Joshi, K., Mangrolia, U., Waghoo, S., Zaidi, G., Rawool, S., Thakare, R. P., Banday, S., Mishra, A. K., Das, G., & Malonia, S. K. (2023). Next-Generation Sequencing Technology: Current Trends and Advancements. *Biology*, *12*(7).
- Sichel, V., Sarah, G., Girollet, N., Laucou, V., Roux, C., Roques, M., Mournet, P., Cunff, L. L., Bert, P. F., This, P., & Lacombe, T. (2023). Chimeras in Merlot grapevine revealed by phased assembly. *BMC Genomics*, *24*(1), 396.
- Slatko, B. E., Gardner, A. F., & Ausubel, F. M. (2018). Overview of Next-Generation Sequencing Technologies. *Current Protocols in Molecular Biology*, *122*(1), e59.
- Sreekantan, L., & Thomas, M. R. (2006). *VvFT* and *VvMADS8*, the grapevine homologues of the floral integrators *FT* and *SOC1*, have unique expression patterns in grapevine and hasten flowering in *Arabidopsis*. *Functional Plant Biology*, *33*(12), 1129–1139.
- Srinivasan, C., & Mullins, M. G. (1981). Physiology of Flowering in the Grapevine — a Review. *American Journal of Enology and Viticulture*, *32*(1), 47–63.
- Staudt, G. (2015). Evaluation of resistance to grapevine powdery mildew (*Uncinula necator* [Schw.] Burr., anamorph *Oidium tuckeri* Berk.) in accessions of *Vitis* species. *Vitis*, *36*, 151–154.
- Tello, J., Mammerler, R., Čajić, M., & Forneck, A. (2019). Major Outbreaks in the Nineteenth Century Shaped Grape Phylloxera Contemporary Genetic Structure in Europe. *Scientific Reports*, *9*(1), 17540.
- The French-Italian Public Consortium. (2009). *12X.0 version of the grapevine reference genome sequence from The French-Italian Public Consortium (PN40024)*. <https://urgi.versailles.inra.fr/Species/Vitis/Data-Sequences/Genome-sequences>.
- This, P., Lacombe, T., & Thomas, M. R. (2006). Historical origins and genetic diversity of wine grapes. *Trends in Genetics*, *22*(9), 511–519.
- Tyson, J. R., O’Neil, N. J., Jain, M., Olsen, H. E., Hieter, P., & Snutch, T. P. (2018). MinION-based

- long-read sequencing and assembly extends the *Caenorhabditis elegans* reference genome. *Genome Research*, 28(2), 266–274.
- Urta, C., Sanhueza, D., Pavez, C., Tapia, P., Núñez-Lillo, G., Minio, A., Miossec, M., Blanco-Herrera, F., Gainza, F., Castro, A., Cantu, D., & Meneses, C. (2023). Identification of grapevine clones via high-throughput amplicon sequencing: a proof-of-concept study. *G3*, 13(9).
- Velt, A., Frommer, B., Blanc, S., Holtgräwe, D., Duchêne, É., Dumas, V., Grimplet, J., Huguene, P., Kim, C., Lahaye, M., Matus, J. T., Navarro-Payá, D., Orduña, L., Tello-Ruiz, M. K., Vitulo, N., Ware, D., & Rustenholz, C. (2023). An improved reference of the grapevine genome reasserts the origin of the PN40024 highly homozygous genotype. *G3*, 13(5).
- Vondras, A. M., Lerno, L., Massonnet, M., Minio, A., Rowhani, A., Liang, D., Garcia, J., Quiroz, D., Figueroa-Balderas, R., Golino, D. A., Ebeler, S. E., Al Rwahnih, M., & Cantu, D. (2021). Rootstock influences the effect of grapevine leafroll-associated viruses on berry development and metabolism via abscisic acid signalling. *Molecular Plant Pathology*, 22(8), 984–1005.
- Vondras, A. M., Minio, A., Blanco-Ulate, B., Figueroa-Balderas, R., Penn, M. A., Zhou, Y., Seymour, D., Ye, Z., Liang, D., Espinoza, L. K., Anderson, M. M., Walker, M. A., Gaut, B., & Cantu, D. (2019). The genomic diversification of grapevine clones. *BMC Genomics*, 20(1), 972.
- Walker, A. R., Lee, E., Bogs, J., McDavid, D. A. J., Thomas, M. R., & Robinson, S. P. (2007). White grapes arose through the mutation of two similar and adjacent regulatory genes. *The Plant Journal*, 49(5), 772–785.
- Walker, A. R., Lee, E., & Robinson, S. P. (2006). Two new grape cultivars, bud sports of Cabernet Sauvignon bearing pale-coloured berries, are the result of deletion of two regulatory genes of the berry colour locus. *Plant Molecular Biology*, 62(4–5), 623–635.
- Wang, Y., Zhao, Y., Bollas, A., Wang, Y., & Au, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology*, 39(11), 1348–1365.
- Wen, J., Lu, L.-M., Nie, Z.-L., Liu, X.-Q., Zhang, N., Ickert-Bond, S., Gerrath, J., Manchester, S. R., Boggan, J., & Chen, Z.-D. (2018). A new phylogenetic tribal classification of the grape family (Vitaceae). *Journal of Systematics and Evolution*, 56(4), 262–272.
- Yakushiji, H., Kobayashi, S., Goto-Yamamoto, N., Tae Jeong, S., Sueta, T., Mitani, N., & Azuma, A. (2006). A skin color mutation of grapevine, from black-skinned Pinot Noir to white-skinned Pinot Blanc, is caused by deletion of the functional *VvmybA1* allele. *Bioscience, Biotechnology, and Biochemistry*, 70(6), 1506–1508.
- Zecca, G., Abbott, J. R., Sun, W.-B., Spada, A., Sala, F., & Grassi, F. (2012). The timing and the mode of evolution of wild grapes (*Vitis*). *Molecular Phylogenetics and Evolution*,

62(2), 736–747.

Zeng, L., Zhang, N., Zhang, Q., Endress, P. K., Huang, J., & Ma, H. (2017). Resolution of deep eudicot phylogeny and their temporal diversification using nuclear genes from transcriptomic and genomic datasets. *The New Phytologist*, *214*(3), 1338–1354.

Zhang, K., Du, M., Zhang, H., Zhang, X., Cao, S., Wang, X., Wang, W., Guan, X., Zhou, P., Li, J., Jiang, W., Tang, M., Zheng, Q., Cao, M., Zhou, Y., Chen, K., Liu, Z., & Fang, Y. (2023). The haplotype-resolved T2T genome of teinturier cultivar Yan73 reveals the genetic basis of anthocyanin biosynthesis in grapes. *Horticulture Research*, *10*(11), uhad205.

Zhou, Y., Minio, A., Massonnet, M., Solares, E., Lv, Y., Beridze, T., Cantu, D., & Gaut, B. S. (2019). The population genetics of structural variants in grapevine domestication. *Nature Plants*, *5*(9), 965–979.

Zou, C., Massonnet, M., Minio, A., Patel, S., Llaca, V., Karn, A., Gouker, F., Cadle-Davidson, L., Reisch, B., Fennell, A., Cantu, D., Sun, Q., & Londo, J. P. (2021). Multiple independent recombinations led to hermaphroditism in grapevine. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(15).

CHAPTER 2: From buds to shoots: insights into grapevine development from the Witch's Broom bud sport

This chapter has been published in *BMC Plant Biology*:

Ritter, E. J., Cousins, P., Quigley, M., Kile, A., Kenchanmane Raju, S. K., Chitwood, D. H., & Niederhuth, C. (2024). From buds to shoots: insights into grapevine development from the Witch's Broom bud sport. *BMC Plant Biology*, 24(283).

AUTHORS AND AFFILIATIONS

Eleanore J. Ritter¹, Peter Cousins², Michelle Quigley^{3,4}, Aidan Kile¹, Sunil K. Kenchanmane Raju^{1,5}, Daniel H. Chitwood^{3,6}, and Chad Niederhuth¹

¹Department of Plant Biology, Michigan State University, East Lansing, MI, USA

²E. & J. Gallo Winery, Modesto, CA, USA

³Department of Horticulture, Michigan State University, East Lansing, MI, USA

⁴Center for Quantitative Imaging, Institute of Energy and the Environment, Penn State University, State College, PA, USA

⁵Center for Genomics and Systems Biology, New York University, Manhattan, NY, USA

⁶Department of Computational Mathematics, Science & Engineering, Michigan State University, East Lansing, MI, USA

AUTHOR CONTRIBUTIONS

CN, EJR, and PC conceptualized and designed the study. PC provided plant material and collected shoot phenotype data. AK and EJR performed leaf landmarking. MQ performed CT scanning of buds. DHC, EJR, and MQ analyzed internal bud morphologies. SKKR isolated DNA and generated libraries for Illumina sequencing. AK and EJR validated potential candidate variants. EJR analyzed the data. CN, DHC, EJR, and PC assisted in data

interpretation. EJR wrote the first draft of the manuscript. All authors assisted with final drafts of the manuscript.

ABSTRACT

Background

Bud sports occur spontaneously in plants when new growth exhibits a distinct phenotype from the rest of the parent plant. The Witch's Broom bud sport occurs occasionally in various grapevine (*Vitis vinifera*) varieties and displays a suite of developmental defects, including dwarf features and reduced fertility. While it is highly detrimental for grapevine growers, it also serves as a useful tool for studying grapevine development. We used the Witch's Broom bud sport in grapevine to understand the developmental trajectories of the bud sports, as well as the potential genetic basis. We analyzed the phenotypes of two independent cases of the Witch's Broom bud sport, in the Dakapo and Merlot varieties of grapevine, alongside wild type counterparts. To do so, we quantified various shoot traits, performed 3D X-ray Computed Tomography on dormant buds, and landmarked leaves from the samples. We also performed Illumina and Oxford Nanopore sequencing on the samples and called genetic variants using these sequencing datasets.

Results

The Dakapo and Merlot cases of Witch's Broom displayed severe developmental defects, with no fruit/clusters formed and dwarf vegetative features. However, the Dakapo and Merlot cases of Witch's Broom studied were also phenotypically different from one another, with distinct differences in bud and leaf development. We identified 968-974 unique genetic mutations in our two Witch's Broom cases that are potential causal variants of the bud sports. Examining gene function and validating these genetic candidates through

PCR and Sanger-sequencing revealed one strong candidate mutation in Merlot Witch's Broom impacting the gene GSVIVG01008260001.

Conclusions

The Witch's Broom bud sports in both varieties studied had dwarf phenotypes, but the two instances studied were also vastly different from one another and likely have distinct genetic bases. Future work on Witch's Broom bud sports in grapevine could provide more insight into development and the genetic pathways involved in grapevine.

BACKGROUND

Bud sports arise when a part of a plant, such as a lateral shoot, develops phenotypic differences from the rest of the parental plant. They typically arise when a somatic mutation occurs within a developing meristem and then spreads throughout the meristem and developing tissue (Foster & Aranzana, 2018). Bud sports are known to arise sporadically in many perennial crops and can be an important source of novel phenotypes, having given rise to many plant cultivars widely grown today. They can be an especially important source of variation in difficult to breed perennial crops, such as grapevine (*Vitis vinifera*), which is challenging to breed due to high genetic heterozygosity and long regeneration times. As a result, beneficial bud sports in grapevines have often propagated to be grown as new varieties. For example, the variety Tempranillo Blanco first arose as a bud sport of Tempranillo Tinto and was clonally propagated to maintain its novel phenotype (Carbonell-Bejerano et al., 2017). Bud sports are not always beneficial and sometimes detrimental to agricultural production, however, such bud sports provide natural mutants that can still be leveraged to study developmental traits that might otherwise not be possible (Foster & Aranzana, 2018).

Grapevines have unique development and physiology compared to many other crops and model systems. They are perennial plants that grow as lianas (also known as woody vines). This growth habit is enabled by tendrils, which are uncommon structures that allow them to climb as they grow. In addition, unlike many plants, their shoot tip does not terminate in an inflorescence, but instead contains an uncommitted primordium that allows the plants to continue growing from the tip (Gerrath et al., 2015). Development within the buds of grapevines themselves is uniquely organized to ensure the successful production of leaves, tendrils, and inflorescences from the primordia. While tendril origin differs on a species basis, grapevine tendrils are modified inflorescences (Gerrath et al., 2015). The switch from inflorescence development to tendril development occurs within the developing buds and is tightly regulated by a mixture of environmental conditions and hormones. Cytokinin signaling, high light, and high temperature promote inflorescence development while gibberellic acid (GA) signaling, low light, and low temperature promote tendril development (Srinivasan & Mullins, 1981). Changes in hormones regulating these structures can have significant impacts on the ability of *V. vinifera* to sexually reproduce, even causing seed abortion (Cheng et al., 2015). However, understanding the regulatory and genetic components involved in grapevine development has proved challenging due to the difficulty of conducting genetic and molecular studies in grapevine.

Witch's Broom (WB) is a bud sport that occurs spontaneously in multiple grapevine varieties. The WB phenotype involves prolific vegetative growth and limited to no production of flowers (Bettiga et al., 2013). In contrast to wild type (WT), the WB bud sport does not easily root from cuttings, although the WB sport may be propagated by grafting. Similar WB bud sport phenotypes in other plant species are usually the result of pathogen

infection, typically by phytoplasma (Jung, 2002; Khan et al., 2002; Montano et al., 2001). However, genetic mutations have been shown to cause WB, as with the WB shoots in *Pinus sibirica* (Zhuk et al., 2015). Cases of WB in grapevine are thought to have arisen through genetic causes and not pathogen infection. Instances of WB in grapevine do not spread within or between plants and have also occurred in plants that tested negative for pathogens. Therefore, WB bud sports in grapevines are thought to have arisen from genetic causes. As a result, the WB bud sport could be valuable for research, providing insight into an aspect of grapevine development and the genetic factors behind it, that would otherwise be near impossible to study. Here, we investigate both the phenotypic effects and the potential genetic underpinnings of two independent cases of the grapevine WB bud sport. Our results demonstrate that the WB bud sport impacts grapevine development from buds to shoots, but in distinct ways in the two cases we studied. Our work also suggests that the basis for the WB bud sports may result from mutations in different genes.

METHODS

Plant material

Two independent cases of WB from two grapevine varieties, Merlot and Dakapo (*Vitis vinifera* L.), were sequenced and phenotyped alongside tissue from WT branches. The Merlot WT and WB samples were derived from the same plant, while the Dakapo WT and WB were derived from two separate plants.

The Merlot WB was identified as a bud sport on a vine of a Merlot plant in a commercial vineyard in Madera, California, USA that was planted in 1994 after being grafted to Harmony rootstock. The vineyard is trained to bilateral cordons, spur pruned, and planted with rows on an east/west orientation. The proband vine was observed in

2013 to have one arm with wild type shoots (the western arm) and one arm with WB shoots (the eastern arm). The plant material both collected and studied come from a mixture of the original proband Merlot vine and cuttings derived from it. The tissue samples used for short read sequencing were collected from the contrasting arms of the original proband Merlot vine for both Merlot WT and Merlot WB. Observations and tissue samples used for long read sequencing of the Merlot WB were from the WB arm of the original proband vine as well. In 2020, budwood was collected from the WB arm of the proband vine and bench grafted to Rupestris St. George rootstock by the commercial nursery Wonderful Nurseries in Wasco, California, USA. The Merlot WB cuttings used for imaging buds were collected (February 2021) from those grafted Merlot WB vines planted in Madera, California, USA in 2020. Cuttings from shoots on the Merlot WT arm of the proband vine were made in 2018 and rooted by the commercial nursery Greenheart Farms in Arroyo Grande, California, USA in individual pots. The vines resulting from those cuttings were planted in Madera, California, USA in 2018 and trained to bilateral cordons and spur pruned. Observations, tissue samples for long read sequencing, and cuttings of Merlot WT were collected from these planted cuttings from the proband vine.

The Dakapo WB was identified as a whole vine sport on a vine in a budwood increase block in Madera, California, USA that was planted in 2011. A budwood increase block is cultivated to provide propagation wood for grafting or cuttings rather than fruit for commercial production. The proband vine was observed in 2013 to demonstrate the WB phenotype, in contrast to nearby Dakapo WT vines of the same age in the same block. Budwood was collected from the proband vine and bench grafted in 2015 onto 140 Ruggeri rootstock by the commercial nursery Duarte Nursery in Hughson, California, USA. The

grafted vines were planted in 2015. Observations and all samples of the Dakapo WB come from a single grafted vine. Observations and all samples of Dakapo WT are from the original Dakapo vines planted in the budwood increase block in 2011.

Phenotyping of the WB bud sport

Shoot and leaf phenotyping was conducted on samples from field grown vines in Madera, California, USA in September 2021. Ten shoots were examined per accession (Merlot WB, Merlot WT, Dakapo WB, Dakapo WT). For WT vines, fertile (with fruit clusters) shoots from retained nodes were observed. For WB vines, shoots from retained nodes were observed. Retained nodes are nodes with dormant buds chosen by professional pruners during dormant pruning as the most likely to produce healthy shoots in an appropriate position during the subsequent growing season and ordinarily the shoots from retained are the most fruitful shoots on a grapevine. Lateral meristem presence and type was recorded for 16 nodes beginning at the basal end of the shoot. The lateral meristem choices were tendril, cluster, and shoot. If a scar was present indicating the loss of the lateral meristem, this was recorded as “scar” since the type of lateral meristem could not be determined by observation. Skipped nodes where no lateral meristem was present were recorded as a “skip”. The length of 16 internodes basal to those nodes was recorded. The maximum blade length, maximum blade width and the petiole length of five fully expanded undamaged leaves at or distal to the cluster zone were recorded from each of the ten shoots per accession.

Leaf landmarking and analysis

Between 12-14 leaves were collected from six shoots per sample from plants in Madera, California, USA in June 2022. The sampled shoots grew from retained nodes.

Leaves were pressed in an herbarium press at Madera, California, USA and shipped in the press to East Lansing, Michigan, USA for scanning and analysis. The leaves were scanned using a CanoScan 9000F Mark II (Canon U.S.A., Inc) at 600 DPI. The leaves were landmarked manually by placing 21 landmarks from Bryson et al. (2020) on leaf scans using ImageJ v1.53k (Abramoff et al., 2004). Scans were saved as x- and y-coordinates in centimeters. The shoelace algorithm, originally described by Meister (1769), was used to calculate leaf, vein, and blade areas using the landmarks. The landmarks were used as the vertices of polygons and the following formula, as described in Chitwood et al. (2021), was used to then calculate the areas (where n represents the number of polygon vertices defined by the landmarked x and y coordinates):

$$\frac{1}{2} |x_1y_2 + x_2y_3 + \dots + x_{n-1}y_n + x_ny_1 - x_2y_1 - x_3y_2 - \dots - x_ny_{n-1} - x_1y_n|.$$

To investigate changes in leaf shape between WT and WB leaves, a generalized Procrustes analysis and a principal components analysis (PCA) was performed using the shapes package v1.2.7 (Dryden & Mardia, 2016) in R v4.2.2 (R Core Team, 2022) and RStudio v2022.12.0.353 (Rstudio Team, 2022), with scaling and rotation. The shapes package v1.2.7 (Dryden & Mardia, 2016) in R and RStudio was also used to test for mean shape differences using a Hotelling's T^2 test.

Data visualization

All plots were made in R using ggplot2 v3.4.2 (Wickham, 2016) and arranged using cowplot v1.1.1 (Wilke, 2021). The R package ggsignif v0.6.4 was used to add significance bars to violin plots (Constantin & Patil, 2021). The R package ggnewscale v0.4.8 was used to plot distinct scales for WT and WB data when needed (Campitelli, 2022).

Bud collection, dissecting, and imaging

Dormant grapevine cuttings were collected in Madera, California, USA in February 2021 and shipped overnight to East Lansing, Michigan, USA. The Dakapo WT and Merlot WT cuttings were between 6-7 mm in diameter, while the Dakapo WB and Merlot WB cuttings were between 4-5 mm in diameter. Cuttings were left at room temperature for 24-72 hours before dissecting. Only live cuttings were used for bud dissection. The buds were dissected using a razor, slicing the buds vertically (parallel to the stem) until the primary, secondary, and tertiary buds could all be seen, but tendrils primordia were still distinguishable. Buds were then imaged with a dissecting microscope.

Buds were also scanned to create 3D X-ray Computed Tomography (CT) reconstructions of internal anatomy. Three individual buds were scanned from Merlot WT cuttings, and four individual buds were scanned from cuttings for the other three samples. The scans were produced using the X3000 system (North Star Imaging) and the included efX software (North Star Imaging). The scans were taken at 75 kV and 100 μ amps with a frame rate of 12.5 frames per second in continuous mode. 2880 projections and 2 frame averages were used. To obtain the maximum voxel size (4.5 μ m), a subpixel scan, which takes 4 scans at half a pixel distance and combines them to get approximately half the voxel size, was used (see scale, Figure 2.5). The 3D reconstruction of the buds was computed with the efX-CT software. efX-View software was used to visualize 2D slices through the 3D reconstructions of the buds.

Whole genome sequencing and alignment

Leaf tissue samples for sequencing were collected from all four accessions (Merlot WB, Merlot WT, Dakapo WB, Dakapo WT) in August 2018. DNA isolation was performed

using the CTAB method as described in Porebski et al. (1997). Library preparation for paired-end (PE) sequencing was performed as in Urich et al. (2015) with slight modification and sequenced on a HiSeq 2500 (Illumina, Inc.) with 150 base pair (bp) PE reads sequenced to 50-58X coverage. The reads were then prepared for downstream analysis, first using cutadapt v3.7 (Martin, 2011) to trim adapters and low-quality bases from the beginning and ends of reads with the following parameters: *q 20,20, --trim-n, -m 30, and -n 3*. The quality of the reads, both before and after trimming, were checked using FastQC v0.11.9 (Andrews, 2010). The trimmed reads were then mapped to the 12X.v2 grapevine reference genome assembly (Canaguier et al., 2017) using BWA-MEM v0.7.17 and the *-M* parameter (Li & Durbin, 2009). Mapped reads were then prepared for variant calling by sorting them with Samtools v1.9 (Danecek et al., 2021) and marking duplicate reads using Picard MarkDuplicates v2.15.0 (Broad Institute, 2017). The reads were then indexed using Samtools v1.9 (Danecek et al., 2021), to enable use with downstream variant callers.

SNP calling and annotation

The GATK v4.0.12.0 (McKenna et al., 2010) pipeline for short variant discovery was used to call SNPs and INDELS in the samples using the BAM files with marked duplicates (DePristo et al., 2011). GATK HaplotypeCaller was used to call SNPs and INDELS in the individual samples. The SNPs and INDELS were combined into one file and genotyped using GATK CombineGVCFs and GenotypeGVCFs, respectively. They were filtered with GATK VariantFiltration (DePristo et al., 2011; Van der Auwera et al., 2013), using the following filters: *MQ<40.00, FS>60.0, QD<2.0, MQRankSum<-12.5, and ReadPosRankSum<-8.0*. These filters were chosen based on GATK's recommendations for hard filtering germline short

variants (Caetano-Anolles, 2024). No additional filtering was done in order to avoid over-filtering and introducing false negatives that would reduce our likelihood of identifying casual variants. ANNOVAR was used to annotate the SNPs and INDELS (Wang et al., 2010) with the Genoscope 12X grapevine genome annotation (Jaillon et al., 2007) lifted to the 12X.v2 grapevine genome assembly (Canaguier et al., 2017) using Liftoff (Shumate & Salzberg, 2021) with the *-copies* parameter to minimize compatibility issues the newest grapevine genome annotation (Canaguier et al., 2017) had with downstream analyses.

Long read sequencing

New tissue was collected for Oxford Nanopore Technologies (ONT) sequencing in July 2021. The tissue samples used were young leaves collected from actively growing shoot tips. The samples were frozen and shipped on dry ice overnight. The MSU Genomics Core extracted DNA from the samples and prepared the sequencing libraries. DNA was isolated from samples using a modified Qiagen Genomic-tip protocol (Qiagen) (Qiagen, 2015) with 5 mg lysing enzyme (0.5 mg/ml; L1412-5G; Sigma-Aldrich, Inc.), 5 mg Pectinase (0.5mg/ml; P2401; Sigma-Aldrich, Inc.), and 500 µl Viscozyme L (5%; V2010-50; MilliporeSigma) added to the lysis buffer. Short read elimination was performed using the Circulomics Short Read Eliminator kit (formerly SS-100-101-01, now SKU 102-208-300; Pacific Biosciences). The size selected DNA was quantified using a Qubit 1.0 Fluorometer (Thermo Fisher Scientific) and the Qubit dsDNA BR (Broad Range) Assay (Q32853; Thermo Fisher Scientific). Barcoded sequencing libraries were then prepared using the Ligation Sequencing Kit 1D (SQK-LSK109; Oxford Nanopore Technologies) and Native Barcoding Expansion Kit (EXP-NBD104; Oxford Nanopore Technologies). The pooled libraries were then loaded on a PromethION FLO-PRO002 (R9.4.1; Oxford Nanopore Technologies) flow

cell and sequenced on a PromethION24 (Oxford Nanopore Technologies), running MinKNOW Release 21.11.7 (Oxford Nanopore Technologies), to 19-31X coverage. Base calling, demultiplexing, and filtering were done using Guppy v5.1.13 (Oxford Nanopore Technologies) with the High Accuracy base calling model. Only reads with a mean Q-score ≥ 9 were kept.

Long read alignment and structural variant calling

Adapters were trimmed from the ONT reads using Porechop v0.2.4 (Wick et al., 2017) with the following parameters: `--min_trim_size 5, --extra_end_trim 2, --end_threshold 80, --middle_threshold 90, --extra_middle_trim_good_side 2, --extra_middle_trim_bad_side 50,` and `--min_split_read_size 300`. NanoLyse v1.2.0 was used to remove ONT reads mapping to the lambda phage genome (De Coster et al., 2018). Low-quality reads and reads shorter than 300 base pairs (bp) were removed using NanoFilt v2.8.0 (De Coster et al., 2018) with the following parameters: `-q 0` and `-l 300`. The quality of the trimmed and filtered reads was analyzed using FastQC v0.11.9 (Andrews, 2010).

The ONT reads were mapped to the 12X.v2 grapevine reference genome assembly (Canaguier et al., 2017) using minimap2 v2.23-r1111 (Li, 2021) two separate times with different parameters based on the needs of downstream programs. For use with sniffles v2.0.6 (Smolka et al., 2024) to call structural variants (SVs), ONT reads were mapped with minimap2 v2.23-r1111 (Li, 2021) and the following parameters: `-ax map-ont --MD`. The mapped reads were sorted with Samtools v1.9 (Danecek et al., 2021). Sniffles v2.0.6 (Smolka et al., 2024) was first run on sorted mapped read files for all samples separately using the `--snf` parameter to generate .snf files for all samples. Sniffles v2.0.6 (Smolka et al., 2024) was then run on the .snf files previously generated for WT and WB samples from the

same variety, running Dakapo and Merlot separately, to create a VCF file with SVs.

The second version of ONT read mapping used minimap2 v2.23-r1111 (Li, 2021) with parameters optimized for use with pbsv v2.8.0 (Pacific Biosciences) (Pacific Biosciences, 2021), an additional SV caller: `-a --MD --eqx -L -O 5,56 -E 4,1 -B 5 --secondary=no -z 400,50 -r 2k -Y`. Samtools v1.9 (Danecek et al., 2021) was used to sort the mapped reads and add read groups. The sorted mapped read files were then used with pbsv v2.8.0 “discover”, running all samples separately to first discover signatures of structural variation and produce a .svsig file. A VCF file with SVs was then generated by running pbsv v2.8.0 “call” (Pacific Biosciences, 2021) with .svsig files for WT and WB samples from the same variety (with Dakapo and Merlot samples run separately) and the 12X.v2 grapevine reference genome assembly (Canaguier et al., 2017).

The SVs generated by sniffles and pbsv were first filtered to remove variants that did not pass the filters applied by the two variant callers. Sniffles performs filtering intrinsically by only keeping SVs 35 bp or longer in length, with a minimum number of supporting reads equal to or above 10% of the sequencing depth (2-3 reads for our samples). Sniffles also applies a “GT” tag for variants where the quality of the genotype is low, and SVs with this tag were filtered out. Pbsv performs filtering intrinsically by only keeping SVs 20 bp minimum in length, with at least 3 supporting reads across all samples and within samples, 1 supporting read per strand total across samples, and supporting reads above 20% of reads mapping to that site per sample. Pbsv also applies filters for variants near gaps in the reference genome or contig ends and for duplication variants with reads that do not fully span the region, which were all filtered out. For total structural variant counts by sample, the filtered VCF files from sniffles and pbsv were then merged

using SURVIVOR v1.0.7 “merge” (Jeffares et al., 2017) to merge SVs identified by both programs that were greater than 30 bp long and within 300 bp of one another. To identify variants with genotypes specific to the WB samples and not present in WT, SnpSift v2017-11-24 (Cingolani et al., 2012) was used with the filtered VCF files to extract out variants either a) only found in the WB sample (homozygous or heterozygous) or b) homozygous in the WB sample and heterozygous in the WT sample. The VCF files filtered both by quality and SnpSift from sniffles and pbsv were then merged using SURVIVOR v1.0.7 “merge” (Jeffares et al., 2017) as described previously. Only SVs that met those two criteria for merging were used for downstream analysis. The genes overlapping with the merged SVs were identified using bedtools v2.30.0 “intersect” (Quinlan & Hall, 2010) and the Genoscope 12X grapevine genome annotation (Jaillon et al., 2007) lifted to the 12X.v2 grapevine genome assembly (Canaguier et al., 2017) using Liftoff v1.6.2 (Shumate & Salzberg, 2021) with the *-copies* parameter.

Candidate gene analysis

To investigate a potential causal gene(s)/variant(s) for the WB budsport in grapevine, all genes with high impact SNPs/INDELs or SVs present in the WB samples and either a) absent in WT (described as “novel” from hereinafter) or b) heterozygous in WT but homozygous in WB, were investigated for gene function by looking into the functions of their closest *Arabidopsis thaliana* ortholog. Variants matching either genotype criteria are described as “genotypically distinct” from hereinafter. In order to understand the putative functions of the genes with SNPs, INDELs, and SVs in the WB samples, diamond v0.8.36 (Buchfink et al., 2015) was used to search for *Arabidopsis* orthologs to the putative causal genes using the Araport 11 *Arabidopsis* annotation (Cheng et al., 2017) with the following

parameters: *--max-target-seqs 1* and *--unal 0*. The list of *Arabidopsis* genes orthologous to WB candidate genes was loaded into RStudio, and the R/Bioconductor package *biomaRt* v2.54.1 (Durinck et al., 2009) was used to obtain gene descriptions from Ensembl Plants (Bolser et al., 2016). The *Arabidopsis* orthologs and the information about their function were then used to prioritize genes involved in developmental, hormone signaling, or other pathways that could potentially result in the WB phenotype. Variants of interest were verified first by looking at mapped reads for all samples in a genome browser to verify that the genetic variants were truly genotypically distinct to the WB sample. Then, polymerase chain reaction (PCR) was used to validate the variant in all samples. The amplified products were Sanger sequenced to verify that the variant called was accurate in both location and genotype.

RESULTS

WB shoot phenotypes

The WB bud sport arises spontaneously in many varieties of grapevine (Bettiga et al., 2013). We characterized two independent cases of WB that occurred at a commercial vineyard in Madera, CA. The first case is a WB mutant of a Merlot grapevine, observed as one arm (the eastern) on a vine in a commercial vineyard block. The adjacent western arm on the same plant is WT. This allowed a direct comparison of WB and WT tissues from the same plant. The second case characterized was in the Dakapo variety and is a WB vine that was identified as a whole plant mutation. As a result, no WT shoots were present on the Dakapo WB plant, so separate, unaffected Dakapo vines from the same propagation batch were used as the WT comparison. In both cases, the bud sport is characterized by vigorous vegetative growth with shortened internodes (Figure 2.1).



Figure 2.1. Photos of wild type and Witch's Broom shoots from a commercial vineyard. **(A)** Photos of Merlot WB and WT on one grapevine plant. WB shoots are the light green shoots in the center of the image, while WT shoots are the darker green shoots on either side of the WB shoots. Merlot WB shoots display prolific growth in comparison to their WT counterparts. **(B)** An up-close photo of Merlot WB shoot, with light green leaves and shortened internodes. **(C)** A side-by-side photo of Dakapo WT (left) and Dakapo WB (right) shoots from different plants. Dakapo WB shoots have shortened internodes and more prolific foliage than their WT counterparts. **(D)** An up-close photo of a Dakapo WB shoot, showing a significantly shortened internode.

Both cases of WB also appear to have issues rooting, with Dakapo WB cuttings rooting less frequently than Dakapo WT cuttings, and Merlot WB cuttings being entirely unable to root

(P. Cousins, unpublished observations). Merlot WB shoots have light green leaves strikingly distinct from WT shoots (Figure 2.1A), while Dakapo WB leaves are similar in color to WT shoots (Figure 2.1C).

Comparison of multiple shoot traits between the WT and WB plants revealed large differences in phenotypes between the two. Both Dakapo and Merlot WB shoots have internodes significantly shorter than their WT counterparts ($t = -21.86$, $df = 230.76$, $P < 0.001$ for Dakapo; $t = -2.93$, $df = 317.25$, $P = 0.003$ for Merlot) (Figure 2.2A).

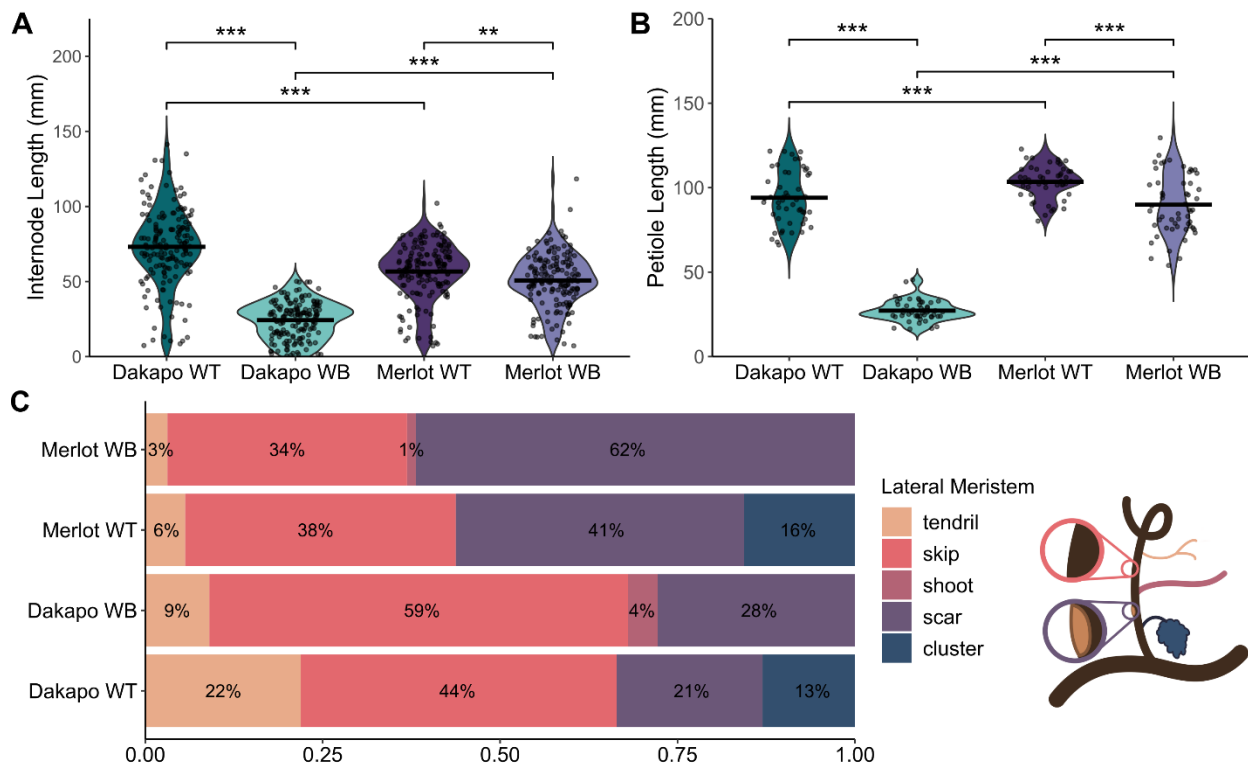


Figure 2.2. Differences in shoot phenotypes between wild type and Witch's Broom samples in Dakapo and Merlot varieties of grapevine. **(A)** A comparison of average internode length and **(B)** petiole length between sample types, collected from 10 shoots each. Mean values were represented by a black line for each sample. Dakapo WB and Merlot WB both have significantly smaller internodes ($P < 0.001^{***}$ and $P < 0.01^{**}$, respectively) and petioles ($P < 0.001^{***}$ for both cases) in comparison to WT plants of the same variety. The WT samples of the two varieties differ as well, with Dakapo WT having longer internodes but shorter petioles than Merlot WT ($P < 0.001^{***}$ for both). Dakapo WB also has significantly smaller internodes and petioles compared to Merlot WB ($P < 0.001^{***}$ for both). **(C)** The percentage of nodes with specific lateral meristem outcomes, collected from 144-160

Figure 2.2 (cont'd)

lateral meristems for each sample. The diagram to the right of the legend shows each of the lateral meristem outcomes both in the color and order they appear on the legend.

The petioles were also smaller in WB plants than WT ($t = -27.72$, $df = 32.91$, $P < 0.001$ for Dakapo; $t = -5.01$, $df = 87.44$, $P < 0.001$ for Merlot) (Figure 2.2B). Our phenotyping also revealed that the Dakapo WB phenotype seems to be more severe than the Merlot WB phenotype. The Dakapo WB internodes are significantly shorter than those of Merlot WB ($t = -15.54$, $df = 281.57$, $P < 0.001$), despite Dakapo WT internodes being longer than Merlot WT internodes ($t = 6.58$, $df = 294.07$, $P < 0.001$) (Figure 2.2A). In addition, the Dakapo WB petioles are also significantly shorter than their Merlot WB counterparts ($t = -25.19$, $df = 69.41$, $P < 0.001$) (Figure 2.2B).

Initial measurements of leaf width and length demonstrated that Dakapo and Merlot WB leaves are significantly shorter and narrower than their WT counterparts when compared at the same node ($P < 0.05$ for width and length at node 4 for Dakapo; $P < 0.05$ for both width and length, for nodes 5-9 for both Dakapo and Merlot cases) (Figure S2.2.1). While initial data collected in 2021 showed that the Dakapo and Merlot WB leaves were typically shorter and narrower than their WT counterparts (Figure S2.2.1), the actual change in leaf area and leaf shape was unknown. Leaves collected and landmarked from all samples in 2022 demonstrated that WB leaf areas were significantly smaller overall than their WT counterparts ($t = 23.49$, $df = 76.98$, $P < 0.001$ for Dakapo; $t = 22.41$, $df = 70.42$, $P < 0.001$ for Merlot) (Figure 2.3A-E).

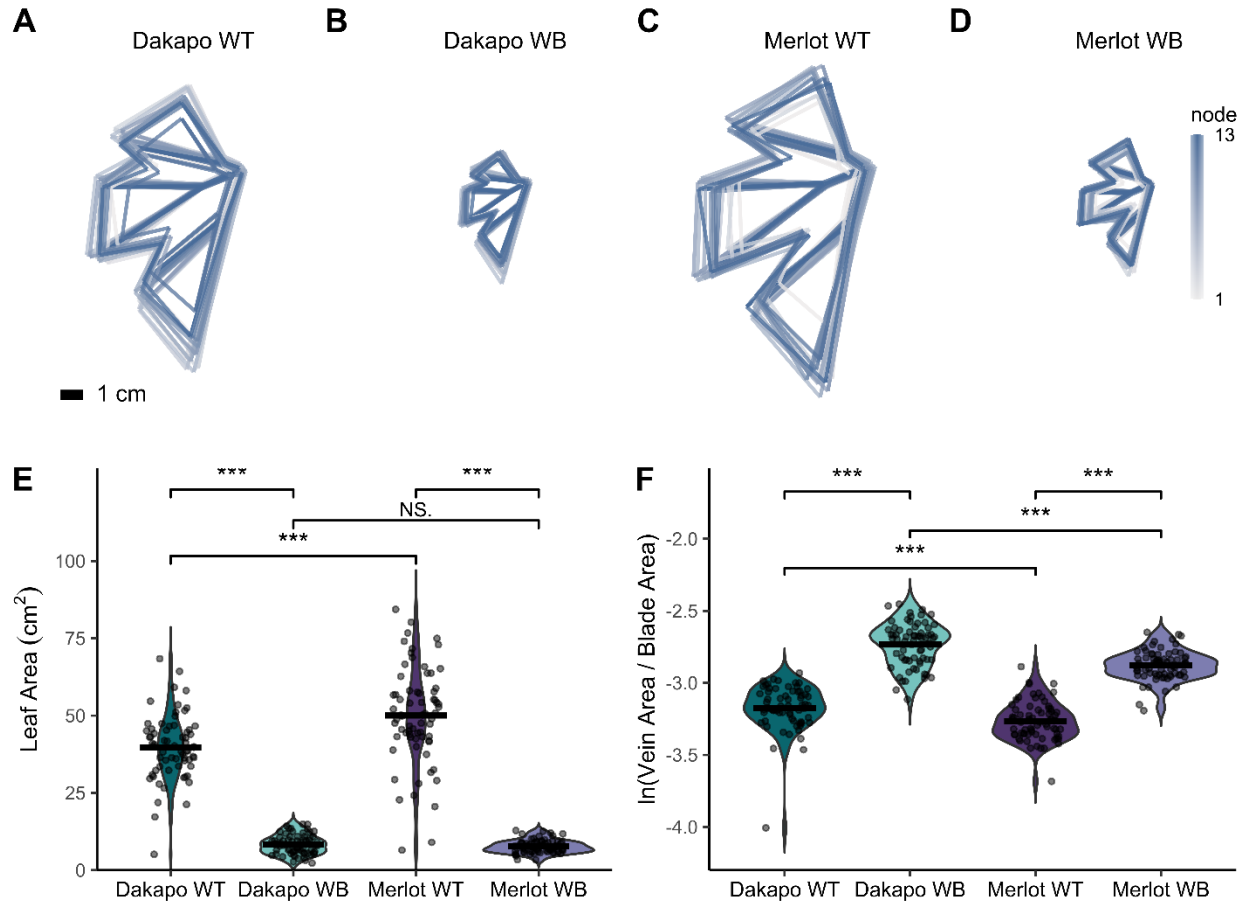


Figure 2.3. Comparing leaf area and the natural log of the ratio of vein-to-blade area between wild type and Witch’s Broom samples in Dakapo and Merlot varieties of grapevine. **(A)** Dakapo WT, **(B)** Dakapo WB, **(C)** Merlot WT, and **(D)** Merlot WB composite leaves generated using leaf landmarks to model leaf shapes for leaves collected across 13 nodes. Composite leaves are colored based on node, from gray (node 1 from the shoot tip) to dark blue (node 13). All samples are to the same scale, and a 1 cm scale bar is provided in the bottom left corner of **(A)**. **(E)** A comparison of leaf area (cm²), as calculated using the shoelace algorithm originally described by Meister (1769) and used in Chitwood et al. (2020) to calculate leaf area in grapevine, with leaf landmark data. Mean leaf area (cm²) is represented by a black line for each sample. Dakapo WB and Merlot WB both have significantly smaller leaves ($P < 0.001^{***}$ for both cases) in comparison to WT plants of the same variety. Merlot WT leaves were larger than Dakapo WT leaves ($P < 0.001^{***}$), however leaf area did not differ between the two WB cases ($P = 0.16$). **(F)** A comparison of the natural log of the ratio of vein-to-blade area, an allometric indicator of leaf size that is typically more sensitive to leaf size changes than leaf area alone. Mean ln (vein-to-blade ratio) is represented by a black line for each sample. Dakapo WB and Merlot WB both have significantly higher vein-to-blade ratios ($P < 0.001^{***}$ for both cases) in comparison to WT plants of the same variety. Dakapo WT leaves have a higher vein-to-blade ratio than Merlot WT leaves ($P < 0.001^{***}$). Dakapo WB leaves have a higher vein-to-blade ratio than Merlot WB leaves ($P < 0.001^{***}$) as well.

To further understand how WB leaf development may differ from typical grapevine leaf development, we calculated the allometric ratio of vein area to blade area. As leaves expand, the blades of leaves expand at a greater rate than the veins (Chitwood et al., 2016). As a result, larger leaves typically have lower vein-to-blade ratios. In addition, the ratio of vein-to-blade area is typically more responsive to subtle changes in leaf shape and development than area alone (Chitwood et al., 2021). As expected, given their small leaves, both Dakapo WB and Merlot WB have significantly higher vein-to-blade ratios in comparison to WT plants of the same variety ($t = -16.67$, $df = 133.14$, $P < 0.001$ for Dakapo; $t = -19.08$, $df = 127.55$, $P < 0.001$ for Merlot) (Figure 2.3F). Dakapo WB leaves also have a higher vein-to-blade ratio than Merlot WB leaves ($t = 6.53$, $df = 120.44$, $P < 0.001$) (Figure 2.3F). This is likely due to very subtle differences in leaf development between the two WB samples that are not captured by comparing leaf area alone, such as differences in vasculature development between the two.

Analyzing the leaf landmark data utilizing a Procrustes analysis and a principal components analysis (PCA) revealed that WB leaves also differ in their shape when compared to their WT counterparts ($H = 13.26$, $P < 0.001$ for Dakapo; $H = 14.07$, $P < 0.001$ for Merlot) (Figure 2.4).

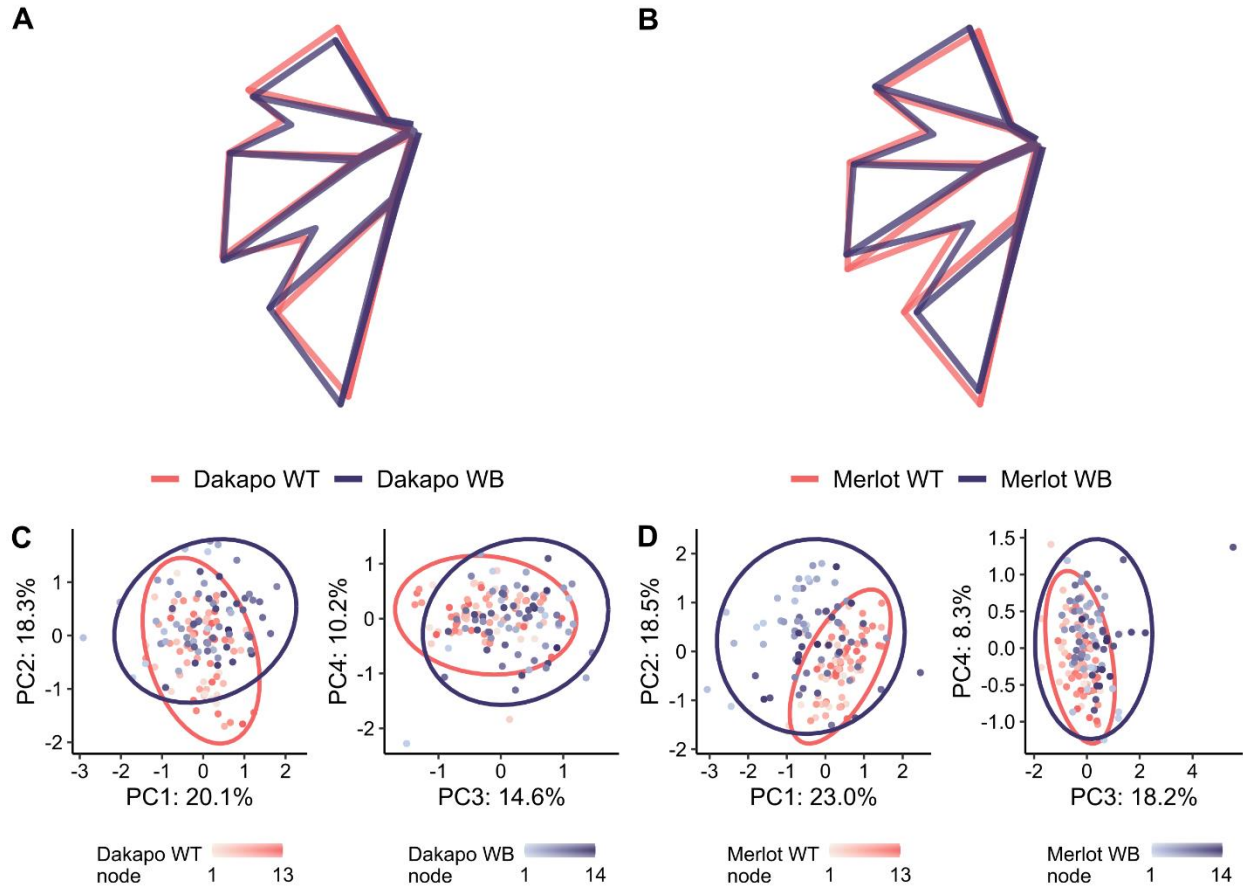


Figure 2.4. Mean leaf shapes rotated and scaled identically for **(A)** Dakapo WT and Dakapo WB, as well as for **(B)** Merlot WT and Merlot WB. **(C and D)** Principal component analysis (PCA) of all leaf shapes, with WT colored in salmon and WB colored in purple, for **(C)** Dakapo and **(D)** Merlot. The node position of the leaves is also shown by shade, with the lightest shade being node 1 (from the shoot tip) and the darkest shade being node 13-14, depending on the sample.

Eigenleaves from the PCA comparing leaf shape between scaled WT and WB leaves (Figure S2.2 and Figure S2.3) revealed the shape features that each PC reflected. The leaf shape variance between WT and WB in both Merlot and Dakapo appears to be due to similar phenotypic changes in the WB leaves. For both varieties, PC2 reflects variance in the depth of the distal sinus, which is deeper in WB samples. WB leaves in both varieties also seem to have a wider petiolar sinus, which is reflected by PC3 in Dakapo and PC4 in Merlot. Additionally, WB plants in both varieties appear to have narrower upper lateral lobes,

which is explained by PC4 in Dakapo and PC1 in Merlot (Figure 2.4). Despite these similarities in how WB leaves differ from WT in the two varieties, WB also appears to impact leaf shape somewhat differently in the two varieties. Dakapo WB leaves appear to have narrower distal sinuses than their WT counterparts, as described by PC1 (Figure 2.4 A, C). Meanwhile, Merlot WB leaves appear to have shorter midveins relative to the rest of leaf features, as explained by PC3 (Figure 2.4 B, D). These two features appear to be specific to WB bud sports of the particular variety.

We also characterized the fates of lateral meristems of the WB bud sports to understand the developmental outcomes of the WB buds. Lateral meristem fates were characterized by the organ or structure that had developed at the nodes, which were either: a) tendrils, b) skips (nodes where no lateral meristem was present), c) shoots, d) scars (nodes where a meristem had formed, but no structure was present when phenotyped), or e) clusters/fruit. These observations revealed that no clusters were developing in the WB shoots. In addition, the WB shoots developed new lateral shoots at 1-4% of nodes, while their WT counterparts did not develop these new lateral shoots at any nodes (Figure 2.2C). New grapevine shoots arise from axillary buds, and lateral shoots typically do not develop. It is possible that the incidence of lateral shoots on the WB bud sports may be due to the mutation directly. Both the presence of the lateral shoots and absence of clusters support that the WB bud sports seem to involve a shift towards vegetative growth and away from reproductive growth. Many of the WB lateral meristems failed to develop properly, with 87% of Dakapo WB buds and 96% of Merlot WB buds failing to develop into tendrils, clusters, or shoots, compared to 65% and 79% in their respective WT counterparts. The higher incidence of skips in Dakapo WB (59%), in

comparison to Dakapo WT (44%), contributes directly to the lack of tendrils and clusters observed. Dakapo WB having more skips present is somewhat unexpected, as grapevines are expected to generally show a phyllotaxy of two successive nodes with a lateral meristem followed by one node without. As a result, we generally expect to see $\frac{1}{3}$ of the nodes studied to be skips. It is possible that the WB mutation in Dakapo causes an unusual phyllotaxy and thus more skips to be present. However, the Merlot WB shoots have about the expected number of skips present (34%). While the characterization of lateral meristem fate demonstrated dominance of vegetative growth in both instances of WB, it also revealed that they may have distinct issues when it comes to lateral meristem development.

Organization and development of WB buds

To investigate the developmental origin and timing of the defects seen across the WB shoots, particularly in lateral meristem fates, dormant winter buds were imaged to identify changes in bud organization. To do so, we imaged dissected grapevine buds with a dissecting scope and whole buds with CT scans. Grapevine dormant winter buds are typically composed of three bud primordia, characterized as primary, secondary, and tertiary, from most developed to youngest respectively. The bud primordia typically house leaf, tendril, and inflorescence primordia (Gerrath et al., 2015). The WT buds for both Dakapo and Merlot varieties had nearly identical organization and structures. The buds and primordia were each at a 45° angle from the stem. All WT buds had three bud primordia in each of the buds as expected. CT scans showed that all WT buds had inflorescence primordia present, with 80% of WT buds having two or more inflorescence primordia present in their primary bud primordia alone. None of the WT buds appeared to have any

organizational defects in the buds, with all primordia appearing to be healthy and properly arranged (Figure 2.5 A, C, Figure S2.4 A, C, and Figure S2.5 A, C).

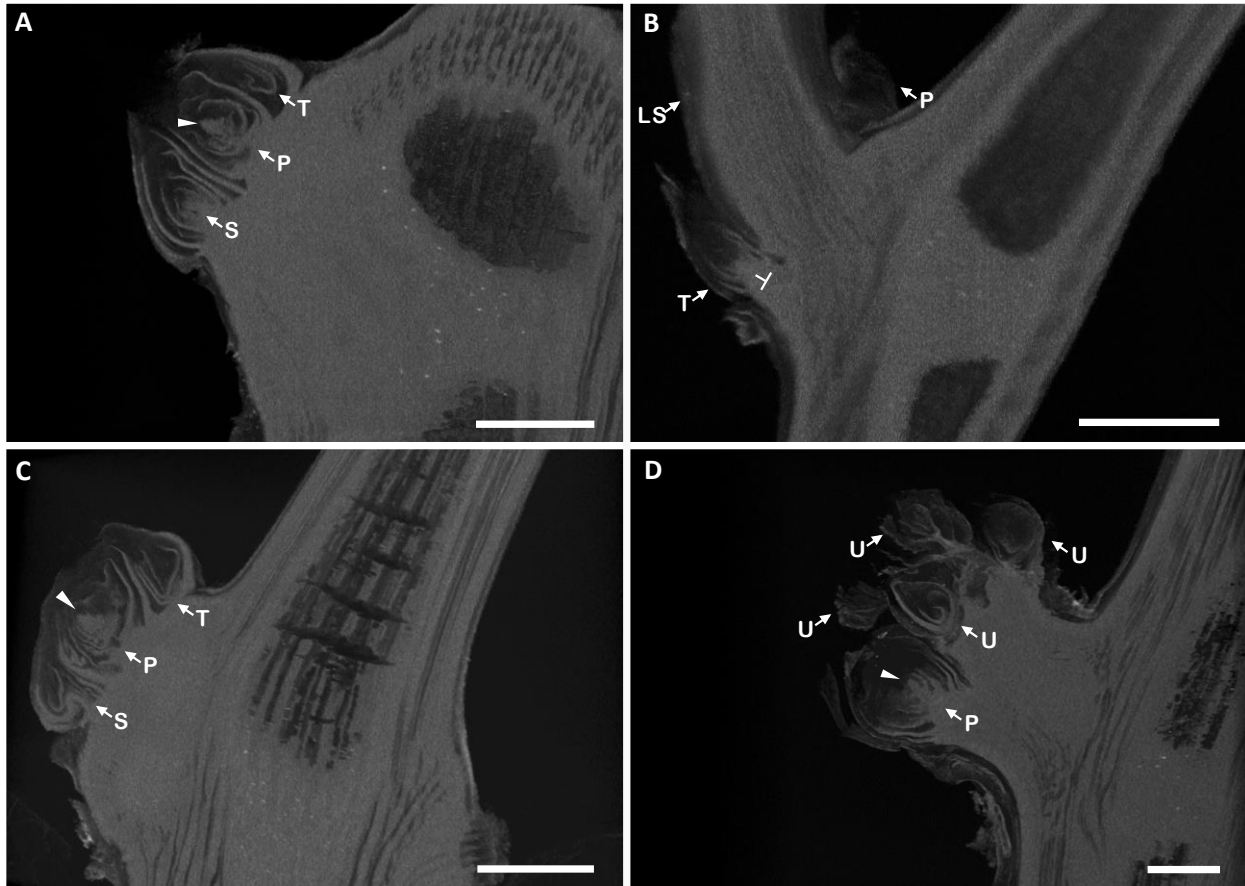


Figure 2.5. Representative CT scans of buds from **(A)** Dakapo WT, **(B)** Dakapo WB, **(C)** Merlot WT, and **(D)** Merlot WB samples. Primary (P), secondary (S), and tertiary (T) bud primordia are indicated if present in the image. The inflorescence primordia are indicated by the solid triangle in the **(A)** Dakapo WT, **(C)** Merlot WT, and **(D)** Merlot WB samples. Only one inflorescence primordium is present in the images, although multiple were seen for both WT samples. The Merlot WB sample shown **(D)** is the only Merlot WB sample scanned with a potential inflorescence primordium present, although the inflorescence primordium seen appears to be deformed due to the edges being smoother than those seen in WT samples **(A and C)**. The lateral shoot stem (LS) is indicated in **(B)** Dakapo WB. The bud primordia in the Merlot WB sample shown are challenging to accurately label, aside from the more-developed primary primordia (P), so we have labeled the additional bud primordia as uncharacterized primordia (U) in **(D)** Merlot WB, which are indicated as well. Scale bar = 1 mm.

In contrast, the WB buds contained multiple organizational defects. Upon examination, about half of the Dakapo WB buds had an initiated lateral shoot stem

extended out of them, about 1 cm long (Figure S2.4B and Figure S2.6). CT scans revealed that this stem appears to be vascular tissue pushing through the bud, disrupting the typical organization (Figure 2.5B and Figure S2.5B). The vascular tissue expanding through the buds sometimes contained a shoot apex on the tip, suggesting that these shoots can produce leaves and other lateral organs. Half of the buds contained an additional change in overall architecture, with the tertiary primordia being perpendicular to the stem (Figure 2.5B). Many of the primordia present in the Dakapo WB buds appeared to be smaller than those in the other samples. Notably, three of four Dakapo WB buds had only one inflorescence primordia present, but the inflorescence primordia appeared deformed in two of the buds scanned.

The Merlot WB buds had drastically different organization from WT buds as well, with the buds containing between 4-8 bud primordia (Figure 2.5D, Figure S2.4D, and Figure S2.5D), in contrast to the 3 consistently found in wild type samples (Figure 2.5A, C, Figure S4 A, C, and Figure S2.5 A, C). Similarly to the Dakapo WB samples, two out of five of the Merlot WB buds had tertiary primordia nearly perpendicular to the stem. In addition, all but one of the Merlot WB buds had no inflorescence primordia. The inflorescence primordium potentially present in the single sample was difficult to confidently identify as such however since it lacks the lobes typically seen in developing inflorescence primordia (Figure 2.5D). As a result, even if this structure is truly an inflorescence primordium, it is extremely deformed. However, none of the Merlot WB buds displayed the vascular tissue expansion seen in the Dakapo WB samples.

Overall, the Dakapo and Merlot WB buds contained phenotypes vastly different from WT and even one another. The WB samples displayed extensive defects in bud organization

and the quantity of inflorescences produced indicating that the WB defects manifested early in bud development. This investigation into the buds of the WB bud sports provided insight into the defects we identified across the shoots of the bud sports. Not only are the shoots failing to develop properly, but the defects are pervasive in the buds and potentially their internal structures, as well.

Genetic variation in WB Bud Sports

To investigate the genetic basis of the WB bud sport, we sequenced DNA from both Dakapo and Merlot WB and WT using both Illumina 150 bp paired-end sequencing and Oxford Nanopore long-read sequencing. After trimming and filtering, the sequencing coverage of the Illumina reads was between 49-57X and the sequencing coverage of the Oxford Nanopore reads was between 18-31X (see Table S2.1 for full sequencing statistics). The read length N50 for the trimmed Oxford Nanopore reads was between 12,890-14,486 bp for the samples. High quality reads were used for mapping to the reference genome and calling variants in each of the samples. For all samples, over 98.2% of Illumina reads and over 99.9% of the Oxford Nanopore reads mapped to the grapevine 12X.v2 reference genome (Canaguier et al., 2017).

SNPs and INDELS were called against the 12X.v2 grapevine reference genome (Canaguier et al., 2017) using Illumina sequencing data. Each sample had between 7.9-8.2 million SNPs/INDELS and high heterozygosity (67.96-71.22%). Most SNPs and INDELS were present in both WT and WB samples of the same variety (94.81-94.97%), however between 409,588-418,818 SNPs/INDELS were entirely novel when compared within-variety. A majority of SNPs and INDELS were either intergenic or not expected to have an impact on gene function (Table S2.2). Of the SNPs and INDELS called in the WB samples,

6,296-6,450 were predicted to have high impact on gene function. Between 597-613 genes impacted by SNPs or INDELS predicted to have a high impact were genotypically distinct in WB bud sports, and these genes were kept as possible causal candidates for the bud sport (Table 2.1).

	Dakapo		Merlot	
	WT	WB	WT	WB
SNPs and INDELS				
Total SNPs/INDELS	7,912,797	7,925,441	8,148,571	8,163,074
Genotypically distinct SNPs/ INDELS ^a	497,531	493,394	495,782	503,973
Novel SNPs/INDELS ^b	410,608	411,425	409,588	418,818
High impact SNPs/INDELS	14,013	13,962	14,032	14,027
Genes impacted by high impact SNPs/INDELS ^c	6,310	6,296	6,450	6,445
Genes impacted by genotypically distinct high impact SNPs/INDELS	611	613	591	597
SVs				
Total SVs	52,119	53,089	54,775	53,912
Genotypically distinct SVs ^a	578	635	691	540
Novel SVs ^b	157	223	224	102
Genes impacted by SVs	15,044	15,134	15,706	15,596
Genes impacted by genotypically distinct SVs	135	136	150	134

Table 2.1. Genetic variants identified in samples when called against the 12X.v2 grapevine reference genome assembly, including SNPs/INDELS and SVs. Novel and genotypically distinct variants were identified by comparing variants intra-variety.

^a Variants genotypically distinct from the sample of the same genotype include entirely novel SNPs, as well as SNPs that have different genotypes when compared intra-variety.

Table 2.1 (cont'd)

^b Novel variants are variants completely absent in the sample of the same variety.

^c High impact SNPs include frameshift deletion or insertion, stop gain/loss, and splicing.

Structural variants were called against the 12X.v2 grapevine reference genome (Canaguier et al., 2017) using long-read sequencing data. Each sample had between 52-55,000 SVs. Deletions were the most common type of SV and accounted for over half of the SVs called. Insertions were the next most common type of SV and accounted for about 47% of total SVs. Inversions, transversions, and duplications were extremely rare, and collectively only accounted for between 1.27-1.66% of all SVs called (Table S2.3). Entirely novel SVs (when compared within variety) were rare as well, with only between 102-224 identified within the samples. Only 635 and 540 SVs were genotypically distinct in Dakapo WB and Merlot WB, respectively. About 15,000 genes had SVs within them for each sample individually. Of the genes containing SVs, 136 and 134 were impacted by genotypically distinct SVs for Dakapo WB and Merlot WB, respectively (Table 2.1).

We identified 974 and 968 high impact SNPs, INDELs, and SVs genotypically distinct in Dakapo WB and Merlot WB respectively that are all potential candidates for the WB bud sport in their respective genotype (Table S2.4). We looked at the gene function for 577 and 561 genes only impacted by high impact mutations in WB samples in Dakapo WB and Merlot WB, respectively. The two WB samples shared 164 genes impacted by high impact variants. All genes in common between the two WB samples were weak candidates with either gene functions unrelated to the WB phenotype or were unsupported by the genome browser and/or PCR validation. As a result, we investigated the potential biological impact and validity of the 974 high impact variants in Dakapo WB and the 968 high impact variants in Merlot WB, separately. To narrow down this list of potential candidates for both

cases of WB, we looked at the function of the genes impacted by these variants and further investigated genes involved in development, growth, or hormone signaling. Most genes impacted by high impact variants were involved in unrelated processes or were of unknown function, however 14 variants in Dakapo WB and 23 variants in Merlot WB were identified as having a high impact on genes involved in development, growth, or hormone signaling. We looked at WT and WB reads mapping at the loci of these variants as an initial pass to ensure that they were accurately genotyped, and only one high impact variant for both Dakapo WB and Merlot WB seemed to truly be genetically distinct to the WB case of interest. PCR validation of these two variants demonstrated that the Dakapo WB variant of interest was present in a heterozygous state in both Dakapo WT and WB and therefore likely not a strong genetic candidate for the Dakapo WB bud sport. However, PCR validation and Sanger sequencing demonstrated that the Merlot WB variant was present in Merlot WB only and was entirely absent in Merlot WT (see Supplemental Methods; Figure S2.7 and Figure S2.8). The PCR-validated variant in Merlot WB is a 3.6 kbp insertion in the intron of GSVIVG01008260001 (VCost.v3 annotation gene ID: Vitvi17g00344 (Canaguier et al., 2017), CRIBI V1 annotation gene ID: VIT_17s0000g03960 (Forcato, 2010)), an ortholog of *Arabidopsis* STOMATAL CYTOKINESIS-DEFECTIVE 1 (Figure S2.9). This variant is heterozygous in Merlot WB and completely absent in Merlot WT samples. A BLASTN search against the 12X.v2 grapevine reference genome assembly (Canaguier et al., 2017) showed that this sequence showed significant similarity to 2,736 sequences within the genome, spread across all 19 chromosomes. The 3.6 kbp insert sequence also contains seven transposable element sequences that account for 98.5% of the sequence, including four *Gypsy* long terminal repeat (LTR) retrotransposons, two uncharacterized LTR

retrotransposons, and one *Mutator* terminal inverted repeats (TIR) retrotransposon (see Supplementary Methods). Of these, one *Gypsy* LTR and one uncharacterized LTR appear twice in the insert sequence adjacent to one-another. We propose that this genetic variant may be the causal mutation for the WB bud sport in the Merlot WB case investigated.

DISCUSSION

Developmental defects in the WB bud sport

Our phenotypic measurements of the Dakapo and Merlot WB bud sports revealed new aspects of the WB phenotype that had previously been unknown. The most striking finding being how different the two instances of WB studied are from one another, with the Dakapo WB shoots having much smaller features in comparison to the Merlot WB (Figure 2.2 A, B and Figure 2.3E). Analysis of lateral meristem fate, leaf shape, and dormant buds further enforced how distinct the two instances of WB are (Figure 2.2C, Figure 2.4, and Figure 2.5 B, D). However, both WB cases were significantly smaller than their wild type counterparts in every trait measured. The WB phenotype also seems to include development defects that have not been previously identified, such as subtle changes in leaf shape in both varieties (Figure 2.4). The phenotypic measurements across the shoots of the WB bud sports show that not only are they smaller than their WT counterparts, but they also have defects in regulating overall shoot and leaf development. Our leaf size and shape data both seem to support that the WB leaves specifically seem to have very distinct developmental trajectories, with a) WB leaf areas not following the negative quadratic trend we expect to see as leaves age (Figure S2.2.10) and b) WB leaves across the shoots having juvenile characteristics, such as deeper sinuses (Bryson et al., 2020) (Figure 2.4). Identifying lateral meristem fates and analyzing internal bud morphologies also clarified

developmental defects within the two instances of WB. These results suggested that the WB phenotype may be largely influenced by issues early on in meristem development, leading to a diverse array of developmental defects.

Investigating the genetic basis of the WB bud sport

Given the phenotypic differences between Dakapo WB and Merlot WB, it is possible that there are multiple genetic means of causing what is colloquially termed a “Witch’s Broom bud sport”. Mutational Witch’s Brooms are poorly described in angiosperms, although they are described from conifers (Zhuk et al., 2015), leading to few likely candidate genes in which mutations may drive the WB phenotype. Due to the large differences in phenotype between the two varieties, as well as none of the shared genes impacted by variants being good candidates for the bud sport, we propose that two different genetic variants cause the WB bud sport in the Dakapo and Merlot cases we investigated.

In Merlot, we identified a putative candidate gene for WB: GSVIVG01008260001. It is highly expressed in most tissue types, including buds, leaves, inflorescences, and roots (Fasoli et al., 2012), making it a promising candidate for a mutation with pleiotropic effects. GSVIVG01008260001 is orthologous to the gene AT1G49040 in *Arabidopsis*, which encodes STOMATAL CYTOKINESIS-DEFECTIVE 1 (AtSCD1). AtSCD1 is involved in the cytokinesis of guard mother cells and other leaf epidermal cells. However, AtSCD1 also appears to play a role in overall plant growth and development. In *Arabidopsis*, *scd1* mutants are smaller than WT plants, have reduced leaf expansion, and defects in flower morphology. The floral buds in *scd1* are smaller than WT due to early abortion in development and are highly branched as well (Falbel et al., 2003). The phenotype of the *scd1* floral buds is similar to the

WB phenotype of Merlot WB buds, which are smaller than WT and also highly branching (Figure 2.5D). The dwarfness and small leaves of *scd1* also match what we see in Merlot WB shoots. The abundant similarities between *scd1* mutants in *Arabidopsis* and the Merlot WB bud sport make GSVIVG01008260001 a strong candidate for one casual gene of the WB bud sport. The insert sequence within GSVIVG01008260001 being almost entirely annotated as TE sequence also provides a clear possible explanation for how this bud sport could arise spontaneously since the TE sequence within the insertion may have led the insertion within this gene. Additionally, no other genes overlapping with SNPs or SVs unique to Merlot WB appear to be strong candidates. Most other genes identified as uniquely impacted by variants in Merlot WB do not appear to be involved in plant growth and development and/or are not truly genetically distinct in Merlot WB. Between the genetic evidence in the Merlot WB grapevine plants and phenotypic similarity to the *Arabidopsis* ortholog (Falbel et al., 2003), we propose GSVIVG01008260001 as a candidate causal gene for the WB bud sport in grapevine.

While we were able to identify a strong candidate in Merlot WB, no strong candidates were identified in Dakapo WB. There are a few complicating factors that contributed to the difficulty of identifying a causal WB candidate in our Dakapo WB sample. For one, grapevine is highly heterozygous, which made it challenging to both accurately call and genotype SNPs and SVs within our samples. In addition, genetic chimeric variability, in which one cell layer has distinct genetic variants in comparison to the other cell layer, has repeatedly been identified in grapevine (Franks et al., 2002; Riaz et al., 2002). The phenotypic manifestation of a chimeric genetic variant depends on the cell layer(s) it is present within (Frank & Chitwood, 2016). As a result, it is possible that the WB causal

variant could be present in both WT and WB sequencing data, but present in distinct cell layer(s) between WT and WB. If the WB causal variant is chimeric in nature, it may not have been identified through our sequencing and variant calling. Finally, it is also possible that the WB bud sport could be the result of an epiallele as well, as was found with the mantled somaclonal variant that arises frequently in oil palm (Ong-Abdullah et al., 2015).

Ultimately, genetic transformation is necessary to prove the causal gene(s) of the WB bud sport. However, it is likely that an inducible mutant will need to be used to circumvent possible lethality due to issues that the bud sports have with rooting. As a result, the natural instances of the WB bud sport studied here provide invaluable natural mutants for studying whole plant development in grapevines. It is possible that the WB bud sport provides insight into developmental defects and interactions between developmental processes that might otherwise be impossible to study due to the inability of the WB bud sports to properly root and produce seed. Studying other occurrences of WB in the future will provide more insight into grapevine development and clarify the extent of the phenotypic and genetic diversity of “Witch’s Broom bud sports”.

Somatic mutations in grapevine shoots and clones

Our paired sequencing of WT and WB tissue from two instances of WB in grapevine also provided insight into somatic mutations both between clones and within plants. All samples had relatively similar counts of sample-unique SNPs when compared within variety (Table 2.1). We found between 349,239-349,533 of clone-specific SNPs in Dakapo (Table S2.5). This is somewhat lower than what other studies performing similar a 1:1 comparison of clones found, which ranged from ~600k-3.3 million SNPs (Urrea et al., 2023; Vondras et al., 2019). This difference is likely due to differences in methods employed, as

these studies compared clone sequencing data to reference genomes of the same variety, while we performed joint variant calling and genotyping against the grapevine reference genome. However, our count is higher than clone-specific SNPs that have been identified when comparing a larger populations of grapevine clones, which ranges from 200 to 30.7k SNPs (Gambino et al., 2017; Urra et al., 2023; Vondras et al., 2019). Studies looking at intra-clonal variation in grapevine have all had different aims and thus different variant calling and filtering approaches, which has likely led to this large range in the number of SNPs identified both between clones and within varieties. Our data also provided insight into intra-organism mutations in grapevine, which have been relatively understudied compared to intra-clonal mutations. Our dataset revealed that the number of somatic mutations within one grapevine plant, when comparing distinct shoots (Merlot WT shoots and Merlot WB shoots), is similar to those found between grapevine clones (Table 2.1), with between 351,018-356,754 shoot-specific SNPs being identified in Merlot (Table S2.5). The counts of shoot-specific SNPs in Merlot is higher than the number of intra-organisms SNPs that have been identified in grapevine (3.2-3.7k) (Sichel et al., 2023) and other plant systems (4.9k SNPs in *Zostera marina* and 44-152k SNPs in *Populus trichocarpa*) (Hofmeister et al., 2020; Yu et al., 2020). This is likely in large part due to the differences in methods used between our study and previous studies due to the differences in the aims of the studies. Given that the main goal of this study was to identify putative causal variants of WB, we did not apply stringent filtering that these previous studies have applied (Hofmeister et al., 2020; Sichel et al., 2023; Yu et al., 2020).

Our long-read sequencing also provided insight into SV somatic mutations, which are relatively understudied in comparison to SNP somatic mutations, especially at the

intra-organism level. We identified between 157-223 clonal-specific SVs in Dakapo, and between 102-224 shoot-specific SVs in Merlot. These findings align with our SNP data and support that the number of intra-organism somatic mutations in grapevine is similar to the number of inter-clone somatic mutations. The actual number of clonal- and shoot-specific SNPs and SVs is likely much lower than what was reported due to sequencing errors, alignment errors, etc. Regardless, these data provide insight into the accumulation of mutations within grapevine and supports the notion that grapevine clonal genetic diversity begins through novel somatic mutation accumulations on grapevine shoots, which are then clonally propagated.

CONCLUSIONS

The WB bud sport provides a natural mutant in which to study developmental defects that might otherwise be impossible to study. Grapevine development is vastly different from that in *Arabidopsis*, and understanding this process and the genetic pathways involved will be invaluable in not only other perennial crop systems, but also in understanding liana development. However, studying the genes involved in grapevine development is difficult due to both traditional breeding and genetic transformation being relatively challenging and time consuming (Campos et al., 2021). Investigating the phenotypic defects and potential genetic basis of the WB bud sport has provided insight in grapevine development from buds to shoots. Future work in WB plants, especially with instances of the bud sport in new varieties and genetic backgrounds, will help deepen our understanding of development in grapevine, as well as other lianas and perennial crops.

ACKNOWLEDGEMENTS

The authors thank Ileana Katzman, Erin Logan, and Brianna Wieferrich for collecting

and preparing grapevine herbarium specimens in 2022. We also thank the Genomics Core at Michigan State University and the Institute for Cyber-Enabled Research at Michigan State University for their services. This work is supported by Michigan State University, the USDA National Institute of Food and Agriculture MICL02572, and the NSF Plant Genome Research Program awards IOS-2310355, IOS-2310356, and IOS-2310357. EJR was supported by the University Distinguished Fellowship at Michigan State University.

REFERENCES

- Abramoff, Magalhães, P. J., & Ram, S. J. (2004). Image processing with ImageJ. *Biophotonics International*, 11(7), 36–42.
- Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>. Accessed 18 October 2023.
- Bettiga, L. J., Smith, R. J., Peacock, W. L., Hembree, K. J., Weber, E. A., & Verdegaal, P. S. (2013). Abiotic Disorders and Injuries of Grapevine. In L. J. Bettiga (Ed.), *Grape Pest Management Third Edition* (pp. 29–46). University of California Agriculture and Natural Resources.
- Bolser, D., Staines, D. M., Pritchard, E., & Kersey, P. (2016). Ensembl Plants: Integrating Tools for Visualizing, Mining, and Analyzing Plant Genomics Data. *Methods in Molecular Biology*, 1374, 115–140.
- Broad Institute. (2017). Picard Toolkit. Broad Institute. <https://broadinstitute.github.io/picard/>. Accessed 20 September 2023.
- Bryson, A. E., Wilson Brown, M., Mullins, J., Dong, W., Bahmani, K., Bornowski, N., Chiu, C., Engelgau, P., Gettings, B., Gomezcano, F., Gregory, L. M., Haber, A. C., Hoh, D., Jennings, E. E., Ji, Z., Kaur, P., Kenchanmane Raju, S. K., Long, Y., Lotreck, S. G., ... Chitwood, D. H. (2020). Composite modeling of leaf shape along shoots discriminates *Vitis* species better than individual leaves. *Applications in Plant Sciences*, 8(12), e11404.
- Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 12(1), 59–60.
- Caetano-Anolles D. (2024). Hard-filtering germline short variants. GATK. <https://gatk.broadinstitute.org/hc/en-us/articles/360035890471-Hard-filtering-germline-short-variants>. Accessed 16 Feb 2024.
- Campitelli E. (2022). ggnewscale: Multiple Fill and Colour Scales in 'ggplot2'. R package version 0.4.8. <https://CRAN.R-project.org/package=ggnewscale>. Accessed 20 September 2023.
- Campos, G., Chialva, C., Miras, S., & Lijavetzky, D. (2021). New Technologies and Strategies for Grapevine Breeding Through Genetic Transformation. *Frontiers in Plant Science*, 12, 767522.
- Canaguier, A., Grimplet, J., Di Gaspero, G., Scalabrin, S., Duchêne, E., Choisne, N., Mohellibi, N., Guichard, C., Rombauts, S., Le Clainche, I., Bérard, A., Chauveau, A., Bounon, R., Rustenholz, C., Morgante, M., Le Paslier, M.-C., Brunel, D., & Adam-Blondon, A.-F. (2017). A new version of the grapevine reference genome assembly (12X.v2) and of

- its annotation (VCost.v3). *Genomics Data*, 14, 56–62.
- Carbonell-Bejerano, P., Royo, C., Torres-Pérez, R., Grimplet, J., Fernandez, L., Franco-Zorrilla, J. M., Lijavetzky, D., Baroja, E., Martínez, J., García-Escudero, E., Ibáñez, J., & Martínez-Zapater, J. M. (2017). Catastrophic Unbalanced Genome Rearrangements Cause Somatic Loss of Berry Color in Grapevine. *Plant Physiology*, 175(2), 786–801.
- Cheng, C., Jiao, C., Singer, S. D., Gao, M., Xu, X., Zhou, Y., Li, Z., Fei, Z., Wang, Y., & Wang, X. (2015). Gibberellin-induced changes in the transcriptome of grapevine (*Vitis labrusca* × *V. vinifera*) cv. Kyoho flowers. *BMC Genomics*, 16(1), 128.
- Cheng, C.-Y., Krishnakumar, V., Chan, A. P., Thibaud-Nissen, F., Schobel, S., & Town, C. D. (2017). Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *The Plant Journal: For Cell and Molecular Biology*, 89(4), 789–804.
- Chitwood, D. H., Mullins, J., Migicovsky, Z., Frank, M., VanBuren, R., & Londo, J. P. (2021). Vein-to-blade ratio is an allometric indicator of leaf size and plasticity. *American Journal of Botany*, 108(4), 571–579.
- Chitwood, D. H., Rundell, S. M., Li, D. Y., Woodford, Q. L., Yu, T. T., Lopez, J. R., Greenblatt, D., Kang, J., & Londo, J. P. (2016). Climate and Developmental Plasticity: Interannual Variability in Grapevine Leaf Morphology. *Plant Physiology*, 170(3), 1480–1491.
- Cingolani, P., Patel, V. M., Coon, M., Nguyen, T., Land, S. J., Ruden, D. M., & Lu, X. (2012). Using *Drosophila melanogaster* as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. *Frontiers in Genetics*, 3, 35.
- Constantin, A. E., & Patil, I. (2021). ggsignif: R package for displaying significance brackets for “ggplot2.” *PsyArxiv*.
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2).
- De Coster, W., D’Hert, S., Schultz, D. T., Cruts, M., & Van Broeckhoven, C. (2018). NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics*, 34(15), 2666–2669.
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernysky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43(5), 491–498.
- Dryden, I. L., & Mardia, K. V. (2016). *Statistical Shape Analysis: With Applications in R*. John Wiley & Sons.

- Durinck, S., Spellman, P. T., Birney, E., & Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*, 4(8), 1184–1191.
- Falbel, T. G., Koch, L. M., Nadeau, J. A., Segui-Simarro, J. M., Sack, F. D., & Bednarek, S. Y. (2003). SCD1 is required for cytokinesis and polarized cell expansion in *Arabidopsis thaliana* [corrected]. *Development*, 130(17), 4011–4024.
- Fasoli, M., Dal Santo, S., Zenoni, S., Tornielli, G. B., Farina, L., Zamboni, A., Porceddu, A., Venturini, L., Bicego, M., Murino, V., Ferrarini, A., Delledonne, M., & Pezzotti, M. (2012). The grapevine expression atlas reveals a deep transcriptome shift driving the entire plant into a maturation program. *The Plant Cell*, 24(9), 3489–3505.
- Forcato, C. (2010). *Gene prediction and functional annotation in the Vitis vinifera genome* [Università degli studi di Padova].
- Foster, T. M., & Aranzana, M. J. (2018). Attention sports fans! The far-reaching contributions of bud sport mutants to horticulture and plant biology. *Horticulture Research*, 5, 44.
- Frank, M. H., & Chitwood, D. H. (2016). Plant chimeras: The good, the bad, and the “Bizzaria.” *Developmental Biology*, 419(1), 41–53.
- Franks, T., Botta, R., Thomas, M. R., & Franks, J. (2002). Chimerism in grapevines: implications for cultivar identity, ancestry and genetic improvement. *Theoretical and Applied Genetics*, 104(2-3), 192–199.
- Gambino, G., Dal Molin, A., Boccacci, P., Minio, A., Chitarra, W., Avanzato, C. G., Tononi, P., Perrone, I., Raimondi, S., Schneider, A., Pezzotti, M., Mannini, F., Gribaudo, I., & Delledonne, M. (2017). Whole-genome sequencing and SNV genotyping of “Nebbiolo” (*Vitis vinifera* L.) clones. *Scientific Reports*, 7(1), 17294.
- Gerrath, J., Posluszny, U., & Melville, L. (2015). *Taming the Wild Grape*. Springer International Publishing.
- Hofmeister, B. T., Denkena, J., Colomé-Tatché, M., Shahryary, Y., Hazarika, R., Grimwood, J., Mamidi, S., Jenkins, J., Grabowski, P. P., Sreedasyam, A., Shu, S., Barry, K., Lail, K., Adam, C., Lipzen, A., Sorek, R., Kudrna, D., Talag, J., Wing, R., ... Schmitz, R. J. (2020). A genome assembly and the somatic genetic and epigenetic mutation rate in a wild long-lived perennial *Populus trichocarpa*. *Genome Biology*, 21(1), 259.
- Jaillon, O., Aury, J.-M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., Vezzi, A., Legeai, F., Huguency, P., Dasilva, C., Horner, D., Mica, E., Jublot, D., Poulain, J., Bruyère, C., ... French-Italian Public Consortium for Grapevine Genome Characterization. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, 449(7161), 463–467.

- Jeffares, D. C., Jolly, C., Hoti, M., Speed, D., Shaw, L., Rallis, C., Balloux, F., Dessimoz, C., Bähler, J., & Sedlazeck, F. J. (2017). Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nature Communications*, *8*, 14061.
- Jung, H. Y. (2002). “*Candidatus* Phytoplasma castaneae”, a novel phytoplasma taxon associated with chestnut witches’ broom disease. *International Journal of Systematic and Evolutionary Microbiology*, *52*(5), 1543–1549.
- Khan, A. J., Botti, S., Al-Subhi, A. M., Gundersen-Rindal, D. E., & Bertaccini, A. F. (2002). Molecular identification of a new phytoplasma associated with alfalfa witches’ broom in oman. *Phytopathology*, *92*(10), 1038–1047.
- Li, H. (2021). New strategies to improve minimap2 alignment accuracy. *Bioinformatics*, *37*(23), 4572–4574.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, *17*(1), 10–12.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303.
- Meister, A. L. F. (1769). *Generalia de genesi figurarum planarum et inde pendentibus earum affectionibus*.
- Montano, H. G., Davis, R. E., Dally, E. L., Hogenhout, S., Pimentel, P., & Brioso, P. S. T. (2001). ‘*Candidatus* Phytoplasma brasiliense’, a new phytoplasma taxon associated with hibiscus witches’ broom disease. *International Journal of Systematic and Evolutionary Microbiology* *51*: 1109–1118.
- Ong-Abdullah, M., Ordway, J. M., Jiang, N., Ooi, S.-E., Kok, S.-Y., Sarpan, N., Azimi, N., Hashim, A. T., Ishak, Z., Rosli, S. K., Malike, F. A., Bakar, N. A. A., Marjuni, M., Abdullah, N., Yaakub, Z., Amiruddin, M. D., Nookiah, R., Singh, R., Low, E.-T. L., ... Martienssen, R. A. (2015). Loss of Karma transposon methylation underlies the mantled somaclonal variant of oil palm. *Nature*, *525*(7570), 533–537.
- Pacific Biosciences. (2021). pbsv. <https://github.com/PacificBiosciences/pbsv>. Accessed 20 September 2023.
- Porebski, S., Bailey, L. G., & Baum, B. R. (1997). Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Molecular Biology Reporter*, *15*(1), 8–15.

- Qiagen. (2015). QIAGEN® Genomic DNA Handbook.
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842.
- R Core Team. (2022). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. <https://www.R-project.org/>. Accessed 20 September 2023.
- Riaz, S., Garrison, K. E., Dangl, G. S., Boursiquot, J.-M., & Meredith, C. P. (2002). Genetic Divergence and Chimerism within Ancient Asexually Propagated Winegrape Cultivars. *Journal of the American Society for Horticultural Science*, *127*(4).
- RStudio Team. (2022). RStudio: Integrated Development Environment for R. Posit Software. <https://posit.co/products/open-source/rstudio/>. Accessed 20 September 2023.
- Shumate, A., & Salzberg, S. L. (2021). Liftoff: accurate mapping of gene annotations. *Bioinformatics*, *37*(12), 1639–1643.
- Sichel, V., Sarah, G., Girollet, N., Laucou, V., Roux, C., Roques, M., Mournet, P., Cunff, L. L., Bert, P. F., This, P., & Lacombe, T. (2023). Chimeras in Merlot grapevine revealed by phased assembly. *BMC Genomics*, *24*(1), 396.
- Smolka, M., Paulin, L. F., Grochowski, C. M., Horner, D. W., Mahmoud, M., Behera, S., Kalef-Ezra, E., Gandhi, M., Hong, K., Pehlivan, D., Scholz, S. W., Carvalho, C. M. B., Proukakis, C., & Sedlazeck, F. J. (2024). Detection of mosaic and population-level structural variants with Sniffles2. *Nature Biotechnology*.
- Srinivasan, C., & Mullins, M. G. (1981). Physiology of Flowering in the Grapevine — a Review. *American Journal of Enology and Viticulture*, *32*(1), 47–63.
- Urich, M. A., Nery, J. R., Lister, R., Schmitz, R. J., & Ecker, J. R. (2015). MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. *Nature Protocols*, *10*(3), 475–483.
- Urra, C., Sanhueza, D., Pavez, C., Tapia, P., Núñez-Lillo, G., Minio, A., Miossec, M., Blanco-Herrera, F., Gainza, F., Castro, A., Cantu, D., & Meneses, C. (2023). Identification of grapevine clones via high-throughput amplicon sequencing: a proof-of-concept study. *G3*, *13*(9).
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics*, *43*(1110), 11.10.1–11.10.33.

- Vondras, A. M., Minio, A., Blanco-Ulate, B., Figueroa-Balderas, R., Penn, M. A., Zhou, Y., Seymour, D., Ye, Z., Liang, D., Espinoza, L. K., Anderson, M. M., Walker, M. A., Gaut, B., & Cantu, D. (2019). The genomic diversification of grapevine clones. *BMC Genomics*, *20*(1), 972.
- Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, *38*(16), e164.
- Wick, R. R., Judd, L. M., Gorrie, C. L., & Holt, K. E. (2017). Completing bacterial genome assemblies with multiplex MinION sequencing. *Microbial Genomics*, *3*(10), e000132.
- Wilke CO. (2021). cowplot: streamlined plot theme and plot annotations for “ggplot2.” R package version 1.1.1. <https://CRAN.R-project.org/package=cowplot>. Accessed 20 September 2023.
- Wickham, H. (2016). Programming with ggplot2. In H. Wickham (Ed.), *ggplot2: Elegant Graphics for Data Analysis* (pp. 241–253). Springer International Publishing.
- Yu, L., Boström, C., Franzenburg, S., Bayer, T., Dagan, T., & Reusch, T. B. H. (2020). Somatic genetic drift and multilevel selection in a clonal seagrass. *Nature Ecology & Evolution*, *4*(7), 952–962.
- Zhuk, E., Vasilyeva, G., & Goroshkevich, S. (2015). Witches’ broom and normal crown clones from the same trees of *Pinus sibirica*: a comparative morphological study. *Trees*, *29*(4), 1079–1090.

APPENDIX A: CHAPTER 2 SUPPLEMENT

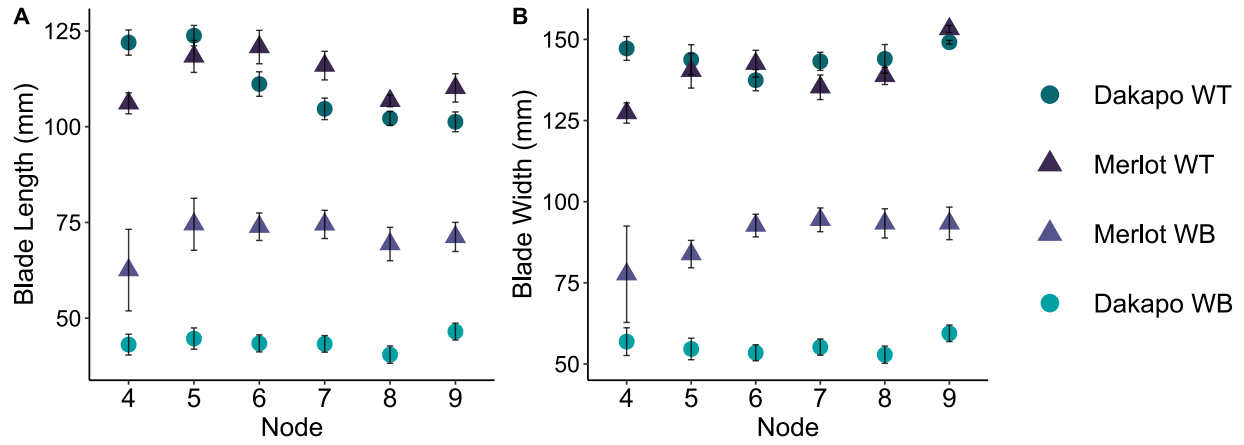


Figure S2.1. Average leaf **(A)** blade length and **(B)** blade width at distinct nodes on the shoots for each sample, collected from 10 shoots each.

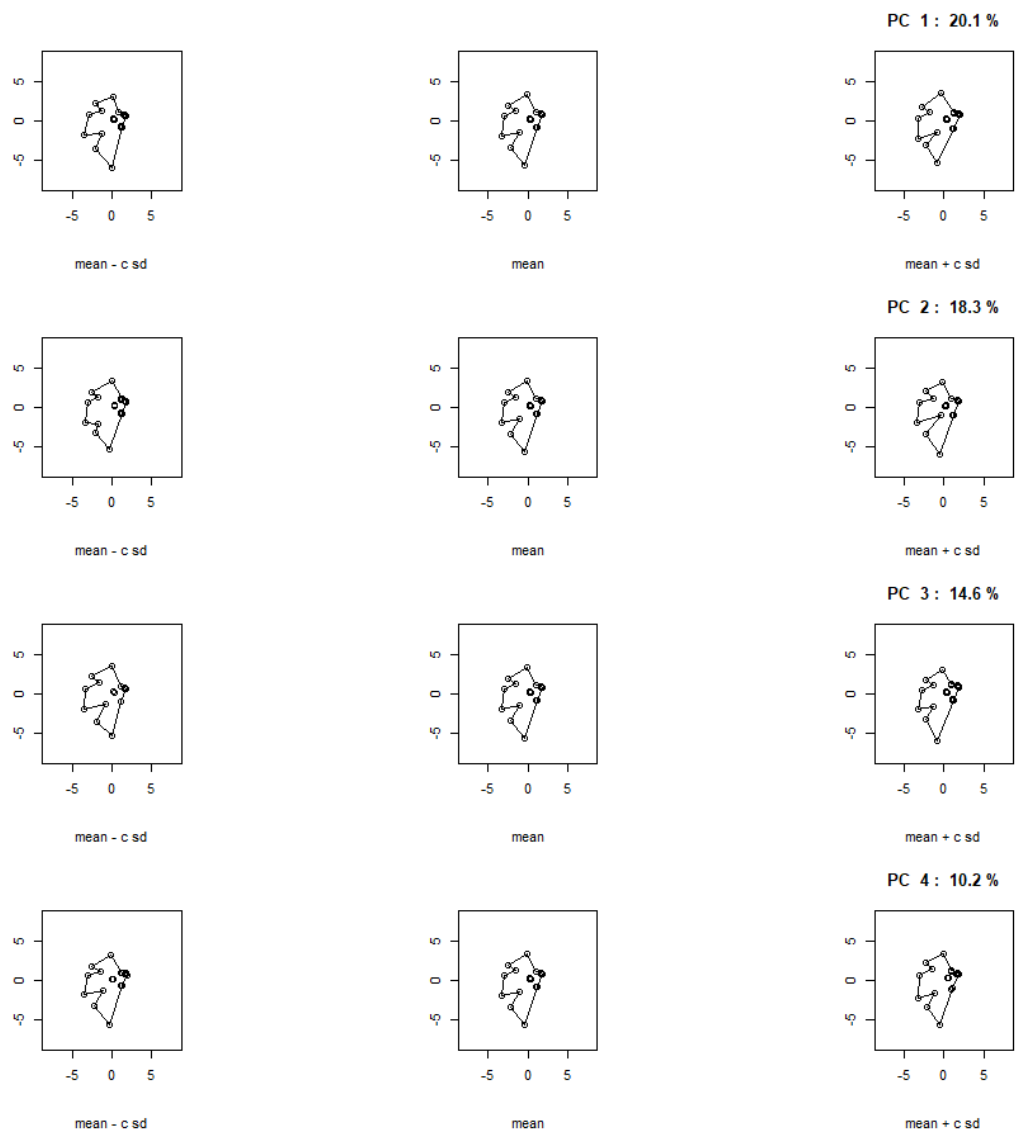


Figure S2.2. Eigenleaves from the PCA comparing leaf shape between scaled Dakapo WT and Dakapo WB leaves, for PC 1-4.

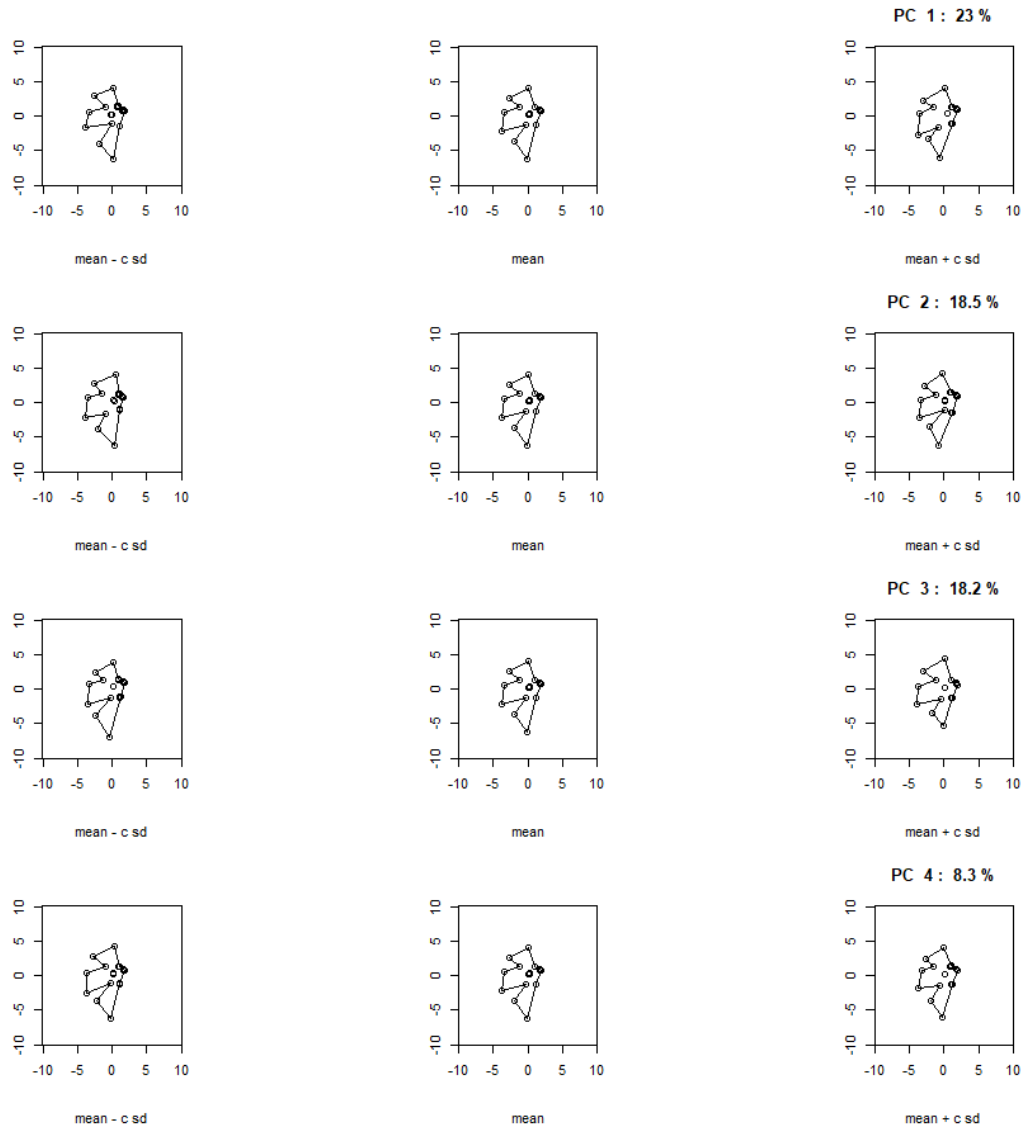


Figure S2.3. Eigenleaves from the PCA comparing leaf shape between scaled Merlot WT and Merlot WB leaves, for PC 1-4.

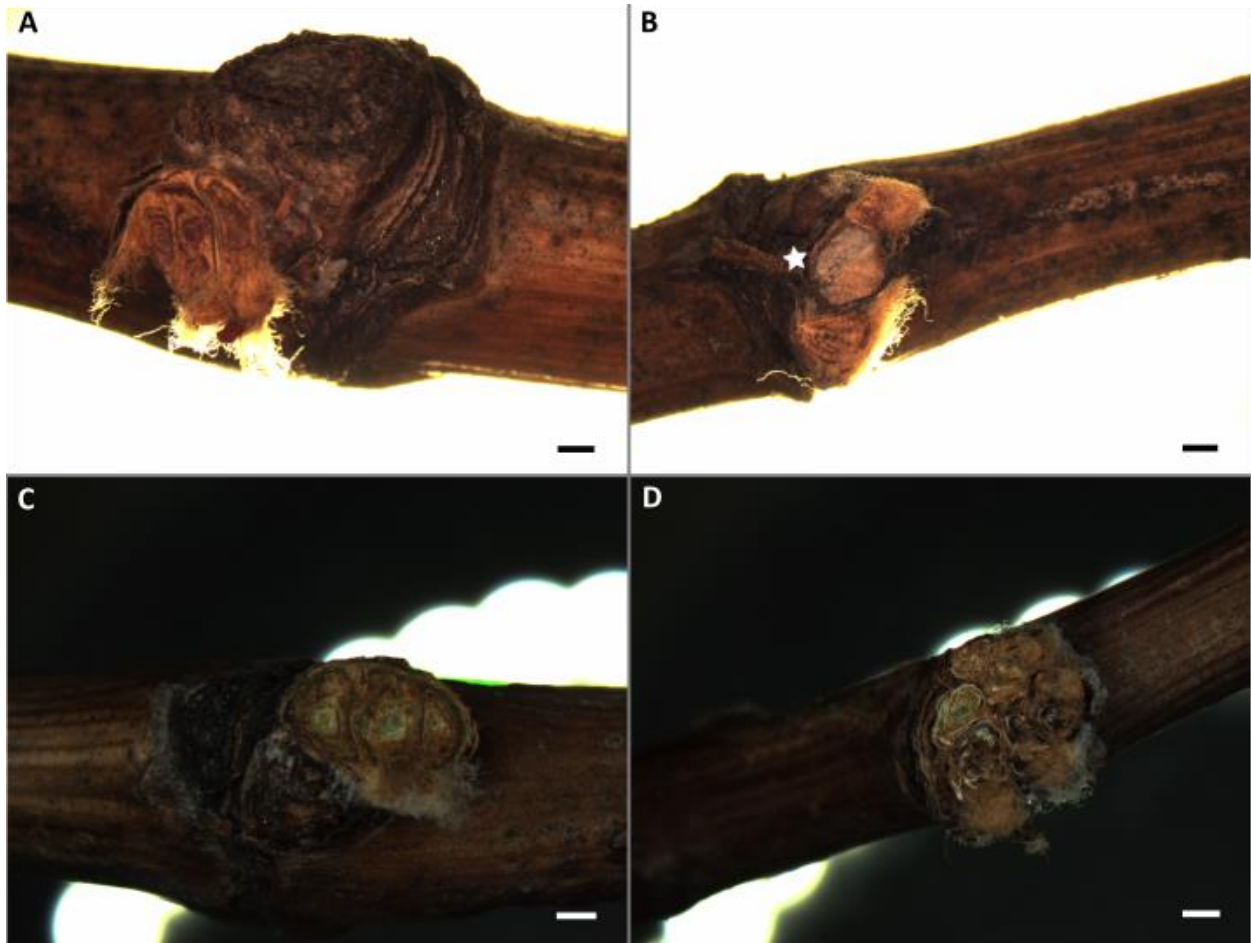


Figure S2.4. Buds imaged using a dissecting microscope for **(A)** Dakapo WT, **(B)** Dakapo WB, **(C)** Merlot WT, and **(D)** Merlot WB samples. The vascular tissue projecting out of the Dakapo WB sample is directly right of the solid star symbol. The scale bars are 1 mm wide.

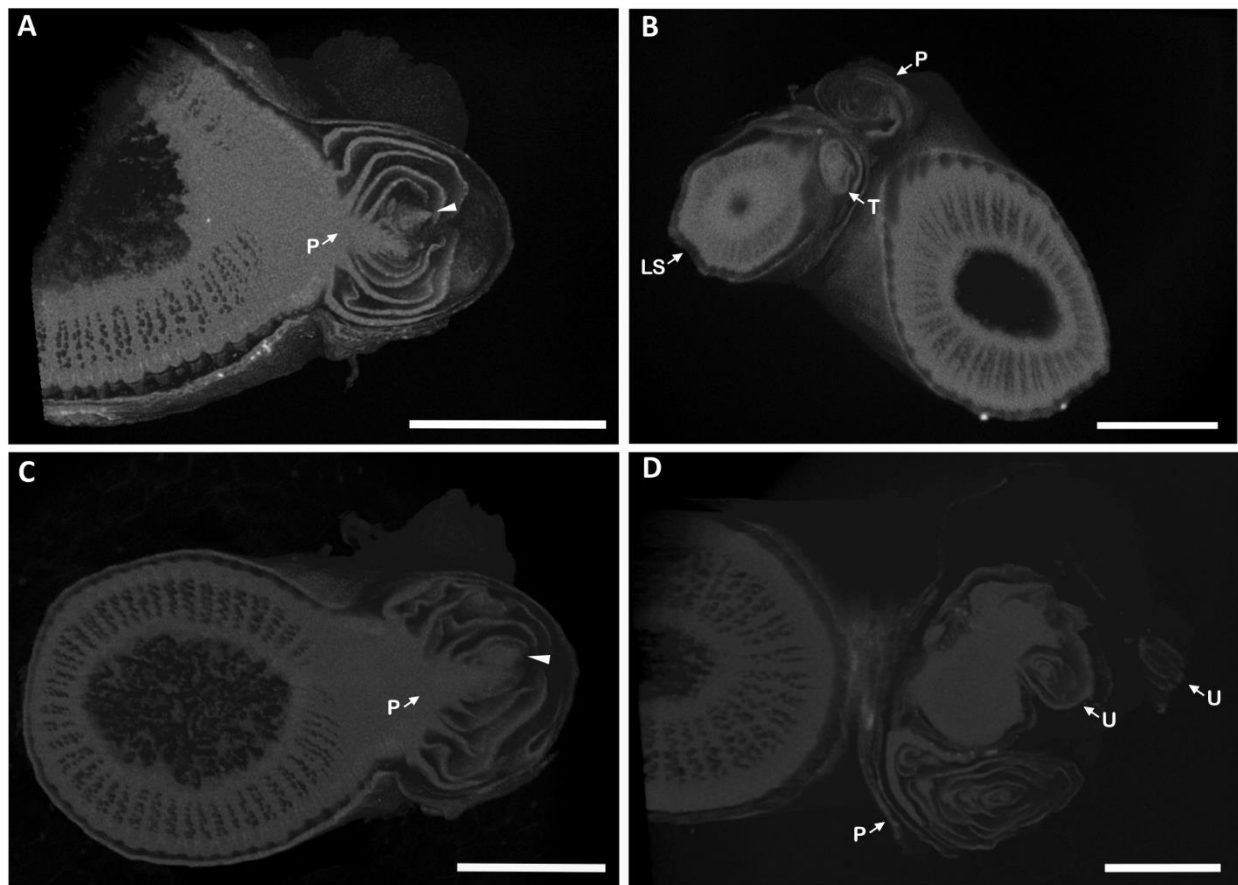


Figure S2.5. CT scans top-down of buds from **(A)** Dakapo WT, **(B)** Dakapo WB, **(C)** Merlot WT, and **(D)** Merlot WB samples. Primary primordia (P) are labeled in all four samples. The inflorescence primordia are indicated by the solid triangle in the **(A)** Dakapo WT and **(C)** Merlot WT samples. The tertiary primordium (T) and the lateral shoot stem (LS) are indicated in **(B)** Dakapo WB and the uncharacterized primordia (U) in **(D)** Merlot WB are indicated as well. Additional bud primordia are present in all samples but obscured due to the angle of the images. Additional inflorescence primordia are present in the WT samples but obscured in the images as well. Scale bar = 2 mm.

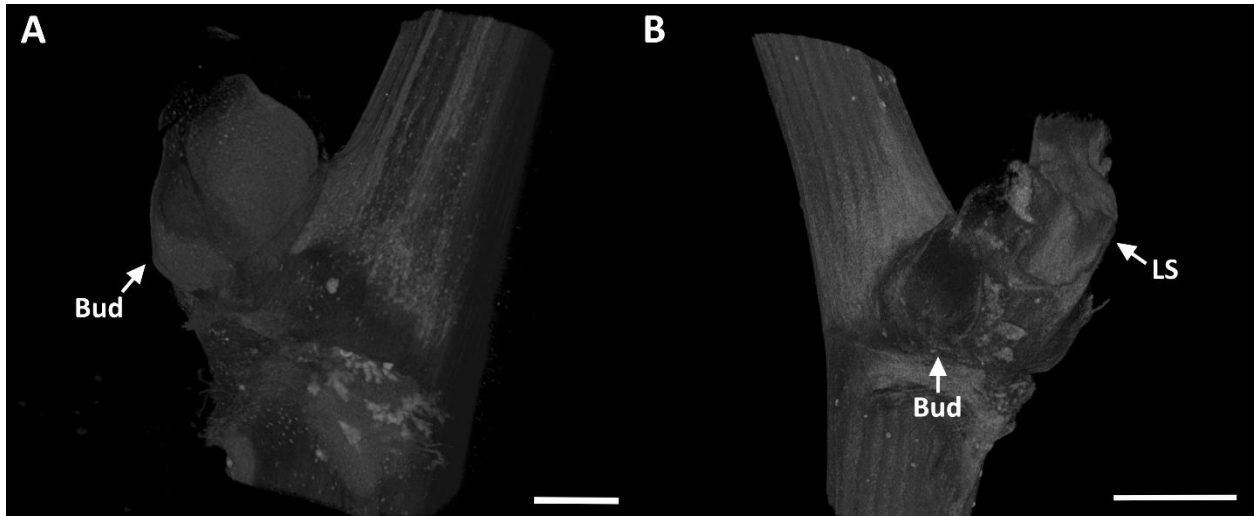


Figure S2.6. CT scans of **(A)** Dakapo WT and **(B)** Dakapo WB stems externally. The buds for both samples are labeled, as well as the initiated lateral shoot stem (LS) in Dakapo WB. In the Dakapo WB sample, an additional bud is present on the other side of the LS but obscured. Scale bar = 2 mm.

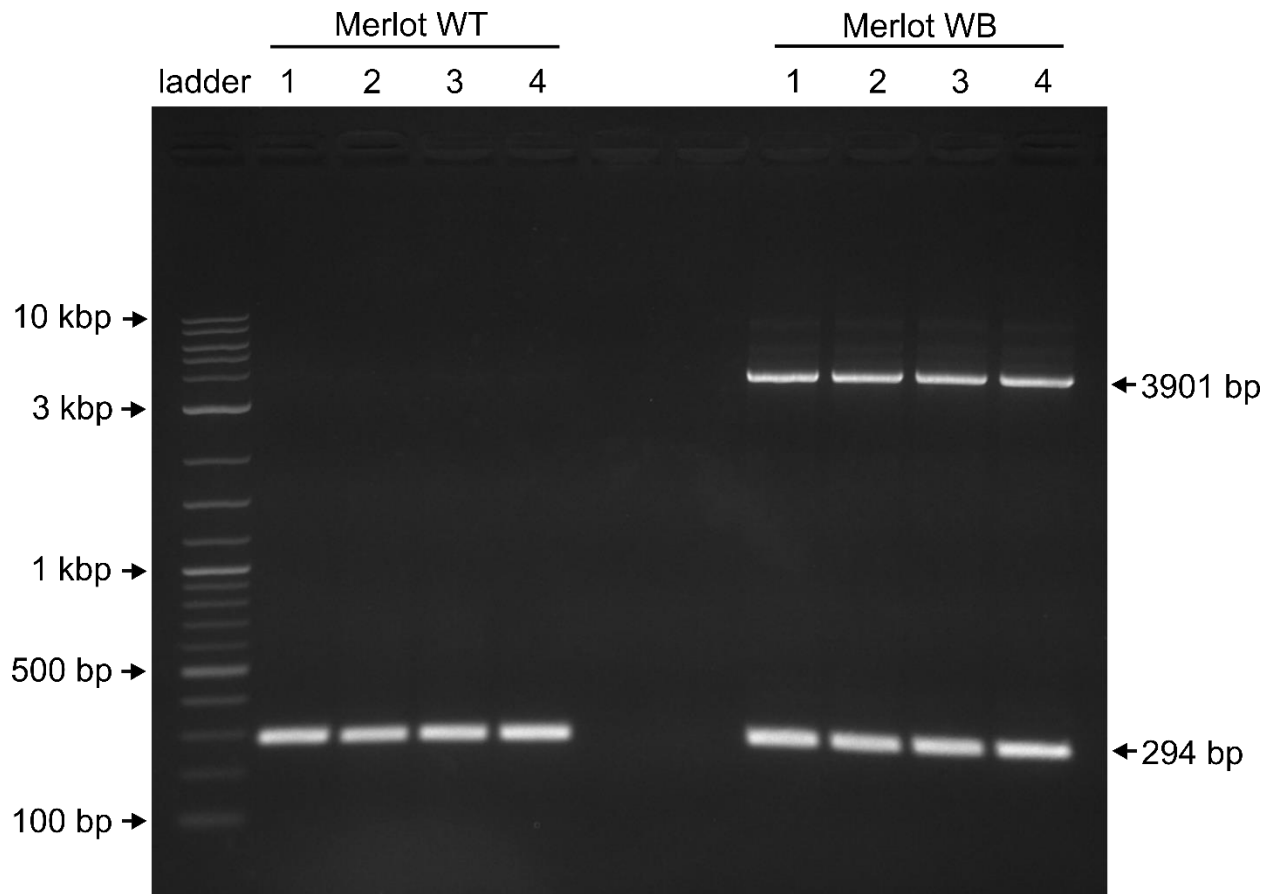


Figure S2.7. Agarose gel (1.2%) electrophoresis of PCR-amplified products using VvSCD1 primers. The ladder lane contains the 1 Kb Plus DNA ladder (NEB). Four technical replicates were run for each sample and were loaded into individual lanes. The amplified wild-type sequence is expected to be 294 bp in length, while the amplified WB sequence with the 3.6 kbp insertion present is expected to be 3901 bp in length. The Merlot WT sample only had bands present at 294 bp. The Merlot WB sample had two bands, one at 294 bp and one at 3901 bp, demonstrating that it is heterozygous for the 3.6 kbp insertion. Sanger sequencing of these individual DNA fragments confirmed that the bands at 294 bp were the amplified wild-type sequence of GSVIVG01008260001. Sanger sequencing also confirmed that the bands at 3901 bp were the amplified sequence of the wild-type sequence as well as the 3.6 kbp insertion within the wild-type sequence.

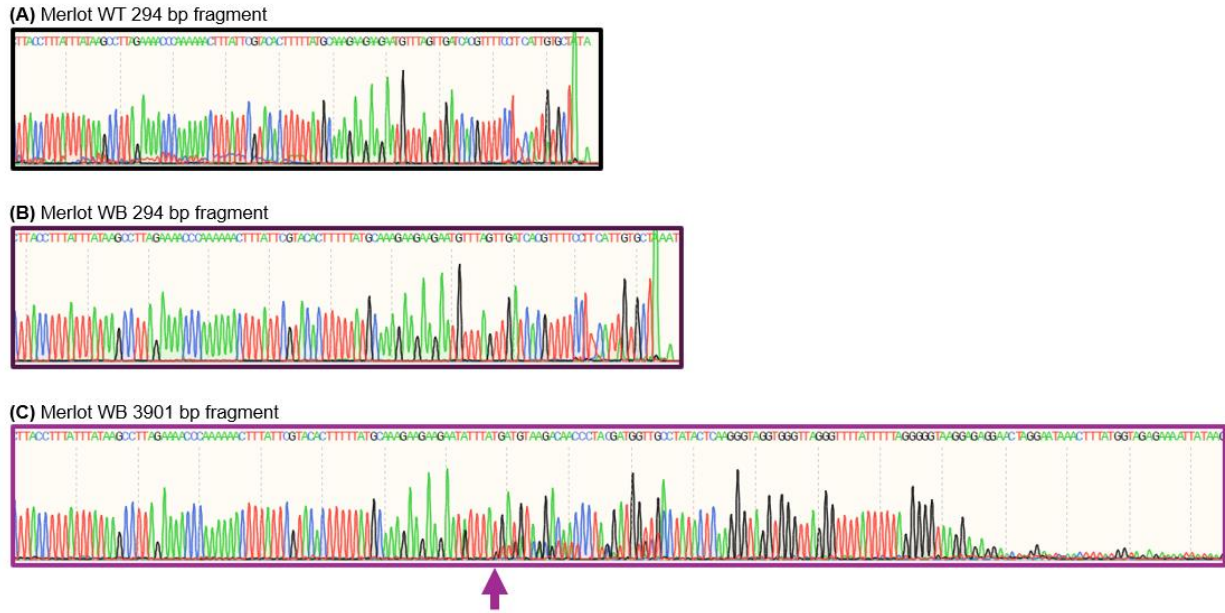


Figure S2.8. Sanger sequencing data of purified fragments from the region around the insertion within GSVIVG01008260001 for the 294 bp fragments amplified in **(A)** Merlot WT, **(B)** Merlot WB, as well as the **(C)** 3901 bp fragment in Merlot WB. Sanger sequencing data generated using the reverse VvSCD1 primer are shown. The sequences shown all start at identical locations within WT sequence and end at the end of the sequence generated through Sanger sequencing. The purple arrow shows approximately where the insertion sequence begins in the Merlot WB 3901 bp fragment.

GSVIVG01008260001

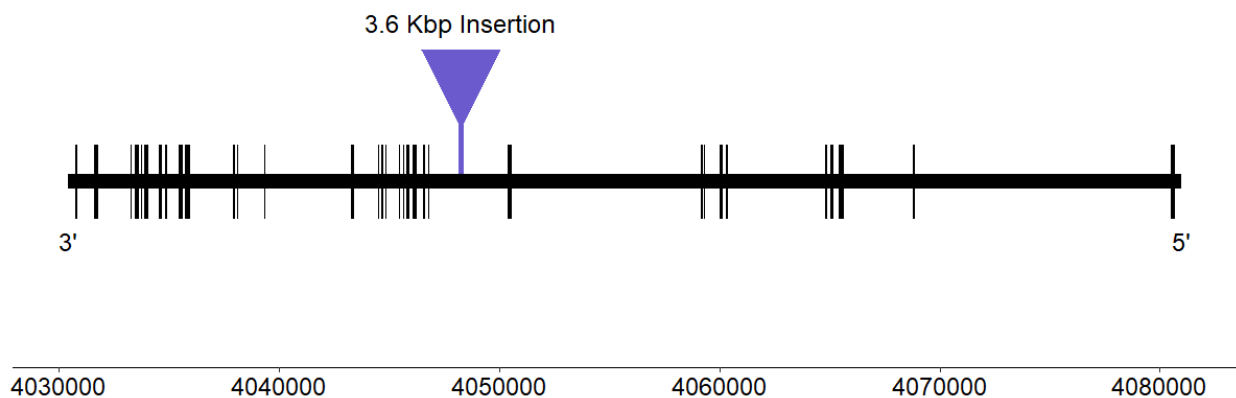


Figure S2.9. A diagram of the gene GSVIVG01008260001, the grapevine ortholog for AtSCD1. Exons are represented by black boxes along the gene body. The location and relative size of the 3.6 Kbp insertion present in Merlot WB is shown by the light purple line and triangle.

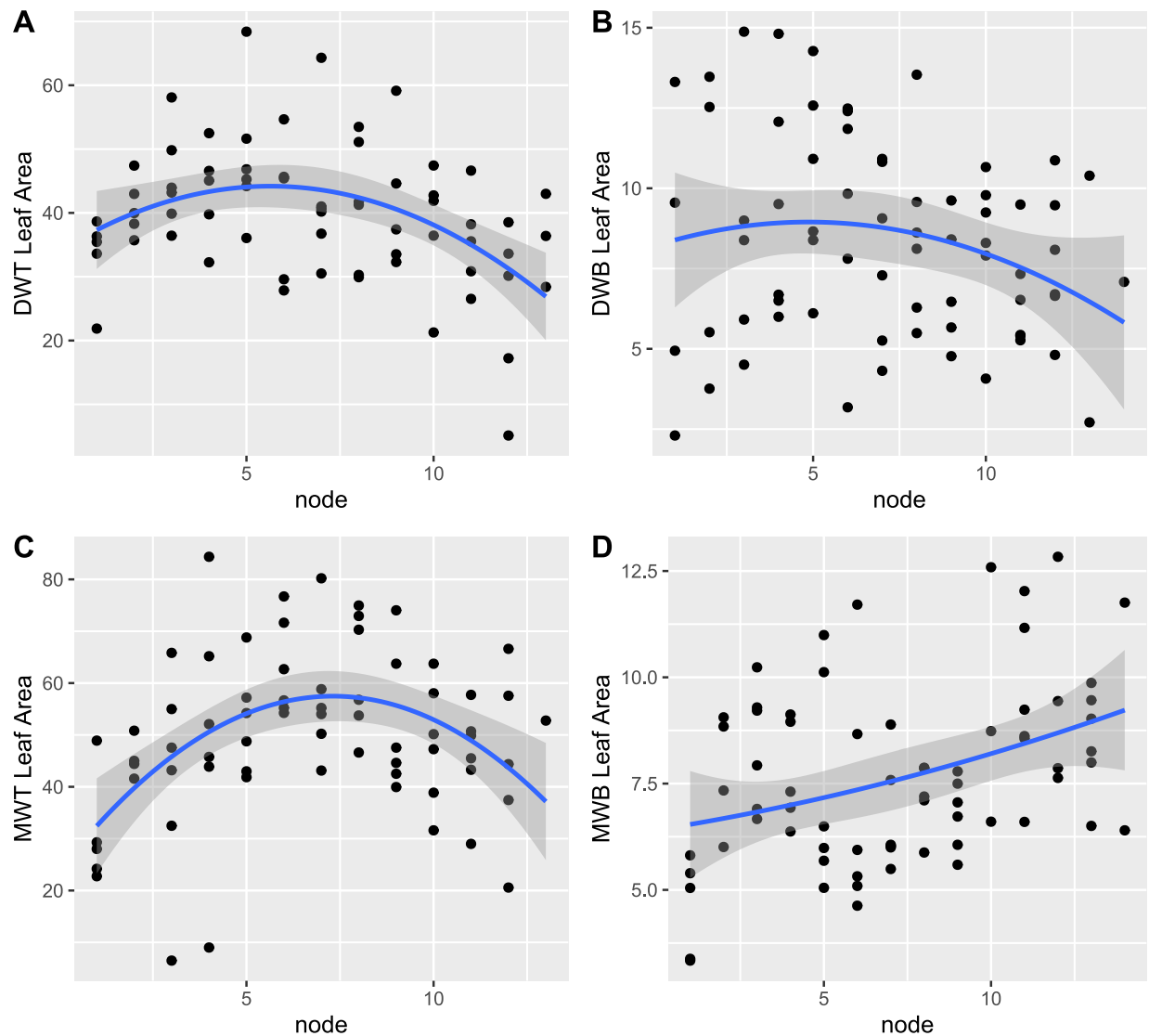


Figure S2.10. The developmental trajectories of leaf area across shoots for **(A)** Dakapo WT, **(B)** Dakapo WB, **(C)** Merlot WT, and **(D)** Merlot WB. The blue line represents the linear model of the formula $y \sim x + x^2$, in which y is leaf area and x is node position. There was significant support for this negative quadratic relationship between leaf area and node in both Dakapo WT and Merlot WT ($P < 0.05$ for both x and x^2 components for both varieties). However, there is not significant support for a negative quadratic relationship between leaf area and node in both Dakapo WB and Merlot WB ($P > 0.05$ for both x and x^2 components for both varieties).

Table S2.1. Summary of sequencing statistics for Illumina and Oxford Nanopore Technologies sequencing data used in this study. Due to length, this table is available in the supplemental files.

Predicted SNP Effects	Dakapo		Merlot	
	WT	WB	WT	WB
downstream	302954	302701	308529	309713
intergenic	4485234	4494808	4559676	4570032
intronic	2451458	2454637	2594527	2597389
ncRNA_exonic	17	22	14	14
splicing	2448	2428	2456	2459
upstream	369709	372057	377116	377983
upstream; downstream	40420	40459	41191	41112
UTR3	68803	68712	70248	70432
UTR5	47161	47189	48194	48313
exonic				
frameshift	7827	7738	7814	7798
nonframeshift	3446	3443	3515	3532
nonsynonymous	115910	115683	117985	117927
stop gain	3382	3404	3391	3389
stop loss	479	485	495	499
synonymous	90373	89946	92738	92715
unknown	148	145	147	162

Table S2.2. Number of SNPs with predicted SNP effects for all four samples individually, when called against the 12X.v2 grapevine reference genome (Canaguier et al., 2017) using Illumina sequencing data.

	Dakapo		Merlot	
	WT	WB	WT	WB
Deletions	27,173	27,420	28,495	28,122
Insertions	24,492	24,911	25,677	25,326
Inversions	65	71	67	63
Transversions	662	754	636	575
Duplications	57	58	56	52

Table S2.3. SV types for all four samples individually, when called against the 12X.v2 grapevine reference genome (Canaguier et al., 2017) using long-read sequencing data.

Table S2.4. Genetic candidates for casual variants of Witch's Broom in Dakapo and Merlot. Due to length, this table is available in the supplemental files.

	Dakapo		Merlot	
	WT	WB	WT	WB
Novel SNPs	349,533	349,239	351,018	356,754
Novel INDELS	61,075	62,186	58,570	62,064

Table S2.5. Novel* SNPs and INDELS for all four samples.

*Novel variants are variants completely absent in the sample of the same variety

SUPPLEMENTARY METHODS

Variant validation

In order to validate the insertion in GSVIVG01008260001, we first performed PCR amplification using primers for the wild-type sequence around the 3.6 kbp insertion. These primers were ~27 bp upstream of the start of the insertion and ~267 bp downstream of the insertion, and their sequences were:

VvSCD1-Forward: AGCACAATGAAGGAAAACGTGA

VvSCD1-Reverse: CTCAACCGGTTACCAAGACGCG

The expected size of the wild-type DNA fragment amplified by these primers was expected to be ~294 bp in length, while the WB DNA fragment was expected to contain the full insertion sequence and be ~3901 bp in length. PCR was performed using the Q5® High-Fidelity DNA Polymerase (M0491; New England Biolabs) and following the manufacturer’s protocol for 25 µL reactions, only modifying the concentration of primers by adding 0.5 µL of 10 µM primers. PCR was performed using both MWT and MWB DNA from ONT sequencing in separate reactions. Amplification was then performed using a Veriti™ 96-Well Fast Thermal Cycler (Cat. No. 4375305; Applied Biosystems™) with the settings shown in Table S2.6.

1 cycle	98°C	30 seconds
	98°C	10 seconds
30 cycles	65°C	30 seconds
	72°C	3 minutes
1 cycle	72°C	7 minutes

Table S2.6. Thermocycler settings for PCR to verify insertion within GSVIVG01008260001. Once complete, the samples were stored at 4°C when not in use. To check the size of the amplified fragment(s), the products were run on a 1.2% Tris-acetate-EDTA (TAE) agarose gel in TAE buffer at 110 V for 1.5 hours. The gel was then imaged using a Axygen Gel Documentation System (GD-1000; Corning) with UV transillumination.

Following successful PCR amplification demonstrating the amplification of fragments of the size expected, we prepared samples for Sanger sequencing. To do so, PCR amplification was performed exactly as described for the initial PCR amplification, but with 100 µL

reactions to increase the final concentration of the amplified DNA fragments. The products were then run on a 1.2% Tris-acetate-EDTA (TAE) agarose gel in TAE buffer at 110 V for 1.5 hours. Once complete, the distinct bands were excised from the gel using a UV transilluminator. DNA was purified from the gel fragments using a QIAquick Gel Extraction Kit (Cat. No. 28704; QIAGEN), following the manufacturer's instructions. The concentration of the purified DNA was checked using a Qubit Broad Range (BR) DNA Assay Kit (Q32850) and an Invitrogen Qubit 4 Fluorometer. Samples were then prepared for Sanger sequencing by combining 10 ng of DNA, 3 μ L of 10 μ M primers, and water to volume (for 12 μ L samples). Three fragments were submitted to the MSU Genomics Core for Sanger Sequencing: (1) a ~294 bp band from MWT, (2) a ~294 bp band from MWB, and (3) a ~3901 bp band from MWB. These fragments were Sanger sequenced using an Applied Biosystems 3730xl Genetic Analyzer using both VvSCD1 primers for two separate runs per fragment. The Sanger sequencing data was viewed using SnapGene Viewer 6.0 and Benchling (2023).

Investigating GSVIVG01008260001 insertion sequence

To see if the 3.6 kbp insertion within GSVIVG01008260001 in Merlot WB showed sequence similarities to transposable elements (TEs), we first added a contig to the 12X.v2 grapevine reference assembly that contained the insertion as well as 50,000 base pairs upstream and downstream of the insertion site. We then ran EDTA v1.9.4 on this modified fasta file with the following flags: *--species others, --step all, --overwrite 1, --sensitive 1, --anno 1, --evaluate 0, and --force 1*. EDTA produced a gff file with the coordinates for high confidence TEs, and we looked into the TEs annotated within the 3.6 kbp insertion sequence.

CHAPTER 3:
Small, but mitey: investigating the molecular genetic basis for mite domatia development and intraspecific variation in *Vitis riparia* using transcriptomics

This chapter is currently in review at *New Phytologist* and available as pre-print on bioRxiv:

Ritter, E. J., Graham, C. D. K., Niederhuth, C., & Weber, M. G. (2024). Small, but mitey: Investigating the molecular genetic basis for mite domatia development and intraspecific variation in *Vitis riparia* using transcriptomics. In *bioRxiv* (p. 2024.03.04.583436).
<https://doi.org/10.1101/2024.03.04.583436>

AUTHORS AND AFFILIATIONS

Eleanore J. Ritter¹, Carolyn D. K. Graham², Chad Niederhuth¹, and Marjorie Gail Weber²

¹Department of Plant Biology, Michigan State University, East Lansing, MI, USA

²Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA

AUTHOR CONTRIBUTIONS

MGW, CN, and EJR conceptualized and designed the study. CDKG provided plant material and characterized domatia traits. EJR carried out the RNA-seq analysis and performed leaf landmarking. MGW, CN, and EJR assisted in data interpretation. EJR wrote the first draft of the manuscript. All authors assisted with final drafts of the manuscript.

SUMMARY

- Here, we investigated the molecular genetic basis of mite domatia, structures on the underside of leaves that house mutualistic mites, and intraspecific variation in domatia size in *Vitis riparia* (riverbank grape).

- Domatia and leaf traits were measured, and the transcriptomes of mite domatia from two genotypes of *V. riparia* with distinct domatia sizes were sequenced to investigate the molecular genetic pathways that regulate domatia development and intraspecific variation in domatia traits.
- Key trichome regulators as well as auxin and jasmonic acid are involved in domatia development. Genes involved in cell wall biosynthesis, biotic interactions, and molecule transport/metabolism are upregulated in domatia, consistent with their role in domatia development and function.
- This work is one of the first to date that provides insight into the molecular genetic bases of mite domatia. We identified key genetic pathways involved in domatia development and function, and uncovered unexpected pathways that provide an avenue for future investigation. We also found that intraspecific variation in domatia size in *V. riparia* seems to be driven by differences in overall leaf development between genotypes.

INTRODUCTION

Mutualisms between plants and arthropods have evolved repeatedly across evolutionary time (Blattner et al., 2001; Bronstein et al., 2006), promoting the evolution of unique, heritable structures in plants that attract, reward, or protect mutualists (Bronstein et al., 2006; Romero & Benson, 2005). Investigating the genetic basis of mutualistic structures provides a valuable lens for understanding how mutualisms evolved. Mite domatia (hereafter “domatia”) are tiny mutualistic plant structures on the underside of leaves that provide shelter for beneficial mites that have received relatively little attention from the genetic perspective despite being produced by many woody plant species. Domatia

facilitate a bodyguard mutualism between plants and mites: mites benefit from the refuge provided by the domatia, which protects them from predators (Grostal & O'Dowd, 1994; Faraji et al., 2002a,b; Norton et al., 2001; Romero & Benson, 2005), and in return plants receive protection from pathogenic fungi and/or herbivory via fungivorous and/or predacious mites (Agrawal & Karban, 1997; Norton et al., 2000; Romero & Benson, 2004). Domatia are common defenses in natural systems: they are present in over 5,000 plant species and make up a large proportion of woody plant species in temperate deciduous forests (e.g., ~50% of woody plant species in forests in Korea (O'Dowd & Pemberton, 1998) and Eastern North America (Willson, 1991)). They are present in several crop plants and have been studied as a pest control strategy in agriculture (Barba et al., 2019; Romero & Benson, 2005). Yet, despite their agricultural and ecological importance, we know relatively little about the genetic underpinnings of mite domatia in plants.

The genus *Vitis* is a powerful group for studying the genetics of domatia due to the heritable variation of domatia presence and size (English-Loeb et al., 2002; Graham et al., 2023) and the genetic and germplasm resources available. In *Vitis*, domatia are constitutive, small, dense tufts of trichomes covering a depression in the leaf surface in the abaxial vein axils, termed “tuft” domatia. Norton et al. (2000) demonstrated that domatia in *Vitis riparia*, a wild grapevine species with relatively large domatia, led to a 48% reduction in powdery mildew in comparison to *V. riparia* plants with blocked domatia, which were inaccessible to mites. Given how effective domatia are as biological control agents in this system, there is interest in understanding domatia in domesticated grapevine (*Vitis vinifera*) and related species.

Two previous studies investigated the genetic basis of domatia in *Vitis* (Barba et al.,

2019; LaPlante et al., 2021). Barba et al. (2019) measured mite abundance, domatia, and general trichome traits in the segregating F₁ family of a complex *Vitis* hybrid cross. They identified multiple QTLs influencing domatia-related traits, including a major QTL on chromosome 1. They also found additional support for a relationship between overall leaf and leaf trichome development, previously demonstrated in *Vitis* (Chitwood et al., 2014). LaPlante et al. (2021) investigated the genetic basis of trichome and domatia traits in a genome-wide association study (GWAS) using a common garden of *V. vinifera* cultivars. They identified a single nucleotide polymorphism (SNP) associated with domatia density near several candidate genes on chromosome 5. Only one gene was identified that was shared in both studies: *Glabrous Inflorescence Stems 2* (VIT_205s0077g01390), which is thought to encode a zinc finger protein that regulates trichome development (LaPlante et al. 2021). The minimal overlap between the two studies is likely due to differences in the scale of genetic diversity investigated in QTL mapping and GWAS. As a result, the various molecular pathways involved in domatia development remain relatively unknown.

While little is known about the development of tuft domatia specifically, work in related structures in other species may provide clues regarding the genes involved in domatia development. Substantial work has characterized the genes involved in the development of trichomes, which are an essential component of tuft domatia. The molecular pathways involved in trichome development have yet to be elucidated in *Vitis*, though they have been characterized in other angiosperms. In addition, previous work has characterized the genes involved in another form of domatia that house ants, called tuber domatia. Tuber domatia are functionally like tuft domatia in providing shelter for mutualistic arthropods in return for defense, but are tubers formed from stem tissue. The genes involved in domatia

development may overlap with those previously implicated in trichome and tuber domatia development, providing additional hypotheses to investigate regarding domatia development in *V. riparia*.

Here, we investigate the molecular genetic mechanisms of development and intraspecific variation in domatia of *V. riparia*, the riverbank grape. We hypothesized that genes differentially expressed in *V. riparia* domatia (i) share similarities with pathways previously identified in trichome development, including TFs and cell wall modification pathways (Dong et al., 2022; Han et al., 2022), (ii) are involved in responses to biotic organisms as has been previously identified with functionally similar tuber domatia (Pu et al., 2021), and (iii) involve auxin-signaling due to its role in both trichome (Han et al., 2022) and tuber domatia development (Pu et al., 2021). We also hypothesized that intraspecific variation in domatia size in *V. riparia* may be driven by differences in overall leaf morphology, as previous work has demonstrated a link between leaf morphology and trichomes in *Vitis* (Barba et al., 2019; Chitwood et al., 2014). We sequenced the transcriptomes of domatia in two *V. riparia* genotypes that differ in their investment in domatia alongside control leaf tissue to identify key pathways involved in domatium development in *Vitis* and landmarked leaves from these two genotypes and compared leaf shapes to identify possible morphological differences that may impact domatia traits.

MATERIALS AND METHODS

Plant material

We pre-selected two genotypes of *V. riparia* that were identified in a previous study to nearly span domatia size variation in the species (English-Loeb & Norton, 2006): genotype 588711, a large domatia genotype (hereafter LDG) and genotype 588710, a small domatia

genotype (hereafter SDG). Hardwood cuttings of each genotype were sourced from the United States Department of Agriculture - Germplasm Resource Information Network (USDA-GRIN) repository in February of 2022. The genotypes were initially collected in Wyoming, USA (SDG) and Manitoba, Canada (LDG). Budding cuttings were potted in March of 2022 and grown in a common garden greenhouse at Michigan State University in East Lansing, Michigan, USA. The vines were watered daily during the growing season with roughly 200 ppm Peters Excel pHLow 15-7-25 High-Mag/High K with Black Iron fertilizer (ICL, Tel-Aviv, Israel) dissolved in water. No pesticides were applied to the plants.

Characterizing domatia traits

To confirm that the two domatia genotypes used in our study statistically differed in domatia size, we collected, dried, and pressed leaves and scored domatia traits using a dissection microscope. Domatia traits were measured on 1-3 fully expanded leaves per plant for five replicate plants per genotype (25 leaves total). To evaluate how domatia size and density changed throughout leaf ontogeny, we collected 5-7 leaves from five plants of each genotype selected to represent the entire leaf lifespan from bud burst to full expansion (55 leaves total). The leaves were scanned while fresh using a CanoScan 9000F Mark II (Canon U.S.A., Inc.) at 1200 DPI.

Two aspects of domatia were measured for both datasets: hair density and size. Domatia hair density scores were assigned using a nine-point scale where 0 represents no hair, and 9 represents a densely packed domatium with no leaf surface visible underneath (Graham et al., 2023). This scale was adapted from the OIV code O-085/U-33 scale, a standard scale for measuring leaf hair density used by grape breeders (IPGRI et al., 1997). Domatium radius was used as a proxy for domatia size, measured on pressed leaves

using an ocular micrometer, and in pixels and converted to mm using the software ImageJ 1.54d on leaf scans (Schneider et al., 2012).

A Mann-Whitney test and a Welch's two-sided t-test were used to test for differences in domatia density and size, respectively, between the two genotypes. The phenotypic data were plotted using ggplot2 v3.4.2 (Wickham, 2016) and cowplot v1.1.1 (Wilke, 2021) in R. The R package ggsignif v0.6.4 was used to add significance bars (Constantin & Patil, 2021). We evaluated domatia development throughout leaf expansion and identified our RNA sampling timepoints by comparing domatia size on leaves that had not fully expanded (ranging from 2.6-5.8 cm in width) to domatia size on larger leaves using a Welch's two-sided t-test (Figure S3.1). All R analyses (including downstream analyses) were run using R v4.2.2 (R Core Team, 2022) and RStudio v2022.12.0.353 (RStudio Team, 2022).

Tissue collection, RNA extraction, and sequencing

RNA samples were collected across two consecutive days at the same time (within one hour from start to finish) from the same plants used for characterizing domatia traits. Samples were collected from young leaves that had not fully expanded (ranging from 2.6-5.8 cm in width), as our domatium ontogeny data demonstrated that domatia were still developing during this time ($P < 0.001$, Figure S3.1). The samples were collected using a circular 1.5 mm hole puncher on domatia and control tissues. Domatia samples were collected at the center of domatia, avoiding veins, and control samples were on laminar tissue 1.5 mm away from the domatia (Figure S3.2). Domatia and control samples were taken from the same leaves, with the first sample (domatia or control) alternated. Between 20-23 tissue samples were taken per plant (across 2-3 leaves) for both domatia and control. Samples were immediately frozen in liquid nitrogen and stored at -80°C.

Samples were pooled into grinding tubes by plant and tissue type and ground using two metal balls per tube in a SPEX™ SamplePrep 2010 Geno/Grinder 2010 at 1750 strokes/minute for 30 seconds. RNA was immediately extracted from pooled samples after grinding using a Spectrum™ Plant Total RNA Kit from Sigma-Aldrich (STRN250) and an On-Column DNase I Digestion Set from Sigma-Aldrich (DNASE70-1SET) to remove DNA. Concentrations were checked with a Qubit High Sensitivity (HS) RNA Assay Kit (Q32852) and an Invitrogen Qubit 4 Fluorometer. The quality of the samples was analyzed using an Agilent 4200 TapeStation, and samples with low quantity and/or quality were discarded.

Fourteen RNA samples with sufficient quantity and quality - paired domatia-control samples from three 588710 plants and four 588711 plants - were prepared for RNA-sequencing (RNA-seq) using the Illumina Stranded mRNA Library Preparation, Ligation Kit with IDT and Illumina Unique Dual Index adapters following the manufacturer's recommendations, except half volume reactions were used. The quantity and quality of the prepared libraries were checked using a Qubit HS dsDNA Assay Kit (Q32851) and the High Sensitivity D1000 ScreenTape assay, respectively. The libraries were pooled and sequenced using one Illumina S4 flow cell lane in a 2x150bp paired-end format and a NovaSeq v1.5 reagent kit (300 cycles). Base calling was done by Illumina Real Time Analysis (RTA) v3.4.4. The output of RTA was demultiplexed and converted to FastQ format with Illumina Bcl2fastq v2.20.0.

RNA read processing and mapping

FastQC v0.11.9 was used to check the quality of the raw RNA-seq reads (Andrews, 2010). One sample failed poly-A capture during library preparation, so both samples from the failed library were excluded from downstream analysis, leaving three paired RNA-seq

samples for each genotype. RNA-seq reads were trimmed using Trimmomatic v0.39 with the *-phred33* flag and the provided NexteraPE-PE.fa adapter file (Bolger et al., 2014). The quality of the trimmed reads was confirmed using FastQC v0.11.9 (Andrews, 2010). After read trimming, there were 26-51 million reads per sample (Table S3.1).

Trimmed RNA-seq reads were mapped to the *V. riparia* genome (Girrollet et al., 2019) using STAR v2.7.0c (Dobin et al., 2013). STAR genome index files were generated by running STAR with the *V. riparia* genome and annotations, using the flags *--runMode genomeGenerate* and *--genomeSAindexNbases 13*. The trimmed reads were mapped using STAR with the generated index files and the *--quantMode GeneCounts* flag. Between 23-46 million reads (84.9-91.6% of trimmed reads) mapped to the *V. riparia* genome (Table S3.1).

Differential expression analysis

Mapped read count outputs from STAR were used for differential expression analysis using DESeq2 v1.38.3 (Love et al., 2014) downloaded from Bioconductor (Huber et al., 2015). DESeq2 was run using the *design = ~genotype + tissue + genotype:tissue*, where tissue was either control or domatia. Genes were considered differentially expressed between groups if the absolute value log₂ fold change was greater than 1 and the adjusted *P*-value was less than 0.05 (Tables S3.2-S3.5). Heatmaps displaying differentially expressed genes (DEGs) were created using the ComplexHeatmap v2.14.0 package (Gu, 2022).

GO term enrichment analysis

Gene ontology (GO) term enrichment analysis was performed on upregulated DEGs. GO term annotations are more robust in *Arabidopsis* (*Arabidopsis thaliana*), so we utilized *Arabidopsis* orthologs for our GO term enrichment analysis. We identified *Arabidopsis* orthologs to *V. riparia* genes using protein sequences for both species with diamond

v2.0.15.153 (Buchfink et al., 2015) and the following flags: *--iterate*, *--max-target-seqs 1*, and *--unal 0*. GO terms of Arabidopsis orthologs for DEGs were used from The Arabidopsis Information Resource (TAIR) (Berardini et al., 2015). Interproscan v5.61-93.0 (Jones et al., 2014) was used to identify PFAM GO terms based on *V. riparia* protein sequences. GO terms for the Arabidopsis orthologs and those generated by PFAM were concatenated, and parent GO terms were added using GO files from the GO knowledgebase v2023-11-15 (Ashburner et al., 2000; Gene Ontology Consortium, 2023). GO term enrichment for upregulated DEGs was assessed by performing a Fisher's exact test with Benjamini-Hochberg correction using TopGO v2.44.0 (Alexa & Rahnenfuhrer, 2023) from Bioconductor (Huber et al., 2015) against the background of *V. riparia* genes. GO terms were considered enriched with a *P*-value < 0.05. GO term enrichment results were plotted using ggplot2 v3.4.2 (Wickham, 2016).

Leaf landmarking and leaf shape analysis

To test whether differences in leaf shape may be related to differences in domatia size between SDG and LDG, we compared leaf shapes between the two. To do so, leaves used for measuring domatia ontogeny were landmarked manually by placing 21 landmarks described in Bryson et al. (2020) on leaf scans using ImageJ v1.53k (Schneider et al., 2012). Landmarks were saved as x- and y-coordinates in centimeters. Comparing differences in leaf shape was performed as described in Ritter et al. (2023) using the shapes package v1.2.7 (Dryden & Mardia, 2016). A Hotelling's T^2 test was used to test for mean shape differences. Average leaves for each sample and PC values were plotted using ggplot2 v3.4.2 (Wickham, 2016) and cowplot v1.1.1 (Wilke, 2021).

RESULTS

Genotypes differ in domatia investment

We characterized domatia traits in two *V. riparia* genotypes previously reported to have different domatia sizes (English-Loeb & Norton, 2006). The two genotypes differed in domatia size, with LDG having larger domatia than SDG ($P < 0.001$) (Figure 3.1). The genotypes had similarly dense domatia ($P = 0.382$).

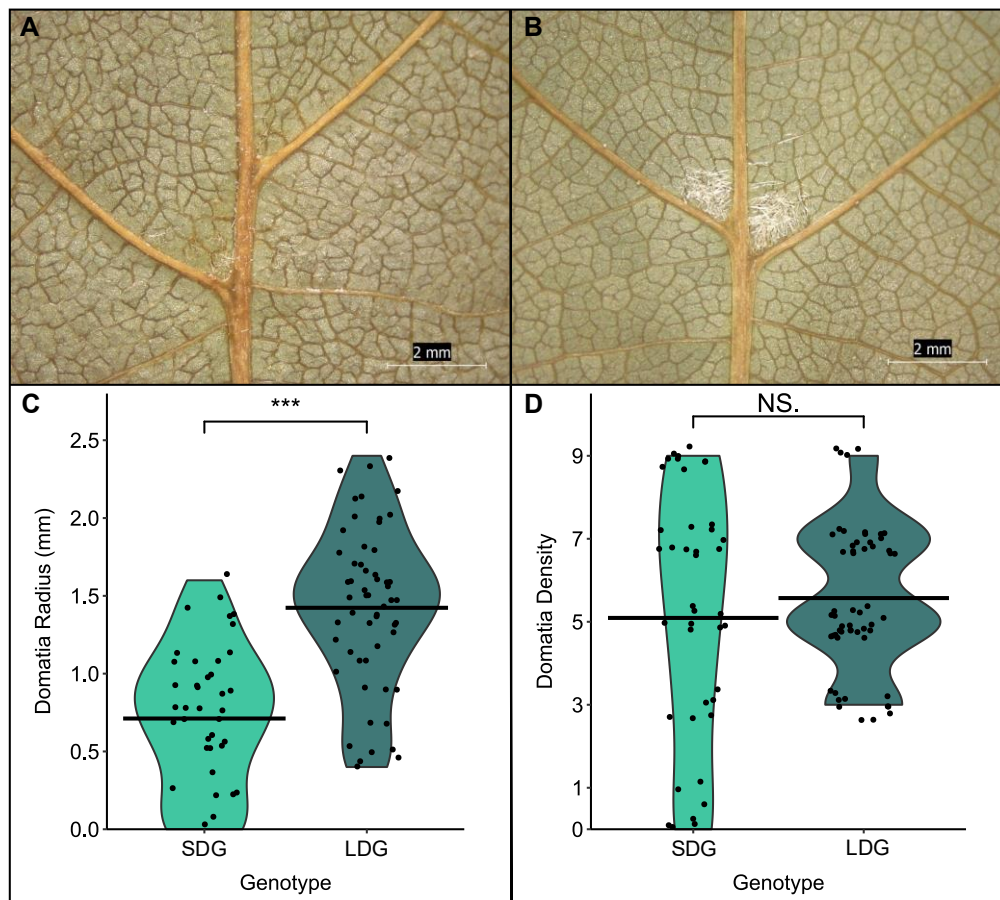


Figure 3.1. Domatia size and density of *V. riparia* SDG and LDG plants. **(A)** Domatia from the SDG. **(B)** Domatia from the LDG. **(C)** The radius of domatia (mm) from SDG and LDG plants, with the average radius of domatia represented by a black line ($***P < 0.001$). **(D)** The domatia density of both genotypes, from 0 (representing essentially no domatium present) to 9 (representing a very dense domatium), with the average density for each genotype represented by a black line (NS., $P = 0.382$).

However, the range of domatia density in the SDG (0-9) was greater than the range of

domatia density in LDG (3-9) due to both absent and very sparse domatia in the SDG (scored 0 and 1, respectively) (Figure 3.1).

Differentially expressed genes in domatia

Differential expression analysis revealed 1,447 and 759 DEGs in the SDG and the LDG domatia compared to control tissue, respectively. Most DEGs were upregulated in the domatia (88.6% in SDG and 94.9% in LDG). There was substantial overlap of DEGs between the two genotypes, with 538 genes (~37% SDG, ~71% LDG) overlapping (Figure S3.3).

GO term enrichment analysis revealed 98 and 58 Biological Process (BP) GO terms enriched in SDG and LDG, respectively (Tables S3.6 and S3.7). Of these, 39 were shared between genotypes (Figure 3.2) and primarily fell into three categories - development, hormone signaling, and responses to stimuli.

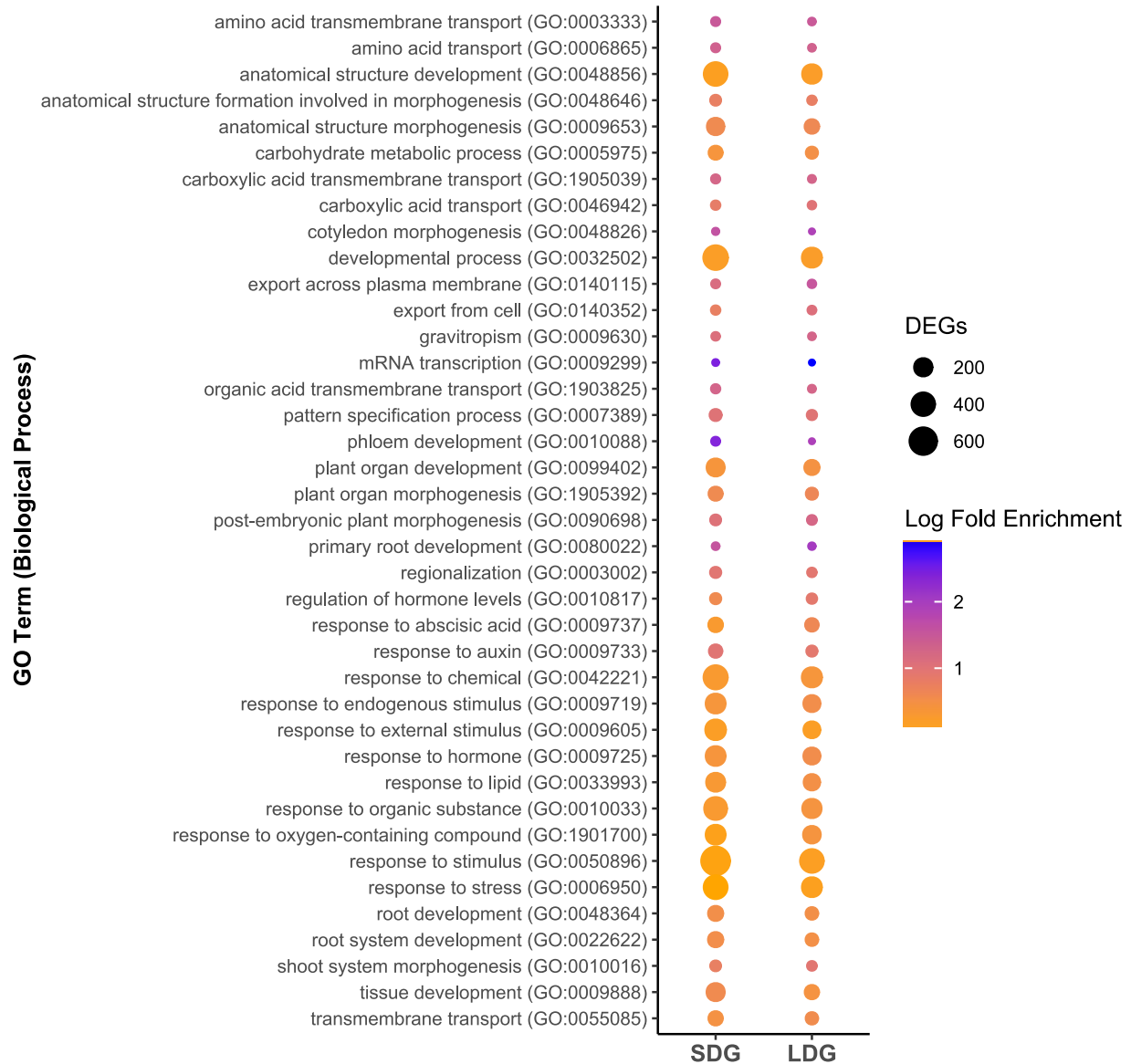


Figure 3.2. Biological Process gene ontology (GO) terms enriched ($P < 0.05$) in both genotypes in differentially expressed genes (DEGs) upregulated in domatia tissue.

DEGs were also enriched for 7 and 8 Cellular Component (CC) GO terms for SDG and LDG, respectively, with 7 of these shared between the genotypes (Figure S3.4), including components in the cell wall, plasma membrane, and regions outside the cell. A total of 23 and 24 Molecular Function (MF) GO terms were enriched for upregulated genes in SDG and LDG, respectively. However, only 12 MF GO terms were enriched in both genotypes (Figure S3.5).

While the MF GO terms enriched were diverse in function, in general, transporter activity and nucleic acid binding were commonly enriched.

Regulators of trichome development are upregulated in domatia

We identified multiple TFs upregulated in domatia that have been shown to regulate trichome initiation, including C2H2 ZFPs and SQUAMOSA promoter-binding protein-like (SPLs) (Zhou et al., 2013; Yue et al., 2018; Zhang et al., 2020; Han et al., 2022). Five genes encoding C2H2 ZFPs, orthologous to *AtNTT* (two *V. riparia* genes), *AtZFP1*, *AtZFP4*, and *AtZFP6*, were upregulated in domatia in both *V. riparia* genotypes. We also found that the *V. riparia* genes orthologous to *AtSPL13A* and *AtSPL12* are upregulated specifically in domatia of one genotype for SDG and LDG, respectively (Figure S3.6).

Cell wall gene expression is upregulated in domatia

As domatia formation requires both the development of trichomes and the depression in the leaf lamina, we hypothesized that cell wall modification genes would be upregulated in domatia. We found that approximately 7% of genes upregulated in domatia are involved in biosynthetic pathways for cell wall components, predominantly hemicelluloses (namely xylan and xyloglucan), pectin, and lignin (Figure 3.3), with upregulated genes in SDG domatia enriched for the biosynthetic processes of all three (Table S3.6).

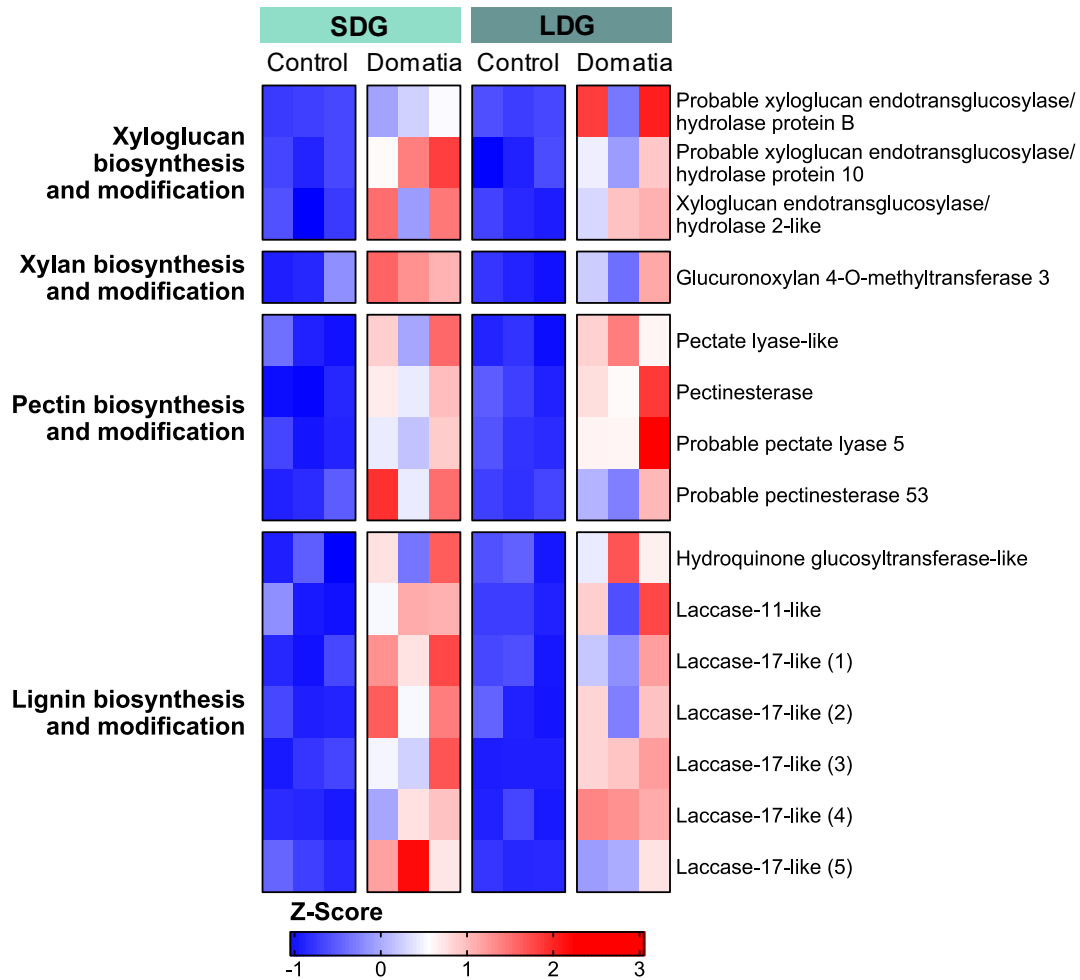


Figure 3.3. A heatmap of genes upregulated in domatia of both genotypes that are involved in cell wall biosynthesis and modification. Each column represents a biological replicate. The cells are colored by Z-score, with blue representing lower expression and red representing higher expression.

These include many genes that do not overlap between genotypes but are within the same gene families, including two different *V. riparia* orthologs to *AtPARVUS*, a key gene in xylan biosynthesis (Lee et al., 2007).

Upregulated genes in domatia mediate interactions with biotic organisms

As identified in tuber domatium (Pu et al., 2021), many genes upregulated in mite domatia are involved in direct defense responses against pathogens. Seventeen genes in SDG and twelve genes in LDG upregulated in domatia are from the *NBS-LRR* family, which are

generally involved in pathogen detection (DeYoung & Innes, 2006), with six genes shared between genotypes (Figure 3.4).

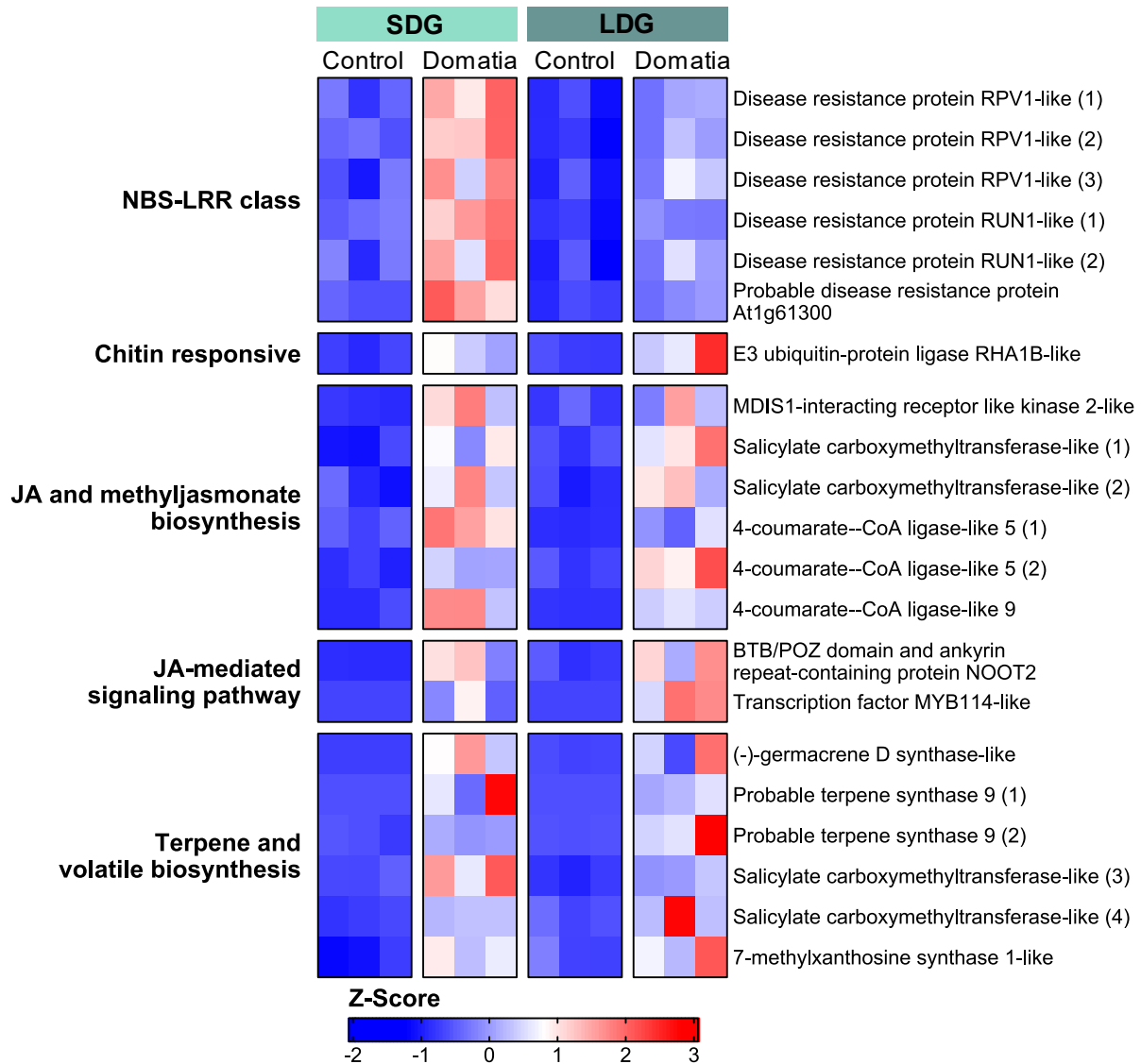


Figure 3.4. A heatmap of genes upregulated in domatia of both genotypes that are involved in interactions with biotic organisms. Each column represents a biological replicate. The cells are colored by Z-score, with blue representing lower expression and red representing higher expression.

Beyond *NBS-LRR* genes, additional genes involved in pathogen and chitin sensing are upregulated in domatia, including genes encoding E3 ubiquitin-protein ligase RHA1B-like (Figure 3.4) and protein LYK5-like (in SDG domatia only).

The jasmonic acid (JA) signaling pathway was previously implicated in both tuber domatia (Pu et al., 2021) and trichome development (Han et al., 2022) and is upregulated in *V. riparia* domatia for both genotypes. Orthologs to *JA carboxyl methyltransferase (JMT)* in Arabidopsis (annotated as encoding salicylate carboxymethyltransferase-like), which plays a crucial role in JA signaling by catalyzing the formation of methyl jasmonate from JA (Seo et al., 2001), are upregulated in domatia, with SDG domatia having two and LDG domatia having three *JMT* orthologs upregulated. Other genes thought to be involved in JA biosynthesis are upregulated in domatia as well, including two genes encoding 4-coumarate--CoA ligase-like 5 proteins (orthologous to *AtOPCL1*) and one encoding 4-coumarate--CoA ligase-like 9 (orthologous to AT5G63380) (Figure 3.4).

Possibly because of JA signaling, genes involved in terpene and volatile synthesis are upregulated in domatia, including genes shown to mediate plant-arthropod interactions. These include the orthologs to *TPS03*, *TPS21*, and *TPS24* in Arabidopsis, all of which are involved in the synthesis of volatile compounds. In addition, two genes upregulated in domatia that encode Salicylate carboxymethyltransferase-like (3-4) are orthologous to the Arabidopsis gene encoding AtBSMT1, which is involved in the production of the volatile compound Methyl Salicylate (MeSa) (Figure 3.4).

Domatia development is likely regulated by auxin signaling

Auxin signaling has been implicated in regulating both trichome development (Han et al., 2022) and tuber domatia development (Pu et al., 2021) and seems to play a role in domatia in *V. riparia*, which have genes upregulated at multiple steps in the auxin signaling pathway (Figure 3.5).

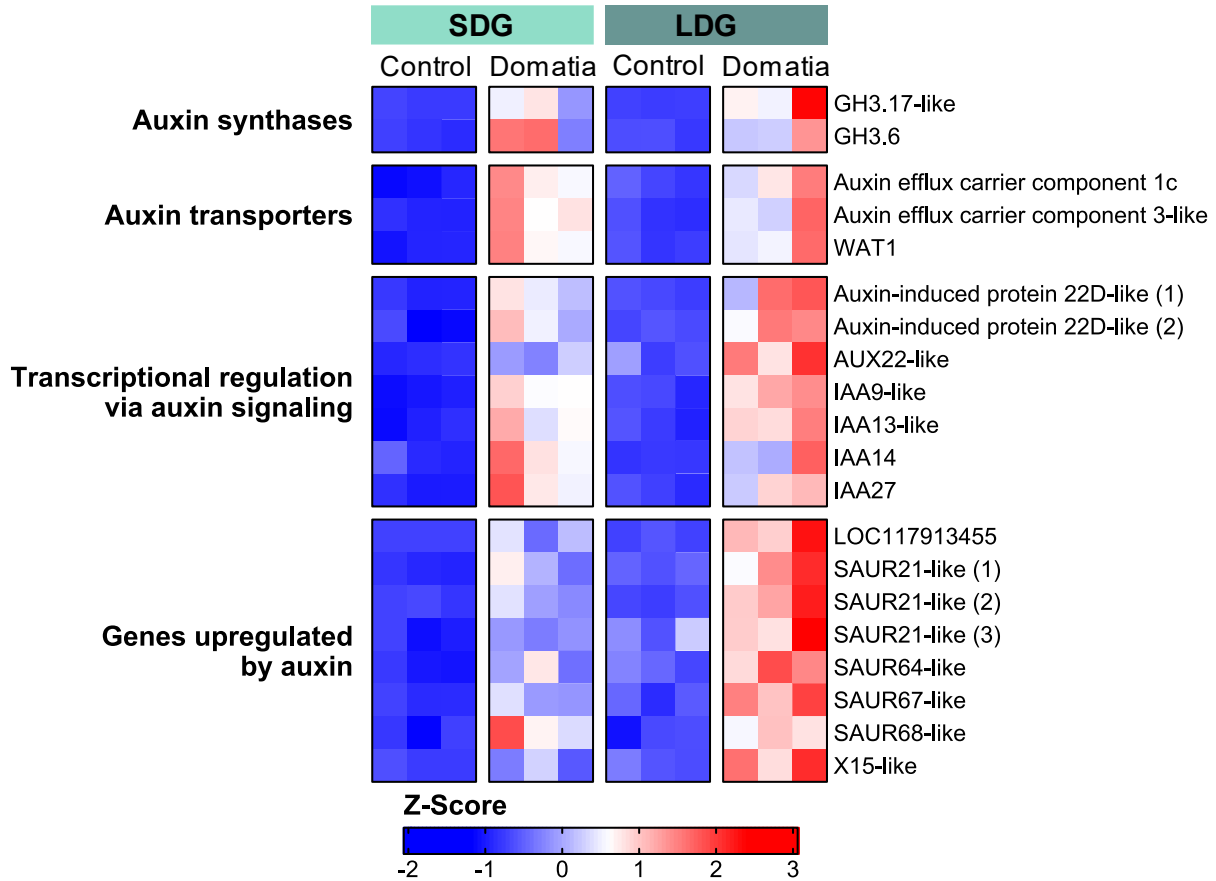


Figure 3.5. A heatmap of genes upregulated in domatia of both genotypes that are involved in auxin signaling. Each column represents a biological replicate. The cells are colored by Z-score, with blue representing lower expression and red representing higher expression.

Two genes encoding auxin synthases (GH3.6 and GH3.17) are upregulated, suggesting that auxin is actively produced during domatia development. In addition, three auxin transporters are upregulated in domatia in both genotypes, including genes encoding auxin efflux carrier components. Several genes involved in auxin responses downstream are upregulated in both genotypes as well, including six *auxin/indole-3-acetic acid* (*Aux/IAA*) genes, which are both involved in transcriptional regulation via auxin signaling (Ulmasov et al., 1997; Leyser, 2018), as well as two genes encoding auxin response factors (ARFs). Genes typically upregulated by auxin are also upregulated in domatia of both genotypes, including eight *small auxin up-regulated RNA* (*SAUR*) genes. Overall, the upregulation of auxin-related

genes suggests that auxin signaling plays a role in regulating domatia development.

Amino acid and carbohydrate transport and carbohydrate metabolism are upregulated in domatia

We observed multiple genes involved with amino acid transport, carbohydrate transport, and carbohydrate metabolism upregulated in domatia in both genotypes. Both genotypes were enriched for amino acid transport (GO:0006865) and amino acid transmembrane transport (GO:0003333), and several amino acid transporters are upregulated in domatia (Figure 3.6).

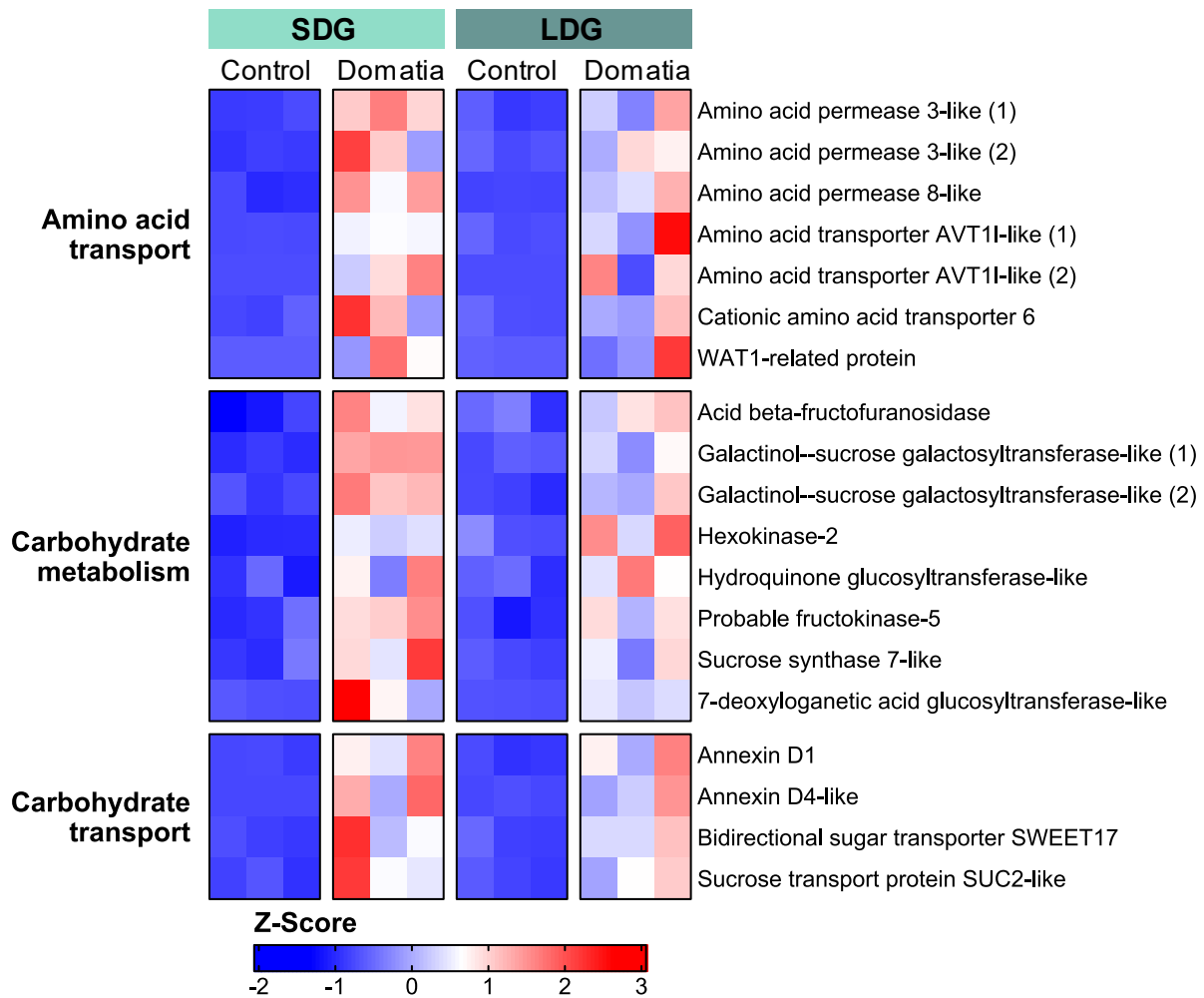


Figure 3.6. A heatmap of genes upregulated in domatia of both genotypes that are involved in the transport and metabolism of amino acids and carbohydrates. Each column

Figure 3.6 (cont'd)

represents a biological replicate. The cells are colored by Z-score, with blue representing lower expression and red representing higher expression.

Domatia also have higher expression of genes involved in carbohydrate metabolism and transport. While many of these genes overlap with those involved in hemicellulose biosynthesis, genes explicitly involved in general sugar metabolism and transport are also upregulated. We see evidence for the metabolism of starch, sucrose, and hexoses through the upregulation of the genes encoding products such as sucrose synthase 7-like, hexokinase-2, and fructokinase-5. Genes involved in the production of secondary metabolites specifically are upregulated in domatia, including two genes encoding galactinol--sucrose galactosyltransferase-like, orthologous to the raffinose synthase AtRS5. Additionally, genes encoding sugar transporters, such as SWEET17, a SUC2-like protein, and Annexin D1, are upregulated (Figure 3.6). Several of these genes are closely related to genes upregulated in the extrafloral nectaries (EFNs) of cotton (*Gossypium hirsutum*) (Chatt et al., 2021). Like domatia, EFNs are mutualistic structures produced by plants, but EFNs provide nectar (rather than housing) to arthropods in return for protection and so the reason for this overlap is unclear. As EFNs are homologous to floral nectaries in many angiosperms (Lee et al., 2005; Weber & Keeler, 2013), we hypothesized that domatia-upregulated genes could also be expressed in grapevine floral tissues and involved in floral development. Supporting this, upregulated genes in LDG domatia are enriched for flower morphogenesis (GO:0048439) (Table S3.7). Comparing the expression of *V. vinifera* orthologs to domatia-upregulated gene expression in floral, leaf, and stem tissue revealed that 98.6% of orthologs were expressed in floral tissue and identified 17 *V. vinifera* orthologs that demonstrated strong preferential expression in floral tissue compared to both leaf and stem tissue

(Supplementary Methods and Figure S3.7). Interestingly, one of the orthologs identified was the transcription factor *VviAGL6a*, which has been shown to play a role in grapevine floral development (Palumbo et al., 2019) and grapevine gall development from phylloxera infection (Schultz et al., 2019).

Intraspecific variation in domatia development

To understand intraspecific variation in domatia size, we identified genes where differences in expression levels between control and domatia tissue differed between the two genotypes. Nineteen genes showed significant genotype-tissue interactions (Figure 3.7).

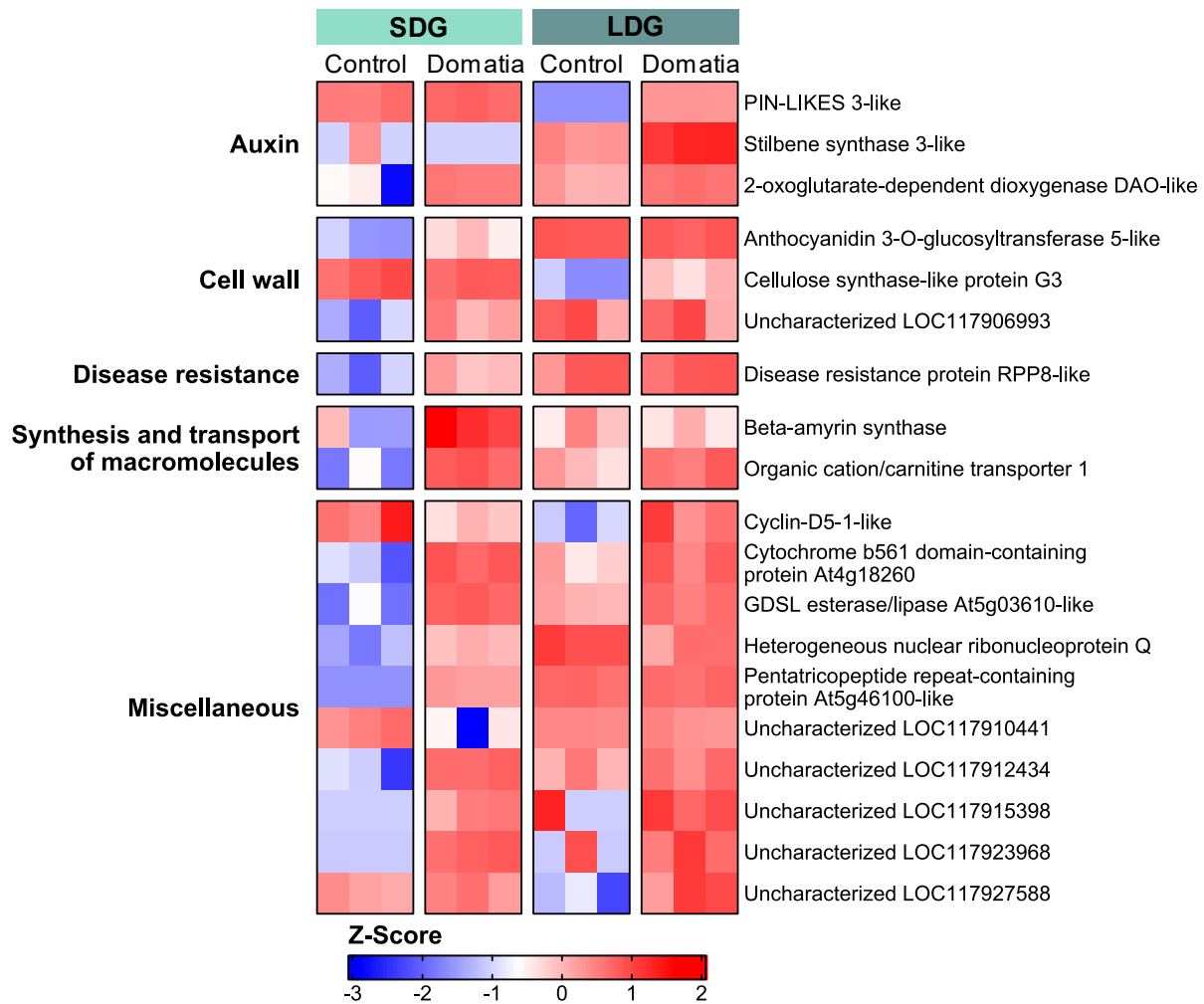


Figure 3.7. A heatmap of genes with significant domatia-genotype interactions between

Figure 3.7 (cont'd)

SDG and LDG. Each column represents a biological replicate. The cells are colored by Z-score, with blue representing lower expression and red representing higher expression.

Of these genes, 14 exhibited tissue-specific expression in SDG only, and five genes exhibited the opposite pattern with tissue-specific expression in LDG only. Nine of these genes are of unknown function, including LOC117912434 and LOC117927588, which are both orthologous to the Arabidopsis gene AT3G18670 and exhibit opposing patterns (the former being “domatia-responsive” in SDG and the latter in LDG) suggesting they may be functionally similar in the two genotypes.

Aside from pathways already implicated in domatia development, cell cycle regulation is implicated in intraspecific variation in domatia development due to the expression of the gene encoding Cyclin-D5-1-like, demonstrating genotype-tissue interactions. The Arabidopsis ortholog for this gene, *AtCYCD5;1*, is part of the D-type cyclin family implicated in regulating DNA replication, the cell cycle, and cellular differentiation (Dewitte et al., 2003; Schnittger et al., 2002).

The other ten genes that exhibited significant genotype-tissue interactions were involved in the processes mentioned above. The gene encoding 2-oxoglutarate-dependent dioxygenase DAO-like, which is orthologous to the gene *AtDAO1* involved in auxin oxidation and auxin-JA crosstalk (Lakehal et al., 2019), exhibited tissue-specific expression in SDG only. The gene encoding Anthocyanidin 3-O-glucosyltransferase 5-like, likely involved in the biosynthesis of the precursors of lignin, and the uncharacterized gene LOC117906993, orthologous to *AtWAK2*, both exhibit tissue-specific expression in SDG only. However, two auxin transporters, the genes encoding stilbene synthase 3-like and PIN-LIKES 3-like, exhibited tissue-specific expression in LDG only. The gene encoding a cellulose synthase-like

protein G3, likely involved in hemicellulose synthesis, exhibited tissue-specific expression in LDG only. Genes exhibiting genotype-tissue interactions involved in disease resistance and the synthesis or transport of molecules only exhibit tissue-specific expression in SDG.

Despite only 19 genes being differentially expressed between SDG domatia and LDG domatia, the two domatia genotypes differ greatly in both phenotype and transcripts upregulated in domatia (when compared to the control tissue). While there is overlap in the genes upregulated in domatia, a considerable number of genes (909 in SDG domatia and 221 in LDG domatia) are not shared when comparing expression between domatia and control tissue between the two genotypes. Of genes differentially expressed in domatia, 35-39% (459 genes in SDG and 227 in LDG) are differentially expressed between the control tissue of the two genotypes (Figure S3.3). Most of these genes (80.6-86.2%) exhibit a specific pattern where they: a) show lower expression levels in the control tissue of one genotype compared to the control leaf tissue of another genotype but are upregulated in domatia tissue of that particular genotype (in contrast to control tissue), while b) the other genotype demonstrates no difference in gene expression levels between control and domatia tissue. This suggests that differences in leaf tissue development may drive differences in domatia traits between genotypes of *V. riparia*. Genes that match this pattern in SDG domatia tissue are enriched for meristem and tissue development ($P < 0.05$), which supports this as well (Figure S3.8).

To better understand the relationship between domatia traits and leaf development, we landmarked leaves from SDG and LDG plants to see if they differed in overall leaf shape, which would demonstrate differences in leaf development between the two genotypes.

Landmarking leaves from SDG and LDG plants revealed that the two genotypes are

significantly different in leaf shape ($H = 2.96$, $P = 0.009$), with SDG leaves having narrower lateral and apical lobes and a narrower upper lateral sinus than LDG (Figure 3.8 and Figure S3.9).

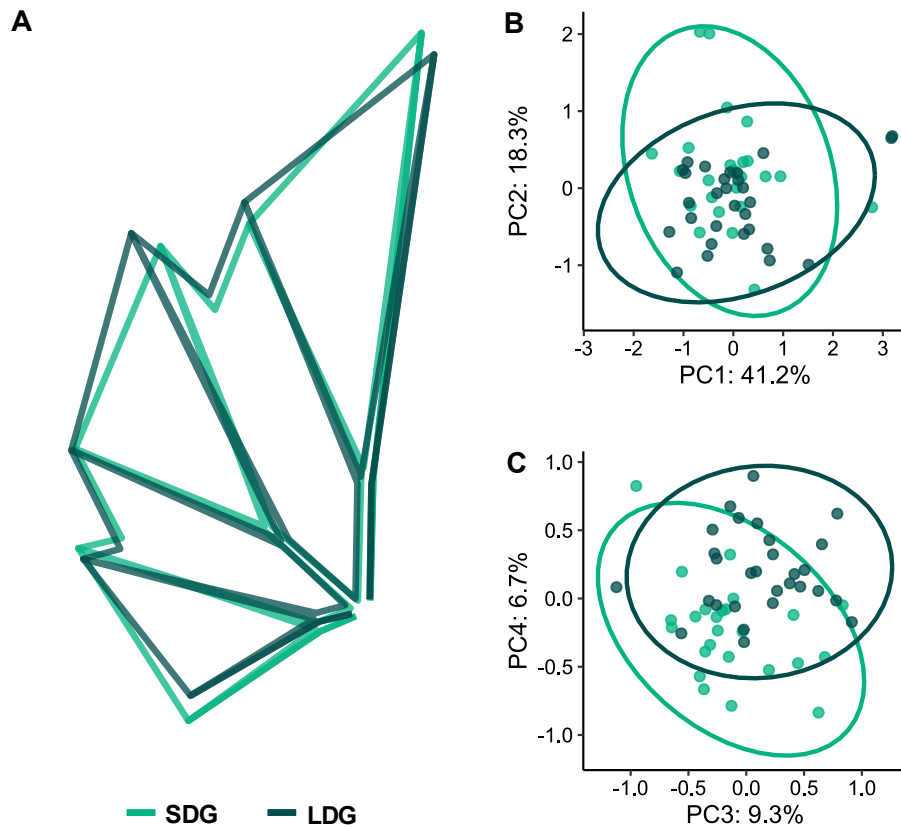


Figure 3.8. Differences in leaf shape between SDG and LDG. **(A)** Mean leaf shapes of SDG leaves (light green) and LDG leaves (dark green) rotated and scaled identically. **(B and C)** Principal component analysis (PCA) of leaf shapes, SDG in light green and LDG in dark green. **(B)** PCs 1 and 2, with PC1 contributing to 41.2% of variation and PC2 to 18.3%. **(C)** PCs 3 and 4, with PC3 contributing to 9.3% of variation and PC4 to 6.7%. In Figure S3.9, Eigenleaves display the morphological characteristics of each PC.

These differences in leaf shape reflect altered angles of vein axils on the leaves where domatia form, which could indirectly impact domatia development and influence domatia size. It is also possible that the genetic differences between genotypes drive differences in leaf shape and are directly responsible for differences in domatia size.

DISCUSSION

Understanding the genetic underpinnings of ecologically important traits is a central goal linking subfields in biology, yet the genetic bases of many ecologically relevant traits remain understudied. Here, we present the first transcriptomic study aimed at understanding the genetic drivers of the development of mite domatia, small structures on the undersides of plant leaves that mediate a powerful and pervasive mutualism between plants and beneficial mites. Several of the genes we identified overlap with genes previously implicated in domatia development in *Vitis*, including *V. riparia* genes encoding TFs thought to regulate trichome development (Barba et al., 2019; LaPlante et al., 2021), as well as Importin Alpha Isoform 1 and GATA Transcription Factor 8 (LaPlante et al., 2021). We found that genes related to domatia development are similar to those involved in trichome and tuber domatia development, including genes involved in trichome regulation, cell wall biosynthesis/modification, plant hormone signaling, and biotic responses. We also found that amino acid transport and carbohydrate metabolism/transport are implicated in the development of domatia. These findings provide insight into the genetic drivers and functioning of this important phenotype.

C2H2 ZFPs and SPLs are likely key regulators of trichome initiation in Vitis domatia

Trichomes have convergently evolved in numerous angiosperm lineages and are regulated by distinct pathways between species (Han et al., 2022; Serna & Martin, 2006), and the genetic pathways of trichome development in *V. riparia* have yet to be uncovered. Our findings, previous genetic research in *Vitis* (Barba et al., 2019; LaPlante et al., 2021), and studies of trichome development in other species (Han et al., 2022), all support C2H2 ZFPs and SPLs as probable key trichome regulators in *V. riparia* and related species. One of the

C2H2 ZFPs identified in previous studies in *Vitis* is orthologous to *AtZFP6*, whose *V. riparia* ortholog was upregulated in domatia in both genotypes in this study. This *V. vinifera* gene, VIT_205s0077g01390, was identified as a candidate gene for hair on leaf blades (Barba et al., 2019) and domatia density (LaPlante et al., 2021). Two additional ZFPs were identified as genetic candidates for leaf trichome traits by Barba et al. (2019). Two *SPLs* were also upregulated in the domatia we sequenced, and Barba et al. (2019) linked the *V. vinifera* gene VIT_15s0021g02290, orthologous to *AtSPL8* in Arabidopsis, to domatia size. The overlap between C2H2 ZFPs and *SPLs* in our dataset, as well as previous quantitative genetic studies in *Vitis* (Barba et al., 2019; LaPlante et al., 2021), suggest that these TFs may play an essential role in regulating trichome and/or domatia development in *Vitis*.

Insights into Vitis domatia cell wall biosynthesis and composition

Our findings provided insight into the biosynthesis and composition of cell walls in domatia. While our domatia samples include both laminar tissue and trichomes, we expect trichomes to have vastly different cell wall composition compared to laminar tissue. Accordingly, DEGs in domatia involved in cell wall biosynthesis provide insight into the composition of trichome cell walls. Gene pathways involved in xyloglucan, xylan, pectin, and lignin biosynthesis were upregulated in domatia tissue. While xyloglucan, pectin, and lignin are fairly common components of cell walls in both normal leaf tissue and trichomes (Bowling et al., 2011; Marks et al., 2008), xylan is not (Bowling et al., 2008, 2011). However, xylan is important for general protection against herbivores and pathogens (Gao et al., 2017; Joo et al., 2021). A previous study investigating loci associated with downy mildew resistance in grapevine also identified a candidate gene involved in xylan biosynthesis (Divilov et al., 2018). It is possible that xylan production in domatia trichomes in *Vitis* enables protection

against downy mildew either directly or indirectly through facilitating mite mutualisms. It is also possible that the upregulation of xylan-related genes is due to the closeness of vascular tissue to domatia and trichomes (Gago et al., 2016; Tilney et al., 2012), which typically have large amounts of xylan (Moore et al., 2014). Future work characterizing the composition of cell walls in domatia trichomes would clarify the specific role of these genes in domatia.

Auxin and JA mediate the development of domatia in V. riparia

We found that auxin and JA genes are heavily upregulated during domatia development. The upregulation of auxin genes is unsurprising, as auxin plays significant roles in cell elongation, leaf trichome development (Xuan et al., 2020) and tuber domatium development (Pu et al., 2021). However, to our knowledge, a connection between JA and domatia has not been studied or shown. Notably, JA is implicated in other plant-arthropod mutualisms, increasing nectar secretion in EFNs (Heil et al., 2001, 2004; Hernandez-Cumplido et al., 2016; Kost & Heil, 2008). JA may mediate plant-mite interactions in some cases, similar to how JA mediates ant-plant interactions in some EFN-bearing species (Heil et al., 2001, 2004; Hernandez-Cumplido et al., 2016; Kost & Heil, 2008). JA could induce the structural development of domatia by regulating trichome development, as shown in a few other species (Han et al., 2022). It could also regulate the release of plant volatiles, as has been shown in other systems (Ament et al., 2004; Degenhardt et al., 2010; Schmelz et al., 2003), and these volatiles could mediate mutualisms with mites through signaling or provide another layer of direct defense (Baldwin, 2010). Alternatively, JA could provide direct defense against bacteria and fungi growing on mite waste within domatia (excrement, exoskeletons, etc.). Future work investigating the impact of JA application on *V. riparia* could clarify JA's role in domatia.

Insights into mite-plant mutualisms in Vitis

Our findings provide insight into possible ways *V. riparia* domatia mediate mite mutualisms. We saw evidence for volatile production through the expression of genes involved in terpenoid synthesis. This includes the ortholog to TERPENE SYNTHASE 21 (AtTPS21) in Arabidopsis. AtTPS21 is involved in the production of (*E*)- β -caryophyllene (Chen et al., 2003a), which mediates both direct defense against microbial pathogens (Cowan, 1999; Huang et al., 2012) and indirect defense against herbivores by attracting natural enemies (Köllner et al., 2008; Rasmann et al., 2005). It is possible that this volatile attracts mites to domatia or is a direct defense to inhibit pathogen growth within domatia.

We also see evidence for methyl salicylate (MeSa) emission through the upregulation of two genes encoding salicylate carboxymethyltransferase-like and one encoding 7-methylxanthosine synthase 1-like, all of which are orthologous to *AtBSMT1* which is responsible for MeSa production (Chen et al., 2003b). MeSa is a common plant volatile typically released after herbivory (Chen et al., 2003b; Snoeren et al., 2010) that repels herbivores (Koschier et al., 2007; Ulland et al., 2008) and attracts predators (De Boer & Dicke, 2004; James & Price, 2004; Mallinger et al., 2011). In grapevine, MeSa attracts the predaceous mite *Typhlodromus pyri* (Gadino et al., 2012) that inhabits leaves (English-Loeb et al., 2002). The upregulation of the *V. riparia* orthologs to *AtBSMT1* suggests that MeSa production and emission may occur in domatia, which could attract predatory mites. Due to their small size, it is challenging to capture domatia-specific volatiles. However, future work investigating volatile emissions from domatia could test hypotheses surrounding the mediation of domatia inhabitancy through volatiles.

Gene expression patterns in domatia share similarities with EFNs

Several genes we identified involved in macromolecule biosynthesis and transport are closely related to genes involved in development and nectar production in EFNs (Chatt et al., 2021; Roy et al., 2017). We found that upregulated genes in LDG domatia are enriched for flower morphogenesis, and *V. vinifera* orthologs of genes upregulated in domatia were expressed in floral tissue compared to leaf and stem tissue. Further, one of the orthologs identified was also related to grapevine gall development from phylloxera infection (Schultz et al., 2019), suggesting that floral pathways have been co-opted in different ways to enable the development of diverse structures like EFNs, galls, and domatia. Understanding the overlap of domatia genes with genes involved in EFN, gall, and floral development may provide insight into potential pathways modified to enable the evolution of plant structures that mediate mutualisms.

The overlap between genes upregulated in domatia and EFNs could also be due to functional similarities. To our knowledge, no studies have suggested that *V. riparia* domatia produce secretions for beneficial mites. However, nectar applied to *V. riparia* leaves increased mite recruitment (Weber et al., 2016), and there is evidence of material exchange between mites and domatia in *Plectroniella armata* (Tilney et al., 2012). The considerable upregulation of sugar and amino acid transport genes (Figure 3.6) and the upregulation of many genes involved in floral development (Figure S3.7) suggests the possibility that these phenomena could be due to material exchange from the plants to the mites and begs future studies investigating the detailed morphologies and functions of *V. riparia* domatia. Alternatively, like EFNs (Chatt et al., 2021), domatia and grapevine trichomes tend to be located near vascular tissue bundles (Gago et al., 2016; Tilney et al., 2012), so macromolecule

biosynthesis and transport gene upregulation may be due to an abundance of vascular tissue in domatia samples.

Intraspecific variation in domatia size may be due to differences in leaf development

Despite varying substantially in domatia size (Figure 3.1C), with LDG domatia nearly two times larger than SDG domatia, we only identified 19 genes differentially expressed between SDG and LDG domatia. However, many genes involved in domatia development overlap with genes differentially expressed between SDG and LDG control tissue (35-39%) (Figure S3.3). Further, the two genotypes varied in domatia traits and overall leaf shape (Figure 3.8 and Figure S3.9). Thus, differences in overall leaf development may shape domatia traits in *V. riparia*. Previous studies in *Vitis* and *Begonia* found a relationship between leaf morphology and leaf trichomes (McLellan, 2005; Chitwood et al., 2014), and Barba et al. (2019) also suggest a link between leaf morphology and trichome development in *Vitis*. Future work investigating domatia and leaf development in *Vitis* together could unravel the molecular genetic mechanisms and developmental processes enabling the potential link between leaf morphology and domatia.

ACKNOWLEDGEMENTS

We would like to thank Grace Fleming, Bruce Martin, Erika LaPlante, Andrew Myers, and other members of the Weber lab for helpful conversations surrounding these research ideas. We are grateful to Dan Chitwood, Emily Josephs, and Robin Buell for helpful discussions on this work and feedback on this manuscript. We thank David Lowry and his research group for feedback on this manuscript as well. We are grateful to the Genomics Core at Michigan State University and the Institute for Cyber-Enabled Research at Michigan State University for their services. We are also grateful to the staff of the Research Greenhouse

Complex at Michigan State University and the staff of the research greenhouses at the University of Michigan for their assistance with plant growth. This work was supported by NSF DEB-1831164 (awarded to MGW). EJR was supported by the University Distinguished Fellowship at Michigan State University.

REFERENCES

- Agrawal A. A., Karban R. (1997). Domatia mediate plant-arthropod mutualism. *Nature*, 387, 562–563.
- Alexa A., Rahnenfuhrer J. (2023). *topGO: Enrichment Analysis for Gene Ontology*. <https://bioconductor.org/packages/topGO>. Accessed 10 March 2024.
- Ament K., Kant M. R., Sabelis M. W., Haring M. A., Schuurink R. C. (2004). Jasmonic acid is a key regulator of spider mite-induced volatile terpenoid and methyl salicylate emission in tomato. *Plant Physiology*, 135, 2025–2037.
- Andrews S. (2010). *FastQC: a quality control tool for high throughput sequence data*. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*, 25(1), 25–29.
- Baldwin I. T. (2010). Plant volatiles. *Current Biology*, 20, R392–R397.
- Barba P., Loughner R., Wentworth K., Nyrop J. P., Loeb G. M., Reisch B. I. (2019). A QTL associated with leaf trichome traits has a major influence on the abundance of the predatory mite *Typhlodromus pyri* in a hybrid grapevine population. *Horticulture Research*, 6, 87.
- Berardini T. Z., Reiser L., Li D., Mezheritsky Y., Muller R., Strait E., Huala E. (2015). The Arabidopsis information resource: Making and mining the ‘gold standard’ annotated reference plant genome. *Genesis*, 53, 474–485.
- Blattner F. R., Weising K., Bänfer G., Maschwitz U., Fiala B. (2001). Molecular analysis of phylogenetic relationships among myrmecophytic *Macaranga* species (Euphorbiaceae). *Molecular Phylogenetics and Evolution*, 19, 331–344.
- Bolger A. M., Lohse M., Usadel B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, 2114–2120.
- Bowling A. J., Maxwell H. B., Vaughn K. C. (2008). Unusual trichome structure and composition in mericarps of catchweed bedstraw (*Galium aparine*). *Protoplasma*, 233, 223–230.
- Bowling A. J., Vaughn K. C., Turley R. B. (2011). Polysaccharide and glycoprotein distribution in the epidermis of cotton ovules during early fiber initiation and growth. *Protoplasma*, 248, 579–590.

- Bronstein J. L., Alarcón R., Geber M. (2006). The evolution of plant-insect mutualisms. *The New Phytologist*, 172, 412–428.
- Bryson, A. E., Wilson Brown, M., Mullins, J., Dong, W., Bahmani, K., Bornowski, N., Chiu, C., Engelgau, P., Gettings, B., Gomezcano, F., Gregory, L. M., Haber, A. C., Hoh, D., Jennings, E. E., Ji, Z., Kaur, P., Kenchanmane Raju, S. K., Long, Y., Lotreck, S. G., ... Chitwood, D. H. (2020). Composite modeling of leaf shape along shoots discriminates *Vitis* species better than individual leaves. *Applications in Plant Sciences*, 8(12), e11404.
- Buchfink B., Xie C., Huson D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 12, 59–60.
- Chatt E. C., Mahalim S.-N., Mohd-Fadzil N.-A., Roy R., Klinkenberg P. M., Horner H. T., Hampton M., Carter C. J., Nikolau B. J. (2021). Nectar biosynthesis is conserved among floral and extrafloral nectaries. *Plant Physiology*, 185, 1595–1616.
- Chen F., Tholl D., D’Auria J. C., Farooq A., Pichersky E., Gershenzon J. (2003a). Biosynthesis and emission of terpenoid volatiles from *Arabidopsis* flowers. *The Plant Cell*, 15, 481–494.
- Chen F., D’Auria J. C., Tholl D., Ross J. R., Gershenzon J., Noel J. P., Pichersky E. (2003b). An *Arabidopsis thaliana* gene for methylsalicylate biosynthesis, identified by a biochemical genomics approach, has a role in defense. *The Plant Journal*, 36, 577–588.
- Chitwood, D. H., Ranjan, A., Martinez, C. C., Headland, L. R., Thiem, T., Kumar, R., Covington, M. F., Hatcher, T., Naylor, D. T., Zimmerman, S., Downs, N., Raymundo, N., Buckler, E. S., Maloof, J. N., Aradhya, M., Prins, B., Li, L., Myles, S., & Sinha, N. R. (2014). A modern ampelography: a genetic basis for leaf shape and venation patterning in grape. *Plant Physiology*, 164(1), 259–272.
- Constantin A. E., Patil I. (2021). ggsignif: R package for displaying significance brackets for ‘ggplot2.’ *PsyArxiv*.
- Cowan M. M. (1999). Plant products as antimicrobial agents. *Clinical Microbiology Reviews*, 12, 564–582.
- De Boer J. G., Dicke M. (2004). The role of methyl salicylate in prey searching behavior of the predatory mite *Phytoseiulus persimilis*. *Journal of Chemical Ecology*, 30, 255–271.
- Degenhardt D. C., Refi-Hind S., Stratmann J. W., Lincoln D. E. (2010). Systemin and jasmonic acid regulate constitutive and herbivore-induced systemic volatile emissions in tomato, *Solanum lycopersicum*. *Phytochemistry*, 71, 2024–2037.
- Dewitte W., Riou-Khamlichi C., Scofield S., Healy J. M. S., Jacqmard A., Kilby N. J., Murray J. A. H. (2003). Altered cell cycle distribution, hyperplasia, and inhibited differentiation

- in Arabidopsis caused by the D-type cyclin CYCD3. *The Plant Cell*, 15, 79–92.
- DeYoung B. J., Innes R. W. (2006). Plant NBS-LRR proteins in pathogen sensing and host defense. *Nature Immunology*, 7, 1243–1249.
- Divilov K., Barba P., Cadle-Davidson L., Reisch B. I. (2018). Single and multiple phenotype QTL analyses of downy mildew resistance in interspecific grapevines. *Theoretical and Applied Genetics*, 131, 1133–1143.
- Dobin A., Davis C. A., Schlesinger F., Drenkow J., Zaleski C., Jha S., Batut P., Chaisson M., Gingeras T. R. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29, 15–21.
- Dong, M., Xue, S., Bartholomew, E. S., Zhai, X., Sun, L., Xu, S., Zhang, Y., Yin, S., Ma, W., Chen, S., Feng, Z., Geng, C., Li, X., Liu, X., & Ren, H. (2022). Transcriptomic and functional analysis provides molecular insights into multicellular trichome development. *Plant Physiology*, 189(1), 301–314.
- Dryden I. L., Mardia K. V. (2016). *Statistical Shape Analysis: With Applications in R*. John Wiley & Sons.
- English-Loeb G., Norton A. (2006). Lack of trade-off between direct and indirect defence against grape powdery mildew in riverbank grape. *Ecological Entomology*, 31, 415–422.
- English-Loeb G., Norton A. P., Walker M. A. (2002). Behavioral and population consequences of acarodomatia in grapes on phytoseiid mites (Mesostigmata) and implications for plant breeding. *Entomologia Experimentalis et Applicata*, 104, 307–319.
- Faraji F., Janssen A., Sabelis M. W. (2002a). The benefits of clustering eggs: the role of egg predation and larval cannibalism in a predatory mite. *Oecologia*, 131, 20–26.
- Faraji F., Janssen A., Sabelis M. W. (2002b). Oviposition patterns in a predatory mite reduce the risk of egg predation caused by prey. *Ecological Entomology*, 27, 660–664.
- Fasoli, M., Dal Santo, S., Zenoni, S., Tornielli, G. B., Farina, L., Zamboni, A., Porceddu, A., Venturini, L., Bicego, M., Murino, V., Ferrarini, A., Delledonne, M., & Pezzotti, M. (2012). The grapevine expression atlas reveals a deep transcriptome shift driving the entire plant into a maturation program. *The Plant Cell*, 24(9), 3489–3505.
- Gadino A. N., Walton V. M., Lee J. C. (2012). Olfactory response of *Typhlodromus pyri* (Acari: Phytoseiidae) to synthetic methyl salicylate in laboratory bioassays. *Journal of Applied Entomology*, 136, 476–480.
- Gago P., Conéjéro G., Martínez M. C., Boso S., This P., Verdeil J.-L. (2016). Microanatomy of leaf trichomes: opportunities for improved ampelographic discrimination of

- grapevine (*Vitis vinifera* L.) cultivars. *Australian Journal of Grape and Wine Research*, 22, 494–503.
- Gao, Y., He, C., Zhang, D., Liu, X., Xu, Z., Tian, Y., Liu, X.-H., Zang, S., Pauly, M., Zhou, Y., & Zhang, B. (2017). Two Trichome Birefringence-Like Proteins Mediate Xylan Acetylation, Which Is Essential for Leaf Blight Resistance in Rice. *Plant Physiology*, 173(1), 470–481.
- Gene Ontology Consortium. (2023). The Gene Ontology knowledgebase in 2023. *Genetics*, 224.
- Girollet N., Rubio B., Lopez-Roques C., Valière S., Ollat N., Bert P.-F. (2019). *De novo* phased assembly of the *Vitis riparia* grape genome. *Scientific Data*, 6, 1–8.
- Graham C. D. K., Forrestel E. J., Schillmiller A. L., Zemenick A. T., Weber M. G. (2023). Evolutionary signatures of a trade-off in direct and indirect defenses across the wild grape genus, *Vitis*. *Evolution*, 77, 2301–2313.
- Grostal R., O’Dowd D. J. (1994). Plants, mites and mutualism: leaf domatia and the abundance and reproduction of mites on *Viburnum tinus* (Caprifoliaceae). *Oecologia*, 97, 308–315.
- Gu Z. (2022). Complex heatmap visualization. *iMeta*, 1, e43.
- Han G., Li Y., Yang Z., Wang C., Zhang Y., Wang B. (2022). Molecular mechanisms of plant trichome development. *Frontiers in Plant Science*, 13, 910228.
- Heil M., Greiner S., Meimberg H., Krüger R., Noyer J.-L., Heubl G., Linsenmair K. E., Boland W. (2004). Evolutionary change from induced to constitutive expression of an indirect plant resistance. *Nature*, 430, 205–208.
- Heil M., Koch T., Hilpert A., Fiala B., Boland W., Linsenmair K. (2001). Extrafloral nectar production of the ant-associated plant, *Macaranga tanarius*, is an induced, indirect, defensive response elicited by jasmonic acid. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 1083–1088.
- Hernandez-Cumplido J., Forter B., Moreira X., Heil M., Benrey B. (2016). Induced floral and extrafloral nectar production affect ant-pollinator interactions and plant fitness. *Biotropica*, 48, 342–348.
- Huang M., Sanchez-Moreiras A. M., Abel C., Sohrabi R., Lee S., Gershenzon J., Tholl D. (2012). The major volatile organic compound emitted from *Arabidopsis thaliana* flowers, the sesquiterpene (*E*)- β -caryophyllene, is a defense against a bacterial pathogen. *The New Phytologist*, 193, 997–1008.
- Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., Bravo, H. C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K. D., Irizarry, R. A.,

- Lawrence, M., Love, M. I., MacDonald, J., Obenchain, V., Oleś, A. K., ... Morgan, M. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods*, 12(2), 115–121.
- IPGRI, UPOV, and OIV. (1997). *Descriptors for Grapevine (Vitis Spp.)*. International Union for the Protection of New Varieties of Plants, Geneva, Switzerland/Office International de la Vigne et du Vin, Paris, France/International Plant Genetic Resources Institute.
- James D. G., Price T. S. (2004). Field-testing of methyl salicylate for recruitment and retention of beneficial insects in grapes and hops. *Journal of Chemical Ecology*, 30, 1613–1628.
- 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. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30(9), 1236–1240.
- Joo, Y., Kim, H., Kang, M., Lee, G., Choung, S., Kaur, H., Oh, S., Choi, J. W., Ralph, J., Baldwin, I. T., & Kim, S.-G. (2021). Pith-specific lignification in *Nicotiana attenuata* as a defense against a stem-boring herbivore. *The New Phytologist*, 232(1), 332–344.
- Köllner T. G., Held M., Lenk C., Hiltbold I., Turlings T. C. J., Gershenzon J., Degenhardt J. (2008). A maize (E)- β -caryophyllene synthase implicated in indirect defense responses against herbivores is not expressed in most American maize varieties. *The Plant Cell*, 20, 482–494.
- Koschier E. H., Hoffmann D., Riefler J. (2007). Influence of salicylaldehyde and methyl salicylate on post-landing behaviour of *Frankliniella occidentalis* Pergande. *Journal of Applied Entomology*, 131, 362–367.
- Kost C., Heil M. (2008). The defensive role of volatile emission and extrafloral nectar secretion for lima bean in nature. *Journal of Chemical Ecology*, 34, 1–13.
- Lakehal A., Dob A., Novák O., Bellini C. (2019). A DAO1-mediated circuit controls auxin and jasmonate crosstalk robustness during adventitious root initiation in *Arabidopsis*. *International Journal of Molecular Sciences*, 20.
- LaPlante E. R., Fleming M. B., Migicovsky Z., Weber M. G. (2021). Genome-wide association study reveals genomic region associated with mite-recruitment phenotypes in the domesticated grapevine (*Vitis vinifera*). *Genes*, 12.
- Lee J.-Y., Baum S. F., Oh S.-H., Jiang C.-Z., Chen J.-C., Bowman J. L. (2005). Recruitment of *CRABS CLAW* to promote nectary development within the eudicot clade. *Development*, 132, 5021–5032.
- Lee C., Zhong R., Richardson E. A., Himmelsbach D. S., McPhail B. T., Ye Z.-H. (2007). The *PARVUS* gene is expressed in cells undergoing secondary wall thickening and is

- essential for glucuronoxyylan biosynthesis. *Plant & Cell Physiology*, 48, 1659–1672.
- Leyser O. (2018). Auxin signaling. *Plant Physiology*, 176, 465–479.
- Love M. I., Huber W., Anders S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15, 1–21.
- Mallinger R. E., Hogg D. B., Gratton C. (2011). Methyl salicylate attracts natural enemies and reduces populations of soybean aphids (Hemiptera: Aphididae) in soybean agroecosystems. *Journal of Economic Entomology*, 104, 115–124.
- Marks M. D., Betancur L., Gilding E., Chen F., Bauer S., Wenger J. P., Dixon R. A., Haigler C. H. (2008). A new method for isolating large quantities of Arabidopsis trichomes for transcriptome, cell wall and other types of analyses. *The Plant Journal*, 56, 483–492.
- McLellan T. (2005). Correlated evolution of leaf shape and trichomes in *Begonia dregei* (Begoniaceae). *American Journal of Botany*, 92, 1616–1623.
- Moore J. P., Nguema-Ona E., Fangel J. U., Willats W. G. T., Hugo A., Vivier M. A. (2014). Profiling the main cell wall polysaccharides of grapevine leaves using high-throughput and fractionation methods. *Carbohydrate Polymers*, 99, 190–198.
- Norton A. P., English-Loeb G., Belden E. (2001). Host plant manipulation of natural enemies: leaf domatia protect beneficial mites from insect predators. *Oecologia*, 126, 535–542.
- Norton A. P., English-Loeb G., Gadoury D., Seem R. C. (2000). Mycophagous mites and foliar pathogens: Leaf domatia mediate tritrophic interactions in grapes. *Ecology*, 81, 490–499.
- O’Dowd D., Pemberton R. (1998). Leaf domatia and foliar mite abundance in broadleaf deciduous forest of north Asia. *American Journal of Botany*, 85, 70.
- Palumbo F., Vannozzi A., Magon G., Lucchin M., Barcaccia G. (2019). Genomics of flower identity in grapevine (*Vitis vinifera* l.). *Frontiers in Plant Science*, 10, 316.
- Pu Y., Naikatini A., Pérez-Escobar O. A., Silber M., Renner S. S., Chomicki G. (2021). Genome-wide transcriptome signatures of ant-farmed *Squamellaria* epiphytes reveal key functions in a unique symbiosis. *Ecology and Evolution*, 11, 15882–15895.
- Rasmann S., Köllner T. G., Degenhardt J., Hiltbold I., Toepfer S., Kuhlmann U., Gershenzon J., Turlings T. C. J. (2005). Recruitment of entomopathogenic nematodes by insect-damaged maize roots. *Nature*, 434, 732–737.
- R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Ritter E. J., Cousins P., Quigley M., Kile A., Kenchanmane Raju S. K., Chitwood D. H.,

- Niederhuth C. (2023). From buds to shoots: Insights into grapevine development from the Witch's Broom bud sport. *bioRxiv*: 2023.09.25.559343.
- Romero G. Q., Benson W. W. (2004). Leaf domatia mediate mutualism between mites and a tropical tree. *Oecologia*, *140*, 609–616.
- Romero G. Q., Benson W. W. (2005). Biotic interactions of mites, plants and leaf domatia. *Current Opinion in Plant Biology*, *8*, 436–440.
- Roy R., Schmitt A. J., Thomas J. B., Carter C. J. (2017). Review: Nectar biology: From molecules to ecosystems. *Plant Science*, *262*, 148–164.
- RStudio Team. (2022). *RStudio: Integrated development environment for R*. Posit software. <https://posit.co/products/open-source/rstudio/>.
- Schmelz E. A., Alborn H. T., Tumlinson J. H. (2003). Synergistic interactions between volicitin, jasmonic acid and ethylene mediate insect-induced volatile emission in *Zea mays*. *Physiologia Plantarum*, *117*, 403–412.
- Schneider C. A., Rasband W. S., Eliceiri K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*, 671–675.
- Schnittger A., Schöbinger U., Bouyer D., Weinl C., Stierhof Y.-D., Hülskamp M. (2002). Ectopic D-type cyclin expression induces not only DNA replication but also cell division in *Arabidopsis* trichomes. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 6410–6415.
- Schultz J. C., Edger P. P., Body M. J. A., Appel H. M. (2019). A galling insect activates plant reproductive programs during gall development. *Scientific Reports*, *9*, 1833.
- Seo H. S., Song J. T., Cheong J. J., Lee Y. H., Lee Y. W., Hwang I., Lee J. S., Choi Y. D. (2001). Jasmonic acid carboxyl methyltransferase: a key enzyme for jasmonate-regulated plant responses. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 4788–4793.
- Serna L., Martin C. (2006). Trichomes: different regulatory networks lead to convergent structures. *Trends in Plant Science*, *11*, 274–280.
- Snoeren T. A. L., Kappers I. F., Broekgaarden C., Mumm R., Dicke M., Bouwmeester H. J. (2010). Natural variation in herbivore-induced volatiles in *Arabidopsis thaliana*. *Journal of Experimental Botany*, *61*, 3041–3056.
- Tilney P. M., van Wyk A. E., van der Merwe C. F. (2012). Structural evidence in *Plectroniella armata* (Rubiaceae) for possible material exchange between domatia and mites. *PloS One*, *7*, e39984.
- Ulland S., Ian E., Mozuraitis R., Borg-Karlson A.-K., Meadow R., Mustaparta H. (2008).

- Methyl salicylate, identified as primary odorant of a specific receptor neuron type, inhibits oviposition by the moth *Mamestra brassicae* L. (Lepidoptera, Noctuidae). *Chemical Senses*, 33, 35–46.
- Ulmasov T., Murfett J., Hagen G., Guilfoyle T. J. (1997). Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements. *The Plant Cell*, 9, 1963–1971.
- Weber M. G., Keeler K. H. (2013). The phylogenetic distribution of extrafloral nectaries in plants. *Annals of Botany*, 111, 1251–1261.
- Weber M. G., Porturas L. D., Taylor S. A. (2016). Foliar nectar enhances plant-mite mutualisms: the effect of leaf sugar on the control of powdery mildew by domatia-inhabiting mites. *Annals of Botany*, 118, 459–466.
- Wickham H. (2016). Programming with ggplot2. In: Wickham H, ed. ggplot2: Elegant Graphics for Data Analysis. Cham: Springer International Publishing, 241–253.
- Wilke C. O. (2021). *cowplot: streamlined plot theme and plot annotations for 'ggplot2'*. <https://CRAN.R-project.org/package=cowplot>.
- Willson M. F. (1991). Foliar shelters for mites in the eastern deciduous forest. *The American Midland Naturalist*, 126, 111–117.
- Xuan L., Yan T., Lu L., Zhao X., Wu D., Hua S., Jiang L. (2020). Genome-wide association study reveals new genes involved in leaf trichome formation in polyploid oilseed rape (*Brassica napus* L.). *Plant, Cell & Environment*, 43, 675–691.
- Yue C., Cao H.-L., Chen D., Lin H.-Z., Wang Z., Hu J., Yang G.-Y., Guo Y.-Q., Ye N.-X., Hao X.-Y. (2018). Comparative transcriptome study of hairy and hairless tea plant (*Camellia sinensis*) shoots. *Journal of Plant Physiology*, 229, 41–52.
- Zhang A., Liu Y., Yu C., Huang L., Wu M., Wu J., Gan Y. (2020). Zinc Finger Protein 1 (ZFP1) is involved in trichome initiation in *Arabidopsis thaliana*. *Agriculture*, 10, 645.
- Zhou Z., Sun L., Zhao Y., An L., Yan A., Meng X., Gan Y. (2013). Zinc Finger Protein 6 (ZFP6) regulates trichome initiation by integrating gibberellin and cytokinin signaling in *Arabidopsis thaliana*. *The New Phytologist*, 198, 699–708.

APPENDIX A: CHAPTER 3 SUPPLEMENT

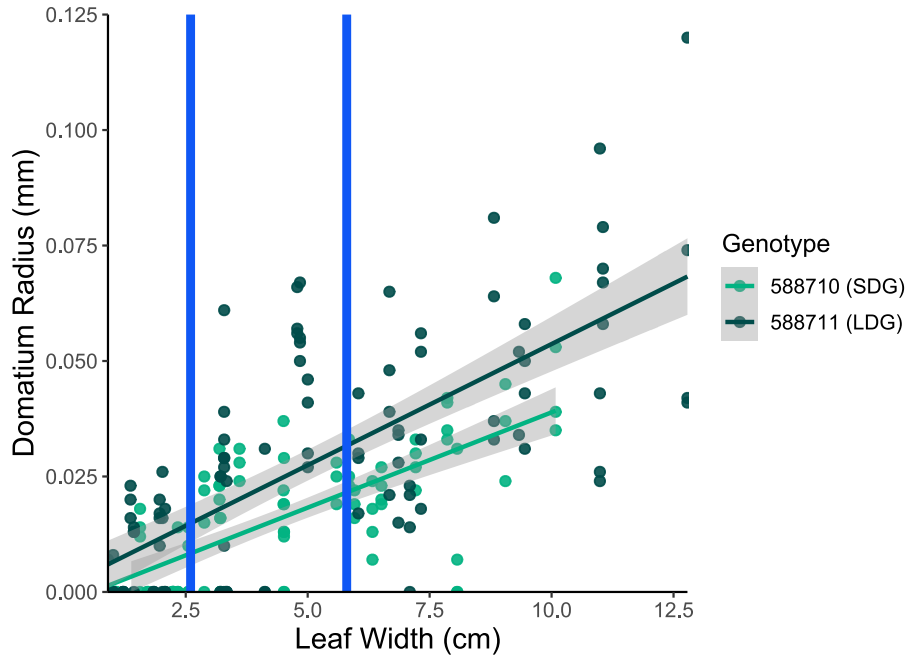


Figure S3.1. Domatia size during leaf expansion. Leaf width (cm) and domatium radius (mm) of leaves collected from bud burst to full expansion for both genotypes of *V. riparia*, 588710 (SDG) and 588711 (LDG). The blue vertical lines represent the lower and upper cutoffs for the leaves that domatia and control tissue were collected from for RNA-sequencing.

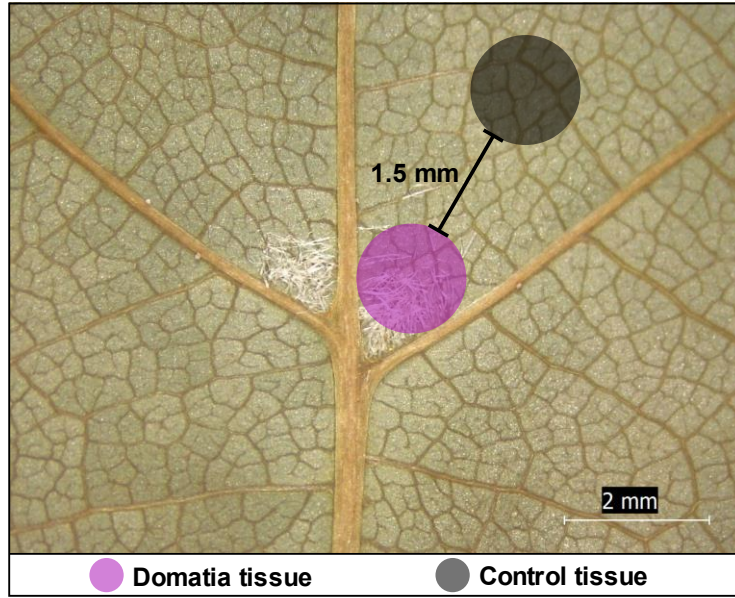


Figure S3.2. Sampling schematic for tissue collected for RNA-sequencing. Domatia and control tissue were collected using a 1.5 mm hole puncher to take circular tissue samples of leaves. The domatia samples were collected by punching out the domatia, avoiding the veins nearby, while the control samples were collected by punching out tissue 1.5 mm directly out from the domatia. The pink circle demonstrates where domatia tissue were collected, and the black circle demonstrates where control tissue was collected.

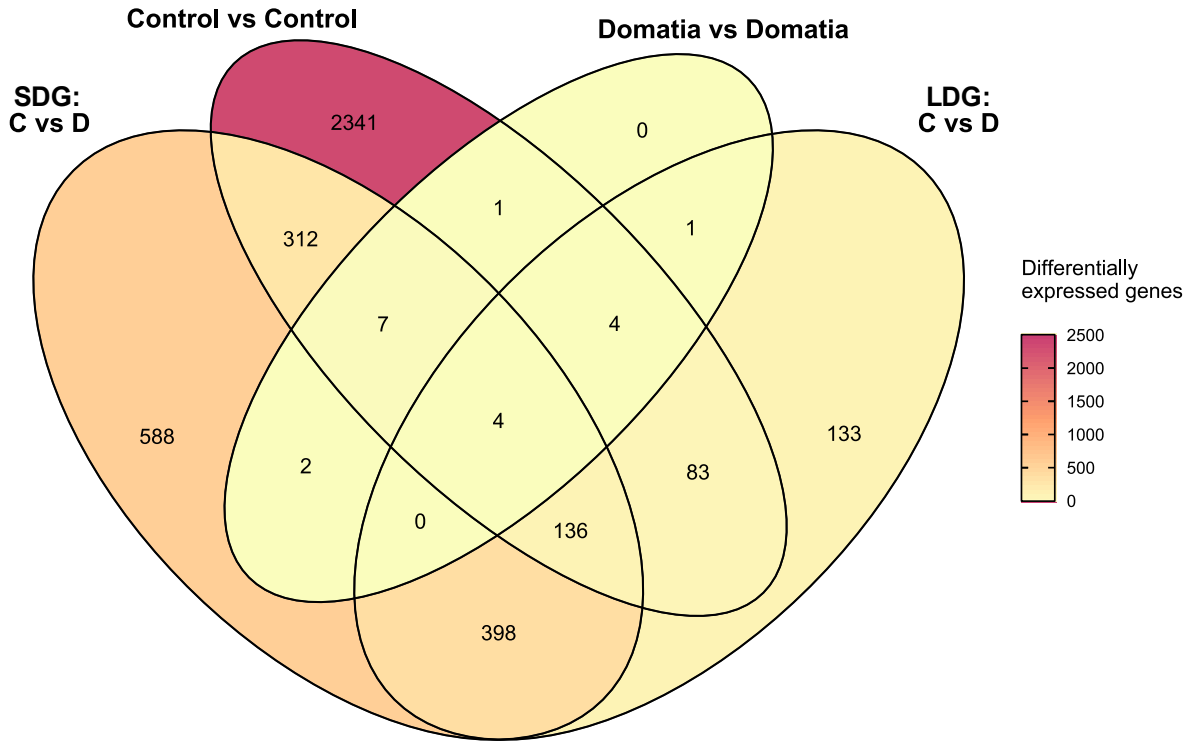


Figure S3.3. The number of differentially expressed that overlap between comparisons. From left to right: control SDG tissue vs domatia SDG tissue (SDG: C vs D), control SDG tissue vs control LDG tissue (Control vs Control), domatia SDG tissue vs domatia LDG tissue (Domatia vs Domatia), and control LDG tissue vs domatia LDG tissue (LDG: C vs D).

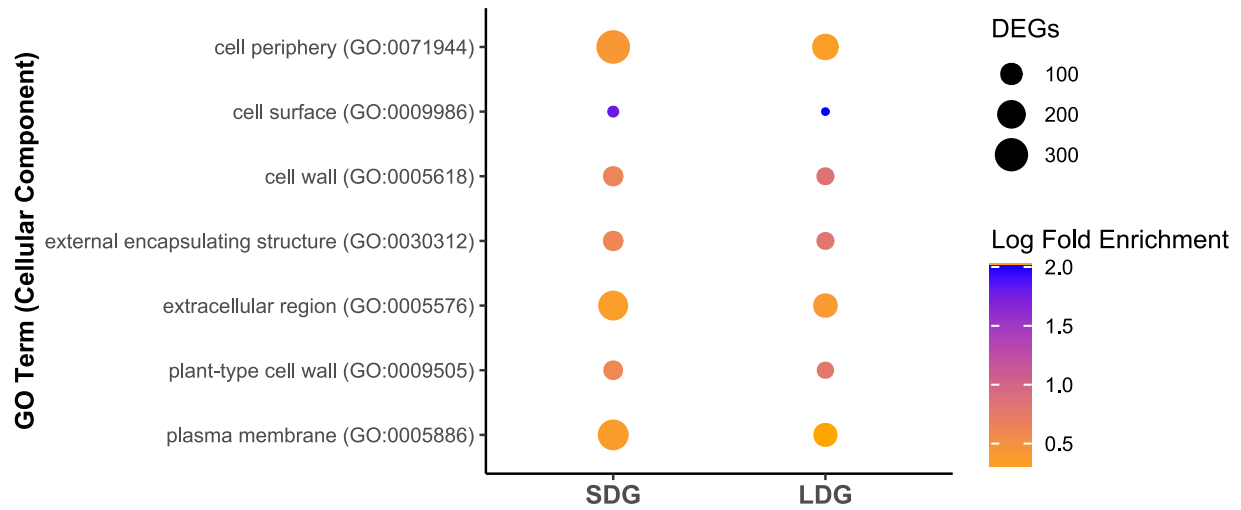


Figure S3.4. Cellular Component GO terms enriched in genes upregulated in domatia. Cellular Component gene ontology (GO) terms enriched ($P < 0.05$) in genes differentially expressed between domatia tissue and control tissue that overlap between both genotypes.

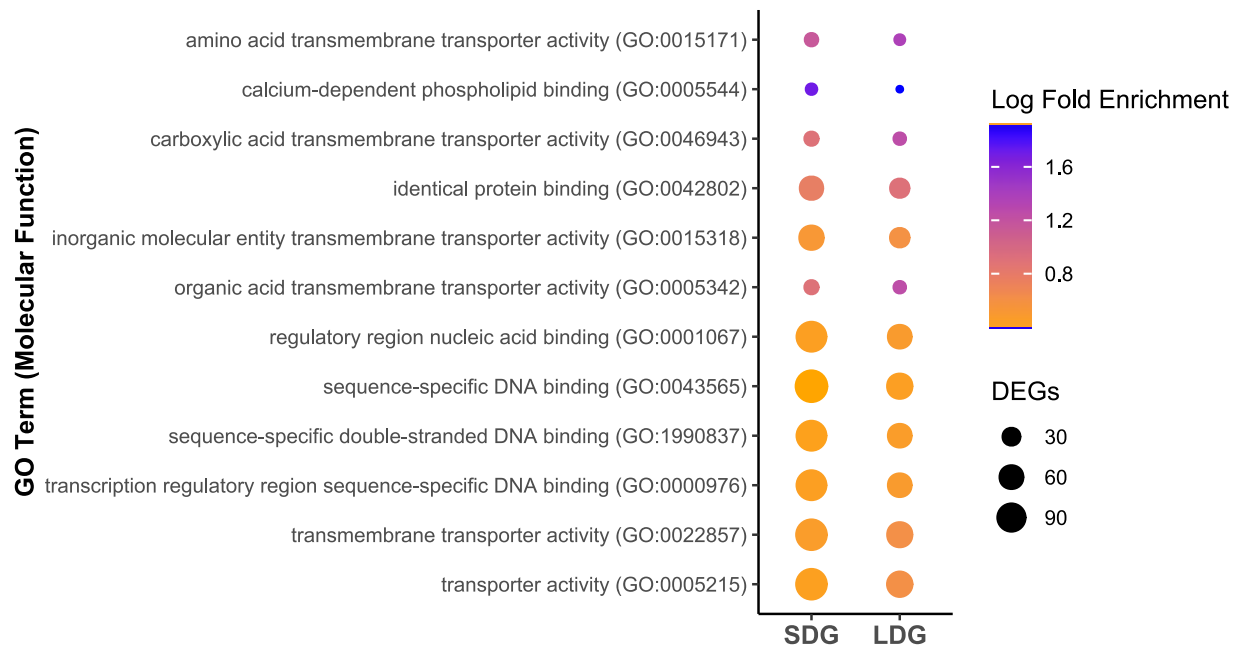


Figure S3.5. Molecular Function GO terms enriched in genes upregulated in domatia. Molecular Function gene ontology (GO) terms enriched ($P < 0.05$) in genes differentially expressed between domatia tissue and control tissue that overlap between both genotypes.

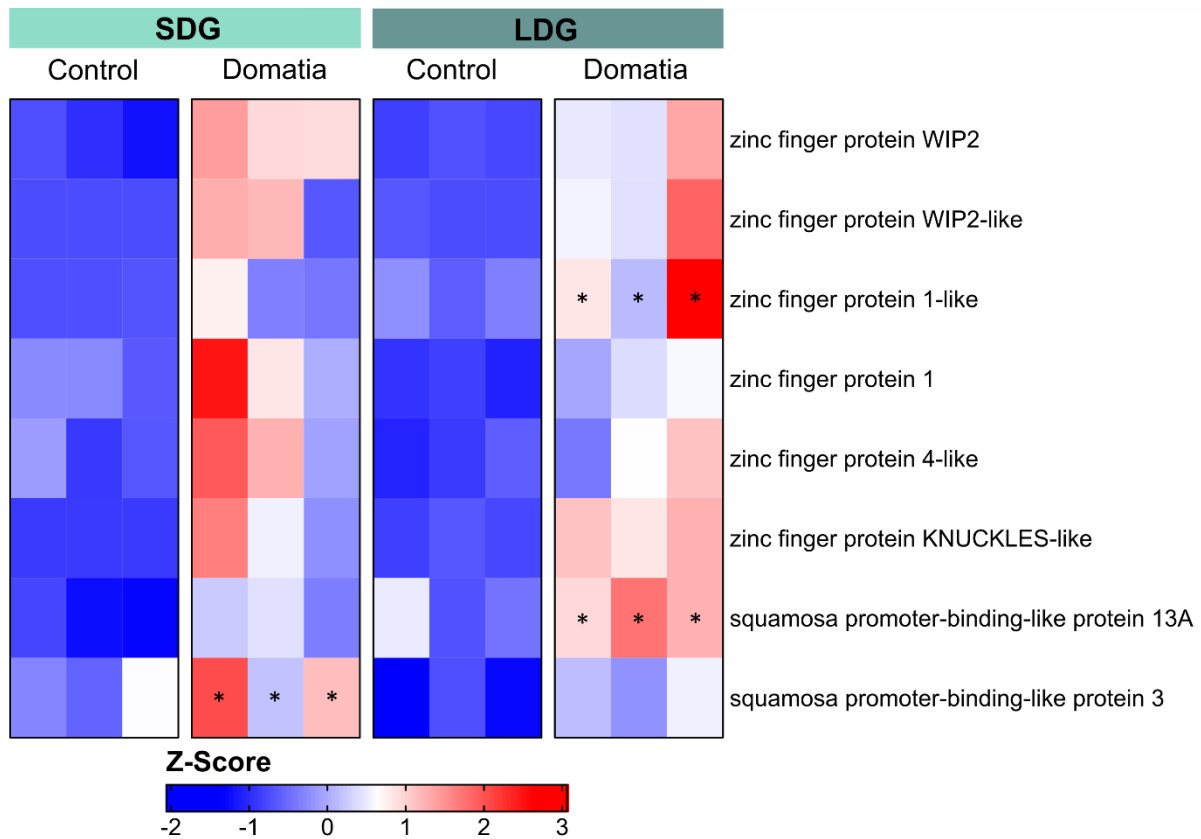


Figure S3.6. C2H2 ZFPs and SPLs upregulated in developing domatia. A heatmap of C2H2 ZFPs and SPLs upregulated in domatia of either both genotypes or one of the genotypes. Each column represents a biological replicate. The cells are colored by Z-score, with blue representing lower expression and red representing higher expression. Genes not differentially expressed in domatia of a particular genotype are denoted with an asterisk.

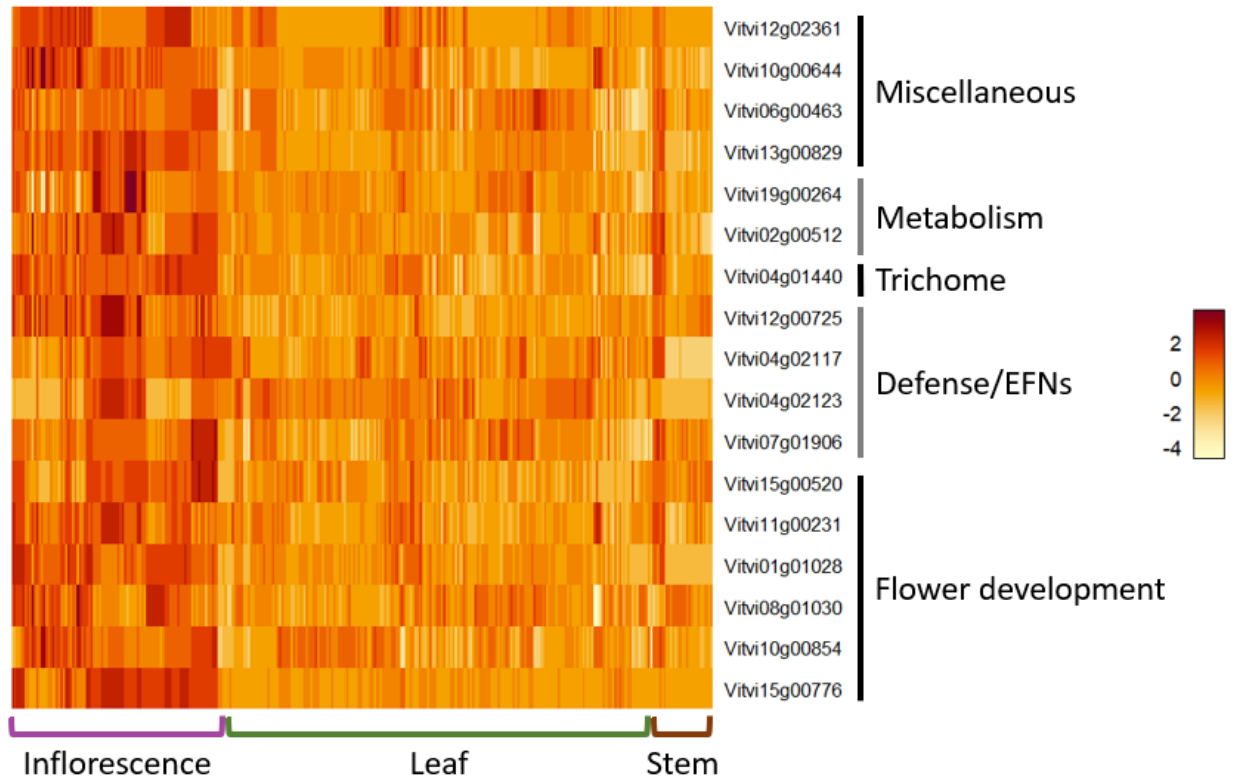


Figure S3.7. Genes upregulated in domatia expressed primarily in floral tissue. A heatmap of the expression of *V. vinifera* orthologs to domatia-upregulated genes for expression that demonstrate preferential expression in floral tissue compared to leaf and stem tissue. Each column represents a biological replicate from GREAT (GRape Expression ATlas). The colors in each cell represent the Z-score, with white representing lower expression and red representing higher expression.

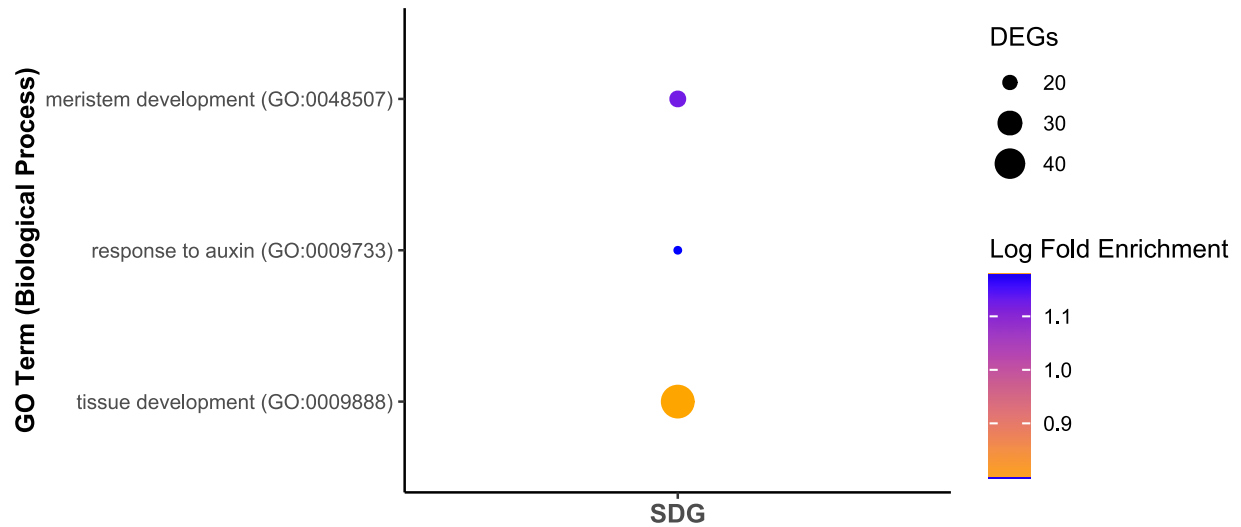


Figure S3.8. Biological Process GO terms enriched in domatia-responsive genes from SDG that are not domatia-responsive in LDG. Biological Process gene ontology (GO) terms enriched ($P < 0.05$) in genes differentially expressed only in SDG domatia that match the following pattern: a) show lower expression levels in the control tissue of SDG compared to the control leaf tissue of LDG, and are upregulated in domatia tissue of SDG (in contrast to control tissue), while b) LDG demonstrates no difference in gene expression levels between control and domatia tissue.



Figure S3.9. Eigenleaves from the PCA comparing leaf shape between scaled SDG and LDG leaves, for PC 1-4.

Table S3.1. Read counts and mapping rates for all RNA-sequencing samples. Raw, trimmed, and mapped reads for all samples, along with mapping rate. For sample names, numbers denote biological sample, with 1,4,5 being 588710/SDG plants and 6,8,9 588711/LDG plants, while "D" corresponds to domatia tissue and "L" corresponds to control tissue. Due to length, this table is available in the supplemental files.

Table S3.2. Differentially expressed genes between SDG control tissue and LDG control tissue. *Vitis riparia* gene names for differentially expressed genes are provided, along with Arabidopsis orthologs and functional descriptions. All DESeq2 results are reported, in addition to both raw and normalized read counts. For sample names, numbers denote biological sample, with 1,4,5 being SDG plants and 6,8,9 LDG plants, while "D" corresponds to domatia tissue and "L" corresponds to control tissue. Due to length, this table is available in the supplemental files.

Table S3.3. Differentially expressed genes between SDG control tissue and SDG domatia tissue. *Vitis riparia* gene names for differentially expressed genes are provided, along with Arabidopsis orthologs and functional descriptions. All DESeq2 results are reported, in addition to both raw and normalized read counts. For sample names, numbers denote biological sample, with 1,4,5 being SDG plants and 6,8,9 LDG plants, while "D" corresponds to domatia tissue and "L" corresponds to control tissue. Due to length, this table is available in the supplemental files.

Table S3.4. Differentially expressed genes between LDG control tissue and LDG domatia tissue. *Vitis riparia* gene names for differentially expressed genes are provided, along with Arabidopsis orthologs and functional descriptions. All DESeq2 results are reported, in addition to both raw and normalized read counts. For sample names, numbers denote biological sample, with 1,4,5 being SDG plants and 6,8,9 LDG plants, while "D" corresponds to domatia tissue and "L" corresponds to control tissue. Due to length, this table is available in the supplemental files.

Table S3.5. Genes that demonstrated significant domatia-genotype interactions between SDG and LDG. *Vitis riparia* gene names for differentially expressed genes are provided, along with Arabidopsis orthologs and functional descriptions. All DESeq2 results are reported, in addition to both raw and normalized read counts. For sample names, numbers denote biological sample, with 1,4,5 being SDG plants and 6,8,9 LDG plants, while "D" corresponds to domatia tissue and "L" corresponds to control tissue. Due to length, this table is available in the supplemental files.

Table S3.6. The Biological Process Gene Ontology (GO) terms enriched ($P < 0.05$) in genes upregulated in SDG domatia. The GO ID and the term are both provided, as well as the total genes annotated in *V. riparia* with the GO term. Significant genes represent the number of genes annotated with the GO term in the upregulated domatia gene set, while expected represents the number of genes expected to have the GO term assuming it is not enriched. The raw weighted Fisher P-values are provided, as well as the adjusted P-value after Benjamini-Hochberg correction. Due to length, this table is available in the supplemental files.

Table S3.7. The Biological Process Gene Ontology (GO) terms enriched ($P < 0.05$) in genes upregulated in LDG domatia. The GO ID and the term are both provided, as well as the total genes annotated in *V. riparia* with the GO term. Significant genes represent the number of genes annotated with the GO term in the upregulated domatia gene set, while expected represents the number of genes expected to have the GO term assuming it is not enriched. The raw weighted Fisher P-values are provided, as well as the adjusted P-value after Benjamini-Hochberg correction. Due to length, this table is available in the supplemental files.

SUPPLEMENTARY METHODS

Because analyses revealed evidence of genes involved in floral structures in domatia,

we asked whether genes involved in domatia development showed specific expression patterns in other tissue types in *Vitis*. We identified *V. vinifera* orthologs to genes differentially expressed in domatia using Diamond v0.8.36 (Buchfink *et al.*, 2015) and the *V. vinifera* reference genome 12X.v2 and its annotation VCost.v3 (Canaguier *et al.*, 2017). The expression patterns of these genes were compared in floral (inflorescence), leaf, and stem tissue using GREAT (GRape Expression ATlas) (<https://great.colmar.inrae.fr/>, unpublished). Genes were identified as being expressed in floral tissue when they had more than 17.45 reads (the 25% quantile of read counts in dataset) expressed in a floral tissue sample. The expression of genes displaying preferential expression in floral tissue was normalized using a Z-score.

REFERENCES

- Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), 59–60.
- Canaguier, A., Grimplet, J., Di Gaspero, G., Scalabrin, S., Duchêne, E., Choisne, N., Mohellibi, N., Guichard, C., Rombauts, S., Le Clainche, I., Bérard, A., Chauveau, A., Bounon, R., Rustenholz, C., Morgante, M., Le Paslier, M.-C., Brunel, D., & Adam-Blondon, A.-F. (2017). A new version of the grapevine reference genome assembly (12X.v2) and of its annotation (VCost.v3). *Genomics Data*, *14*, 56–62.

CHAPTER 4: The assembly and annotation of two teinturier grapevine varieties, Dakapo and Rubired

This chapter is in preparation for submission to *GigaScience* as a Data Note.

AUTHORS AND AFFILIATIONS

Eleanore Ritter¹, Noé Cochetel², Andrea Minio², Peter Cousins³, Dario Cantu^{2,4}, and Chad Niederhuth¹

¹Department of Plant Biology, Michigan State University, East Lansing, Michigan, USA

²Department of Viticulture and Enology, University of California Davis, Davis, California, USA

³E. & J. Gallo Winery, Modesto, CA, USA

⁴Genome Center, University of California Davis, Davis, California, USA

AUTHOR CONTRIBUTIONS

CN envisioned the project, secured the funding, and supervised research on Dakapo. DC envisioned the project, secured the funding, and supervised research on Rubired. EJR assembled and annotated the Dakapo genome, designed and executed comparative analyses within the study, and wrote the first draft of the manuscript. NC and AM assembled and annotated the Rubired genome. PC led Dakapo vine cultivation and tissue sample collection and helped conceptualize comparative analyses within the study.

ABSTRACT

Background

Teinturier grapevine varieties were first described in the 16th century and have persisted due to their deep pigmentation. Unlike most other grapevine varieties, teinturier varieties produce berries with pigmented flesh due to anthocyanin production within the flesh. As a

result, teinturier varieties are of interest not only for their ability to enhance the pigmentation of wine blends but also for their health benefits. Here, we assembled and annotated the Dakapo and Rubired genomes, two teinturier varieties.

Findings

For Dakapo, we used a combination of Nanopore sequencing, Illumina sequencing, and scaffolding to the existing grapevine genome assembly to generate a final assembly of 508.5 Mbp with an N50 scaffold length of 25.6 Mbp and a BUSCO score of 98.0%. A combination approach of *de novo* annotation and lifting over annotations from the existing grapevine reference genome resulted in the annotation of 36,940 genes in the Dakapo assembly. For Rubired, PacBio HiFi reads were assembled, scaffolded, and phased to generate a diploid assembly with two haplotypes 474.7-476.0 Mbp long. The diploid genome has an N50 scaffold length of 24.9 Mbp and a BUSCO score of 98.7%, and both haplotype-specific genomes are of similar quality. *De novo* annotation of the diploid Rubired genome yielded annotations for 56,681 genes.

Conclusions

The Dakapo and Rubired genome assemblies and annotations will provide genetic resources for future investigations into berry flesh pigmentation and other traits of interest in grapevine.

CONTEXT

Domesticated grapevine (*Vitis vinifera*) is the fifth most produced fruit globally (FAO, 2023), with 80.1 million tonnes produced in 2022 alone (Statistics Department of the International Organisation of Vine and Wine, 2022). Various grapevine varieties have been bred since its domestication ~11,000 years ago (Dong et al., 2023), mostly for winemaking

purposes. This has resulted in the selection of numerous diverse phenotypes with significant variation in traits, including berry color and aromatic compounds, as well as more utilitarian traits like yield or biotic and abiotic stress resistance. Berry color is of particular importance in wine grapes due to how it influences wine color and quality. Both consumers and experts prefer red wines with darker colorations (Parpinello et al., 2009; Sáenz-Navajas et al., 2011), making strong pigmentation in berries advantageous for wine producers. Pigmentation typically occurs only in the skin of ripened grapevine berries, with most grapevine varieties having white-colored flesh. The pigmentation within the berry skin is due to the production of anthocyanins, which are colored flavonoids that also act as antioxidants (Flamini et al., 2013). As anthocyanins significantly influence both the quality of wines and their health benefits, the genetic and molecular pathways involved in anthocyanin produced in berry skin have been of high interest and well-characterized (He et al., 2010).

Teinturier (also known as “dyer”) varieties produce berries with pigmented skin and flesh, as well as pigmented leaves. They are highly favorable for use in red wine blends, in which they provide a deeper color. They also remain valuable resources for understanding the production of anthocyanins outside of berry skin. Dakapo and Rubired are two teinturier varieties that are widely grown, and Rubired was the 8th-most crushed grapevine variety in California in 2022 (California Department of Food and Agriculture, 2023). Both varieties are descendants of Teinturier du Cher but of distinct generations. Dakapo was initially bred through a cross between Deckrot and Blauer Portugieser, with Deckrot being a direct descendant of Teinturier du Cher. Rubired is a hybrid grapevine variety bred through a cross between Tinto Cão and Alicante Ganzin, with Alicante Gazin

being a fourth-generation descendent of Teinturier du Cher. While both Dakapo and Rubired are descendants of Teinturier du Cher, their Teinturier du Cher ancestors likely differ based on previous genetic work in teinturier grapes (Röckel et al., 2020) (Figure 4.1).

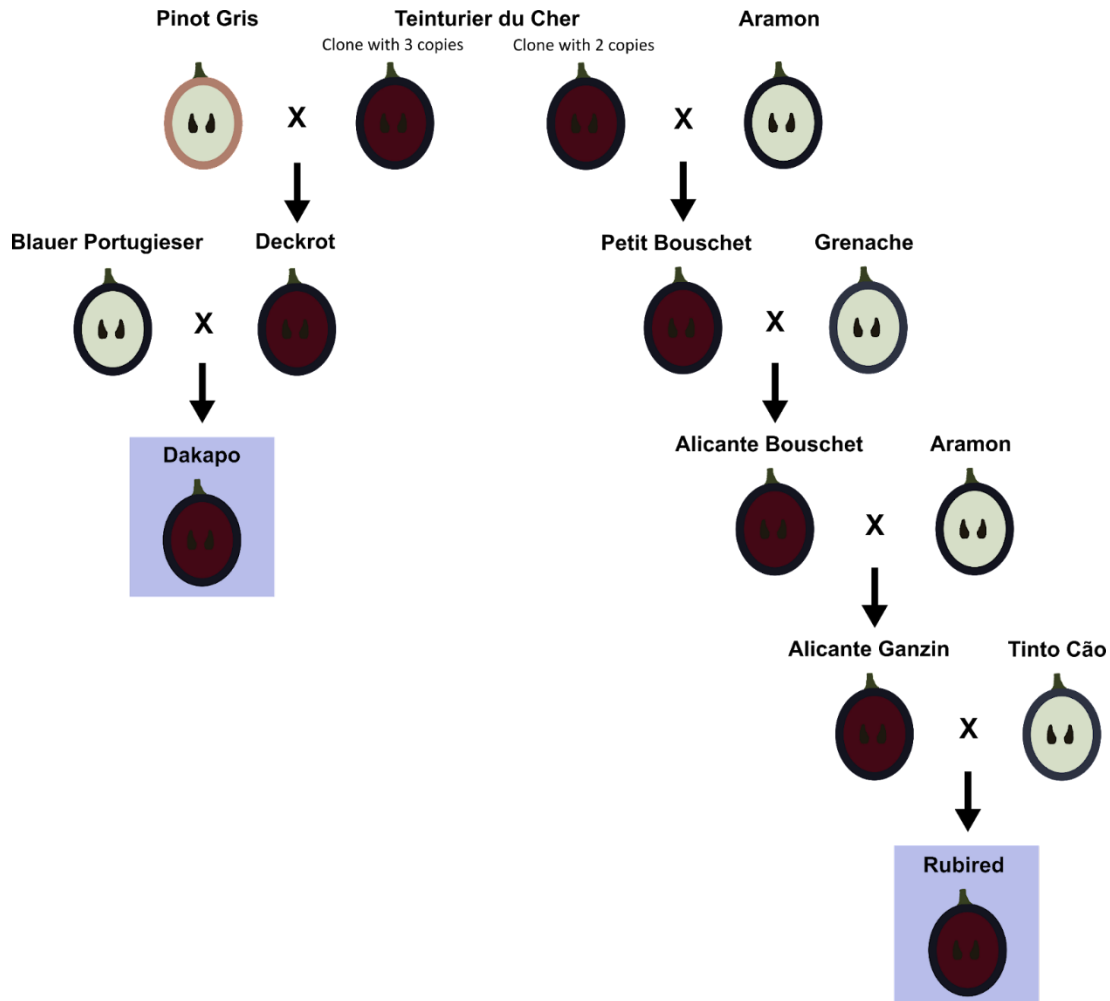


Figure 4.1. The ancestry of Dakapo and Rubired, with berry skin and flesh color shown. Dakapo and Rubired are thought to have been bred from Teinturier du Cher clones with differing copy numbers of a 408 bp repeat within the promoter of *VvMybA1*, which is noted in the figure.

Teinturier varieties have substantially higher anthocyanin content in their berries than non-teinturier varieties due to the accumulation of anthocyanins within their berry

flesh. Previous work showed that the juice produced with Dakapo berries had 39-91 times more anthocyanin content than commercial red grape juice (Fröhling et al., 2012). Teinturier varieties themselves vary in anthocyanin content and the profiles of anthocyanins present within the berry flesh (Kőrösi et al., 2022; Röckel et al., 2020). Previous studies have made progress in investigating the genetic basis of berry flesh pigmentation and variation in overall anthocyanin production within teinturier grapes. A previous study (Röckel et al., 2020) demonstrated that a 408 bp repeat in the promoter of the gene *VvMybA1* is directly linked to increased anthocyanin production in teinturier berries. They also found that the copy number of this 408 bp sequence varied between varieties and that varieties derived from Teinturier du Cher had either two, three, or five copies of this 408 bp sequence within the promoter region. Increased copies of these repeats were correlated with increased anthocyanin content within berry skin and flesh (Röckel et al., 2020). While this study greatly illuminated the genetic basis of increased anthocyanin production in Teinturier du Cher descendants, it is still unclear why different teinturier grapes have distinct anthocyanin profiles in berry flesh, regardless of the number of copies of the 408 bp sequence they have (Kőrösi et al., 2022). For example, the concentration of a specific type of anthocyanin, Cyanidin-3-*O*-glucoside, can vary from 1.0-41.7 mg/L when comparing the anthocyanin content within the flesh of teinturier berries from different varieties (Kőrösi et al., 2022). While overall anthocyanin content does correlate with the number of repeats upstream of *VvMybA1* (Röckel et al., 2020), large differences in concentrations of specific anthocyanins exist between teinturier varieties with the same number of copies of the 408 bp repeat (Kőrösi et al., 2022). The assembly and annotation of the genome of the Yan73 teinturier grapevine variety were recently

generated and provided additional insight into the regulation of anthocyanin accumulation in Yan73 berry flesh (Zhang et al., 2023). However, the lack of additional genomic resources for teinturier grapes has inhibited further investigations into differences between teinturier varieties, and the genetic basis for these large differences in anthocyanin composition remains unclear as a result.

Here, we sequenced, assembled, and annotated the Dakapo and Rubired genomes to provide additional resources for understanding teinturier varieties and to further enable their use in breeding programs. These genomes will greatly facilitate future work into understanding the regulation of anthocyanins within berry flesh. Beyond anthocyanin production, Dakapo and Rubired have also been utilized in researching other traits in grapevine. A QTL mapping population of Dakapo x Cabernet Sauvignon has been established and utilized for investigating *Botrytis* bunch rot in grapevine (Herzog et al., 2021). Additionally, Rubired is notable for being highly resistant to *Xylella fastidiosa*, which causes Pierce's disease in grapevine (Rashed et al., 2011, 2013; Wallis et al., 2013). As a result, we believe these high-quality reference genome assemblies and annotations will be a useful resource for the grapevine and plant science communities.

METHODS

Plant material

Vitis vinifera plants of the Dakapo variety were planted in Madera, California, USA in 2011. Young leaf tissue samples for Oxford Nanopore Technologies (ONT) long-read sequencing were collected in July 2021. The samples were frozen and shipped on dry ice overnight. Plant material used in this study was also utilized in a previous study (Ritter et al., 2023) as the "Dakapo WT" samples. For Rubired tissue, young leaves were collected

from the accession Rubired FPS clone 02 maintained by the Foundation Plant Services at the University of California, Davis.

DNA extraction and sequencing of Dakapo tissue

High molecular weight DNA was extracted and a sequencing library was prepared for ONT sequencing by the Genomics Core at Michigan State University as previously described (Ritter et al., 2023), using the Oxford Nanopore Technologies Ligation Sequencing Kit (SQK-LSK109). The library was sequenced on a PromethION FLO-PRO002 flow cell (R9.4.1; Oxford Nanopore Technologies) on a PromethION24 (Oxford Nanopore Technologies) running MinKNOW Release 21.11.7 (Oxford Nanopore Technologies), resulting in 61.9 Gbps of sequence (~123.8X coverage) with a read length N50 of 13.2 kbp. Base calling and demultiplexing were performed using Guppy v5.1.13 (Oxford Nanopore Technologies) with the High Accuracy base calling model. An additional 9.5 Gbps (~18.9X coverage) of ONT sequencing data and 25.6 Gbps (~51.3X coverage) of Illumina paired-end sequencing data previously published from the “Dakapo WT” sample described previously (Ritter et al., 2023) were utilized for this study (see Table S4.1 for full sequencing statistics).

Dakapo genome assembly

Raw ONT sequencing data from this study and previous work with the same plant material (Ritter et al., 2023) were combined. Adapters were trimmed using Porechop v0.2.4 (Wick et al., 2017) with the following settings: `--min_trim_size 5`, `--extra_end_trim 2`, `--end_threshold 80`, `--middle_threshold 90`, `--extra_middle_trim_good_side 2`, `--extra_middle_trim_bad_side 50`, and `--min_split_read_size 300`. Reads mapping to the lambda phage genome were removed using NanoLyse v1.2.0 (De Coster et al., 2018). NanoFilt

v2.8.0, with the flags *-q 0* and *-l 300*, was used to remove low-quality reads and reads shorter than 300 base pairs (bp) (De Coster et al., 2018). The quality of the reads was analyzed using FastQC v0.11.9 (Andrews, 2010), NanoStat v1.6.0 (De Coster et al., 2018), and NanoPlot v1.38.0 (De Coster & Rademakers, 2023). ONT reads were then assembled using Flye v2.8.3-b1695 (Kolmogorov et al., 2019) for two iterations. One round of polishing was performed on the assembly using the ONT reads with Racon v1.4.20 (Vaser et al., 2017) and the following settings: *--include-unpolished, -m 8, -x -6, -g -8, and -w 500*. The assembly was then scaffolded to the 12X.v2 grapevine genome assembly (Canaguier et al., 2017) using RagTag v2.0.1 “scaffold” (Alonge et al., 2022) with the following settings: *-f 1000, -d 100000, -i 0.2, -a 0.0, -s 0.0, -r, -g 100, and -m 100000*.

Paired-end Illumina reads were used for the final polishing of the scaffolded assembly. These were first trimmed using Trimmomatic v0.39 (Bolger et al., 2014) to remove adapters and low-quality sequences, with the following settings used: *-phred33, ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10:4:TRUE, LEADING:3, TRAILING:3, SLIDINGWINDOW:4:15, and MINLEN:30*. The reads were then mapped to the draft genome assembly using BWA-MEM v0.7.17-r1188 (Li, 2013) with the *-M* flag. PCR duplicate reads were removed using Picard MarkDuplicates v2.15.0 (Broad Institute, 2017) with the *--REMOVE_DUPLICATES TRUE* flag. The mapped reads with marked duplicates were then used to polish the draft assembly using Pilon v1.24 (Walker et al., 2014) with the *--fix all* flag used to correct all errors identified and the *--diploid* flag. Two iterations of Pilon polishing were performed.

Following polishing, haplotigs were removed using Purge Haplotigs v1.1.2 (Roach et al., 2018). To do so, all prepped ONT reads were mapped to the Dakapo draft assembly

using minimap2 v2.23-r1111 (Li, 2021) with the flags *-ax map-ont* and *-L*, and *purge_haplotigs hist* was then run with default settings to generate a read-depth histogram of these mapped reads. Based on the histogram generated, *purge_haplotigs cov* was run with the previous output file and the following flags: *-low 15*, *-mid 88*, and *-high 195*. Finally, *purge_haplotigs purge* was run using the previous output file to purge haplotigs from the Dakapo draft assembly (Roach et al., 2018).

Before finalizing the Dakapo genome assembly, chr00 was split apart manually at gaps since it is an artificial chromosome of unmapped contigs from the 12X.v2 grapevine genome assembly (Canaguier et al., 2017) used for scaffolding. The assembly was also searched for microbial contamination using the *gather-by-contig.py* script adapted from (Brown, 2018) that utilizes sourmash and its pre-built database “GTDB R06-RS202 genomic representatives” (Brown & Irber, 2016). No contamination was found from this process. The chromosome names were maintained from the scaffolding to the 12X.v2 grapevine genome assembly (Canaguier et al., 2017). All other contigs were sorted and renamed in order of length, including the contigs split apart from chr00, using the custom script *sort_rename_fasta.sh*.

To assess the quality of the Dakapo genome assembly, we used BUSCO v5.2.2 (Manni et al., 2021) to check the completeness of the assembly when compared to the *eudicots_odb10* dataset at each step of genome assembly and polishing. Assembly statistics were calculated using *assembly-stats v1.0.1* (*Assembly-Stats*, 2020). Finally, the quality of repetitive sequences and intergenic space was also assessed by calculating the long terminal repeats (LTR) Assembly Index (LAI) for the Dakapo assembly using LTRs

annotated by the Extensive de-novo TE Annotator (EDTA) (see below for transposable element annotation methods).

Dakapo genome annotation

Transposable elements (TEs) and repeats in the Dakapo genome assembly were annotated using EDTA v1.9.4 (Ou et al., 2019) with the following flags: *--species others, --step all, --overwrite 1, --sensitive 1, --anno 1, --evaluate 0, and --force 1.*

MAKER was used for *de novo* annotation of genes in the Dakapo genome. Before running MAKER, RNA-seq reads from diverse tissues in grapevine and protein sequences from related species were used to provide initial support for gene models. To do so, RNA-seq samples from Perazzolli et al. (2012), Da Silva et al. (2013), Minio et al. (2019), Vannozzi et al. (2021), Daldoul et al. (2022), and Ma et al. (2023) were downloaded from the NCBI Sequence Read Archive (SRA) using fasterq-dump v2.10.7 from the sra-toolkit (SRA Toolkit Development Team, 2020) (see Table S4.2 for the SRA IDs of the specific files used). Trimmomatic v0.39 (Bolger et al., 2014) was used to trim adapters from Illumina RNA-seq reads with the flags: *--phred33* and *ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10*. These were then mapped to the Dakapo genome assembly using HISAT2 v2.2.1 (Kim et al., 2019) with the *--phred33* flag. PacBio RNA-seq reads were mapped to the Dakapo genome assembly using minimap2 v2.23-r1111 (Li, 2021) with the flags *-ax splice:hq* and *-uf*. Transcripts from these mapped RNA-seq reads were assembled using StringTie v2.2.1 (Shumate et al., 2022) with the following flags: *-c 1, -f 0.01, -m 200, -a 10, -j 1, -M 1, -s 4.75* (for mapped Illumina reads) or *1.5* (for mapped PacBio reads), and *-g 50* (for mapped Illumina reads) or *0* (for mapped PacBio reads). The output files for all RNA-seq samples

were converted to gff3 files using gffread v0.12.7 (Shumate et al., 2022), combined, and then sorted using gff3_sort v2.1.0 (Chen et al., 2019).

Before protein sequences were aligned to the Dakapo genome assembly, repeats in the assembly were masked using RepeatMasker v4.1.2-p1 with the TE library generated using EDTA and the following flags: *-e rmbblast, -s, -norna, -xsmall, -gff, -html, and -source*. Protein sequences from the Arabidopsis (*Arabidopsis thaliana*) Araport11 annotation (Cheng et al., 2017), the *Oryza sativa* Release 7 annotation (Kawahara et al., 2013), and the Viridiplantae UniProtKB/Swiss-Prot reviewed protein sequence dataset from UniProt release 2023_05 (UniProt Consortium, 2023) were aligned to the masked Dakapo genome assembly using exonerate v2.4.0 (Slater & Birney, 2005) with the following flags: *--model protein2genome, --bestn 5, --minintron 10, --maxintron 5000, --querychunktotal 5, --targetchunktotal 10, --showtargetgffyes, --showalignment no, --showvulgar no, --ryo ">%qi length=%ql alnlen=%qal\n>%ti length=%tl alnlen=%tal\n"*. The outputs for each dataset were combined, reformatted using the custom script reformat_exonerate_protein_gff.pl, and sorted using gff3_sort v2.1.0 (Chen et al., 2019).

MAKER v3.01.04 (Holt & Yandell, 2011) was initially run on the Dakapo genome assembly with the gff files generated through transcript assembly and protein sequence alignment. These initial annotations generated by MAKER were then used to train SNAP and AUGUSTUS. To train SNAP, maker2zff from MAKER was first used to convert genes to the ZFF format with the flag *-x 0.1*. This input was used with SNAP v2013_11_29 (Korf, 2004) to first categorize genes by running the command *fathom* to produce reformatted files, followed by the command *forge* to estimate parameters. Hidden Markov Models (HMMs) were created using *hmm-assembler.pl* from SNAP (Korf, 2004). To train

AUGUSTUS, maker2zff from MAKER was first used to convert genes to the ZFF format with the following flags: *-c 0.5, -e 0.5, -o 0.5, -a 0, -t 0, -l 200, and -x 0.2*. The *fathom* command from SNAP (Korf, 2004) followed by the custom script *fathom_to_genbank.pl* were then run to reformat the files and keep only 600 randomly sampled annotations. Fasta files of the subsetted genes were then generated using the custom script *get_subset_of_fastas.pl*. These subsetted genes were split into training and test files, and then *autoAug.pl* from AUGUSTUS v3.4.0 (Stanke et al., 2008) was run to produce batch scripts that were then run. This step was repeated, using the following flags with *autoAug.pl*: *-useexisting* and *--index=1*. The sensitivity and specificity of the AUGUSTUS HMMs were evaluated by running the *augustus* command. A second round of MAKER v3.01.04 (Holt & Yandell, 2011) was then run using the HMMs from SNAP and AUGUSTUS to produce gene annotations.

We then filtered annotations and flagged genes that may have been misannotated as transposons using methods described previously (Kollar et al., 2023). To ensure that our annotations were as complete as possible, we used Liftoff v1.6.2 to transfer annotations from the PN40024.v4 grapevine genome assembly (Velt et al., 2023) to the Dakapo genome. We then used the methods described previously (Kollar et al., 2023) to assign “pseudogene” and “gene” labels to the lifted genes based on the confidence of the lifted gene model.

Finally, gene functions were assigned to each annotated gene by first using InterProScan v5.66-98.0 (Jones et al., 2014) to assign Pfam domains and corresponding gene ontology (GO) terms using the following flags: *-appl pfam, -goterms, -pa, -dp, -iprlookup, -t p, and -f TSV*. Then, Arabidopsis orthologs were identified by running DIAMOND v2.0.15.153 (Buchfink et al., 2015) with protein sequences from Dakapo and the

TAIR10 Arabidopsis annotation (Lamesch et al., 2012) and the following flags: `--evaluate 1e-6`, `--max-hsps 1`, `--max-target-seqs 5`, and `--outfmt 0`. The results from InterProScan, DIAMOND, and Arabidopsis gene functions and GO terms of orthologs (Berardini et al., 2015) were all combined to generate a file with functional descriptions for each gene using the custom script `create_functional_annotation_file.pl`.

DNA extraction and sequencing of Rubired tissue

High molecular weight genomic DNA was extracted using the method previously described (Chin et al., 2016). The HiFi library preparation and sequencing were performed as previously described (Minio et al., 2022). The HiFi library fraction with a length >15 kbp was sequenced in two SMRT cells on a PacBio Sequel IIe platform at the DNA Technology Core Facility, University of California, Davis. The sequencing generated 31.3 Gbp sequences corresponding to 62.6X coverage with an N50 of 11.5 kbp.

Rubired genome assembly

The pseudomolecules of the Rubired genome were assembled, phased, and scaffolded as described previously (Minio et al., 2022). Briefly, after testing multiple Hifiasm v.0.16.1-r374 (Cheng et al., 2021) parameters, the best assembly obtained with the configuration `'-a 4 -k 41 -w 71 -f 25 -r 4 -s 0.7 -D 3 -N 100 -n 25 -z 20'` was selected consisting of 273 contigs with an N50 = 12.9 Mb. An integrated phasing and scaffolding procedure further led to the construction of chromosome-scale pseudomolecules using HaploSync (Minio et al., 2022) combined with a high-density consensus map (Zou et al., 2020). Quality and completeness of the assembly were assessed as described above for the Dakapo genome.

Rubired genome annotation

The structural and functional annotation of the Rubired genome followed the exhaustive annotation pipeline described (Cochetel et al., 2021). Briefly, high-quality Iso-Seq data from *V. vinifera* Cabernet Sauvignon (Minio et al., 2019), quality-based filtered RNA-Seq data from *V. rupestris* (Cochetel et al., 2023), and external databases were used to generate a collection of assemblies, alignments, *ab initio* predictions, and transcript/protein evidence. The resulting consensus gene models were then polished, filtered, and functionally annotated following the annotation workflow established previously (Cochetel et al., 2021).

Investigating the VvMybA1 sequences in the Dakapo and Rubired genomes

Teinturier grape varieties are known to have tandem copies of a 408 bp sequence within the promoter region of *VvMybA1*, a key gene in anthocyanin biosynthesis, which leads to increased anthocyanin accumulation within their berries. Dakapo and Rubired both contain 3 and 2 copies of this repeat, respectively (Röckel et al., 2020). To investigate the sequence similarities of these repeat sequences in our assembled genome, we first used blastn from BLAST v2.10.0+ (Camacho et al., 2009) to search for the locations of *VvMybA1* within the Dakapo genome and both Rubired haplotypes. We then extracted the sequences of *VvMybA1* and the 10 kbp surrounding the gene from the genomes using bedtools v2.27.1 *getfasta* (Quinlan & Hall, 2010). We searched these sequences for the 408 bp repeat identified previously (Röckel et al., 2020) both manually and using blastn from BLAST v2.10.0+ (Camacho et al., 2009).

Exploring synteny among various grapevine genomes

GENESPACE v1.3.1 (Lovell et al., 2022) was used with MCScanX (Wang et al., 2012) and Orthofinder v2.5.5 (Emms & Kelly, 2019) to align and plot protein sequences from the following grapevine genomes: Dakapo, Rubired, Cabernet Franc (Minio et al., 2022), Cabernet Sauvignon (Minio et al., 2022), Chardonnay (Minio et al., 2024), and Pinot Noir (Cantu Lab, n.d.). Individual chromosomes from the Dakapo and Rubired assemblies were aligned using MUMmer v4.0.0rc1 (Marçais et al., 2018). To do so, first, the *nucmer* command was run with default settings. The command *delta-filter* was then run with the following flags: *-i 90 -l 5000*. Finally, plots were generated using the *mummerplot* command.

DATA DESCRIPTION AND QUALITY CONTROL

The Dakapo genome assembly and annotation

The Dakapo genome was assembled using ONT reads representing 142.4X coverage (based on a genome size of 500 Mbps). This draft assembly was then scaffolded to the 12X.v2 grapevine reference genome (Canaguier et al., 2017) and polished using both ONT reads (142.4X coverage) and Illumina reads (51.3X coverage) to produce the final genome assembly of 508.5 Mbp. It is comprised of 19 chromosomes and 542 unplaced contigs, with 96.3% of the Dakapo assembly sequence located on the chromosomes and 2,644 gaps of unknown sequence. The final genome assembly is highly contiguous, with an N50 of 25.6 Mbp, slightly higher than the most recent PN40024.v4 assembly (Velt et al., 2023). The Dakapo assembly has a high BUSCO (Benchmarking Universal Single-Copy Orthologue) score of 98% complete BUSCOs (94.7% single-copy BUSCOs and 3.3% duplicated BUSCOs), similar to prior PN40024 reference assemblies (97.7% for 12X.v2 (Canaguier et al., 2017) and 98.3% for PN40024.v4 (Velt et al., 2023)). In addition, the Dakapo genome received a

raw LAI score of 12.22 and thus contains a reference-quality assembly of repetitive/intergenic sequences (Ou et al., 2018) (Table 4.1).

	Dakapo	Rubired whole assembly	Rubired haplotype 1	Rubired haplotype 2	12X.v2	PN40024.v4
Assembly size (Mbp)	508.5	983.8	476.0	474.7	486.2	475.6
Number of contigs	561	185	19	19	20	22
N50 (Mbp)	25.6	24.9	24.7	24.9	24.3	24.4
Number of gaps	2,644	97	38	59	15,325	4,019
Total complete BUSCO	98.0%	98.7%	98.3%	97.3%	97.7%	98.3%
Raw LAI	12.22	N/A*	15.22	15.62	9.4**	13.97

Table 4.1 Assembly statistics of the Dakapo and Rubired genome assemblies presented here, as well as assembly statistics of the 12X.v2 (Canaguier et al., 2017) and PN40024.v4 (Velt et al., 2023) grapevine reference genome assemblies. The Rubired whole assembly contains both haplotypes and unplaced sequences.

*The raw LAI score was not calculated for the whole assembly due to high sequence similarity between haplotypes which would prevent an accurate calculation.

**Previously calculated (Ou et al., 2018).

The Dakapo genome was annotated using a combination of *de novo* annotations using MAKER (Holt & Yandell, 2011) and annotations lifted from PN40024.v4 (Velt et al., 2023) using Liftoff (Shumate & Salzberg, 2021). This resulted in 36,940 genes being annotated. We also annotated both TEs and repeat sequences and found that these comprised 45.38% of the genome, similar to what has previously been reported in grapevine (41.4-51.1% (Jaillon et al., 2007; Minio, Massonnet, Figueroa-Balderas, Castro, et al., 2019; Zhou et al., 2019)). LTRs make up a majority of the repetitive sequences

annotated in the Dakapo genome and comprise 30.48% of the genome, with *Gypsy* LTRs specifically being the most abundant type, comprising 12.88% of the Dakapo genome sequence (Table S4.3).

The Rubired genome assembly and annotation

The Rubired genome was sequenced with highly accurate long-read sequencing generating 62.6X HiFi coverage (using a haploid genome size of 500 Mbp as reference). Pseudomolecules were constructed by scaffolding and phasing the assembly using HaploSync (Minio, Cochetel, Vondras, et al., 2022) generating two haplotypes comprising 19 chromosomes and averaging to a total length of ~475 Mbp. With complete BUSCO scores of 98.3% and 97.3% for haplotype 1 and haplotype 2, respectively (between 95.4-96.3% single copy BUSCOs and between 1.9-2.0% duplicated BUSCOs), and only 33 Mb of unplaced sequences in the diploid assembly of the Rubired genome, the Rubired assembly is highly complete. Both genomes for the two Rubired haplotypes also have high raw LAI scores (15.22 for haplotype 1 and 15.62 for haplotype 2), demonstrating that the diploid Rubired genome contains a reference-quality assembly of repetitive/intergenic sequences that are likely more complete than the 12X.v2 (Canaguier et al., 2017)3] and PN40024.v4 (Velt et al., 2023) assemblies (Table 4.1). The gene annotation resulted in 56,681 genes showing a chromosome anchoring of 97.7% further supporting the reference quality of the assembly. Overall, the genome was composed of 50.46% of repetitive sequences with a clear accumulation in the unplaced sequences with 74.34% of its sequences annotated as repeats. The repeat distribution was similar to the Dakapo genome with *Gypsy* LTRs as the predominant repeats type corresponding to 13.91% of the genome sequence (Table S4.4).

RE-USE POTENTIAL

Grapevine varieties have been bred to produce berries in a variety of colors, commonly divided into red-, black-, and white-skinned berries that typically have white flesh. However, there are several varieties of teinturier grapes, which contain pigmented skin and pigmented flesh, including the Dakapo and Rubired varieties sequenced here. By assembling these genomes, we fully assembled *VvMybA1* and the tandem repeat associated with anthocyanin content in teinturier grapes (Röckel et al., 2020). As expected, we found three tandem copies of this repeat within the promoter region of *VvMybA1* (the *VvMybA1t3* allele) in the Dakapo genome, exactly as described previously (Röckel et al., 2020). All three repeats contain identical 408 bp repeat sequences (Figure 4.2A). In addition, the Rubired haplotype 2 assembly contained two tandem copies of this repeat at the exact location (the *VvMybA1t2* allele) as expected (Röckel et al., 2020), with both copies containing the same 408 bp sequence as those in Dakapo (Figure 4.2B). The Rubired haplotype 1 assembly did not contain teinturier-associated alleles but instead contained the *VvMybA1a* allele responsible for white berry skin color (Kobayashi et al., 2004), as expected based on previous findings (Röckel et al., 2020). The *VvMybA1a* allele is distinct from teinturier alleles and other functional *VvMybA1* alleles due to the presence of the *Gret1* retrotransposon upstream of coding sequences (Kobayashi et al., 2004). However, it does contain the 408 bp repeat upstream of the *Gret1* retrotransposon (Röckel et al., 2020). This repeat sequence is not perfectly identical to the repeat in Dakapo or Rubired haplotype 2 and instead contains three single base pair mutations within the sequence (Figure 4.2C).

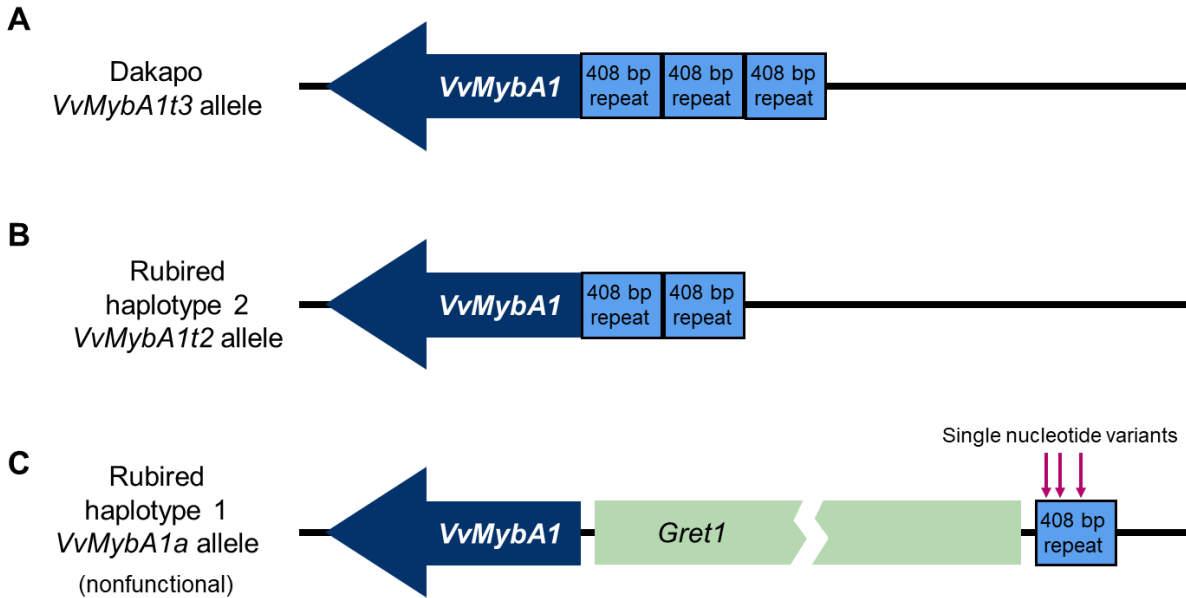


Figure 4.2. Diagrams of the *VvMybA1* alleles in **A)** the Dakapo assembly, **B)** the Rubired haplotype 2 assembly, and **C)** the Rubired haplotype 1 assembly. *VvMybA1* is represented by the dark blue arrow and the 408 bp repeats are shown in light blue boxes. Dakapo contains three tandem copies of the 408 bp repeat, while the Rubired haplotype 2 assembly contains two tandem copies. The Rubired haplotype 1 assembly contains the nonfunctional *VvMybA1a* allele with the *Gret1* retrotransposon shown upstream of *VvMybA1* in a light green box, truncated to fit in the figure. The three single nucleotide variants within the 408 bp repeat of the nonfunctional allele in Rubired haplotype 1 are indicated by arrows.

Beyond fully sequencing the *VvMybA1* alleles of Dakapo and Rubired, these genomes will enable more insight into grapevine berry color by providing two high-quality teinturier grapevine genomes for future studies. As previously mentioned, teinturier grapes differ in the composition of total anthocyanins produced, and this phenomenon does not seem to be driven by differences in *VvMybA1* alleles (Körösi et al., 2022). These genomes will provide resources for investigating the genetic mechanisms driving this phenomenon. Focusing on the berry color locus on chromosome 2 (Azuma et al., 2009; Doligez et al., 2002; Walker et al., 2007) and the anthocyanin locus on chromosome 14 (Matus et al., 2017), in particular, may provide insight into the regulation of specific anthocyanin molecules within the flesh of teinturier berries.

The Dakapo and Rubired genomes and annotations will also offer additional resources for future work in grapevine. Beyond berry flesh color, the Dakapo and Rubired genomes will also provide resources for investigating additional traits. For example, Dakapo is both frost-susceptible (Lisek, 2012) and *Botrytis*-susceptible (Herzog et al., 2021), while Rubired is notably highly mildew-resistant (Doster, 1985). A QTL-mapping population generated through a cross between Dakapo x Cabernet Sauvignon has also been previously established (Herzog et al., 2021), so the availability of this reference genome will greatly aid future studies with this population. These genomes will ultimately provide new resources for investigating a variety of grapevine traits, enabling advances in grapevine breeding and agriculture and allowing for comparisons between grapevine genomes (Figure 4.3).

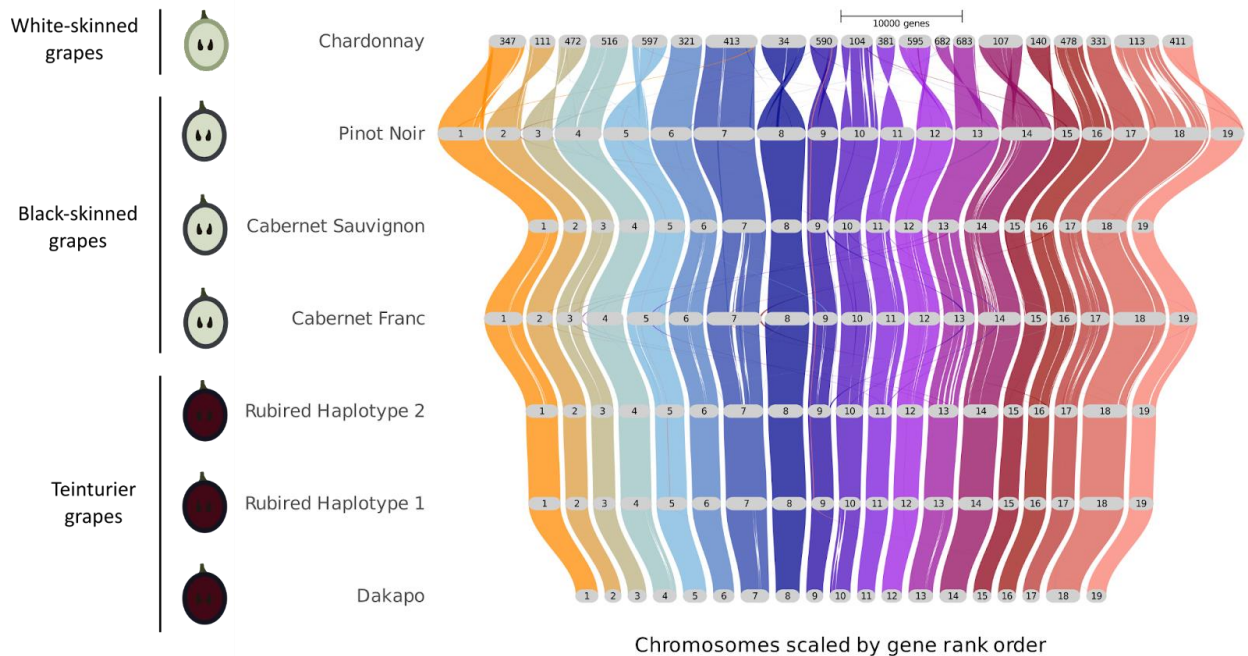


Figure 4.3. Synteny of several grapevine genomes with chromosome-scale assemblies, organized by berry and flesh color.

Initial comparisons of these assemblies to other grapevine genomes even revealed a large (1.82 Mbp) inversion within Dakapo (Figure 4.4) that contains 274 genes (Table S4.5).

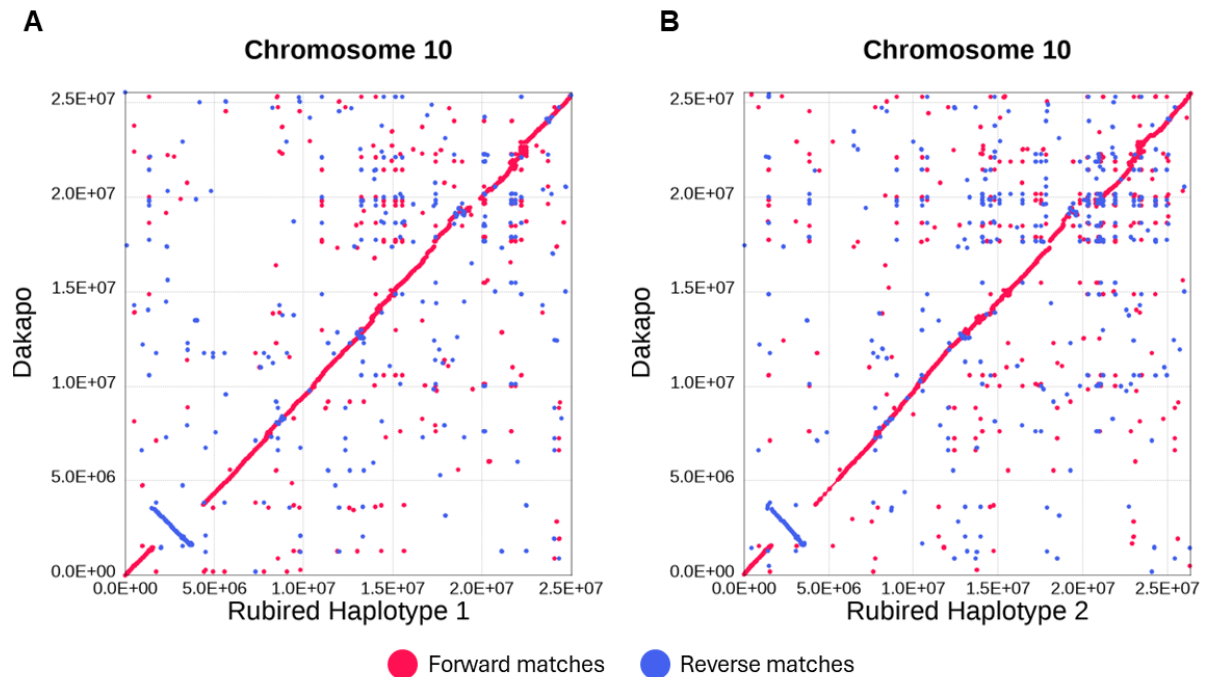


Figure 4.4. Alignment of chromosome 10 in Dakapo versus chromosome 10 in **A)** the Rubired haplotype 1 assembly and **B)** the Rubired haplotype 2 assembly, showing the 1.82 Mbp inversion present in Dakapo. The dotplots show forward matches in red and reverse matches in blue.

Inversions can cause changes in gene expression depending on various genetic factors (Kollar et al., 2023; Loveland et al., 2021; Puig et al., 2015), so we were interested in whether the Dakapo chromosome 10 inversion could contribute to Dakapo's increased cold susceptibility and/or increased pathogen susceptibility. Several genes within the inversion do appear to be involved in cold- and/or pathogen-responsive pathways, including VvDak_v1.10g0003381, whose *Arabidopsis* ortholog (AT3G07650) regulates the expression of genes within the cold acclimation pathway (Li et al., 2021), and VvDak_v1.10g0003951, whose *Arabidopsis* and rice orthologs (AT4G03960 and *OsPFA-DSP2*, respectively) negatively regulate pathogen response pathways (He et al., 2012). The implications of this

inversion remain unclear; however, future research could unveil the potential phenotypic impacts of this inversion.

Grapevine is a useful model system due to its unique life and domestication history and is one of few lianas (woody vines) with robust genomic resources. In addition, grapevine breeding and propagation have been ongoing for millennia, resulting in a fascinating array of phenotypes and an abundance of accumulated somatic variants. The assemblies and annotations of the Dakapo and Rubired genomes add to a growing number of grapevine genomes that will provide valuable tools for both grapevine breeders and geneticists.

ACKNOWLEDGMENTS

We are grateful to Dan Chitwood, Emily Josephs, and Robin Buell for helpful discussions on this work and feedback on this manuscript. We are grateful for the Genomics Core at Michigan State University, the Institute for Cyber-Enabled Research at Michigan State University, and the UC Davis DNA Technology Core for their services. We would like to thank Kevin Childs for providing guidance and custom scripts for genome annotation. We acknowledge Rosa Figueroa-Balderas for processing the samples, extracting the nucleic acids, and preparing the sequencing libraries for the Rubired genome. The Dakapo genome work was supported by Michigan State University and the USDA National Institute of Food and Agriculture MICL02572. Oxford Nanopore Technologies provided sequencing for the Dakapo assembly. The Rubired genome was supported by E. & J. Gallo Winery and the NSF grant #1741627.

REFERENCES

- 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(1), 258.
- Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>. Accessed 18 October 2023.
- Assembly-Stats*. (2020). <https://github.com/sanger-pathogens/assembly-stats>. Accessed 1 January 2024.
- Azuma, A., Kobayashi, S., Goto-Yamamoto, N., Shiraishi, M., Mitani, N., Yakushiji, H., & Koshita, Y. (2009). Color recovery in berries of grape (*Vitis vinifera* L.) “Benitaka”, a bud sport of “Italia”, is caused by a novel allele at the *VvmybA1* locus. *Plant Science*, 176(4), 470–478.
- Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E., & Huala, E. (2015). The Arabidopsis information resource: Making and mining the “gold standard” annotated reference plant genome. *Genesis*, 53(8), 474–485.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120.
- Broad Institute. (2017). Picard Toolkit. Broad Institute. <https://broadinstitute.github.io/picard/>. Accessed 20 September 2023.
- Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 12(1), 59–60.
- California Department of Food and Agriculture. (2023). *California Grape Crush Final Report March 10, 2023*.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10, 421.
- Canaguier, A., Grimplet, J., Di Gaspero, G., Scalabrin, S., Duchêne, E., Choisne, N., Mohellibi, N., Guichard, C., Rombauts, S., Le Clainche, I., Bérard, A., Chauveau, A., Bounon, R., Rustenholz, C., Morgante, M., Le Paslier, M.-C., Brunel, D., & Adam-Blondon, A.-F. (2017). A new version of the grapevine reference genome assembly (12X.v2) and of its annotation (VCost.v3). *Genomics Data*, 14, 56–62.
- Cantu Lab. (n.d.). *Vitis vinifera* cv. Pinot Noir cl. FPS123. <https://www.grapegenomics.com/pages/VvPinNoir/VvPinNoir123/>. Accessed 16 February 2024.

- Cheng, C.-Y., Krishnakumar, V., Chan, A. P., Thibaud-Nissen, F., Schobel, S., & Town, C. D. (2017). Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *The Plant Journal*, *89*(4), 789–804.
- 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.
- Chen, M.-J. M., Lin, H., Chiang, L.-M., Childers, C. P., & Poelchau, M. F. (2019). The GFF3toolkit: QC and Merge Pipeline for Genome Annotation. *Methods in Molecular Biology*, *1858*, 75–87.
- Chin, C.-S., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion, G. T., Clum, A., Dunn, C., O'Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., Cramer, G. R., Delledonne, M., Luo, C., Ecker, J. R., Cantu, D., Rank, D. R., & Schatz, M. C. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. *Nature Methods*, *13*(12), 1050–1054.
- Cochetel, N., Minio, A., Guarracino, A., Garcia, J. F., Figueroa-Balderas, R., Massonnet, M., Kasuga, T., Londo, J. P., Garrison, E., Gaut, B. S., & Cantu, D. (2023). A super-pangenome of the North American wild grape species. *Genome Biology*, *24*(1), 290.
- Cochetel, N., Minio, A., Massonnet, M., Vondras, A. M., Figueroa-Balderas, R., & Cantu, D. (2021). Diploid chromosome-scale assembly of the *Muscadinia rotundifolia* genome supports chromosome fusion and disease resistance gene expansion during *Vitis* and *Muscadinia* divergence. *G3*, *11*(4).
- De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., & Van Broeckhoven, C. (2018). NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics*, *34*(15), 2666–2669.
- De Coster, W., & Rademakers, R. (2023). NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics*, *39*(5).
- Doligez, A., Bouquet, A., Danglot, Y., Lahogue, F., Riaz, S., Meredith, C., Edwards, K., & This, P. (2002). Genetic mapping of grapevine (*Vitis vinifera* L.) applied to the detection of QTLs for seedlessness and berry weight. *Theoretical and Applied Genetics*, *105*(5), 780–795.
- Dong, Y., Duan, S., Xia, Q., Liang, Z., Dong, X., Margaryan, K., Musayev, M., Goryslavets, S., Zdunić, G., Bert, P.-F., Lacombe, T., Maul, E., Nick, P., Bitskinashvili, K., Bisztray, G. D., Drori, E., De Lorenzis, G., Cunha, J., Popescu, C. F., ... Chen, W. (2023). Dual domestications and origin of traits in grapevine evolution. *Science*, *379*(6635), 892–901.
- Doster, M. A. (1985). Effects of leaf maturity and cultivar resistance on development of the powdery mildew fungus on grapevines. *Phytopathology*, *75*(3), 318.

- Emms, D. M., & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*(1), 238.
- FAO. (2023). *Agricultural production statistics 2000–2022* (FAOSTAT Analytical Briefs No. 79).
- Flamini, R., Mattivi, F., De Rosso, M., Arapitsas, P., & Bavaresco, L. (2013). Advanced knowledge of three important classes of grape phenolics: anthocyanins, stilbenes and flavonols. *International Journal of Molecular Sciences*, *14*(10), 19651–19669.
- Fröhling, B., Patz, C. D., Dietrich, H., & Will, F. (2012). Anthocyanins, total phenolics and antioxidant capacities of commercial red grape juices, black currant and sour cherry nectars. *Fruit Process*, *3*, 100–104.
- He, F., Mu, L., Yan, G.-L., Liang, N.-N., Pan, Q.-H., Wang, J., Reeves, M. J., & Duan, C.-Q. (2010). Biosynthesis of anthocyanins and their regulation in colored grapes. *Molecules*, *15*(12), 9057–9091.
- He, H., Su, J., Shu, S., Zhang, Y., Ao, Y., Liu, B., Feng, D., Wang, J., & Wang, H. (2012). Two homologous putative protein tyrosine phosphatases, OsPFA-DSP2 and AtPFA-DSP4, negatively regulate the pathogen response in transgenic plants. *PLoS One*, *7*(4), e34995.
- Herzog, K., Schwander, F., Kassemeyer, H.-H., Bieler, E., Dürrenberger, M., Trapp, O., & Töpfer, R. (2021). Towards Sensor-Based Phenotyping of Physical Barriers of Grapes to Improve Resilience to Botrytis Bunch Rot. *Frontiers in Plant Science*, *12*, 808365.
- Holt, C., & Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*, *12*, 491.
- Jaillon, O., Aury, J.-M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., Vezzi, A., Legeai, F., Huguency, P., Dasilva, C., Horner, D., Mica, E., Jublot, D., Poulain, J., Bruyère, C., ... Wincker, P. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, *449*(7161), 463–467.
- 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. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240.
- Kawahara, Y., de la Bastide, M., Hamilton, J. P., Kanamori, H., McCombie, W. R., Ouyang, S., Schwartz, D. C., Tanaka, T., Wu, J., Zhou, S., Childs, K. L., Davidson, R. M., Lin, H., Quesada-Ocampo, L., Vaillancourt, B., Sakai, H., Lee, S. S., Kim, J., Numa, H., ...

- Matsumoto, T. (2013). Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice*, 6(1), 4.
- Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, 37(8), 907–915.
- Kobayashi, S., Goto-Yamamoto, N., & Hirochika, H. (2004). Retrotransposon-induced mutations in grape skin color. *Science*, 304(5673), 982.
- Kollar, L. M., Stanley, L. E., Raju, S. K. K., Lowry, D. B., & Niederhuth, C. E. (2023). The role of breakpoint mutations, supergene effects, and ancient nested rearrangements in the evolution of adaptive chromosome inversions in the yellow monkey flower, *Mimulus guttatus*. In *bioRxiv* (p. 2023.12.06.570460).
<https://doi.org/10.1101/2023.12.06.570460>
- Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology*, 37(5), 540–546.
- Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinformatics*, 5, 59.
- Kőrösi, L., Molnár, S., Teszlák, P., Dörnyei, Á., Maul, E., Töpfer, R., Marosvölgyi, T., Szabó, É., & Röckel, F. (2022). Comparative Study on Grape Berry Anthocyanins of Various Teinturier Varieties. *Foods*, 11(22).
- Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D. L., Garcia-Hernandez, M., Karthikeyan, A. S., Lee, C. H., Nelson, W. D., Ploetz, L., Singh, S., Wensel, A., & Huala, E. (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research*, 40(Database issue), D1202–D1210.
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. In *arXiv [q-bio.GN]*. arXiv. <http://arxiv.org/abs/1303.3997>
- Li, H. (2021). New strategies to improve minimap2 alignment accuracy. *Bioinformatics*, 37(23), 4572–4574.
- Lisek, J. (2012). Winter frost injury of buds on one-year-old grapevine shoots of cultivars and interspecific hybrids in Poland. *Folia Horticulturae*, 24(1), 97–103.
- Li, Y., Shi, Y., Li, M., Fu, D., Wu, S., Li, J., Gong, Z., Liu, H., & Yang, S. (2021). The CRY2-COP1-HY5-BBX7/8 module regulates blue light-dependent cold acclimation in Arabidopsis. *The Plant Cell*, 33(11), 3555–3573.
- Loveland, J. L., Lank, D. B., & Küpper, C. (2021). Gene Expression Modification by an Autosomal Inversion Associated With Three Male Mating Morphs. *Frontiers in Genetics*, 12, 641620.

- Lovell, J. T., Sreedasyam, A., Schranz, M. E., Wilson, M., Carlson, J. W., Harkess, A., Emms, D., Goodstein, D. M., & Schmutz, J. (2022). GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *eLife*, *11*.
- Manni, M., Berkeley, M. R., Seppey, M., Simao, F. A., & Zdobnov, E. M. (2021). BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. In *arXiv [q-bio.GN]*. arXiv. <http://arxiv.org/abs/2106.11799>
- Marçais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L., & Zimin, A. (2018). MUMmer4: A fast and versatile genome alignment system. *PLoS Computational Biology*, *14*(1), e1005944.
- Matus, J. T., Cavallini, E., Loyola, R., Höll, J., Finezzo, L., Dal Santo, S., Violet, S., Commisso, M., Roman, F., Schubert, A., Alcalde, J. A., Bogs, J., Ageorges, A., Tornielli, G. B., & Arce-Johnson, P. (2017). A group of grapevine MYBA transcription factors located in chromosome 14 control anthocyanin synthesis in vegetative organs with different specificities compared with the berry color locus. *The Plant Journal*, *91*(2), 220–236.
- Minio, A., Cochetel, N., Figueroa-Balderas, R., & Cantu, D. (2024). *Grapegenomics.com - Genome release: Vitis vinifera cv. Chardonnay cl. 04 v2.0*. <https://doi.org/10.5281/zenodo.10578344>
- Minio, A., Cochetel, N., Massonnet, M., Figueroa-Balderas, R., & Cantu, D. (2022). HiFi chromosome-scale diploid assemblies of the grape rootstocks 110R, Kober 5BB, and 101–14 Mgt. *Scientific Data*, *9*(1), 1–8.
- Minio, A., Cochetel, N., Vondras, A. M., Massonnet, M., & Cantu, D. (2022). Assembly of complete diploid-phased chromosomes from draft genome sequences. *G3*, *12*(8).
- Minio, A., Massonnet, M., Figueroa-Balderas, R., Castro, A., & Cantu, D. (2019). Diploid Genome Assembly of the Wine Grape Carménère. *G3*, *9*(5), 1331–1337.
- Minio, A., Massonnet, M., Figueroa-Balderas, R., Vondras, A. M., Blanco-Ulate, B., & Cantu, D. (2019). Iso-Seq Allows Genome-Independent Transcriptome Profiling of Grape Berry Development. *G3*, *9*(3), 755–767.
- Ou, S., Chen, J., & Jiang, N. (2018). Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Research*, *46*(21), e126.
- Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J. R. A., Hellinga, 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*(1), 275.

- Parpinello, G. P., Versari, A., Chinnici, F., & Galassi, S. (2009). Relationship among sensory descriptors, consumer preference and color parameters of Italian Novello red wines. *Food Research International*, 42(10), 1389–1395.
- Puig, M., Casillas, S., Villatoro, S., & Cáceres, M. (2015). Human inversions and their functional consequences. *Briefings in Functional Genomics*, 14(5), 369–379.
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842.
- Rashed, A., Daugherty, M. P., & Almeida, R. P. P. (2011). Grapevine genotype susceptibility to *Xylella fastidiosa* does not predict vector transmission success. *Environmental Entomology*, 40(5), 1192–1199.
- Rashed, A., Kwan, J., Baraff, B., Ling, D., Daugherty, M. P., Killiny, N., & Almeida, R. P. P. (2013). Relative susceptibility of *Vitis vinifera* cultivars to vector-borne *Xylella fastidiosa* through time. *PLoS One*, 8(2), e55326.
- Ritter, E. J., Cousins, P., Quigley, M., Kile, A., Kenchanmane Raju, S. K., Chitwood, D. H., & Niederhuth, C. (2023). From buds to shoots: Insights into grapevine development from the Witch's Broom bud sport. In *bioRxiv* (p. 2023.09.25.559343). <https://doi.org/10.1101/2023.09.25.559343>
- Roach, M. J., Schmidt, S. A., & Borneman, A. R. (2018). Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics*, 19(1), 460.
- Röckel, F., Moock, C., Braun, U., Schwander, F., Cousins, P., Maul, E., Töpfer, R., & Hausmann, L. (2020). Color Intensity of the Red-Fleshed Berry Phenotype of *Vitis vinifera* Teinturier Grapes Varies Due to a 408 bp Duplication in the Promoter of *VvmybA1*. *Genes*, 11(8).
- Sáenz-Navajas, M.-P., Echavarri, F., Ferreira, V., & Fernández-Zurbano, P. (2011). Pigment composition and color parameters of commercial Spanish red wine samples: linkage to quality perception. *European Food Research and Technology*, 232(5), 877–887.
- Shumate, A., & Salzberg, S. L. (2021). Liftoff: accurate mapping of gene annotations. *Bioinformatics*, 37(12), 1639–1643.
- Shumate, A., Wong, B., Pertea, G., & Pertea, M. (2022). Improved transcriptome assembly using a hybrid of long and short reads with StringTie. *PLoS Computational Biology*, 18(6), e1009730.
- Slater, G. S. C., & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, 6, 31.

- SRA Toolkit Development Team. (2020). *sra-tools*. <https://github.com/nbi/sra-tools>. Accessed 21 January 2024.
- Stanke, M., Diekhans, M., Baertsch, R., & Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*, *24*(5), 637–644.
- Statistics Department of the International Organisation of Vine and Wine. (2022). *Annual Assessment of the World Vine and Wine Sector in 2022*. International Organisation of Vine and Wine.
- Titus Brown, C. (2018). *Detecting microbial contamination in long-read assemblies (from known microbes)*. <http://ivory.idyll.org/blog/2018-detecting-contamination-in-long-read-assemblies.html>. Accessed 17 January 2023.
- Titus Brown, C., & Irber, L. (2016). sourmash: a library for MinHash sketching of DNA. *Journal of Open Source Software*, *1*(5), 27.
- UniProt Consortium. (2023). UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, *51*(D1), D523–D531.
- Vaser, R., Sović, I., Nagarajan, N., & Šikić, M. (2017). Fast and accurate *de novo* genome assembly from long uncorrected reads. *Genome Research*, *27*(5), 737–746.
- Velt, A., Frommer, B., Blanc, S., Holtgräwe, D., Duchêne, É., Dumas, V., Grimplet, J., Huguency, P., Kim, C., Lahaye, M., Matus, J. T., Navarro-Payá, D., Orduña, L., Tello-Ruiz, M. K., Vitulo, N., Ware, D., & Rustenholz, C. (2023). An improved reference of the grapevine genome reasserts the origin of the PN40024 highly homozygous genotype. *G3*, *13*(5).
- Walker, A. R., Lee, E., Bogs, J., McDavid, D. A. J., Thomas, M. R., & Robinson, S. P. (2007). White grapes arose through the mutation of two similar and adjacent regulatory genes. *The Plant Journal*, *49*(5), 772–785.
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, *9*(11), e112963.
- Wallis, C. M., Wallingford, A. K., & Chen, J. (2013). Effects of cultivar, phenology, and *Xylella fastidiosa* infection on grapevine xylem sap and tissue phenolic content. *Physiological and Molecular Plant Pathology*, *84*, 28–35.
- Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., Lee, T.-H., Jin, H., Marler, B., Guo, H., Kissinger, J. C., & Paterson, A. H. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, *40*(7), e49.

- Wick, R. R., Judd, L. M., Gorrie, C. L., & Holt, K. E. (2017). Completing bacterial genome assemblies with multiplex MinION sequencing. *Microbial Genomics*, 3(10), e000132.
- Zhang, K., Du, M., Zhang, H., Zhang, X., Cao, S., Wang, X., Wang, W., Guan, X., Zhou, P., Li, J., Jiang, W., Tang, M., Zheng, Q., Cao, M., Zhou, Y., Chen, K., Liu, Z., & Fang, Y. (2023). The haplotype-resolved T2T genome of teinturier cultivar Yan73 reveals the genetic basis of anthocyanin biosynthesis in grapes. *Horticulture Research*, 10(11), uhad205.
- Zhou, Y., Minio, A., Massonnet, M., Solares, E., Lv, Y., Beridze, T., Cantu, D., & Gaut, B. S. (2019). The population genetics of structural variants in grapevine domestication. *Nature Plants*, 5(9), 965–979.
- Zou, C., Karn, A., Reisch, B., Nguyen, A., Sun, Y., Bao, Y., Campbell, M. S., Church, D., Williams, S., Xu, X., Ledbetter, C. A., Patel, S., Fennell, A., Glaubitz, J. C., Clark, M., Ware, D., Londo, J. P., Sun, Q., & Cadle-Davidson, L. (2020). Haplotyping the *Vitis* collinear core genome with rhAmpSeq improves marker transferability in a diverse genus. *Nature Communications*, 11(1), 413.

APPENDIX A: CHAPTER 4 SUPPLEMENT

Table S4.1. Summary of sequencing statistics for reads used to assemble and polish the Dakapo and Rubired genomes. Due to length, this table is available in the supplemental files.

Table S4.2. SRA IDs of RNA-seq files used for annotating the Dakapo genome. Due to length, this table is available in the supplemental files.

Table S4.3. Repeat classes of transposable elements and repetitive sequences annotated in the Dakapo genome. Due to length, this table is available in the supplemental files.

Table S4.4. Repeat classes of transposable elements and repetitive sequences annotated in the diploid Rubired whole genome assembly. Due to length, this table is available in the supplemental files.

Table S4.5. Genes (including pseudogenes) within the chromosome 10 inversion of the Dakapo grapevine variety. Due to length, this table is available in the supplemental files.

CHAPTER 5: Concluding remarks

THE WITCH'S BROOM BUD SPORT IN GRAPEVINE

In Chapter 2, I investigated the genetic basis of the Witch's Broom (WB) bud sport in grapevine and characterized the phenotypes of two independent cases of WB (i.e., the Dakapo and Merlot varieties) over developmental time. Unexpectedly, the results suggest that bud sports described as "Witch's Broom" in grapevine may in fact refer to distinct bud sports with superficially similar but distinct phenotypes. This finding demonstrates the importance and usefulness of thoroughly characterizing lab-generated and natural mutants like WB. In addition, I identified a strong candidate gene for the WB in Merlot: GSVIVG01008260001. An essential next step for this work is using genetic transformation to knock out this gene and characterize the resulting phenotype in order to prove the genetic basis of this case of WB. In Dakapo, additional work comparing gene expression between various tissues in Dakapo WT and Dakapo WB could potentially narrow down the causal gene and, at the very least, give insight into the genetic pathways altered in the development of Dakapo WB.

DOMATIA IN *VITIS*

In Chapter 3, I identified key genetic pathways likely involved in regulating the development of domatia in *Vitis riparia*. I identified probable regulators of trichome development, including several C2H2 ZFPs and SPLs. Genetic transformation of these genes would clarify whether they are involved in trichome development in *Vitis*, and, if so, whether they are domatia-specific or responsible for trichome formation across the leaf surface. We also found evidence that auxin and JA signaling regulate domatia development.

While a link between auxin and domatia development was expected given auxin's involvement in both regulating trichome development (Xuan et al., 2020) and cell elongation (Velasquez et al., 2016), JA has not been implicated in domatia in *Vitis* to my knowledge. JA could play a number of potential roles in domatia, including regulating trichome development (Han et al., 2022), the synthesis of volatiles that could attract mites (Ament et al., 2004; Degenhardt et al., 2010; Schmelz et al., 2003), and/or direct defense pathways (Baldwin, 2010). Future work applying JA to *V. riparia* plants and examining changes in domatia phenotypes as well as mite recruitment would clarify the exact impacts JA has on domatia.

Beyond identifying potential regulators of domatia development, the findings of this work revealed possible aspects of domatia biology that beg for future investigation. For one, xylan biosynthesis appears to be upregulated in developing domatia, and it remains unclear whether this is due to xylan within the trichomes or vascular tissue being present beneath domatia in *Vitis*. Future work investigating the biochemical composition of trichomes in *Vitis* would determine if xylan is present. Regardless, this finding demonstrates that detailed microscopy of *Vitis* domatia would help improve our understanding of domatia development and function. In addition, this work suggests but does not confirm that domatia produce volatiles. Given these findings and the fact that plants commonly produce volatiles to attract beneficial arthropods (Baldwin, 2010), it is likely that *Vitis* domatia are producing volatiles. Future work collecting and analyzing volatiles from domatia would clarify whether domatia produce volatiles, making way for research aimed at understanding their role (i.e., attracting beneficial mites, among other possibilities) and the possible variation in volatiles produced among genotypes, species, and/or biotic and abiotic

conditions. Finally, the overlap of genes involved in domatia and EFN development, including those involved in carbohydrate and amino acid transport and/or metabolism, and the upregulation of several floral genes in developing domatia, raise additional questions warranting future investigation. For one, without detailed characterization of cellular-level *Vitis* domatia morphologies, it remains impossible to know whether *Vitis* domatia could produce food, in addition to shelter, to the beneficial mites that inhabit them. Past work using microscopy to characterize *Plectroniella armata* domatia anatomy provided evidence for material exchange between plants and mites despite it previously being thought that no material exchange occurred in this system (Tilney et al., 2012). Domatia in *Vitis* could be similar, and detailed imaging of *Vitis* domatia may help shed light on this matter. Additionally, floral genes may have been coopted for domatia development, as has been shown in *Vitis* gall and tendril development (Gerrath et al., 2015; Schultz et al., 2019). Investigating domatia development in other *Vitis* species would help identify the core genes involved in domatia development and possibly clarify how domatia evolved, including whether floral pathways were coopted for domatia development.

This work also suggested a strong link between leaf and domatia development that may drive intraspecific variation in domatia phenotypes within *V. riparia*. Additional work analyzing leaf shapes and domatia traits within *V. riparia* and even between different *Vitis* species could further clarify the drivers of this relationship.

Overall, our findings revealed new questions regarding domatia development and functioning in *V. riparia*, demonstrating the need for detailed morphological and biochemical studies on domatia in *Vitis*. Additional studies in other *Vitis* species will also be critical for characterizing how domatia evolved within the genus.

TEINTURIER GRAPEVINE VARIETIES

In Chapter 4, I presented the assembly and annotation of two teinturier grapevine varieties, including the Dakapo variety, which I assembled and annotated. These genomes and annotations will provide valuable resources for future investigations by grapevine researchers. The assembly of the Dakapo genome revealed a large, unexpected inversion in chromosome 10 that was not present in other publicly available chromosome-scale assemblies for domesticated grapevine varieties. However, another teinturier grapevine variety, Yan73, also has a large inversion within chromosome 10 (Zhang et al., 2023). The length and location of this inversion in Yan73 were not reported, so it is impossible to determine whether this inversion is the same as the one present in Dakapo without access to the Yan73 genome. However, future genetic work could clarify this inversion's origin and impact on the expression of the genes within it. In particular, sequencing the parents of Dakapo—Deckrot and Blauer Portugieser—would reveal if this inversion was inherited from either parent or if it is a novel insertion in Dakapo resulting from either chromosome breakage or ectopic recombination. Additionally, measuring the expression of genes within the inversion in Dakapo, compared to a variety without the inversion, would determine whether the inversion is suppressing, enhancing, or having no impact on gene expression levels. Ultimately, these genomes will help facilitate research in important grapevine traits and provide exciting avenues for more generally understanding mutations and genome evolution in grapevine.

REFERENCES

- Ament, K., Kant, M. R., Sabelis, M. W., Haring, M. A., & Schuurink, R. C. (2004). Jasmonic acid is a key regulator of spider mite-induced volatile terpenoid and methyl salicylate emission in tomato. *Plant Physiology*, *135*(4), 2025–2037.
- Baldwin, I. T. (2010). Plant volatiles. *Current Biology*, *20*(9), R392–R397.
- Degenhardt, D. C., Refi-Hind, S., Stratmann, J. W., & Lincoln, D. E. (2010). Systemin and jasmonic acid regulate constitutive and herbivore-induced systemic volatile emissions in tomato, *Solanum lycopersicum*. *Phytochemistry*, *71*(17–18), 2024–2037.
- Gerrath, J., Posluszny, U., & Melville, L. (2015). *Taming the Wild Grape*. Springer International Publishing.
- Han, G., Li, Y., Yang, Z., Wang, C., Zhang, Y., & Wang, B. (2022). Molecular Mechanisms of Plant Trichome Development. *Frontiers in Plant Science*, *13*, 910228.
- Schmelz, E. A., Alborn, H. T., & Tumlinson, J. H. (2003). Synergistic interactions between volicitin, jasmonic acid and ethylene mediate insect-induced volatile emission in *Zea mays*. *Physiologia Plantarum*, *117*(3), 403–412.
- Schultz, J. C., Edger, P. P., Body, M. J. A., & Appel, H. M. (2019). A galling insect activates plant reproductive programs during gall development. *Scientific Reports*, *9*(1), 1833.
- Tilney, P. M., van Wyk, A. E., & van der Merwe, C. F. (2012). Structural evidence in *Plectroniella armata* (Rubiaceae) for possible material exchange between domatia and mites. *PloS One*, *7*(7), e39984.
- Velasquez, S. M., Barbez, E., Kleine-Vehn, J., & Estevez, J. M. (2016). Auxin and Cellular Elongation. *Plant Physiology*, *170*(3), 1206–1215.
- Xuan, L., Yan, T., Lu, L., Zhao, X., Wu, D., Hua, S., & Jiang, L. (2020). Genome-wide association study reveals new genes involved in leaf trichome formation in polyploid oilseed rape (*Brassica napus* L.). *Plant, Cell & Environment*, *43*(3), 675–691.
- Zhang, K., Du, M., Zhang, H., Zhang, X., Cao, S., Wang, X., Wang, W., Guan, X., Zhou, P., Li, J., Jiang, W., Tang, M., Zheng, Q., Cao, M., Zhou, Y., Chen, K., Liu, Z., & Fang, Y. (2023). The haplotype-resolved T2T genome of teinturier cultivar Yan73 reveals the genetic basis of anthocyanin biosynthesis in grapes. *Horticulture Research*, *10*(11), uhad205.