2022

Development of Genomic Resources in *Vitis Riparia* for Discoveries on Pre- And Post-Transcriptional Molecular Regulators of Early Induction into Endodormancy

Michael Robben
South Dakota State University, mikerobbens@gmail.com

DEVELOPMENT OF GENOMIC RESOURCES IN *VITIS RIPARIA* FOR DISCOVERIES ON PRE- AND POST-TRANSCRIPTIONAL MOLECULAR REGULATORS OF EARLY INDUCTION INTO ENDODORMANCY

BY

MICHAEL ROBBEN

A dissertation submitted in partial fulfillment of the requirements for

the Doctor of Philosophy

Major in Biological Sciences

Specialization in Plant Molecular Biology

South Dakota State University

2022

DISSERTATION ACCEPTANCE PAGE

Michael W Robben

This dissertation is approved as a creditable and independent investigation by a candidate for the Doctor of Philosophy degree and is acceptable for meeting the dissertation requirements for this degree. Acceptance of this does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department.

Anne Fennell
Advisor Date

David Wright
Department Head Date

Nicole Lounsbery, PhD
Director, Graduate School Date

ACKNOWLEDGEMENTS

I would like to thank Dr. Anne Fennell for all the help that she has provided me during my Ph.D. here at South Dakota State University. Her insight has made me a better scientist and professional and her help will be vital to my further career in the field of molecular biology. I also would like to thank my committee members past and present, Dr. Jose Gonzalez, Dr. Sen Subramanian, Dr. Qin Ma, and Dr. Sanjeev Anand for all their help in shaping my graduate studies. Dr. Gonzalez and Dr. Subramanian especially for their assistance as co-mentors during my graduate research. I have received the help and advice of my fellow lab mates over the course of my graduate studies and would like to also thank Dr. Dilmini Alahakoon, Dr. Turhan Yılmaz, Dr. Shuchi Smita, Dr. Sagar Patel, Priya Swaminathan, Karoline

Woodhouse, Zachary Helget, Seyma Bokuz, Roberto Villegaz-Diaz, and Prakriti Sharma. Lastly, I would like to thank the National Science Foundation for their financial support through the grant “NSF/EPSCoR Cooperative Agreement #IIA-1355423the South Dakota Agriculture Experiment Station, BioSNTR, South Dakota Research and Innovation Center and the National Institute of Food and Agriculture, U.S. Department of Agriculture, Hatch project under SD00H449-12”.

TABLE OF CONTENTS

ABBREVIATIONS

%	Percent
°C	Degrees Celsius
ABA	Abscisic Acid
AP2	Apetala 2
ARF	Auxin Response Factor
bHLH	Basic Helix Loop Helix
Bp	Base Pairs
CBF	C-repeat Binding Factor
ChIP-seq	Chromatin Immunoprecipitation sequencing
CMT	Chromomethylase
COR	Cold Regulated Genes
DAM	Dormancy Associated MADS-box
DCL1	Dicer-Like 1
DEG	Differentially Expression Genes
DML3	Demeter-Like 3
DNA	Deoxyribonucleic Acids
DOG1	Delay of Germination 1
ERF	Ethylene Response Factor
F2	2nd Generation Progeny
FDR	False Discovery Rate
fMRI	Functional Magnetic Resonance Imaging
DREB	Dehydration Response Element Binding Factor
FT	Flowering Time
GA	Gibberellic Acid
GBS	Genotype-by-Sequencing
GSEA	Gene Set Enrichment Analysis
HAT5	Histone Acetyltransferase
HSTFB2B	Heat Shock Transcription Factor B 2B
IAA	Indole 3'-Acetic Acid
ICE	Inducer of CBF
JA	Jasmonic Acid
Kb	Kilobase (1,000 bp)
LD	Long Day
LFC	Log Fold Change
LFY	Leafy
LTR	Long Terminal Repeat
Mb	Megabase (1,000,000 bp)
min	Minutes
miRNA	microRNA

mRNA	Messenger RNA
Mya	Million Years Ago
N50	Median length of Contigs
ncRNA	Non-coding RNA
Nt	Nucleotide
PCA	Principle Component Analysis
PHYA	Phytochrome A
PLT	Plethora
qPCR	Quantitative Polymerase Chain Reaction
QTL	Quantitative Trait Loci
RISC	RNA-induced Silencing Complex
RNA	Ribonucleic Acid
RNA-seq	RNA-sequencing
SAM	Shoot Apical Meristem
SD	Short Day
SNP	Single Nucleotide Polymorphism
SPL	Squamosa Promoter-Like
spp.	Species
SVM	Support Vector Machine
SVP/AGL	Short Vegetative Phase/Agamous-Like
TCP	Teosinte Branched
TF	Transcription Factor
TSS	Transcription Start Site
VRN1	Vernalization 1
WGCNA	Weighted Gene Co-Expression Analysis
µg	Microgram
µL	microliter

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

DEVELOPMENT OF GENOMIC RESOURCES IN *VITIS RIPARIA* FOR THE INVESTIGATION OF MOLECULAR REGULATORS OF EARLY INDUCTION INTO ENDODORMANCY, A STUDY OF PRE- AND POST-TRANSCRIPTIONAL REGULATION

MICHAEL ROBBEN

2022

Grapevine is one of the most important fruit crops in the world, responsible for billions in global sales annually. The largest threat to grapevine and other crop production is global climate change resulting human activities. This brings unpredictable and drastic changes in ambient air temperatures to many climates in which grapes are grown. Lower temperatures and inclement weather are already responsible for millions in lost revenue due to tissue damage of established plants. Thus, protecting grapevine crops from weather-related damage is the biggest concern to growers aside from pathogen- and disease-related crop damage. The primary mechanism for winter survival in woody perennial plants is bud endodormancy, a state of hibernation that is activated in response to decreasing temperatures and photoperiod. The current understanding of this process is limited, but it is believed that induction into

endodormancy is controlled by a combination of hormones and transcriptional regulators internal to the cell.

Grapevines have variable resistance to cold depending on species. Of the approximately 80 identified grapevine species, North American and Asian grapevines have more enhanced winter survival. *Vitis riparia,* the riverbank grapevine, is one of the most resistant of the genus and has been identified to enter endodormancy at longer day lengths. Investigating why *V. riparia* responds differently may reveal key genes and molecular mechanisms needed for photoperiod induced endodormancy induction. To investigate this species-specific response, we first sought to establish a genome assembly for this non-model species. Sequencing and assembly of DNA from *V. riparia* resulted in 69,616 scaffolds at an N50 of 518,740. Reference, mapping, and non-homologous estimates of misassembly suggest that this draft assembly is of a high quality. cDNA sequence prediction from multiple RNA-seq studies resulted in 40,019 genes. Variations in gene families demonstrated that there were genetic differences between *V. riparia* and *V. vinifera* which could explain the difference in response to photoperiod and winter survival.

One of the best indicators in plants of the physiological response to external regulators is changes in gene expression. We measured changes in expression during endodormancy transition in two F_2 genotypes at multiple time periods of exposure to short day (SD, 13h) and long day (LD, 15h) photoperiods. Expression of genes associated with cell cycle control and phenylpropanoid biosynthesis were downregulated in response to SD treatment. The F_2-110

genotype which more closely resembled *V. riparia* had greater natural expression of auxin signaling genes than the F_2-040. This was further confirmed by co-expression networks that were highly correlated with short day induced endodormancy transition and F2-110 genotypes.

Regulation of endodormancy induction is a primary concern for this study. We performed small-RNA seq to find miRNA that were differentially regulated during dormancy transition. A machine learning based prediction of miRNA identified 216 regulatory sequences in the non-model *V. riparia* genome. We found that miRNA families 166 and 167 were predominantly upregulated during dormancy transition. This coincided with downregulation of cell cycle control genes and suppression of cyclins and expansins by the MYB3R1 transcription factor. Motif enrichment of gene co-expression clusters identified PLETHORA 1 as a major regulator of the stem cell state during dormant conditions.

These results suggest that auxin is a major regulator of endodormancy through control of cell differentiation in the bud apical meristem. Auxin signaling may therefore also be a contributor to the enhanced dormancy response in *V. riparia* due to an increased sensitivity to auxin in the buds. Further research is needed to determine auxin's role in regulation of the process of endodormancy and what effect it has in crop winter survival.

1 Introduction and literature review

1.1 Vitis Riparia

Grapevines are one of the most economically important fruit crops in the world. They are grown on every continent in a wide variety of climates and conditions (Arnold & Schnitzler, 2020). All grapevines are fruit-bearing woody perennial vine plants that belong to the genus *Vitis*, found naturally growing in North America, Asia, Europe and Africa (Keller, 2015). The majority of grapevine used in viticulture belongs to the species *Vitis vinifera*, a European native plant that has been selected for thousands of years in wine production (NASS, 2021).

With the exception of *V. labrusca*, North American species are generally not used in fruit production because they lack desirable production of metabolites. However, North American species are economically important rootstock vines because of their propensity for disease resistance to biotrophic agents, like those that cause powdery mildew disease, and insect pathogen disease (Qiu et al., 2015; Smith et al., 2018). Phylloxera is a soil pest pathogen that was introduced to Europe through trade with the Americas, which resulted in the loss of a significant amount of French vineyards in the mid-1800's (Granett et al., 2001; Ordish, 1972). Pathogenic infection is characterized by the adult insect invading roots and infecting the plant with a potent toxin that often results in tissue death and necrosis (Granett et al., 2001). This epidemic was mitigated by grafting scion tissue to the root system of resistant North American species such

as *V. riparia, Vitis rupestris*, and *Vitis berlandieri*, and is still the process used today to prevent phylloxera infestation.

Although hybridization with phylloxera resistant North American species was attempted, it generally introduced undesirable aromas in the wine, and so grafting is still predominantly used to protect vineyards from the pathogen (Yin et al., 2019). North American species are the most used rootstock, with 90% of vinifera grafted onto 10 genotypes (Gautier et al., 2020). *V. riparia* is especially represented in rootstock genotypes with 52% of them having *V. riparia* ancestry.

1.2 Endodormancy and cold tolerance

Each year grape vines are lost to winter death because they have limited resistance to cold. Up to 75% of primary bud loss has been observed in certain cultivars grown in the Midwest during winter months (Atucha et al., 2018). Grapevines protect bud tissue through a process of differentiation and protective changes made to the bud prior to the winter season called endodormancy. Endodormancy is the process in which grapevines shut down growing tissue for the winter in specialized bud organs (Arora et al., 2003). In most species of trees, the plant sets terminal buds, however, in grapevine, axillary buds on the primary nodes undergo differentiation to the endodormant state (Keller, 2015; Leduc et al., 2014). After winter, shoot apical meristems (SAM) break out of the hard bud scale and develop into new vines that bear inflorescence for the growing season.

Endodormancy is regulated by signals internal to the bud, but is initiated by a combination of photoperiod and temperature (LANG & G. A, 1987).

Wake and Fennell (Wake & Fennell, 2000), demonstrated this in the North American grapevine *V. riparia*, showing that short day (SD) photoperiod exposure will result in full bud endodormancy and cessation of paradormant growth. While some species are able to induce dormancy through decreased temperature alone (Li et al., 2005), an experiment in black cottonwood demonstrated that short pulses of red light during dark phase were enough to inhibit dormancy in a genotypic dependent manner indicating that photoperiod is a major contributor (Howe et al., 1996). Likewise, PHYA is a major light sensitive protein implicated in endodormancy and has been associated with changes in abscisic acid (ABA) and ethylene levels (Rohde et al., 2002; Ruonala et al., 2006). Experiments in Birch have shown that expression of cold tolerance related genes is enhanced under SD photoperiods (Puhakainen et al., 2004).

Multiple gene pathways and regulatory networks are associated with cold tolerance (Cooke et al., 2012);(Wisniewski et al., 2018). Physiological impacts of cold tolerance genes include overcooling of intracellular fluids, acclimation to cold, production of cryoprotectants and dormancy (Gusta & Wisniewski, 2013; Wisniewski et al., 2018). A major pathway in cold tolerance acquisition involves the CBF transcription factor which controls expression of Cold Regulated (COR) genes (Park et al., 2018). Expression of C-REPEAT BINDING FACTOR (CBF) is

controlled by multiple pathways that conduct signals received by cold, light, and clock responses in the cell (Shi et al., 2018).

Tissue damage due to early deacclimation is a present threat to grape growers (Pagter & Arora, 2013) especially as climate change alters average winter temperatures worldwide. A study in China showed that in particularly cold years, vineyards can lose up to 23% of vines (Li, 2015). This makes breeding cold resistant traits into commercial varieties the focus of some studies (Wang et al., 2020). Certain species of grapevine are naturally more resistant to the cold, with *V. riparia* showing greater winter survival than other European and North American species (Londo & Martinson, 2015). There are multiple possibilities that could contribute to this phenotypic resistance. fMRI studies of *V. riparia* axillary buds show earlier bud tissue formation and undercooling during dormancy which allows the plant to survive temperatures as low as -37 C (Fennell & Line, 2001). Other labs have identified possible genetic sources of enhanced cold resistance. The CBF4 gene is a unique DEHYDRATION RESPONSE ELEMENT (DREB) that was identified in *V. riparia* and *V. vinifera* and may have an effect on cold resistance in the species (Dong et al., 2013; Xiao et al., 2008).

1.3 Genome assembly

Large scale genomic studies of endodormancy require accurate genome assemblies of the species of interest. Sequencing of plant genomes has become more common for genomic studies in crop science thanks to decreasing costs of sequencing and more training in bioinformatics. Sequencing of the first plant

species began in the 2000's and the genomes for rice and maize serve as the basis for many genomic studies in a variety of species (International Rice Genome Sequencing Project, 2005; Ming et al., 2008; Schnable et al., 2009). To this day, 798 unique plant genomes have been fully sequenced, yet this represents less than 0.2% of known extant plant species (Marks et al., 2021). *V. vinifera* was one of the first plant genomes to be sequenced and assembled back in 2007 (Jaillon et al., 2007). There are many efforts currently to sequence more genomes in the genus *Vitis*, however, they are focused on species and cultivars involved in viticulture (Vondras et al., 2019). Such sequencing efforts have allowed researchers to discover quantitative and qualitative genetic traits in grapevine such as those that control sex determination (Zou et al., 2021).

The first eukaryotic genome ever assembled, yeast, was put together using a combination of shotgun sequencing and brute force assembly (Giani et al., 2020). Since then, sequencing technology has been improved, allowing researchers to sequence millions of reads at once at a high accuracy of read quality. The reference sequence genome for *V. vinifera* is based on the 'PN40024' accession of Pinot Noir (Jaillon et al., 2007). In a measure of heterozygosity, this genome was shown to be much more homozygous than *V. riparia* which increased the ease of assembly (Patel et al., 2018). Short read Illumina sequencing has been shown to conserve less heterozygosity than long read technologies, but recent assembly trials have shown that certain assembler

technologies can conserve the genetic differences between homologous chromosomes (Kajitani et al., 2019).

1.4 RNA seq

Gene expression during endodormancy transition is the most important indicator of changes made for differentiation and cold resistance. The post-transcriptional content of mRNA within the bud indicates which genes are important for endodormancy because of their increased or decreased expression in relation to paradormant buds. In the past, such indicators were monitored by techniques which reverse transcribed mRNA into cDNA on a per gene basis and measured levels of that gene through techniques such as blots and qPCR (Raso & Biassoni, 2014). With the advent of next generation high-throughput sequencing, new techniques which leverage big data enable researchers to evaluate the expression of all genes simultaneously (Raso & Biassoni, 2014; Wang et al., 2009).

Multiple studies over the past two decades have utilized this technology to examine differential expression in endodormancy. Meta-analysis of RNA-Seq experiments in Prunus species shows common responses to endodormancy transition (Canton et al., 2021). Many differentially expressed genes were connected to hormone regulation and response including ABA, Ethylene, and Auxin. Other RNA sequencing studies in aspen (Böhlenius et al., 2006), poplar (Olsen, 2010), peony (Mornya & Cheng, 2011), and leafy spurge (Doğramacı et al., 2013) found that gene expression of photoperiod, circadian clock, and

hormone genes were upregulated under dormant conditions. A comparative RNA-seq analysis of different dormancy states in grapevine found that genes differentially expressed in endodormant compared to paradormant conditions were enriched for ABA and GA hormone regulation pathways (Khalil-Ur-Rehman et al., 2019). DORMANCY ASSOCIATED MADS-BOX transcription factors (DAMs) are differentially expressed at great amounts during endodormancy and are found differentially expressed in transcriptomic studies of bud break in grapevines (Shangguan et al., 2020). Further transcriptomic studies will reiterate what is already known and reveal new genes and pathways involved in regulation of endodormancy.

1.5 miRNA

Regulation of endodormancy transition likely can occur post-transcriptionally as has been shown before in flowering (Wisniewski et al., 2018). miRNAs are short 20-24 nt RNA sequences that are the reverse complement of sections of transcribed mRNAs (B. Zhang et al., 2006). These small RNA sequences regulate gene expression post-transcriptionally by binding to their corresponding mRNA molecules and either signaling them for destruction or preventing them from being translated into proteins. miRNAs are derived from larger pieces of transcribed pre-miRNAs, which are processed into the regulatory form by DICER-LIKE 1 (DCL1) and taken into the RNA-induced Silencing Complex (RISC). The processing of small RNA sequences makes prediction of

regulatory sequences difficult but a library of miRNAs was established in *V. vinifera* by small RNA sequencing (Belli Kullan et al., 2015).

miRNAs have been confirmed to play roles in regulation of endodormancy. In studies done in peach and pear, miRNAs 6285 and 6390 have been found to negatively regulate the expression of CBF, DAMs and Abscisic Acid (ABA) related genes (Niu et al., 2016; Yu et al., 2021). miR156, miR159, and miR167 are differentially expressed in all of these studies as well as a dormancy study in peony (Y. Zhang et al., 2018). These miRNAs have been linked to regulation of SQUAMOSA promoter-binding protein-like (SPL) and APETALA 2 (AP2) genes which control tissue differentiation and flowering. Similar miRNA and genes were identified in a small-RNA sequencing experiment done in Tea (Jeyaraj et al., 2014; Qu et al., 2021). Much of the knowledge of miRNA regulation has been inferred from similar studies in seed dormancy (Huo et al., 2016), and more large-scale analysis of miRNA expression in bud dormancy are needed to identify important miRNAs.

1.6 Transcription factors

Genes necessary for regulation of endodormancy are likely controlled transcriptionally by cis-regulatory binding elements (Liu et al., 1999). Transcription factors are DNA binding proteins that attach to chromatin upstream of specific gene segments. This either promotes or suppresses RNA polymerase activity at the Transcription Start Site (TSS). Transcription factors are typically activated by signaling pathways in response to binding of cell surface receptors.

Certain signaling pathways and cis-regulatory elements are associated with endodormancy. Common transcription factors differentially expressed in many endodormancy data sets are transcription factors of the MADS-Box family (Canton et al., 2021; Moser et al., 2020; Niu et al., 2016). Certain MADS-Box transcription factors are colloquially known as Dormancy Associated MADS-Box genes (Bielenberg et al., 2004, 2008). Expression of these SVP1/AGL24-like transcription factors are needed to regulate expression of FLOWERING TIME (FT) and LEAFY (LFY) which control clock dependent cell-cycle regulation (Cooke et al., 2012). However, while PHYTOCHROME (PHYA, PHYB, and PHYC) signaling is associated with short day photoperiod (Kozarewa et al., 2010), no signaling pathway has been directly associated with photoperiod sensing in bud tissue.

1.7 References

Arnold, C., & Schnitzler, A. (2020). Ecology and Genetics of Natural Populations of North American Vitis Species Used as Rootstocks in European Grapevine Breeding Programs. *Frontiers in Plant Science*, *11*, 866.

Arora, R., Rowland, L. J., & Tanino, K. (2003). Induction and release of bud dormancy in woody perennials: A science comes of age. *HortScience: A Publication of the American Society for Horticultural Science*, *38*(5), 911–921.

Atucha, A., Hedtcke, J., Workmaster, B. A., & Others. (2018). Evaluation of

cold-climate interspecific hybrid wine grape cultivars for the upper Midwest. *J. Am. Pomol. Soc*, *72*, 80–93.

Belli Kullan, J., Lopes Paim Pinto, D., Bertolini, E., Fasoli, M., Zenoni, S., Tornielli, G. B., Pezzotti, M., Meyers, B. C., Farina, L., Pè, M. E., & Mica, E. (2015). miRVine: a microRNA expression atlas of grapevine based on small RNA sequencing. *BMC Genomics*, *16*, 393.

Bielenberg, D. G., Wang, Y. (eileen), Li, Z., Zhebentyayeva, T., Fan, S., Reighard, G. L., Scorza, R., & Abbott, A. G. (2008). Sequencing and annotation of the evergrowing locus in peach [Prunus persica (L.) Batsch] reveals a cluster of six MADS-box transcription factors as candidate genes for regulation of terminal bud formation. *Tree Genetics & Genomes*, *4*(3), 495–507.

Bielenberg, D. G., Wang, Y., Fan, S., Reighard, G. L., Scorza, R., & Abbott, A. G. (2004). A deletion affecting several gene candidates is present in the Evergrowing peach mutant. *The Journal of Heredity*, *95*(5), 436–444.

Böhlenius, H., Huang, T., Charbonnel-Campaa, L., Brunner, A. M., Jansson, S., Strauss, S. H., & Nilsson, O. (2006). CO/FT regulatory module controls timing of flowering and seasonal growth cessation in trees. *Science*, *312*(5776), 1040–1043.

Canton, M., Forestan, C., Bonghi, C., & Varotto, S. (2021). Meta-analysis of RNA-Seq studies reveals genes with dominant functions during flower bud endo- to eco-dormancy transition in Prunus species. *Scientific Reports*, *11*(1), 13173.

Cooke, J. E. K., Eriksson, M. E., & Junttila, O. (2012). The dynamic nature of bud dormancy in trees: environmental control and molecular mechanisms. *Plant, Cell & Environment*, *35*(10), 1707–1728.

Doğramacı, M., Foley, M. E., Chao, W. S., Christoffers, M. J., & Anderson, J. V. (2013). Induction of endodormancy in crown buds of leafy spurge (Euphorbia esula L.) implicates a role for ethylene and cross-talk between photoperiod and temperature. *Plant Molecular Biology*, *81*(6), 577–593.

Dong, C., Zhang, M., Yu, Z., Ren, J., Qin, Y., Wang, B., Xiao, L., Zhang, Z., & Tao, J. (2013). Isolation and expression analysis of CBF4 from Vitis amurensis associated with stress. *Agricultural Sciences in China / Sponsored by the Chinese Academy of Agricultural Sciences*, *04*(05), 224–229.

Fennell, A., & Line, M. J. (2001). Identifying differential tissue response in grape (Vitis riparia) during induction of endodormancy using nuclear magnetic resonance imaging. *Journal of the American Society for Horticultural Science. American Society for Horticultural Science*, *126*(6), 681–688.

Gautier, A., Cookson, S. J., Lagalle, L., Ollat, N., & Marguerit, E. (2020). Influence of the three main genetic backgrounds of grapevine rootstocks on petiolar nutrient concentrations of the scion, with a focus on phosphorus. *OENO One*, *54*(1), 1–13.

Giani, A. M., Gallo, G. R., Gianfranceschi, L., & Formenti, G. (2020). Long walk to genomics: History and current approaches to genome sequencing

and assembly. *Computational and Structural Biotechnology Journal*, *18*, 9–19.

Granett, J., Walker, M. A., Kocsis, L., & Omer, A. D. (2001). Biology and management of grape phylloxera. *Annual Review of Entomology*, *46*, 387–412.

Gusta, L. V., & Wisniewski, M. (2013). Understanding plant cold hardiness: an opinion. *Physiologia Plantarum*, *147*(1), 4–14.

Howe, G. T., Gardner, G., Hackett, W. P., & Furnier, G. R. (1996). Phytochrome control of short-day-induced bud set in black cottonwood. *Physiologia Plantarum*, *97*(1), 95–103.

Huo, H., Wei, S., & Bradford, K. J. (2016). DELAY OF GERMINATION1 (DOG1) regulates both seed dormancy and flowering time through microRNA pathways. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(15), E2199–E2206.

International Rice Genome Sequencing Project. (2005). The map-based sequence of the rice genome. *Nature*, *436*(7052), 793–800.

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., Hugueney, 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.

Jeyaraj, A., Chandran, V., & Gajjeraman, P. (2014). Differential expression of microRNAs in dormant bud of tea [Camellia sinensis (L.) O. Kuntze]. *Plant Cell Reports*, *33*(7), 1053–1069.

Kajitani, R., Yoshimura, D., Okuno, M., Minakuchi, Y., Kagoshima, H., Fujiyama, A., Kubokawa, K., Kohara, Y., Toyoda, A., & Itoh, T. (2019). Platanus-allee is a de novo haplotype assembler enabling a comprehensive access to divergent heterozygous regions. *Nature Communications*, *10*(1), 1702.

Keller, M. (2015). *The Science of Grapevines: Anatomy and Physiology*. Academic Press.

Khalil-Ur-Rehman, M., Wang, W., Dong, Y., Faheem, M., Xu, Y., Gao, Z., Guo Shen, Z., & Tao, J. (2019). Comparative Transcriptomic and Proteomic Analysis to Deeply Investigate the Role of Hydrogen Cyanamide in Grape Bud Dormancy. *International Journal of Molecular Sciences*, *20*(14). https://doi.org/10.3390/ijms20143528

Kozarewa, I., Ibáñez, C., Johansson, M., Ogren, E., Mozley, D., Nylander, E., Chono, M., Moritz, T., & Eriksson, M. E. (2010). Alteration of PHYA expression change circadian rhythms and timing of bud set in Populus. *Plant Molecular Biology*, *73*(1-2), 143–156.

LANG, & G. A. (1987). Dormancy : A new universal terminology. *HortScience: A Publication of the American Society for Horticultural Science*, *25*, 817–820.

Leduc, N., Roman, H., Barbier, F., Péron, T., Huché-Thélier, L., Lothier, J.,

Demotes-Mainard, S., & Sakr, S. (2014). Light Signaling in Bud Outgrowth and Branching in Plants. *Plants*, *3*(2), 223–250.

Li, S. (2015). Grapevine breeding and genetics in China: History, current status and the future. *Acta Horticulturae*, *1082*, 165–176.

Li, S., Martin, L. T., Pezeshki, S. R., & Shields, F. D. (2005). Responses of black willow (Salix nigra) cuttings to simulated herbivory and flooding. *Acta Oecologica*, *28*, 173–180.

Liu, L., White, M. J., & MacRae, T. H. (1999). Transcription factors and their genes in higher plants functional domains, evolution and regulation. *European Journal of Biochemistry / FEBS*, *262*(2), 247–257.

Londo, J., & Martinson, T. (2015). Geographic trend of bud hardiness response in Vitis Riparia. *Acta Horticulturae*, *1082*, 299–304.

Marks, R. A., Hotaling, S., Frandsen, P. B., & VanBuren, R. (2021). Representation and participation across 20 years of plant genome sequencing. *Nature Plants*. https://doi.org/10.1038/s41477-021-01031-8

Ming, R., Hou, S., Feng, Y., Yu, Q., Dionne-Laporte, A., Saw, J. H., Senin, P., Wang, W., Ly, B. V., Lewis, K. L. T., Salzberg, S. L., Feng, L., Jones, M. R., Skelton, R. L., Murray, J. E., Chen, C., Qian, W., Shen, J., Du, P., … Alam, M. (2008). The draft genome of the transgenic tropical fruit tree papaya (Carica papaya Linnaeus). *Nature*, *452*(7190), 991–996.

Mornya, P., & Cheng, F. (2011). The levels of hormone and carbohydrate in autumn and non-autumn flowering tree peonies. *Canadian Journal of Plant Science. Revue Canadienne de Phytotechnie*, *91*(6), 991–998.

Moser, M., Asquini, E., Miolli, G. V., Weigl, K., Hanke, M.-V., Flachowsky, H., & Si-Ammour, A. (2020). The MADS-Box Gene MdDAM1 Controls Growth Cessation and Bud Dormancy in Apple. *Frontiers in Plant Science*, *11*, 1003.

NASS. (2021). *California Grape Acreage Report, 2020 Summary*. USDA.

Niu, Q., Li, J., Cai, D., Qian, M., Jia, H., Bai, S., Hussain, S., Liu, G., Teng, Y., & Zheng, X. (2016). Dormancy-associated MADS-box genes and microRNAs jointly control dormancy transition in pear (Pyrus pyrifolia white pear group) flower bud. *Journal of Experimental Botany*, *67*(1), 239–257.

Olsen, J. E. (2010). Light and temperature sensing and signaling in induction of bud dormancy in woody plants. *Plant Molecular Biology*, *73*(1-2), 37–47.

Ordish, G. (1972). *The Great Wine Blight*.

Pagter, M., & Arora, R. (2013). Winter survival and deacclimation of perennials under warming climate: physiological perspectives. *Physiologia Plantarum*, *147*(1), 75–87.

Park, J., Lim, C. J., Shen, M., Park, H. J., Cha, J.-Y., Iniesto, E., Rubio, V., Mengiste, T., Zhu, J.-K., Bressan, R. A., Lee, S. Y., Lee, B.-H., Jin, J. B., Pardo, J. M., Kim, W.-Y., & Yun, D.-J. (2018). Epigenetic switch from repressive to permissive chromatin in response to cold stress. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(23), E5400–E5409.

Patel, S., Lu, Z., Jin, X., Swaminathan, P., Zeng, E., & Fennell, A. Y. (2018). Comparison of three assembly strategies for a heterozygous seedless

grapevine genome assembly. *BMC Genomics*, *19*(1), 57.

Puhakainen, T., Li, C., Boije-Malm, M., Kangasjärvi, J., Heino, P., & Palva, E. T. (2004). Short-day potentiation of low temperature-induced gene expression of a C-repeat-binding factor-controlled gene during cold acclimation in silver birch. *Plant Physiology*, *136*(4), 4299–4307.

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, H., Liu, Y., Jiang, H., Liu, Y., Song, W., & Chen, L. (2021). Identification and characterization of miRNAs associated with sterile flower buds in the tea plant based on small RNA sequencing. *Hereditas*, *158*(1), 26.

Raso, A., & Biassoni, R. (2014). Twenty years of qPCR: a mature technology? *Methods in Molecular Biology* , *1160*, 1–3.

Rohde, A., Prinsen, E., De Rycke, R., Engler, G., Van Montagu, M., & Boerjan, W. (2002). PtABI3 impinges on the growth and differentiation of embryonic leaves during bud set in poplar. *The Plant Cell*, *14*(8), 1885–1901.

Ruonala, R., Rinne, P. L. H., Baghour, M., Moritz, T., Tuominen, H., & Kangasjärvi, J. (2006). Transitions in the functioning of the shoot apical meristem in birch (Betula pendula) involve ethylene. *The Plant Journal: For Cell and Molecular Biology*, *46*(4), 628–640.

Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T. A., Minx, P., Reily, A. D.,

Courtney, L., Kruchowski, S. S., Tomlinson, C., Strong, C., Delehaunty, K., Fronick, C., Courtney, B., … Wilson, R. K. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science*, *326*(5956), 1112–1115.

Shangguan, L., Chen, M., Fang, X., Xie, Z., Gong, P., Huang, Y., Wang, Z., & Fang, J. (2020). Comparative transcriptome analysis provides insight into regulation pathways and temporal and spatial expression characteristics of grapevine (Vitis vinifera) dormant buds in different nodes. *BMC Plant Biology*, *20*(1), 390.

Shi, Y., Ding, Y., & Yang, S. (2018). Molecular Regulation of CBF Signaling in Cold Acclimation. *Trends in Plant Science*, *23*(7), 623–637.

Smith, H. M., Clarke, C. W., Smith, B. P., Carmody, B. M., Thomas, M. R., Clingeleffer, P. R., & Powell, K. S. (2018). Genetic identification of SNP markers linked to a new grape phylloxera resistant locus in Vitis cinerea for marker-assisted selection. *BMC Plant Biology*, *18*(1), 360.

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.

Wake, C. M. F., & Fennell, A. (2000). Morphological, physiological and dormancy responses of three Vitis genotypes to short photoperiod. *Physiologia Plantarum*, *109*(2), 203–210.

Wang, Z., Chai, F., Zhu, Z., Kirabi Elias, G., Xin, H., Liang, Z., & Li, S. (2020). The inheritance of cold tolerance in seven interspecific grape

populations. *Scientia Horticulturae*, *266*, 109260.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews. Genetics*, *10*(1), 57–63.

Wisniewski, M., Nassuth, A., & Arora, R. (2018). Cold Hardiness in Trees: A Mini-Review. *Frontiers in Plant Science*, *9*, 1394.

Xiao, H., Tattersall, E. A. R., Siddiqua, M. K., Cramer, G. R., & Nassuth, A. (2008). CBF4 is a unique member of the CBF transcription factor family of Vitis vinifera and Vitis riparia. *Plant, Cell & Environment*, *31*(1), 1–10.

Yin, L., Clark, M. D., Burkness, E. C., & Hutchison, W. D. (2019). Grape Phylloxera (Hemiptera: Phylloxeridae), on Cold-Hardy Hybrid Wine Grapes (Vitis spp.): A Review of Pest Biology, Damage, and Management Practices. *Journal of Integrated Pest Management*, *10*(1). https://doi.org/10.1093/jipm/pmz011

Yu, J., Bennett, D., Dardick, C., Zhebentyayeva, T., Abbott, A. G., Liu, Z., & Staton, M. E. (2021). Genome-Wide Changes of Regulatory Non-Coding RNAs Reveal Pollen Development Initiated at Ecodormancy in Peach. *Frontiers in Molecular Biosciences*, *8*, 612881.

Zhang, B., Pan, X., Cobb, G. P., & Anderson, T. A. (2006). Plant microRNA: a small regulatory molecule with big impact. *Developmental Biology*, *289*(1), 3–16.

Zhang, Y., Wang, Y., Gao, X., Liu, C., & Gai, S. (2018). Identification and characterization of microRNAs in tree peony during chilling induced dormancy release by high-throughput sequencing. *Scientific Reports*, *8*(1),

4537.

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). https://doi.org/10.1073/pnas.2023548118

2 Assembly and analysis of the *V. riparia* draft genome

2.1 Abstract

Large scale bioinformatic analysis of plant experiments require high quality genome assemblies that have complete genic regions to allow for a wide array of methodologies. In most non-model systems, complete genomes are not available for use in analysis and researchers must rely on reference genomes that may share little to no homology with the genomic system they are interested in. *V. vinifera* is the most common reference sequence in grapevine bioinformatics, yet it is an inbred line with high homozygosity for grapevine species when compared to collected *Vitis* accessions. To create a reference genome that is more representative of heterozygous North American grapevines, an assembly of *V. riparia* Michx 'Manitoba 37' was created from short read sequencing. The assembly resulted in 69,616 contigs with an N50 of 518,740 bp and was found to be of high quality with high between-chromosome homology with *V. vinifera* 'PN40024' and low predicted misassembly by non-homologous methodologies. A gene prediction by RNA-seq data found about 40,000 genes that contained 96% of BUSCO genes. Predicted genes held high homology with *V. vinifera* and variation in transcription factors suggested phenotypic differences between vitis species. Alignment of markers from an F2 mapping population revealed several structural variations and genes involved with important flower

development and summer lateral growth cessation. This is a high-quality genome that can be used for future genomics experiments in *V. riparia* and other North American grapevines.

2.2 Introduction

Plant research has been using genomics more extensively over the past couple of decades. High quality wheat and corn genomes have solved multiple biological questions that are important to crop science and agricultural development (Brenchley et al. 2012; Haberer et al. 2005). Most of the plant genomes that have been sequenced using high-throughput techniques are considered model organisms that represent entire clades or families. However, some crop species important for food production have only just recently undergone whole genome sequencing and assembly (Maccaferri et al. 2019; Edger et al. 2019). *Vitis vinifera*, a European grapevine, was first sequenced back in 2007, and has been used extensively in horticultural genomics research since (Jaillon et al. 2007).

While the reference genome is useful for fruit research, it is insufficient to represent rootstock species of the vitis genus. Rootstocks are grown in every country as a method for phylloxera resistance (Gautier et al. 2020) and are primarily North American in origin. However, these North American species are non-model grapevines and are much more heterozygous than the cultivar of *V. vinifera* used in the reference assembly (PN40024). They are also believed to be evolutionarily divergent from *V. vinifera*, separating from a common ancestor in

North America roughly 12 Million Years Ago (Mya) (Wan et al. 2013). The genotypic difference between North American species like *V. riparia* and the reference genome is likely vast as *V. riparia* shows different fruit qualities, leaf shape, and stress tolerance properties (Hemstad and Luby 2000).

In order to do any type of genomic research in North American rootstock varieties, we require a genome that is more representative of the clade. To remedy this, *V. riparia* Michx 'Manitoba 37' was obtained from the breeding program at the University of Minnesota, (St. Paul, MN, USA) and later placed in the USDA ARS Germplasm Repository at Geneva, New York under the identifier of PI588259. We analyzed an Illumina short-read sequencing assembly of DNA extracted from young leaves. We coupled this with analysis of predicted gene models from RNA-seq data to find sources of increased stress tolerance and rootstock traits. Finally, we use markers from an F2 mapping population to identify structural variants that are evolutionarily distinct to *V. riparia*.

2.3 Materials and Methods

2.3.1 *V. riparia* Michx. 'Manitoba 37' materials

V. riparia Michx. 'Manitoba 37' (identified as 'PI588259' in USDA Germplasm Repository, Geneva, NY, USA) was used for sequencing. The genetic relationship of *V. riparia* 'Manitoba 37' to other *V. riparia* genotypes was analyzed using a data set extracted from genotype data collected from multiple species housed at the Geneva USDA-ARS grape germplasm repository (Klein et al. 2018). To identify highly specific SNPs, VCFtools filters were applied to

156,799 SNPs from 74 unique *V. riparia* genotypes, keeping those found in at least 50% of the individuals, resulting in 54,029 SNPs (Danecek et al. 2011). We then removed six *V. riparia* genotypes with missing data at greater than 30% of the total SNPs. A high stringency filter was applied to the remaining SNPs keeping all SNPs found in 95% or greater of the *V. riparia* providing 1,485 highly specific SNPs. SNPRelate R package (Zheng et al. 2012) was used to calculate the principal components of the specific SNPs data set and plotted the 68 individuals using ggplot2 (Wickham 2016).

2.3.2 DNA sequencing and pre-processing of reads

One centimeter diameter new leaves of greenhouse grown vines were used for DNA extraction and sequencing. A total of nine paired-end libraries were constructed with insert sizes of 346, 473, 478 by Illumina I and 250, 450, 600, 3–5 kb, 8–10 kb, 15–20 kb by Illumina Hiseq 2500 sequencer. In total, 2,295.4 M raw reads were generated with 658.4X coverage and read length from 100 nt–260 nt. The k-mer analysis was carried out with Jellyfish with 19 bp k-mers using only 262.3X coverage of filtered reads. The genome's heterozygosity and other results were obtained with GenomeScope (Vurture et al. 2017). All filtered reads used for *de novo* genome assembly were mapped back to our assembly using *bowtie2* (Langmead and Salzberg 2012). The SAM files of the *bowtie2* mapping results were converted to BAM files using SAMtools, and then the alignment statistics were obtained using the *flagstat* option of SAMtools (Li et al. 2009).

2.3.3 *V. riparia* ‘Manitoba 37’ *de novo* heterozygous genome assembly and assembly evaluation

A total of 1,313.7 M filtered reads were used for de novo genome assembly and constructed with the PLATANUS assembler (Patel et al. 2018; Kajitani et al. 2014). The quality of the assembly was further assessed by three independent methods. (1) The percentage filtered reads were mapped back to the *V. riparia* ‘Manitoba 37’ genome using a zero mismatch. (2) The REAPR program (Hunt et al. 2013) which measures the number of times that there is low mapped mate-paired read coverage of any specific site to predict potential errors in contig assembly. (3) The *V. riparia* ‘Manitoba 37’ assembly quality was further characterized by generating a dot plot of *V. vinifera* ‘PN40024’ and *V. riparia* assembly using the D-genie program (Cabanettes and Klopp 2018) which plotted a sorted and denoised global alignment of the two assemblies.

2.3.4 Plant transcription factors prediction and phylogenetic tree of gene families

Using all predicted protein sequences from *V. riparia* ‘Manitoba 37’ assembly and *V. vinifera* ‘PN40024’ (12X.1, V2 and 12X.2, V3) annotation we predicted Plant transcription factors with PlantTFDB (4.0) (Jin et al. 2017) and compared them directly through motif and phylogenetic analysis. Annotation of subgroups in MYB and ERF TFs was done through a BLAST alignment to the Arabidopsis TFs predicted in PlantTFDB (Dubos et al. 2010; Nakano et al. 2006).

The neighbor-joining tree for the MYB and ERF TFs was generated from a clustal alignment of all genes.

2.3.5 Alignment of F2 GBS markers to *V. riparia* ‘Manitoba 37’ and *V. vinifera* ‘PN40024’ 12X.2

The predicted *V. riparia* ‘Manitoba 37’ gene set and *V. vinifera* ‘PN40024’ annotation were aligned to the PFAM database using HMMer (Finn, Clements, and Eddy 2011). The matching annotations were used to divide genes into families or domains. Then the differences in gene number for each gene family at each position of the corresponding gene on the *V. vinifera* chromosome were plotted with RCircos (Zhang, Meltzer, and Davis 2013). The GBS genetic markers from a F2 mapping population (VRS-F_2), derived from a self of an individual F1 genotype from a cross of *V. riparia* ‘Manitoba 37’ and hybrid cultivar ‘Seyval’ (Yang et al. 2016), were then aligned to *V. vinifera* ‘PN40024’ 12X.2 and a pseudo-chromosomal assembly of *V. riparia* ‘Manitoba 37’. The *V. riparia* pseudo-chromosomal assembly was based on its genomic alignment to *V. vinifera*, using the bowtie2 aligner. R programming was used to find the common marker set and plot the markers mapping to chromosomes using RCircos. The gene containing regions of *V. vinifera* 12X.2 and *V. riparia* ‘Manitoba 37’ were extracted from between QTL markers for previously predicted QTL flanking markers for female sex and summer lateral cessation in response to decreasing photoperiod phenotype on chromosome 2 and scaled to markers shared

between the species. The LOD score of the species-specific markers were then plotted using scaled LOD values to present species protein domain distribution.

2.4 Results

2.4.1 Genetic analysis of *V. riparia* ‘Manitoba 37’

The genetic relationship of *V. riparia* ‘Manitoba 37’ relative to other *V. riparia* in the USDA ARS Germplasm Repository, Geneva, NY USA indicated that ‘Manitoba 37’ is representative of the *V. riparia* collected throughout its native range (Fig 2.1a and 2.1b). Principal component analysis (PCA) of SNP data from 68 *V. riparia* samples demonstrated that *V. riparia* diversity is best described as two separate clusters. Both *V. riparia* ‘Manitoba 37’ and the important rootstock cultivar *V. riparia* ‘Gloire de Montpellier’ are in the primary cluster, and a secondary cluster is made up of species from the Northwestern edge of *V. riparia*’s range. Using Illumina HiSeq (Illumina, USA) short reads and three mate-pair libraries of varying insert sizes, we generated 2,295.4 M raw reads for the *V. riparia* diploid genome draft assembly. A 1.39% heterozygosity was estimated from the unprocessed short reads with the Jellyfish plot showing the heterozygous peak slightly lower than the homozygous peak (Fig 2.2).

2.4.2 Draft assembly and comparison

The draft assembly was a final length of 494.6 Mb among 69,616 scaffolds, with an N50 of 518,740 bp and a scaffold N of 3.57% (Table 2.1). The

V. riparia ‘Manitoba 37’ and *V. vinifera* PN40024 alignment was shown to contain high between-chromosome homology (Fig 2.3a) while the *V. riparia* alignment to itself was near linear (Fig 2.3b), indicating no major erroneous duplication among contigs. In addition to the Assemblathon statistics, 96% of the filtered reads mapped back to the *V. riparia* ‘Manitoba 37’ genome assembly with zero mismatch.

REAPR analysis of assembly accuracy using mate-paired reads found evidence for potential mis-assembly in no more than 16% of the scaffolds. While we found few mistakes in assembly of reads (only 16% of contigs had errors), we did observe low incidence of error free bases (maximum 42.11%). The reason for this low rate is unknown, however, REAPR is a relatively new tool in plant genome development and may not be properly calibrated for the high heterozygosity of this genome. Misassembly events appeared to be fairly linear with scaffold length (Fig 2.4a). Implementing a size cutoff for scaffolds longer than the insert size resulted in a decrease in the number of common misassembly sites between all mate-pair libraries (Fig 2.4a and 2.4b). There were roughly 200 sites of misassembly in each chromosome (Fig 2.5a) and these sites were not colocated with inversions or translocations with the reference sequence (Fig 2.5b).

2.4.3 Analysis of genes and transcription factors

We predicted genes from the assembled genome with various RNA-seq datasets from a multitude of tissues and experiments, resulting in the identification of 40,019 genes (Table 2.1). A BUSCO prediction resulted in 96%

prediction of BUSCO genes which is greater than the 92% found in *V. vinifera* ‘PN40024’. We performed a HMMer annotation of predicted genes and found that there were families that differed in gene copy number between species of vitis. This included greater duplication of gag retrotransposons in *V. riparia* when compared to *V. vinifera* (Fig 2.6a). Phylogenetic analysis revealed that paralogous duplication of retrotransposons was subfamily dependent in each species (Fig 2.6c). This was also confirmed by the difference in copy number of *“DNA integration protein”* annotated genes (Fig 2.6b).

Using data from the PlantTFDB, we predicted transcription factors from *V. riparia* genes. Most families were well conserved with *V. vinifera,* however, some families showed large copy number variation (Patel et al., 2020). TF families like the LATERAL ORGAN BINDING DOMAIN (LBD) family were well conserved but had some variation and we identified genes that did not belong to any known class (Fig 2.6d). Pure counts of ETHYLENE RESPONSE FACTOR (ERF) genes were nearly half that found in *V. vinifera* (Table 2.2). All predicted ERFs fell into 10 subfamilies and aligned to *V. vinifera* ERF genes, and we identified that most of the paralogous duplications in this family occurred in subfamily 9 (Fig 2.7). When we looked at the C-REPEAT BINDING FACTOR (CBF) genes in this family we found that the *V. riparia* CBF genes lacked a hydrophobic motif that was found in *V. vinifera* CBF genes (Fig 2.8). MYB genes were not as drastically different between *V. riparia, V. vinifera* ‘PN40024’, and *V. vinifera* ‘Sultana’ spp. when compared to the ERF family, but we did find duplication of MYBs in subgroup 6 (Fig 2.9).

2.4.4 F2 mapping markers alignment to the *V. riparia* genome

Single Nucleotide Polymorphism (SNP) marker sequences from an F2 mapping population derived from a single F1 (generated by crossing *V. riparia* ‘Manitoba 37’ (female; grandmother) and the cultivar ‘Seyval’ (male; grandfather)) were used to evaluate the utility of the *V. riparia* genome as a reference assembly. Alignment of the SNP marker sequences to *V. riparia* ‘Manitoba 37’ and *V. vinifera* ‘PN40024’ chromosomes indicated that 89.5% of the markers aligned to both species, while about 6.2% and 4.2% mapped uniquely to *V. riparia* or *V. vinifera*, respectively (Fig 2.10a). The number of markers that mapped was not proportional to genome or chromosome size between the two species indicating an even distribution across the genome (Table 2.3). Putative rearrangements, needing further study to verify, were noted on sections of chromosomes 5, 6, and 8, as well as between chromosomes 14 and 15 (Fig 2.10b).

Using the aligned SNP markers and phenotype data for flower sex and the summer lateral shoot cessation photoperiod response, we identified and aligned genes between flanking markers of the respective QTLs (Fig. 2.11a and 2.11b) in the *V. riparia* pseudo-chromosomes and the *V. vinifera* ‘PN40024’ chromosomes. Genes in common between the species with similar position alignment were noted underneath markers with high LOD scores.

2.5 Discussion

The draft genome of the *V. riparia* ‘Manitoba 37’ represents a positive step in bioinformatic research of North American Grapevine species. We have shown that our assembly reflects closely related species of *V. riparia* from different populations by geographical location in the US. This is important as some North American accessions can be more closely related to East Asian ones than other North Americans (Liang et al. 2019). It was also more central to populations than the recently sequenced *V. riparia* ‘Gloire de Montpellier’ (Girollet et al. 2019). Another concern for bioinformatic analysis in Grapevine is the high homozygosity found in the *V. vinifera* ‘PN40024’ reference assembly that does not represent the highly heterozygous nature of other species in the genus (Velasco et al. 2007; Jaillon et al. 2007). Our genome is highly heterozygous when looking at k-mer content, especially when compared to varieties of *V. vinifera* such as ‘Sultanina’ (Patel et al. 2018).

This draft genome sequence, assembled using high quality Illumina reads (>369X coverage), provides a valuable resource for marker development and breeding efforts using wild germplasm. The assembly of 495 Mb in 69,616 scaffolds has an N50 of 518 kb which is greater than the N50 value reported for the *V. vinifera* ‘Sultanina’ genome assembled using Illumina data (Patel et al. 2018). The closest *Vitis* representative to our draft genome is *V. riparia* ‘Riparia Gloire de Montpellier’, a widely used rootstock variety. Interestingly, the long-read genome assembly of the ‘Gloire’ variety reports 33.9% repetitive sequences, far less than the 46% repeat sequences we detected in ‘Manitoba 37’

(which is similar to *V. vinifera* ‘PN40024’ (41.4%)) (Girollet et al. 2019). We observed similar chromosomal collinearity between *V. riparia* ‘Manitoba 37’ and *V. vinifera* ‘PN40024’, which has been similarly demonstrated in the long-read *V. riparia* ‘Gloire de Montpellier’ with many inversions like those on chromosome 6 and 7, conserved between the two *V. riparia* genomes.

Because of these structural variations we were concerned that homology-based estimates of misassembly would not accurately reflect the quality of the genome. To remedy this, we utilized REAPR, a non-homology mapping-based estimate of misassembly to estimate the overall quality of our assembly (Hunt et al. 2013). This demonstrated to us that we had about 200 misassembly events per chromosome but these didn’t appear to coincide with any structural variation and were more likely misjoins from homologous super scaffold construction. Indeed, the strong similarity of the LBD transcription factor family, between *V. riparia* and *V. vinifera* provide further evidence of assembly quality (Grimplet et al. 2017). The validation, completeness and accuracy of *V. riparia* ‘Manitoba 37’ genome features indicate that using only short-read Illumina sequences a high-quality *V. riparia* genome assembly was developed.

V. riparia is typically used in breeding programs to incorporate abiotic stress tolerance traits into new hybrid cultivars. Thus, particular attention was paid to examining the MYB and ERF transcription factors that influence gene regulation and have a strong role in abiotic stress tolerance phenotypes (Nakano et al. 2006). The MYB family of transcription factor genes were explored specifically because of their importance to color, flavor and chemistry in

grapevine species. Phylogenetic reconstruction of MYB subgroups found that the subgroups 4, 13, and 24 that had potential paralogous duplications were involved in ABA response, flavanol synthesis, secondary growth and anther development (Gonzalez et al. 2008). The only homologues with deletions in *V. riparia* was MYB113 of subgroup 6 which has been shown to regulate the production of anthocyanins in a BASIC HELIX-LOOP-HELIX (bHLH) dependent manner. We also noted an increased number of bHLH genes predicted in the *V. riparia* 'Manitoba 37' genome assembly (by > 20 genes). bHLH genes in *V. riparia* are known to produce high amounts of predominantly diglucoside anthocyanin derivatives as opposed to monoglucoside derivatives in *V. vinifera* (Liang et al. 2008). This result adds genomic context to one of the key issues facing acceptance of hybrid derived grape products as the presence of diglucoside derivatives is considered an indication of low quality in hybrid wines (Manns, Coquard Lenerz, and Mansfield 2013). Analysis of ERF genes using alignment and motif comparison between *V. riparia* and *V. vinifera* found functional similarities between members of each subfamily from both genomes. However, there were many instances of duplications present in *V. vinifera* 'PN40024' that were not present in our assembly, such as in subfamily IX where some duplications presented with different motifs. CBF was of particular interest as *V. riparia* CBF4 lacked a hydrophobic motif which could be the reason *V. riparia* has higher cold stress resistance due to altered trans-activation (Xiao et al. 2008; Vazquez-Hernandez et al. 2017).

Retrotransposon activity has long been associated with diversification of species clades. We observed a lower number of genes associated with transposases and retrotransposons in the *V. riparia* ‘Manitoba 37’ and *V. vinifera* ‘Sultanina’ than in the *V. vinifera* 12X.2. All three species seemed to share common ancestors for each Long Terminal Repeat (LTR) retrotransposon gene but experienced paralogous gene duplication at different rates in each clade. We posit that this change in retrotransposons could have played some impact on the divergent evolution of the species, as it has been found previously that Tvv1 transposon markers could accurately distinguish between North American species and *V. vinifera* cultivars (Sant’Ana et al. 2012).

The SNP markers that were developed using Genotype-by-Sequence (GBS) of the F2 population in comparison with *V. vinifera* ‘PN40024’ 12X.1 allowed further analysis of the *V. riparia* ‘Manitoba 37’ assembly. Aligning these SNP markers with the pseudo chromosomes of the grandparents, *V. riparia* ‘Manitoba 37’ and the *V. vinifera* ‘PN40024’ 12X.2 chromosomes showed that the F2 population more closely modeled *V. riparia* ‘Manitoba 37’. This can be expected since informative SNPs were predicted using the F2 grandparents and the male parent ‘Seyval’ has a complex pedigree including *V. vinifera* and other species. The presence of markers that aligned to chromosome 20 of *V. vinifera*, un-assembled scaffolds, but to other chromosomes on *V. riparia* may give us a better indication of the actual genomic position of those scaffolds on *V. vinifera* ‘PN40024’ assembly. By using the markers that mapped to different chromosomes in the female grandparent and *V. vinifera* as a representative

portion of the male grandparent, we found evidence for potential large genomic alterations between these species that may have occurred during the evolution and geographic isolation 3.5–9.5 million years ago. When we look at areas of both genomes containing QTL's we can see that large translocations between chromosomes shows missing genes found between flanking markers in *V. riparia* relative to *V. vinifera*, thus impacting the resulting observed phenotype. This shows the potential power of sequencing and assembling a genetic grandparent of a F2 population in identifying the genetic basis of QTL regions.

In conclusion, we present a high coverage short-read draft genome sequence of the wild grapevine species *V. riparia*. This genome represents the second genome assembly of this critically important species and the first representative of a locally adapted stress tolerant genotype. The *V. riparia* 'Manitoba 37' genome assembly provides an important resource for comparative genomic and genetic marker studies. This *V. riparia* 'Manitoba 37' genome has already proven useful for the development of molecular markers in North American breeding programs and will serve as an important tool in the development of genomics-assisted selection for grapevine improvement, particularly for traits associated with abiotic and biotic stress resistance.

2.6 References

Brenchley, Rachel, Manuel Spannagl, Matthias Pfeifer, Gary L. A. Barker, Rosalinda D'Amore, Alexandra M. Allen, Neil McKenzie, et al. 2012. "Analysis of the Bread Wheat Genome Using Whole-Genome Shotgun

Sequencing.” *Nature* 491 (7426): 705–10.

Cabanettes, Floréal, and Christophe Klopp. 2018. “D-GENIES: Dot Plot Large Genomes in an Interactive, Efficient and Simple Way.” *PeerJ* 6 (June): e4958.

Danecek, Petr, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo, Robert E. Handsaker, et al. 2011. “The Variant Call Format and VCFtools.” *Bioinformatics* 27 (15): 2156–58.

Dubos, Christian, Ralf Stracke, Erich Grotewold, Bernd Weisshaar, Cathie Martin, and Loïc Lepiniec. 2010. “MYB Transcription Factors in Arabidopsis.” *Trends in Plant Science* 15 (10): 573–81.

Edger, Patrick P., Thomas J. Poorten, Robert VanBuren, Michael A. Hardigan, Marivi Colle, Michael R. McKain, Ronald D. Smith, et al. 2019. “Origin and Evolution of the Octoploid Strawberry Genome.” *Nature Genetics* 51 (3): 541–47.

Finn, Robert D., Jody Clements, and Sean R. Eddy. 2011. “HMMER Web Server: Interactive Sequence Similarity Searching.” *Nucleic Acids Research* 39 (Web Server issue): W29–37.

Gautier, Antoine, Sarah J. Cookson, Loïc Lagalle, Nathalie Ollat, and Elisa Marguerit. 2020. “Influence of the Three Main Genetic Backgrounds of Grapevine Rootstocks on Petiolar Nutrient Concentrations of the Scion, with a Focus on Phosphorus.” *OENO One* 54 (1): 1–13.

Girollet, Nabil, Bernadette Rubio, Céline Lopez-Roques, Sophie Valière, Nathalie Ollat, and Pierre-François Bert. 2019. “De Novo Phased Assembly

of the Vitis Riparia Grape Genome." *Scientific Data* 6 (1): 127.

Gonzalez, Antonio, Mingzhe Zhao, John M. Leavitt, and Alan M. Lloyd. 2008. "Regulation of the Anthocyanin Biosynthetic Pathway by the TTG1/bHLH/Myb Transcriptional Complex in Arabidopsis Seedlings." *The Plant Journal: For Cell and Molecular Biology* 53 (5): 814–27.

Grimplet, Jérôme, Diana Pimentel, Patricia Agudelo-Romero, Jose Miguel Martinez-Zapater, and Ana Margarida Fortes. 2017. "The LATERAL ORGAN BOUNDARIES Domain Gene Family in Grapevine: Genome-Wide Characterization and Expression Analyses during Developmental Processes and Stress Responses." *Scientific Reports* 7 (1): 15968.

Haberer, Georg, Sarah Young, Arvind K. Bharti, Heidrun Gundlach, Christina Raymond, Galina Fuks, Ed Butler, et al. 2005. "Structure and Architecture of the Maize Genome." *Plant Physiology* 139 (4): 1612–24.

Hemstad, P. R., and James J. Luby. 2000. "Utilization of Vitis Riparia for the Development of New Wine Varieties with Resistance to Disease and Extreme Cold." In *VII International Symposium on Grapevine Genetics and Breeding*, 487–90. International Society for Horticultural Science.

Hunt, Martin, Taisei Kikuchi, Mandy Sanders, Chris Newbold, Matthew Berriman, and Thomas D. Otto. 2013. "REAPR: A Universal Tool for Genome Assembly Evaluation." *Genome Biology* 14 (5): R47.

Jaillon, Olivier, Jean-Marc Aury, Benjamin Noel, Alberto Policriti, Christian Clepet, Alberto Casagrande, Nathalie Choisne, et al. 2007. "The Grapevine Genome Sequence Suggests Ancestral Hexaploidization in Major

Angiosperm Phyla.” *Nature* 449 (7161): 463–67.

Jin, Jinpu, Feng Tian, De-Chang Yang, Yu-Qi Meng, Lei Kong, Jingchu Luo, and Ge Gao. 2017. “PlantTFDB 4.0: Toward a Central Hub for Transcription Factors and Regulatory Interactions in Plants.” *Nucleic Acids Research* 45 (D1): D1040–45.

Kajitani, Rei, Kouta Toshimoto, Hideki Noguchi, Atsushi Toyoda, Yoshitoshi Ogura, Miki Okuno, Mitsuru Yabana, et al. 2014. “Efficient de Novo Assembly of Highly Heterozygous Genomes from Whole-Genome Shotgun Short Reads.” *Genome Research* 24 (8): 1384–95.

Klein, Laura L., Allison J. Miller, Claudia Ciotir, Katie Hyma, Simon Uribe-Convers, and Jason Londo. 2018. “High-Throughput Sequencing Data Clarify Evolutionary Relationships among North American Vitis Species and Improve Identification in USDA Vitis Germplasm Collections.” *American Journal of Botany* 105 (2): 215–26.

Langmead, Ben, and Steven L. Salzberg. 2012. “Fast Gapped-Read Alignment with Bowtie 2.” *Nature Methods* 9 (4): 357–59.

Liang, Zhenchang, Shengchang Duan, Jun Sheng, Shusheng Zhu, Xuemei Ni, Jianhui Shao, Chonghuai Liu, et al. 2019. “Whole-Genome Resequencing of 472 Vitis Accessions for Grapevine Diversity and Demographic History Analyses.” *Nature Communications* 10 (1): 1190.

Liang, Zhenchang, Benhong Wu, Peige Fan, Chunxiang Yang, Wei Duan, Xianbo Zheng, Chunyan Liu, and Shaohua Li. 2008. “Anthocyanin Composition and Content in Grape Berry Skin in Vitis Germplasm.” *Food*

Chemistry 111 (4): 837–44.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. 2009. “The Sequence Alignment/Map Format and SAMtools.” *Bioinformatics* 25 (16): 2078–79.

Maccaferri, Marco, Neil S. Harris, Sven O. Twardziok, Raj K. Pasam, Heidrun Gundlach, Manuel Spannagl, Danara Ormanbekova, et al. 2019. “Durum Wheat Genome Highlights Past Domestication Signatures and Future Improvement Targets.” *Nature Genetics* 51 (5): 885–95.

Manns, David C., Céline T. M. Coquard Lenerz, and Anna Katharine Mansfield. 2013. “Impact of Processing Parameters on the Phenolic Profile of Wines Produced from Hybrid Red Grapes Maréchal Foch, Corot Noir, and Marquette.” *Journal of Food Science* 78 (5): C696–702.

Nakano, Toshitsugu, Kaoru Suzuki, Tatsuhito Fujimura, and Hideaki Shinshi. 2006. “Genome-Wide Analysis of the ERF Gene Family in Arabidopsis and Rice.” *Plant Physiology* 140 (2): 411–32.

Patel, Sagar, Zhixiu Lu, Xiaozhu Jin, Padmapriya Swaminathan, Erliang Zeng, and Anne Y. Fennell. 2018. “Comparison of Three Assembly Strategies for a Heterozygous Seedless Grapevine Genome Assembly.” *BMC Genomics* 19 (1): 57.

Sant’Ana, G. C., J. L. Ferreira, H. S. Rocha, A. Borém, M. Pasqual, and G. M. A. Cançado. 2012. “Comparison of a Retrotransposon-Based Marker with Microsatellite Markers for Discriminating Accessions of Vitis Vinifera.”

Genetics and Molecular Research: GMR 11 (2): 1507–25.

Vazquez-Hernandez, Maria, Irene Romero, M. I. Escribano, Carmen Merodio, and M. T. Sanchez-Ballesta. 2017. “Deciphering the Role of CBF/DREB Transcription Factors and Dehydrins in Maintaining the Quality of Table Grapes Cv. Autumn Royal Treated with High CO2 Levels and Stored at 0°C.” *Frontiers in Plant Science* 8 (September): 1591.

Velasco, Riccardo, Andrey Zharkikh, Michela Troggio, Dustin A. Cartwright, Alessandro Cestaro, Dmitry Pruss, Massimo Pindo, et al. 2007. “A High Quality Draft Consensus Sequence of the Genome of a Heterozygous Grapevine Variety.” *PloS One* 2 (12): e1326.

Vurture, Gregory W., Fritz J. Sedlazeck, Maria Nattestad, Charles J. Underwood, Han Fang, James Gurtowski, and Michael C. Schatz. 2017. “GenomeScope: Fast Reference-Free Genome Profiling from Short Reads.” *Bioinformatics* 33 (14): 2202–4.

Wan, Yizhen, Heidi R. Schwaninger, Angela M. Baldo, Joanne A. Labate, Gan-Yuan Zhong, and Charles J. Simon. 2013. “A Phylogenetic Analysis of the Grape Genus (Vitis L.) Reveals Broad Reticulation and Concurrent Diversification during Neogene and Quaternary Climate Change.” *BMC Evolutionary Biology* 13 (July): 141.

Wickham, Hadley. 2016. “ggplot2: Elegant Graphics for Data Analysis.” Springer-Verlag New York. https://ggplot2.tidyverse.org.

Xiao, Huogen, Elizabeth A. R. Tattersall, Mahbuba K. Siddiqua, Grant R. Cramer, and Annette Nassuth. 2008. “CBF4 Is a Unique Member of the CBF

Transcription Factor Family of Vitis Vinifera and Vitis Riparia." *Plant, Cell & Environment* 31 (1): 1–10.

Yang, Shanshan, Jonathan Fresnedo-Ramírez, Qi Sun, David C. Manns, Gavin L. Sacks, Anna Katharine Mansfield, James J. Luby, et al. 2016. "Next Generation Mapping of Enological Traits in an F2 Interspecific Grapevine Hybrid Family." *PloS One* 11 (3): e0149560.

Zhang, Hongen, Paul Meltzer, and Sean Davis. 2013. "RCircos: An R Package for Circos 2D Track Plots." *BMC Bioinformatics* 14 (August): 244.

Zheng, Xiuwen, David Levine, Jess Shen, Stephanie M. Gogarten, Cathy Laurie, and Bruce S. Weir. 2012. "A High-Performance Computing Toolset for Relatedness and Principal Component Analysis of SNP Data." *Bioinformatics* 28 (24): 3326–28.

Table 2.1 Assembly statistics for the draft genome of *V. riparia* Michx. 'Manitoba 37' (Patel et al. 2018).

Assembly statistics	Details	*V. riparia assembly (≥500 bp)*
	Number of scaffolds	69,616
	Total size of scaffolds	494,682,949
	Longest scaffold	5,123,774
	Number of scaffolds >1 K nt	31,418
	Number of scaffolds >10 K nt	1760
	Number of scaffolds >100 K nt	742
	Number of scaffolds >1 M nt	97
	Scaffold %N	3.57
	N50 scaffold length	518,740
	NG50 scaffold length	535,518
	N50 contig length	61,142
Gene prediction		
	Total CDS and protein	40,019
	Total CDS bp	39,395,553
	Mean CDS length	984.4
	Longest CDS length	16,443
	Total protein length	13,093,122
	Mean protein length	327.2
	Longest protein length	5480

Table 2.2 Prediction of transcription factors in *V. vinifera* ‘PN40024’ 12x.2 and *V. riparia* Michx ‘Manitoba 37’.

Transcription factor	*V. vinifera*	*V. riparia*
ERF	126	56
AP2	18	21
MYB	152	165
MYB-Related	73	108

Table 2.3 Mapping of GBS markers for F2 marker population on *V. vinifera* PN40024, 12X.2 and *V. riparia* 'Manitoba 37'.

V. vinifera	Total Aligned Markers	*V. riparia*		Total Aligned Markers	Common Markers	*V. vinifera* Markers	*V. riparia* Markers
chr1	784	**chr1**	676		613	171	63
chr2	707	**chr2**	860		671	36	189
chr3	721	**chr3**	701		653	68	48
chr4	1295	**chr4**	1205		1081	214	124
chr5	1224	**chr5**	1106		881	343	225
chr6	944	**chr6**	911		680	264	231
chr7	1093	**chr7**	1076		957	136	119
chr8	1108	**chr8**	1186		959	149	227
chr9	944	**chr9**	993		758	186	235
chr10	501	**chr10**	646		423	78	223
chr11	715	**chr11**	637		563	152	74
chr12	886	**chr12**	1026		851	35	175
chr13	1160	**chr13**	873		763	397	110
chr14	1024	**chr14**	933		702	322	231
chr15	242	**chr15**	504		172	70	332
chr16	828	**chr16**	821		770	58	51
chr17	842	**chr17**	949		783	59	166
chr18	1308	**chr18**	1505		1265	43	240
chr19	739	**chr19**	809		611	128	198
chrUn	0	**chr20**	6		0	0	6
NA	NA	**chr21**	2		NA	NA	2

Figure 2.1 Principal component analysis of informative SNPs in 68 *V. riparia* individuals in the USDA ARS Geneva New York germplasm repository. Symbols represent sample origin by state (United States) and Canadian province as noted in the USDA Germplasm Resource Information Network. Open squares indicate individuals with unknown geographic origin. *V. riparia* 'Manitoba 37' and *V. riparia* 'Gloire de Montpellier' (unknown geographic origin) are represented by blue diamond and black square, respectively. a) Principal components one and two, 10.6% and 5.9% of variation, respectively. b) Presents principal components one and four, 10.6% and 4.5% of variation, respectively.

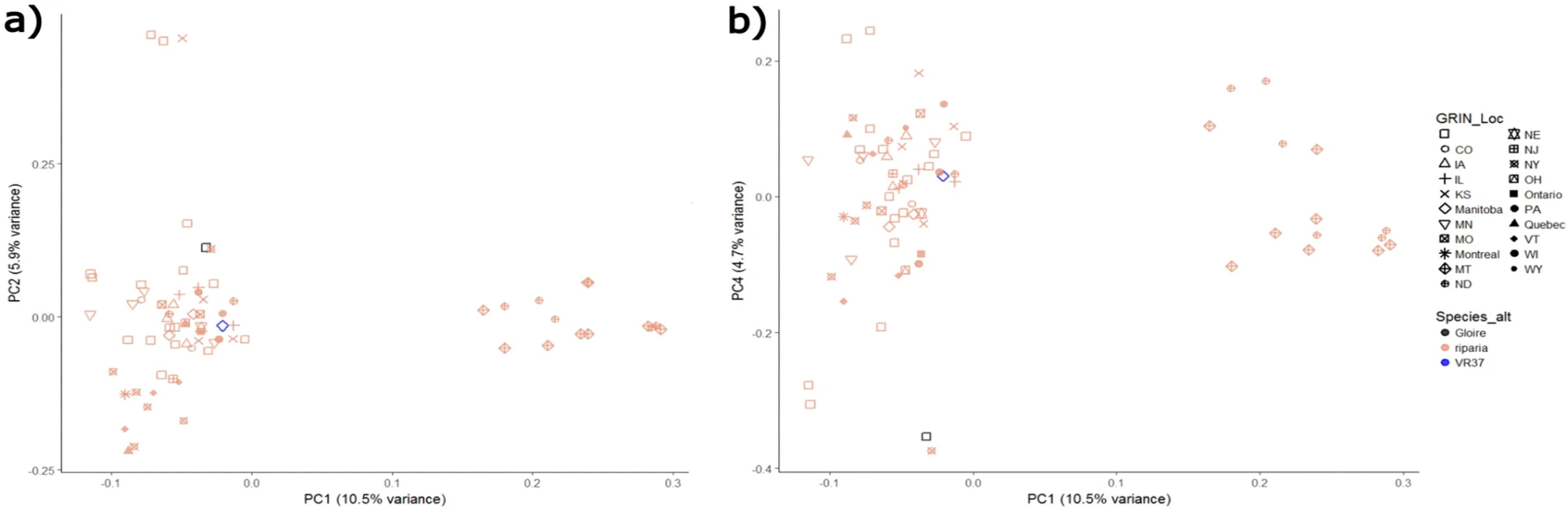


Figure 2.2 K-mer analysis of *V. riparia* ‘Manitoba 37’ genome. The 19 k-mer was carried out with 262.3X coverage by Jellyfish and heterozygosity obtained by GenomeScope. The first peak located at coverage 89X corresponds to the heterozygous peak and the second peak at 184X corresponds to the homozygous peak.

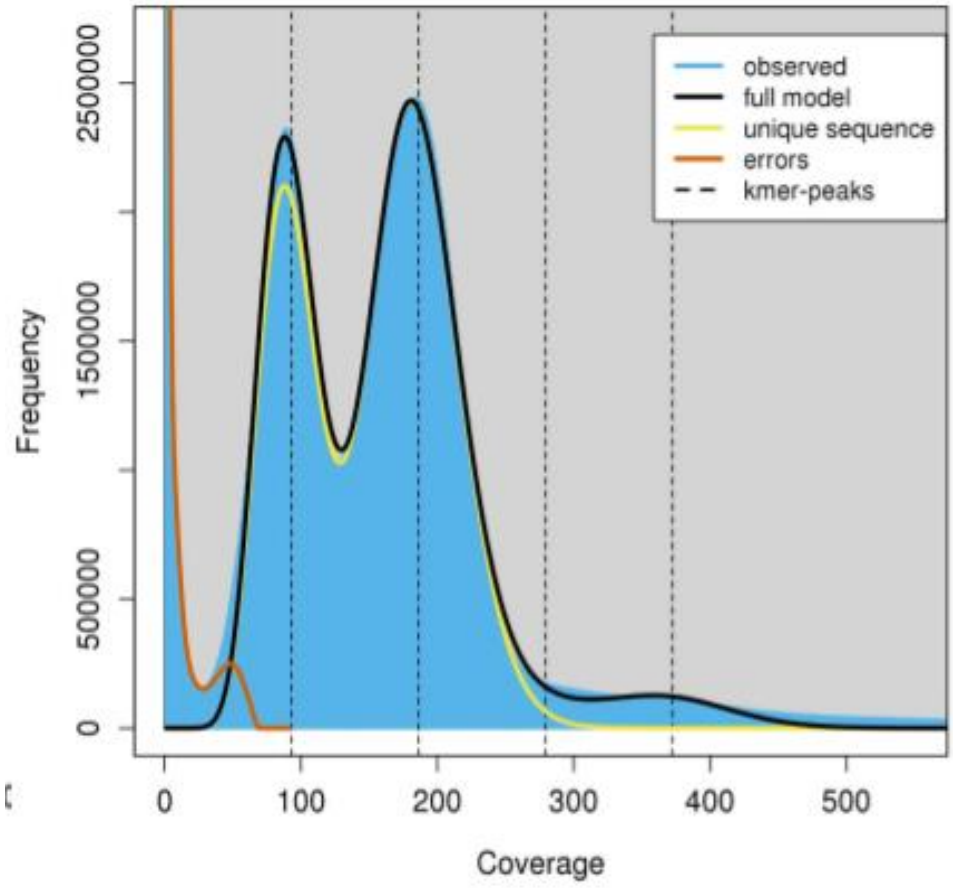


Figure 2.3 Dot plot of global alignment between genomes using D-Genies. Each point represents a homologous alignment predicted with minimap2. a) *V. vinifera* PN40024 on the x-axis and *V. riparia* 'Manitoba 37' on the y-axis. b) *V. riparia* 'Manitoba 37' aligned to self on the x and y axis.

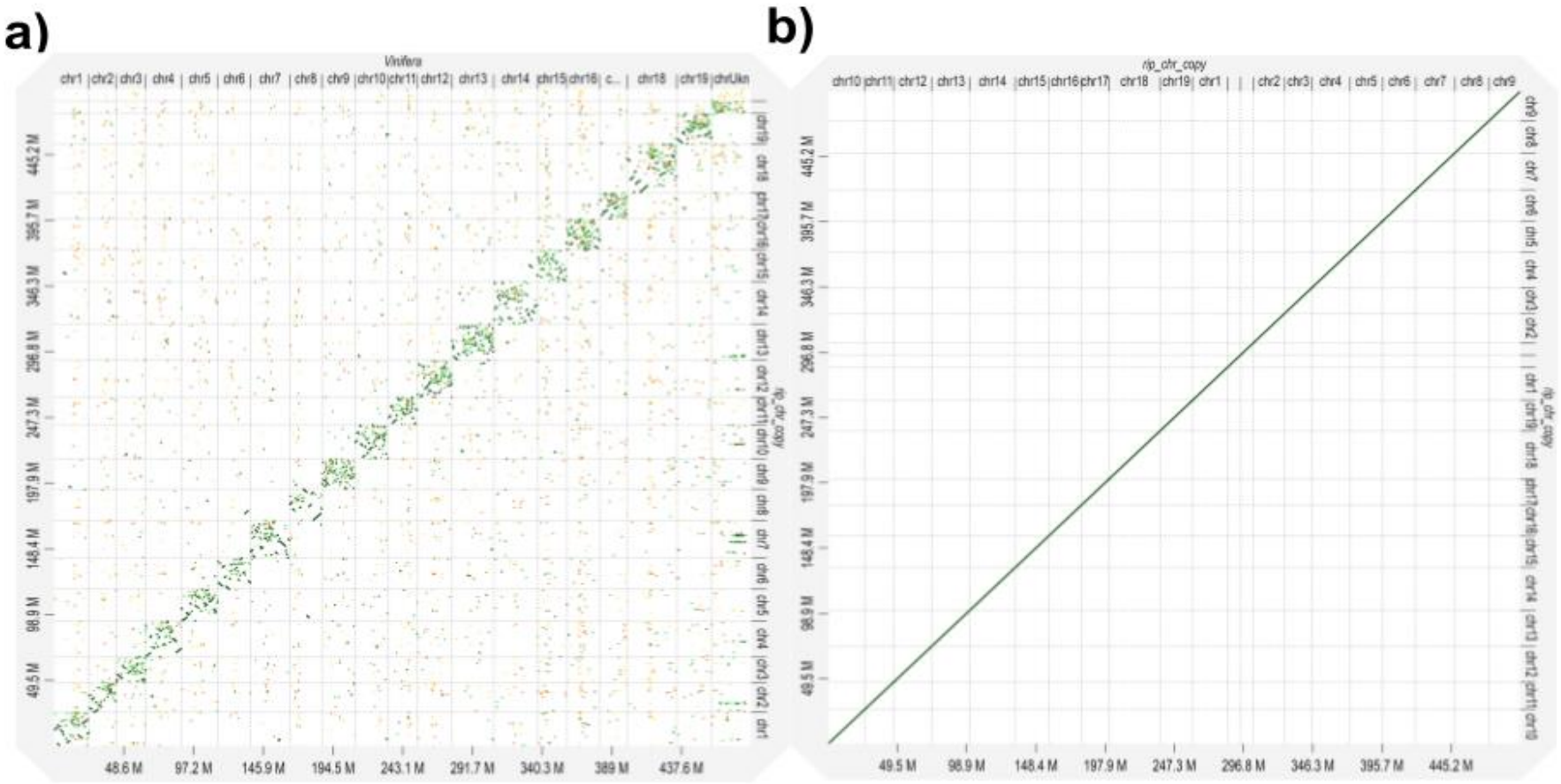


Figure 2.4 Predicted misassembly sites using the REAPR program. a) Number of sites with low mate-pair coverage plotted against length of contig, separated by insert size of mate-pair library. b) Venn diagram showing similar site position based on insert size of library before removal of contigs smaller than insert size. c) Venn diagram showing similar site position based on insert size (3-5 kb, 8-10 kb, and 15-20 kb) of library after removal of contigs smaller than insert size.

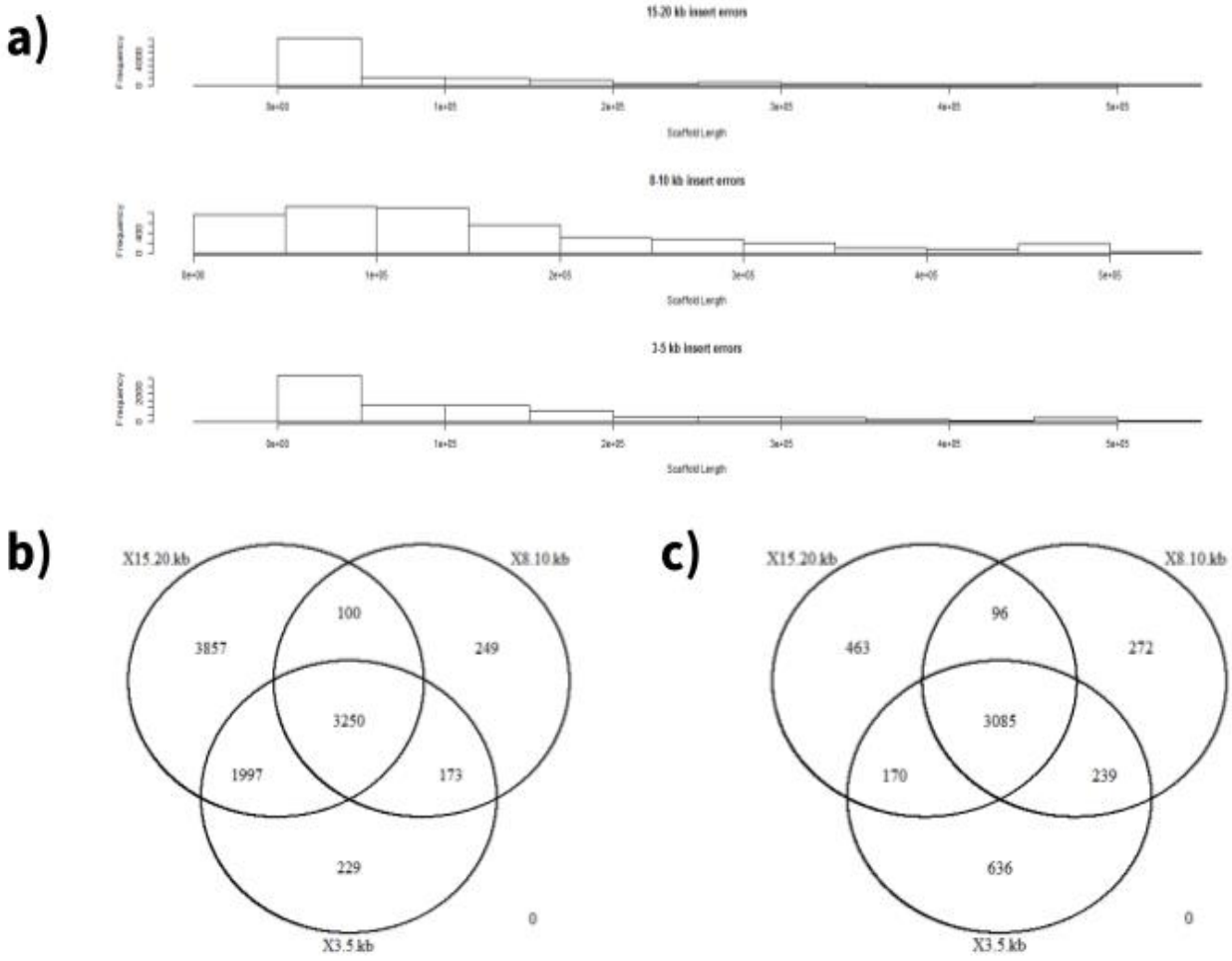


Figure 2.5 Misassembly sites mapped based on chromosomal positions. a) Sites mapped to position of contig on pseudo chromosome. c) Sites mapped to pseudo chromosome 10 of *V. riparia* ‘Manitoba 37’ aligned to homologous alignment with *V. vinifera* ‘PN40024’. Sites are marked in green.

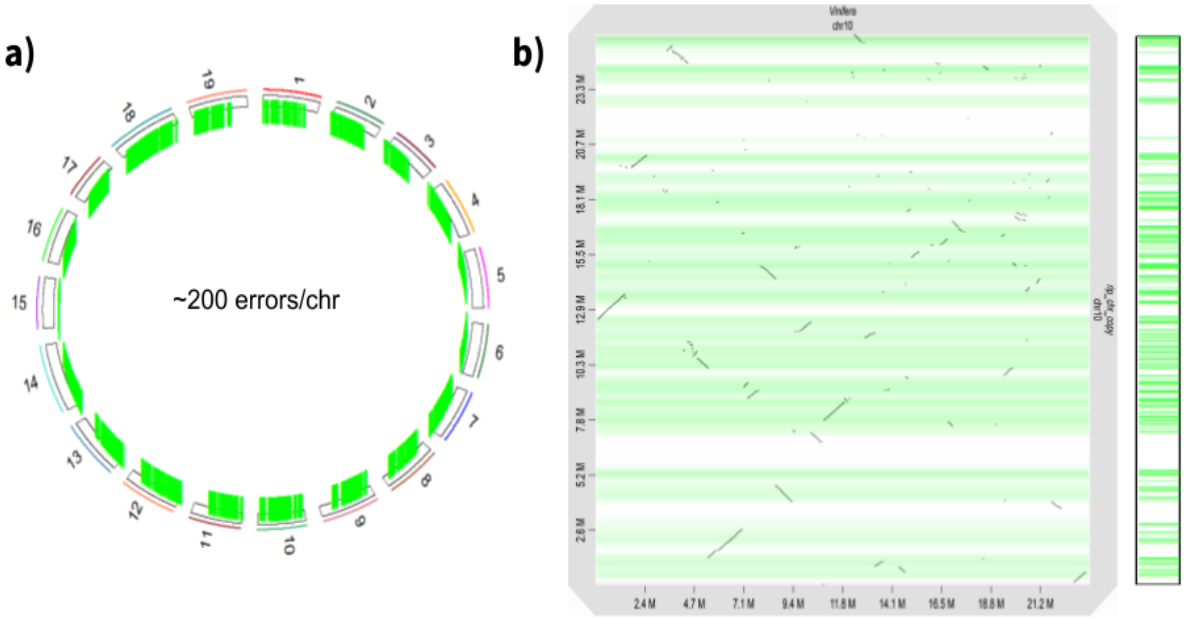


Figure 2.6 Analysis of predicted gene annotation. a) Heat Map displaying the differences in numbers of genes per a protein family or b) Gene ontology (GO) term. c) Protein alignment of members of the “retrotrans_gag 2” family (LTR retrotransposons). Monophyletic clades containing genes from *V. vinifera* ‘PN40024’ and *V. riparia* ‘Manitoba 37’ were annotated and the node containing the most recent common ancestor was indicated by black triangle. d) Protein alignment of LBD family proteins, bootstrap values are at nodes.

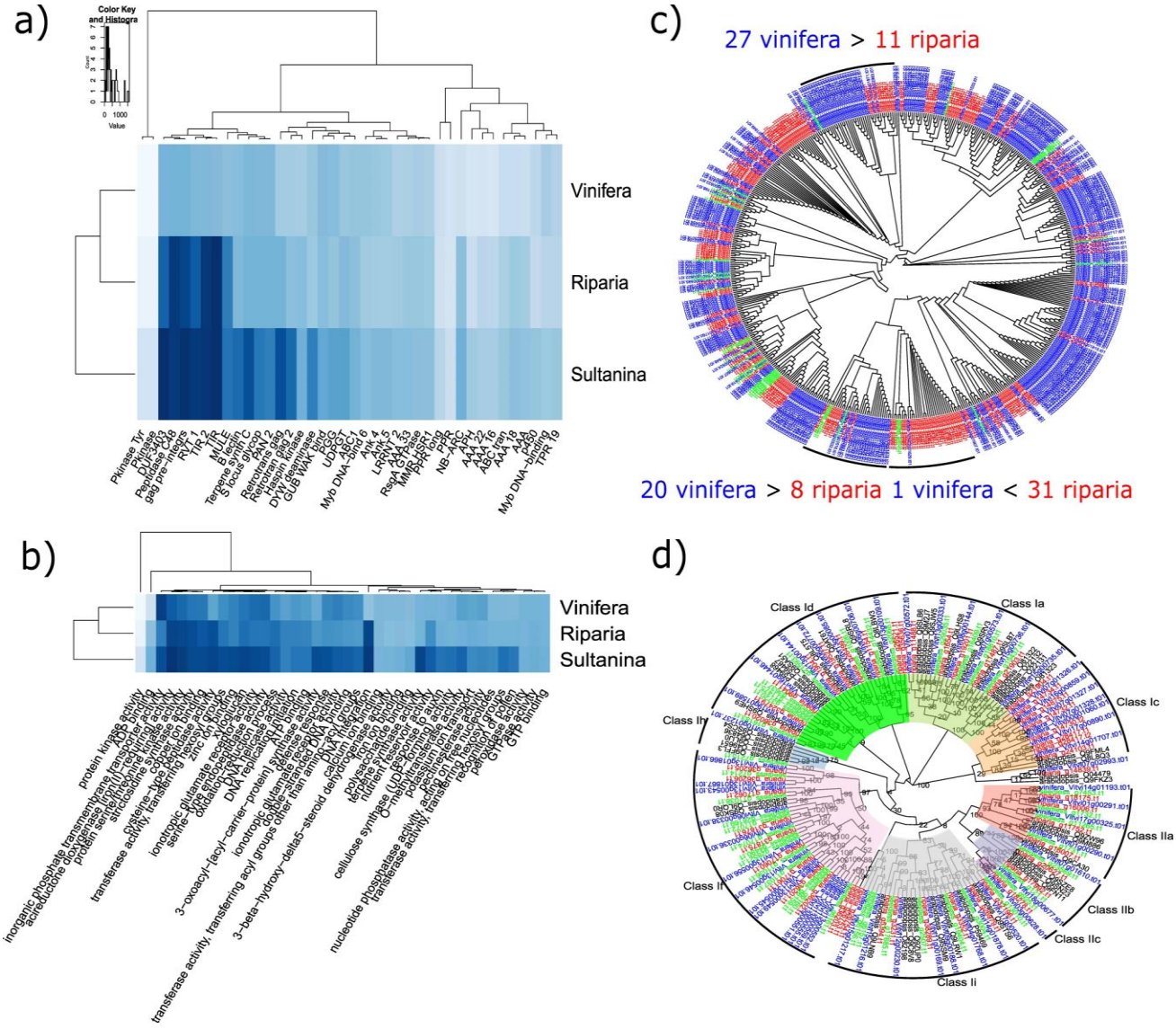


Figure 2.7 Phylogenetic tree displaying the evolutionary relationship between ERF family transcription factors in *V. riparia* 'Manitoba 37' and *V. vinifera* 'PN40024'. Gene subfamilies annotated based on alignment to *A. thaliana* ERF genes. Predicted protein motifs using meme-suite tools plotted next to each gene. Confidence of each node represented in bootstrap values.

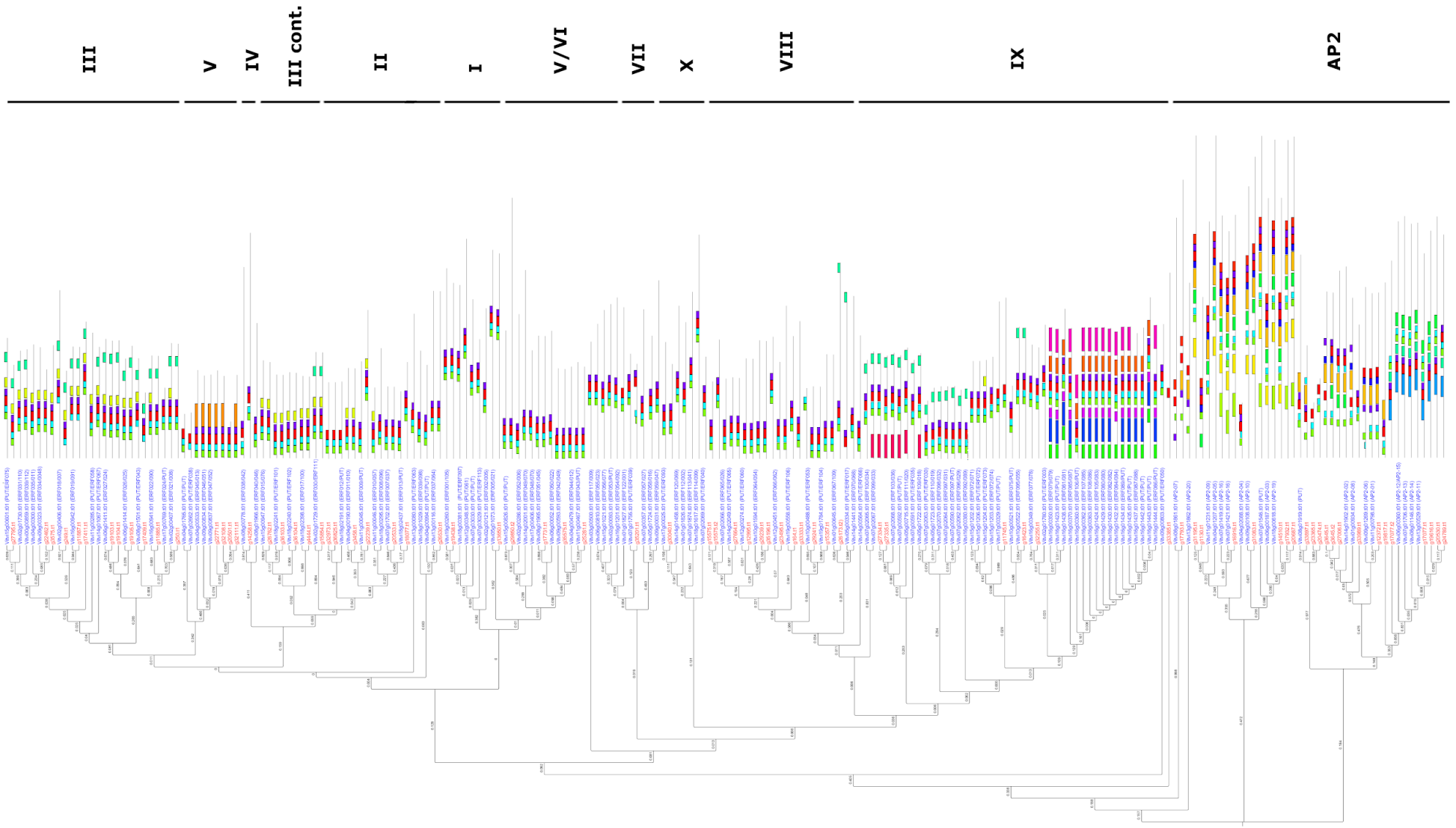


Figure 2.8 CBF genes from an ERF phylogenetic analysis. Hydrophobic motif that is missing in *V. riparia* 'Manitoba 37' is boxed in red. The motif logo for this missing motif is displayed in the top right of the figure. The CBF4 gene that was previously predicted in *V. riparia* is annotated in both *V. riparia* 'Manitoba 37' and *V. vinifera* 'PN40024'.

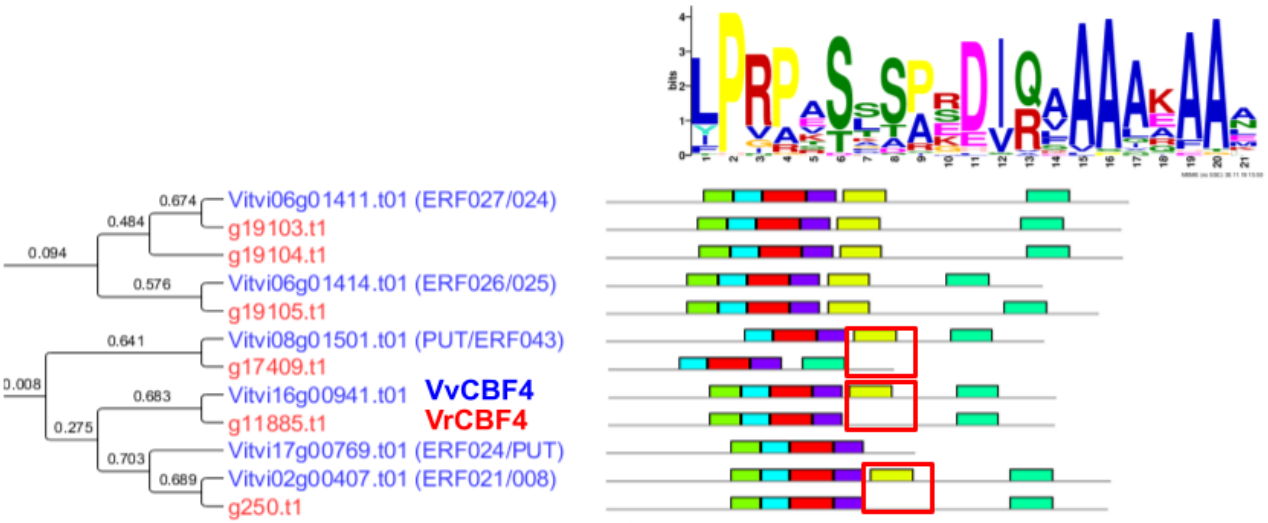


Figure 2.9 Phylogeny of the MYB family of transcription factors in *V. riparia* 'Manitoba 37'. *V. vinifera* 'PN40024' MYB genes labeled in blue and *V. riparia* 'Manitoba 37' MYB genes labeled in red. MYB genes labeled with green dots and MYB-related genes labeled with black dots. Subgroup 6 which experienced the most paralogous duplication has been annotated.

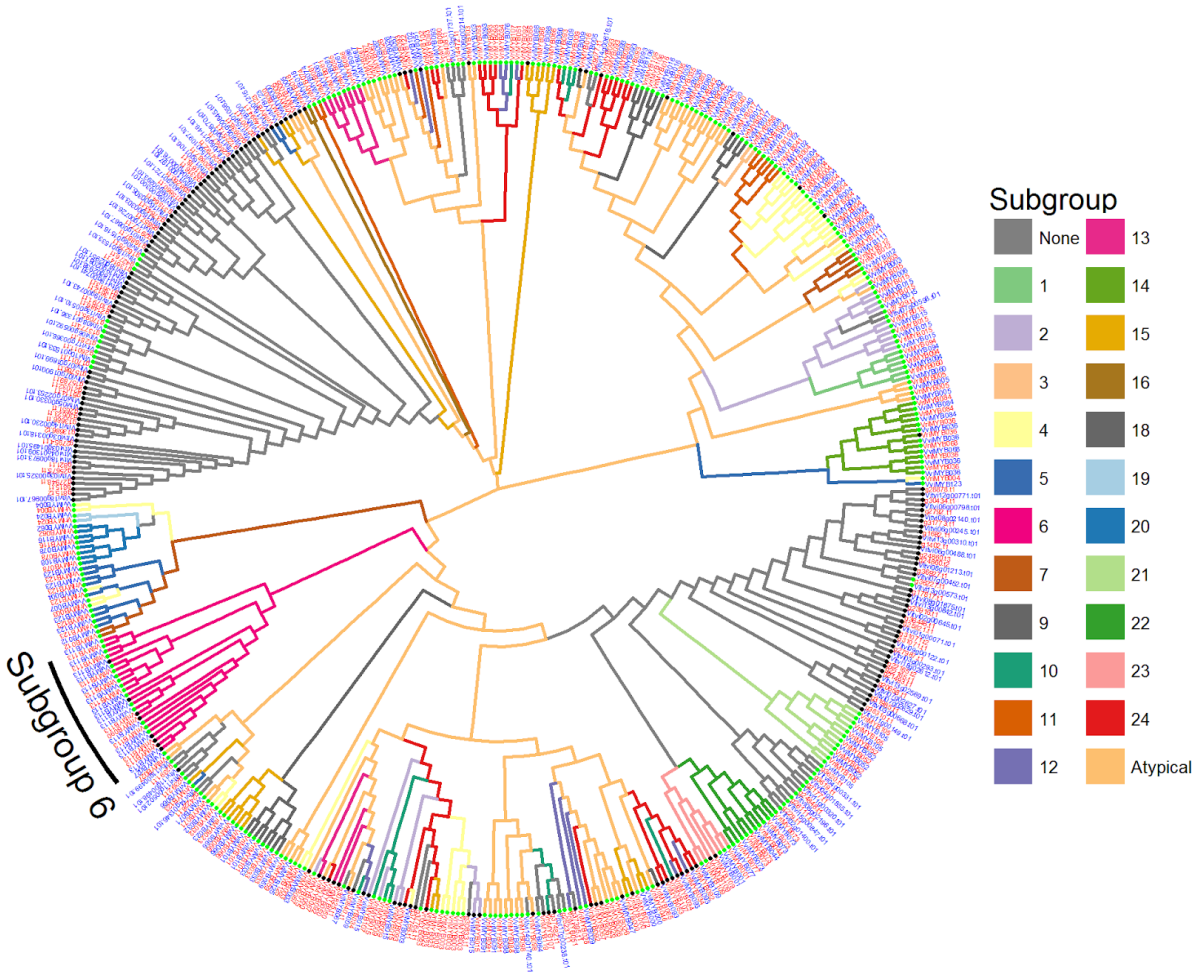


Figure 2.10 Alignment of GBS markers from a F2 mapping population to *V. riparia* 'Manitoba 37'. a) Venn diagram showing which markers aligned to both genomes. We observed that 2.36% of markers did not align to either genome. b) Markers were mapped based on chromosomal positions in *V. riparia* and *V. vinifera* and linked between the genomes to identify chromosomal translocations.

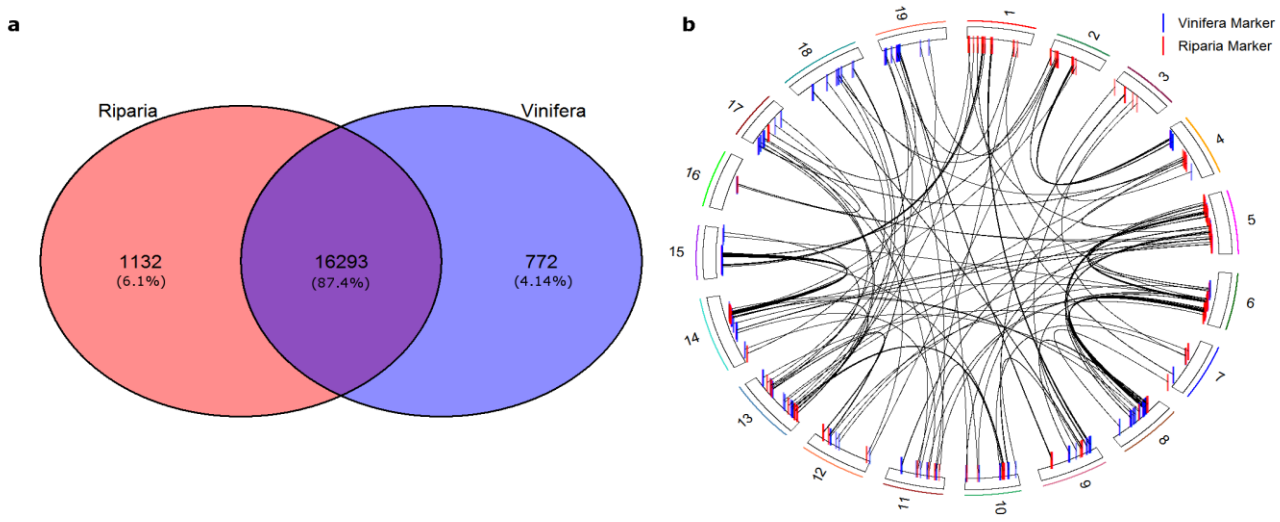
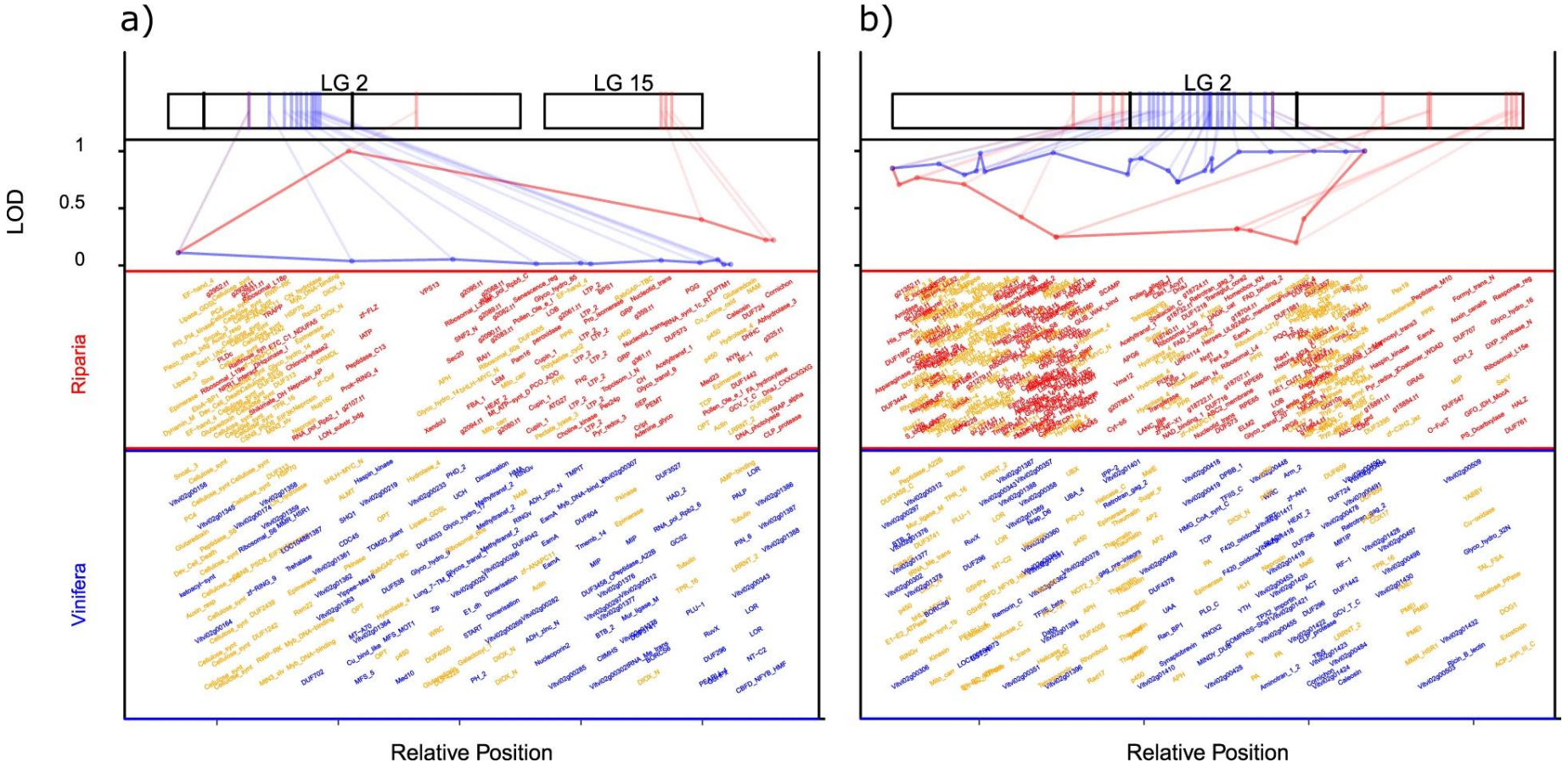


Figure 2.11 Visual representation of QTL found in assembled grapevine genomes aligned to normalized genic regions. The QTL represented are for a) summer lateral critical photoperiod growth cessation and b) flower type. In the top–middle row, markers found within a genomic region of each genome between the two flanking markers are plotted by their relative position along the x axis and the LOD score along the y axis. The LOD score at each marker is normalized and represented as values between 0 and 1 of the maximum LOD score. F2 markers present in *V. riparia* are colored red and those present in *V. vinifera* are blue. In the two bottom panels, genes are annotated by PFAM name, and genes within the QTL in both genomes are colored orange. In the top row, marker positions within the F2 genetic map are connected to the same marker position on each genome.



3 RNA-seq reveals bud endodormancy expression unique to grapevine F2 genotypes

3.1 Abstract

Endodormancy signals a major change in RNA expression within bud tissue. We hypothesized that this expression would be different in grapevines that have different phenotypic responses to photoperiod induced endodormancy. To examine this difference in the expression state we performed RNA-seq on bud tissue from multiple F2 genotypes treated with both LD (15 h) and SD (13 h) photoperiod at 28- and 42-day time points. Greenhouse studies combined with general differential expression revealed that F2-110 was more genetically similar to *V. riparia* than *Vitis spp.* 'Seyval Blanc'. Differentially expressed genes related to endodormancy were different between genotypes with F2-110 showing increased expression of many stress related and pathogen response genes. We confirmed that F2-110 had higher expression of plant defense genes, as well as higher expression of genes involved in auxin signaling pathways. We found that differentially expressed genes clustered into patterns of expression which correlated highly to genotype and photoperiod. Genotype specific clusters are also correlated to auxin signaling pathways which suggest that F2-110 are more sensitive to auxin. This is confirmed in F2-110 by greater relative transcriptional regulation through PLT1, an auxin dependent regulator of the stem cell state. Clusters that correlated to SD treatment were specific to ethylene response

pathways, highlighting the purpose of this pathway in endodormancy. The pathways and gene networks identified in this study are consistent with findings from other large scale transcriptomic studies of endodormancy and further investigation may reveal more about the regulation of dormancy in grapevine.

3.2 Introduction

Winter survival is one of the most important factors in fruit crop agriculture as losses of crop plants due to frost damage greatly affects annual yield for growers (Atucha et al. 2018). Physiological processes are known to protect plants from tissue damage in the winter, namely through changes to the paradormant bud, such as closure of the apoplast from the stem phloem and hardening of the bud scales (Anderson et al. 2005). These physiological changes have been shown to start in the fall season in response to decreasing temperature and day length and are crucial for preparing the bud to resist cold temperatures through isolation of the bud tissue and overcooling internal fluids to prevent freezing damage (Anderson et al. 2005; LANG and G. A 1987; A. Fennell and Mathiason 2002). This process that takes place in the terminal bud of trees has been termed “endodormancy” and is suggested to be controlled by internal signals within the bud. While grapevine sets axillary buds instead of terminal buds at the end of the growing season, grapevine buds still undergo similar tissue differentiation during the early stages of endodormancy induction. Endodormancy has been suggested to be necessary for cold hardiness during the winter season in *V. vinifera* (Rubio et al. 2016). North American species of

grapevine have been shown to have higher winter survival than *V. vinifera* (Londo and Martinson 2015) and the North American grapevine, *V. riparia*, has been found to enter endodormancy at longer day lengths than *V. vinifera*, illustrating the importance of genetics in the process of endodormancy (Wake and Fennell 2000).

Endodormancy is brought about by drastic adjustment to internal gene expression as has been demonstrated previously by multiple studies. Clock-related and photosynthesis genes are heavily impacted in pear and grapevine undergoing dormancy transition in early December (Liu et al. 2012; Vergara, Noriega, and Pérez 2021). Chromatin modification is also a commonly seen factor related to the endodormant condition, which has been demonstrated in differentially expressed gene clusters in pear, apricot, and sweet chestnut (Santamaría et al. 2009; J. Yu et al. 2020). The role of secondary metabolites in endodormancy is so far unknown but significant enrichment of genes involved in metabolic pathways was found in the differentially expressed genes of alfalfa (Du et al. 2018). This, combined with the significant amount of evidence suggesting that phenylpropanoids are involved in the endodormant response make them a focus of research in both seed and bud dormancy (Buer and Muday 2004; Buer, Muday, and Djordjevic 2008; Debeaujon et al. 2001; A. Y. Fennell et al. 2015). Previous high-throughput experiments of grapevine are of significant interest, as genes and pathways in other *vitis* species could provide insight into North American grapevine's response to endodormancy. The pathways involved in these studies have roots in energy and carbohydrate metabolism, hormones,

oxidative stress, signal transduction and cell growth (Khalil-Ur-Rehman et al. 2019, 2017; Noriega and Pérez 2017; Mathiason et al. 2009; Min et al. 2017; Sudawan et al. 2016).

While these types of large-scale studies are important in understanding differential expression during dormancy, few studies have been aimed at induction into endodormancy by photoperiod signaling (A. Y. Fennell et al. 2015). Sequencing the gene expression in species that have differing regulation of endodormancy induction will highlight genes and pathways that are required for the regulation of endodormancy induction. In the current study, we perform RNA-seq in buds of *V. riparia* and *V. vinifera* derived genotypes after 28- and 42-day treatment under LD and SD conditions. We used F2 genotypes derived from a *V. riparia* 'Michx' x *V. vinifera* 'Seyval' cross because they displayed exaggerated short-day response to endodormancy. We will examine changes in gene expression due to photoperiod treatment and compare these changes at different timepoints and how they differ between genotypes to identify genes and pathways that are necessary for endodormancy induction.

3.3 Materials and Methods

3.3.2 Plant material and photoperiod treatment

Two genotypes from an F2 mapping population resulting from the cross of *V. riparia* 'Michx.' x *V. vinifera* 'Seyval' were selected for response to photoperiod (Yang et al. 2016). These potted and spur pruned-vines were grown in LD (15 h)

at 25/20 ± 3°C day/night temperatures with 600–1400 mol m−2 s−1 photosynthetic photon flux in a climate-controlled unshaded glass greenhouse (En Tech Control Systems Inc., Montrose, Minn.) in Brookings, South Dakota (44.3 N). Thirty days post bud break, 25 plants each were randomly divided into a split plot design photoperiod treatment of 15 h LD or 13 h SD which was provided by an automated, white-covered black-out system (735 ft2 × 12 ft ceiling height; Van Rijn Enterprises LTD; Grassie, Ontario). Bud tissue was harvested from 5 plants each at 7-, 14-, 21-, 28-, and 42-day time periods, placed in liquid nitrogen, and stored at -80°C for RNA extraction. Bud break was evaluated in single node canes incubated at 15 h LD in water by measuring the number of replicates that reach EL5 stage of leaf growth after harvest. Harvested plants were also allowed to regrow under LD greenhouse conditions and measured for bud break in a similar manner after 2 weeks of growth.

3.3.2 RNA extraction and sequencing

RNA from 28- and 42-day SD and LD photoperiod treated buds was extracted according to manufacturer instructions from the RNeasy plant mini kit (Qiagen, Valencia CA) with 2% Polyvinylpyrrolidone (PVP) added to the extraction buffer. DNA was removed by incubation with 1 unit per microgram (µg) RNase-free DNase (Promega, Madison WI) at 37°C for 30 min. RNA quality was assessed by gel electrophoresis coupled with analysis of 1 µL by Nanodrop UV-vis (Thermo Fisher scientific, Waltham MA). Prior to library preparation, quality and quantity were confirmed by Agilent (Santa Clara, CA) 2100 Bioanalyzer RNA 6000 nano chip. cDNA libraries were constructed using Truseq RNA Library prep

kit (Illumina, San Diego CA) and ran on an Illumina NextSeq 550 using two flow cells.

3.3.3 Sequence processing and alignment

Raw reads from sequencing were processed for quality with FASTQC and trimmed with Trimmomatic (Bolger, Lohse, and Usadel 2014). Processed reads were aligned to a recent long read assembly of the *V. riparia* genome (Not yet published), and gene counts were determined from alignment using HTSeq 0.9.1 (Anders, Pyl, and Huber 2015). Counts were filtered for low read specificity and transformed using both Rlog and variance stabilizing (Ntd) transformations in the R programming language (R Core Team 2017).

3.3.4 Differential expression analysis and statistics

We applied a *3-way Anova* to examine the bud break in each genotype and plotted values using the ggplot2 package (Wickham 2016). Differential expression was evaluated between combinations of treatment conditions, and a negative binomial statistical test was applied through the DESeq2 package (Love, Huber, and Anders 2014) and corrected with a False Discovery Rate (FDR). Differentially Expressed Genes (DEG) were filtered for Log Fold Change (LFC) values greater than 1 and an alpha of 0.01. Differential expression within each time period followed a simple LD*SD design, however for differential expression between genotypes we applied a mixed model: $= geno + trt + per + geno{:}trt + geno{:}per + trt{:}per + geno{:}trt{:}per$, where *geno* is genotype, *trt* is photoperiod treatment, and *per* is time point. For examination of

statistical enrichment of KEGG pathways in differentially expressed genes, we applied a gene set enrichment (GSEA) algorithm through the clusterprofiler package (G. Yu et al. 2012). Genes were organized into modules based on similar patterns of expression using WGCNA to determine network involvement (Langfelder and Horvath 2008).

3.4 Results

3.4.1 SD induced endodormancy displays differential growth and expression patterns in F2 genotypes

Under SD (13h) endodormant treatments, axillary grapevine buds on *V. riparia* (VR37) do not break out of paradormancy as well as *V. vinifera* ‘Seyval’ (Fig 3.1a). The F2 genotypes that we have selected from a cross between these species show exaggerated phenotypes of growth cessation (Fig 3.1b). We observed a statistically significant reduction in bud break of F2-110 SD treated buds after only 14 days of treatment while F2-040 didn’t see the same reduction until 42 days of SD treatment (Fig 3.2). The principal component analysis (PCA) of sequenced bud mRNA showed variation in expression between LD and SD treated samples was different in the 28 day F2-110 but not in the 28-day F2-040 (Fig 3.3). Certain genes such as aquaporins and heat shock proteins showed higher read counts specific to SD treatments, but varied by genotype and time point, indicating that they are specific to endodormant conditions (Fig 3.4). Additional analysis of the data showed us that there was no correlation between

differential expression of genes and higher count numbers giving us an unbiased sample to test differential expression (Fig 3.5).

3.4.2 Differential expression of genes at 28 days of SD photoperiod showed greater transcriptional activity in F2-040 than in F2-110

We examined differential expression of genes in SD treated samples as compared to LD treated samples and then compared results in each genotype (Appendix 1a, 1b). The F2-110 had an overall lower number of differentially expressed genes than F2-040 (Fig 3.6). Increased transcriptional activity suggests that the F2-040 is still undergoing endodormant transformation. We found that most of the upregulated DEGs in F2-040 were related to microtubule movement, phenylpropanoid biosynthesis and signaling pathways (Fig 3.7). Likewise, protein kinase genes had inverse patterns of expression, downregulated in the F2-110 and upregulated in F2-040. Specific genes that were downregulated in the F2-110 but upregulated in the F2-040 at 28 days of SD photoperiod treatment included light responsive NON-PHOTOTROPHIC HYPOCOTYL 3 (NPH3) genes, expansins and NAC transcription factors (Table 3.1). Significantly enriched pathways identified in F2-040 through GSEA included phenylpropanoid biosynthesis, DNA replication, pectin, and gluconeogenesis (Fig 3.8a). If F2-040 at 28 days is undergoing transformation into endodormant buds, we can consider that phenylpropanoids play a major role in this process and a lesser role in maintenance of the endodormant state as shown in the reduced number of differentially expressed phenylpropanoid pathway genes in 28 day SD

treated F2-110 (Fig 3.7). Conversely the most important pathways involved in endodormant maintenance, that were differentially expressed in the F2-110 but not F2-040 included auxin mediated signaling pathways, multidrug ABC transport, pectin, and protein kinases (Fig 3.8b). While pectin pathways were upregulated in the F2-040, they were downregulated in the F2-110.

3.4.3 Differential expression of genes at 42 days SD photoperiod treatment

At 42 days of SD photoperiod treatment, we observed significant reduction in bud break in both genotypes, suggesting that they are both in an endodormant state (Appendix 1c, 1d and Fig 3.2). This is reflected in comparative differential gene expression where we saw almost identical numbers of differentially expressed genes between the two genotypes (Fig 3.9). The high transcriptional activity (>2,000 DEGs) seen in both genotypes at 42 days of SD photoperiod treatment. The majority of pathways that had increased enrichment of DEGs in 42 day SD treated F2-040 saw rampant downregulation, and these pathways represented cell growth and death, microtubule, and pectin synthesis pathways (Fig 3.10). Upregulated pathways for the 42 day F2-040 included MYB transcription factors and single reactions typical in secondary metabolite production. Phenylpropanoid biosynthesis genes remained partially upregulated in 42 day SD treated F2-040 buds but not F2-110 buds. This corresponded to significant enrichment in pathway gene sets of F2-040 DEGs, confirming the significant upregulation of MYB transcription factors and single reactions typical

of secondary metabolite production, and downregulation of pectin modification, heat shock, drought response, and auxin signaling genes (Fig 3.11). Unfortunately, no pathways were found to be significantly enriched in F2-110 endodormant samples.

3.4.4 Auxin signaling pathways are required for genotype specific endodormancy response

Finally, we looked at differential expression of genes between the F2 genotypes across all libraries. Because we had multiple treatments that could affect expression, we applied a mixed model to test the interactions of multiple treatment factors. We found that there was a higher upregulation of genes in the F2-110 when compared to the F2-040 (Appendix 1e and Fig 3.12). Any of these genes could be needed for F2-110 to enter endodormancy at a faster rate than F2-040 so we compared DEGs and pathways to the other RNA-seq comparisons. Much like in the 28 day SD photoperiod results, we saw downregulation of protein kinases in F2-110 and upregulation of phenylpropanoid biosynthesis and auxin mediated signaling genes (Fig 3.13). This could indicate that genes in these pathways are naturally higher expressed in the F2-110. Phenylpropanoid biosynthesis genes were then found significantly enriched in upregulated pathways, stressing its importance in endodormancy induction (Fig 3.14). GSEA also showed the importance of differential expression of transport genes, as well as the downregulation of Jasmonate, K+ transporters, and NBS-LRR biotic stress response genes.

3.4.5 Auxin and phenylpropanoid biosynthesis networks enriched in gene co-expression clusters

We applied WGCNA to identify over 150 clusters of genes that are expressed similarly among treatments. These gene clusters may be involved in important networks or controlled by the same regulatory pathways. When we analyzed clusters that were highly correlated to SD photoperiod treatment ($r^2 > 0.7$; Fig 3.15), we found that a majority of the clusters were related to ethylene-mediated signaling pathways which includes stress response genes like ERF transcription factors (Fig 3.16a-b). There was a much greater correlation of clusters to F2 genotypes than to treatments ($r^2 > 0.9$; Fig 3.17). Such clusters could tell us reasons why F2-110 is more responsive to SD photoperiod treatment. Genes found in F2-040 specific clusters were typically expressed at higher rates than in F2-040 samples and were more enriched in protein kinase pathways (Fig 3.18a and 3.18c). Genes found in F2-110 specific clusters were more expressed in F2-110 samples and were more enriched in flavonoid biosynthesis and auxin signaling pathways (Fig 3.18b and 3.18d).

3.5 Discussion

In this study, we decided to investigate gene expression of photoperiod induced endodormancy using high-throughput RNA-seq technologies of LD and SD photoperiod treated grapevine buds. This experimental design is intended to elucidate genes and pathways that are important for induction into the endodormant condition. We have further improved on this design by including F2

genotypes of a *V. riparia* x *V. vinifera* derived cross in the RNA-seq design. The advantage of this system is that recombination could better reveal genetic relationships to the phenotype of early endodormancy induction (Espinosa-Soto, Hernández, and Posadas-García 2021). We included two time points to view endodormancy at different points in transition.

This experimental design led us to find major differences in expression between the two genotypes. In the RNA-seq comparison between LD and SD photoperiod treated samples at each time point, this meant a greater differential expression in F2-040 than F2-110 at 28 days but not 42 days. This is consistent with the greater DEGs in endodormancy transition seen in other RNA-seq studies (A. Y. Fennell et al. 2015) and leads us to believe that F2-040 at 28 days of SD treatment is just entering endodormancy while F2-110 has already been through it. What we currently cannot explain is why there is a very high DEG count at 42 days for both genotypes, but it could be because endodormancy maintenance has different patterns of gene expression than paradormancy. Pathway involvement differed in both genotypes, but we saw many common threads. Pathways involving cell growth were suppressed, unsurprisingly, because they aren’t needed to maintain a dormant bud meristem tissue (Sudawan et al. 2016). This was apparent with the downregulation of key genes like cyclins and DNA replication genes.

Genes differentially expressed between the two genotypes such as expansin and NAC (Table 3.1) contributed to the genotype specific SD photoperiod response seen in the F2-110. Similar to expression of genes in the

F2-040 due to SD photoperiod treatment, the F2-110 had a naturally higher differential expression of phenylpropanoid biosynthesis genes. This is a pathway that has been identified to be differentially expressed in previous bud endodormancy experiments (A. Y. Fennell et al. 2015). A study in grapevine that looked at metabolites found increased amounts of flavonoids like kaempferol, quercetin, and procyanidins in endodormant plants as a way of regulating auxin transport and catabolism through multidrug resistant ABC-transporters (Conrad et al. 2019; Lewis et al. 2007). This connects to what we see in genotype and endodormant specific expression of flavonoid, ABC transporter, and auxin signaling pathway genes. We observed similar pathways when we looked at networks of genes that were co-expressed at similar levels across samples (Fig 3.16 and Fig 3.18).

One thing that was rarely seen in other studies was the involvement of plant pathogen pathways in clusters related to endodormancy. This could be one way that F2-110 is able to respond to cold stress better than F2-040. Some studies have linked plant pathogen response to Salicylic Acid, which improves cold stress tolerance by reducing oxidative degradation of lipids and membrane permeability in grapevine (Dempsey et al. 2011; Miura and Tada 2014; Wang and Li 2006). Clusters that were highly correlated to photoperiod endodormancy also included genes in the ethylene response pathway. Aside from the ICE, CBF, COR response pathway to cold stress, experiments in birch have shown that ethylene insensitive mutants have delayed dormancy under SD conditions (Ruonala et al. 2006).

In all, this RNA-seq provides us with an interesting path to investigate the regulation of genes involved in endodormancy in *V. riparia*. We identified and confirmed the involvement of pathways like cell growth and death, microtubule, ABA response and ERF signaling that have been previously identified in other studies (Khalil-Ur-Rehman et al. 2017; A. Y. Fennell et al. 2015; Min et al. 2017). Our findings illuminate a greater role for phenylpropanoids and auxin in the regulation of the endodormant state. Further investigation is needed, however, to determine the exact purpose of these genes in endodormancy induction in grapevine.

3.6 References

Anderson, James V., Russ W. Gesch, Ying Jia, Wun S. Chao, and David P. Horvath. 2005. “Seasonal Shifts in Dormancy Status, Carbohydrate Metabolism, and Related Gene Expression in Crown Buds of Leafy Spurge.” Plant, Cell & Environment 28 (12): 1567–78.

Anders, Simon, Paul Theodor Pyl, and Wolfgang Huber. 2015. “HTSeq--a Python Framework to Work with High-Throughput Sequencing Data.” Bioinformatics 31 (2): 166–69.

Atucha, Amaya, Janet Hedtcke, Beth Ann Workmaster, and Others. 2018. “Evaluation of Cold-Climate Interspecific Hybrid Wine Grape Cultivars for the Upper Midwest.” J. Am. Pomol. Soc 72: 80–93.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. “Trimmomatic: A

Flexible Trimmer for Illumina Sequence Data.” Bioinformatics 30 (15): 2114–20.

Buer, Charles S., and Gloria K. Muday. 2004. “The Transparent testa4 Mutation Prevents Flavonoid Synthesis and Alters Auxin Transport and the Response of Arabidopsis Roots to Gravity and Light.” The Plant Cell 16 (5): 1191–1205.

Buer, Charles S., Gloria K. Muday, and Michael A. Djordjevic. 2008. “Implications of Long-Distance Flavonoid Movement in Arabidopsis Thaliana.” Plant Signaling & Behavior 3 (6): 415–17.

Conrad, Anna O., Jiali Yu, Margaret E. Staton, Jean-Marc Audergon, Guillaume Roch, Veronique Decroocq, Kevin Knagge, et al. 2019. “Association of the Phenylpropanoid Pathway with Dormancy and Adaptive Trait Variation in Apricot (Prunus Armeniaca).” Tree Physiology 39 (7): 1136–48.

Debeaujon, I., A. J. Peeters, K. M. Léon-Kloosterziel, and M. Koornneef. 2001. “The TRANSPARENT TESTA12 Gene of Arabidopsis Encodes a Multidrug Secondary Transporter-like Protein Required for Flavonoid Sequestration in Vacuoles of the Seed Coat Endothelium.” The Plant Cell 13 (4): 853–71.

Dempsey, D’maris Amick, A. Corina Vlot, Mary C. Wildermuth, and Daniel F. Klessig. 2011. “Salicylic Acid Biosynthesis and Metabolism.” The

Arabidopsis Book / American Society of Plant Biologists 9 (December): e0156.

Du, Hongqi, Yinghua Shi, Defeng Li, Wenna Fan, Yanhua Wang, Guoqiang Wang, and Chengzhang Wang. 2018. “Proteomics Reveals Key Proteins Participating in Growth Difference between Fall Dormant and Non-Dormant Alfalfa in Terminal Buds.” Journal of Proteomics 173 (February): 126–38.

Espinosa-Soto, Carlos, Ulises Hernández, and Yuridia S. Posadas-García. 2021. “Recombination Facilitates Genetic Assimilation of New Traits in Gene Regulatory Networks.” Evolution & Development 23 (5): 459–73.

Fennell, Anne, and Katherine Mathiason. 2002. “Early Acclimation Response in Grapes (VITIS).” In Plant Cold Hardiness: Gene Regulation and Genetic Engineering, edited by Paul H. Li and E. Tapio Palva, 93–107. Boston, MA: Springer US.

Fennell, Anne Y., Karen A. Schlauch, Satyanarayana Gouthu, Laurent G. Deluc, Vedbar Khadka, Lekha Sreekantan, Jerome Grimplet, Grant R. Cramer, and Katherine L. Mathiason. 2015. “Short Day Transcriptomic Programming during Induction of Dormancy in Grapevine.” Frontiers in Plant Science 6 (November): 834.

Khalil-Ur-Rehman, Muhammad, Long Sun, Chun-Xia Li, Muhammad Faheem, Wu Wang, and Jian-Min Tao. 2017. “Comparative RNA-Seq Based Transcriptomic Analysis of Bud Dormancy in Grape.” BMC Plant Biology 17

(1): 18.

Khalil-Ur-Rehman, Muhammad, Wu Wang, Yang Dong, Muhammad Faheem, Yanshuai Xu, Zhihong Gao, Zhen Guo Shen, and Jianmin Tao. 2019. “Comparative Transcriptomic and Proteomic Analysis to Deeply Investigate the Role of Hydrogen Cyanamide in Grape Bud Dormancy.” International Journal of Molecular Sciences 20 (14). https://doi.org/10.3390/ijms20143528.

Langfelder, Peter, and Steve Horvath. 2008. “WGCNA: An R Package for Weighted Correlation Network Analysis.” BMC Bioinformatics 9 (December): 559.

LANG, and G. A. 1987. “Dormancy : A New Universal Terminology.” HortScience: A Publication of the American Society for Horticultural Science 25: 817–20.

Lewis, Daniel R., Nathan D. Miller, Bessie L. Splitt, Guosheng Wu, and Edgar P. Spalding. 2007. “Separating the Roles of Acropetal and Basipetal Auxin Transport on Gravitropism with Mutations in Two Arabidopsis Multidrug Resistance-like ABC Transporter Genes.” The Plant Cell 19 (6): 1838–50.

Liu, Guoqin, Wanshun Li, Penghua Zheng, Tong Xu, Lijuan Chen, Dongfeng Liu, Sayed Hussain, and Yuanwen Teng. 2012. “Transcriptomic Analysis of ‘Suli’ Pear (Pyrus Pyrifolia White Pear Group) Buds during the Dormancy by

RNA-Seq.” BMC Genomics 13 (December): 700.

Londo, J., and T. Martinson. 2015. “Geographic Trend of Bud Hardiness Response in Vitis Riparia.” Acta Horticulturae, no. 1082 (April): 299–304.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” Genome Biology 15 (12): 550.

Mathiason, Kathy, Dong He, Jérôme Grimplet, J. Venkateswari, David W. Galbraith, Etti Or, and Anne Fennell. 2009. “Transcript Profiling in Vitis Riparia during Chilling Requirement Fulfillment Reveals Coordination of Gene Expression Patterns with Optimized Bud Break.” Functional & Integrative Genomics 9 (1): 81–96.

Min, Zhuo, Xianfang Zhao, Runyu Li, Bohan Yang, Min Liu, and Yulin Fang. 2017. “Comparative Transcriptome Analysis Provides Insight into Differentially Expressed Genes Related to Bud Dormancy in Grapevine (Vitis Vinifera).” Scientia Horticulturae 225 (November): 213–20.

Miura, Kenji, and Yasuomi Tada. 2014. “Regulation of Water, Salinity, and Cold Stress Responses by Salicylic Acid.” Frontiers in Plant Science 5 (January): 4.

Noriega, Ximena, and Francisco J. Pérez. 2017. “ABA Biosynthesis Genes Are Down-Regulated While Auxin and Cytokinin Biosynthesis Genes Are Up-Regulated During the Release of Grapevine Buds From Endodormancy.”

Journal of Plant Growth Regulation 36 (4): 814–23.

R Core Team. 2017. “R: A Language and Environment for Statistical Computing.” Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rubio, Sebastián, Débora Dantas, Ricardo Bressan-Smith, and Francisco J. Pérez. 2016. “Relationship Between Endodormancy and Cold Hardiness in Grapevine Buds.” Journal of Plant Growth Regulation 35 (1): 266–75.

Ruonala, Raili, Päivi L. H. Rinne, Mourad Baghour, Thomas Moritz, Hannele Tuominen, and Jaakko Kangasjärvi. 2006. “Transitions in the Functioning of the Shoot Apical Meristem in Birch (Betula Pendula) Involve Ethylene.” The Plant Journal: For Cell and Molecular Biology 46 (4): 628–40.

Santamaría, Ma Estrella, Rodrigo Hasbún, Ma José Valera, Mónica Meijón, Luis Valledor, Jose L. Rodríguez, Peter E. Toorop, Ma Jesús Cañal, and Roberto Rodríguez. 2009. “Acetylated H4 Histone and Genomic DNA Methylation Patterns during Bud Set and Bud Burst in Castanea Sativa.” Journal of Plant Physiology 166 (13): 1360–69.

Sudawan, Boonyawat, Chih-Sheng Chang, Hsiu-Fung Chao, Maurice S. B. Ku, and Yung-Fu Yen. 2016. “Hydrogen Cyanamide Breaks Grapevine Bud Dormancy in the Summer through Transient Activation of Gene Expression and Accumulation of Reactive Oxygen and Nitrogen Species.” BMC Plant Biology 16 (1): 202.

Vergara, Ricardo, Ximena Noriega, and Francisco J. Pérez. 2021. “VvDAM-SVPs Genes Are Regulated by FLOWERING LOCUS T (VvFT) and Not by ABA/low Temperature-Induced VvCBFs Transcription Factors in Grapevine Buds.” Planta 253 (2): 31.

Wake, Carol M. F., and Anne Fennell. 2000. “Morphological, Physiological and Dormancy Responses of Three Vitis Genotypes to Short Photoperiod.” Physiologia Plantarum 109 (2): 203–10.

Wang, Li-Jun, and Shao-Hua Li. 2006. “Salicylic Acid-Induced Heat or Cold Tolerance in Relation to Ca2+ Homeostasis and Antioxidant Systems in Young Grape Plants.” Plant Science: An International Journal of Experimental Plant Biology 170 (4): 685–94.

Wickham, Hadley. 2016. “ggplot2: Elegant Graphics for Data Analysis.” Springer-Verlag New York. https://ggplot2.tidyverse.org.

Yang, Shanshan, Jonathan Fresnedo-Ramírez, Qi Sun, David C. Manns, Gavin L. Sacks, Anna Katharine Mansfield, James J. Luby, et al. 2016. “Next Generation Mapping of Enological Traits in an F2 Interspecific Grapevine Hybrid Family.” PloS One 11 (3): e0149560.

Yu, Guangchuang, Li-Gen Wang, Yanyan Han, and Qing-Yu He. 2012. “clusterProfiler: An R Package for Comparing Biological Themes among Gene Clusters.” Omics: A Journal of Integrative Biology 16 (5): 284–87.

Yu, Jiali, Anna O. Conrad, Véronique Decroocq, Tetyana Zhebentyayeva,

Daniel E. Williams, Dennis Bennett, Guillaume Roch, et al. 2020. “Distinctive Gene Expression Patterns Define Endodormancy to Ecodormancy Transition in Apricot and Peach.” Frontiers in Plant Science 11 (February): 180.

Table 3.1 Inverse expression of DEGs between F2-110 and F2-040 genotypes in 28 day photoperiod treated buds. Log fold change represents up- or down-regulation of genes in the SD photoperiod in F2-040.

Gene Name	Log Fold Change	Annotation
Expansin EXPA17	3.63	Auxin-mediated signaling pathway
Beta-glucosidase	2.42	Cyanoamino acid metabolism
NAC 74	2.37	NAC transcription factor
NPH3	2.26	Light signaling
Pectinesterase	2.19	Pectin modification
C2H2 zinc finger	2.15	C2H2 family transcription factor
SmD3	2.06	mRNA biosynthesis
BTB/POZ NPH3	2.01	Ubiquitin-mediated proteolysis
Unknown	1.82	Lipase family
bHLH	1.63	bHLH family transcription factor
Unknown	1.58	Unknown
Dicyanin	1.44	Oxidative stress response
WAK kinase	1.41	Protein Kinase
NAC 34/35	1.37	NAC transcription factor
MLO6	1.12	Cell death
Nodulin MtN3	-1.17	Root development

Figure 3.1 Photographs showing bud break and growth in grapevine after 28 days of LD or SD photoperiod treatment. a) Comparison of growth in *V. vinifera* ‘Seyval’, *V. riparia* ‘Michx’ (VR37) and an F1 cross of the two species. b) Comparison of growth in two F2 genotypes derived from the F1 cross. Photographs were taken 2 weeks after harvesting and represent viability of paradormant buds.

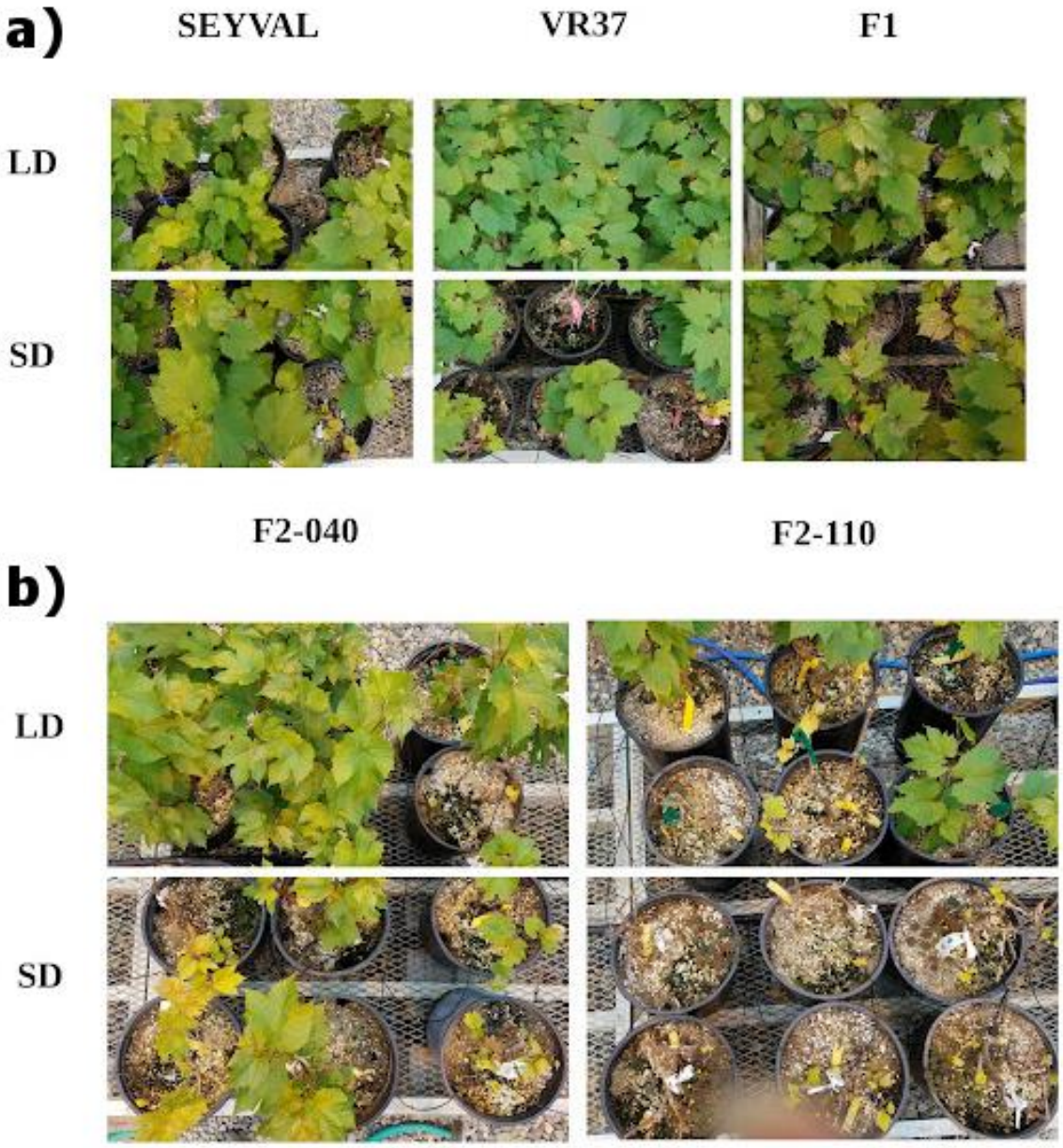
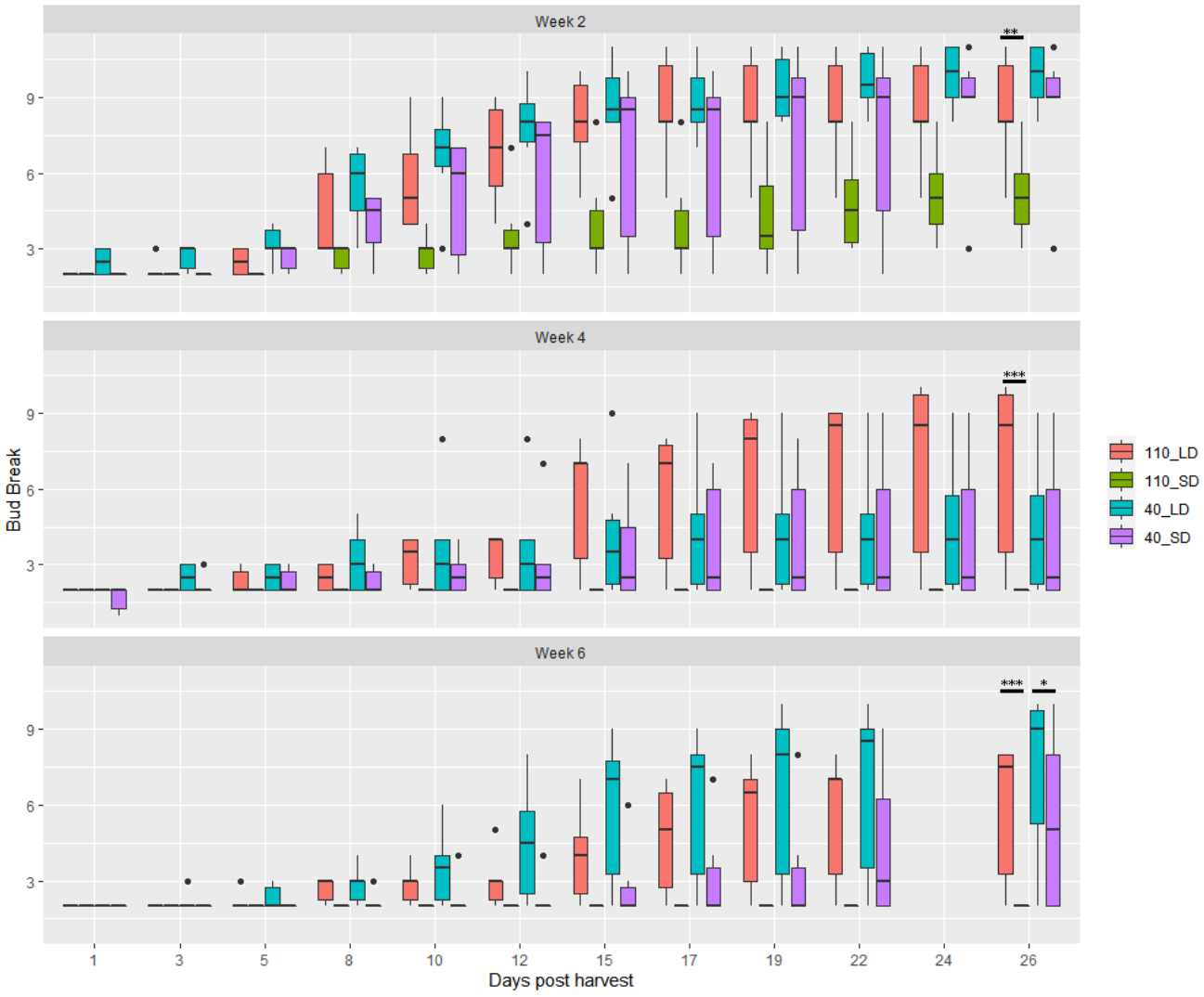


Figure 3.2 Measurement of bud break in F2 hybrids in the days following 14, 28, or 42 day (2, 4, and 6 weeks respectively) treatment with LD and SD photoperiod conditions. Bud break was measured by the number of buds that reached an EL5 stage of growth after a set number of days of incubation in room temperature water post harvest. Data collected from 5th node cuttings incubated at greenhouse temperatures in water solution. Statistics represent pairwise t-test comparisons derived from a multivariate 3-way ANOVA.



**** p < 0.05***
***** p < 0.01***
****** p < 0.001***

Figure 3.3 Principal component analysis of sequenced bud tissue treated with LD and SD photoperiods at multiple time points. Count data transformed by Rlog to normalize patterns of expression.

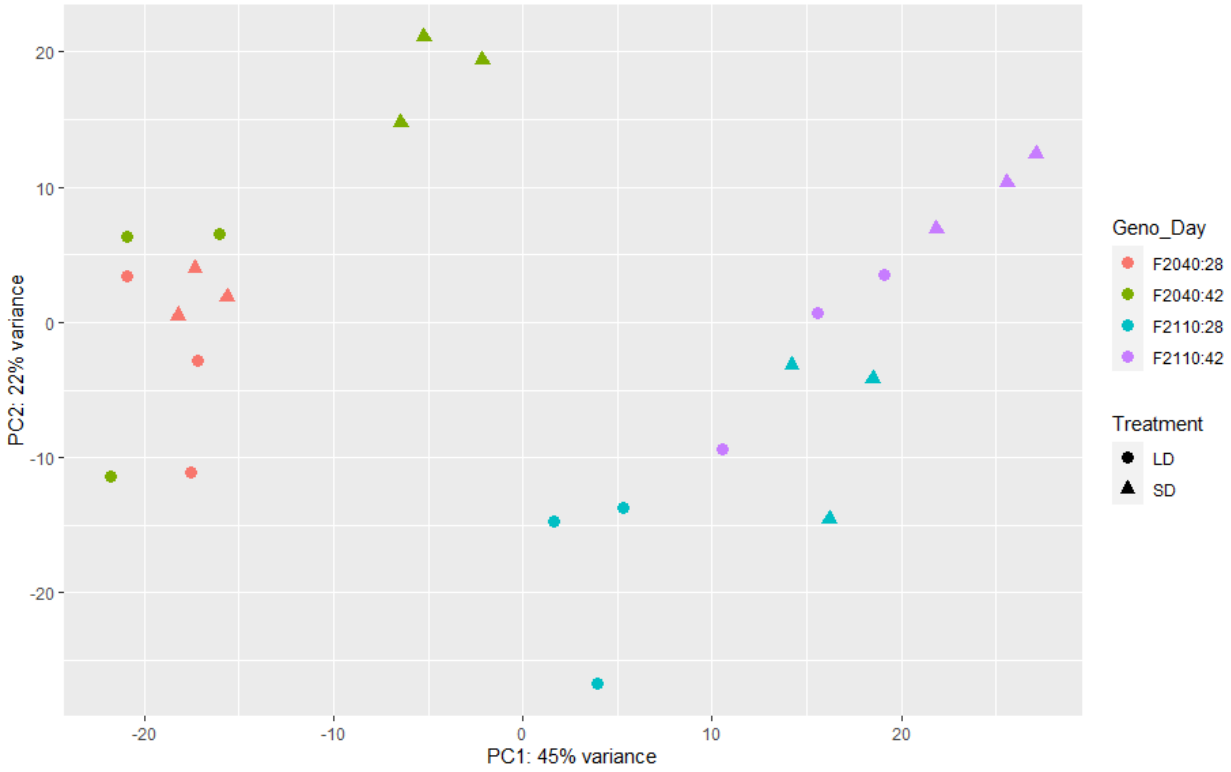


Figure 3.4 Heatmap of most variably expressed count data from photoperiod treated bud tissue. Expression transformed by variance stabilizing methods (ntd) to be more comparable between libraries. Intensity of color represents the general level of gene expression and column bars indicate genotype, time period, and treatment factor levels for each sample.

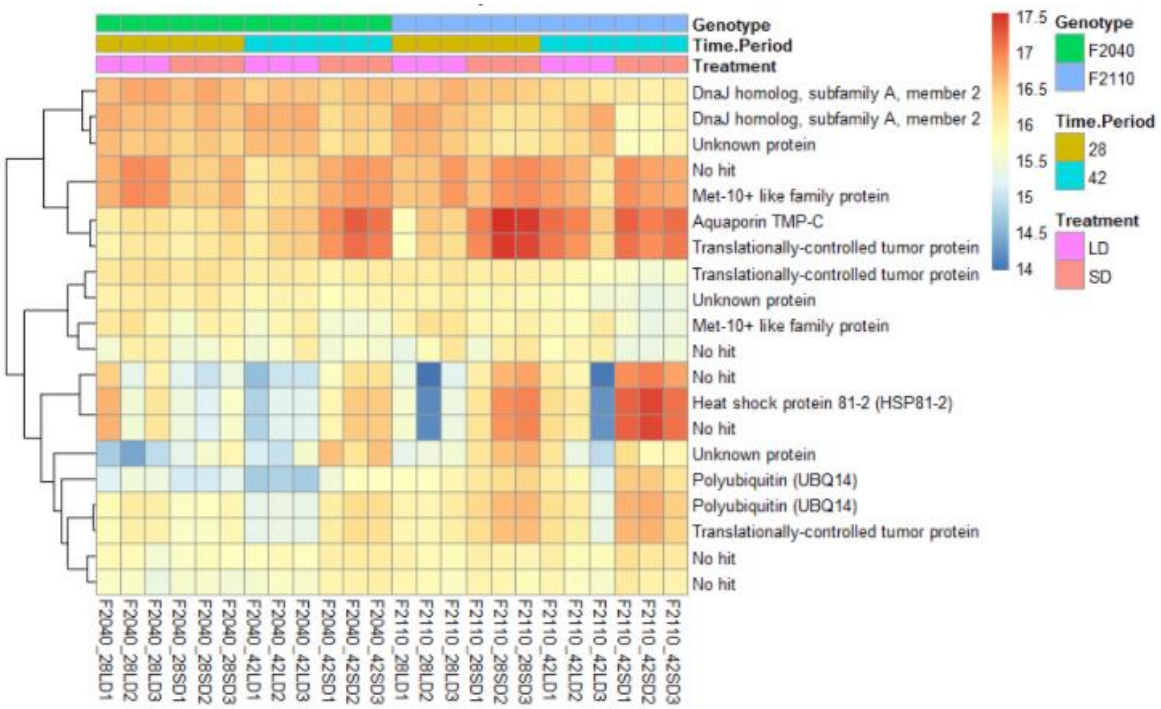


Figure 3.5 MA plots show that differential expression is not influenced by count number. Ntd transformed data sets both show that transformation does not affect log fold change to count ratio. Differential expression of genes at 28 days in the F2-040 genotype is plotted on the y-axis.

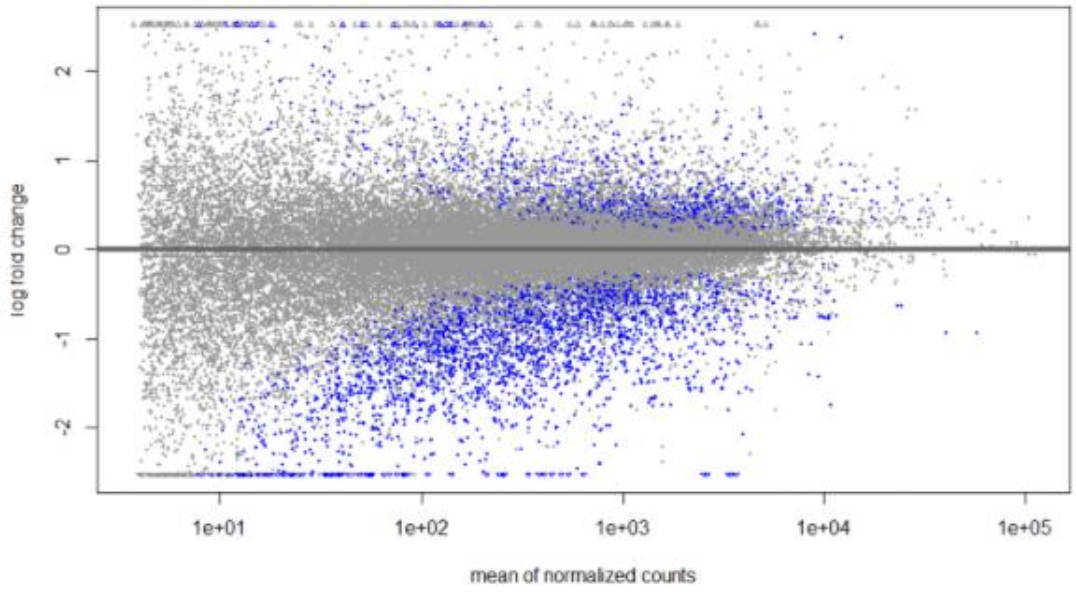


Figure 3.6 Differential expression in F2-040 and F2-110 under SD photoperiod treatment at 28 days of treatment. DEG represents up- or down-regulation of genes in SD photoperiod treatments in pure counts of genes.

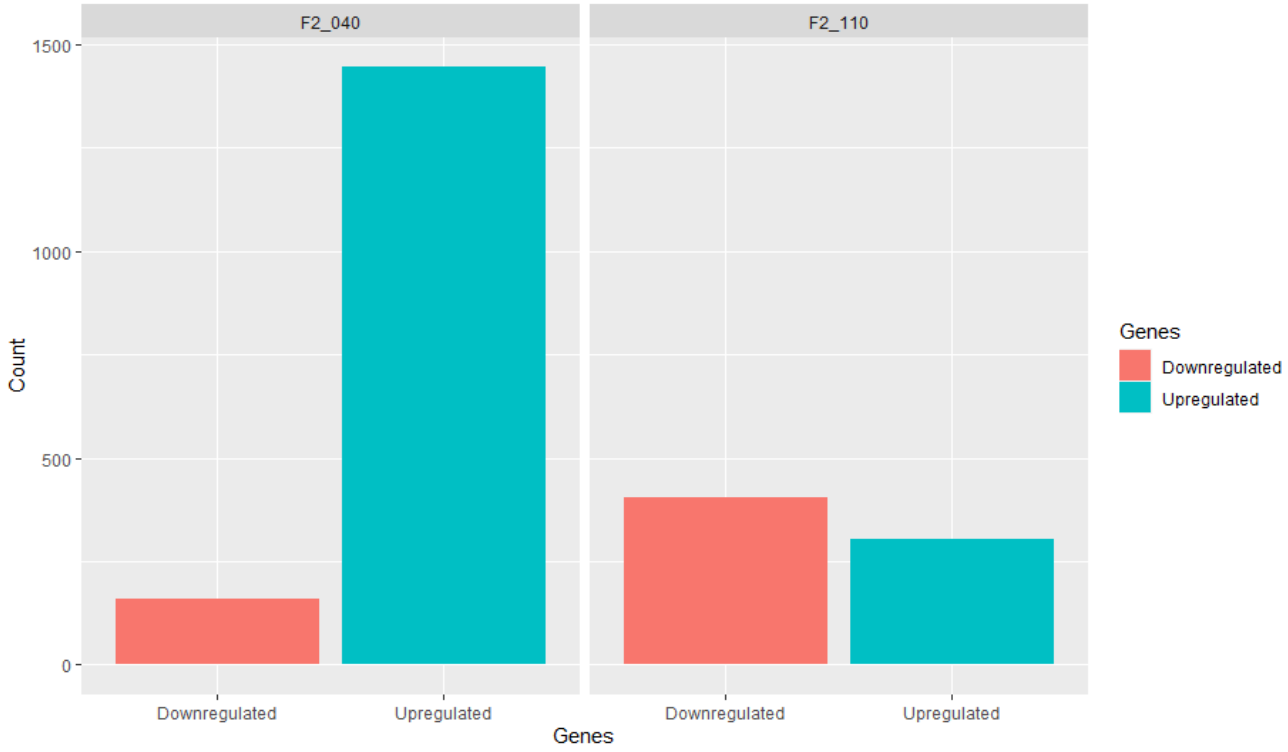


Figure 3.7 Network involvement of differentially expressed genes in F2-110 and F2-040 at 28 days of SD photoperiod treatment. Pure count of genes in each pathway represented on the x-axis separated by up- or down-regulation.

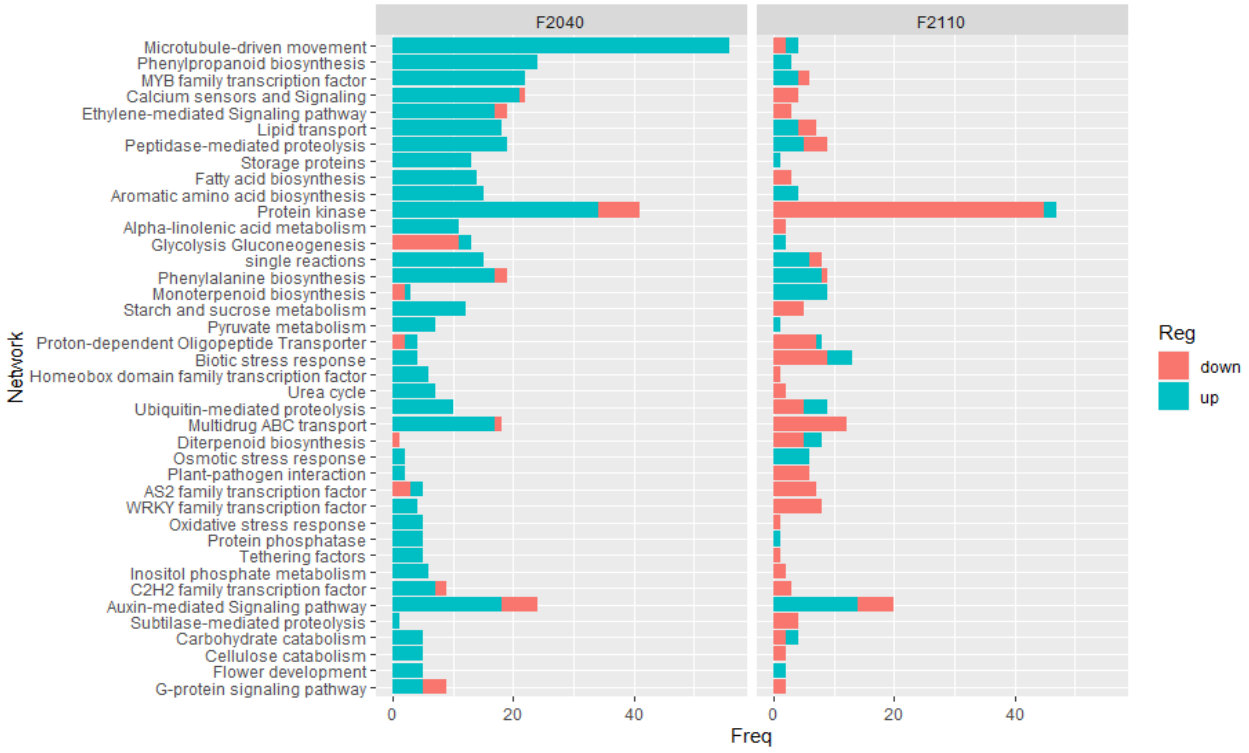


Figure 3.8 Pathway enrichment for differentially expressed genes at 28 days of photoperiod treatment. Enrichment of DEG in genotypes a) F2-040 or b) F2-110 after 28 days of SD photoperiod treatment using GSEA. Activation or suppression of pathways based on up- or down-regulation in SD photoperiod treated samples.

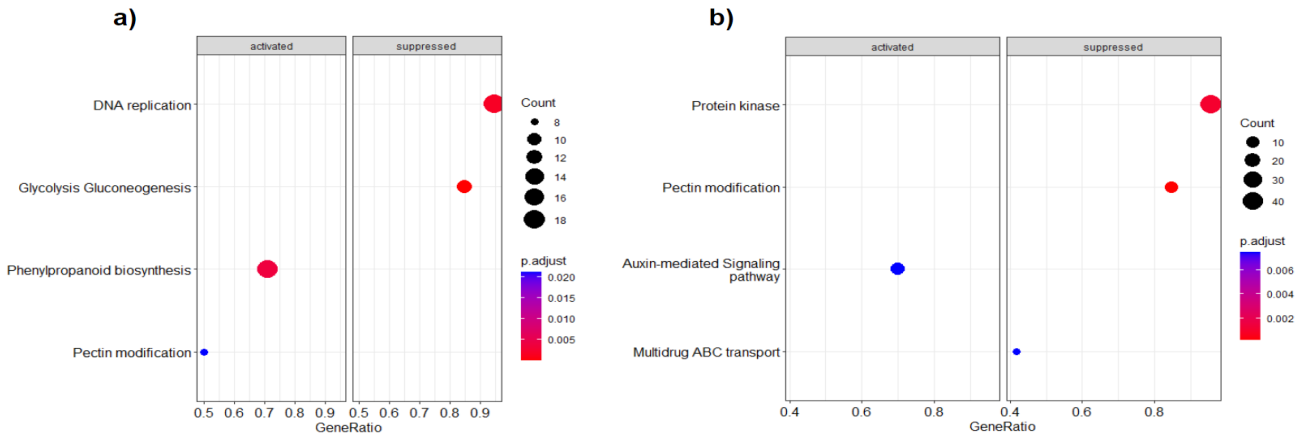


Figure 3.9 Differential expression in F2-040 and F2-110 under SD photoperiod treatment at 42 days of treatment. DEG represents up- or down-regulation of genes in SD photoperiod treatments in pure counts of genes.

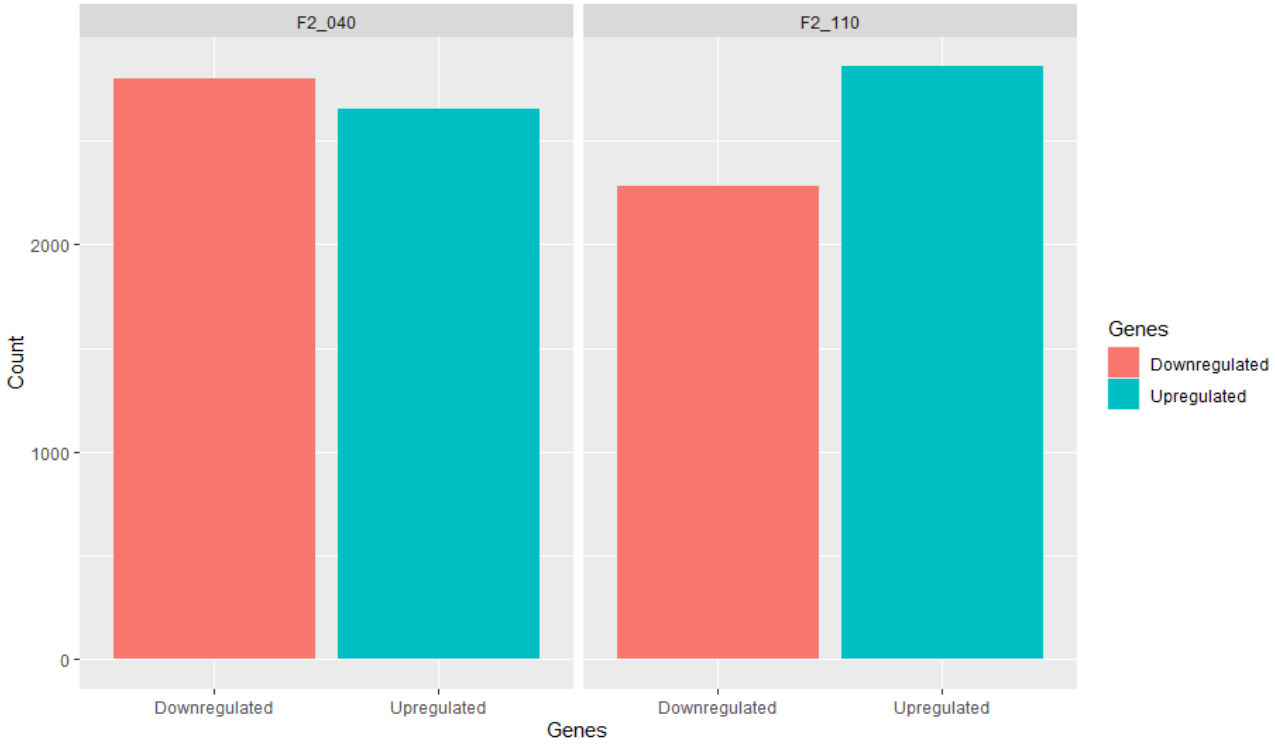


Figure 3.10 Network involvement of differentially expressed genes in F2-040 and F2-110 at 42 days of SD photoperiod treatment. Pure count of genes in each pathway represented on the x-axis divided by up- or down-regulation.

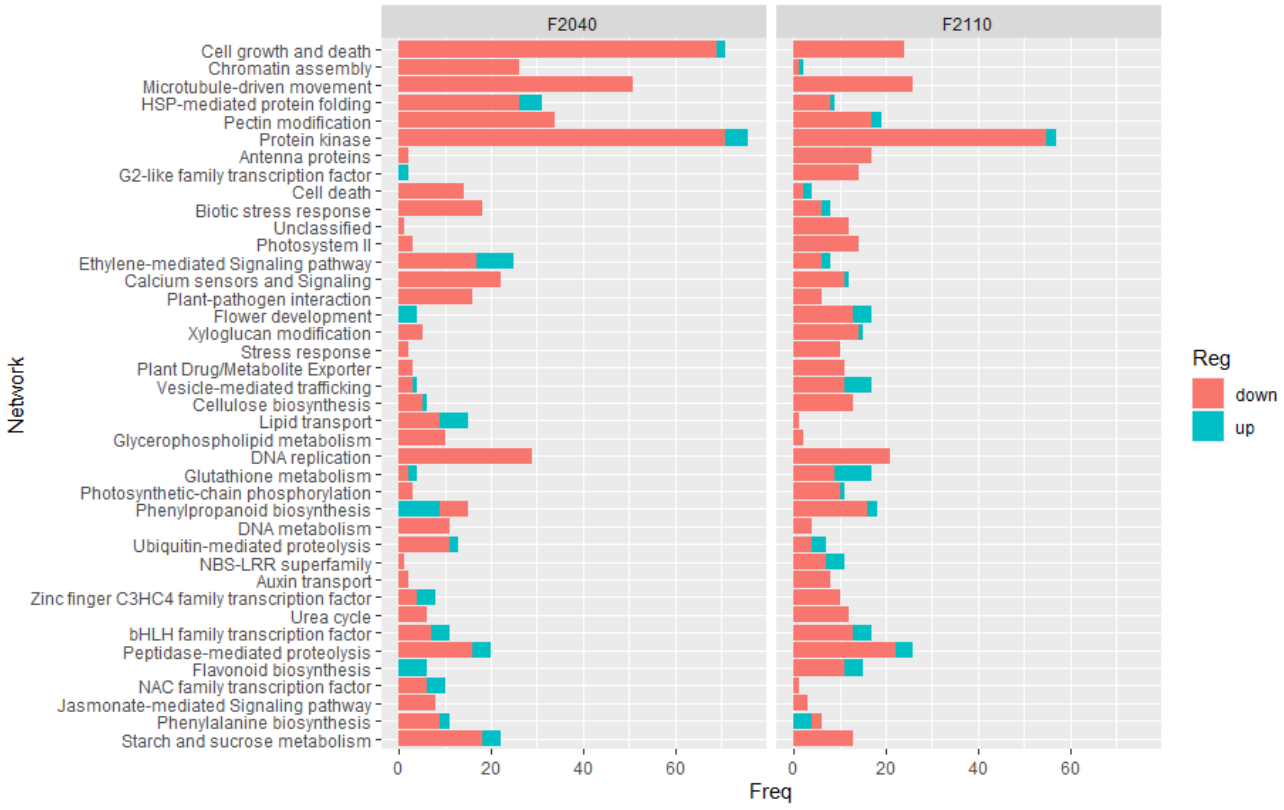


Figure 3.11 Enrichment of pathways in differentially expressed genes in the F2-040 genotype at 42 days using GSEA. Activation or suppression of pathways based on up- or down-regulation in SD photoperiod treated samples.

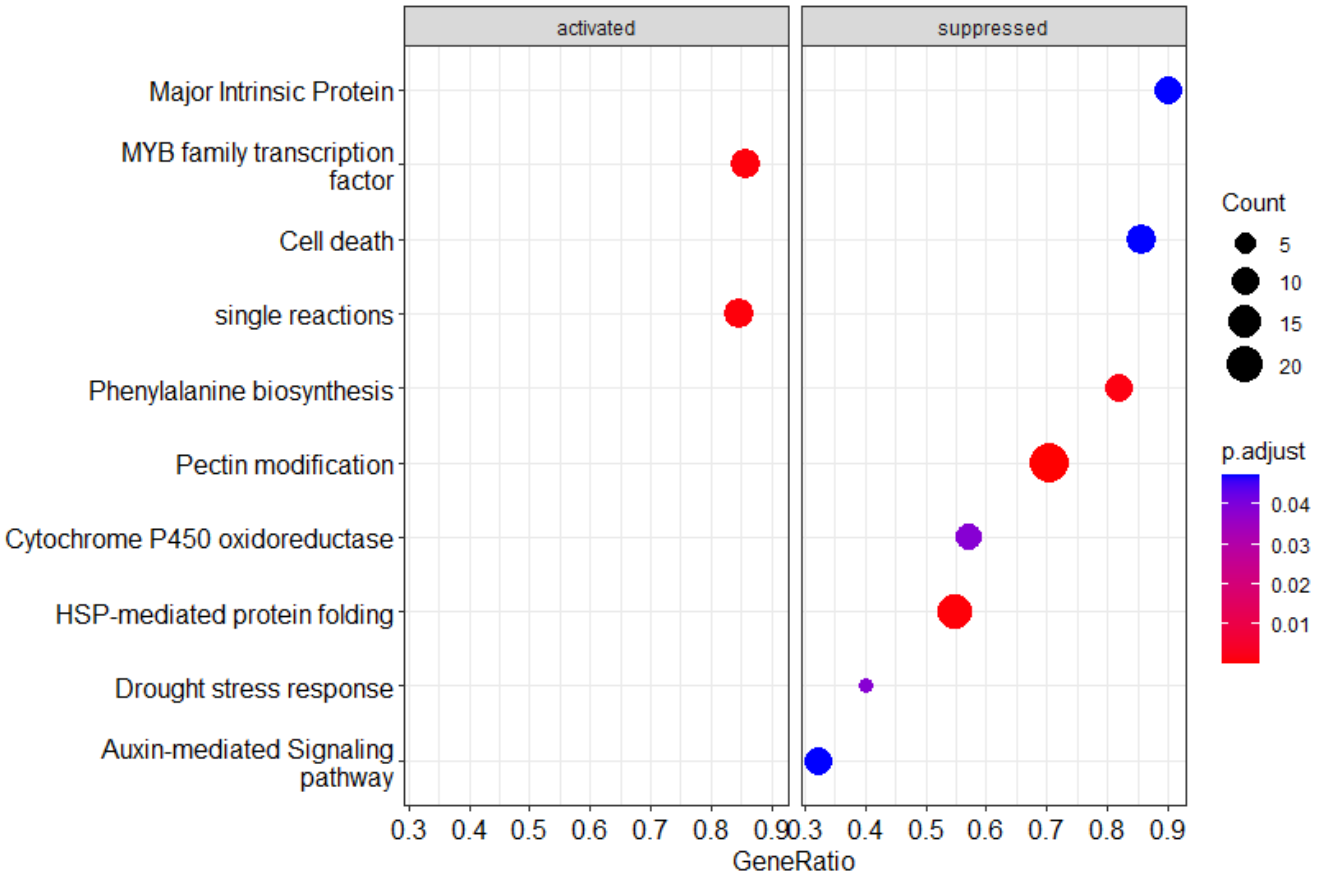


Figure 3.12 Differential expression between F2-040 and F2-110 using a mixed model as seen in section 3.3.4. DEG represents up- or down-regulation of genes in F2-110 genotypes relative to F2-040 represented in pure counts of differentially expressed genes.

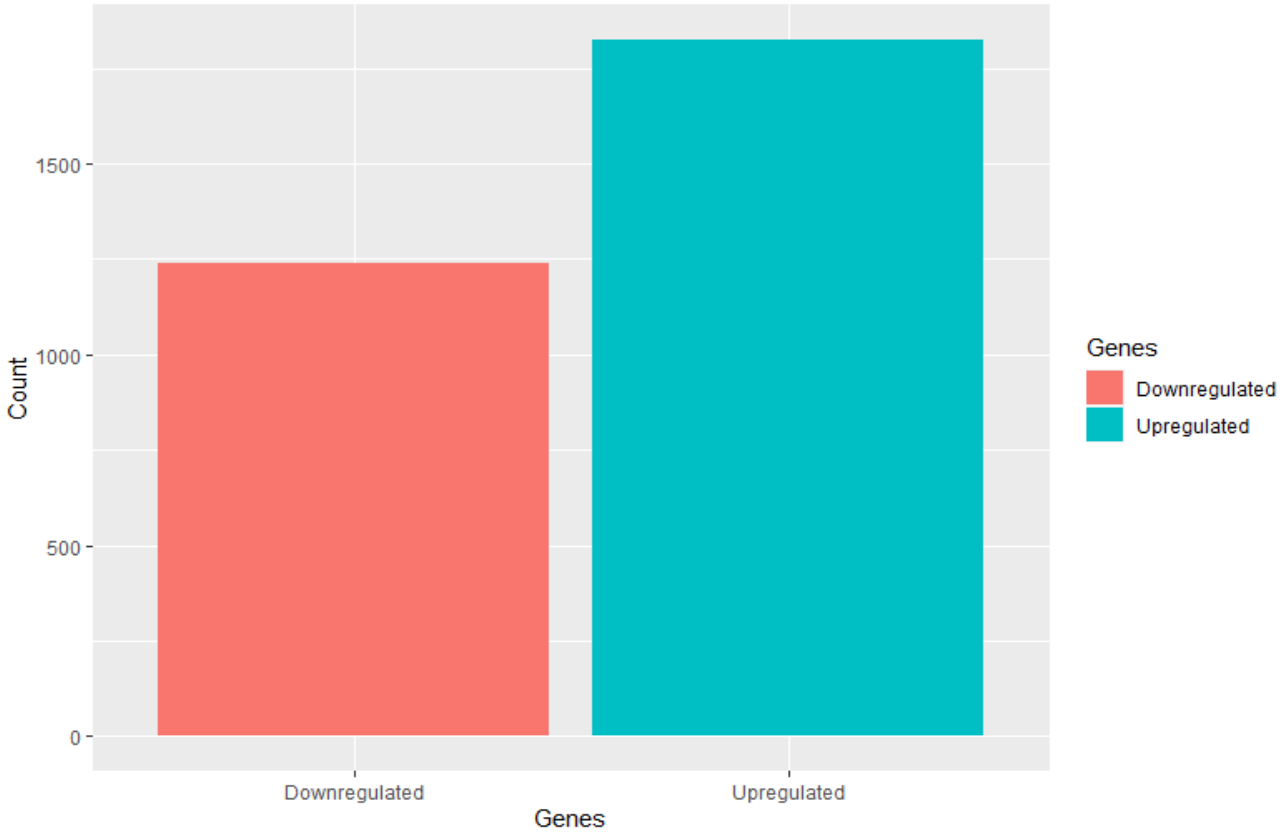


Figure 3.13 Number of differentially expressed genes in pathways between F2-110 and F2-040 genotypes. Pure count of genes in each pathway represented on the x-axis divided by up- or down-regulation in the F2-110 genotype.

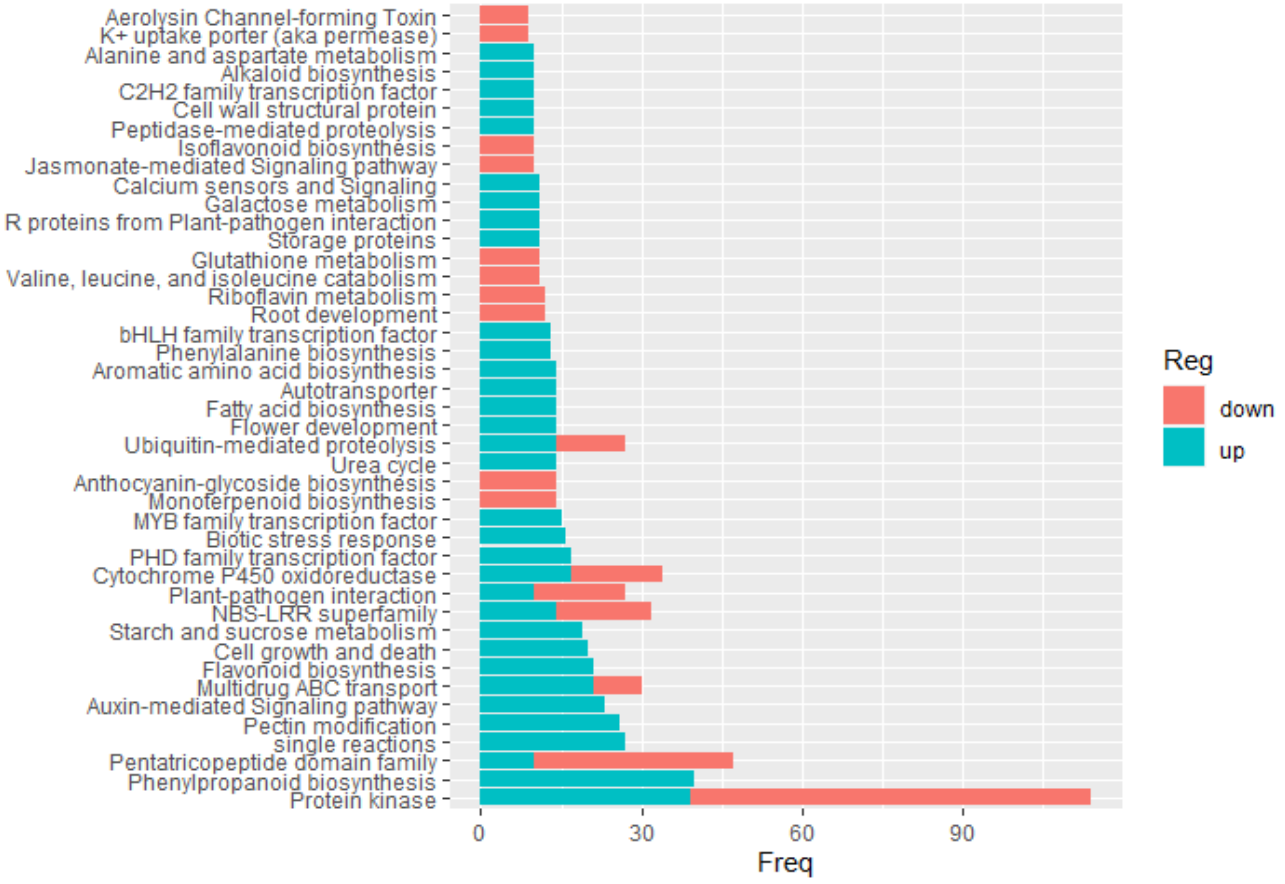


Figure 3.14 Enrichment of pathways in differentially expressed genes between F2-040 and F2-110 genotypes. Activation or suppression of pathways based on up- or down-regulation in the F2-110 genotype relative to F2-040.

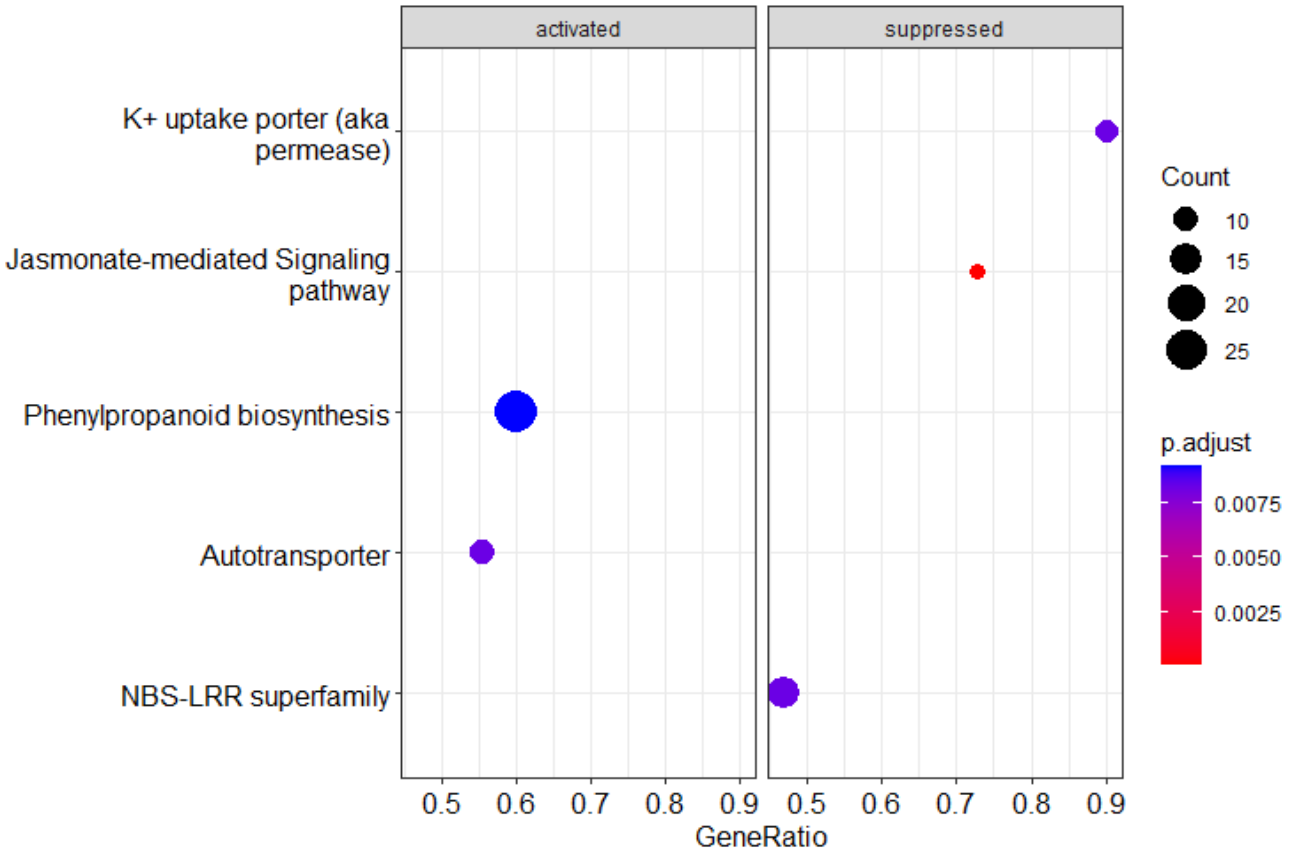


Figure 3.15 Heatmap reveals correlation between weighted gene co-expression clusters and photoperiod treatments. Pearson correlations to photoperiod treatment displayed in red and blue to represent positive and negative values respectively. Modules represented on the y-axis by randomly assigned color identifier.

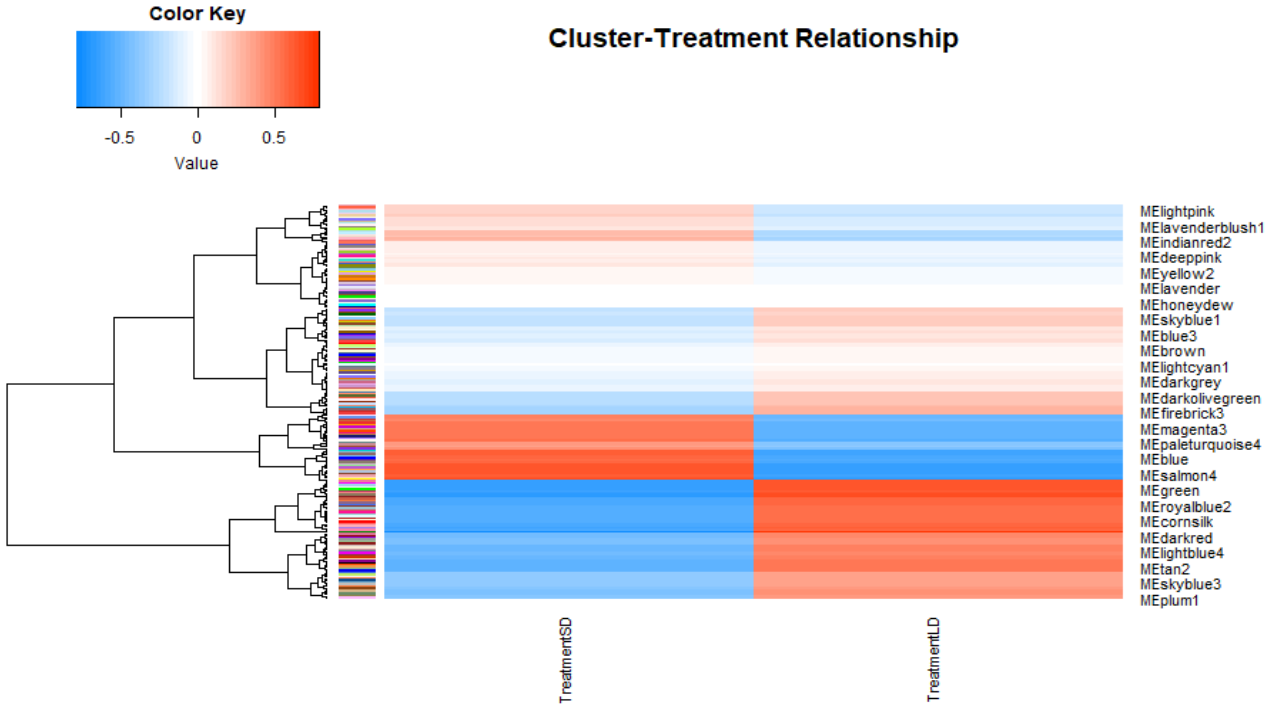


Figure 3.16 Clusters highly correlated (Pearson correlation > 0.7) to photoperiod treatment show gene expression specific to endodormancy conditions. Ntd normalized expression value of each gene is plotted on the y-axis as a line plot between each condition to show patterns of expression for each cluster. Expression of genes in clusters correlated to a) SD photoperiod and b) LD photoperiod. Network involvement of genes from clusters correlated to c) SD photoperiod d) LD photoperiod.

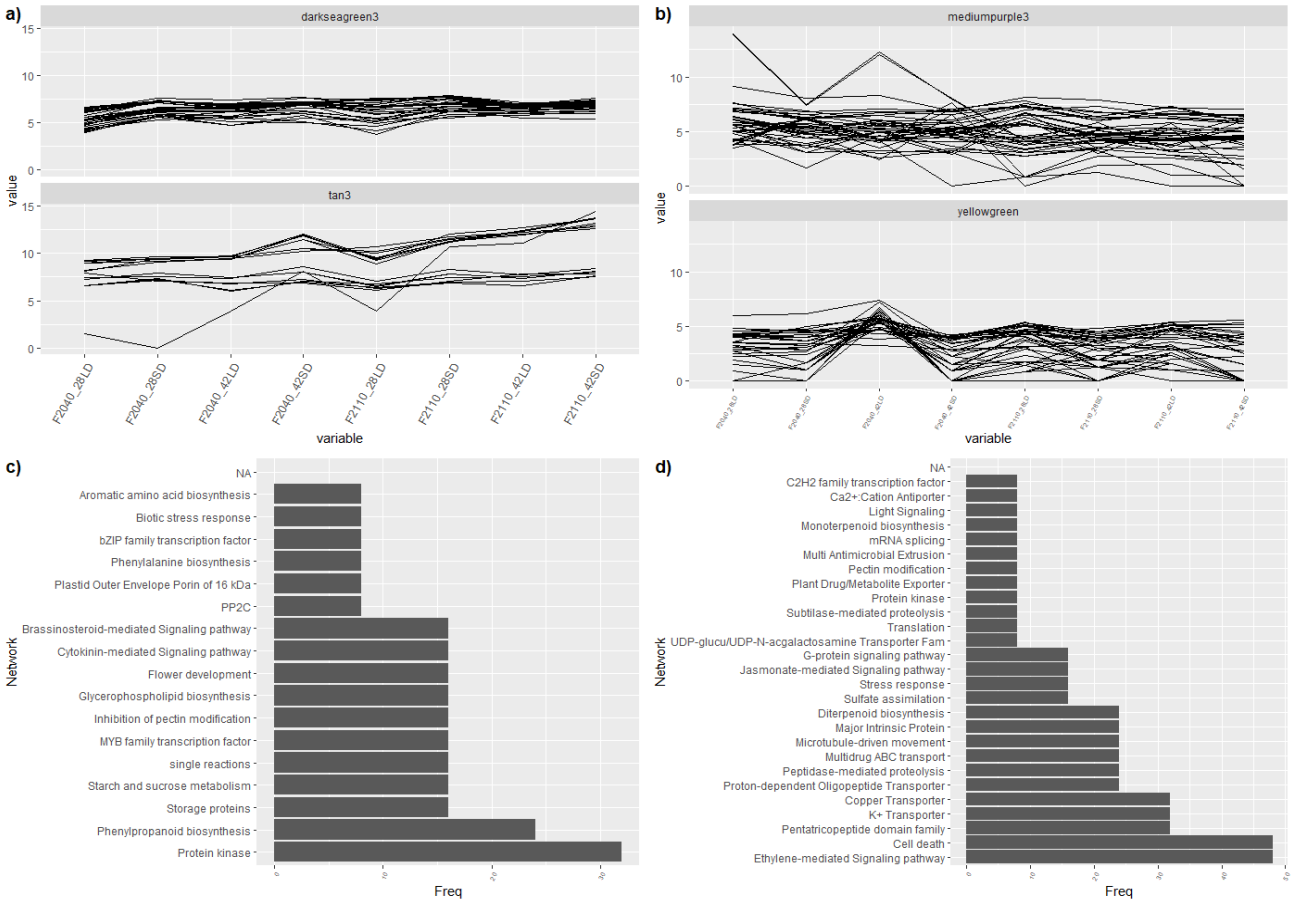


Figure 3.17 Heatmap reveals correlation between weighted gene co-expression clusters and genotypes. Pearson correlations to genotype displayed in red and blue to represent positive and negative values respectively. Modules represented on the y-axis by randomly assigned color identifier.

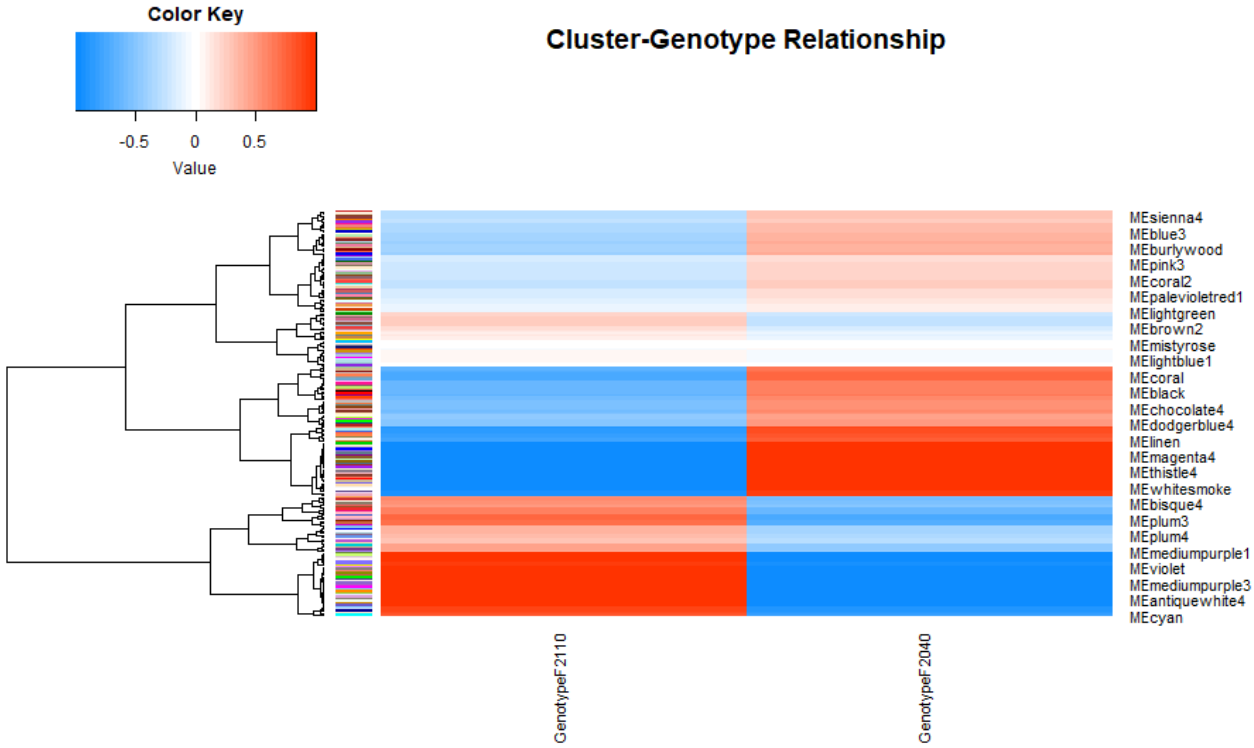
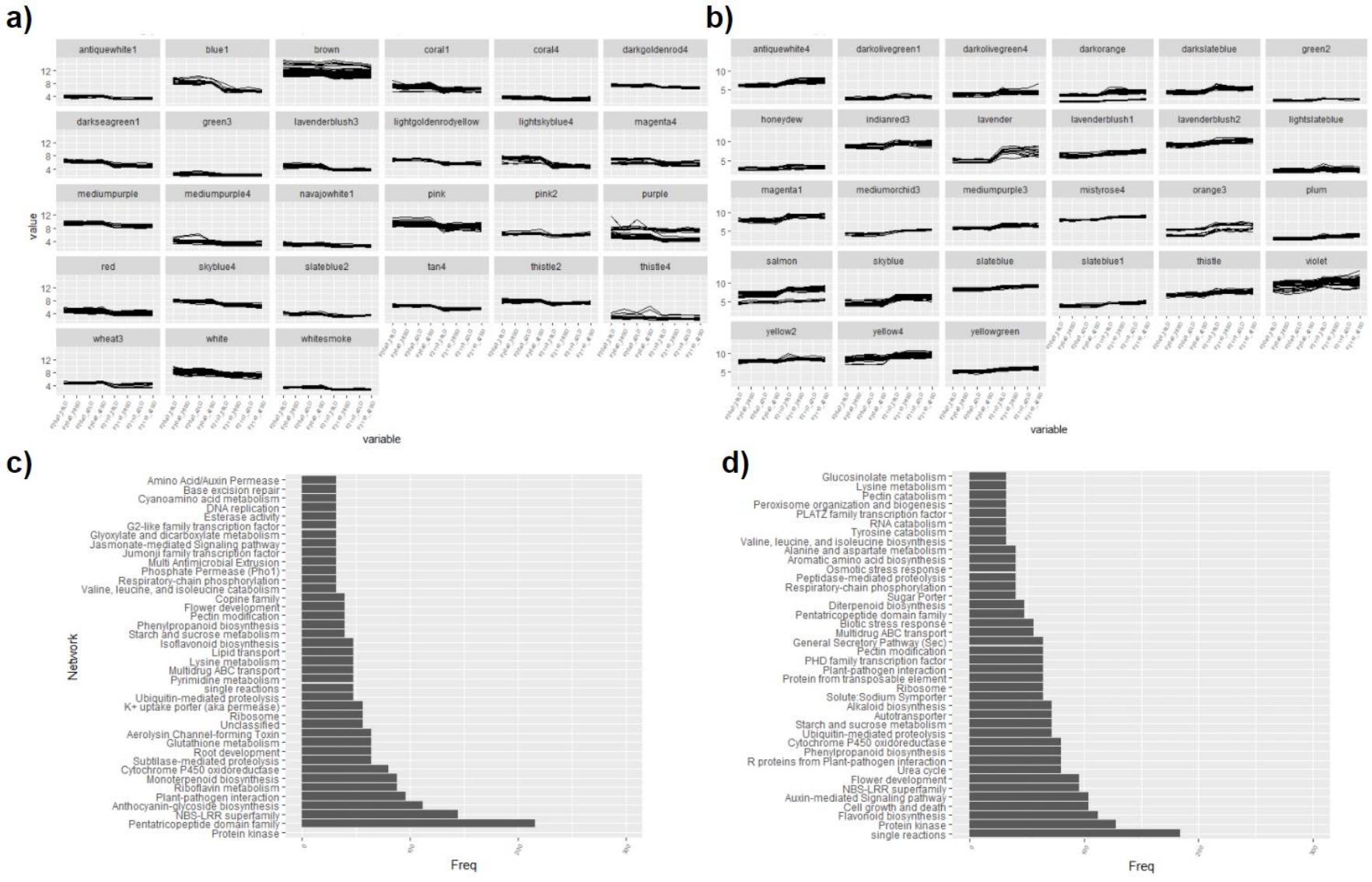


Figure 3.18 Clusters strongly correlated (Pearson correlation > 0.9) to genotypes show expression specific to either F2 genotype. Ntd normalized expression value of each gene is plotted on the y-axis as a line plot between each condition to show patterns of expression for each cluster. Expression of genes in clusters correlated to a) F2-040 and b) F2-110. Network involvement of genes from clusters in c) F2-040 and d) F2-110.



4 Prediction of miRNA and transcription factor regulators in grapevine bud endodormancy

4.1 Abstract

Regulation of the endodormant state in buds relies on a combination of temperature and photoperiod signals. The molecular pathways that enable transition of the bud tissue to a winter resistant organ have yet to be identified. We sought to investigate the post-transcriptional and pre-transcriptional regulators of the endodormant bud in short day (SD) induced endodormancy. We predicted roughly 200 regulatory miRNAs' specific to the novel *Vitis riparia* genome and used differential expression to identify several miRNA families needed for bud dormancy transition. miR families like miR166 and miR167 control important cell cycle genes that are needed for suppression of cell division and cell growth. Cross regulation of families miR166, 167, and 156 with MYB3R1 is needed to regulate cell cycle control genes. When we enriched F2-110 specific gene clusters for regulatory cis binding elements we found that PLT1, an auxin responsive DNA binding element, to be the most significantly enriched transcription factor motif. This could suggest a higher sensitivity to auxin in *V. riparia* that enables it to enter endodormancy at earlier stages than other *vitis* spp. These miRNAs and transcription factors represent targets for future research on endodormancy induction.

4.2 Introduction

The endodormant state is a complex process of tissue differentiation that maintains a dormant shoot apical meristem for the entirety of the winter season (Horvath et al. 2003). This is antithetical to the state of terminal buds during growing seasons which is termed paradormancy. During spring and summer these flowering buds in most tree species remain mitotically active and capable of forming new branches or flowers. Grapevine demonstrates similar bud activity with axillary buds that form on each node being capable of flowering or growing into new vines. At the end of winter, in response to longer days and higher temperatures, grapevine buds break and release new shoots. Proximal buds during paradormancy experience apical dominance asserted by distal buds that keep the tissue in a semi-dormant state, still capable of producing shoots (LANG and G. A 1987). It is thought that auxin, as well as other signaling hormones and metabolites are responsible for maintaining this apical dominance (Beveridge, Symons, and Turnbull 2000). However, it is not believed that apical dominance regulates bud growth in the endodormant state.

Endodormancy is slightly less understood and is believed to be controlled by a combination of internal and external signals including ABA, among other hormones, and MADS-Box transcription factors (Horvath et al. 2003). Photoperiod plays a big role in these signals, influencing decreasing levels of Indoleacetic Acid (IAA) and increasing levels of ABA (L. Zhang et al. 2012). ABA can affect cell cycle genes through regulation of DORMANCY ASSOCIATED MADS-Box (DAM) genes and SHORT VEGETATIVE PHASE (SVP)/AGAMOUS-

Like (AGL) transcription factors (Pan et al. 2021). Many of these proteins are downstream of the cold response ABA/CBF pathway, but are not dependent upon CBF to induce endodormancy in response to SD photoperiod (Li et al. 2019; Vergara, Noriega, and Pérez 2021). Many DAMs are known to control histone methyltransferases for H3K27me3 such as DEMETER-Like protein 3 (DML3) (Chinnusamy, Gong, and Zhu 2008).

Post-transcriptional regulation of gene expression has been found to play a significant role in gene expression during endodormancy transition. CBF and DAM genes have both been shown to be regulated by miR6390 which is upregulated during dormancy release (Niu et al. 2016). (Huo, Wei, and Bradford 2016), recognized that miR156 and miR172 regulate seed dormancy in a DELAY OF GERMINATION 1 (DOG1) dependent manner. Another study in peony found that miR156, 159, 167, and 172 are all upregulated on release from endodormancy. Unfortunately, no small RNA-Seq studies have looked at the induction into endodormancy. In the present study, we perform small-RNA seq to identify regulators of the process. Combined with analysis of previous RNA-seq studies, we will correlate these findings with transcription factor enrichment to get a complete picture of the regulation of endodormancy induction in *V. riparia*.

4.3 Materials and Methods

4.3.1 Plant materials and photoperiod treatments

Six-year-old spur pruned ecodormant *V. riparia* ‘Manitoba 37′ Michx. (PI588259)) vines were repotted and grown in long photoperiod (LD, 15 h) at

25/20 ± 3 °C day/night temperatures with 600–1400 mol m−2 s−1 photosynthetic photon flux in a climate-controlled unshaded glass greenhouse (EnTech Control Systems Inc., Montrose, MN, USA) in Brookings, South Dakota (44.3 N). Grapevines were grown for 30 days post bud break reaching shoot lengths of 10–15 nodes. Three replicate ten-vine experimental units were randomly assigned to each photoperiod treatment of continued LD (paradormancy) or short photoperiod (SD, 13 h; endodormancy) as previously described (Fennell et al. 2015). After 28 days, three replicate bud samples for each photoperiod were harvested into separate tubes of liquid nitrogen and stored at −80 °C until small RNA or total RNA extraction.

4.3.2 Small RNA library construction, sequencing and processing

Total RNA was isolated using Plant RNA Reagent (Invitrogen/Thermo Fisher Scientific, Carlsbad, CA, USA), and low molecular weight RNA was purified from total RNA by PEG8000/NaCl precipitation (Accerbi et al. 2010; Meyers and Green 2010). Two small RNA libraries were constructed: (1) a pool of LD and (2) a pool of SD buds as described (Lu, Meyers, and Green 2007) with the following: RNA oligos for RNA ligation:5′ RNA adapter: 5′-GUUCAGAGUUCUACAGUCCGACGAUC-3′, 3′ RNA adapter: 5′-pUCGUAUGCCGUCUUCUGCUUG-idT-3′ (p, phosphate; idT, inverted deoxythymidine); DNA oligo for reverse transcription: RT-primer (5′-CAAGCAGAAGACGGCATACGA-3′); DNA oligos for PCR amplification:5′ PCR primer (5′-AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAGTCCGA-

3′), 3′ PCR primer (5′-CAAGCAGAAGACGGCATACGA-3′). The Illumina sequencing of small RNA libraries from LD and SD led to the generation of 7,761,608 and 6,530,774 non redundant sequences, respectively. Filtering the low quality reads and adaptor contaminations resulted in 7,591,740 (LD) (total read count 24,452,821) and 6,352,594 (SD) (total read count 22,206,946) unique sequences that were retained for further analysis.

4.3.3 miRNA prediction and differential expression analysis

All 13,944,334 unique small RNA reads were analyzed against the *V. riparia* genome. We found 10,665,915 of the reads aligned to the sequenced assembly and we extracted 350 bp to either side of the alignment. After filtering for size and other non-coding RNA (ncRNA) classes, we ran potential sequences through the miRdeep workflow (An et al. 2013) resulting in 7,901 predicted pre-miRNA sequences. Further filtering by the structural Support Vector Machine (SVM) based prediction method miRfinder (Huang et al. 2007) provided us with a final set of 264 regulatory sequences. We used five different machine learning algorithms (Triplet-SVM plant and animal (Xue et al. 2005); HeteromiRPred (Lertampaiporn et al. 2013); miRPred (Brameier and Wiuf 2007); and RegSVM (Tran et al. 2015)) to validate our predicted sequences. Predicted regulatory sequences were grouped by miR family and statistical analysis of differential expression determined significance of log fold change.

4.3.4 Target prediction for abundant miRNAs

Empirical parameters were used with an in-house perl script to run Patscan (http://blog.theseed.org/servers/2010/07/scan-for-matches.html accessed on 4 October 2019) and RNAduplex (Lorenz et al. 2011) for recognition of potential targets of abundant miRNAs using *V. vinifera* as reference mRNAs. The empirically inferred parameters are tuned with maximum one mismatch at position 2–9, no mismatch at position 10–11 and 4 mismatches from position 12 to end (Schwab et al. 2005; Allen et al. 2005). The output was parsed to identify hits on complementary strands with <3 consecutive mismatches and relative minimum free energy (MFE) ≥70% compared to perfectly complementary target genes. Cytoscape software was used to visualize the miRNA-target regulatory networks (Shannon et al. 2003).

4.3.5 Transcription factor motif enrichment

V. vinifera specific transcription factor, regulated gene sets from enriched motif predictions in TFDB (downloaded 12/10/2018) were used in a GSEA analysis through the clusterProfiler package (Yu et al. 2012). Predictions of protein interactions in *V. vinifera* were obtained from BIOgrid and limited to interactions with a confidence score greater than 700. Protein interactions combined with Protein-DNA interactions and miRNA-target predictions were constructed into a network using Cytoscape complimented with RNA-seq log fold change scores. The promoter regions of clustered differentially expressed genes were extracted and analyzed through programs in the Meme suite (CentriMo

(Bailey and Machanick 2012) and FIMO (Grant, Bailey, and Noble 2011)) to enrich DNA-binding motifs.

4.4 Results

4.4.1 Novel miRNA predicted from sequenced bud small RNA

Generation of small RNA from Illumina sequencing resulted in roughly 12,000,000 unique reads, the majority of which likely originated from the degradome and are non-regulatory in nature (Lu, Meyers, and Green 2007). To identify the sequences that are regulatory we utilized sequence mapping techniques to predict pri-miRNA that formed hairpin loop structures. We were able to identify 57 precursor sequences with miRNA's that annotated exactly to previously known sequences, 200 that were similar to known sequences, and 7 that were novel previously uncharacterized (Fig 4.1). Because prediction relies only on characterization of hairpin structure and other biochemical aspects, we utilized machine learning tools to validate predicted sequences (Fig 4.2a). A combination of 5 machine learning algorithms led to a consensus of 99 validated miRNA sequences (Fig 4.2b). Three of these sequences were novel miRNA, and an examination of the secondary structure showed a typical hairpin consistent with confirmed regulatory sequences (Fig 4.3).

4.4.2 Differential expression of miRNA in endodormant bud

miRNAs are a post-transcriptional regulator of the endodormant state and increased expression of this class of small RNAs are associated with

downregulation of target genes important for endodormancy induction. When we examined the differential expression of predicted miRNAs at 28 days of SD photoperiod in *V. riparia* buds, we found that miR166 and 167 were the most significantly downregulated (Appendix 1f). We also found that while most predicted miRNAs were downregulated under SD photoperiod conditions (65%), suggesting that an increase in expression of target genes is required for endodormancy induction, miR156 and miR397 were among the few miRNAs that were upregulated. miRNAs target specific genes for downregulation, so we examined the inverse expression patterns of differentially expressed mRNAs targeted by identified miRNAs.

A previous RNA-seq study (Smita et al. 2021) found that twice as many genes were upregulated as were downregulated during endodormancy (Fig 4.4). Analysis of the inverse expression of these genes found that downregulation of miR166 and miR167 resulted in an increase in expression of thaumatin and powdery mildew resistance genes while upregulation of miR156 correlated with a downregulation of SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 9 (SPL9) and other cell cycle control genes.

4.4.3 Enrichment of transcription factor regulatory motifs during endodormancy transition

We employed enrichment methods of differentially expressed genes to identify important transcription factors. GSEA using regulatory gene sets found that MYB3R1, Heat Shock Transcription Factor B2B (HSTFB2B), and Teosinte

Branched 2 (TCP2) (Fig 4.5). MYB3R1 is an important regulator of the cell cycle process in endodormancy and has significant cross-talk with differentially expressed genes and miRNAs (Fig 4.6). AURORA1, a gene downstream of MYB3R1, was found to be downregulated downstream of MYB3R1 and miR156i upregulation. We then looked at the enrichment of motifs in gene clusters that were important for SD photoperiod induced bud dormancy, as identified in Chapter 3 (Fig 3.16). Many enriched motifs were related to homeobox genes or NAC, REDUCED VERNALIZATION RESPONSE (VRN1), and Histone Acetyltransferase 5 (HAT5) transcription factors (Table 4.1). When we examined relative enrichment of transcription factors in F2-110 gene clusters we discovered a very significant enrichment of PLT1 using Arabidopsis binding motifs (Table 4.2) and SPL9 using Vinifera binding motifs (Table 4.3). Expression of PLT1 appeared to be high in F2-040 at 28 days of SD photoperiod but returned to normal levels by 42 days, while expression in F2-110 remained stable at both time points (Fig. 4.7).

4.5 Discussion

The regulation of the endodormant process has long been a secret that plant physiologists have attempted to solve. Many expression studies have identified several regulators in their roles on bud endodormancy maintenance like DAM genes and ABA related genes (Pan et al. 2021); however, no study has connected upregulated genes and the factors that control their expression during dormancy transition, such as photoperiod or temperature. In our experiment, we

have looked at pre-transcriptional and post-transcriptional regulators of gene expression. We utilized multiple models to examine bud dormancy including SD induced dormancy of *V. riparia* and SD induced dormancy of F2 mapping genotypes.

Small RNA sequencing of total RNA from endodormant *V. riparia* resulted in about 12,000,000 unique reads less than 50 bp long. Alignment of reads to conserved miRNAs from various plant species yielded only about 0.5% mapping and we saw similar low alignment in other miRNA-sequencing experiments (Niu et al. 2016). We decided to first use various methods of machine learning to validate *in silico* which reads were regulatory in nature. This led to the prediction of 264 regulatory miRNAs, and we validated 37% of those using 5 different models. We identified 7 novel regulatory sequences that can be helpful for future research in miRNA regulation. Only two novel sequences were differentially expressed in endodormant conditions, however many previously characterized miRNA families were differentially expressed like miR156, 166, 167, and 169. Some of these miRNAs are very important for many dormancy-like processes. miR156 is a key regulator of vernalization and flowering phase transition by regulation of the SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) genes (Wang, Czech, and Weigel 2009; Spanudakis and Jackson 2014). As one of the only miRNAs upregulated in the dormant condition it seems to be needed to downregulate SPL4 expression. The prevailing theory of flowering control involves increased miR156 abundance in the vegetative phase reducing SPL abundance which is needed for activation of flowering genes like LEAFY (LFY)

(Spanudakis and Jackson 2014). The fact that miR156 is also affected by cold stress could be one of the reasons why axillary buds need cold weather for chilling fulfillment in bud release (Zhang et al., 2018; Myking and Heide 1995). Other miRNAs like miR166 and miR167 have come up in other endodormancy studies in pear and peony (Bai et al. 2016; Y. Zhang et al. 2018). miR167 is known to target Auxin Response Factor (ARF) genes which regulate vegetative growth in the SAM (Mallory, Bartel, and Bartel 2005). We identified that miR 166 and 167 could potentially inversely regulate the expression of Thaumatin and other biotic stress resistance genes, and the downregulation of miR166/167 which results in the upregulation of these biotic stress resistance genes is congruent with our results finding that biotic stress response genes are upregulated in the SD photoperiod treated buds during endodormancy transition (Fig 3.7).

There is strong evidence that regulation of endodormancy requires a multitude of different factors. Hormones like ABA, Ethylene, JA, and Auxin have strong connections to dormancy phase change (Liu and Sherif 2019). We have also found evidence that epigenetic changes may be connected to cold temperature related changes in gene expression (Singh et al. 2019). This included increased expression of Chromodomain Methyltransferases (CMT) that altered the epigenetic state of DNA. Transcription factors involved in endodormancy may lead to changes in gene expression downstream of cold response and photoperiod signaling. We performed GSEA enrichment of transcription factor motifs in differentially expressed genes in endodormant *V.*

riparia and found that MYB3R1, HSTFB2B, and TCP2 were significantly enriched in suppressed genes. We found that there was regulation of interacting proteins upstream of MYB3R1 by differentially expressed miRNAs. We found that increased activity of MYB3R1 was concomitant with reduced expression of cell cycle control genes. When we looked at motifs of genes belonging to F2-110 expression clusters, there was significant enrichment of PLT1 and SPL9. Earlier we showed that downregulation of miR156 during endodormancy induction was inversely correlated with upregulation of SPL9, a process that induced season phase transition in *V. vinifera* buds (Díaz-Riquelme et al. 2012). PLT1 is auxin regulated (Ding and Friml 2010) and when we examined count data, we saw that there was an initial spike in expression of PLT1. This suggests that auxin sensitivity increases either because of or as a cause of endodormancy induction and then tapers into endodormant maintenance. The increase in expression of phenylpropanoid genes that we observed under endodormant conditions correlates with this hypothesis because phenylpropanoids relate to increased auxin transport (Peer and Murphy 2007). We believe that this represents a complex regulatory network, in which hormones, epigenetics, miRNA and transcription factors all cooperate to transition the buds into a vegetatively inactive SAM.

4.6 References

Accerbi, Monica, Skye A. Schmidt, Emanuele De Paoli, Sunhee Park, Dong-Hoon Jeong, and Pamela J. Green. 2010. “Methods for Isolation of Total

RNA to Recover miRNAs and Other Small RNAs from Diverse Species." Methods in Molecular Biology 592: 31–50.

Allen, Edwards, Zhixin Xie, Adam M. Gustafson, and James C. Carrington. 2005. "microRNA-Directed Phasing during Trans-Acting siRNA Biogenesis in Plants." Cell 121 (2): 207–21.

An, Jiyuan, John Lai, Melanie L. Lehman, and Colleen C. Nelson. 2013. "miRDeep*: An Integrated Application Tool for miRNA Identification from RNA Sequencing Data." Nucleic Acids Research 41 (2): 727–37.

Bailey, Timothy L., and Philip Machanick. 2012. "Inferring Direct DNA Binding from ChIP-Seq." Nucleic Acids Research 40 (17): e128.

Bai, Songling, Takanori Saito, Akiko Ito, Pham Anh Tuan, Ying Xu, Yuanwen Teng, and Takaya Moriguchi. 2016. "Small RNA and PARE Sequencing in Flower Bud Reveal the Involvement of sRNAs in Endodormancy Release of Japanese Pear (Pyrus Pyrifolia 'Kosui')." BMC Genomics 17 (March): 230.

Beveridge, C. A., G. M. Symons, and C. G. Turnbull. 2000. "Auxin Inhibition of Decapitation-Induced Branching Is Dependent on Graft-Transmissible Signals Regulated by Genes Rms1 and Rms2." Plant Physiology 123 (2): 689–98.

Brameier, Markus, and Carsten Wiuf. 2007. "Ab Initio Identification of Human microRNAs Based on Structure Motifs." BMC Bioinformatics 8 (December): 478.

Chinnusamy, Viswanathan, Zhizhong Gong, and Jian-Kang Zhu. 2008. “Abscisic Acid-Mediated Epigenetic Processes in Plant Development and Stress Responses.” Journal of Integrative Plant Biology 50 (10): 1187–95.

Díaz-Riquelme, José, Jérôme Grimplet, José M. Martínez-Zapater, and María J. Carmona. 2012. “Transcriptome Variation along Bud Development in Grapevine (Vitis Vinifera L.).” BMC Plant Biology 12 (October): 181.

Ding, Zhaojun, and Jirí Friml. 2010. “Auxin Regulates Distal Stem Cell Differentiation in Arabidopsis Roots.” Proceedings of the National Academy of Sciences of the United States of America 107 (26): 12046–51.

Fennell, Anne Y., Karen A. Schlauch, Satyanarayana Gouthu, Laurent G. Deluc, Vedbar Khadka, Lekha Sreekantan, Jerome Grimplet, Grant R. Cramer, and Katherine L. Mathiason. 2015. “Short Day Transcriptomic Programming during Induction of Dormancy in Grapevine.” Frontiers in Plant Science 6 (November): 834.

Grant, Charles E., Timothy L. Bailey, and William Stafford Noble. 2011. “FIMO: Scanning for Occurrences of a given Motif.” Bioinformatics 27 (7): 1017–18.

Horvath, David P., James V. Anderson, Wun S. Chao, and Michael E. Foley. 2003. “Knowing When to Grow: Signals Regulating Bud Dormancy.” Trends in Plant Science 8 (11): 534–40.

Huang, Ting-Hua, Bin Fan, Max F. Rothschild, Zhi-Liang Hu, Kui Li, and

Shu-Hong Zhao. 2007. “MiRFinder: An Improved Approach and Software Implementation for Genome-Wide Fast microRNA Precursor Scans.” BMC Bioinformatics 8 (September): 341.

Huo, Heqiang, Shouhui Wei, and Kent J. Bradford. 2016. “DELAY OF GERMINATION1 (DOG1) Regulates Both Seed Dormancy and Flowering Time through microRNA Pathways.” Proceedings of the National Academy of Sciences of the United States of America 113 (15): E2199–2206.

LANG, and G. A. 1987. “Dormancy : A New Universal Terminology.” HortScience: A Publication of the American Society for Horticultural Science 25: 817–20.

Lertampaiporn, Supatcha, Chinae Thammarongtham, Chakarida Nukoolkit, Boonserm Kaewkamnerdpong, and Marasri Ruengjitchatchawalya. 2013. “Heterogeneous Ensemble Approach with Discriminative Features and Modified-SMOTEbagging for Pre-miRNA Classification.” Nucleic Acids Research 41 (1): e21.

Li, Jianzhao, Xinhui Yan, Qinsong Yang, Yunjing Ma, Bo Yang, Juan Tian, Yuanwen Teng, and Songling Bai. 2019. “PpCBFs Selectively Regulate PpDAMs and Contribute to the Pear Bud Endodormancy Process.” Plant Molecular Biology 99 (6): 575–86.

Liu, Jianyang, and Sherif M. Sherif. 2019. “Hormonal Orchestration of Bud Dormancy Cycle in Deciduous Woody Perennials.” Frontiers in Plant

Science 10 (September): 1136.

Lorenz, Ronny, Stephan H. Bernhart, Christian Höner Zu Siederdissen, Hakim Tafer, Christoph Flamm, Peter F. Stadler, and Ivo L. Hofacker. 2011. “ViennaRNA Package 2.0.” Algorithms for Molecular Biology: AMB 6 (November): 26.

Lu, Cheng, Blake C. Meyers, and Pamela J. Green. 2007. “Construction of Small RNA cDNA Libraries for Deep Sequencing.” Methods 43 (2): 110–17.

Mallory, Allison C., David P. Bartel, and Bonnie Bartel. 2005. “MicroRNA-Directed Regulation of Arabidopsis AUXIN RESPONSE FACTOR17 Is Essential for Proper Development and Modulates Expression of Early Auxin Response Genes.” The Plant Cell 17 (5): 1360–75.

Meyers, Blake C., and Pamela J. Green, eds. 2010. Plant MicroRNAs: Methods and Protocols. Humana Press.

Myking, T., and O. M. Heide. 1995. “Dormancy Release and Chilling Requirement of Buds of Latitudinal Ecotypes of Betula Pendula and B. Pubescens.” Tree Physiology 15 (11): 697–704.

Niu, Qingfeng, Jianzhao Li, Danying Cai, Minjie Qian, Huimin Jia, Songling Bai, Sayed Hussain, Guoqin Liu, Yuanwen Teng, and Xiaoyan Zheng. 2016. “Dormancy-Associated MADS-Box Genes and microRNAs Jointly Control Dormancy Transition in Pear (Pyrus Pyrifolia White Pear Group) Flower Bud.” Journal of Experimental Botany 67 (1): 239–57.

Pan, Wenqiang, Jiahui Liang, Juanjuan Sui, Jingru Li, Chang Liu, Yin Xin, Yanmin Zhang, et al. 2021. “ABA and Bud Dormancy in Perennials: Current Knowledge and Future Perspective.” Genes 12 (10). https://doi.org/10.3390/genes12101635.

Peer, Wendy Ann, and Angus S. Murphy. 2007. “Flavonoids and Auxin Transport: Modulators or Regulators?” Trends in Plant Science 12 (12): 556–63.

Schwab, Rebecca, Javier F. Palatnik, Markus Riester, Carla Schommer, Markus Schmid, and Detlef Weigel. 2005. “Specific Effects of microRNAs on the Plant Transcriptome.” Developmental Cell 8 (4): 517–27.

Shannon, Paul, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. 2003. “Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks.” Genome Research 13 (11): 2498–2504.

Singh, Rajesh Kumar, Pal Miskolczi, Jay P. Maurya, and Rishikesh P. Bhalerao. 2019. “A Tree Ortholog of SHORT VEGETATIVE PHASE Floral Repressor Mediates Photoperiodic Control of Bud Dormancy.” Current Biology: CB 29 (1): 128–33.e2.

Smita, Shuchi, Michael Robben, Anup Deuja, Monica Accerbi, Pamela J. Green, Senthil Subramanian, and Anne Fennell. 2021. “Integrative Analysis of Gene Expression and miRNAs Reveal Biological Pathways Associated

with Bud Paradormancy and Endodormancy in Grapevine.” Plants 10 (4). https://doi.org/10.3390/plants10040669.

Spanudakis, Eleonora, and Stephen Jackson. 2014. “The Role of microRNAs in the Control of Flowering Time.” Journal of Experimental Botany 65 (2): 365–80.

Tran, Van Du T., Sebastien Tempel, Benjamin Zerath, Farida Zehraoui, and Fariza Tahi. 2015. “miRBoost: Boosting Support Vector Machines for microRNA Precursor Classification.” RNA 21 (5): 775–85.

Vergara, Ricardo, Ximena Noriega, and Francisco J. Pérez. 2021. “VvDAM-SVPs Genes Are Regulated by FLOWERING LOCUS T (VvFT) and Not by ABA/low Temperature-Induced VvCBFs Transcription Factors in Grapevine Buds.” Planta 253 (2): 31.

Wang, Jia-Wei, Benjamin Czech, and Detlef Weigel. 2009. “miR156-Regulated SPL Transcription Factors Define an Endogenous Flowering Pathway in Arabidopsis Thaliana.” Cell 138 (4): 738–49.

Xue, Chenghai, Fei Li, Tao He, Guo-Ping Liu, Yanda Li, and Xuegong Zhang. 2005. “Classification of Real and Pseudo microRNA Precursors Using Local Structure-Sequence Features and Support Vector Machine.” BMC Bioinformatics 6 (December): 310.

Yu, Guangchuang, Li-Gen Wang, Yanyan Han, and Qing-Yu He. 2012. “clusterProfiler: An R Package for Comparing Biological Themes among

Gene Clusters.” Omics: A Journal of Integrative Biology 16 (5): 284–87.

Zhang, Li, Yi Wang, Xinzhong Zhang, Min Zhang, Deguo Han, Changpeng Qiu, and Zhenhai Han. 2012. “Dynamics of Phytohormone and DNA Methylation Patterns Changes during Dormancy Induction in Strawberry (Fragaria × Ananassa Duch.).” Plant Cell Reports 31 (1): 155–65.

Zhang, Yuxi, Yanyan Wang, Xuekai Gao, Chunying Liu, and Shupeng Gai. 2018. “Identification and Characterization of microRNAs in Tree Peony during Chilling Induced Dormancy Release by High-Throughput Sequencing.” Scientific Reports 8 (1): 4537.

Table 4.1 Transcription factor motifs enriched in co-expression clusters that correlated with SD photoperiod induced endodormancy. Motifs originated from experimental predictions in *Arabidopsis thaliana*.

Transcription Factor	total_sites	adj_p-value
LMI1	117	1.50E-06
ATHB20	125	2.60E-05
AT5G60130	125	3.90E-04
ATHB40	133	4.30E-04
ATHB18	132	1.30E-03
ATHB53	128	1.50E-03
LMI1	128	1.50E-03
VRN1	93	1.50E-03
VRN1	124	1.70E-03
AT2G20110	127	2.20E-03
At1g64620	134	4.40E-03
At4g38000	134	4.40E-03
AT1G76870	133	5.10E-03
HAT5	130	5.30E-03
GTL1	134	5.30E-03
ATHB13	133	5.50E-03
HAT5	130	8.30E-03
ATHB13	134	8.80E-03

Table 4.2 *A. thaliana* transcription factor motifs enriched in F2-110 co-expression clusters relative to F2-040 clusters.

Gene ID	consensus	p-value	adj_p-value
PLT1	KGCACGVWTHYCGAGRHRD	0.00000718	0.00094
At5g08330 (TCP21)	WGTGGGMCCCACNW	0.000014	0.00387
HSFC1	WKCTTCTAGAAGCTTCT	0.000099	0.00573
At2g45680 (TCP9)	GTGGGHCCCAC	0.000031	0.00591
At1g72010 (TCP22)	WWGTGGGHCCCAC	0.000031	0.00695
HSFB4	AGAAGCTTCTAGAAG	0.000053	0.00698
AGL6	TTWCCAAAAAWGGAAAAWW	0.0000445	0.0106

Table 4.3 *V. vinifera* transcription factor motifs enriched in F2-110 co-expression clusters relative to F2-040 clusters.

Transcription Factor	p-value	adj_p-value
TCP22	3.87E-03	1.11E+00
HSFB4	6.98E-03	2.00E+00
AGL6	1.06E-02	3.04E+00
MYB62	1.54E-02	4.40E+00
SPL9	3.04E-02	8.71E+00

Figure 4.1 Prediction of regulatory small RNA sequences from small RNA-seq datasets.

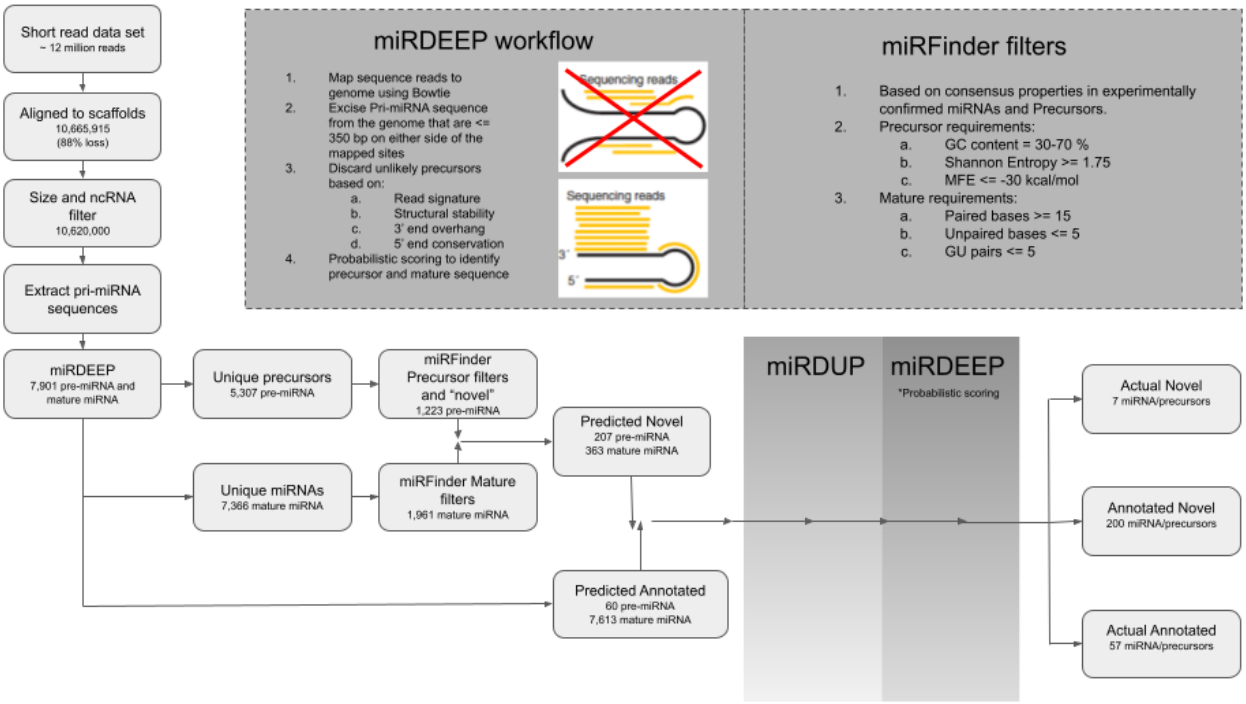


Figure 4.2 Validation of predicted miRNA sequences by machine learning methodologies. a) Percent of validated miRNAs out of 264 predicted sequences (positive/total, conserved, annotated, and novel sets) using five machine learning models. b) Venn diagram showing consensus of confirmed miRNAs between the 5 machine learning models.

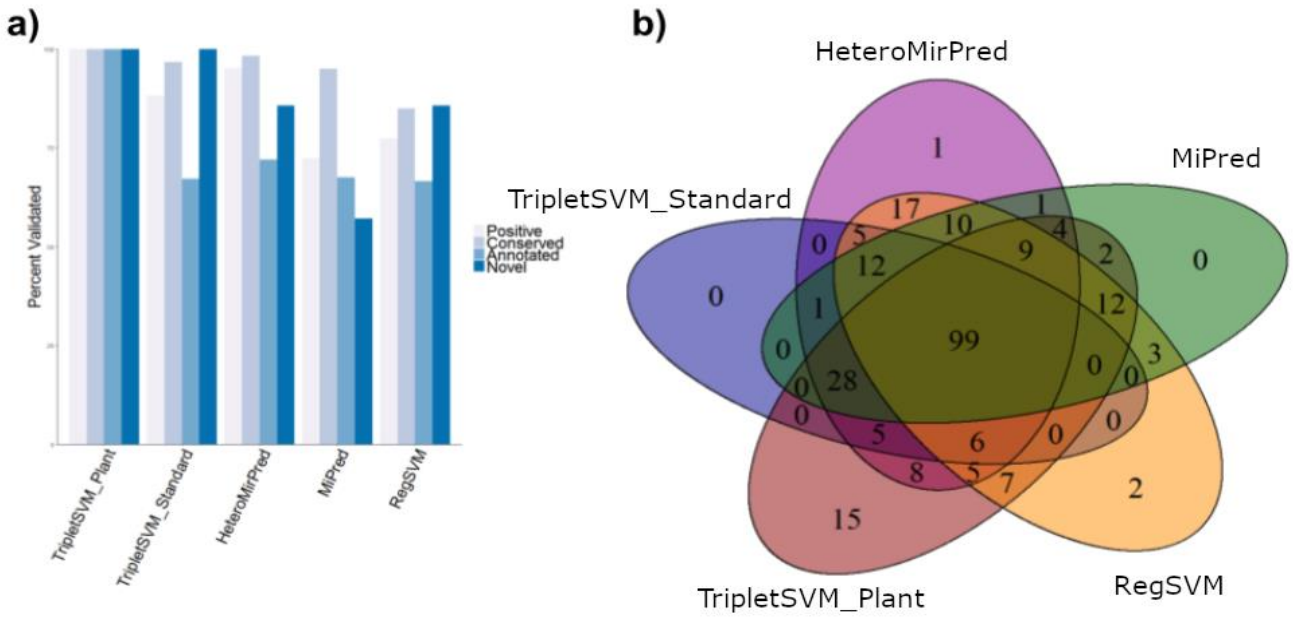


Figure 4.3 Predicted novel miRNA sequences that were confirmed to be regulatory by five independent machine learning programs. Full pre-miRNA are shown in folded secondary hairpin structures. Names represent internal naming of novel miRNA and are not officially named prior to submission.

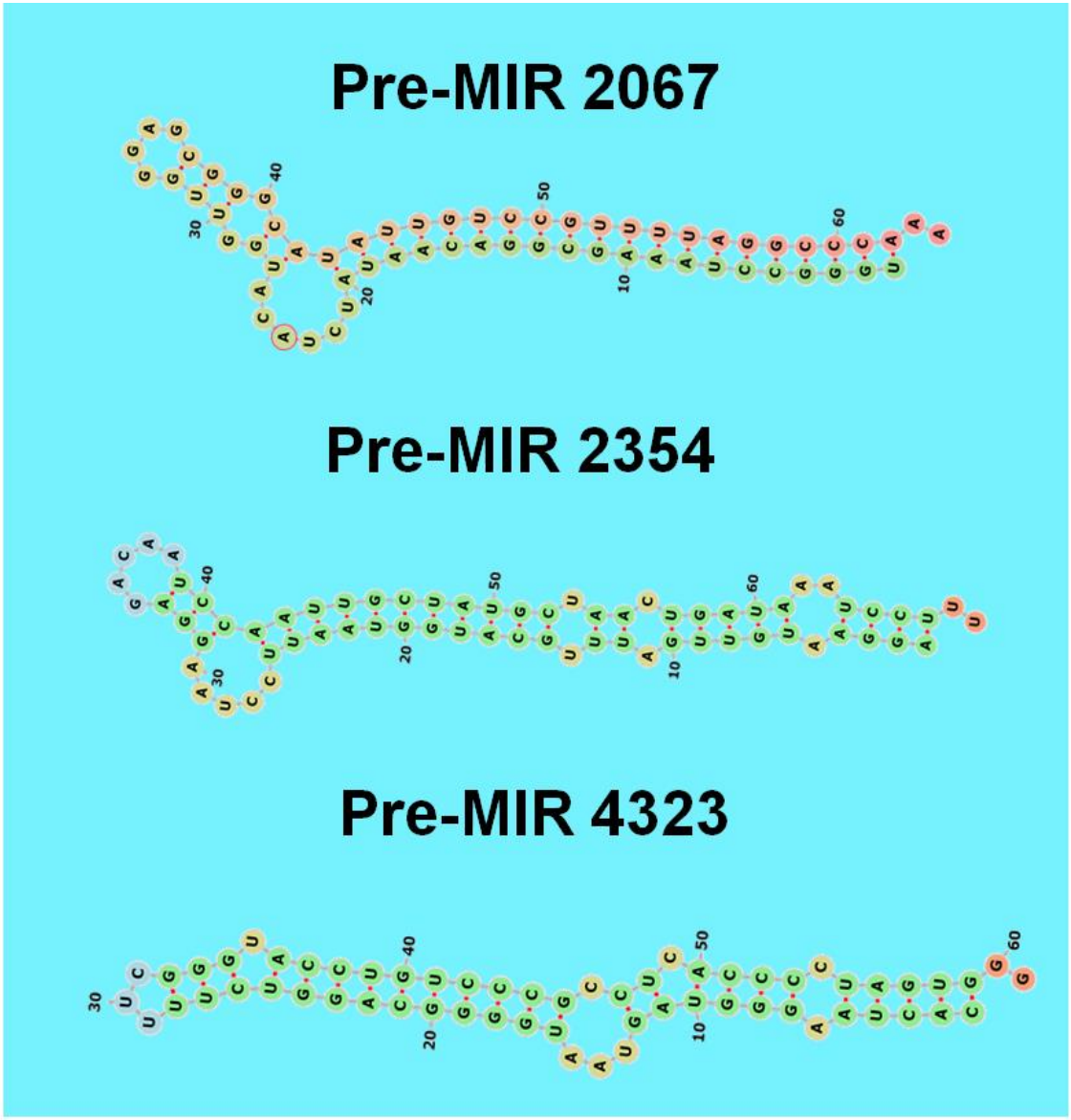


Figure 4.4 Differentially expressed genes in *V. riparia* buds aligned to *V. vinifera* in SD photoperiod treatment as relative to LD photoperiod treatment. Genes were filtered for LFC >= 1 and an alpha <= 0.01.

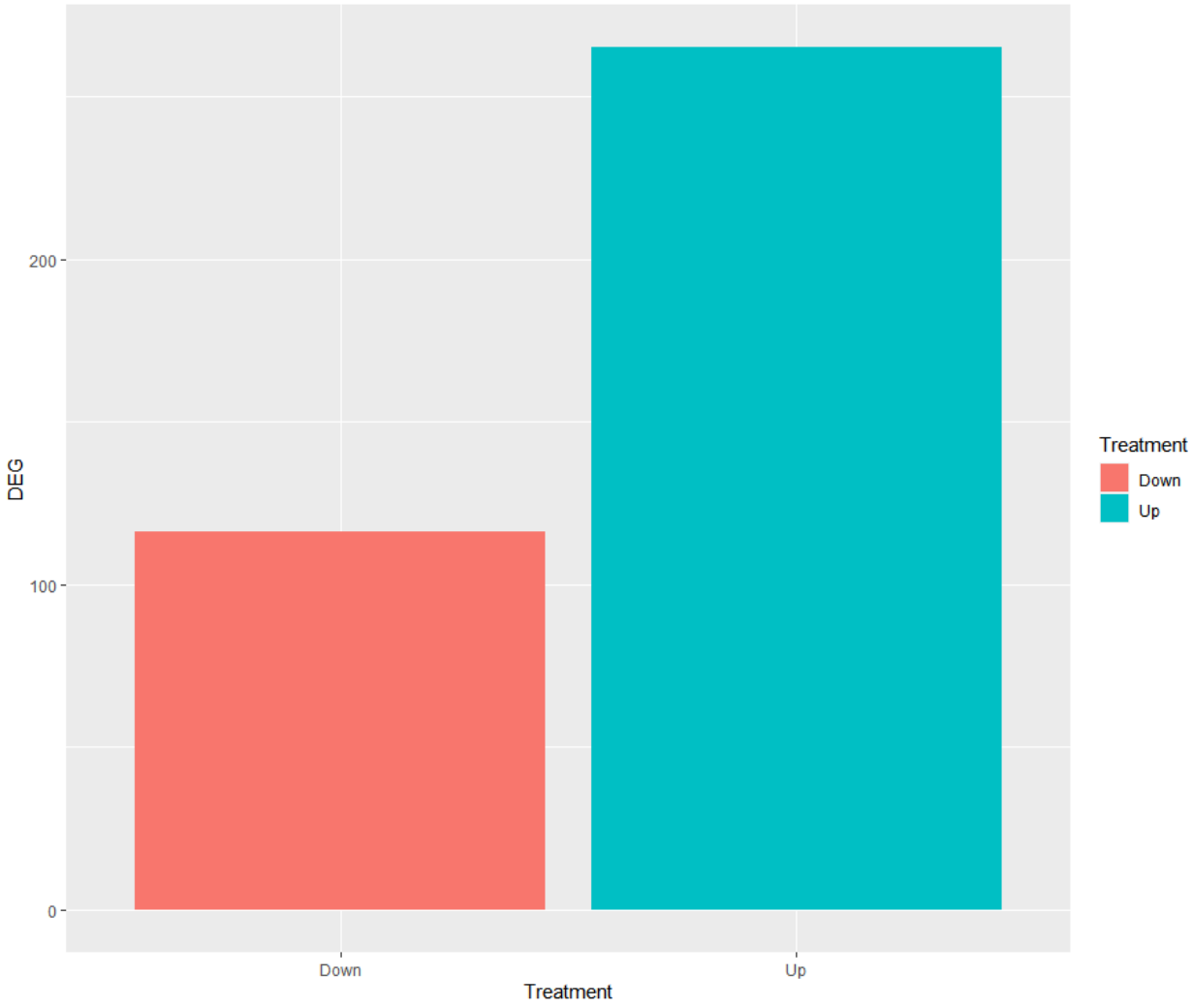


Figure 4.5 Three transcription factors regulate SD photoperiod induced differential gene expression in V. riparia bud tissue.Transcription factor regulation was determined using GSEA of transcription factor regulatory gene sets. Transcription factors represented by name in V2 assembly annotation. Differential expression of genes in SD treated *V. riparia* buds in comparison to LD treated buds.

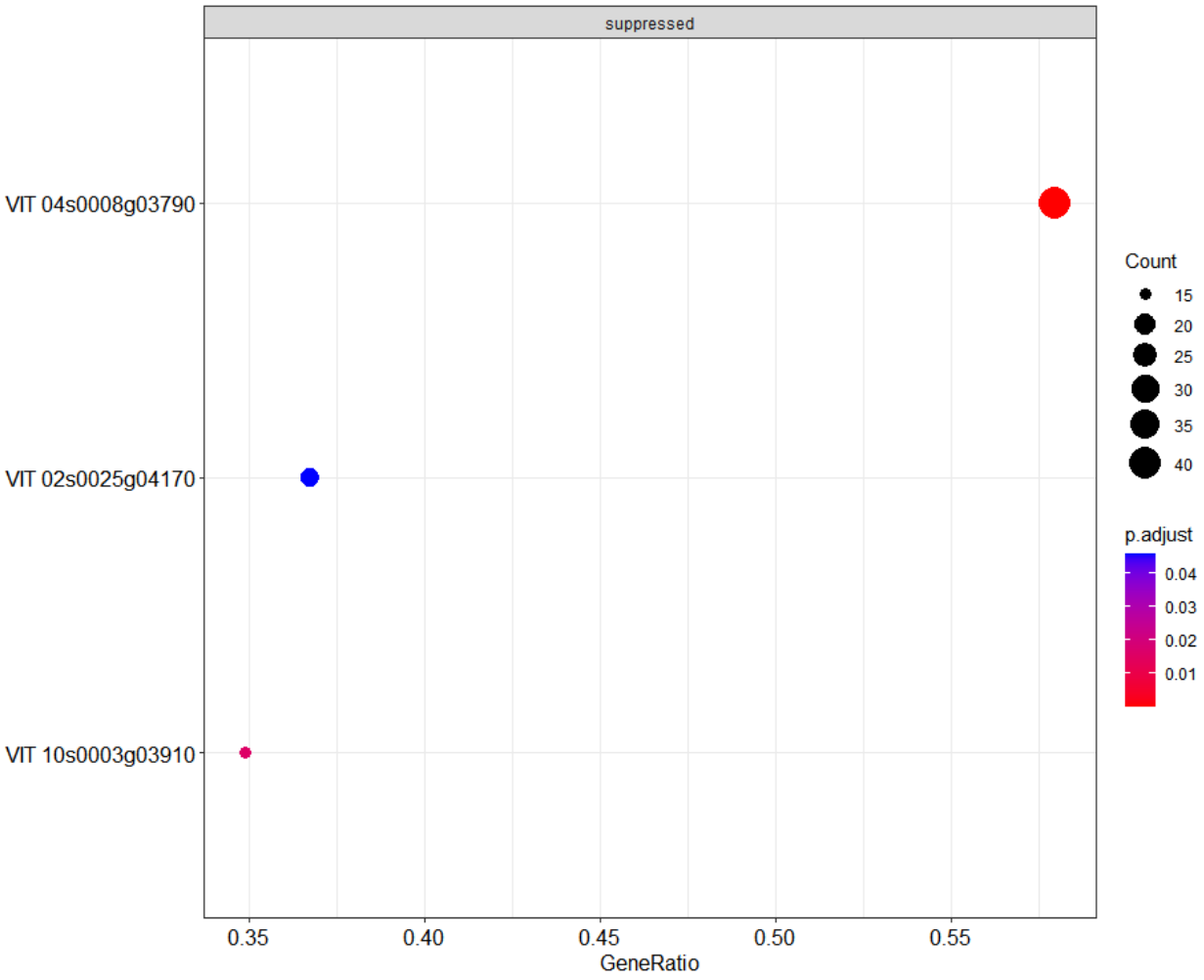


Figure 4.6 MYB3R1 interaction network showing the relationship between regulators of SD induced endodormancy. Transcriptional targets, protein interactors, and miRNA nodes are represented by squares, circles, and triangles, respectively. MYB3R1-target interactions are denoted by yellow edges, MYB3R1-protein interactions by green edges, protein-protein interactions by red edges, and miRNA-target by blue edges. The node fill color (blue (negative) to red (positive)) indicates the log fold change expression values for differentially expressed genes and nodes with green fill color were not differentially expressed in SD relative to LD buds.

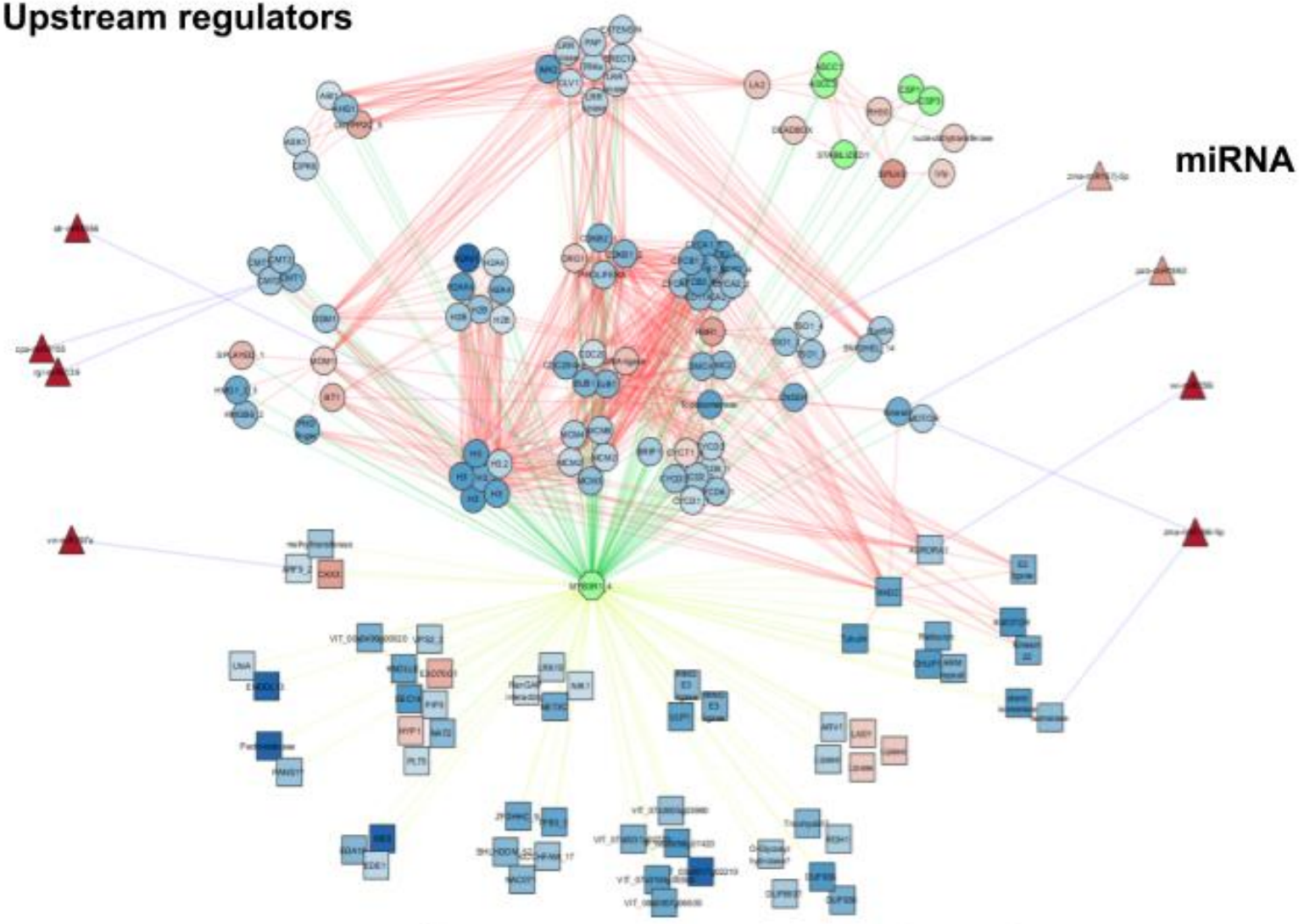
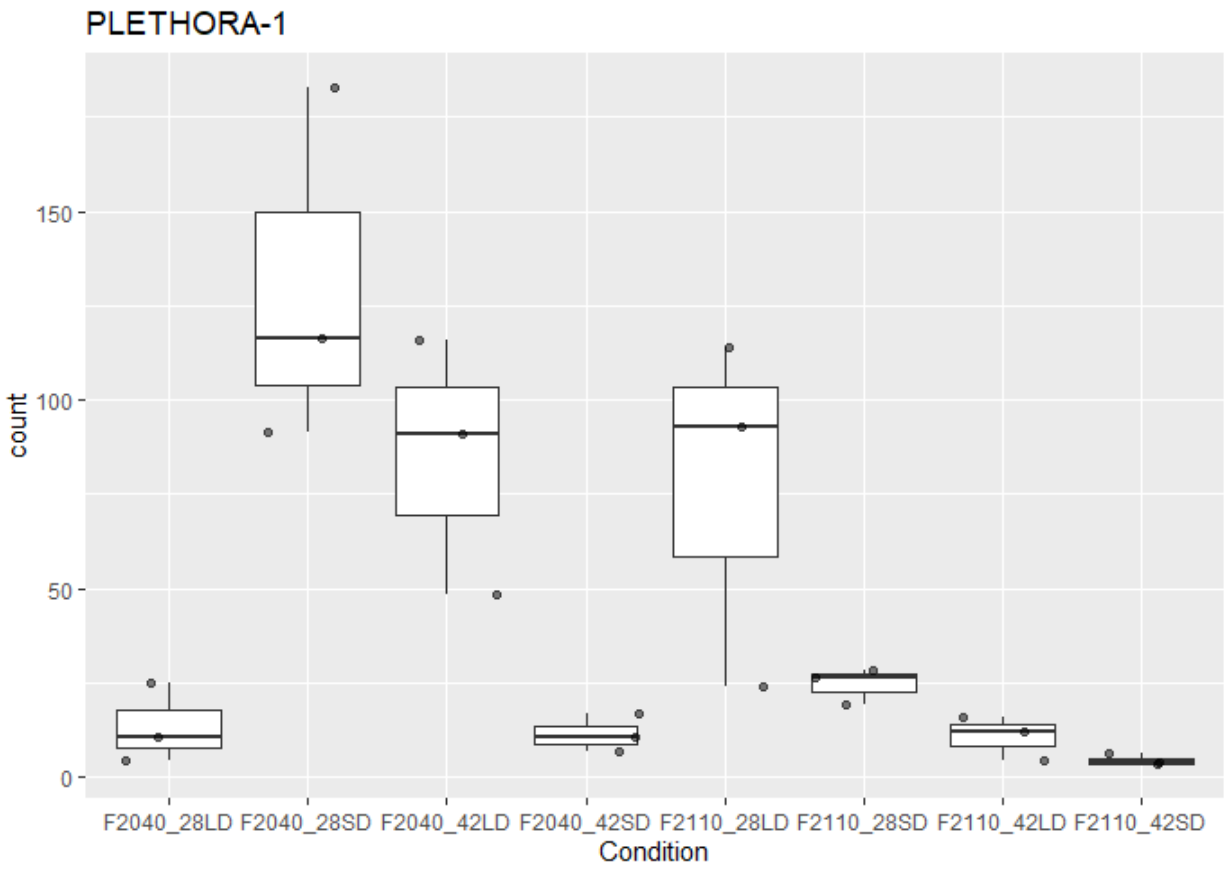


Figure 4.7 PLETHORA (PLT1) gene count data for VRS-F2 siblings under long and short photoperiod conditions. Counts are normalized by Rlog transformation and plotted by exact counts in each replicate as points and distribution represented by box plot.



5 Future steps

Several breeding programs already exist for increasing winter survival in grapevine crops. However, without experiments to elucidate the mechanisms that control endodormancy transition, such endeavors will be difficult. Through greenhouse experiments, we have confirmed that F2 genotypes from a previously generated mapping population show varied phenotypes for endodormancy response. This F2 population may be useful for studying and creating markers that can identify increased cold tolerance and endodormancy response.

In this study we have suggested several genes and gene pathways such as ERF signaling, auxin signaling, and phenylpropanoid biosynthesis that could be involved in both endodormancy and enhanced photoperiod response in V. riparia. Many of the genes we recognized have roots in auxin signaling. PLT1 being upregulated by Auxin Response Factors (ARFs) means that changes in its expression may correspond to changes in auxin levels or bud sensitivity to auxin. We noticed that expression of PLT1 peaks at 28 days in F2-040 and tapers to normal levels by 42 days. Likely, the bud becomes more susceptible to auxin signaling at the start of endodormancy induction acting like a switch to turn off the stem-cellness of the SAM. Phenylpropanoids which we found upregulated in endodormant samples may influence this by increasing auxin transport.

In order to confirm this auxin response to photoperiod signals, we would need to measure relative auxin concentrations in different tissues at each stage of endodormancy induction. We could also generate auxin insensitive mutants to

determine if there is a difference in grapevine endodormancy induction. The concentrations of certain phenylpropanoids are also important to look at, and we need to measure the concentration of phenylpropanoids that have already been linked to endodormancy like kaempferol and quercetin.

As many of these possible regulators of the endodormant state are transcription factors, ChIP-seq experiments would tell us the exact effect that they have in phase transition. PLT1, MYB3R1, AGL6, and SPL9 are all possible targets of this research. Although working in a non-model organism, it is hard to find antibodies specific to the protein of interest for immunoprecipitation experiments. miRNAs are easier to confirm, the general expression of miR166 and miR167 can be measured by qPCR. Thanks to our validation of predicted miRNA, we have an accurate sequence to construct primers for this purpose.

Unfortunately, many of these transcription factors and interactors have little experimental evidence of protein-protein interactions. A useful experiment that would provide us insight as to molecular interactions at the time of endodormancy induction are experiments like Yeast 2-Hybrid, Affinity Purification, or other modern experimental measurements of protein interactions. We hope that by conducting further research we can give meaningful results to growers and improve winter survival for grapevines.