GENETIC DISSECTION OF DISEASE RESISTANCE AND PEST RELATED TRAITS IN HYBRID GRAPEVINE FAMILIES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Paola Leonor Barba Burgos

August 2015

GENETIC DISSECTION OF DISEASE RESISTANCE AND PEST RELATED TRAITS IN HYBRID GRAPEVINE FAMILIES

Paola Leonor Barba Burgos, Ph. D.

Cornell University 2015

The genus *Vitis* is highly diverse and includes more than 60 species. In scion breeding, hybrid grapevines are created to combine fruit quality of *V. vinifera* cultivars with a wide range of adaptive traits from their wild relatives. Among these, resistance to disease is the most representative example. In this work, the goals were to dissect genetic regions that control disease resistance and pest-related traits in *V. rupestris* B38, *V. cinerea* B9, *V. vinifera* 'Chardonnay', Illinois 547-1 and the complex hybrid 'Horizon'. For this, we studied the segregation of powdery mildew (*Erysiphe necator*) and phomopsis (*Diaporthe ampelina*) disease intensities; predatory mite (*Phytoseiid*) abundance; and trichome trait densities, such as domatia, leaf hairs and bristles, in bi-parental F_1 families. We used genotyping-by-sequencing (GBS) to rapidly identify single nucleotide polymorphisms across the whole genome. In order to make full use of this technology in heterozygous crosses, we developed novel methods to generate dense genetic maps and validated them by correctly localizing the sex locus. Multiple QTL mapping was used to identify genetic regions controlling resistance traits. For powdery mildew, we identified a moderate effect locus controlling susceptibility on chromosome 9 of the cultivar, 'Chardonnay'. Two qualitative loci, *Rda*2 and *Rda*1, were found to be responsible for phomopsis cane and berry rot resistance on chromosome 7 of

‘Horizon’ and chromosome 15 of *V. cinerea* B9, respectively. Further, transcriptome analyses in *V. cinerea* B9 characterized a resistance reprograming and expression QTL mapping (eQTL) narrowed down *Rda*1 candidate genes to three NBS-LRR genes. Genetic determinism of leaf trichomes was more complex, with several major and minor QTL controlling trait expression on leaf blades and veins. Among these, a QTL located on chromosome 1 of Illinois 547-1 had major effect on all trichome traits and was also correlated with the abundance of predatory mites. While providing knowledge of the mechanisms of powdery mildew resistance, phomopsis cane resistance and abundance of predatory mites, this work also delivered tools for the creation of dense genetic maps in heterozygous crosses and for marker assisted selection in grapevine breeding.

BIOGRAPHICAL SKETCH

Paola Barba was born to Davioleta del Carmen Burgos and Vicente Barba in Santiago de Chile. Although, Paola lived most of her entire life in a city, since her childhood she often traveled south into natural areas. The beauty, peace and diversity of life awoke a sense of passion for nature that still remains alive and continues to inspire her.

Paola started her studies in engineering at the Faculty of Physical Science and Mathematics (FCFM - Facultad de Ciencias Físicas y Matemáticas) of the University of Chile in 1997. After three years of the common engineer lectures she moved towards a specialization in biotechnology. There, her interest focused upon molecular biology guided by Professor Oriana Salazar, who gave her the opportunity to accomplish her first research project, and later accepted her as thesis advisor. During the 6 years of the Engineering program, Paola was twice recognized for ranking in the top 5% of the engineering school. She also enjoyed being an assistant teacher of physics and courses in the biotechnology program. In 2003 she obtained her degree of Civil Engineer in Biotechnology with the highest score.

Plant science came into her life in 2007 when she moved to work at the Institute of Agricultural Investigations (INIA) under the leadership of Ph.D. Humberto Prieto. There, she made substantial contributions to projects related to stone fruit transformation for plum pox virus resistance, and grape genomics in response to *Botrytis cinerea*. During this period she realized the need to complement her career with knowledge in plant sciences.

Funded by the Becas Chile, in 2009 Paola started graduate studies in Plant Breeding at Cornell University, where she obtained a M.S. degree in 2012 for her work on resistance to powdery mildew in hybrid *Vitis* crosses, and adapting genotyping-by-sequencing technology for this heterozygous genus. In 2011 she was awarded a scholarship by Becas Chile to pursue a Ph.D. degree at Cornell.

“Son los genes, son los genes” decía mi abuelo, cada vez que una de sus nietas hacía algo que lo hacía sentir orgulloso

ACKNOWLEDGMENTS

To my colleagues and advisors: The great grape bunch: Bruce Reisch, Pat Wallace, Steve Luce and Elizabeth Takacs. The Buckler lab: Ed Buckler, Sara Miller, Jeff, Glaubitz, Fei Lu, Karl Kremling, Alberto Romero, Kelly Swarts, Cinta Romay, among others. VitisGen: Qi Sun, Jason Londo, Shanshan Yang and Jonathan Fresnedo. The Cadle-Davidson team: Lance Cadle-Davidson, Jacquelyn Lillis, Jub Siraprapa Brooks and Erin Galarneau. Collaborators at the Biotechnology Resource Center: Katie Hyma, Sharon Mitchell, Peter Schweitzer and Linda Cote. To plant pathology co-authors and colleagues: Judith Burr, and Wayne Wilcox and to the domatians Rebecca Loughner, Karen Wentworth, Gregory Loeb and Jan Nyrop. Thanks for all the support, knowledge and efforts that allowed this work to reach this point.

To my friends, family and people who supported my soul and mind when I was not working: mom, dad and my sister Carolina, Diego Muñoz, Ricardo A. Daziano, Alberto Romero, Vanesa Soto, Julio I Sanchez, Gracia Montilla-Bascon, Eliel May, Maria de los Angeles Miccono, Macarena Garrido and Alvaro Sequeida. Special thanks to Eduardo Carrillo for his support, advice, patience and love.

USA agencies that funded this work: CONICYT through Becas Chile; The USDA-NIFA VitisGen project (Award No. 2011-51181-30635); The USDA Viticulture Consortium – East; The New York Wine & Grape Foundation and the Lake Erie Regional Grape Processor's Fund; as well as the Michael Nolan Endowment Fund.

TABLE OF CONTENTS

CHAPTER 1

GRAPEVINE POWDERY MILDEW RESISTANCE AND SUSCEPTIBILITY LOCI IDENTIFIED ON A HIGH-RESOLUTION SNP MAP[1]

Abstract

Improved efficacy and durability of powdery mildew resistance can be enhanced via knowledge of the genetics of resistance and susceptibility coupled with the development of high-resolution maps to facilitate the stacking of multiple resistance genes and other desirable traits. We studied the inheritance of powdery mildew (*Erysiphe necator*) resistance and susceptibility of wild *Vitis rupestris* B38 and cultivated *V. vinifera* 'Chardonnay', finding evidence for quantitative variation. Molecular markers were identified using genotyping-by-sequencing, resulting in 16,833 single nucleotide polymorphisms (SNPs) based on alignment to the *V. vinifera* PN40024 reference genome sequence. With an average density of 36 SNPs/Mbp and uniform coverage of the genome, this 17K set was used to identify 11 SNPs on chromosome 7 associated with a resistance locus from *V. rupestris* B38 and ten SNPs on chromosome 9 associated with a locus for susceptibility from

[1] Barba, P., Cadle-Davidson, L., Harriman, J., Glaubitz, J., Brooks, S., Hyma, K., Reisch, B. (2014). "Grapevine powdery mildew resistance and susceptibility loci identified on a high-resolution SNP map." Theor Appl Genet **127**: 73-84. Contributions: Paola Barba carried out the project and analysis described; Lance Cadle-Davidson and Bruce Reisch provided advice and direction; Bruce Reisch provided genetic populations and access to field and lab resources; James Harriman and Jeffrey C. Glaubitz processed sequence data and called SNPs; Siraprapa Brooks provided advice and adjusted the protocol for library construction; and Katie Hyma advised on the use of LD analysis and linkage map construction.

‘Chardonnay’ using single marker association and linkage disequilibrium analysis. Linkage maps for *V. rupestris* B38 (1,146 SNPs) and ‘Chardonnay’ (1,215 SNPs) were constructed and used to corroborate the ‘Chardonnay’ locus named *Sen1* (*Susceptibility to* Erysiphe necator *1*), providing the first insight into the genetics of susceptibility to powdery mildew from *V. vinifera.* The identification of markers associated with a susceptibility locus in a *V. vinifera* background can be used for negative selection among breeding progenies. This work improves our understanding of the nature of powdery mildew resistance in *V. rupestris* B38 and ‘Chardonnay’, while applying next-generation sequencing tools to advance grapevine genomics and breeding.

Introduction

Powdery mildew (PM) resistance differs among and within *Vitis* species (Cadle-Davidson et al., 2011). While most European *V. vinifera* cultivars are susceptible to the PM fungus *Erysiphe necator*, North American species such *as V. rupestris, V. riparia, V. aestivalis, V. cinerea* and *V. rotundifolia* are generally considered to be resistant (Alleweldt et al. 1991; Pearson 1988). The nature of PM resistance in *V. rupestris* or susceptibility in ‘Chardonnay’ are not completely understood, but early studies of the inheritance of PM resistance in *V. rupestris* suggested that the trait was controlled by a polygenic system (Boubals 1961), while a recent study of the mechanism of resistance classified one *V. rupestris* genotype as having partial resistance with a low incidence of programmed cell death (PCD) (Feechan et al. 2010).

Dominant single loci and quantitative trait loci (QTL) may control different mechanisms of plant-pathogen interactions. In grapevine PM, dominant single locus resistance is typically thought to be related to a gene-for-gene interaction (Jones and Dangl 2006; Feechan et al. 2010; Coleman et al. 2009). In most cases, this type of resistance confers complete protection against specific races of the pathogen but also induces strong selective pressure over pathogen populations, which could potentially overcome resistance, as has been observed in grapevines (Cadle-Davidson et al. 2011; Peressotti et al. 2010). However, quantitative resistance is described as less likely to be overcome as it is due to the cumulative effect of several loci that might play a role at different stages of the plant immune response (Poland et al. 2009). For a vineyard that is expected to be productive for 15 to 20 years, durable resistance is desired.

Wild species harbor an assortment of undesirable flavors and viticultural traits, and thus introgression of PM resistance alleles is usually accompanied by off-flavors and traits that are not desired by consumers and growers. Moreover, the genetic nature of the resistance mechanism adds complexity, as quantitative traits may require introgression of several minor loci in order to reach the desired level of resistance. Marker assisted selection (MAS) helps to overcome these constraints. Molecular markers can be used to select resistant genotypes, avoid susceptibility alleles, combine traits, reduce linkage drag and stack several loci, while markers distributed along the genome may help to recover the cultivated background (Dalbó et al. 2001; Eibach et al. 2007; Mahanil et al. 2011; Di Gaspero and Cattonaro 2009). However, as more traits

are tracked and combined, greater marker resolution and accuracy are needed to identify desirable recombination events.

The relevance of molecular markers to grapevine genetics has driven the development of a common set of markers and genetic maps (Riaz et al. 2004; Adam-Blondon et al. 2004; Doligez et al. 2006). Nowadays, the International Grape Genome Program refers to an integrated map containing more than 400 SSR markers (http://www.vitaceae.org/index.php/Maps_and_Markers) in addition to a dense genetic linkage map anchored to the ‘Pinot noir’ genome with 1006 markers (Troggio et al. 2007). Physical maps have also been developed, such as the *V. vinifera* grapevine reference genome for a nearly homozygous selection, PN40024 (Jaillon et al. 2007), 'Cabernet Sauvignon' (Moroldo et al. 2008) and 'Pinot noir' (Velasco et al. 2007). Next-Generation Sequencing (NGS) has been employed recently for the construction of a genetic map with 1,643 SNPs derived from a cross of Z180 (*V. monticola* × *V. riparia*) and Beihong (*V. vinifera* × *V. amurensis*) (Wang et al. 2012) and to develop a SNP chip with an array of nearly 9,000 SNPs based on the sequence of 10 cultivated *V. vinifera* cultivars and 7 wild species (Myles et al. 2010; Myles et al. 2011). This genotyping microarray has been used successfully to identify *V. vinifera* markers flanking the introgressed PM resistance locus *Ren4* (Mahanil et al. 2011).

Whole genome sequencing and NGS have boosted genomic research in several plant species (Deulvot et al. 2010; Poland et al. 2012; Xie et al. 2010; Morrell et al. 2012). Single nucleotide polymorphism (SNP) markers can be identified from short reads generated by NGS, either by aligning to a reference genome or by *de novo* assembly (Nielsen et al. 2011). The adoption of

techniques, such as Reduced Representation Libraries (RRLs) to lower genome complexity (Van Tassell et al. 2008; Wiedmann et al. 2008; Barbazuk et al. 2005), and bar-coded adapters to allow pooling hundreds of samples in a single sequencing lane has significantly reduced the cost per marker and per sample (Elshire et al. 2011).

Nowadays, Genotyping-by-Sequencing (GBS) provides a simple and robust procedure for simultaneous SNP discovery and genotyping through pooled barcoded RRLs, Illumina sequencing and SNP calling based on alignment of short reads. As a result, thousands of markers with low coverage are obtained (Elshire et al. 2011), which should be sufficient to infer linkage in bi-parental populations and for QTL mapping (Davey et al. 2011). Due to its speed, low cost, and reduced ascertainment bias, GBS is a good strategy for simultaneous discovery and assay of SNPs suitable for rapid development of dense maps in segregating populations.

In this study, we present the discovery of a dense set of SNPs using the GBS procedure with an F_1 grapevine population. SNPs were positioned in the 12X *V. vinifera* PN40024 reference genome (Jaillon et al. 2007) and tested for association with PM severity. Linkage maps were constructed using a subset of SNPs. QTL interval mapping confirmed the discovery of a susceptibility QTL from ‘Chardonnay’.

Materials and Methods

Plant material

Seeds from the cross of *V. rupestris* B38 (resistant) with *V. vinifera* 'Chardonnay' (susceptible) were obtained in 2008, then stratified, germinated in a greenhouse, and planted to a field nursery (Geneva, NY) 0.46 m apart within rows and 1.52 m between rows in 2009. At the end of the growing season, vines were pruned and stored at 4 ºC in the dark over the winter. All 85 vines were planted 1.2 m apart within a single vineyard row in Geneva. A control block was placed at the head of the row with: *V. vinifera* 'Chardonnay', V. hybrid 'Chancellor' (Seibel 5163 x Seibel 880), *V. rupestris* B38, *V.* hybrid 'Horizon' ('Seyval' x 'Schuyler') and the PM resistant selection, NY88.0514.04. A susceptible control ('Chardonnay') was placed after every 15 seedling vines. Downy mildew was controlled in 2010 and 2011 using the fungicide Captan 80WPG, which is not registered for control of PM, and does not provide commercially acceptable PM control in the field.

Quantification of powdery mildew severity

Powdery mildew infection was evaluated on parents and progeny by visual evaluation of field-grown vines over 3 years. Disease was allowed to progress naturally, and foliage of each vine was evaluated as follows: in 2009 a visual scale from 0 to 3 was used near the end of the growing season (0: absent, 1: less than 5 small spots, 2: five to twenty spots, growing, 3: widespread, dense sporulation); In 2010, on August 9, August 20, August 30, September 7, and September 20 using the IPGRI scale established by the Organisation

Internationale de la Vigne et du Vin (IPGRI et al. 1997); and in 2011 on July 7 and August 18, again using the IPGRI scale.

DNA extraction, library preparation, and sequencing

DNA was extracted from two unexpanded leaves (less than 1 cm^2) from each parent and progeny vine using either the DNeasy® 96 Plant Kit (Qiagen) or the DNeasy® Plant Mini Kit (Qiagen) and quantified using Quant-iT™ PicoGreen ® dsDNA Kit (Invitrogen). Whole genome amplification was performed using 10 ng of DNA and the Illustra™ GenomiPhi™ V2 DNA Amplification Kit (GE Healthcare). Amplified DNA (1.0 µg) was plated and dried using a vacuum centrifuge.

Dried DNA was resuspended and digested at 75ºC for 2 h using a 10 µl mix containing 4 units of *ApeKI* restriction endonuclease (New England Biolabs, Ipswich, MA) and 1 µl of 10X NEBuffer 3, then cooled on ice. Forty-eight unique barcode adapters were used to track individual DNA samples (Elshire et al. 2011). Dried barcode adapters were resuspended by pipetting 40 µl of a ligation mix containing 4 units of T4 DNA Ligase (Promega) in 2X rapid ligation buffer (Promega). Resuspended barcode adapters were mixed with cooled digested DNA. Ligation was performed at room temperature for 60 min, followed by incubation at 65ºC for 30 min to inactivate the enzyme and then cooling on ice until to the next step. Ligation products were purified using 90 µl of Agencourt AMPure (Beckman Coulter) beads per the manufacturer's instructions and eluted in 35 µl of EB Buffer (Qiagen). PCR was performed by adding the following to 10 µl of the eluted ligation product: 22.5 µl of water, 2 µl of dNTPs (10 mM), 5 µl of Primer mix (5' AATGATACGGCGACCACCGAG

ATCACTCTTTCCCTACACGACGCTCTTCCGATCT 3’ and 5’ AAGCAGAAGA

CGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC
T 3’, 5 µM each), 10 µl of 5X buffer and 0.5 µl of Phusion DNA polymerase (Finnzymes). Amplifications were performed by initial denaturation at 98ºC for 30 s; 18 cycles of 10 s at 98ºC, 30 s at 65ºC and 30 s at 72ºC; and a final step of 72ºC for 5 min. PCR products were purified using Agencourt AMPure beads and EB Buffer as described above. The result was a sequencing library for each DNA sample, which were individually quantified using Quant-iT™ PicoGreen ® dsDNA Kit (Invitrogen). Libraries with concentrations lower than 10 ng/µl were repeated either from re-extracted DNA or by repeating the PCR step in triplicate and pooling 3 PCR products during the last AMPure elution step. For twenty arbitrarily selected libraries across the range of concentrations, size distribution and proportion of adapter dimers were quality checked using an Experion™ Automated Electrophoresis System (Bio-Rad).

The barcoded libraries were normalized and two pools of up to 48 samples were prepared with a final concentration of 0.5 ng/µl each. Each pool was sequenced on a single flow cell lane using Illumina instruments at the Cornell University Life Sciences Core Laboratories Center: the first pool on an Illumina Genome Analyzer III (GA3) and the second on an Illumina HiSeq 2000. In both cases, single end sequencing was performed with a read length of 100 bp.

SNP calling, localization and distribution on the reference genome

The raw sequence data were processed into SNP genotype files in HapMap format using the TASSEL 3.0 GBS pipeline (Glaubitz et al. 2012). First, all of the 100 bp reads that contained a known barcode along with the expected *Ape*KI cut site remnant were converted into 64 base sequence tags (where, barring sequencing errors, each tag represented an allele) by trimming off the barcode along with excess 3' nucleotides. Reads containing N's within the first 64 bases after the barcode were rejected. Reads that contained either the beginning of the common (non-barcoded) adapter (from short restriction fragments) or a full *Ape*KI site (from incomplete digest or chimera formation) within the first 64 bases after the barcode were truncated accordingly. A master tag list was constructed comprising all tags that were observed at least 10 times across all of the samples. These tags were then aligned to the 12X *V. vinifera* PN40024 reference genome (Jaillon et al. 2007) using a Burrows – Wheeler Aligner (BWA) (Li and Durbin 2009) with default parameters. Tags located at the same, unique position on the grapevine reference genome and containing no more than two SNPs relative to the reference were then aligned against each other, which, along with information from the barcodes indicating which samples each tag was observed in, allowed SNP genotypes to be called. SNPs were output only if 40% or more of the samples were covered by at least one of the tags at the corresponding locus and if the minor allele frequency (MAF) was at least 0.1. Indel polymorphisms were ignored as were any additional rare alleles beyond the major and minor alleles. Subsequently, additional filtering was applied so that only SNPs with a MAF between 0.15

and 0.35 were retained, according to the expected MAF of 0.25 for markers in an AB:AA configuration in the parents.

Marker density and single marker association analyses

More stringent filters were used for the marker density and trait association analyses. Only sites with less than 20% missing data were retained. Out of the remaining genotypes, there were 4% for which only the minor allele was sequenced; these were imputed as heterozygotes. The SNPs were then further filtered to retain only SNPs with MAF 0.25 $\pm$ 0.05. These filters and minimal imputation resulted in a data set consisting of 16,833 SNPs (17K SNP set). Pearson's correlation (r) was used to determine correspondence between the number of SNPs per chromosome and its physical size. A bin analysis was used to estimate SNP density across the *V. vinifera* PN40024 reference genome, by counting the number of SNPs in contiguous windows of 500 kb.

To find associations of single SNPs with PM severity, a general linear model was used (TASSEL 3.0, Bradbury et al. 2007). The false discovery rate was controlled according to the Benjamini-Hochberg procedure implemented in the multtest package in R (Pollard et al.2004). For site by all linkage disequilibrium (LD) analysis, markers were assigned to either 'Chardonnay' (9,187 SNPs) or *V. rupestris* B38 (6,213 SNPs) according to the sequence information of the parents for a given site. Markers with missing data in both parents were not considered in this analysis (1,433 SNPs). Linkage disequilibrium (D') between individual markers and parental SNPs was determined using the "site by all" option in TASSEL 4.1.18

Linkage maps

To construct linkage maps with JoinMap 4.1 (Van Ooijen 2006), the 17K SNP set was filtered down to 2,543 SNPs, a number below the 3,000 marker threshold of the current version of the software. Prior to the filtering process, four individuals (43, 68, 69 and 71) were discarded due to their high proportion of missing data. Filtering was based on parental information, physical distance, physical location, LD and missing data. First we selected markers based on parental information: markers that were homozygous, had missing data or were heterozygous in both parents were discarded, as well as markers located on non-aligned chromosomes, obtaining a set of 5,592 and 7,197 markers for *V. rupestris* B38 and 'Chardonnay'. We further discard markers located within 64 bp of each other obtaining two parental SNP sets of 3,502 and 4,631 markers for *V. rupestris* B38 and 'Chardonnay'. These parental SNP sets were used for whole genome LD analysis in TASSEL 4.1.18; markers that were not in LD with their physical chromosome (378 and 1,003) were discarded. Finally we filtered out markers that were positively correlated within 500 bp and sites with more than 10% missing data. Significant hits from the single marker association test were included, as well as markers located in the random portion of chromosome 7. We obtained a set of 1,222 and 1,321 markers for the *V. rupestris* B38 and 'Chardonnay' maps, respectively.

Linkage groups (LG) were determined by JoinMap 4.1 using a minimum LOD score of 6.0, and numbered according the physical chromosome numbers. While 'Chardonnay' SNPs resulted in 19 LG, *V. rupestris* B38 SNPs were clustered in 20 LG. Maps and the order of markers were generated using the regression mapping algorithm with the Haldane function and default

parameters. Two LG with SNPs from both ends of chromosome 1 were joined after each map was ordered and oriented.

Maps resulting from JoinMap were further analyzed by R/QTL software (Broman et al. 2013) using a 4 way cross format with missing information. First, we analyzed the number of crossovers per individuals as a function of missing data in order to identify possible outliers (Appendix 1-1). Individuals 2, 36, 45, 50, 55, 62, 66, 70, 74, and 76 were discarded due to higher proportion of crossing over, which may indicate pollen contamination or sample mix up during DNA manipulation stages. Genotyping errors were identified by visual inspection of plot.geno maps, according to Ward et al. 2013, and replaced with missing data. Distances were recalculated by est.map using Kosambi function. Problematic markers were determined with the droponemarker command; a few markers with higher LOD scores, that increased the map size and whose genetic and physical positions were in conflict were removed. Finally, distances were recalculated using the est.map function as described above.

QTL analysis

QTL mapping was performed with the one-dimension scan function, *scanone*, of R/QTL software using a normal model, Haley-Knott regression method and default parameters. Multipoint genotype probabilities were calculated beforehand using *calc.genoprob* with step = 1 and default parameters. LOD significance scores were determined by permutation tests (10,000.) Position was refined using the *refineqtl* function, and the presence of supplementary QTLs was corroborated with the *addqtl* command.

Results

Field evaluation of PM severity

Disease progression varied from year to year. In 2009, the population mean score was 2.1 (of 3 max) on October 8. In 2010, disease progression was evaluated 5 times with the IPGRI scale (1 to 9); with population mean scores of 4.7 on August 9, 6.7 on August 20, 7.8 on August 30 and September 7, and 7.6 on September 20. In 2011 the disease progressed earlier, with population disease severity mean scores of 3.7 on July 12 and 8.0 on August 18 (Figure 1-1).

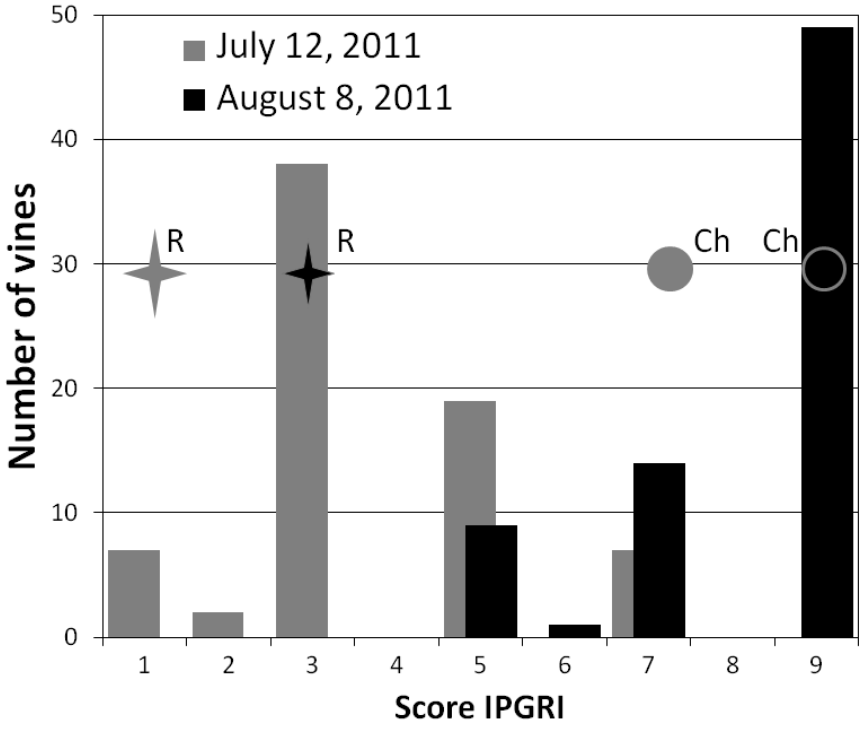


Figure 1-1. Distribution of powdery mildew scores among the progeny for July 12, 2011 (grey) and August 18, 2011 (black). Scores for parents *Vitis rupestris* B38 (star) and 'Chardonnay' (circle) are shown on the x-axis for both dates.

Genotyping: Sequencing, SNP calling and SNP selection

Averages of 712,400 reads per vine (n= 46) and 2,787,000 reads per vine (n= 42) were obtained per sequencing batch, for the Illumina GA3 and HiSeq machines, respectively. The distribution of the number of reads obtained per vine sample is shown in Figure 1-2. The TASSEL SNP call containing 42,172 SNPs had a correlation (r) of 0.9 between SNPs per chromosome and chromosome physical size in bp. Further filtering and minimal imputation lead to a less redundant set of 16,833 SNPs (17K SNP set) with r = 0.82 (Table 1-1).

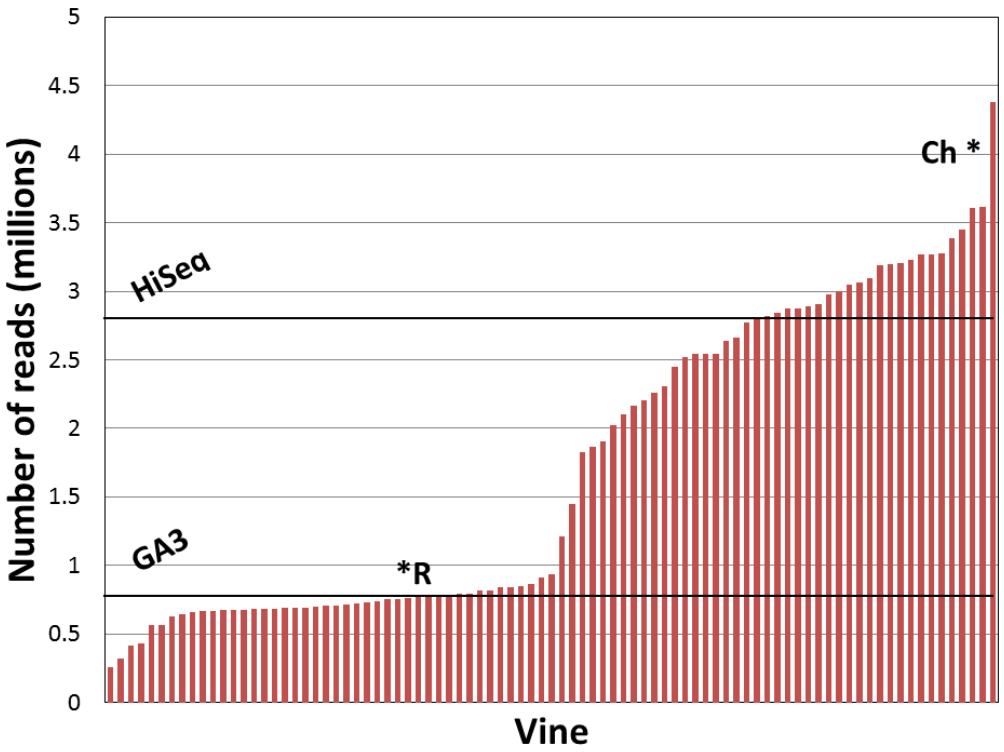


Figure 1-2 The number of reads obtained per vine. Numbers of reads obtained for parents *Vitis rupestris* B38 (R) and *Vitis vinifera* 'Chardonnay' (Ch) are indicated with *. Sample average values for each pooled library are indicated by horizontal lines.

Table 1-1. Number of SNPs per chromosome of the reference genome PN40024 in the 17K SNP set. SNPs were selected based upon maximum missing data of 20%, minor allele frequency of 0.25 ± 0.05, and inferring that rare allele homozygotes are actually heterozygous.

Chr[a]	Number of SNPs		Chr[a]	Number of SNPs
1	679		16	862
2	591		17	526
3	655		18	1,054
4	652		19	976
5	1,018		20	898
6	743		1_random	14
7	634		3_random	42
8	1,068		5_random	6
9	811		7_random	69
10	652		10_random	17
11	559		12_random	47
12	936		13_random	117
13	1,024		16_random	21
14	1,178		17_random	20
15	810		18_random	154
			Total	16,833

[a] Chr indicates chromosome location, Chr20 contigs have not been assigned to a chromosome, and suffix '_random' corresponds with unassembled portions of the indicated reference chromosome.

Distribution and density of 17K SNP set on the grapevine reference genome

The 17K SNP set showed similar marker densities among chromosomes, ranging from a mean value of one SNP every 36 kb on chromosome 4 to one SNP every 21 kb on chromosome 8. In order to identify local increases or decreases in marker density, the *V. vinifera* PN40024 reference genome was divided into 863 bins of 500 kb and the number of SNPs within each bin was determined.

Within chromosomes, SNP density varied. While 606 (70%) of the 500 kb bins had a moderate number (10-50) of SNPs, there were 240 bins (28%) with fewer than 10 SNPs, 7 bins with zero SNPs and 10 bins with a SNP density from 51 to 106 (Figure 1-3). The 17K SNP set had a mean value of 18 SNPs per 500 kb bin.

Single marker association test and LD analysis

Among all dates scored, 22 markers were significantly associated with PM severity at $\alpha = 0.05$ after multiple test correction. The strongest association was found between SNP S8_19258484 from 'Chardonnay' and PM severity evaluated on August 18, 2011, with a corrected p-value of 0.0255. The proportion of the phenotypic variance explained by each significant marker ranged from 0.209 to 0.278 (Table 1-2.) A subset of 10 alleles from 'Chardonnay' all led to increased PM severity, with estimated effects between 1.39 and 1.58. One minor allele from 'Chardonnay' in repulsion with the other 10 significant minor alleles (Appendix 1-2) reduced PM severity with an

estimated effect of -1.43. *V. rupestris* B38 alleles always reduced PM severity by 1.50 to 1.72.

Figure 1-3 SNP density across the *Vitis vinifera* 12X grapevine reference genome PN40024 (Jaillon et al. 2007). Each block represents a 500 kb bin of the *Vitis vinifera* 12X grapevine reference genome. The color scale on the right represents the number of SNPs located within each 500 kb bin. Bins with no markers are framed with grey color. From the 17K SNP set, 95% of markers were distributed among assembled chromosomes 1 to 19, with a mean value of 18 SNPs / 500 kb. Correlation (r) between chromosome size and number of markers per chromosome was 0.82.

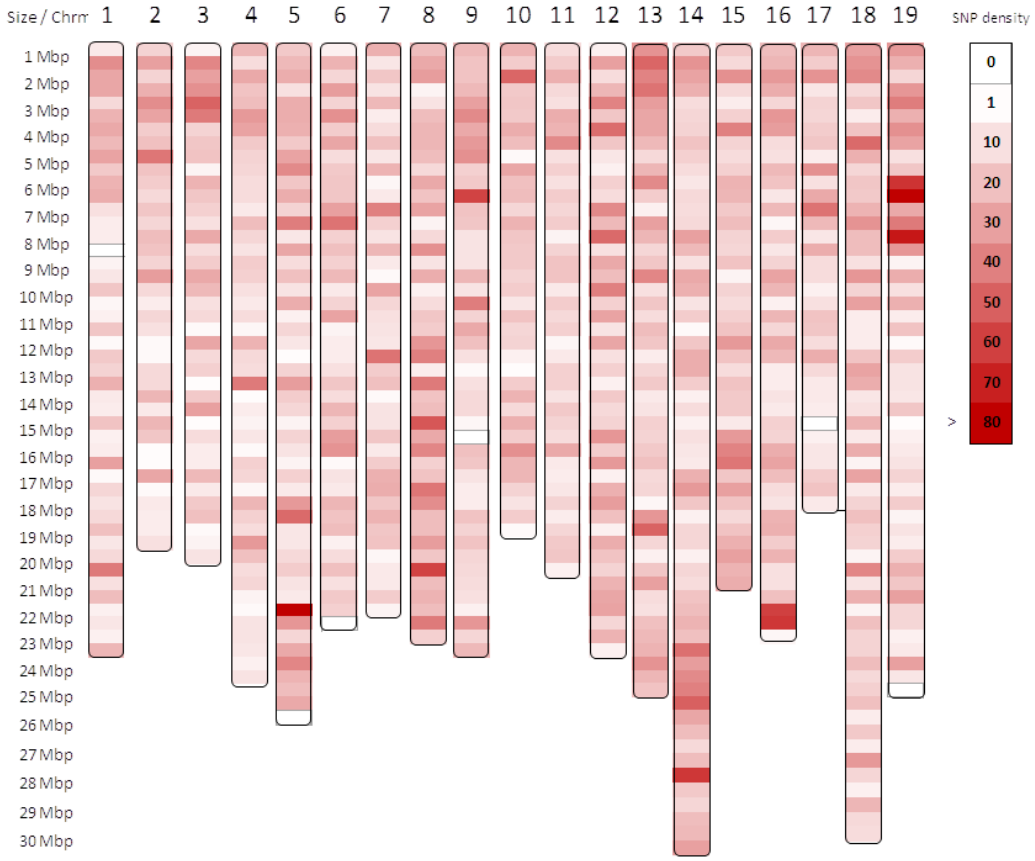


Table 1-2. Summary statistics for markers significantly associated with powdery mildew severity, at α = 0.05**.** A General lineal model was used to test for single marker associations with powdery mildew reactions using TASSEL 3.0.

	Single marker association test			LD analysis	
Marker[a]	p-value (BH)[b]	R^2	Allele effect	Parent	Locus LD (Chr)[c]
S7_14758877	0.0497	0.21	-1.54	*Vitis rupestris* B38	7
S13_18425381	0.0497	0.23	-1.50		7
S14_16921119	0.0497	0.26	-1.62		7
S20_17736100	0.0497	0.21	-1.51		7
S20_23796628	0.0497	0.23	-1.60		7
S20_23819240	0.0497	0.26	-1.65		7
S20_23819354	0.0497	0.28	-1.72		7
S20_32360020	0.0497	0.21	-1.54		7
S27_1104742	0.0497	0.22	-1.53		7
S27_1104824	0.0497	0.22	-1.53		7
S10_16893872	0.0497	0.25	-1.68		NA
S15_12704457	0.0263	0.25	-1.43	*Vitis vinifera* 'Chardonnay'	9
S2_790346	0.0361	0.27	1.41		9
S8_19258484	0.0255	0.27	1.58		9
S8_19258518	0.0315	0.24	1.39		9
S9_10531863	0.0441	0.21	1.42		9
S9_13661499	0.0361	0.23	1.53		9
S9_18099474	0.0361	0.23	1.53		9
S13_8723867	0.0375	0.21	1.46		9
S15_5224226	0.0315	0.23	1.50		9
S16_11260816	0.0315	0.22	1.47		9
S16_11260842	0.0315	0.22	1.47		9

[a] Marker name corresponds with chromosome location in the PN40024 reference genome. S20 markers have not been assigned to a chromosome, and S27 corresponds to unassembled portions of chromosome 7.

[b] The false discovery rate was controlled according to the Benjamini-Hochberg (BH) procedure.

[c] Locus LD reports the reference chromosome to which the marker was in linkage disequilibrium. S10_16893872 was not in LD with any PN40024 reference chromosome (NA).

The associated markers appeared to represent at least 10 different chromosomes based on the physical PN40024 reference sequence (Table 1-2.) To quickly test whether they genetically map to fewer loci, LD analysis was conducted. SNPs were genetically assigned to a chromosome if they were in significant LD with other markers on that chromosome (*e.g.*, Figure 1-4). For 11 of the 22 markers significantly associated with PM severity, the chromosomal placement from this LD analysis conflicted with the physical chromosomal assigned by alignment to the *V. vinifera* PN40024 reference genome (Table 1-2). Markers that aligned to chromosome 20 were not considered to be conflicting, as this “chromosome” is a collection of sequence contigs that have not yet been placed on any of the 19 chromosomes in the reference genome. All 11 significant SNPs coming from ‘Chardonnay’ were found to be in LD with other markers on chromosome 9. For *V. rupestris* B38, 10 out of 11 SNPs were found to be in LD with chromosome 7 (Figure 1-4); SNP S10_16893872 was not in LD with any chromosome (Appendix 1-3).

Figure 1-4 Linkage disequilibrium of a single SNP with whole genome SNPs from the corresponding parent, measured as D'. SNP S16_11260816 (panel A) is shown as a representative example of linkage to chromosome 9 in 'Chardonnay'. SNP S20_32360020 (Panel B) is shown as a representative example of linkage to chromosome 7 in *Vitis rupestris* B38. X-axis indicates SNP positions based on alignment to physical map (PN40024). All significant markers from Table 1-2 are shown in Appendix 1-3.

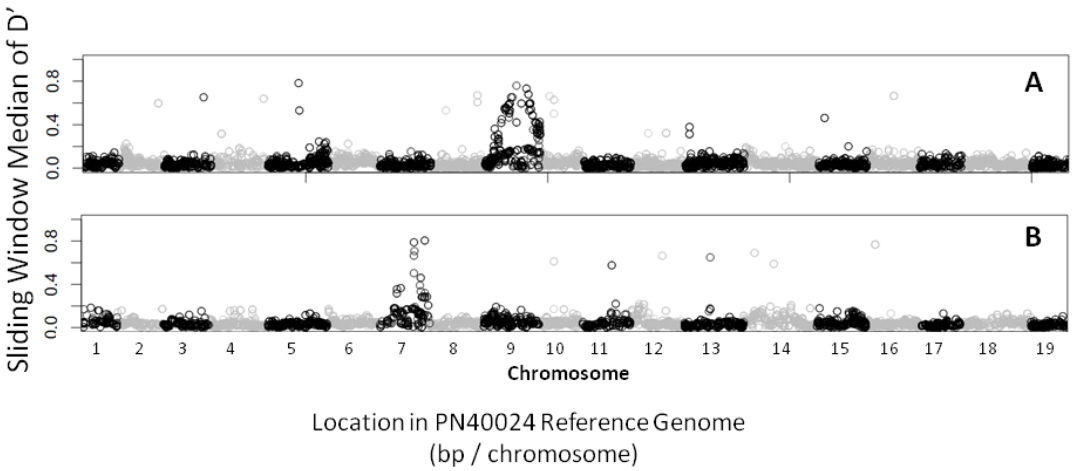


Table 1-3. Number of SNPs and total genetic distance (cM) of linkage groups (LG) in *V. rupestris* B38 and 'Chardonnay' maps. Linkage maps were constructed using Joinmap 4.1

	V. rupestris B38		'Chardonnay'	
LG	Number of SNPs	Total genetic distance (cM)	Number of SNPs	Total genetic distance (cM)
1	44	116.5	49	95.3
2	39	89.2	48	89.9
3	41	64.7	52	80.9
4	64	74.1	14	62.5
5	65	70.9	79	118.4
6	60	76.8	61	92.8
7	70	84.3	67	154.4
8	73	84.5	96	135.5
9	72	91.0	77	94.3
10	47	60.1	55	109.9
11	51	71.6	37	135.0
12	68	96.3	79	124.1
13	96	94.0	88	108.9
14	70	114.0	91	97.4
15	54	112.6	62	76.6
16	68	77.9	71	101.5
17	44	65.6	31	54.7
18	53	109.3	88	158.2
19	67	91.9	70	76.8
Total	1,146	1,645.3	1,215	1,967.4

QTL analysis

Using interval mapping, the susceptibility QTL from ‘Chardonnay’ was confirmed for 2 seasons (Figure 1-5). For *V. rupestris* B38, markers with significant scores from the single marker association test also had the highest LOD scores in interval mapping, but did not exceed the significance threshold. A summary of p-values and LOD scores for markers associated with the significant loci at each time point is shown in Table 1-4.

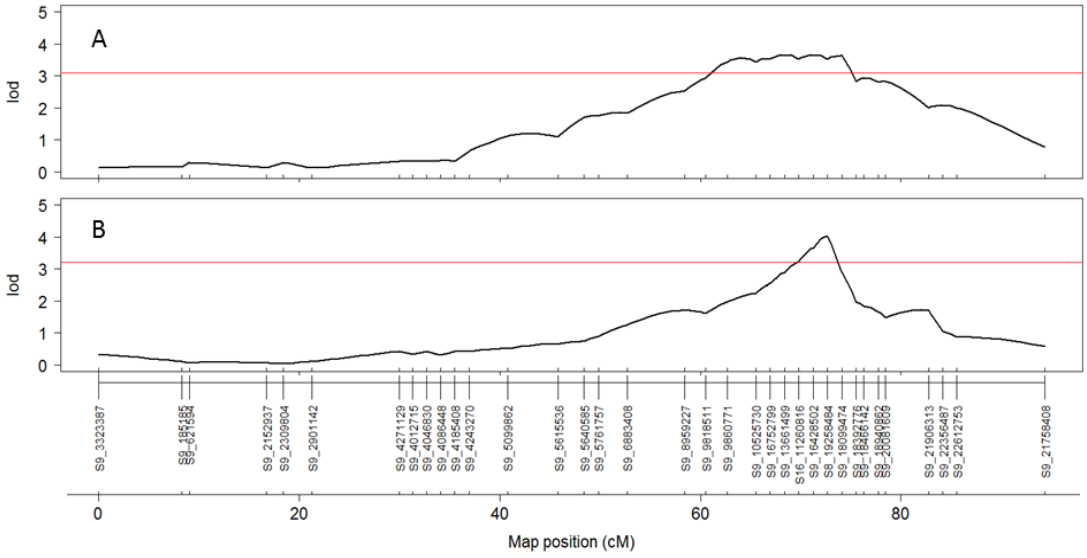


Figure 1-5 Interval mapping of a QTL for powdery mildew susceptibility on chromosome 9 in ‘Chardonnay’ on August 20, 2010 (A) and August 18, 2011 (B). Significance was calculated by permutation tests (10,000). X-axis indicates SNP positions based on the linkage map. Significant markers S15_12704457, S2_790346, and S8_19258518 are genetically redundant and map with significant marker S8_19258484 and marker S13_8723867 was genetically redundant and map with significant marker S9_13661499. Redundant markers do not appear in this figure.

Table 1-4. Summary statistics for selected markers at the four most significant field evaluation dates**.** A General Lineal Model was used to test for single marker associations with powdery mildew phenotype (-log(p-value)) and interval QTL mapping was performed using the scanone function in R/QTL (LOD).

Marker[a]	-log(p-value) (BH)[b]	LOD[c]	-log(p-value) (BH)[b]	LOD[c]
V. rupestris B38 markers	August 9, 2010		July 12, 2011	
S7_14758877	2.96	NA	4.45	NA
S13_18425381	2.85	1.83	4.58	2.32
S14_16921119	1.75	1.36	4.45	2.32
S20_17736100	3.07	NA	4.42	NA
S20_23796628	3.51	1.83	4.74	2.32
S20_23819240	3.62	NA	4.94	NA
S20_23819354	4.13	2.45	5.42	2.75
S20_32360020	3.63	1.83	4.45	2.32
S27_1104742	3.81	NA	4.60	NA
S27_1104824	3.27	NA	4.60	NA
S10_16893872	0.93	NA	4.82	NA
'Chardonnay' markers	August 20, 2010		August 18, 2011	
S15_12704457	1.68	3.53	5.20	4.04
S2_790346	1.64	3.53	4.54	4.04
S8_19258484	2.45	3.53	5.52	4.04
S8_19258518	1.73	3.53	4.79	4.04
S9_10531863	2.67	NA	4.38	NA
S9_13661499	3.29	3.63	4.52	2.90
S9_18099474	2.70	3.63	4.52	2.90
S13_8723867	3.05	3.63	4.48	2.90
S15_5224226	2.63	NA	4.78	NA
S16_11260816	2.55	3.53	4.69	3.24
S16_11260842	2.55	NA	4.69	NA

[a] Marker name corresponds with chromosome location in the PN40024 reference genome. S20 markers have not been assigned to a chromosome, and S27 corresponds to unassembled portions of chromosome 7.
[b] For single marker associations, the false discovery rate was controlled according to the Benjamini-Hochberg (BH) procedure.
[c] For interval QTL mapping, LOD was calculated using the Halley-Knott regression method. Significance thresholds ($\alpha = 0.05$) were obtained by permutation tests (10,000) and correspond to 3.06, 3.04, 3.09 and 3.19 for PM evaluation on August 9 (2010), July 12 (2011), August 20 (2010) and August 18 (2011), respectively.

Discussion

In this work we present the first application of GBS in a F1 family from two highly heterozygous grapevines. QTL analysis was performed in two stages: First, SNPs generated by GBS were filtered based on their segregation ratio among the offspring and the percentage of missing data to develop a stringent set of 16,833 high quality SNPs distributed evenly across the genome for which single marker associations with PM severity could be tested. Later, linkage maps with 1,146 and 1,215 SNPs were generated for *V. rupestris* B38 and 'Chardonnay', respectively, and interval mapping was performed to corroborate association test results. Use of the single marker association test and interval mapping led to the identification of the first QTL for PM susceptibility in *V. vinifera* 'Chardonnay', located on chromosome 9. While single marker tests identified SNPs associated with PM resistance from *V. rupestris* B38, interval mapping LOD scores were below the significance threshold. The results presented here justify naming the novel QTL from 'Chardonnay' *Sen*1 (Susceptibility to *Erysiphe necator* 1).

One strategy to select SNPs that are useful molecular markers is to determine the parental genotypes based on deep sequencing of their libraries (Davey et al. 2011) and then use this information to select SNPs that segregate at the expected ratio. As an alternative approach, we conducted shallow sequencing of parents and progeny and followed a strategy of selecting SNPs based on the segregation ratio among the progeny. In this work we analyzed only biallelic SNPs with MAF of 0.25 ($\pm$ 0.05), as a strategy for a simple and rapid

selection of reliable markers. As a consequence, several potentially informative SNPs were not considered, including biallelic SNPs with segregation distortion or that are heterozygous in both parents, and tri-allelic or quadra-allelic SNPs. The selection of the reference genome-based approach also limited the nature of the SNPs detected, as sequences from *V. rupestris* B38 with more than 2 polymorphisms from the reference genome were discarded at the alignment step. We favored the use of the reference-based pipeline to include all progeny in the analysis, as the non-reference pipeline (Lu et al. 2013) requires deeper reads to obtain similar results. Moreover, because LD extends for long blocks of the genome in a F_1 family, SNPs located in diverse regions that were excluded at the SNP calling stage could still be in linkage with nearby common regions.

Despite these constraints, the conservative selection strategy used here proved to be sufficient to give a robust set of 17K markers with good coverage of the *V. vinifera* PN40024 reference genome. The 17K SNP set was distributed across the grapevine reference genome, covering the entire length of each chromosome, with an average density of 36 SNPs/Mbp (Figure 1-3). Distribution analysis of the number of SNP markers located within 500 kb bins showed a pattern of continuous variation with some outlier 500 kb bins with high numbers of SNPs and some 500 kb bins with few or no SNPs. Continuous distribution of the counts indicated that the RRL created with the enzyme *ApeKI* succeeded in reducing the complexity of the genome without introducing a significant positional bias of the SNPs. Outlier 500 kb bins could be a minimal source of error and may be explained by several factors, including: errors in the physical map of the grapevine reference genome;

differences between the reference genome and parental genomes; and local enrichment of repetitive DNA for which reads would have been discarded (Figure 1-3).

In this study, SNPs derived from NGS data were prone to errors associated with the complexity of the grape genome, reference genome errors or lack of representation, the sequencing technology, and the characteristics of the GBS protocol. Thus reliable markers had to be filtered from among the initial collection of SNPs. Pooling samples in one sequencing lane lowers the read coverage for each SNP marker when compared with other genotyping strategies that use a single lane of NGS per sample (Myles et al. 2010), leading to an increase in the amount of missing data. Successful application of GBS has been reported in homozygous lines of maize and barley (Elshire et al. 2011; Poland et al. 2012) but the implementation of this technique in the F_1 progeny of two heterozygous parents required new approaches, as the error rate increases when heterozygous markers are miscalled as homozygous if only one allele has been sequenced.

Due to Joinmap 4.1 restrictions on the maximum number of markers that can be analyzed, the 17K SNP set was filtered using a selection criterion of MAF between 0.2 and 0.3. As a result, markers with segregation distortion were discarded from linkage analysis; hence, it is not possible to distinguish if regions with low numbers of markers were due to genomic diversity or segregation distortion. Despite the presence of a 32 cM gap in LG 1 of the *V. rupestris* B38 map, these maps represent an improvement over current grapevine linkage maps (Adam-Blondon et al. 2004; Doligez et al. 2006; Di Gaspero et al. 2007; Mahanil et al. 2011; Wang et al. 2012), with an average

spacing between markers of 1.52 and 1.87 cM and average maximum spacing (per chromosome) of 11.0 and 12.2 cM for *V. rupestris* B38 and 'Chardonnay' maps, respectively. These results suggest possible chromosome locations for contigs of the random chromosome (here called 20) and unaligned portions of chromosome 7.

SNPs obtained via the GBS approach showed the potential for high-throughput marker discovery and highlight the need for improvements in the creation of linkage maps, such as use of larger populations to resolve dense marker sets as well as better computational and statistical algorithms to phase, impute, order and genetically map larger sets of markers.

QTL were analyzed using two methods: Single marker association tests followed by LD analysis using the 17K SNP set, and interval mapping using parental linkage maps. In both methods, we found 2 loci associated with PM resistance or susceptibility, but with different levels of significance (Table 1-4). The QTL *Sen*1, from 'Chardonnay', associated with susceptibility to PM, was confirmed by both methods.

For *V. rupestris* B38, the single marker association test resulted in 11 significant alleles (p-value = 0.047) associated with reduced PM severity. In interval mapping, LOD scores for this locus were below threshold (Table 1-4), which is not surprising since the corrected (BH) p-values from the single marker association test were already just marginally lower than $\alpha = 0.05$. It is possible that either the elimination of 10 individuals reduced the statistical power of the test, or removed spurious associations. No major genes for resistance were found in either of the analyses, consistent with the quantitative

segregation patterns observed (Figure 1-1). As a consequence, we hypothesize that *V. rupestris* B38 has a quantitative resistance to PM due the action of minor QTLs. The association of these minor loci with PM resistance needs to be confirmed with greater power and resolution. To this end, we are characterizing *V. rupestris* B38 resistance in additional F_1 families.

In 'Chardonnay' the single marker association test revealed 10 markers linked to increased PM severity during one evaluation period (August 18, 2011) with estimated allele effects between 1.39 and 1.58. Only one minor allele in repulsion with the other 10 significant minor alleles from ‘Chardonnay’ (S15_12704457, Appendix 1-2) was associated with reduced PM severity, with an estimated effect of -1.43 evaluated early in the season, on July 12, 2011. Interval mapping confirmed this susceptibility QTL but now on two evaluation dates: August 20, 2010 and August 18, 2011.

The use of a General Linear Model (GLM) for single marker association tests allowed for rapid screening of a dense SNP dataset, without a requirement for linkage maps. This GLM approach is particularly valuable as it can accommodate a larger number of markers than standard software used for linkage mapping. In the current study, many of the markers that significantly predicted powdery mildew severity would have been discarded by the filters we used to develop the genetic linkage map. Since we had already identified associated markers by GLM, we could add these back to the genetic linkage map in order to saturate the QTL regions. However, GLM should not be used alone, as it does not take in to account population structure, or spurious individuals (eg, pollen contamination or sample mix-up) that can have a large

effect on the results. Therefore, interval mapping or a Mixed Linear Model needs to be used to confirm results.

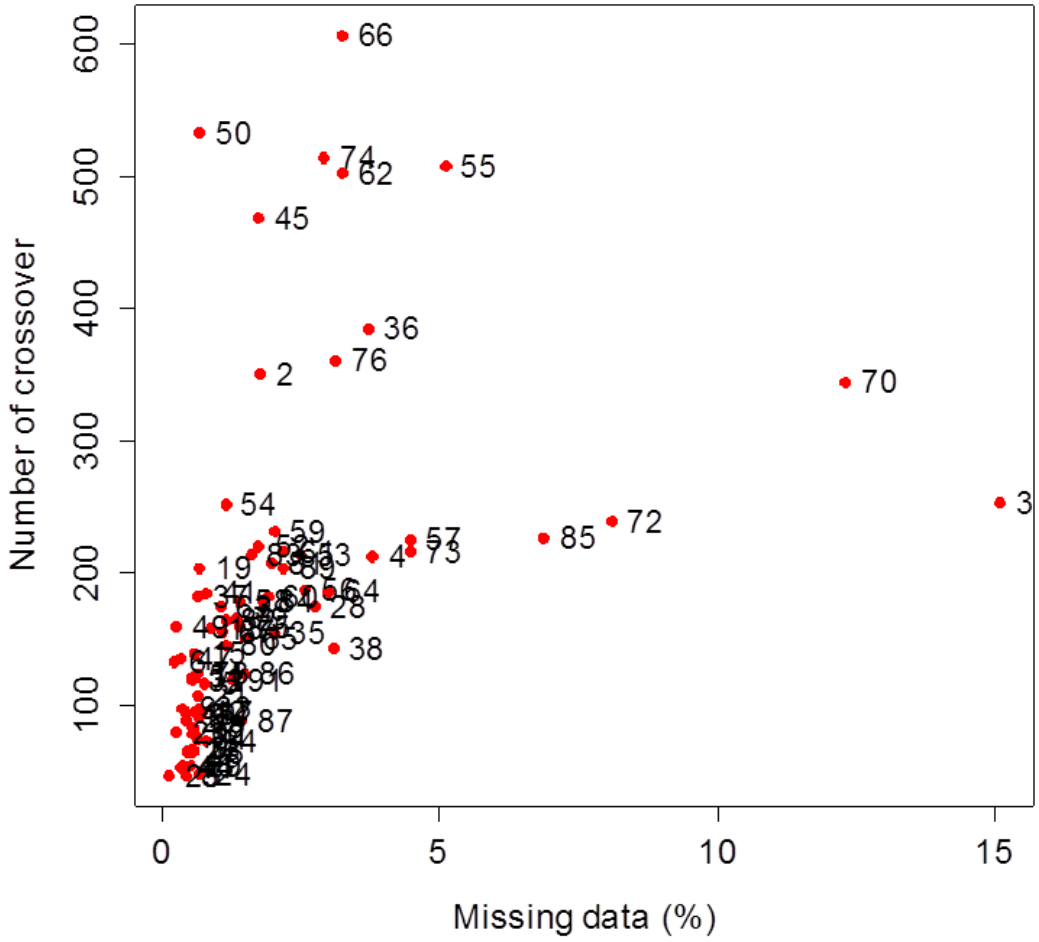
Based on homology with the *V. vinifera* PN40024 reference genome, significant markers obtained with a general linear model were initially assigned to locations on different chromosomes (Table 1-2). We used LD (D') to resolve this conflict and to assess the occurrence of major genome rearrangements. Analysis of LD of each individual marker against the whole set of SNPs from their parents revealed linkage to chromosome 7 for 10 of 11 significant SNPs from *V. rupestris* B38, and to chromosome 9 for all significant SNPs from 'Chardonnay' (Figure 1-4, Appendix 1-3). The genetic position of SNP S10_16893872 from *V. rupestris* B38 was not possible to establish due to lack of LD with any other marker. Whole genome LD analysis for 3,502 SNPs from *V. rupestris* B38 and 4,631 SNPs from 'Chardonnay' confirmed overall linkage of markers within chromosomes, with no major rearrangements in their genomes. A few markers in conflict with their physical alignment were observed in all chromosomes, indicating that this phenomenon is not specific to QTL regions. This suggests either rare errors in the *V. vinifera* PN40024 genome, genetic diversity between *Vitis* cultivars or species, or errors at the alignment step. Misalignment within the grapevine reference genome has also been found in other mapping populations (Wang et al. 2012) and differences within varieties of the same species has also been reported for other crops such as maize (Ganal et al. 2011).

In this work we present a high resolution SNP map for a biparental family derived from the cross of *V. rupestris* B38 and ‘Chardonnay’. Based on trait distribution, single marker association tests, and interval mapping, we found no evidence for qualitative inheritance of PM resistance, but we identified one moderate QTL for PM susceptibility, *Sen1* from ‘Chardonnay’, which constitutes a novel source of PM susceptibility as it is located on a chromosome where no grapevine – PM interaction loci have been described previously. Higher statistical power would be needed to identify remaining QTLs with small effects.

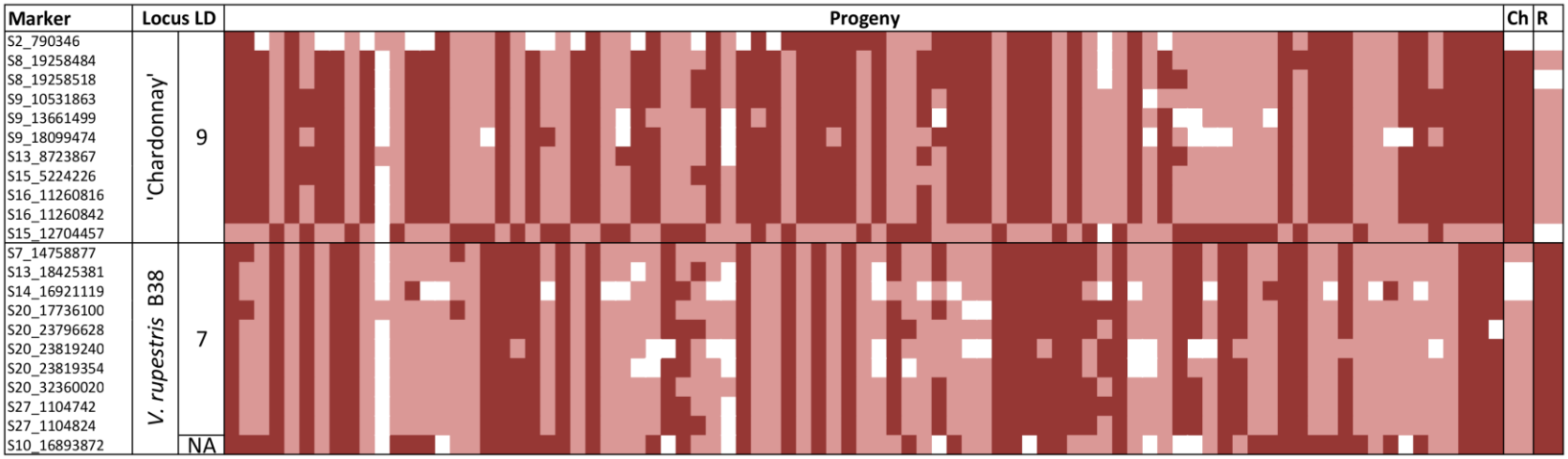
These results present new tools for grapevine MAS. A high resolution SNP set and dense linkage maps across the entire grapevine genome could be useful for retaining the *V. vinifera* background during the introgression of traits from non-cultivated relatives. In addition, GBS proved to be a useful method for high-throughput genotyping in heterozygous hybrid crosses, and its application in MAS could allow simultaneously genotyping of major, moderate or minor effect QTLs, such as *Sen*1 presented here, to develop long-lasting PM tolerance.

APPENDIX

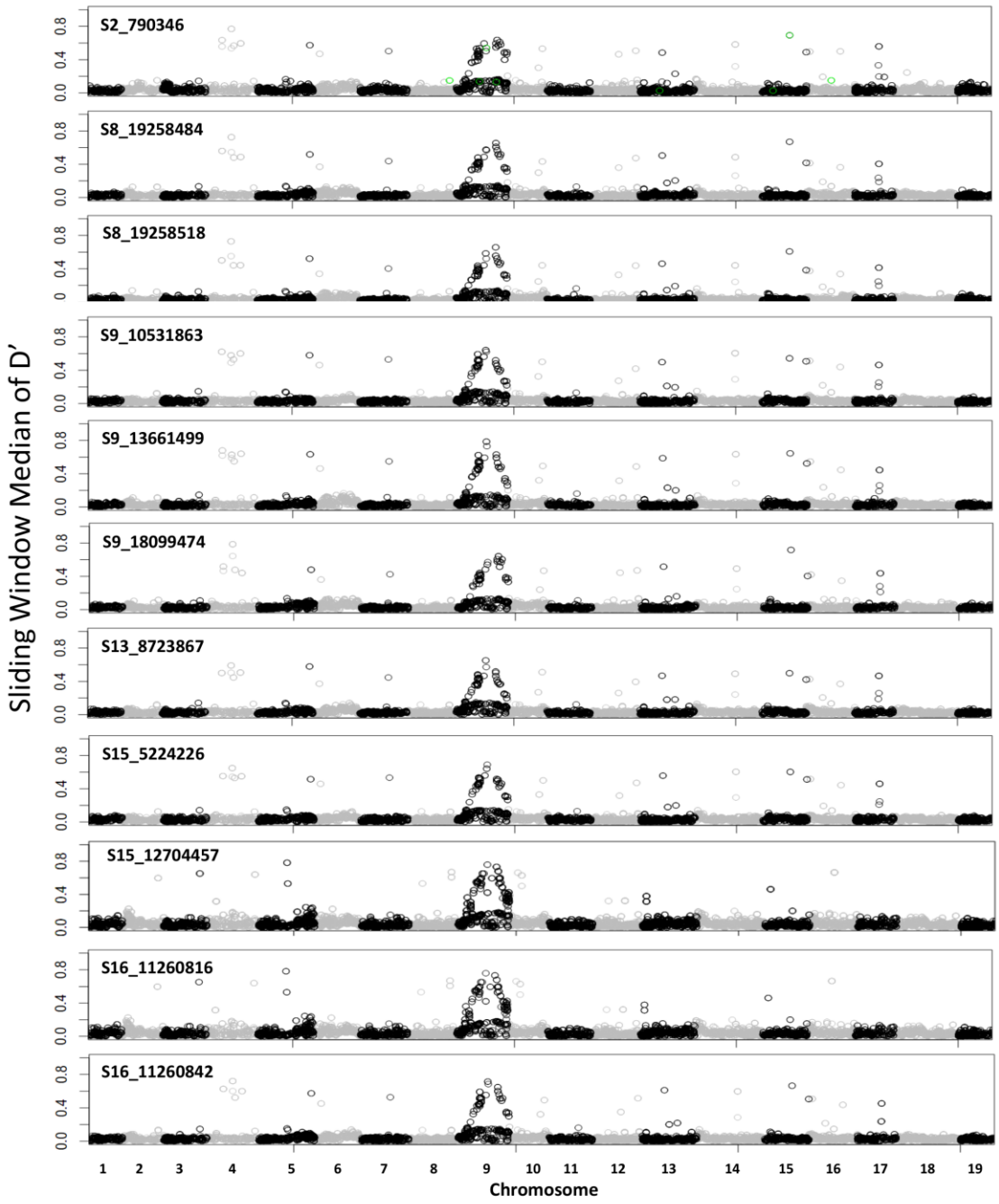
Appendix 1-1: Number of crossovers per individual as a function of percentage of missing data. Individuals 2, 36, 45, 50, 55, 62, 66, 70, 74 and 76 showing an increased proportion of crossing over were discarded from the linkage analysis.



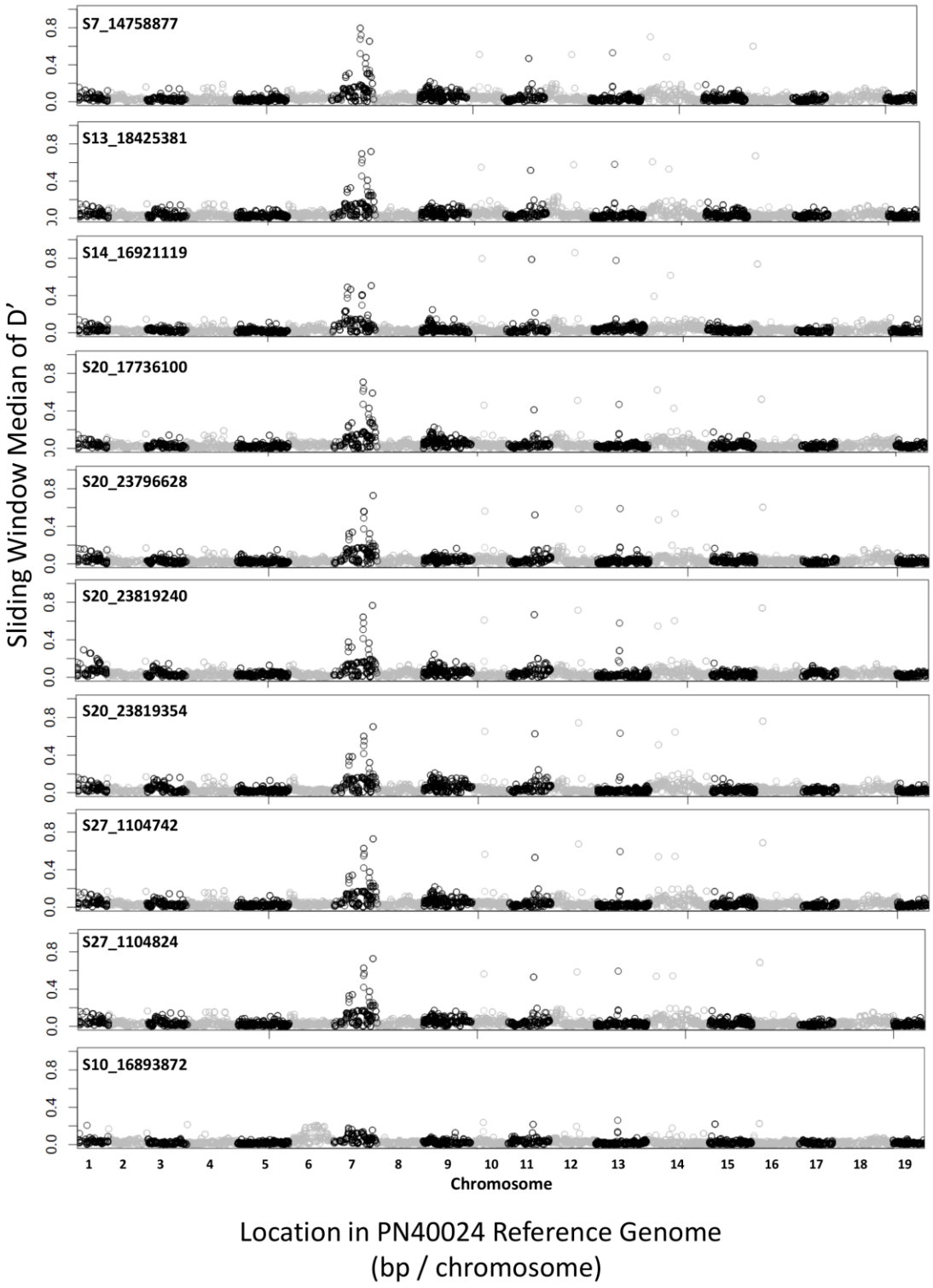
Appendix 1-2: The allelic state of each significant SNP in each of the parents and progeny. For each progeny, 'Chardonnay' (Ch) and *Vitis rupestris* B38 (R), each of these biallelic SNPs is coded as dark red or light red, with a blank for missing data. Marker name corresponds with chromosome location in the PN40024 reference genome. S20 markers have not been assigned to a chromosome, and S27 corresponds to unassembled portions of chromosome 7. Locus LD reports the reference chromosome to which the marker was in linkage disequilibrium. S10_16893872 was not in LD with any PN40024 reference chromosome.



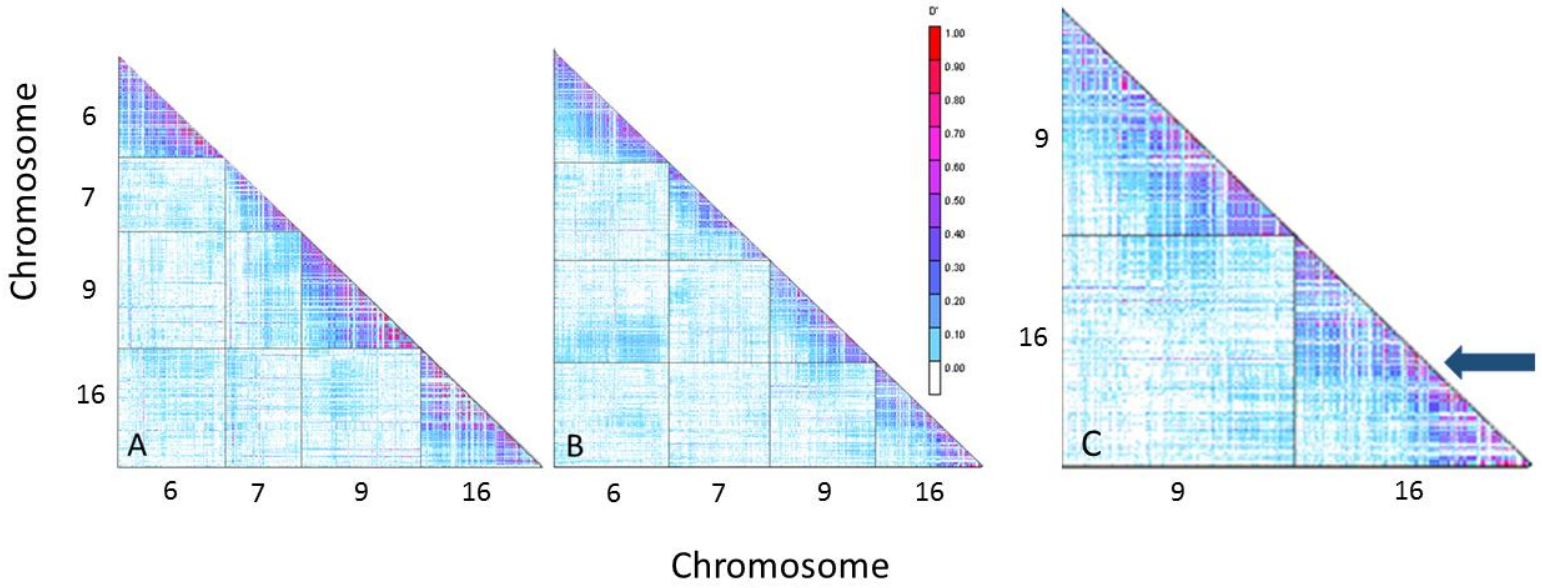
Appendix 1-3. Linkage disequilibrium of single SNPs with whole genome SNPs from the corresponding parent, measured as D'. X-axis indicates SNP position based on alignment to the physical map (PN40024). Additional information for each of these significant SNPs is provided in Table 2.


S2_790346
S8_19258484
S8_19258518
S9_10531863
S9_13661499
S9_18099474
S13_8723867
S15_5224226
S15_12704457
S16_11260816
S16_11260842
Sliding Window Median of D'
0.0
0.4
0.8
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
Chromosome
Location in PN40024 Reference Genome
(bp / chromosome)

Appendix 1-3. Continued



Appendix 1-4: Linkage disequilibrium (LD) analysis based on D′ calculations for a full matrix of selected chromosomes of *V. rupestris* B38 (A) and ‘Chardonnay’ (B). Chromosomes 7 and 9 contain SNPs associated with powdery mildew resistance and chromosomes 6 and 16 are representative of the pattern observed among the remaining chromosomes. X- and Y-axes indicate SNP positions based on alignment to the physical map (PN40024). Markers with high D' values are in LD (red to purple) as commonly seen by the within chromosome comparisons. Panel (C) represents an enlargement of the comparison between chromosomes 9 and 16 of ‘Chardonnay’. The arrow shows the position of significant marker S16_11260816. Alignment to the physical map placed this marker in chromosome 16, but LD analysis and the linkage map placed it on chromosome 9.



REFERENCES

Adam-Blondon AF, Roux C, Claux D, Butterlin G, Merdinoglu D, This P (2004) Mapping 245 SSR markers on the *Vitis vinifera* genome: a tool for grape genetics. Theor Appl Genet 109 (5):1017-1027.

Alleweldt G, Spiegel-Roy P, Reisch B (1991) Grapes (*Vitis*). In, vol 290. Acta Hort (ISHS), pp 291-330

Barbazuk WB, Bedell JA, Rabinowicz PD (2005) Reduced representation sequencing: A success in maize and a promise for other plant genomes. BioEssays 27 (8):839-848

Boubals D (1961) Étude des causes de la résistance des Vitacées à l'Oïdium de la vigne (*Uncinula necator* (Schw.) Burr.) et de leur mode de transmission héréditaire. Ann Amélior Plant 11:401-500

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23 (19):2633-2635

Broman KW, Wu H, Sen S, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics 19:889–890

Browning BL, Browning SR (2009) A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. AJHG 84 (2):210-223

Cadle-Davidson L, Mahanil S, Gadoury DM, Kozma P, Reisch BI (2011) Natural infection of *Run1*-positive vines by naïve genotypes of *Erysiphe necator*. Vitis 50:173-175

Coleman C, Copetti D, Cipriani G, Hoffman S, Kozman P, Kovacs L, Morgante M, Testolin R, Di Gaspero G (2009) The powdery mildew resistance gene *REN1* co-segregates with an NBS-LRR gene cluster in two Central Asian grapevines. BMC Genetics 10. doi:8910.1186/1471-2156-10-89

Dalbó MA, Ye GN, Weeden NF, Wilcox WF, Reisch BI (2001) Marker-assisted selection for powdery mildew resistance in grapes. J Am Soc Hortic Sci 126(1):83-89

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet 12 (7):499-510

Deulvot C, Charrel H, Marty A, Jacquin F, Donnadieu C, Lejeune-Henaut I, Burstin J, Aubert G (2010) Highly-multiplexed SNP genotyping for genetic mapping and germplasm diversity studies in pea. BMC Genomics 11:468

Di Gaspero G, Cattonaro F (2009) Application of genomics to grapevine improvement. Aust J Grape Wine Res 16:122-130

Di Gaspero G, Cipriani G, Adam-Blondon AF, Testolin R (2007) Linkage maps of grapevine displaying the chromosomal locations of 420 microsatellite markers and 82 markers for *R* -gene candidates. Theor Appl Genet 114 (7):1249-1263

Doligez A, Adam-Blondon A, Cipriani G, Di Gaspero G, Laucou V, Merdinoglu D, Meredith C, Riaz S, Roux C, This P (2006) An integrated SSR map of grapevine based on five mapping populations. Theor Appl Genet 113 (3):369-382

Eibach R, Zyprian E, Welter LJ, Topfer R (2007) The use of molecular markers for pyramiding resistance genes in grapevine breeding. Vitis 46 (3):120-125

Elshire R, Glaubitz J, Sun Q, Poland J, Kawamoto K, Buckler E, Mitchell S (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE.doi:10.1371/journal.pone.0019379

Feechan A, Kabbara S, Dry IB (2010) Mechanisms of powdery mildew resistance in the *Vitaceae* family. Mol Plant Path doi:10.1111/j.1364-3703.2010.00668.x

Ganal MW, Durstewitz G, Polley A, Bérard A,Buckler ES, Charcosset A, Clarke JD, Graner E-M, Hansen M, Joets J, Le Paslier M-C, McMullen MD, Montalent P, Rose M, Schön C-C, Sun Q, Walter H, Martin OC, Falque M (2011) A large maize (*Zea mays* L.) SNP genotyping array: Development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. PLoS ONE doi:10.1371/journal.pone.0028334

Glaubitz J, Casstevens T, Elshire R, Harriman J, Casstevens T (2012) TASSEL 3.0 Genotyping by Sequencing (GBS) pipeline documentation. http://www.maizegenetics.net/tassel/docs/TasselPipelineGBS.pdf Accessed 23 November 2012

IPGRI, UPOV, OIV (1997) Descriptors for grapevine (*Vitis* spp.). International Union for the Protection of New Varieties of Plants, Geneva, Switzerland/Office International de la Vigne et du Vin, Paris, France/International Plant Genetic Resources Institute, Rome, Italy

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, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Fabbro CD, Alaux M, Gaspero GD, Dumas V, Felice N, Paillard S, Juman I, Moroldo M, Scalabrin S, Canaguier A, Clainche IL, Malacrida G, Durand E, Pesole G, Laucou V, Chatelet P, Merdinoglu D, Delledonne M, Pezzotti M, Lecharny A, Scarpelli C, Artiguenave F, Pè ME, Valle G, Morgante M, Caboche M, Adam-Blondon A-F, Weissenbach J, Quétier F, Wincker P (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449 (7161):463-467.

Jones JDG, Dangl JL (2006) The plant immune system. Nature 444 (7117):323-329

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25 (14):1754-1760

Lu F, Lipka AE, Glaubitz J, Elshire R, Cherney JH, Casler MD, Buckler ES, Costish DE (2013) Switchgrass genomic diversity, ploidy, and evolution: Novel insights from a network-based SNP discovery protocol. PLoS Genet 9(1): e1003215

Mahanil S, Ramming DW, Cadle-Davidson M, Owens CL, Garris A, Myles S, Cadle-Davidson L (2011) Development of marker sets useful in the early selection of *Ren4* powdery mildew resistance and seedlessness for table and raisin grape breeding. Theor Appl Genet. doi:10.1007/s00122-011-1684-7

Moroldo M, Paillard S, Marconi R, Fabrice L, Canaguier A, Cruaud C, De Berardinis V, Guichard C, Brunaud V, Le Clainche I, Scalabrin S, Testolin R, Di Gaspero G, Morgante M, Adam-Blondon A (2008) A physical map of the heterozygous grapevine 'Cabernet Sauvignon' allows mapping candidate genes for disease resistance. BMC Plant Biol doi:10.1186/1471-2229-8-66

Morrell PL, Buckler ES, Ross-Ibarra J (2012) Crop genomics: advances and applications. Nat Rev Genet 13 (2):85-96

Myles S, Boyko AR, Owens CL, Brown PJ, Grassi F, Aradhya MK, Prins B, Reynolds A, Chia J-M, Ware D, Bustamante CD, Buckler ES (2011) Genetic

structure and domestication history of the grape. Proc Natl Acad Sci USA 108 (9):3530-3535

Myles S, Chia J-M, Hurwitz B, Simon C, Zhong GY, Buckler E, Ware D (2010) Rapid genomic characterization of the genus *Vitis*. PLoS ONE e8219.doi:10.1371/journal.pone.0008219

Nielsen R, Paul JS, Albrechtsen A, Song YS (2011) Genotype and SNP calling from next-generation sequencing data. Nat Rev Genet 12 (6):443-451

Pearson R (1988) Compendium of grape diseases. American Phytophatological Society APS, Minnesota, USA

Peressotti E, Wiedemann-Merdinoglu S, Delmotte F, Bellin D, Di Gaspero G, Testolin R, Merdinoglu D, Mestre P (2010) Breakdown of resistance to grapevine downy mildew upon limited deployment of a resistant variety. BMC Plant Biology doi:10.1186/1471-2229-10-147

Poland JA, Balint-Kurti PJ, Wisser RJ, Pratt RC, Nelson RJ (2009) Shades of gray: the world of quantitative disease resistance. Trends Plant Sci 14 (1):21-29

Poland JA, Brown PJ, Sorrells ME, Jannink J-L (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE doi:10.1371/journal.pone.0032253

Pollard KS, Gilbert HN, Ge Y, Taylor S, Dudoit (2004) S multtest: Resampling-based multiple hypothesis testing. R package version 2.12.0.

Riaz S, Dangl GS, Edwards KJ, Meredith CP (2004) A microsatellite marker based framework linkage map of *Vitis vinifera* L. Theor Appl Genet 108 (5):864-872

Troggio M, Malacarne G, Coppola G, Segala C, Cartwright DA, Pindo M, Stefanini M, Mank R, Moroldo M, Morgante M, Grando MS, Velasco R (2007) A dense single-nucleotide polymorphism-based genetic linkage map of grapevine (*Vitis vinifera* L.) anchoring Pinot Noir Bacterial Artificial Chromosome contigs. Genetics 176 (4):2637-2650

Van Ooijen JW (2006). JoinMap ® 4, Softwa re for the calculation of genetic linkage maps in experimental populations. Kyazma. Wageningen, Netherlands.

Van Tassell CP, Smith TPL, Matukumalli LK, Taylor JF, Schnabel RD, Lawley CT, Haudenschild CD, Moore SS, Warren WC, Sonstegard TS (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nat Meth 5 (3):247-252

Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, Pruss D, Pindo M, FitzGerald LM, Vezzulli S, Reid J, Malacarne G, Iliev D, Coppola G, Wardell B, Micheletti D, Macalma T, Facci M, Mitchell JT, Perazzolli M, Eldredge G, Gatto P, Oyzerski R, Moretto M, Gutin N, Stefanini M, Chen Y, Segala C, Davenport C, Demattè L, Mraz A, Battilana J, Stormo K, Costa F, Tao Q, Si-Ammour A, Harkins T, Lackey A, Perbost C, Taillon B, Stella A, Solovyev V, Fawcett JA, Sterck L, Vandepoele K, Grando SM, Toppo S, Moser C, Lanchbury J, Bogden R, Skolnick M, Sgaramella V, Bhatnagar SK, Fontana P, Gutin A, Van de Peer Y, Salamini F, Viola R (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS ONE doi:10.1371/journal.pone.0001326

Wang N, Fang L, Xin H, Wang L, Li S (2012) Construction of a high-density genetic map for grape using next generation restriction-site associated DNA sequencing. BMC Plant Biology doi:10.1186/1471-2229-12-148

Ward JA, Bhangoo J, Fernández-Fernández F, Moore P, Swanson JD, Viola R., Velasco R, Bassil N, Weber CA, Sargent DJ (2013) Saturated linkage map construction in *Rubus idaeus* using genotyping by sequencing and genome-independent imputation. BMC Genomics doi:10.1186/1471-2164-14-2

Wiedmann R, Smith T, Nonneman D (2008) SNP discovery in swine by reduced representation and high throughput pyrosequencing. BMC Genetics doi:10.1186/1471-2156-9-81

Xie W, Feng Q, Yu H, Huang X, Zhao Q, Xing Y, Yu S, Han B, Zhang Q (2010) Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing. Proc Natl Acad Sci USA. doi:10.1073/pnas.1005931107

CHAPTER 2

HETMAPPS: HETEROZYGOUS MAPPING STRATEGY FOR HIGH RESOLUTION GENOTYPING-BY-SEQUENCING MARKERS[1]

Abstract

Genotyping by sequencing (GBS) provides opportunities to generate high-resolution genetic maps at a low genotyping cost, but for highly heterozygous species, missing data and heterozygote undercalling complicate the creation of GBS genetic maps. To overcome these issues, we developed a publicly-available, modular approach called HetMappS (Heterozygous Mapping Strategy), which functions independently of parental genotypes and corrects for genotyping errors associated with heterozygosity. For linkage group formation, HetMappS includes both a reference-guided synteny pipeline and a reference-independent *de novo* pipeline. The *de novo* pipeline can be utilized for under-characterized or high diversity families that lack an appropriate reference. We applied both HetMappS pipelines in five half-sib F_1 families involving genetically diverse *Vitis spp*. Starting with at least 116,466 putative SNPs per family, the HetMappS pipelines identified 10,440 to 17,267 phased pseudo-testcross (Pt) markers and generated high-confidence maps. Pt marker density exceeded crossover resolution in all cases; up to 5,560 non-

[1] Conceived of and designed the experiments (Qi Sun, Bruce Reisch, Lance Cadle-Davidson), developed pipelines (Katie Hyma, Paola Barba, Minghui Wang), generated data (Bruce Reisch, Paola Barba, Katie Hyma, Sharon Mitchell, Charlotte Acharya, Jason Londo), analyzed and interpreted the data (Katie Hyma, Paola Barba, Qi Sun, Lance Cadle-Davidson), wrote the paper (Paola Barba, Katie Hyma, Lance Cadle-Davidson).

redundant markers were used to generate parental maps ranging from 1,047 cM to 1,696 cM. The number of markers used was strongly correlated with family size in both *de novo* and synteny maps (r = 0.92 and 0.91, respectively). Comparisons between allele and tag frequencies suggested that many markers were in tandem repeats and mapped as single loci, while markers in regions of more than two repeats were removed during map curation. Both pipelines generated similar genetic maps, and genetic order was strongly correlated with the reference genome physical order in all cases. Independently created genetic maps from shared parents exhibited nearly identical results. Flower sex was mapped in three families, and correctly localized to the known sex locus in all cases. The HetMappS pipeline could have wide application for genetic mapping in highly heterozygous species, and its modularity provides opportunities to adapt portions of the pipeline to other family types, genotyping technologies or applications.

Introduction

High throughput sequencing provides opportunities for generating high-resolution genetic maps at a low per-sample genotyping cost. While whole genome sequencing is currently a feasible and cost-effective approach to genotyping organisms with smaller genomes, the genotyping of larger genomes benefit from reduced representation library (RRL) approaches. The low cost and feasibility make this an ideal marker system for use in both model and non-model organisms, including specialty crops like grapevine.

While reduced representation approaches such as genotyping-by-sequencing (GBS) (reviewed in (Davey et al., 2011)) have been applied successfully for genetic map creation in inbred line (Spindel et al., 2013, Poland et al., 2012, Hart & Griffiths, 2015), organisms with highly heterozygous genomes, particularly high-diversity species, present additional computational challenges. The primary challenge for analysis of GBS data from heterozygous species stems from its primary advantage: GBS is a highly multiplexed, shallow sequencing strategy designed to simplify library production and minimize per sample cost. Shallow sequencing coverage results in missing data, genotyping error, and under-calling of heterozygous sites. While imputation of missing genotypes is practical in homozygous samples with known haplotypes and sequence order, imputation is error-prone for heterozygous samples of diverse materials or those lacking a suitable reference genome to infer order (Ward et al., 2013, Swarts et al., 2014).

Traditionally, genetic map construction in heterozygous crosses has utilized low-resolution markers with a low amount of missing data, infrequent genotyping error, and high quality parental genotype data (Grattapaglia & Sederoff, 1994). In contrast, GBS queries thousands of markers with a high proportion of missing data and variable genotyping error, and some datasets lack parental genotypes. Recent genotyping methods have been developed to overcome the problems of low sequence coverage and the lack of parental genotypes (Xie et al., 2010), but are optimized for inbred samples and not for heterozygous, outcrossing species. GBS data generated from crosses between two heterozygous parents are further complicated by difficulty in linkage group (LG) formation and phasing of markers, especially when high

quality genotypes of the parents is not available. Methods are available for LG formation and phasing, but most are either computationally infeasible for large data sets (Stam, 1993), and thus require filtering SNPs prior to analysis (Barba et al., 2014, Gardner et al., 2014), or have high error rates associated with low coverage, missing data and genotype error (Jansen et al., 2001, Van Ooijen, 2011). Other software provide tools for LG formation but require markers to be previously phased (Broman et al., 2003). Methods for haplotype discovery in diverse non-pedigree related populations can be applied, but many of these methods are computationally intractable for large datasets, or are strongly affected by genotyping error (Browning & Browning, 2011).

Grapevine exemplifies the high genetic diversity and heterozygosity of many specialty crops. Most specialty crop geneticists use low-throughput marker platforms, of which simple sequence repeats (SSRs) are most common. SSRs have the advantage of being transferable across diverse germplasm (Vezzulli et al., 2008a, Di Gaspero et al., 2000) and can provide more polymorphisms per marker than biallelic SNPs (Rafalski, 2002). However, logistical limits on multiplexing in data collection and analysis make SSRs an expensive marker technology for genetic mapping, primarily due to time involved and labor costs. As a result, SSR maps are typically low resolution, ranging from 100 to 600 markers per genome (Duchêne et al., 2012, Vezzulli et al., 2008b, Adam-Blondon et al., 2004, Mejia et al., 2007). Linkage disequilibrium decay in grapevines is rapid, with multilocus r^2 values declining down to 0.1 within 2.7 cM when measured as r^2 between SSRs (Barnaud et al., 2010) and faster than humans, *Arabidopsis* and maize when measured as r^2 between SNPs (Myles et al., 2011). In marker assisted selection it is

desirable to maintain linkage between marker and QTLs in a diverse genetic background, hence denser genetic maps are needed. SNP genotyping microarrays provided an improved option for higher resolution maps with 1,000 markers or more and low labor cost, but high technology cost. However, genotyping via microarrays suffers from ascertainment bias, and in high diversity species like grapevine, markers useful for one family are often not transferable even to closely related families because of flanking unknown SNPs (Mahanil et al., 2012, Miller et al., 2013). Recently, NGS has been successfully applied to marker discovery and genetic mapping in grapevines, but these SNP sets have been filtered down to meet the limits of the genetic mapping software used (Wang et al., 2012, Barba et al., 2014, Chen et al., 2015).

Hermaphroditism is a predominant domestication-related trait that gave rise to cultivated *Vitis vinifera* (Antcliff, 1980), with wild grapevines being dioecious. Flower sex is due to a single major locus with three alleles controlling male (M), hermaphrodite (H) and female (*f*) flower sex, where M is dominant over H, which is dominant over *f*. Genetic mapping has located the sex locus to chromosome 2, closely linked to the SSR locus VVIB23 (Dalbó et al., 2000, Marguerit et al., 2009, Riaz et al., 2006, Battilana et al., 2013). Two independent research groups have fine mapped the sex locus between 4.91 and 5.05 Mbp (Fechter et al., 2012) and between 4.89 and 5.04 Mbp (Picq et al., 2014) of the 12x.0 version of the PN40024 reference genome (Jaillon et al., 2007, Adam-Blondon et al., 2011a).

In the current study, we developed a modular computational pipeline for constructing and curating genetic maps from grape GBS data, with the option

of using synteny to assign markers to chromosomes, or *de novo* assignment to LGs. We selected families expected to be challenging - wide crosses involving several highly diverse *Vitis* spp. The method does not rely on known parental genotypes, but rather on the predictable genetic hallmarks of F_1 progeny, and this, along with utilizing a measure of genotype quality to mitigate potential heterozygote under-calling and correct putative genotyping errors, allows us to retain a higher number of markers than relying on high quality parental genotypes alone. We present results of the synteny based and *de novo* pipelines and curation for five half-sib families, and demonstrate localization of the flower sex locus on chromosome 2.

Materials and methods

Plant material and phenotype

Five half sib, interspecific F_1 families were generated using the following parental genotypes: *V. vinifera* 'Chardonnay' (hermaphrodite), *V. cinerea* B9 (male), *V. rupestris* B38 (female) and the hybrids 'Horizon' ('Seyval' x 'Schuyler', whose pedigree includes *V. vinifera, V. labrusca, V. aestivalis* and *V. rupestris,* hermaphrodite) and Illinois 547-1 (*V. rupestris* B38 x *V. cinerea* B9, male). The 'Horizon' x Illinois 547-1 family resulted from a cross made in 1988 (Dalbó et al., 2000) , and was enlarged with additional seedlings from a cross made in 1996. Additionally, two crosses were made in 2008: *V. rupestris* B38 x 'Chardonnay' and *V. rupestris* B38 x 'Horizon'; and two crosses were made in 2009: 'Horizon' x *V. cinerea* B9 and 'Chardonnay' x *V. cinerea* B9. Seeds were stratified and germinated the following year, and seedlings were

grown in an irrigated field nursery for one season followed by transplantation to a permanent vineyard in Geneva, New York.

For each progeny vine, flower sex was determined by visual observation during flowering time and fruit set, during two consecutive years. Fruit formation was confirmed later in the season. Male vines (including the parents Illinois 547-1 and *V. cinerea* B9) rarely set fruit, whereas hermaphrodite vines (including the parents 'Chardonnay' and 'Horizon') and the female parent *V. rupestris* B38 produced berries with fertile seeds.

Sample collection and DNA extraction

For each vine, a single small leaf (less than 1cm diameter) was harvested and placed in one tube of a Costar 96-well cluster tube collection plate (Corning, Corning NY, USA). Each 96-well plate received up to 91 unique samples plus two sets of duplicated and a blank well to serve as quality controls. The location of the blank well was unique for each plate in order to independently confirm the identity of plate after sequencing. Leaf tissue was maintained at 4 °C from harvest until delivery to the laboratory. Upon delivery, two stainless steel genogrinder beads were placed in each tube and the entire plate was frozen at -80 °C. When completely frozen, two 96-well plates were agitated at 2x400 speed for 1 minute in a Geno/Grinder 2000 (OPS Diagnostics LLC, Lebanon NJ, USA). Plates were then stored at -80 °C until processing with DNeasy 96-well DNA extraction kits (Qiagen, Valencia CA, USA). The following modifications were made from the manufacturer's protocol to improve DNA quality and quantity: 1) PVP-40 (2% w/v) was added to the AP1 lysis buffer prior to heating of the buffer, and 2) the agitation step following

AP1 addition was amended to include visual inspection of each 8-tube strip for complete re-suspension of the tissue pellet by hand or vortex. For the F_1 family *V. rupestris* B38 x 'Chardonnay', whole genome amplification was performed using 10 ng of DNA and the Illustra™ GenomiPhi™ V2 DNA Amplification Kit (GE Healthcare) to obtain 1.0 µg of dried DNA per sample. Then, 48-plex GBS library preparation and sequencing was completed, as described previously (Barba et al., 2014). For the other four F_1 families, DNA was quantified using the QuantiFlor dsDNA System (Promega) and processed as described below.

Library preparation

384-plex GBS libraries were prepared using a protocol modified from (Elshire et al., 2011). *Ape*KI barcode and common adapters (3ng each) were transferred into four 96-well plates, a different barcode adapter per well, and dried in a speed-vac. Sequences and barcodes comprising these four 96-plex adapters are presented in Appendix 2-1. DNA (100ng each) from four 96-well DNA plates described above was transferred to one of the four 96-plex adapter layouts, such that each DNA sample was assigned a unique barcode adapter. DNA digestion, adapter ligation, 96-well sample pooling, sample clean-up and PCR were performed as described in (Elshire et al., 2011) except DNA samples were digested with 1U *Ape*KI, and 18µL pooled template DNA was used in each of four PCRs (one PCR per 96-plex adapter set). The resulting four 96-plex GBS libraries were quantified using a fluorometer (Qubit®, Life Technologies, Grand Island NY, USA), diluted to 2 nM and combined in equal volumes for sequencing. Single-end sequences (100bp)

were collected on the HiSeq2000 (Illumina Inc., San Diego CA, USA) at the Institute of Biotechnology, Genomics Facility, Cornell University, Ithaca, NY.

SNP calling

A dataset derived from GBS sequencing of 8,353 *Vitis* spp. samples from the USDA-NIFA Specialty Crops Research Initiative *Vitis*Gen project (www.vitisgen.org) was used to call SNPs with the TASSEL-GBS pipeline, version 3.0.139 (Glaubitz et al., 2014), an extension to the Java program TASSEL (Bradbury et al., 2007). This *Vitis*Gen dataset included the four F_1 families (*V. rupestris* B38 x 'Horizon' with 215 individuals, 'Horizon' x Illinois 547-1 with 366 individuals, *V. vinifera* 'Chardonnay' x *V. cinerea* B9 with 148 individuals and 'Horizon' x *V. cinerea* B9 with 162 individuals) described above. Sequence data for the F_1 family *V. rupestris* B38 x *V. vinifera* 'Chardonnay' was previously generated, and SNPs were called independently for this family as described previously (Barba et al., 2014).

Tags (sorted, trimmed and collapsed de-barcoded sequence reads) were filtered and merged to generate a list of unique sequence tags for the *Vitis*Gen dataset. Then, tags were aligned to the 12X.0 *Vitis vinifera* reference genome PN40024 (Jaillon et al., 2007, Adam-Blondon et al., 2011a) using BWA version 0.6.2-r126 (Li & Durbin, 2009), with a maximum edit distance of 0.04. Only tags aligned to unique positions were used by the TASSEL-GBS pipeline during SNP calling. Chromosome names in the SAM file were modified for compatibility with the TASSEL-GBS pipeline in the following manner: Leading "EG:" was removed from chromosome names, trailing "_random" was replaced with "00", and "Un" was replaced with "999". The BWA generated SAM file was

converted to the Tags on Physical Map (TOPM), and individual Tags-by-Taxa (TBT) files were created for each individual sequencing lane.

SNPs were called from the TOPM and TBT files using the TASSEL-GBS pipeline. Genotype assignment during SNP calling follows (Etter et al., 2011), where likelihood scores for each possible genotype were calculated according to formula 3.8, and the most likely genotype was assigned. The GQ (genotype quality) score or phred-scaled confidence that the true genotype is the one provided in GT, is calculated according to GATK software (McKenna et al., 2010). Plugins and values used are described in Appendix 2-2.

Resulting VCF files containing genotype information were compressed with bgzip and indexed with tabix version 0.2.5 (r964) (Li, 2011). Separate files for each chromosome were concatenated into a single chromosome with the vcftools (version 0.1.9) utility vcf-concat. Individual datasets for each F_1 family were extracted from the *Vitis*Gen dataset using VCFtools (version 0.1.10) (Danecek et al., 2011). The VCF files produced by the TASSEL-GBS pipeline differ from the standard format in that the allele order is listed in major/minor order, rather than in reference/alternate order.

Quality control within F_1 families

To identify potential sources of contamination such as pollen impurity or mislabeling, each bi-parental family was considered for quality control separately. Genotypes were filtered on genotype quality (GQ ≥ 98), and relatedness was calculated using VCFtools, invoking options --GQ 98 and --

relatedness (Danecek et al., 2011). Relatedness of each progeny against each parent was visualized using R (R Development Core Team, 2013). Progeny that failed to have near zero relatedness to both parents and or failed to cluster with most other progeny were flagged for removal. Mendelian errors were also calculated with the PLINK option --mendel (Purcell et al., 2007). The proportion of male incompatible Mendelian errors (eg progeny genotypes carrying an allele not present in the male) to female incompatible Mendelian errors (eg progeny genotypes carrying an allele not present in the female) was calculated; progeny that showed elevated paternal to maternal incompatibilities were marked as pollen contaminants (including self-pollination) and flagged for removal. Individuals were removed based on the intersection of the relatedness and Mendelian errors results.

HetMappS pipeline

Quality-filtered GBS SNP datasets for each F_1 family were analyzed independently with the HetMappS pipeline using two approaches (Figure 2-1). Following (A) pseudo-testcross marker identification, marker grouping and ordering for each family was performed using one of two pipelines: (B1) a synteny pipeline, for which markers were initially separated into chromosomal groups based on alignment to the reference genome, filtered based on linkage, and then phased; or (B2) *de novo* genetic map pipeline, with linkage group (LG) formation based solely on progeny genotypes, regardless of alignment position, followed by phasing within LG. For both pipelines, although optionally for the synteny pipeline, (C) genetic ordering was carried out using MSTMap (Wu et al., 2008), and resulting maps were exported as cross files for analysis in R/qtl (Broman et al., 2003).

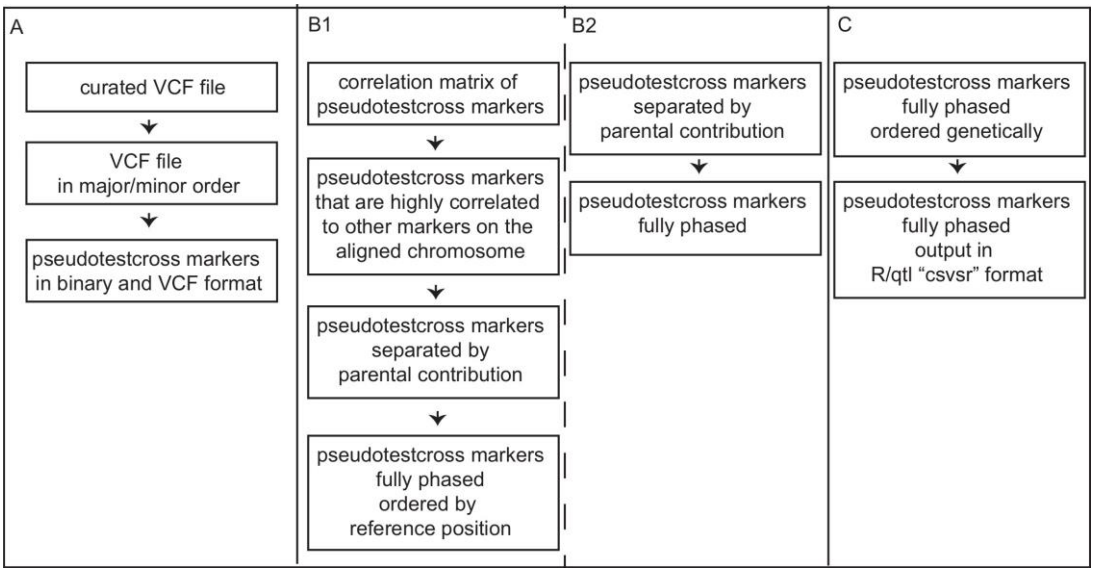


Figure 2-1. Overview of the HetMappS pipelines. (A) shared initial steps resulting in identification of pseudo-testcross markers, (B) linkage group creation and phasing steps, either (B1) synteny or (B2) *de novo*, and (C) genetic ordering and formatting for R/qtl.

Parent-independent identification of pseudo-testcross (Pt) markers

Pseudo-testcross (Pt) markers (markers that are heterozygous in one parent and homozygous in the other, eg, AA x AB (Grattapaglia & Sederoff, 1994)) were identified based on segregation patterns within the progeny. Inferring Pt markers based on progeny data eliminated the need for deeply sequenced parental genotypes of high quality and thereby maximized the number of retained markers. When provided, parental and grandparental genotypes with genotype quality (GQ) > 98 were retained for downstream validation.

Genotyping errors, including those associated with low coverage were then removed or corrected in the following manner. First, progeny genotypes were curated to remove alleles with less than 5% allele frequency (putative

sequencing errors) and to remove SNPs with a genotyping rate less than 50%. SNPs were selected based on their allele frequencies, genotype frequencies and levels of heterozygosity. SNPs with two segregating alleles (of frequency ≥ 5%), two or more segregating genotypes (of frequency ≥ 5%), minor allele frequencies of 0.25 ± 0.125, and major allele frequencies of 0.75 ± 0.125 were retained. For retained SNPs, genotypes were error corrected as follows: 1) AA genotypes (homozygous for major allele) with GQ < 98 were converted to missing and marked as masked, as they are potentially heterozygous genotypes that have been under sampled, 2) BB genotypes (homozygous minor allele) with GQ ≥ 98 were converted to missing and marked as errors, as this genotype cannot exist for Pt markers, 3) BB genotypes with GQ < 98 were assumed to be under sampled heterozygotes, converted to AB (heterozygous) and marked as corrected. Finally, SNPs were curated again to remove sites with genotyping rate less than 50% or with an error rate greater than 5%, measured as proportion of genotypes BB with GQ ≥ 98. The remaining markers are presumed Pt markers. A VCF file and “binary” file (with genotypes encoded as 1 or 0, indicating presence or absence of the minor allele) were generated for progeny and progenitor datasets. Progenitor and progeny genotypes were analyzed independently. All progenitor genotypes with GQ < 98 were filtered, and remaining genotypes at putative Pt loci were output.

.

Synteny option (reference-guided approach)

For the HetMappS Synteny Pipeline, Pt markers were initially grouped into physical chromosomes based on the coordinates of the reference genome; thus, SNPs identified from alignments with unassembled portions of the reference genome were not considered for further analysis. As described below, for each physical chromosome, a clustering method was used to separate maternal and paternal LGs and to determine phases within each LG. Markers were assigned to 12X.0 version of PN40024 reference physical chromosome using the physical position indicated in the VCF file (see SNP calling for details on the alignment step). The genetic linkage of SNPs within each chromosomal group was tested to discard mis-assigned markers using an r^2 matrix based on presence/absence of the minor allele. A filtering parameter ("diff") was set up as follows: Mean correlation between each query SNP with target SNPs across each of the 19 grapevine chromosomes was calculated and ranked. A query SNP was retained only if the mean correlation with the reference-assigned chromosome was at least two times (diff = 2) higher than the second mean chromosome correlation. All other SNPs were discarded.

Following correlation-based filtering, SNPs were separated into two or more LGs per chromosome, corresponding to segregation of the minor alleles, using the R package "WGCNA" (Langfelder & Horvath, 2008). Originally created for analysis of microarray data, this package was used to produce an adjacency matrix with the square of the marker correlations derived from binary genotype data. A topological overlap matrix is calculated from this adjacency matrix using the WGCNA function "TOMdist" and used for hierarchical clustering. For

each chromosome, LGs were created by cutting of the resulting dendrogram into distinct groups of SNPs. A static cut of the dendrogram with height 0.9 for the *Vitis*Gen families and 0.825 for the *V. rupestris* B38 x 'Chardonnay' family, with a minimum cluster size of 30, resolved two or more LGs per chromosome, which were later validated with parental genotype data for each family.
After LG formation, the same procedure was used to phase SNPs within each LG. However, the adjacency matrix was calculated as the marker correlation, with negative correlations converted to 0, as markers within the same phase are expected to be only positively correlated. Potentially mis-assigned markers, correlated with SNPs in both parental groups, were removed using "diff" = 2. A static cut of the dendrogram with height 0.9 and minimum cluster size of 30 resolved two phases for each LG for each family.

De novo option (non-reference approach)

The *de novo* pipeline differs from the above in that this approach does not rely on the physical position of the detected SNPs in the *Vitis* reference genome to create LGs. As such, this tool is a valuable addition for researchers studying heterozygous species where the reference genome is under construction, not accurate, or expected to be too diverged for the synteny pipeline.
LGs are formed through the clustering method described for the synteny-based LG formation, but across the entire dataset rather than within each chromosome. In order to determine the combination of cut height and minimum cluster size that results in LGs representing each chromosome for each parental map, the resulting dendrogram was cut for all pairwise combinations of a number of different static heights (0.95, 0.9375, 0.925, 0.9125, 0.9, 0.8875, 0.875, 0.8625 and 0.85,) and minimum cluster sizes (50,

100, 150, 200, 250, and 300). Then, LGs defined for each set of parameters were tested for mis-assigned markers using a correlation-based filtering similar to that described for chromosome assignment, but using LG assignment in lieu of chromosome alignment (here diff = 2, using correlation squared). For each family, a cut height and minimum cluster size of 50 was chosen for which at least 38 LGs were created, corresponding to the 19 chromosomes from each parent, and that maximized the proportion of markers assigned to a LG. After LG formation, the assignment of phases within each LG was perform as described above for the synteny pipeline, with the exception of filtering for mis-assigned markers, which was already performed as described above, during the LG formation step in the *de novo* pipeline.

Standardization of genetic maps

LGs were renamed and grouped into parental maps of 19 LGs, according to the grapevine community standards. First, correspondence of LG and its marker's physical position on the 12X.0 PN40024 reference genome (Jaillon et al., 2007, Adam-Blondon et al., 2011a) was determined. In the synteny pipeline this information is used to generate LGs, and was incorporated in the pipeline LG name, in a format 'chr number'_'cluster', where each chromosome/cluster combination represents a single LG. In the *de novo* pipeline, LGs were named arbitrarily, as the physical position was not considered to generate them. The *de novo* pipeline's random LG name was replaced by a physical chromosome number (as contained in the input VCF file) according to the chromosome with highest representation in the physical position of the SNPs. Each LG was assigned to the parent in whom segregation of the minor allele was observed for the majority of the markers.

LGs originating from the female parent were labeled from 1 to 19, and LGs originating from the male parent were labeled from 20 to 38.

Within each LG, phased clusters were renamed when grandparental information was available or when two LGs were joined. In the first case, the origin of each phased cluster can be assigned to a biological grandparent in whom segregation of the minor allele was observed for the majority of the markers. Here, for the 'Horizon' x Illinois 547-1 family, *V. rupestris* B38 and *V. cinerea* B9 genotypes were used to validate proper phasing for LGs corresponding to the Illinois 547-1 parental contribution for each chromosome. In the second case, when markers from the same physical chromosome are separated in two or more LG and no grandparental genotype is available, it is not possible to know *a priori* which two phases are in coupling. If these LG are joined with their phases in repulsion, an expansion of the genetic distances will be observed. Here, both combinations of phases for the two joining LG were tested, and the one that provide the shorter genetic distance was selected. When there are only two LGs per chromosome and grandparent genotypes data is unavailable, the phases remain arbitrary.

Ordering markers

Within each parental map, markers can be ordered by either the physical positions (synteny pipeline) or genetically.

Physical coordinates were assigned at the SNP calling step based on the *V. vinifera* reference genome (Jaillon et al., 2007) version 12X.0 (Adam-Blondon et al., 2011a) and then converted from the 12x.0 to 12x.2 (URGI, 2014) version, using one of the following approaches: For SNP coordinates, the 200

base pair context sequence flanking each SNP was retrieved from the 12x version, and was aligned to the 12x.2 version, using bwa version 0.7.5a-r405 (Li & Durbin, 2009). The new coordinates were inferred from the alignment coordinates. Only uniquely aligned sequences were converted. For sequence tags, both the start and stop positions were converted, as described for SNP coordinates.

Genetic ordering of the markers was achieved by computing the minimum spanning tree of an associated graph with MSTMap (Wu et al., 2008), with the following parameters: population_type = DH, distance_function = “kosambi”, cut_off_p_value = 2, no_map_dist = 15, no_map_size = 5, missing_threshold = 0.5, estimation_before_clustering = no, detect_bad_data = yes, objective_function = ML. Following the initial ordering step, redundant markers, suspicious genotypes identified by MSTMap, and markers that resulted in double crossovers for more than 20% of progeny were discarded. The remaining markers were re-ordered with the same MSTMap parameters as above.

Maps were then converted to R/qtl’s “csvsr” format for curation and QTL mapping. Data were encoded in both the 4-way cross format (“4way”) and the backcross (“BC”) format in order to access the full functionality of R/qtl (Broman et al., 2003).

Curation of genetic maps

Genetic maps in backcross format were independently curated in R/qtl in five steps.

1) Markers with ≥50% missing data and individuals with either >50% missing genotypes or with twice as many crossovers as the family mean were removed.

2) For each LG, marker order was evaluated by visual inspection of rf/LOD plots in R/qtl. In cases where a group of markers were in higher linkage with non-adjacent markers, marker order was re-assigned using the R/qtl function *switch.order*, or re-estimated using the R/qtl function *orderMarkers* after manually dropping conflicting markers identified by either high mean recombination fraction or low mean LOD related to their neighboring markers.

3) Genotyping error was determined as the value that maximizes the log likelihood estimate as described (Broman, 2010). The order of backwards LGs was inverted for correspondence with the SNP's physical position, and genetic distances for all LG were re-estimated using the Kosambi function.

4) The effect of dropping one marker at a time was estimated using *droponemarker* function in a sliding window of nine markers within each LG. A LOD difference threshold was determined for each map by plotting a histogram of the results and selecting the LOD difference value that removes the upper tail of the distribution. Suspicious markers that increase genetic distances over 2 cM and had LOD difference values above the selected

threshold were removed. Genetic distances and recombination fractions were re-estimated.

5) Finally, each LG was manually inspected for suspicious markers, including those creating gaps larger than 2cM or with high mean recombination fraction or low mean LOD related to their neighboring markers (S8 Table). Suspicious markers were removed, maps were rippled with a window of seven SNPs, and genetic distances were recalculated using the Kosambi mapping function Minor allele frequency (MAF) and minor tag frequency (MTF) were calculated for both, SNPs in final map and SNPs removed during curation using the VCF file generated at the SNP call stage. For each marker MAF was determined using --freq2 command in VCFtools (Danecek et al., 2011), and MTF was calculated as the read depth of the alternative allele divided by read depth of all alleles at the locus.

Mapping the flower sex locus

R/qtl software (Broman et al., 2003) was used to map the flower sex locus in the three F_1 families segregating for the trait, using both synteny and *de novo* curated maps in a 4-way format. Multipoint genotype probabilities were calculated using the *calc.prob* function with step = 1. A one-dimension scan was performed using the *scanone* function, with a binomial model, the Haley-Knott regression method and an error.prob = 0.01. LOD significance thresholds were determined by permutation tests (1,000). *Makeqtl* and *fitqtl* functions were used to fit a single locus model. Locus position was refined using the *refineqtl* function, and the presence of additional loci was tested with

the *addqtl* command. QTL confidence intervals were determined using a 1.8 LOD threshold.

Results

GBS SNP calling and family level quality control

From the 8,353 *Vitis*Gen samples included in SNP calling, 16,708,678 tags were identified. From these, 9,179,721 tags (54.94%) aligned to unique positions in the PN40024 reference genome, 1,190,427 (7.12%) aligned to multiple positions, and 6,338,530 (37.94%) could not be aligned. Overall genome coverage of all *Vitis*Gen tags uniquely aligned to the 12x.0 assembly was 13,774,386 bp, representing nearly 2.8% of it. In total, there were 301,506 sequenced intervals ranging from 12 to 979 bp in length, with a mean length of 103. Of these intervals, 262,082 (86%) were variable, resulting in a total of 1,881,000 putative SNPs from the 8,353 samples. The four F_1 families described here had a range of 852,885 to 1,219,257 tags present in each dataset, with alignment rates ranging from 66% to 73%, resulting in between 300,773 and 449,840 unfiltered SNPs.

All progeny were tested for relatedness to the parents and for Mendelian errors, in order to identify pollen contaminants, self-hybridization, and mislabeling. Individuals derived from pollen contamination or self-hybridization were more related to the mother and less to the father than true progeny, as shown for eight outliers in the ‘Horizon’ x Illinois 547-1 dataset (Figure 2-2). Additionally, seven of these eight individuals had a high ratio of male

incompatible Mendelian errors compared to female incompatible Mendelian errors. From zero to eight individuals were removed from each family, 1.7% of progeny across the families. After removing individuals, SNPs that were invariant or missing in all remaining individuals were removed.

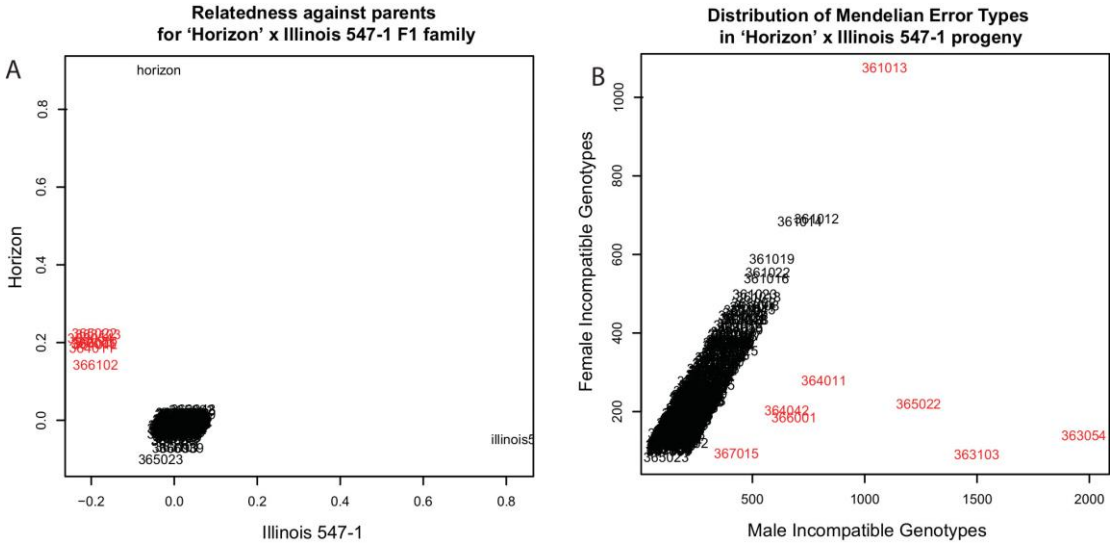


Figure 2-2. Relatedness to parents and Mendelian errors in the F_1 family 'Horizon' x Illinois 547-1. A) Analysis of progeny relatedness to parents demonstrated that most progeny had expected relatedness values near (0,0), whereas 8 individuals were more related to 'Horizon' (emasculated hermaphrodite parent) and less related to Illinois547-1 (pollen parent) and were thus removed for downstream analysis. B) Mendelian error analysis indicated that 7 of these same individuals were enriched for male incompatible genotypes.

Parent-independent identification of pseudo-testcross (Pt) markers

Pt markers were identified based on expected segregation of alleles (3:1) and genotypes (1:1) across progeny, filtering genotypes based on genotype quality scores to correct genotyping errors associated with heterozygote under-calling. Results were consistent across the four *Vitis*Gen F_1 families, with a

mean of 39% of SNP sites being removed due to low genotyping rate (< 50%) (Table 2-1). After first removing putative sequencing errors (alleles with < 5% MAF), 32% of SNPs became monomorphic. After further targeted genotype correction, 14% had incorrect segregation patterns, 12% had low genotyping rate, and only a 0.07% of the sites were rejected due to high error rate (i.e. sites with > 5% of high quality homozygote minor allele genotypes). At the end of this step, between 12,089 and 17,002 Pt markers were retained for construction of genetic maps, representing on average 4.2% of the starting number of putative SNPs.

Table 2-1. Number of SNPs after each stage of pseudo-testcross (Pt) marker identification for four F_1 families.

Stage	***V. rupestris* B38 x 'Horizon'**	**'Horizon' x Illinois 547-1**	**'Chardonnay' x *V. cinerea* B9**	**'Horizon' x *V. cinerea* B9**
Initial	337,365	449,840	300,773	331,356
Low genotyping rate	119,850 (36%)	177,955 (40%)	116,467 (39%)	130,042 (39%)
Monomorphic after sequencing error correction	115,865 (34%)	160,424 (36%)	84,141 (28%)	92,409 (28%)
Incorrect segregation patterns	45,381 (13%)	43,938 (10%)	51,775 (17%)	47,630 (14%)
Low genotyping rate after genotype error correction	41,141 (12%)	50,169 (11%)	36,084 (12%)	45,757 (14%)
High error rate	290 (0.09%)	352 (0.08%)	187 (0.06%)	181 (0.05%)
HetMappS Pt Markers Output	14,838 (4.4%)	17,002 (3.8%)	12,089 (4.0%)	15,337 (4.6%)

Results of HetMappS synteny pipeline

The synteny pipeline was executed using the HetMappS Pt markers identified above (Table 2-1). From each family, 6% of markers aligned to 'random' chromosomes (12X.0 version of PN40024 reference genome) were discarded. Another 5-7% were removed due to stronger genetic linkage with another chromosome than to the chromosome to which they aligned. Between 5 and 11% of markers could not be resolved, because they were correlated to multiple chromosomes, and the difference between correlations did not meet the thresholds imposed. The majority of SNPs (77-82%) were found to agree with the reference-assigned chromosome and were retained for downstream analyses (Table 2-2).

Table 2-2. Number of SNPs at each stage of correlation based chromosome assignment for four F_1 families analyzed with the synteny pipeline.

Stage	*V. rupestris* B38 x 'Horizon'	'Horizon' x Illinois 547-1	'Chardonnay' x *V. cinerea* B9	'Horizon' x *V. cinerea* B9
Initial	14,838	17,002	12,089	15,337
Markers on random chromosomes	902 (6%)	1,060 (6%)	731 (6%)	897 (6%)
In linkage with another chromosome	979 (7%)	1,196 (7%)	664 (5%)	901 (6%)
In linkage with multiple chromosomes	1,344 (9%)	773 (5%)	1,378 (11%)	1,466 (10%)
In linkage with aligned chromosome	11,613 (78%)	13,973 (82%)	9,316 (77%)	12,073 (79%)

Markers from each chromosome were further separated into linkage groups (LGs) corresponding to the minor allele contribution from each parent, through hierarchical clustering and dendrogram cutting (Figure 2-3). For all *Vitis*Gen families, a static cut height of 0.9 resolved at least one LG per chromosome per parent. In some cases, chromosomal groups were split into more than one LG per parent, resulting in a larger number of LGs than the expected 38 per grapevine family (Table 2-3). These extra groups can be fused together at the ordering step if the true classification can be determined, either by looking for recombination fractions lower than 0.5 (Broman, 2010) among groups or by reference to the parental contribution of the minor allele. Results shown here were based on the latter approach.

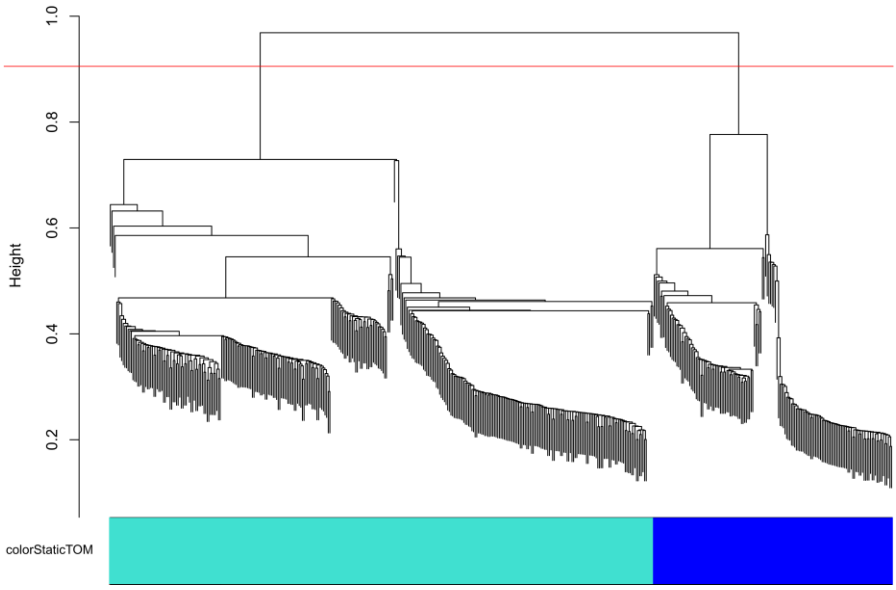


Figure 2-3. Separation of chromosomes into linkage groups (LGs) in the HetMappS synteny pipeline. This dendrogram was created from hierarchical clustering of a topological overlap matrix (as implemented by WGCNA) for markers on chromosome 2 in the F_1 family 'Horizon' x Illinois 547-1. LGs result from cutting the dendrogram with height 0.9 and minimum cluster size 30, creating one LG for each parent.

Table 2-3. Number of SNPs and linkage groups (LG) during LG and phase assignment for four F_1 families analyzed with the synteny pipeline.

F_1 family (# individuals)	# SNPs in chromoso-mal groups	# LGs	Split LGs (Parent)	# SNPs in phased LGs, female	# SNPs in phased LGs, male	# ordered SNPs after filtering
V. rupestris B38 x 'Horizon' (215)	11,613 (78%)	41	1,4,7 'Horizon'	3,449 (23%)	7,616 (51%)	5,897 (40%)
'Horizon' x Illinois 547-1 (366)	13,973 (82%)	40	1,7 'Horizon'	5,666 (33%)	7,521 (44%)	9,519 (56%)
'Chardonnay' x *V. cinerea* B9 (148)	9,316 (77%)	38	none	4,765 (39%)	4,044 (33%)	4,482 (37%)
'Horizon' x *V. cinerea* B9 (162)	12,073 (79%)	40	1 'Horizon'	7,503 (49%)	3,644 (24%)	5,185 (34%)

All percentages are relative to the number of Pt markers entering the synteny pipeline.

Markers from each LG were phased using hierarchical clustering and dendrogram cutting, with a static cut height of 0.9. For all LGs, markers resolved well into 2 phases. No markers were removed at this stage. Following phasing, and prior to ordering, LGs with markers from same physical chromosome were identified and joined, using the parental and physical information. In all cases it was possible to reconstruct 19 LGs per parental map. Each LG was assigned to the parent in whom segregation of the minor allele was observed for the majority of the markers.

Markers resulting from the synteny pipeline can simply retain their physical position on the reference genome (Jaillon et al., 2007) or can be ordered genetically. Here, we take both approaches and additionally convert tag alignment and SNP coordinates from the 12x.0 assembly (Adam-Blondon et al., 2011b) used for SNP call, to the more recent 12x.2 assembly (URGI, 2014).

Genetic ordering was performed with MSTmap (Wu et al., 2008). After the first round of ordering, at each genetic position the marker with the most information was retained. The total number of markers per F1 family varied between 4,482 and 9,519 (Table 2-3), depending on the number of progeny. The number of progeny highly correlated with the total number of non-redundant markers (r = 0.99).

Unique tag alignment density, mean tag depth, SNP density entering the pipeline, and minor allele frequency (MAF) of SNPs entering the pipeline were visualized on 1 MB sliding windows with a 100 KB slide using Circos (Krzywinski et al., 2009). Additionally, for both parents, phased SNP density (output of HetMappS synteny pipeline), MAF, and recombination frequency (obligate crossovers per progeny per MB) were visualized across these windows (Figure 2-4 and Appendix 2-3, 2-4 and 2-5).

Minor allele frequencies for SNPs entering the pipeline were 18% for the ‘Horizon’ x Illinois 547-1 family, and between 20% and 23% for the remaining three families prior to filtering and Pt marker identification. Following Pt marker identification, MAF of Pt markers was 25±1% (Figure 2-4 and Appendix 2-3, 2-4 and 2-5). Minor allele frequency was also calculated across genomic

windows of 1Mb size (with a 100Kb slide). The mean MAF within these windows ranged from 13% to 37%. Between 17 and 41 contiguous regions of segregation distortion (MAF < 20% or MAF > 30%) were discovered for each parental map, ranging from 0.1 to 3.1 Mb.

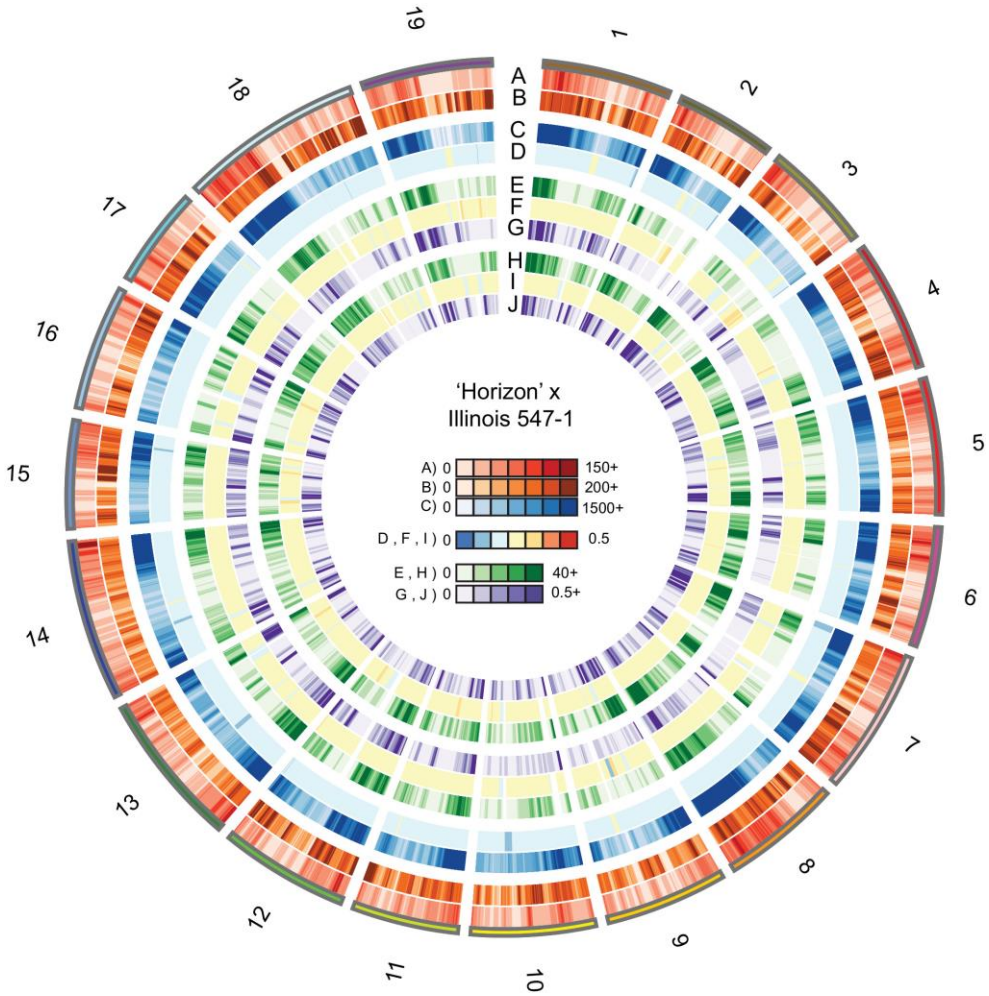


Figure 2-4. Visualization of 'Horizon' x Illinois 547-1 genomic data. Data are shown on 1 Mb windows with a 100 Kb slide. A) Number of unique tags aligned, B) Mean tag depth calculated as total tag depth over number of unique tags aligned, C) Density of SNPs entering the HetMappS pipeline, D) Minor allele frequency (MAF) of SNPs entering the pipeline, E-J) SNP output from the synteny pipeline: E) SNP density 'Horizon', F) MAF 'Horizon' SNPs, G) Recombination frequency 'Horizon', calculated as the number of obligate crossovers per progeny per Mb, H) SNP density Illinois 547-1, I) MAF Illinois 547-1 SNPs, J) recombination frequency Illinois 547-1, calculated as the number of obligate crossovers per progeny per Mb

Results of HetMappS de novo pipeline

The de novo pipeline was initiated using the HetMappS Pt output markers identified in Table 2-1. LGs were resolved by hierarchical clustering and dendrogram cutting of all Pt markers simultaneously, cutting the dendrogram (Figure 2-5) at a height that resulted in 38 or more LGs while maximizing the number of markers retained. The optimal cut height varied slightly among the four VitisGen families, with values between 0.925 and 0.95, and the number of resulting LGs obtained also varied, between 39 and 43 (Table 2-4). These extra groups can be fused together at the ordering step if the true classification can be determined, either by looking for recombination fractions lower than 0.5 (Broman, 2010) among groups or by reference to the parental contribution of the minor allele and physical location on a reference genome. Results shown here were based on the latter approach.

Markers from each LG were phased, as described for the synteny pipeline. For all F_1 families, a static cut height of 0.9 and static min size of 10 resolved all LGs into 2 phases. Split LGs with markers from same physical chromosome were identified and joined prior to genetic ordering, using parental and physical information. In all cases, it was possible to reconstruct 19 LGs per parental map.

Genetic ordering was performed with MSTmap (Wu et al., 2008). After the first round of ordering, at each genetic position the marker with the most information was retained. The total number of markers per F_1 family varied between 5,343 and 11,080 (Table 2-4), depending on the number of progeny.

The number of progeny was highly correlated with the total number of non-redundant markers (r = 0.99).

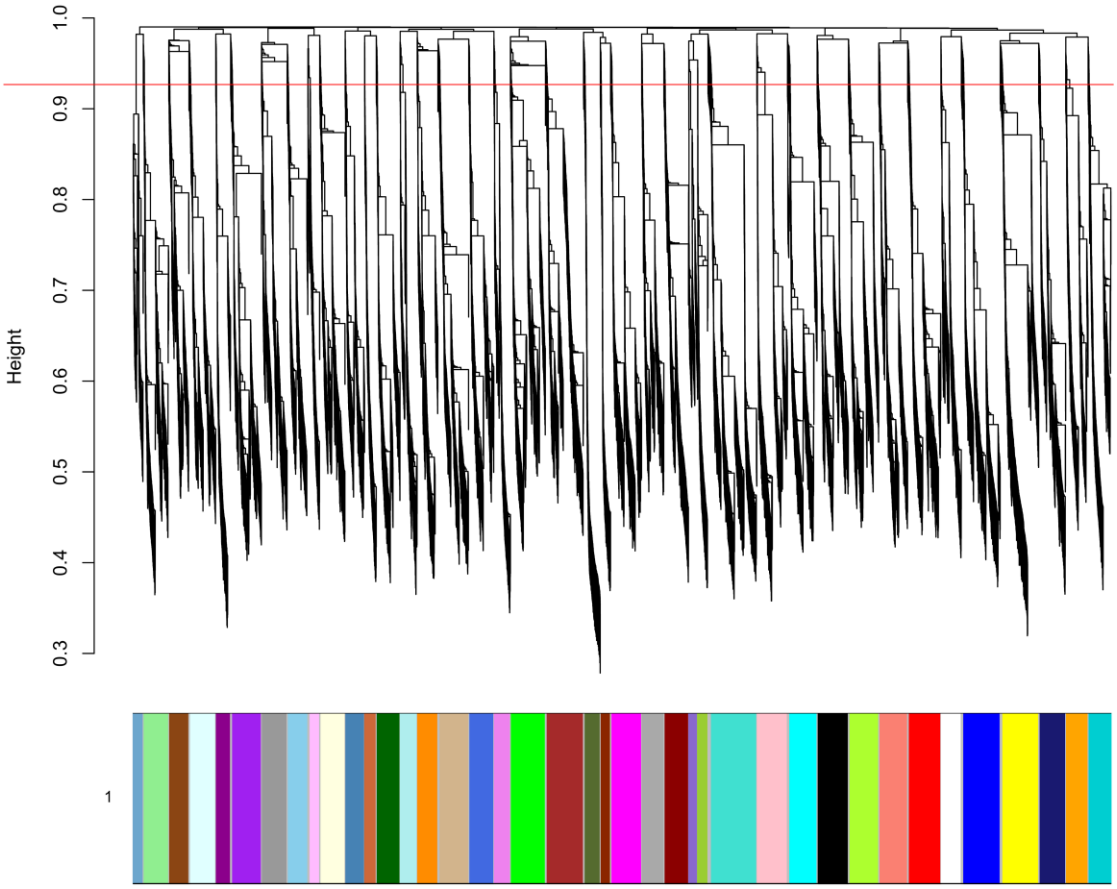


Figure 2-5. Linkage group (LG) formation in the HetMappS *de novo* pipeline. A representative dendrogram is shown for the F_1 family 'Horizon' x Illinois 547-1, created from hierarchical clustering of a topological overlap matrix (as implemented in WGCNA), and subsequent cutting of the dendrogram. A cut height of 0.925 and minimum LG size of 50 resulted in 91% of the initial markers (15,464) separating into 40 LGs. Two pairs of LGs were joined in subsequent steps to create 2 parental maps with 19 LGs each.

Table 2-4. Number of SNPs and linkage groups (LG) during LG and phase assignment for four F_1 families analyzed with the *de novo* pipeline.

F_1 family (size)	Cut height	# LGs	Split LG (parent)	# SNPs in LGs	# Filtered SNP in LG	# SNPs phased female	# SNPs phased male	# ordered SNP after filtering
V. rupestris B38 x 'Horizon' (215)	0.95	43	8, 18 *V. rupestris* B38 1,4,7 'Horizon'	13,829 (93%)	13,027 (88%)	4,200 (28%)	8,827 (59%)	7,018 (47%)
'Horizon' x Illinois 547-1 (366)	0.925	40	1,7 'Horizon'	16,115 (95%)	15,464 (91%)	6,670 (39%)	8,794 (52%)	11,080 (65%)
'Chardonnay' x *V. cinerea* B9 (148)	0.9375	39	17 'Chardonnay'	10,914 (90%)	10,440 (86%)	5,513 (46%)	4,927 (41%)	5,343 (44%)
'Horizon' x *V. cinerea* B9 (162)	0.925	41	1,4,7 'Horizon'	14,028 (92%)	13,260 (87%)	8,806 (57%)	4,454 (29%)	6,221 (41%)

Percentages below SNP number are relative to the number of pseudo-testcross markers entering the *de novo* pipeline.

HetMappS analysis of a pre-*Vitis*Gen family

Prior to the implementation of standardized sampling and genotyping methods for the *Vitis*Gen project, the family *V. rupestris* B38 x 'Chardonnay' was genotyped using GBS with four key differences for the pre-*Vitis*Gen samples: 1) whole genome amplification, 2) 48-plex libraries, 3) purification and size selection with AMPure beads, and 4) SNP calling independent of the 8,353 sample *Vitis*Gen build. Given the half-sib family experimental design, we analyzed this closely-related family for comparison to the *Vitis*Gen approach.

For this pre-*Vitis*Gen family, fewer raw SNPs were identified (116,466) and more progeny (13 out of 88) were filtered out during family level quality control. While there were fewer raw SNPs for this family, 14.8% of them were retained during Pt marker identification, compared to 4.2% of the markers in the *Vitis*Gen families, resulting in 17,267 Pt markers for *V. rupestris* B38 x 'Chardonnay', the highest of the five families tested. This was primarily due to fewer markers with low genotyping rate or <5% allele frequency, and fewer invariant sites after masking errors. However, this family had more markers with unexpected segregation ratios.

In the HetMappS Synteny Pipeline, only 49% of Pt markers for the *V. rupestris* B38 x 'Chardonnay' family were linked with reference-assigned chromosomes, due to frequent linkage with other chromosomes or with multiple chromosomes. For LG assignment, this family required a more relaxed cut height (0.825) than the *Vitis*Gen families (0.9), and nine LGs were split on the parental maps. Due to the small population size of 88 progeny, only 2,669 markers were ordered in the Synteny map.

In the HetMappS *de novo* Pipeline, the static cut height for the *V. rupestris* B38 x 'Chardonnay' family (0.875) was again outside the range of the *Vitis*Gen families (0.925 and 0.95), and the majority of LGs were split, resulting in 71 LGs. Furthermore, only the F_1 family *V. rupestris* B38 x 'Chardonnay' lost markers (1,807) during phasing. Some clustered markers had an atypical linkage pattern not seen in the *Vitis*Gen families, resulting in five large LGs with poorly defined linkage (Appendix 2-6). These five LGs were removed due their atypical linkage pattern, resulting in 66 LGs in the *V. rupestris* B38 x 'Chardonnay' map.

Results of genetic map curation

Genetic maps from both synteny and *de novo* pipelines were evaluated in terms of marker order and total size. An average of 1 and 2 LGs per map showed major order problems in synteny and *de novo* pipelines, respectively, and all genetic maps derived from MSTMap showed larger genetic distances than the ~100 cM per LG that is typical for *Vitis* F_1 families.

Initial genetic distances ranged from 1,891 cM (1,231 SNPs) for the female synteny map of the *V. rupestris* B38 x 'Chardonnay' family, up to 8,311 cM (4,450 SNPs) for the male *de novo* map of the 'Horizon' x Illinois 547-1 family (Table 2-5 and 2-6, respectively). Map curation with R/qtl was effective in reducing map size in both synteny and *de novo* pipelines, from an average distance of 2,906 cM and 4,878 cM to average distances of 1,286 cM and 1,351 cM respectively. The number of markers in the final map was highly correlated with the number of individuals in the family both synteny and de novo pipeline (r = 0.92 and 0.91, respectively). Final maps from both synteny and *de novo* pipelines resulted in good correlations between the genetic order and physical position of SNPs, and also had good coverage of the physical genome (Figure 2-6).

Table 2-5. Synteny map distances and numbers of markers throughout the curation process for pre-*Vitis*Gen and *Vitis*Gen families.

F_1 family (# individuals)		Initial # SNP	Initial Genetic distance (cM)	# SNP after order	Genetic distance (cM) after order	# SNP after auto drop	Genetic dist. (cM) after auto drop	Final # SNP	Final genetic dist. (cM)
V. rupestris B38 x 'Chardonnay' (69)	Fem.	1,231	1,891	1,231	1,671	1,210	1,526	925	1,275
	male	1,438	2,379	1,373	2,021	1,333	1,748	1,026	1,468
V. rupestris B38 x 'Horizon' (211)	Fem.	2,092	2,600	2,092	2,589	2,046	2,023	1,924	1,313
	male	3,805	3,525	3,805	3,511	3,732	2,667	3,493	1,583
'Horizon' x Illinois 547-1 (344)	Fem.	4,158	2,856	4,158	2,996	4,089	2,294	3,875	1,229
	male	5,361	4,462	5,361	4,695	5,264	2,604	4,992	1,211
'Chardonnay' x *V. cinerea* B9 (140)	Fem.	2,375	2,239	2,375	2,226	2,332	1,879	2,108	1,233
	male	2,107	2,436	2,077	2,262	2,032	1,869	1,869	1,180
'Horizon' x *V. cinerea* B9 (145)	Fem.	3,297	4,563	2,913	4,036	2,796	3,004	2,600	1,320
	male	1,888	2,110	1,653	1,873	1,610	1,435	1,520	1,047
Average:		2,775	2,906	2,704	2,788	2,644	2,105	2,433	1,286
Percentage of Initial:				97%	96%	95%	72%	88%	44%

Table 2-6. *De novo* map distances and numbers of markers throughout the curation process for pre-*Vitis*Gen and *Vitis*Gen families.

F_1 family (# individuals)		Initial # SNP	Initial genetic distance (cM)	# SNP after order	Genetic distance (cM) after order	# SNP after auto drop	Genetic dist. (cM) after auto drop	Final # SNP	Final genetic dist. (cM)
V. rupestris B38 x 'Chardonnay' (69)	Fem.	1,594	3,280	1,558	2,469	1,471	1,782	1,067	1,322
	male	1,772	3,303	1,663	2,348	1,577	1,879	1,199	1,459
V. rupestris B38 x 'Horizon' (211)	Fem.	2,568	4,523	2,568	4,247	2,441	2,661	2,225	1,388
	male	4,450	8,312	4,450	7,880	4,268	5,681	3,889	1,696
'Horizon' x Illinois 547-1 (344)	Fem.	4,857	6,547	4,784	5,873	4,635	4,194	4,316	1,286
	male	6,223	6,152	6,106	5,186	5,936	3,548	5,560	1,314
'Chardonnay' x *V. cinerea* B9 (140)	Fem.	2,772	3,474	2,772	3,240	2,676	2,446	2,394	1,275
	male	2,571	3,661	2,537	3,328	2,441	2,377	2,177	1,293
'Horizon' x *V. cinerea* B9 (145)	Fem.	3,886	5,971	3,560	3,913	3,396	2,297	3,118	1,347
	male	2,336	3,554	2,271	3,084	2,161	2,133	1,956	1,125
Average:		3,303	4,878	3,227	4,157	3,100	2,900	2,790	1,351
Percentage of Initial:				98%	85%	94%	59%	84%	28%

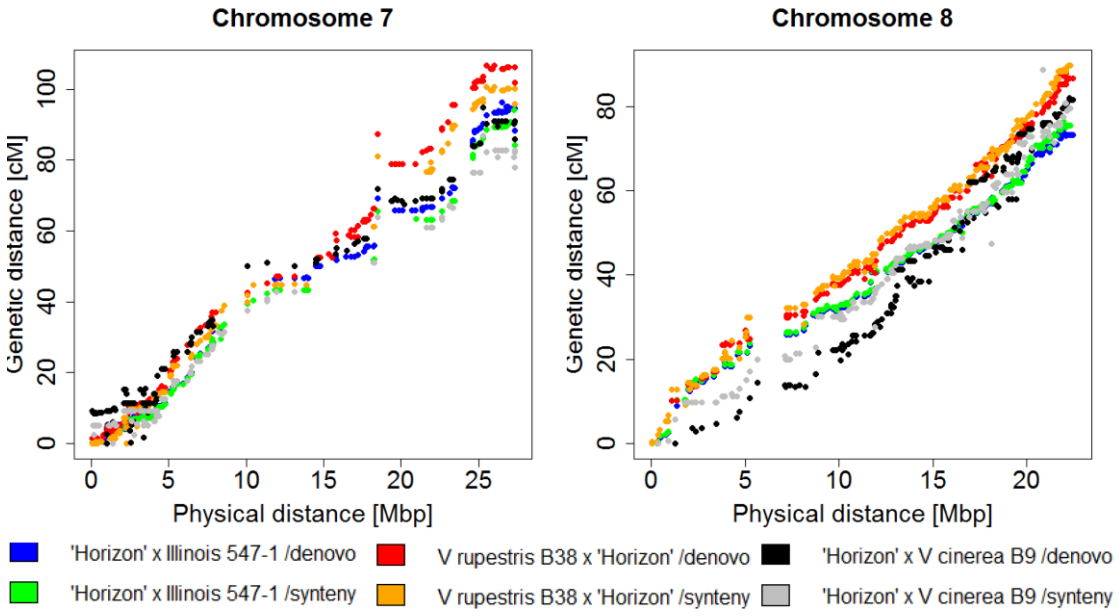


Figure 2-6. Physical position (PN40024 12X.2) vs genetic distances in six independent maps for chromosomes 7 (A) and 8 (B) of 'Horizon'. Genetic maps were generated with both synteny and de novo pipelines from three independent F_1 families: *V. rupestris* B38 x 'Horizon', 'Horizon' x *V. cinerea* B9 and 'Horizon' x Illinois 547-1.

While the minor allele frequency (MAF) in the final maps showed a normal distribution centered near 0.25 (Figure 2-7, panel A, dotted line), the minor tag frequency (MTF, measured at read depth level of alleles) suggested a bimodal distribution with peak modes of 0.25 and 0.125 (Figure 2-7, panel A, solid line). In contrast, markers removed during the map curation process were distributed across all MTFs with a peak lower than 0.125 and with no correlation to the MAF distribution (Fig. 7, panel B, dotted line).

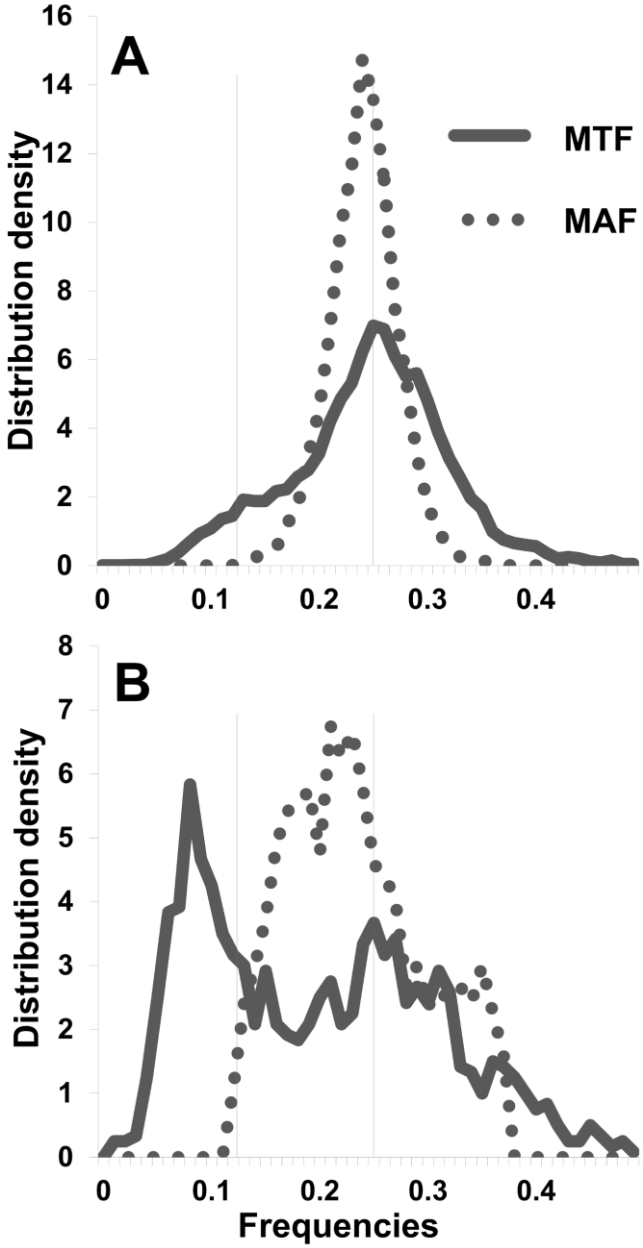


Figure 2-7. Comparison of the distribution of minor tag frequency (MTF) and minor allele frequencies (MAF) in final and spurious SNPs. MTF (continuous line) and MAF (dotted line) distributions are shown for A) the final map (9,876 SNPs) and B) markers removed during curation (1,204 SNPs) in 'Horizon' x Illinois 547-1 *de novo* map.

Localization of the flower sex locus

Among the five interrelated families, flower sex segregated as a major locus following the current genetic model of dominance M > H > *f* for three families: 'Horizon' x Illinois 547-1, 'Horizon' x *V. cinerea* B9 and 'Chardonnay' x *V. cinerea* B9, with a 1:1 ratio of male:hermaphrodite ($\chi^2(1)$ = 3.1935, 0.7042 and 2.3478, respectively). Crosses made with the female parent *V. rupestris* B38 (*f f*) showed no segregation of flower sex with all progeny being hermaphroditic, indicating that that 'Chardonnay' (Fernandez et al., 2006) and 'Horizon' are homozygous for the hermaphrodite allele (HH). In all three crosses and both pipelines, the flower sex locus consistently mapped to a physical position between 4.75 Mb to 5.39 Mb of the PN40024 version 12X.0 (Table 2-7), further supporting previous genetic analyses of flower sex (Fechter et al., 2012, Battilana et al., 2013, Picq et al., 2014). Combining the genetic maps from these three families in alignment with the reference genome provided a higher resolution of the flower sex locus (Figure 2-8).

Table 2-7. Genetic mapping statistics for the flower sex locus using both synteny and *de novo* HetMappS pipelines for three F_1 families.

F1 family (# individuals)	Map Type	LOD Peak	LOD thres-hold[1]	Locus Position (cM)	Nearest markers*	Confidence Interval ± 1.8 LOD (cM)	% of Variation Explained
'Horizon' x Illinois 547-1	*de novo*	96.1	3.26	25.0	S2_4855222	24.0 – 25.8	72.8
	synteny	95.8	3.26	25.5	S2_4855222	24.0 – 27.0	72.7
'Horizon' x *V. cinerea* B9	*de novo*	39.3	3.29	29.0	S2_5125806 S2_5390838	27.0 – 31.0	70.8
	synteny	38.6	3.23	27.0	S2_5125806 S2_5390838	25.1 – 31.0	70.2
'Chardonnay' x *V. cinerea* B9	*de novo*	42.8	3.28	30.4	S2_4745220	28.0 – 31.0	74.8
	synteny	39.7	3.13	26.0	S2_4745220 S2_5121461	25.0 – 30.0	72.2

[1] LOD threshold determined by permutation test (1,000) at $\alpha = 0.05$.
* Single or two flanking markers are provided. Marker name indicates the position on the physical PN40024 genome (version 12X.0) in format S(chromosome)_(position in bp)

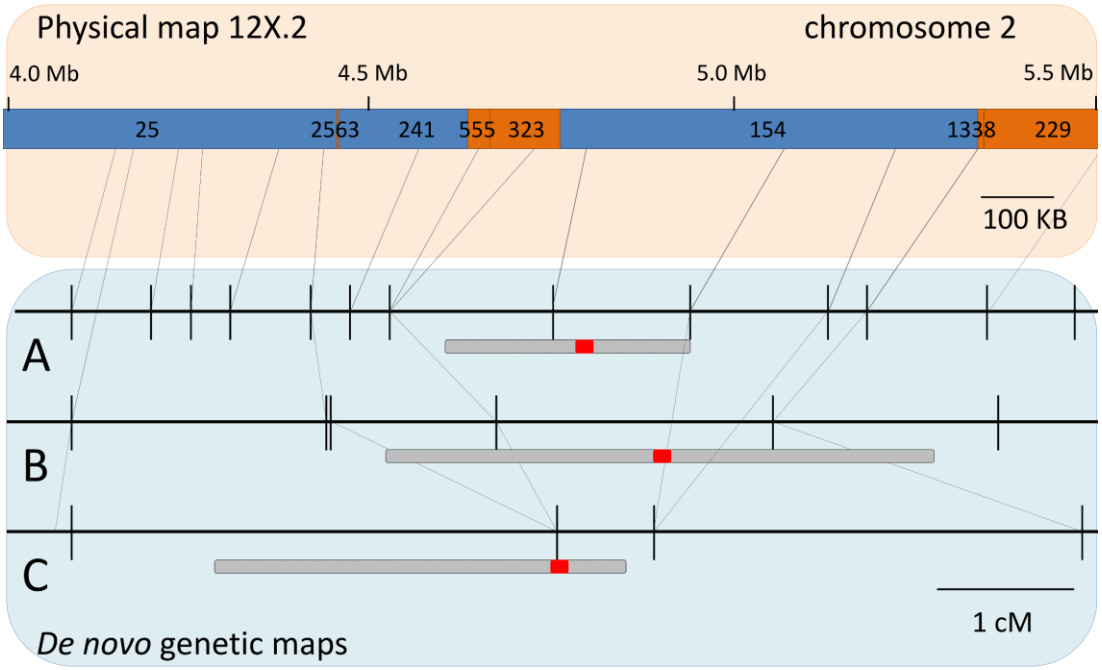


Figure 2-8. Localization of the flower sex locus using *de novo* maps from three families. The PN40024 version 12X.2 reference surrounding the sex locus is shown. Numbers above blue and orange sections indicate scaffold id. Blue scaffolds were located in chromosome 2 and orange scaffolds were located in "unknown" chromosomes in the previous version 12X.0 of PN40024. Connecting lines indicate physical position for SNPs in three *de novo* maps: A) 'Horizon' x Illinois 547-1, B) 'Horizon' x *V. cinerea* B9 and C) 'Chardonnay' x *V. cinerea* B9. For each map, localization of flower sex locus is shown. Shaded areas indicating 1.8 LOD confidence intervals and solid red areas indicate position of the maximum LOD.

Discussion

The modular HetMappS pipeline handles large datasets for heterozygous F_1 families and provides corrections for common GBS problems like large proportion of missing data and heterozygote under-calling. The method does not rely on parental genotypes, although they can be utilized for validation. The choice between synteny and *de novo* linkage group (LG) formation allows the leveraging of existing genome information where available, but also creates a viable option to develop genetic maps for families that include complex hybrid backgrounds and wild species. Unlike previous methods (Gardner et al., 2014, Barba et al., 2014), the *de novo* option can handle a large number of markers for LG creation without the aid of a reference genome. The pipeline is modular, and it can be combined in total or in portion with other filtering or ordering strategies.

Starting from 300,000 to 450,000 markers of variable quality, we were able to create saturated maps with genetic distances comparable to standard SSR-based maps (Adam-Blondon et al., 2004), containing up to 20 times more markers (4,992 and 5,560 SNPs for the synteny and *de* novo pipelines respectively). The number of markers in the final map was highly correlated with the number of individuals in the family both for the synteny pipeline and *de novo* pipeline (r = 0.92 and 0.91, respectively). This suggests that recombination events (hence, the number of progeny genotyped) are currently the limiting factor for the creation of saturated genetic maps in crosses of heterozygous parents.

Here we used the *V. vinifera* reference genome for alignment of tags, representing a reduced section (2.8%) of the physical assembly. This approach of analyzing 64 bp tags generated by *Ape*KI (theoretically cutting every 1024 bp) could be expected to represent 6.2% of the physical assembly. The reduced number of tags could be due to PCR bias, imperfect restriction digestion, methylation, repetitive sequences resulting in multiple alignment of tags, and non-reference sequences. For the four *Vitis*Gen families, alignment rates ranged between 66 to 72% of the tags aligning to a unique position of the reference genome, 11 to 12% were multiply aligned, and between 16 to 22% of the tags were unaligned. Unaligned tags can contain poor quality reads and low complexity sequences, but can also contain divergent or species-specific sequences that are not represented in the reference genome. This unaligned pool of tags could be re-analyzed when additional reference genomes become available that are more relevant to the mapping families. Previously, a genotyping microarray revealed similar but more extreme challenges resulting from the large genetic diversity of Vitis (Myles et al., 2015). The vast majority (over 85%) of SNP microarray markers designed based on *V. vinifera* sequences were not informative in each breeding population. Thus, GBS appears superior to SNP microarrays for high-diversity species and will increase in utility as additional genomes are sequenced.

The distribution of GBS tags across the genome showed a remarkable consistency for the four families (track A in Figure 2-4 and Appendix 2-3, 2-4 and 2-5). Tags were concentrated in the arms of the chromosomes and sparse in centromeric regions, likely due to repetitive sequences in centromeric regions that were multiply aligned and enrichment of methylated sequences

that were not digested by *Ape*KI. Alternative explanations for lower proportion of tags in some regions of the reference genome could be divergence between the wild and hybrid vines with the *V. vinifera* genome. There are also regions with higher proportion of tags (chromosomes 8, 15, 18 and 19) that are consistent among the families. Areas with high mean tag depth (track B in Figure 2-4 and Appendix 2-3, 2-4 and 2-5) may indicate repeated (but uniquely aligning) fragments in the wild and hybrid *Vitis* genomes that are not represented in the *V. vinifera* reference genome.

Family-level quality control is an essential step, given that marker identification is based on progeny segregation patterns, and LG formation and phasing are heavily influenced by the amount of linkage present. Individuals who do not share the same parents will distort segregation ratios and dilute the linkage patterns, thereby reducing resolution.

Because of shallow sequence coverage, the error rate associated with heterozygote under-calling is high, which is problematic both for linkage analysis and for identifying, phasing, and ordering markers. Genotype filtering in HetMappS was tailored to retain the maximum amount of high-quality genotypes and improve the ability to order markers (Lu et al., 2013). The vast majority of SNPs were filtered due to low genotyping rate (51%) and putative sequencing error (32%). Only 13% of the markers were filtered due to non-Pt segregation ratios and just 0.07% for high error rate (ie, homozygous for minor allele with GQ>98). These results are consistent with the expected output for the GBS technique. With only a 4.2% of the initial markers being retained for linkage map creation, our pipeline was effective at reducing the number of markers in a single step, with decisions based largely on SNP quality. On

average, loci had 6.0% of genotypes corrected (sd 10%), and 21.4% of sites masked (sd 31.3%). This could be explained by sequencing error or heterozygote under sampling.

Pt marker coverage correlates well with SNP coverage (r = 0.86, Table 2-1). Some regions had a lower than expected number of Pt markers for one or both parents based on the mean Pt and SNP coverage across windows, although those regions with less than 10% of the mean Pt marker coverage were primarily low coverage regions. Of the 622 regions across 8 parental maps where Pt marker coverage was less than 10% of the expected number, SNP coverage was at most 17% of the mean number of SNPs per 1Mb window, and 98% were under 10%.

Genetic maps generated were consistent between pipelines within populations, as well as maps for the common parent in the half-sib families examined (Figure 2-6). After LG formation, phasing and ordering, the *de novo* pipeline retained on average an extra 7% of markers compared to the synteny pipeline (49.3% and 42.3%, respectively). This can be attributed to markers located on unassembled chromosomes (6.06%) or markers that were assigned to a different chromosome during the alignment step (6.31%), since both of these categories were filtered previous to LG formation in the synteny pipeline. However, the *de novo* pipeline had more apparent problems with initial ordering (Tables 5 and 6). The *de novo* approach should be useful when there are major structural variations relative to the reference genome, as it makes no assumptions about synteny with the reference genome. However, depending on the amount of linkage between markers, the *de novo* pipeline may be more error-prone for LG creation and may require more downstream

filtering. There is greater control associated with the choice of dendrogram cutting options using the *de novo* pipeline, and while more stringent cut parameters ensure fewer problematic markers, it can also result in complete or partial loss of LGs.

Dendrogram cut height in the synteny and *de novo* pipelines may explain subtle differences in the maps obtained for each pipeline. In both pipelines, dendrograms were cut with a static height across each set of Pt markers. In synteny LG formation, markers belonging only to 19 chromosomal groups are identified and separated previous to dendrogram formation, hence 19 dendrograms were cut at the same height to determine parental segregation within each chromosome (Figure 2-3). This two-step process contrasts with the *de novo* pipeline, where all 38 chromosomes need to be separated by a static cut of the same dendrogram (Figure 2-5). In most cases, the expected 38 LGs were not resolved perfectly by the *de novo* pipeline (Table 2-4); some LGs were fused while others split in two, reflecting technical and biological realities. For example, while some chromosomes were only split in one pipeline, chromosome 1 and 7 in the parental maps of 'Horizon' tended to split into two LGs in both the *de novo* pipeline (Table 2-4) and the synteny pipeline (Table 2-3). The gaps observed in some LGs may be related with the evolutionary history of the *Vitis* genome; it is known that the bottom part of chromosome 7 in *V. vinifera* corresponds to the extra chromosome found in the *V. rotundifolia* (Doligez et al., 2006, Blanc et al., 2012).

Correlations between results obtained independently from the synteny and *de novo* pipelines suggests that both methods are robust for map creation, and that the maps generated from the two pipelines correspond well in most cases.

For each family, the total number of SNPs retained could be explained by factors like number of progeny or dendrogram cut height. However, for parental maps within each family, hybrid parents retained a higher proportion of markers (51.3, 48.9, 44.2, and 33.3%), compared with wild species (23.2, 23.8, and 33.5%), and *V. vinifera* 'Chardonnay' was intermediate (39.4%, Table 2-3). This likely reflects the better suitability of the PN40024 reference genome for *V. vinifera* alleles versus wild *Vitis* as well as greater genetic diversity within hybrids. Lack of diversity within *V. vinifera* cultivars has been shown previously (Myles et al., 2011).

Genetic maps often require manual curation as spurious markers can distort genetic distances and marker order. This process cannot be fully automated, as LOD scores for each marker may vary along the LGs, especially at the extremes of chromosome arms, where markers are in weaker linkage than SNPs located at the center of the chromosome (Broman, 2010). With hundreds of markers per LG it is not feasible to examine site by site. In order to standardize the process we followed a five-step procedure for map curation and developed scripts to ease this step.

With average distances of 2,906 and 4,878 cM, non-curated genetic maps generated from synteny or *de novo* pipelines were larger than expected. As reference, parental grapevine SSR maps span between 1,172 and 1,406 cM (Adam-Blondon et al., 2004), while an integrated SSR map of 257 SSR markers in five populations had a total length of 1,485 cM (Doligez et al., 2006). A dense map of 994 SNP loci in a *V. vinifera* cross showed a total length of 1,245 cM (Troggio et al., 2007). Other maps created with RAD or

GBS were between 1,381 and 1,967 cM (Chen et al., 2015, Barba et al., 2014).

Presence of spurious markers was the main cause of map inflation. In the auto drop step, removing an average of 5% and 6% of markers (synteny and *de novo* maps, respectively) led to an average map reduction of 24% and 26% of the initial genetic length. SNPs located at the extremes of the LG were not automatically removed to avoid unnecessary trimming of the maps. Manually removing an extra 7% and 10% of markers (synteny and *de novo* maps, respectively) led to further average reduction of genetic length of 28% and 31%. Prior to these, the ordering step had minimal impact, both in terms of markers removed (3% and 2% in synteny and *de novo* maps, respectively) and genetic distance reduction (4% and 15%). This suggests that the minimization of an associated spanning tree implemented by MSTMap provided a good marker order, even in data sets with high percentage of missing data (up to 50% for these datasets) (Wu et al., 2008).

As expected, final maps distances were not correlated with the number of progeny or with the number of markers in the final synteny or *de novo* maps ($r = -0.13$ and $r = 0.00$, respectively). In contrast, the number of progeny was highly correlated with the number of markers retained for each map ($r = 0.92$ and 0.91, respectively). Total genetic distances in final maps averaged 1,286 cM for 2,433 loci in the synteny pipeline and 1,351 cM for 2,790 loci in the *de novo* pipeline.

Tags located in repeated regions can produce marker artifacts. Identical PCR products originating from repeated regions will collapse into one single tag, representing only one allele in the genotype output. SNPs associated with these may not show distortion at the allele frequency level. In order to identify these markers, it is necessary to consider the frequency of the read depth for the minor allele, as a proxy for the frequency of the minor tag within a locus, contained in the VCF file.

While the minor allele frequency (MAF) in the final maps showed a normal distribution centered near 0.25 (Figure 2-7 A, dotted line), the minor tag frequency (MTF) suggested a bimodal distribution with larger variance (Figure 2-7 A, solid line). The predominant MTF mode peak was centered in 0.25, which is consistent with the observed MAF. The peak for a smaller second mode was located around 0.125, which is consistent with the 1:7 segregation expected for a minor tag located in one of two duplicated regions. This second peak was not observed in the MAF distribution (Figure 2-7 A, dotted line), suggesting that most of the repeats underlying these 1:7 MTFs on the final map were tightly linked and inherited as normal Pt markers.

Markers removed during the map curation process were distributed across all MTFs, with a peak at frequencies lower than 1:7. The MAF distribution (Fig. 7B, dotted line) of these markers had a broad distribution from 0.1 to 0.4 and did not correlate with the MTF distribution (solid line). This result suggests that a proportion of spurious markers may be located in repeated regions that are not represented in the ‘PN40024’ genome. Hence, the associated tags were not removed after the alignment step. Markers located in repeated segments or markers in tandem regions of more than two repeats could inflated genetic

distance calculations due to independent assortment or occasional recombination within the tandem repeats.

We used the flower sex locus to demonstrate the utility of both synteny and *de novo* maps in three populations segregating for the trait. Overall, the localization of markers nearest to the QTL peak marker was between 4.75 and 5.39 Mbp (Table 2-7). In the largest population a SNP located at 4.85 Mbp had the highest LOD score, very close to locations reported by fine mapping and BAC sequencing: between 4.91 and 5.05 Mbp (Fechter et al., 2012) and between 4.89 and 5.04 Mbp (Picq et al., 2014), respectively. Correspondence between the genetic order and physical order of the markers in the sex locus was high (Figure 2-8). In maps generated by the *de novo* pipelines, SNPs located in un-assembled scaffolds from 12x.0 version of the genome were incorporated solely by linkage. A comparison with the newest 12x.2 assembly corroborated the inclusion of scaffolds 2563, 555, 323, 1338 and 229 in chromosome 2. Furthermore, our results suggest that two small additional scaffolds, 1,344 and 1,682, belong to the sex locus. According to gene annotations in version 12X.2 of the PN40024 reference genome (Vitulo et al., 2014), these scaffolds contain three genes: VIT_200s1344g00010 (unknown protein), VIT_200s1682g00020 (cytochrome p450) and VIT_200s1682g00010 (primary amine oxidase). GBS markers associated with scaffold 233 did not map to chromosome 2, even though this scaffold was previously localized in the sex locus region (Fechter et al., 2012, Picq et al., 2014).

Analysis of the sex locus demonstrates the power of our approach to leverage linkage, sequencing, and alignment to the reference genome to quickly characterize a locus. SNPs anchored to the reference genome allowed us to

correlate the physical position with the LOD peak without the need of BAC clones, while the *de novo* pipeline allowed incorporation of scaffolds that were not assembled in the reference genome. Further, divergent sequences discarded during read alignment and SNP calling could be analyzed to leverage more information once the region has been narrowed down.

Many traits of interest in breeding applications are introgressed from wild species. In grapevine this is exemplified by adaptive traits like disease resistance or abiotic stress tolerance. Despite occasional physical gaps in our maps, our results show that it is possible to obtain SNPs across the whole genome in crosses involving wild species. Divergent regions will still be in linkage with SNPs from common loci in a F_1 cross, as linkage usually extends through a long portion of the chromosomes. After curated framework maps are created, additional markers may be added.

The optimal parameters for marker identification, LG formation and phasing, and filtering steps may vary depending not only on the underlying biology of the data set, but also on technical aspects of the dataset, such as genotyping methods, depth of coverage, method of DNA preparation and enzyme choice. For example, the pre-*Vitis*Gen population *V. rupestris* B38 x 'Chardonnay' differed in several key aspects from the other four families. The sequence coverage for this family was higher, and DNA was prepared with a whole genome amplification (WGA) step. The WGA protocol results in a loss of methylation. Because the enzyme used here for genome complexity reduction, *ApeKI*, is a methylation sensitive enzyme, some tags may have targeted inactive transposons, repetitive regions, centromeres, or other regions that are usually methylated. The pre-*Vitis*Gen approach resulted in

more Pt markers, but also more markers with unexpected segregation ratios, an atypical linkage pattern, and split LGs. Because both parents were used for *Vitis*Gen maps, we conclude these complications were more likely due to technique than to biology.

In summary, HetMappS is a publicly-available, modular approach that overcomes the limitations of applying GBS for genetic mapping in highly heterozygous species. The synteny and *de novo* pipelines generate similar maps that match the physical genome with a 5- to 20-fold increase in marker resolution relative to existing grapevine genetic maps. HetMappS can accelerate the discovery of candidate genes underlying traits and enhance the accuracy of genome-wide marker assisted selection in breeding programs, and can be adapted for other applications.

APPENDIX

Appendix 2-1: Sequences and barcodes comprising four 96-plex adapter sets used to generate 384-plex GBS libraries

Adapter Set	Barcode	Top Strand Sequence (5'-3')	Bottom Strand Sequence (5'-3')
ApeKI_ E	AGGC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAGGC	CWGGCCTAGATCGGAAGAGCGT CGTGTAGGGAAAGAGTGT
ApeKI_ E	GATT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGATT	CWGAATCAGATCGGAAGAGCGT CGTGTAGGGAAAGAGTGT
ApeKI_ E	ACCGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTACCGT	CWGACGGTAGATCGGAAGAGCG TCGTGTAGGGAAAGAGTGT
ApeKI_ E	CGTCA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCGTCA	CWGTGACGAGATCGGAAGAGCG TCGTGTAGGGAAAGAGTGT
ApeKI_ E	TCGCA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTCGCA	CWGTGCGAAGATCGGAAGAGCG TCGTGTAGGGAAAGAGTGT
ApeKI_ E	CGCAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCGCAT	CWGATGCGAGATCGGAAGAGCG TCGTGTAGGGAAAGAGTGT
ApeKI_ E	TCATAGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTCATAGT	CWGACTATGAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	TTACGAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTTACGAT	CWGATCGTAAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	GGCTAGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGGCTAGA	CWGTCTAGCCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	ACAATGGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTACAATGGA	CWGTCCATTGTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	ACAAGAGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTACAAGAGT	CWGACTCTTGTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	GAACATGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGAACATGA	CWGTCATGTTCAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	AGCATT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAGCATT	CWGAATGCTAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ E	CTCCGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCTCCGA	CWGTCGGAGAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ E	TTGGCA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTTGGCA	CWGTGCCAAAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ E	CCACGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCCACGT	CWGACGTGGAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ E	GATGTC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGATGTC	CWGGACATCAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ E	TGTTAC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGTTAC	CWGGTAACAAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ E	CAGTTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCAGTTA	CWGTAACTGAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ E	GCCTAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCCTAT	CWGATAGGCAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ E	AGTGGC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAGTGGC	CWGGCCACTAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ E	TGACCT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGACCT	CWGAGGTCAAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ E	TTGCAC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTTGCAC	CWGGTGCAAAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ E	CTAGCT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCTAGCT	CWGAGCTAGAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ E	AATCGTT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAATCGTT	CWGAACGATTAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT

Appendix 2-1 Continued

Adapter Set	Barcode	Top Strand Sequence (5'-3')	Bottom Strand Sequence (5'-3')
ApeKI_ E	CTATGGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCTATGGA	CWGTCCATAGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	TACGGTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTACGGTA	CWGTACCGTAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	ACTATGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTACTATGT	CWGACATAGTAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	CGTGAAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCGTGAAT	CWGATTCACGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	TTGCAGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTTGCAGA	CWGTCTGCAAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	AACTTGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAACTTGT	CWGACAAGTTAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	TGACGTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGACGTA	CWGTACGTCAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	GCTATAA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCTATAA	CWGTTATAGCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	ATCGTAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTATCGTAT	CWGATACGATAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	TACTGAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTACTGAT	CWGATCAGTAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	CTTGAGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCTTGAGA	CWGTCTCAAGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	TCAAGTT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTCAAGTT	CWGAACTTGAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	GATCATA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGATCATA	CWGTATGATCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	GCATTGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCATTGA	CWGTCAATGCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	CAGGTAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCAGGTAT	CWGATACCTGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	TGCAATA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGCAATA	CWGTATTGCAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	ATATCGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTATATCGT	CWGACGATATAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	AGTCTAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAGTCTAT	CWGATAGACTAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	GTCTGAA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTCTGAA	CWGTTCAGACAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	ATCAGTT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTATCAGTT	CWGAACTGATAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	CAGTTGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCAGTTGA	CWGTCAACTGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	TGTGCAA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGTGCAA	CWGTTGCACAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	CGACAGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCGACAGT	CWGACTGTCGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT

Appendix 2-1 Continued

Adapter Set	Barcode	Top Strand Sequence (5'-3')	Bottom Strand Sequence (5'-3')
ApeKI_ E	ACGTGTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTACGTGTA	CWGTACACGTAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	GATGCAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGATGCAT	CWGATGCATCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	CTAATGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCTAATGT	CWGACATTAGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	GTCGATA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTCGATA	CWGTATCGACAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	TATACGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTATACGT	CWGACGTATAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	GCGTAAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCGTAAT	CWGATTACGCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	AGCGTTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAGCGTTA	CWGTAACGCTAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	ATCCGGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTATCCGGA	CWGTCCGGATAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	TCAGTAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTCAGTAT	CWGATACTGAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	CAATGTT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCAATGTT	CWGAACATTGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	GTTACGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTTACGA	CWGTCGTAACAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	TGCATAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGCATAT	CWGATATGCAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	CAAGAAGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCAAGAAGT	CWGACTTCTTGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	GTCATGGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTCATGGT	CWGACCATGACAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	AACAGTGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAACAGTGA	CWGTCACTGTTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	GTGCAAGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTGCAAGA	CWGTCTTGCACAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	CAATAGGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCAATAGGA	CWGTCCTATTGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	TGCAGTGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGCAGTGT	CWGACACTGCAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	AGGCTAGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAGGCTAGA	CWGTCTAGCCTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	CTAGTGGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCTAGTGGT	CWGACCACTAGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	GCTAGTGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCTAGTGT	CWGACACTAGCAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	AGTTGGCA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAGTTGGCA	CWGTGCCAACTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	TCGCAAGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTCGCAAGT	CWGACTTGCGAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT

Appendix 2-1 Continued

Adapter Set	Barcode	Top Strand Sequence (5'-3')	Bottom Strand Sequence (5'-3')
ApeKI_ E	ACGTGTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTACGTGTA	CWGTACACGTAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	GATGCAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGATGCAT	CWGATGCATCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	CTAATGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCTAATGT	CWGACATTAGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	GTCGATA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTCGATA	CWGTATCGACAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	TATACGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTATACGT	CWGACGTATAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	GCGTAAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCGTAAT	CWGATTACGCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	AGCGTTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAGCGTTA	CWGTAACGCTAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	ATCCGGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTATCCGGA	CWGTCCGGATAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	TCAGTAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTCAGTAT	CWGATACTGAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	CAATGTT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCAATGTT	CWGAACATTGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	GTTACGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTTACGA	CWGTCGTAACAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	TGCATAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGCATAT	CWGATATGCAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	CAAGAAGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCAAGAAGT	CWGACTTCTTGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	GTCATGGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTCATGGT	CWGACCATGACAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	AACAGTGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAACAGTGA	CWGTCACTGTTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	GTGCAAGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTGCAAGA	CWGTCTTGCACAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	CAATAGGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCAATAGGA	CWGTCCTATTGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	TGCAGTGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGCAGTGT	CWGACACTGCAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	AGGCTAGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAGGCTAGA	CWGTCTAGCCTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	CTAGTGGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCTAGTGGT	CWGACCACTAGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	GCTAGTGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCTAGTGT	CWGACACTAGCAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	AGTTGGCA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAGTTGGCA	CWGTGCCAACTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	TCGCAAGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTCGCAAGT	CWGACTTGCGAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT

Appendix 2-1 Continued

Adapter Set	Barcode	Top Strand Sequence (5'-3')	Bottom Strand Sequence (5'-3')
ApeKI_ E	CGATGTGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCGATGTGT	CWGACACATCGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	AACGTAGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAACGTAGA	CWGTCTACGTTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	CTCACGGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCTCACGGA	CWGTCCGTGAGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	TAGCGTGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTAGCGTGT	CWGACACGCTAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	ACGTAAGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTACGTAAGA	CWGTCTTACGTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	CGTATGGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCGTATGGT	CWGACCATACGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	GTACGTGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTACGTGT	CWGACACGTACAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	TTCGAAGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTTCGAAGA	CWGTCTTCGAAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	AATACGGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAATACGGA	CWGTCCGTATTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	TGACTGGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGACTGGT	CWGACCAGTCAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	GCGGATGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCGGATGT	CWGACATCCGCAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	CATTGAGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCATTGAGA	CWGTCTCAATGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	GTAACAGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTAACAGA	CWGTCTGTTACAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	AGCTTGGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAGCTTGGT	CWGACCAAGCTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	ACAGATGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTACAGATGA	CWGTCATCTGTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	CAGTTGGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCAGTTGGT	CWGACCAACTGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	TGCAAGAA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGCAAGAA	CWGTTCTTGCAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	ACTCGAGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTACTCGAGA	CWGTCTCGAGTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	GGAGCTGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGGAGCTGT	CWGACAGCTCCAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	CTGAGTGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCTGAGTGT	CWGACACTCAGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	GATCAGAA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGATCAGAA	CWGTTCTGATCAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	TGCATAGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGCATAGA	CWGTCTATGCAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	GAACGAAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGAACGAAT	CWGATTCGTTCAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT

Appendix 2-1 Continued

Adapter Set	Barcode	Top Strand Sequence (5'-3')	Bottom Strand Sequence (5'-3')
ApeKI_ E	TTGGCGGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTTGGCGGA	CWGTCCGCCAAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ E	CGCCGCAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCGCCGCAT	CWGATGCGGCGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	CGAA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCGAA	CWGTTCGAGATCGGAAGAGCGT CGTGTAGGGAAAGAGTGT
ApeKI_ F	GTCC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTCC	CWGGGACAGATCGGAAGAGCGT CGTGTAGGGAAAGAGTGT
ApeKI_ F	ACAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTACAT	CWGATGTAGATCGGAAGAGCGT CGTGTAGGGAAAGAGTGT
ApeKI_ F	GCACT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCACT	CWGAGTGCAGATCGGAAGAGCG TCGTGTAGGGAAAGAGTGT
ApeKI_ F	AGCTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAGCTA	CWGTAGCTAGATCGGAAGAGCG TCGTGTAGGGAAAGAGTGT
ApeKI_ F	CAGCT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCAGCT	CWGAGCTGAGATCGGAAGAGCG TCGTGTAGGGAAAGAGTGT
ApeKI_ F	TATCTGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTATCTGA	CWGTCAGATAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	TTCAAGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTTCAAGT	CWGACTTGAAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	GAGCAGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGAGCAGT	CWGACTGCTCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	TAACGAGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTAACGAGA	CWGTCTCGTTAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	ATACAGGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTATACAGGA	CWGTCCTGTATAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	CACAGAGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCACAGAGT	CWGACTCTGTGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	TAGTGC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTAGTGC	CWGGCACTAAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ F	GTTCCA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTTCCA	CWGTGGAACAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ F	TGAATC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGAATC	CWGGATTCAAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ F	CGCCAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCGCCAT	CWGATGGCGAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ F	ATCGTC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTATCGTC	CWGGACGATAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ F	GTCACT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTCACT	CWGAGTGACAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ F	CATCGC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCATCGC	CWGGCGATGAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ F	AGGTCT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAGGTCT	CWGAGACCTAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ F	CCTGCA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCCTGCA	CWGTGCAGGAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT

Appendix 2-1 Continued

Adapter Set	Barcode	Top Strand Sequence (5'-3')	Bottom Strand Sequence (5'-3')
ApeKI_ F	GTACTC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTACTC	CWGGAGTACAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ F	ACCTGC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTACCTGC	CWGGCAGGTAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ F	CTTGAC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCTTGAC	CWGGTCAAGAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ F	CAGCGTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCAGCGTA	CWGTACGCTGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	AGTTCGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAGTTCGA	CWGTCGAACTAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	GCTAATT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCTAATT	CWGAATTAGCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	ACGCTGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTACGCTGA	CWGTCAGCGTAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	TAAGCTT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTAAGCTT	CWGAAGCTTAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	CTTAGAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCTTAGAT	CWGATCTAAGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	TGATACA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGATACA	CWGTGTATCAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	ACCGAGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTACCGAGT	CWGACTCGGTAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	GTACTTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTACTTA	CWGTAAGTACAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	ATGTCAA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTATGTCAA	CWGTTGACATAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	CACAGGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCACAGGT	CWGACCTGTGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	GGTGGCA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGGTGGCA	CWGTGCCACCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	CTGAATT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCTGAATT	CWGAATTCAGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	TCGTTAA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTCGTTAA	CWGTTAACGAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	TGATCAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGATCAT	CWGATGATCAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	CGTCGGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCGTCGGA	CWGTCCGACGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	GACTATT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGACTATT	CWGAATAGTCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	TAGGTCA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTAGGTCA	CWGTGACCTAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	GCAACGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCAACGT	CWGACGTTGCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	ACTTGAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTACTTGAT	CWGATCAAGTAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT

Appendix 2-1 Continued

Adapter Set	Barcode	Top Strand Sequence (5'-3')	Bottom Strand Sequence (5'-3')
ApeKI_ F	TCTGATA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTCTGATA	CWGTATCAGAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	ATAGTCA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTATAGTCA	CWGTGACTATAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	TAGCCAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTAGCCAT	CWGATGGCTAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	AGGAGTC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAGGAGTC	CWGGACTCCTAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	GTAGAGC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTAGAGC	CWGGCTCTACAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	GACCTAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGACCTAT	CWGATAGGTCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	CATTAGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCATTAGT	CWGACTAATGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	ATTAGCA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTATTAGCA	CWGTGCTAATAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	GAATCTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGAATCTA	CWGTAGATTCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	TGTCATT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGTCATT	CWGAATGACAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	GCCAGAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCCAGAT	CWGATCTGGCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	CCAGGTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCCAGGTA	CWGTACCTGGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	TGGCAAC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGGCAAC	CWGGTTGCCAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	ATGATCT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTATGATCT	CWGAGATCATAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	CATGTTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCATGTTA	CWGTAACATGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	TGTAAGC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGTAAGC	CWGGCTTACAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	ACGTGTGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTACGTGTGT	CWGACACACGTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	CAAGTGAA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCAAGTGAA	CWGTTCACTTGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	GTTAACGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTTAACGA	CWGTCGTTAACAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	TCGTGGAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTCGTGGAT	CWGATCCACGAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	AGCGATAA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAGCGATAA	CWGTTATCGCTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	CCACCAGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCCACCAGT	CWGACTGGTGGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	GGCGTGTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGGCGTGTA	CWGTACACGCCAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT

Appendix 2-1 Continued

Adapter Set	Barcode	Top Strand Sequence (5'-3')	Bottom Strand Sequence (5'-3')
ApeKI_ F	AAGACAGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAAGACAGT	CWGACTGTCTTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	TTGTGCGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTTGTGCGA	CWGTCGCACAAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	GAAGACAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGAAGACAT	CWGATGTCTTCAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	TGTCAAGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGTCAAGA	CWGTCTTGACAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	GCGAGGTT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCGAGGTT	CWGAACCTCGCAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	CGCTTGAA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCGCTTGAA	CWGTTCAAGCGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	ATCGGTGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTATCGGTGT	CWGACACCGATAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	CCTACCGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCCTACCGA	CWGTCGGTAGGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	GGACAATA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGGACAATA	CWGTATTGTCCAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	AACGGAAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAACGGAAT	CWGATTCCGTTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	TCGATGGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTCGATGGT	CWGACCATCGAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	GTTCATTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTTCATTA	CWGTAATGAACAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	GAATCGAA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGAATCGAA	CWGTTCGATTCAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	CATGGTGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCATGGTGT	CWGACACCATGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	TGGCTATT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGGCTATT	CWGAATAGCCAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	ACCACCGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTACCACCGT	CWGACGGTGGTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	CTATGTAA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCTATGTAA	CWGTTACATAGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	AACGAGTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAACGAGTA	CWGTACTCGTTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	CCGACCAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCCGACCAT	CWGATGGTCGGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	TGTCTTAA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGTCTTAA	CWGTTAAGACAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	GTATACTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTATACTA	CWGTAGTATACAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	TGTGCAGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGTGCAGT	CWGACTGCACAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	ATGCTGGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTATGCTGGT	CWGACCAGCATAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT

Appendix 2-1 Continued

Adapter Set	Barcode	Top Strand Sequence (5'-3')	Bottom Strand Sequence (5'-3')
ApeKI_ F	GCAAGTAA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCAAGTAA	CWGTTACTTGCAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	CAGGAATA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCAGGAATA	CWGTATTCCTGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	GTCTGGAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTCTGGAT	CWGATCCAGACAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	CACCACGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCACCACGT	CWGACGTGGTGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	TCATTGTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTCATTGTA	CWGTACAATGAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ F	GGTTCATT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGGTTCATT	CWGAATGAACCAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	GCGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCGA	CWGTCGCAGATCGGAAGAGCGT CGTGTAGGGAAAGAGTGT
ApeKI_ G	TGCT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGCT	CWGAGCAAGATCGGAAGAGCGT CGTGTAGGGAAAGAGTGT
ApeKI_ G	CCTC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCCTC	CWGGAGGAGATCGGAAGAGCGT CGTGTAGGGAAAGAGTGT
ApeKI_ G	TCAGC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTCAGC	CWGGCTGAAGATCGGAAGAGCG TCGTGTAGGGAAAGAGTGT
ApeKI_ G	CATGC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCATGC	CWGGCATGAGATCGGAAGAGCG TCGTGTAGGGAAAGAGTGT
ApeKI_ G	GCTAC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCTAC	CWGGTAGCAGATCGGAAGAGCG TCGTGTAGGGAAAGAGTGT
ApeKI_ G	CTGAC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCTGAC	CWGGTCAGAGATCGGAAGAGCG TCGTGTAGGGAAAGAGTGT
ApeKI_ G	TCTAGGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTCTAGGA	CWGTCCTAGAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	ACACGGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTACACGGT	CWGACCGTGTAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	GACGTGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGACGTGA	CWGTCACGTCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	CGGCAGGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCGGCAGGT	CWGACCTGCCGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	GCGCGTGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCGCGTGA	CWGTCACGCGCAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	ACCGCT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTACCGCT	CWGAGCGGTAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ G	GCTCAC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCTCAC	CWGGTGAGCAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ G	CTATGC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCTATGC	CWGGCATAGAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ G	CAGCCT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCAGCCT	CWGAGGCTGAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ G	GCCATC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCCATC	CWGGATGGCAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT

Appendix 2-1 Continued

Adapter Set	Barcode	Top Strand Sequence (5'-3')	Bottom Strand Sequence (5'-3')
ApeKI_ G	CGCTCA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCGCTCA	CWGTGAGCGAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ G	ACGCTC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTACGCTC	CWGGAGCGTAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ G	CTGATC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCTGATC	CWGGATCAGAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ G	CCAGTC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCCAGTC	CWGGACTGGAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ G	TACGCC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTACGCC	CWGGGCGTAAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ G	CCGTAC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCCGTAC	CWGGTACGGAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ G	TCGACC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTCGACC	CWGGGTCGAAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ G	ACGGCAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTACGGCAT	CWGATGCCGTAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	TTCTTGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTTCTTGA	CWGTCAAGAAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	GGACCTT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGGACCTT	CWGAAGGTCCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	GTGTACA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTGTACA	CWGTGTACACAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	TAGCGGC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTAGCGGC	CWGGCCGCTAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	CGCTTAA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCGCTTAA	CWGTTAAGCGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	GATAGCT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGATAGCT	CWGAGCTATCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	AGAGCCT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAGAGCCT	CWGAGGCTCTAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	TACGAGC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTACGAGC	CWGGCTCGTAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	TTGACTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTTGACTA	CWGTAGTCAAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	CGCCGAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCGCCGAT	CWGATCGGCGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	AATTGGC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAATTGGC	CWGGCCAATTAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	ATACTGC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTATACTGC	CWGGCAGTATAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	GCGGCTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCGGCTA	CWGTAGCCGCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	CGTTATA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCGTTATA	CWGTATAACGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	TTAGACT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTTAGACT	CWGAGTCTAAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT

Appendix 2-1 Continued

Adapter Set	Barcode	Top Strand Sequence (5'-3')	Bottom Strand Sequence (5'-3')
ApeKI_ G	AGGTTAC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAGGTTAC	CWGGTAACCTAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	GGCAACT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGGCAACT	CWGAGTTGCCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	TCACCGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTCACCGA	CWGTCGGTGAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	AAGGTTC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAAGGTTC	CWGGAACCTTAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	CTAGCAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCTAGCAT	CWGATGCTAGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	GCCGACC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCCGACC	CWGGGTCGGCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	AGTACTT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAGTACTT	CWGAAGTACTAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	TAATTGC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTAATTGC	CWGGCAATTAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	GCTGGAC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCTGGAC	CWGGTCCAGCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	GTCATCA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTCATCA	CWGTGATGACAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	CAGTGAC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCAGTGAC	CWGGTCACTGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	GCACACT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCACACT	CWGAGTGTGCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	AAGTCTT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAAGTCTT	CWGAAGACTTAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	TGTATCA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGTATCA	CWGTGATACAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	CTGGAAC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCTGGAAC	CWGGTTCCAGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	ATTCGAC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTATTCGAC	CWGGTCGAATAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	CAAGATC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCAAGATC	CWGGATCTTGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	TCGCGCA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTCGCGCA	CWGTGCGCGAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	AGTGTGC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAGTGTGC	CWGGCACACTAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	GAATTCT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGAATTCT	CWGAGAATTCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	AAGATCGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAAGATCGA	CWGTCGATCTTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	CGAAGAAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCGAAGAAT	CWGATTCTTCGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	TCTGATTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTCTGATTA	CWGTAATCAGAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT

Appendix 2-1 Continued

Adapter Set	Barcode	Top Strand Sequence (5'-3')	Bottom Strand Sequence (5'-3')
ApeKI_ G	ATCTCTGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTATCTCTGA	CWGTCAGAGATAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	AAGCAGAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAAGCAGAT	CWGATCTGCTTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	TGCGGATT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGCGGATT	CWGAATCCGCAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	CTCCTCGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCTCCTCGA	CWGTCGAGGAGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	GGTACTAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGGTACTAT	CWGATAGTACCAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	ACAGGATA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTACAGGATA	CWGTATCCTGTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	GATCTGGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGATCTGGT	CWGACCAGATCAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	CTGAAGAA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCTGAAGAA	CWGTTCTTCAGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	TCCGCCAA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTCCGCCAA	CWGTTGGCGGAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	ATATTCGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTATATTCGT	CWGACGAATATAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	CGGTGATT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCGGTGATT	CWGAATCACCGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	TATGCTGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTATGCTGA	CWGTCAGCATAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	GAACAGTT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGAACAGTT	CWGAACTGTTCAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	ACGATTAA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTACGATTAA	CWGTTAATCGTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	GGTTACGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGGTTACGT	CWGACGTAACCAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	TTACGGTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTTACGGTA	CWGTACCGTAAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	CACGCCAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCACGCCAT	CWGATGGCGTGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	TCGAGATA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTCGAGATA	CWGTATCTCGAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	AGAACTGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAGAACTGA	CWGTCAGTTCTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	GTTCTAAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTTCTAAT	CWGATTAGAACAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	CCTTAGAA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCCTTAGAA	CWGTTCTAAGGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	GACTGTGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGACTGTGT	CWGACACAGTCAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	CGCGCCTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCGCGCCTA	CWGTAGGCGCGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT

Appendix 2-1 Continued

Adapter Set	Barcode	Top Strand Sequence (5'-3')	Bottom Strand Sequence (5'-3')
ApeKI_ G	TTGCATAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTTGCATAT	CWGATATGCAAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	GCTATAGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCTATAGA	CWGTCTATAGCAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	ACAGCCTT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTACAGCCTT	CWGAAGGCTGTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	CTGTGGTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCTGTGGTA	CWGTACCACAGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	TGCTTAAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGCTTAAT	CWGATTAAGCAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	AATGACGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAATGACGA	CWGTCGTCATTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	AGAACGAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAGAACGAT	CWGATCGTTCTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	GTGCGATT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTGCGATT	CWGAATCGCACAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	CTAATTGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCTAATTGA	CWGTCAATTAGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ G	TAGCAGTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTAGCAGTA	CWGTACTGCTAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	AACA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAACA	CWGTGTTAGATCGGAAGAGCGT CGTGTAGGGAAAGAGTGT
ApeKI_ H	CTGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCTGT	CWGACAGAGATCGGAAGAGCGT CGTGTAGGGAAAGAGTGT
ApeKI_ H	TAAC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTAAC	CWGGTTAAGATCGGAAGAGCGT CGTGTAGGGAAAGAGTGT
ApeKI_ H	CTCGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCTCGA	CWGTCGAGAGATCGGAAGAGCG TCGTGTAGGGAAAGAGTGT
ApeKI_ H	GACTC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGACTC	CWGGAGTCAGATCGGAAGAGCG TCGTGTAGGGAAAGAGTGT
ApeKI_ H	GCATCC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCATCC	CWGGGATGCAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ H	CGTACC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCGTACC	CWGGGTACGAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ H	ATTCAGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTATTCAGT	CWGACTGAATAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	TATGTAC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTATGTAC	CWGGTACATAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	AGACAAGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAGACAAGT	CWGACTTGTCTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	GGCCGAGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGGCCGAGT	CWGACTCGGCCAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	AACAAGGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAACAAGGT	CWGACCTTGTTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	ACGTC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTACGTC	CWGGACGTAGATCGGAAGAGCG TCGTGTAGGGAAAGAGTGT

Appendix 2-1 Continued

Adapter Set	Barcode	Top Strand Sequence (5'-3')	Bottom Strand Sequence (5'-3')
ApeKI_ H	CGATC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCGATC	CWGGATCGAGATCGGAAGAGCG TCGTGTAGGGAAAGAGTGT
ApeKI_ H	ATGCGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTATGCGT	CWGACGCATAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ H	TCCAGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTCCAGT	CWGACTGGAAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ H	GCTTGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCTTGA	CWGTCAAGCAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ H	TCCGTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTCCGTA	CWGTACGGAAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ H	TCGGAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTCGGAT	CWGATCCGAAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ H	GTCTTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTCTTA	CWGTAAGACAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ H	CGGAGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCGGAGT	CWGACTCCGAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ H	CGTCTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCGTCTA	CWGTAGACGAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ H	CACGTT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCACGTT	CWGAACGTGAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ H	GTTAGC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTTAGC	CWGGCTAACAGATCGGAAGAGC GTCGTGTAGGGAAAGAGTGT
ApeKI_ H	ATGACGC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTATGACGC	CWGGCGTCATAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	TATTGCA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTATTGCA	CWGTGCAATAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	GGATATC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGGATATC	CWGGATATCCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	CCGAGCT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCCGAGCT	CWGAGCTCGGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	TTGATAC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTTGATAC	CWGGTATCAAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	CACGACT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCACGACT	CWGAGTCGTGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	ATCTAGC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTATCTAGC	CWGGCTAGATAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	TGAATTC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGAATTC	CWGGAATTCAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	CGGCTCA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCGGCTCA	CWGTGAGCCGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	GGTACAC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGGTACAC	CWGGTGTACCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	AATGTCT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAATGTCT	CWGAGACATTAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	TCATGAC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTCATGAC	CWGGTCATGAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT

Appendix 2-1 Continued

Adapter Set	Barcode	Top Strand Sequence (5'-3')	Bottom Strand Sequence (5'-3')
ApeKI_ H	TTCGGAC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTTCGGAC	CWGGTCCGAAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	GCTTAGC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCTTAGC	CWGGCTAAGCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	ACTGCCA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTACTGCCA	CWGTGGCAGTAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	TAGTACT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTAGTACT	CWGAGTACTAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	GACACTC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGACACTC	CWGGAGTGTCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	CGAGTAC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCGAGTAC	CWGGTACTCGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	ATATGTC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTATATGTC	CWGGACATATAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	TGGACCT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGGACCT	CWGAGGTCCAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	CATATGC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCATATGC	CWGGCATATGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	TTCGCCA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTTCGCCA	CWGTGGCGAAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	AGTTACT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAGTTACT	CWGAGTAACTAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	GTGCCAC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTGCCAC	CWGGTGGCACAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	ACCGGTC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTACCGGTC	CWGGACCGGTAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	CGGTAGC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCGGTAGC	CWGGCTACCGAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	GCAGTTC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCAGTTC	CWGGAACTGCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	TATAGTC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTATAGTC	CWGGACTATAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	ACTCCGC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTACTCCGC	CWGGCGGAGTAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	GTATTAC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTATTAC	CWGGTAATACAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	TGAGCGC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGAGCGC	CWGGCGCTCAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	GAGTCGC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGAGTCGC	CWGGCGACTCAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	ATTGATC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTATTGATC	CWGGATCAATAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	TCCACGC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTCCACGC	CWGGCGTGGAAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	GTTATTC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTTATTC	CWGGAATAACAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT

Appendix 2-1 Continued

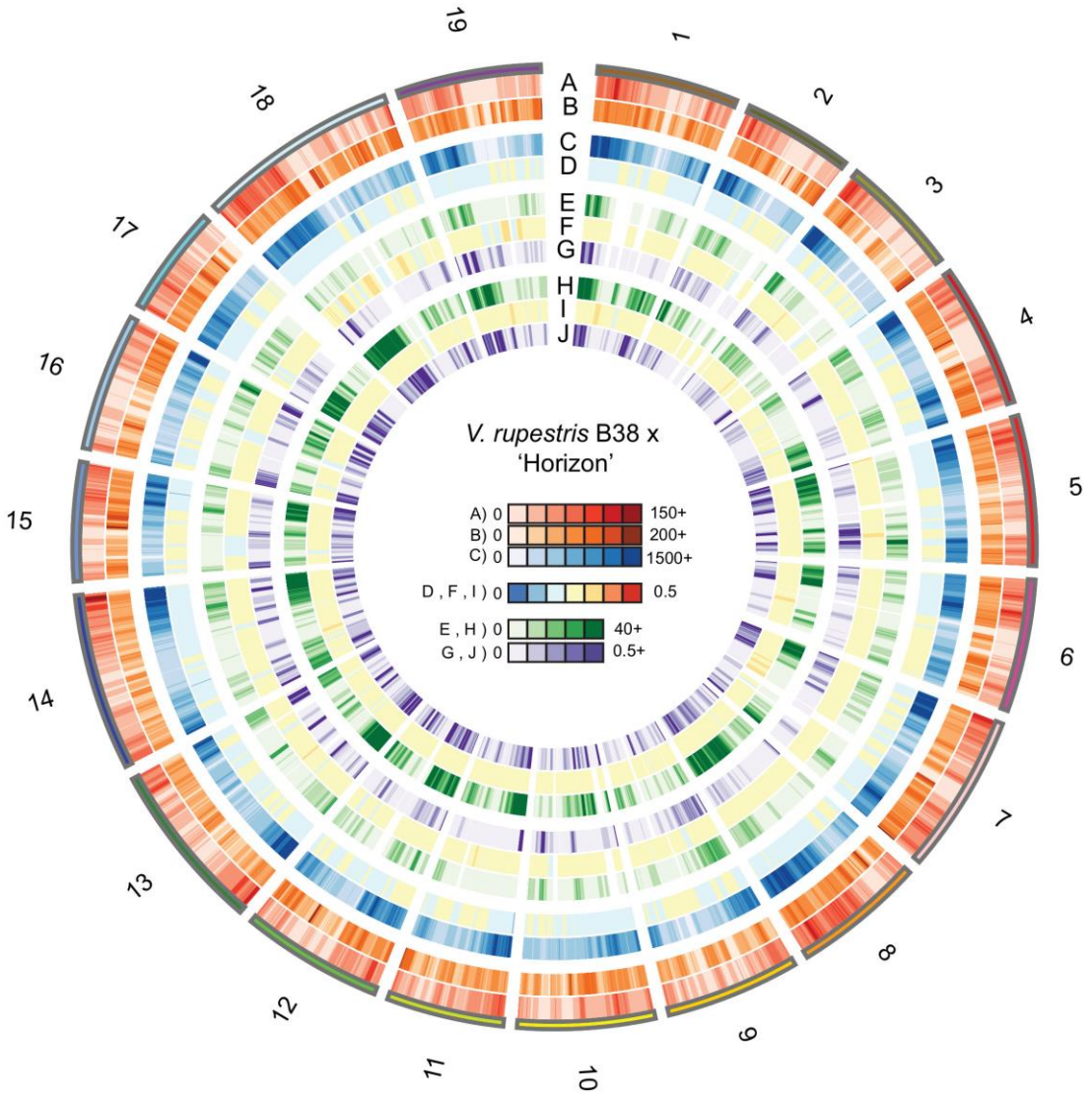
Adapter Set	Barcode	Top Strand Sequence (5'-3')	Bottom Strand Sequence (5'-3')
ApeKI_ H	ATGCACC	ACACTCTTTCCCTACACGACG CTCTTCCGATCTATGCACC	CWGGGTGCATAGATCGGAAGAG CGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	TCCGATAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTCCGATAT	CWGATATCGGAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	GCCTCCGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCCTCCGA	CWGTCGGAGGCAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	AGTCGTGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAGTCGTGT	CWGACACGACTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	CATGTATT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCATGTATT	CWGAATACATGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	TGAAGCAA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTGAAGCAA	CWGTTGCTTCAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	GCATTGGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCATTGGT	CWGACCAATGCAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	AAGCGATA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAAGCGATA	CWGTATCGCTTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	GTCAATAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTCAATAT	CWGATATTGACAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	CGGCCGTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCGGCCGTA	CWGTACGGCCGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	TTAGTCGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTTAGTCGA	CWGTCGACTAAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	GACTCAAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGACTCAAT	CWGATTGAGTCAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	ACTAGGAA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTACTAGGAA	CWGTTCCTAGTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	CGCAACTT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCGCAACTT	CWGAAGTTGCGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	TCGGTTGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTCGGTTGA	CWGTCAACCGAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	AATTCTGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAATTCTGT	CWGACAGAATTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	CTAGAGTA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCTAGAGTA	CWGTACTCTAGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	GCGCGCAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCGCGCAT	CWGATGCGCGCAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	CGATTAGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCGATTAGA	CWGTCTAATCGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	ATTCAGTT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTATTCAGTT	CWGAACTGAATAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	TTGACAAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTTGACAAT	CWGATTGTCAAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	TACGTTAA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTACGTTAA	CWGTTAACGTAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	AGATACGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAGATACGA	CWGTCGTATCTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT

Appendix 2-1 Continued

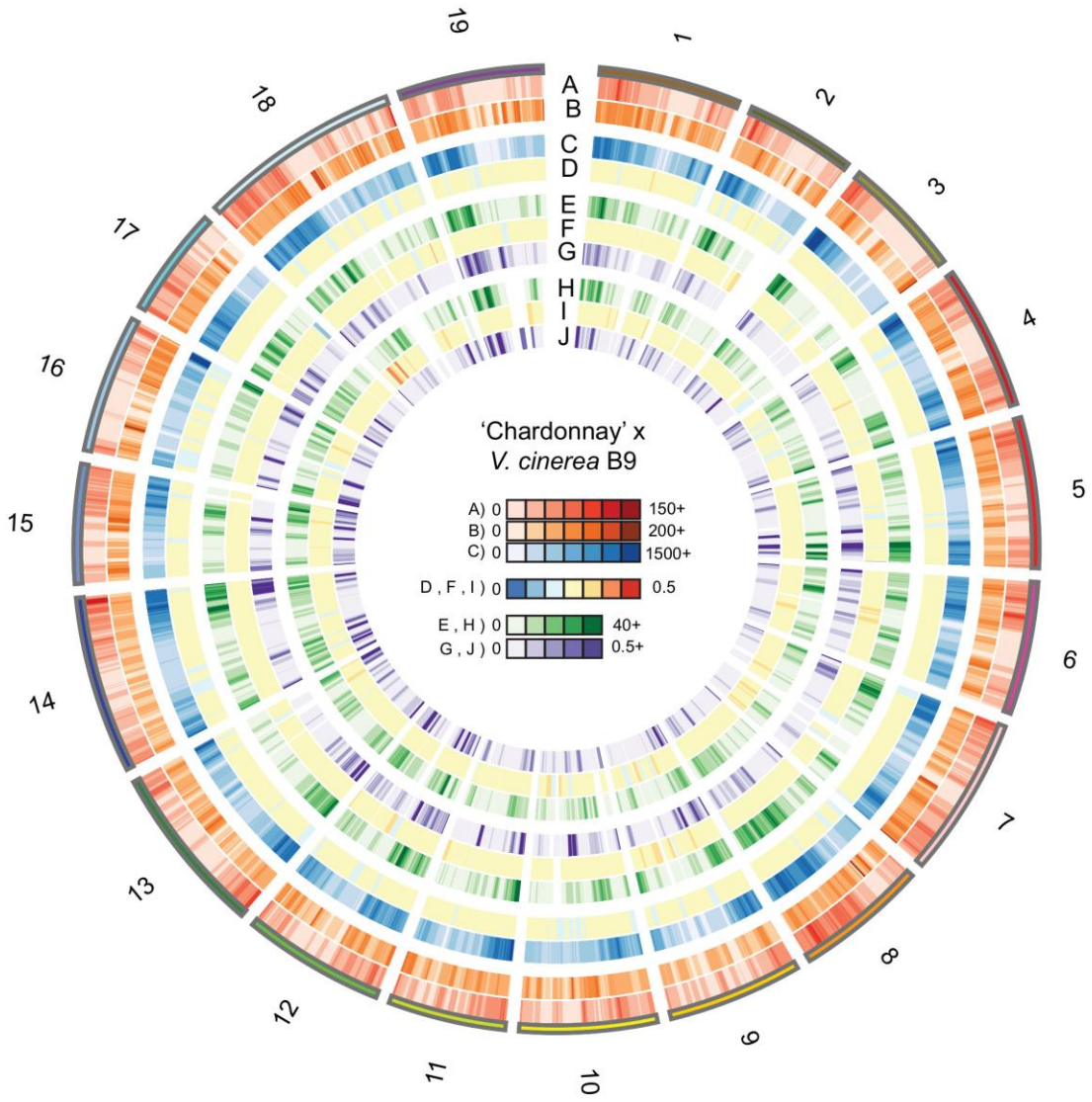
Adapter Set	Barcode	Top Strand Sequence (5'-3')	Bottom Strand Sequence (5'-3')
ApeKI_ H	GCTGGATT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGCTGGATT	CWGAATCCAGCAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	GACACCTT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGACACCTT	CWGAAGGTGTCAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	CCGCCTAA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCCGCCTAA	CWGTTAGGCGGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	TTCTTGGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTTCTTGGA	CWGTCCAAGAAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	CAACGCTT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCAACGCTT	CWGAAGCGTTGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	AGCAGATA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAGCAGATA	CWGTATCTGCTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	TCTTAGGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTCTTAGGT	CWGACCTAAGAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	GTGGCTAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTGGCTAT	CWGATAGCCACAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	AGACTGAA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAGACTGAA	CWGTTCAGTCTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	CATAATGA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCATAATGA	CWGTCATTATGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	GTGTCAGT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTGTGTCAGT	CWGACTGACACAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	AGTGGCTT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTAGTGGCTT	CWGAAGCCACTAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	CCACTGCA	ACACTCTTTCCCTACACGACG CTCTTCCGATCTCCACTGCA	CWGTGCAGTGGAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_ H	TACAGGAT	ACACTCTTTCCCTACACGACG CTCTTCCGATCTTACAGGAT	CWGATCCTGTAAGATCGGAAGA GCGTCGTGTAGGGAAAGAGTGT
ApeKI_C ommon	None	CWGAGATCGGAAGAGCGGTT CAGCAGGAATGCCGAG	CTCGGCATTCCTGCTGAACCGCT CTTCCGATCT

Appendix 2-2: TASSEL-GBS plugins (Glaubitz et al., 2014) used in HetMappS

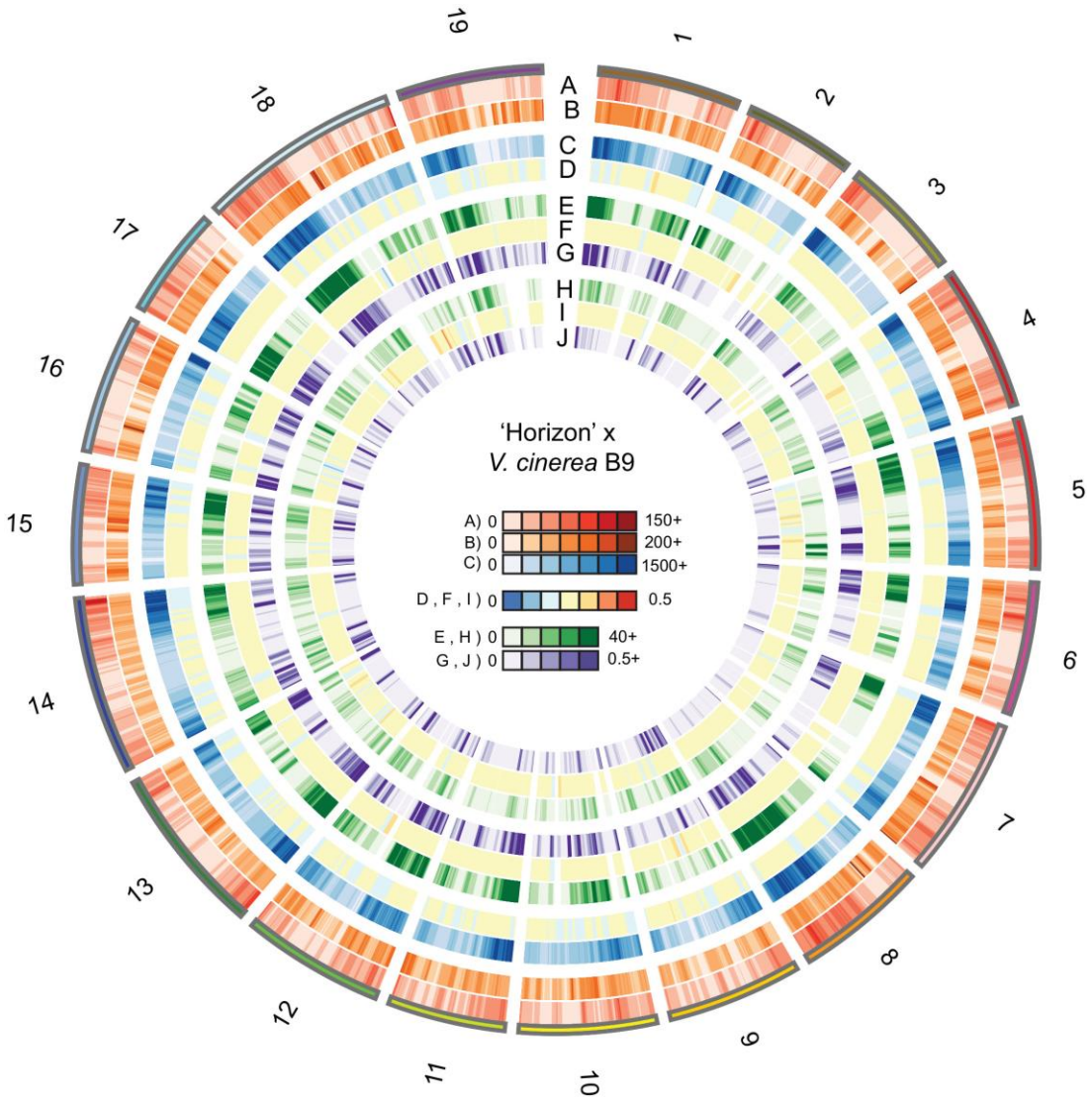
Tag generation Plugin	**Option**	**Value**	**Description**
FastqToTagCountPlugin	-s	300,000,000	Maximum number of good reads per lane. Default: 200000000
FastqToTagCountPlugin	-c	1	Minimum number of times a tag must be present to be output. Default: 1
MergeMultipleTagCount Plugin	-c	5	Minimum number of times a tag must be present to be output. Default: 1
Tags by taxa Plugin	**Option**	**Value**	**Description**
FastqToTBTPlugin	-c	5	Minimum taxa count within a fastq file for a tag to be output. Default: 1
FastqToTBTPlugin	-s	300,000,000	Max good reads per lane. (Optional. Default is 200,000,000).
FastqToTBTPlugin	-y	-y	Output in TBTByte format (counts from 0-127) instead of TBTBit (0 or 1).
MergeTagsByTaxaFiles Plugin	-s	300,000,000	Maximum number of tags the TBT can hold while merging (default: 200,000,000). Reduce this only if you run out of memory (omit the commas).
MergeTagsByTaxaFiles Plugin	-x	-x	Merges tag counts of taxa with identical names if set to -x. Not performed by default
SNP calling Plugin	**Option**	**Value**	**Description**
tbt2vcfPlugin	-ak	4	Maximum number of alleles that are kept for each marker across the population; default: 3
tbt2vcfPlugin	-mnMAF	0	Minimum minor allele frequency (default: 0.0)
tbt2vcfPlugin	-mnLCov	0	Minimum locus coverage (proportion of Taxa with a genotype) (default: 0.0)
MergeDuplicateSNP_vcf_Plugin	-ak	4	Maximum number of alleles that are kept for each marker across the population; default: 3



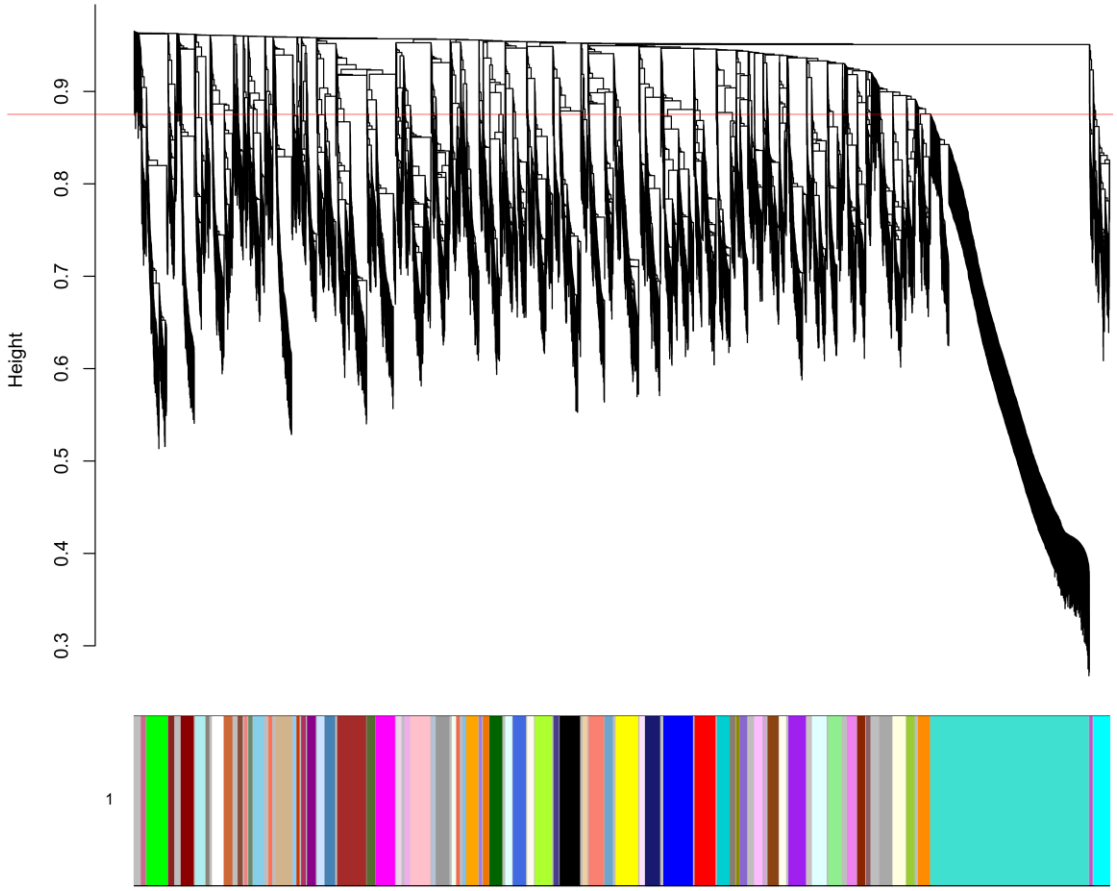
Appendix 2-3: Visualization of *V. rupestris* B38 x 'Horizon' genomic data. Data are shown on 1 Mb windows with a 100 Kb slide. A) Number of unique tags aligned, B) Mean tag depth calculated as total tag depth over number of unique tags aligned, C) Density of SNPs entering the HetMappS pipeline, D) Minor allele frequency (MAF) of SNPs entering the pipeline, E-J) SNP output from the synteny pipeline: E) SNP density 'Horizon', F) MAF 'Horizon' SNPs, G) Recombination frequency 'Horizon', calculated as the number of obligate crossovers per progeny per Mb, H) SNP density Illinois 547-1, I) MAF Illinois 547-1 SNPs, J) recombination frequency Illinois 547-1, calculated as the number of obligate crossovers per progeny per Mb.



Appendix 2-4: Visualization of 'Chardonnay' x *V. cinerea* B9 genomic data. Data are shown on 1 Mb windows with a 100 Kb slide. A) Number of unique tags aligned, B) Mean tag depth calculated as total tag depth over number of unique tags aligned, C) Density of SNPs entering the HetMappS pipeline, D) Minor allele frequency (MAF) of SNPs entering the pipeline, E-J) SNP output from the synteny pipeline: E) SNP density 'Horizon', F) MAF 'Horizon' SNPs, G) Recombination frequency 'Horizon', calculated as the number of obligate crossovers per progeny per Mb, H) SNP density Illinois 547-1, I) MAF Illinois 547-1 SNPs, J) recombination frequency Illinois 547-1, calculated as the number of obligate crossovers per progeny per Mb.



Appendix 2-5: Visualization of 'Horizon' x *V. cinerea* B9 genomic data. Data are shown on 1 Mb windows with a 100 Kb slide. A) Number of unique tags aligned, B) Mean tag depth calculated as total tag depth over number of unique tags aligned, C) Density of SNPs entering the HetMappS pipeline, D) Minor allele frequency (MAF) of SNPs entering the pipeline, E-J) SNP output from the synteny pipeline: E) SNP density 'Horizon', F) MAF 'Horizon' SNPs, G) Recombination frequency 'Horizon', calculated as the number of obligate crossovers per progeny per Mb, H) SNP density Illinois 547-1, I) MAF Illinois 547-1 SNPs, J) recombination frequency Illinois 547-1, calculated as the number of obligate crossovers per progeny per Mb.



Appendix 2-6: Atypical linkage pattern in pre-*Vitis*Gen family *V. rupestris* B38 x 'Chardonnay'. Dendrogram created from hierarchical clustering of topological overlap matrix for SNPs derived from whole genome amplified DNA of the *V. rupestris* B38 x 'Chardonnay' F_1 family, displaying an atypical linkage pattern. This dendrogram was cut at 0.875 height and 5 linkage groups were discarded before proceeding to the phasing step.

REFERENCES

Adam-Blondon, A.-F., Jaillon, O., Vezzulli, S., Zharkikh, A., Troggio, M., Velasco, R., (2011a). Genome sequence initiatives. In: A-F, A.-B., JM, M.-Z., Kole, C. (Eds.), Genetics, genomics and breeding of grapes. Science Publishers.

Adam-Blondon, A.F., Boulay, M., Martinez-Zapater, J.M., (2011b). Future Prospects. In: Adam-Blondon, A.F., Martinez-Zapater, J.M., Kole, C. (Eds.), Genetics, Genomics and breeding of grapes. Science Publishers, Enfield, New Hampshire.

Adam-Blondon, A.F., Roux, C., Claux, D., Butterlin, G., Merdinoglu, D., This, P., (2004). Mapping 245 SSR markers on the *Vitis vinifera* genome: a tool for grape genetics. Theor. Appl. Genet. (109): 1017-1027.

Antcliff, A.J., (1980). Inheritance of sex in *Vitis*. Annales de l'Amelioration des Plantes (30): 113-122.

Barba, P., Cadle-Davidson, L., Harriman, J., Glaubitz, J., Brooks, S., Hyma, K., Reisch, B., (2014). Grapevine powdery mildew resistance and susceptibility loci identified on a high-resolution SNP map. Theor Appl Genet (127): 73-84.

Barnaud, A., Laucou, V., This, P., Lacombe, T., Doligez, A., (2010). Linkage disequilibrium in wild French grapevine, *Vitis vinifera* L. subsp. *silvestris*. Heredity (104): 431-437.

Battilana, J., Lorenzi, S., Moreira, F., Moreno-Sanz, P., Failla, O., Emanuelli, F., Grando, M.S., (2013). Linkage Mapping and Molecular Diversity at the Flower Sex Locus in Wild and Cultivated Grapevine Reveal a Prominent SSR Haplotype in Hermaphrodite Plants. Mol Biotechnol (54): 1031-1037.

Blanc, S., Wiedemann-Merdinoglu, S., Dumas, V., Mestre, P., Merdinoglu, D., (2012). A reference genetic map of *Muscadinia rotundifolia* and identification of Ren5, a new major locus for resistance to grapevine powdery mildew. Theor Appl Genet (125) 1663-1675.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., Buckler, E.S., (2007). TASSEL: software for association mapping of complex traits in diverse samples. Bioinf (23): 2633-2635.

Broman, K.W., (2010). Genetic map construction with R/qtl. 2010 November [cited 11 March 2015] in R/qtl [Internet] Available: http://www.rqtl.org/tutorials/geneticmaps.pdf.

Broman, K.W., Wu, H., Sen, Ś., Churchill, G.A., (2003). R/qtl: QTL mapping in experimental crosses. Bioinf (19): 889-890.

Browning, S.R., Browning, B.L., (2011). Haplotype phasing: existing methods and new developments. Nat Rev Genet (12): 703-714.

Chen, J., Wang, N., Fang, L.-C., Liang, Z.-C., Li, S.-H., Wu, B.-H., (2015). Construction of a high-density genetic map and QTLs mapping for sugars and acids in grape berries. BMC Plant Bio (15): 1-14.

Dalbó, M., Ye, G., Weeden, N., Steinkellner, H., Sefc, K., Bl, R., (2000). A gene controlling sex in grapevines placed on a molecular marker-based genetic map. Genome (43): 333-340.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., Group, G.P.A., (2011). The variant call format and VCFtools. Bioinf (27): 2156-2158.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., Blaxter, M.L., (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet (12): 499-510.

Di Gaspero, G., Peterlunger, E., Testolin, R., Edwards, K.J., Cipriani, G., (2000). Conservation of microsatellite loci within the genus Vitis. Theor Appl Genet (101): 301-308.

Doligez, A., Adam-Blondon, A., Cipriani, G., Di Gaspero, G., Laucou, V., Merdinoglu, D., Meredith, C., Riaz, S., Roux, C., This, P., (2006). An integrated SSR map of grapevine based on five mapping populations. Theor Appl Genet (113): 369-382.

Duchêne, E., Butterlin, G., Dumas, V., Merdinoglu, D., (2012). Towards the adaptation of grapevine varieties to climate change: QTLs and candidate genes for developmental stages. Theor. Appl. Genet. (124): 623-635.

Elshire, R., Glaubitz, J., Sun, Q., Poland, J., Kawamoto, K., Buckler, E., Mitchell, S., (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE (6): e19379.

Etter, P., Bassham, S., Hohenlohe, P., Johnson, E., Cresko, W., (2011). SNP discovery and genotyping for evolutionary genetics using RAD sequencing. In: Orgogozo, V., Rockman, M.V. (Eds.), Molecular Methods for Evolutionary Genetics. Humana Press, pp. 157-178.

Fechter, I., Hausmann, L., Daum, M., Rosleff Sörensen, T., Viehöver, P., Weisshaar, B., Töpfer, R., (2012). Candidate genes within a 143 kb region of the flower sex locus in *Vitis*. Mol Genet Genomics (287): 247-259.

Fernandez, L., Doligez, A., Lopez, G., Thomas, M.R., Bouquet, A., Torregrosa, L., (2006). Somatic chimerism, genetic inheritance, and mapping of the fleshless berry (flb) mutation in grapevine (*Vitis vinifera* L.). Genome (49): 721-728.

Gardner, K.M., Brown, P., Cooke, T.F., Cann, S., Costa, F., Bustamante, C., Velasco, R., Troggio, M., Myles, S., (2014). Fast and cost-effective genetic mapping in apple using Next-Generation Sequencing. G3 (Bethesda) (4): 1681-1687.

Glaubitz, J.C., Casstevens, T.M., Lu, F., Harriman, J., Elshire, R.J., Sun, Q., Buckler, E.S., (2014). TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. PLoS ONE (9): e90346.

Grattapaglia, D., Sederoff, R., 1994. Genetic Linkage Maps of Eucalyptus grandis and Eucalyptus urophylla Using a Pseudo-Testcross: Mapping Strategy and RAPD Markers. Genetics (137): 1121-1137.

Hart, J.P., Griffiths, P.D., (2015). Genotyping-by-Sequencing enabled mapping and marker development for the By-2 potyvirus resistance allele in common bean. Plant Gen. doi:10.3835/plantgenome2014.09.0058

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., Billault, A., Segurens, B., Gouyvenoux, M., Ugarte, E., Cattonaro, F., Anthouard, V., Vico, V., Fabbro, C.D., Alaux, M., Gaspero, G.D., Dumas, V., Felice, N., Paillard, S., Juman, I., Moroldo, M., Scalabrin, S., Canaguier, A., Clainche, I.L., Malacrida, G., Durand, E., Pesole, G., Laucou, V., Chatelet, P., Merdinoglu, D., Delledonne, M., Pezzotti, M., Lecharny, A., Scarpelli, C., Artiguenave, F., Pè, M.E., Valle, G., Morgante, M., Caboche, M., Adam-

Blondon, A.-F., Weissenbach, J., Quétier, F., Wincker, P., (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature (449): 463-467.

Jansen, J., de Jong, A.G., van Ooijen, J.W., (2001). Constructing dense genetic linkage maps. Theor. Appl. Genet. (102): 1113-1122.

Krzywinski, M., Schein, J., Birol, İ., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., Marra, M.A., (2009). Circos: An information aesthetic for comparative genomics. Genome Res (19): 1639-1645.

Langfelder, P., Horvath, S., (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinf (9): 559.

Li, H., (2011). Tabix: fast retrieval of sequence features from generic TAB-delimited files. Bioinf (27): 718-719.

Li, H., Durbin, R., (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinf (25): 1754-1760.

Lu, F., Lipka, A.E., Glaubitz, J., Elshire, R., Cherney, J.H., Casler, M.D., Buckler, E.S., Costich, D.E., (2013). Switchgrass genomic diversity, ploidy, and evolution: Novel insights from a network-based SNP discovery protocol. PLoS Genet (9): e1003215.

Mahanil, S., Ramming, D.W., Cadle-Davidson, M., Owens, C.L., Garris, A., Myles, S., Cadle-Davidson, L., (2012). Development of marker sets useful in the early selection of *Ren4* powdery mildew resistance and seedlessness for table and raisin grape breeding. Theor Appl Genet (124): 23-33.

Marguerit, E., Boury, C., Manicki, A., Donnart, M., Butterlin, G., Némorin, A., Wiedemann-Merdinoglu, S., Merdinoglu, D., Ollat, N., Decroocq, S., (2009). Genetic dissection of sex determinism, inflorescence morphology and downy mildew resistance in grapevine. Theor Appl Genet (118): 1261-1278.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., ((2010)). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. (20): 1297-1303.

Mejia, N., Gebauer, M., Munoz, L., Hewstone, N., Munoz, C., Hinrichsen, P., (2007). Identification of QTLs for seedlessness, berry size, and ripening date

in a. seedless x seedless table grape progeny. Am. J. Enol. Vitic (58), 499-507.

Miller, A.J., Matasci, N., Schwaninger, H., Aradhya, M.K., Prins, B., Zhong, G.-Y., Simon, C., Buckler, E.S., Myles, S., (2013). *Vitis* Phylogenomics: hybridization intensities from a SNP array outperform genotype calls. PLoS ONE (8): e78680.

Myles, S., Boyko, A.R., Owens, C.L., Brown, P.J., Grassi, F., Aradhya, M.K., Prins, B., Reynolds, A., Chia, J.-M., Ware, D., Bustamante, C.D., Buckler, E.S., (2011). Genetic structure and domestication history of the grape. Proc Natl Acad Sci USA (108): 3530-3535.

Myles, S., Mahanil, S., Harriman, J., Gardner, K., Franklin, J., Reisch, B., Ramming, D., Owens, C., Li, L., Buckler, E., Cadle-Davidson, L., (2015). Genetic mapping in grapevine using SNP microarray intensity values. Mol Breeding (35): 1-12.

Picq, S., Santoni, S., Lacombe, T., Latreille, M., Weber, A., Ardisson, M., Ivorra, S., Maghradze, D., Arroyo-Garcia, R., Chatelet, P., This, P., Terral, J.F., Bacilieri, R., (2014). A small XY chromosomal region explains sex determination in wild dioecious *V. vinifera* and the reversal to hermaphroditism in domesticated grapevines. BMC Plant Bio (14): 17.

Poland, J.A., Brown, P.J., Sorrells, M.E., Jannink, J.-L., (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE (7): e32253.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., Sham, P.C., (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet (81): 559-575.

Rafalski, A., (2002). Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol (5): 94-100.

Riaz, S., Krivanek, A.F., Xu, K., Walker, M.A., (2006). Refined mapping of the Pierce's disease resistance locus, PdR1, and Sex on an extended genetic map of *Vitis rupestris* x *V arizonica*. Theor Appl Genet (113): 1317-1329.

Spindel, J., Wright, M., Chen, C., Cobb, J., Gage, J., Harrington, S., Lorieux, M., Ahmadi, N., McCouch, S., (2013). Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP markers and new

value to traditional bi-parental mapping and breeding populations. Theor. Appl. Genet. (126): 2699-2716.

Stam, P., 1993. Construction of integrated genetic linkage maps by means of a new computer package: Join Map. Plant J (3): 739-744.

Swarts, K., Li, H., Romero Navarro, J.A., An, D., Romay, M.C., Hearne, S., Acharya, C., Glaubitz, J.C., Mitchell, S., Elshire, R.J., Buckler, E.S., Bradbury, P.J., (2014). Novel methods to optimize genotypic imputation for low-coverage, Next-Generation Sequence data in crop plants. Plant Gen. doi:10.3835/plantgenome2014.05.0023.

R development core team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Troggio, M., Malacarne, G., Coppola, G., Segala, C., Cartwright, D.A., Pindo, M., Stefanini, M., Mank, R., Moroldo, M., Morgante, M., Grando, M.S., Velasco, R., (2007). A dense single-nucleotide polymorphism-based genetic linkage map of grapevine (*Vitis vinifera* L.) anchoring Pinot Noir Bacterial Artificial Chromosome contigs. Genetics (176): 2637-2650.

URGI, (2014). 12X.2 version of the grapevine reference genome sequence from The French-Italian Public Consortium (PN40024). [Internet]. Available: https://urgi.versailles.inra.fr/Species/Vitis/Data-Sequences/Genome-sequences.

Van Ooijen, J.W., (2011). Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. Genetics Res (93): 343-349.

Vezzulli, S., Micheletti, D., Riaz, S., Pindo, M., Viola, R., This, P., Walker, M.A., Troggio, M., Velasco, R., (2008)a. A SNP transferability survey within the genus *Vitis*. BMC Plant Bio (8): 128.

Vezzulli, S., Troggio, M., Coppola, G., Jermakow, A., Cartwright, D., Zharkikh, A., Stefanini, M., Grando, M., Viola, R., Adam-Blondon, A.-F., Thomas, M., This, P., Velasco, R., (2008)b. A reference integrated map for cultivated grapevine (*Vitis vinifera* L.) from three crosses, based on 283 SSR and 501 SNP-based markers. Theor Appl Genet (117): 499-511.

Vitulo, N., Forcato, C., Carpinelli, E., Telatin, A., Campagna, D., D'Angelo, M., Zimbello, R., Corso, M., Vannozzi, A., Bonghi, C., Lucchin, M., Valle, G., (2014). A deep survey of alternative splicing in grape reveals changes in the

splicing machinery related to tissue, stress condition and genotype. BMC Plant Biol (14): 99.

Wang, N., Fang, L., Xin, H., Wang, L., Li, S., (2012). Construction of a high-density genetic map for grape using next generation restriction-site associated DNA sequencing. BMC Plant Bio (12): 148.

Ward, J., Bhangoo, J., Fernandez-Fernandez, F., Moore, P., Swanson, J., Viola, R., Velasco, R., Bassil, N., Weber, C., Sargent, D., (2013). Saturated linkage map construction in *Rubus idaeus* using genotyping by sequencing and genome-independent imputation. BMC Genomics (14): 2.

Wu, Y., Bhat, P.R., Close, T.J., Lonardi, S., (2008). Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. PLoS Genet (4): e1000212.

Xie, W., Feng, Q., Yu, H., Huang, X., Zhao, Q., Xing, Y., Yu, S., Han, B., Zhang, Q., ((2010)). Parent-independent genotyping for constructing an ultrahigh-density linkage map based on population sequencing. Proc Natl Acad Sci USA (107): 10578–10583.

CHAPTER 3

A GENETIC LOCUS ASSOCIATED WITH ABUNDANCE OF THE PREDATORY MITE, *TYPHLODROMUS PYRI,* HAS A MAJOR INFLUENCE ON LEAF TRICHOME TRAITS IN *VITIS*[1]

Abstract

Abundance of predatory phytoseiid mites, *Typhlodromus pyri*, important biological control agents of spider mite pests in numerous crops, is positively influenced by the density of leaf trichomes and tuft-form domatia in vein axils. Abundance of *T. pyri* and non-glandular trichomes was measured in a segregating F_1 family derived from the cross of the complex *Vitis* hybrid, 'Horizon', with Illinois 547-1 (*V. rupestris* B38 x *V. cinerea* B9). One major quantitative trait locus (QTL) located on chromosome 1 of Illinois 547-1 explained 23% of the variation in phytoseiid abundance and co-localized with QTL that explained similar amounts of variance in trichome traits, such as domatia rating (21%), domatia size (16%), leaf bristle density (37% in veins and 33% in blades), and leaf hair density (20% in veins and 15% in blades). Another nine QTL distributed among chromosomes 1, 2, 5, 8 and 15 were associated solely with trichome density, and explained 7-17% of the phenotypic variation. While these results show that the abundance of

[1] Contributions: Jan Nyrop, Gregory Loeb and Bruce Reisch conceived of and designed experiments; Bruce Reisch provided genetic populations and access to the field; Jan Nyrop and Gregory Loeb provided lab resources; Rebecca Loughner and Karen Wentworth collected and analyzed phenotype data; Paola Barba analyzed data and wrote the paper.

predatory mites can be influenced by genomic regions of the host, co-segregation of this QTL with loci controlling trichome traits may explain the positive correlation observed in previous phenotypic and ecological studies. Combined, our results provide evidence of the genetic control of abundance of a biological control agent at different trophic level in plants, and provide information for breeding grapevines with a more favorable habitat for biological control agents.

Introduction

Leaf morphological traits in plants can positively or negatively affect plant fitness through direct effect on the second trophic level (plant feeding insects) and indirectly by directly affecting the abundance of the third trophic level (natural enemies of plant feeding insects) (tri-trophic interactions *sensu* Price et al., 1980). Leaf trichomes have been particularly well studied in this regard. In multiple plant systems, leaf trichomes have been shown to have a direct negative effect on the performance of small herbivorous arthropods (direct defense) (Levin, 1973, Southwood, 1986, Mauricio, 1998, Handley et al., 2005, Plett et al., 2010, Valverde et al., 2001). This includes both non-glandular epidermal trichomes (bristles and hairs) and glandular trichomes that release sticky and/or toxic substances (Duffey, 1986). In some cases, leaf trichomes positively affect preference or performance of herbivorous arthropods, although more commonly, performance can be enhanced indirectly by negatively affecting performance and predation rates of natural enemies (Cortesero et al., 2000). This is best shown with glandular trichomes

(Duffey, 1986, Kauffman & Kennedy, 1989, Kennedy, 2003, Simmons & Gurr, 2005). The relationship between non-glandular trichomes and natural enemies such as predatory and mycophagous mites has frequently been shown to be positive (reviewed in Walter, 1996, Schmidt, 2014) although counter examples also exist (Camporese & Duso, 1996, Krips et al., 1999, Seelmann et al., 2007, Schmidt, 2014).

The mechanistic basis of the positive direct effect of non-glandular trichomes on natural enemy mites is most likely related to providing shelter from their natural enemies (intraguild predators), but also trichomes may increase the availability of alternative food sources such as pollen (reviewed by Schmidt, 2014). Non-glandular trichomes can be widely distributed on the leaf blade, or concentrated along veins or in vein axils on the abaxial leaf surface. Non-glandular trichomes concentrated in the vein axils are referred to as tuft-form acarodomatia to reflect the common observation that they are frequently occupied by predatory and mycophagous species of mites as opposed to phytophagous species (Pemberton & Turner, 1989, Brouwer et al., 1990, Walter, 1996, Walter & O'Dowd, 1995). Trichomes that make up tuft-form acarodomatia (simplified from here on as domatia) appear similar in shape and size to non-glandular trichomes found on leaf veins or the leaf blade, although the underlying embryogenesis or genetics have not been explicitly explored.

Predatory mites in the families Phytoseiidae and Stigmatidae are well known as important biological control agents of spider mites and other arthropod

pests in agriculture (Hoy, 2011, Duso et al., 2012), and mycophagous mites in the family Tydeidae have been shown to be important consumers of some fungal pathogens such as grape powdery mildew (English-Loeb et al., 2007). *Typhlodromus pyri* are generalist phytoseiids that prey on grapevine pests, including the European red mite *Panonychus ulmi* and the grape rust mite *Calepitrimerus vitis* (Nalepa) (Duso et al., 2012, Loeb et al., 2015), but also feed on pollen and fungal spores. Studies have shown that the presence of trichomes and domatia influence the abundance of some species of generalist phyotoseiid mites even greater than the abundance of prey (Karban et al., 1995, English-Loeb et al., 2002). Generalist phytoseiids often make greater use of and lay more eggs on plants with pubescent leaves, compared to plants with glabrous leaves (Roda et al., 2001). *T. pyri* and some other generalist phytoseiid mites show a strong preference for leaves containing non-glandular trichomes and are more abundant in grape and apple cultivars with pubescent leaves (Duso et al., 1991, Loughner et al., 2008). In an assemblage study, plant patches containing higher proportions of glabrous plants had significantly lower populations of *T. pyri* six weeks after inoculation, suggesting that trichomes are required to maintain the overall abundance of these predatory mites (Loughner et al., 2010).

Genes regulating the development of non-glandular trichomes in the model plant *Arabidopsis thaliana* are well characterized (Pattanaik et al., 2014, Hauser, 2014). The GLABROUS2 (GL2) homeodomain protein, required for trichome morphogenesis, can be activated or repressed by several transcription factors (Table 3-1), which have been shown to result in either positive or negative effects, respectively, on trichome development.

Table 3-1: List of proteins regulating GLABROUS2 (GL2) protein, required for trichome morphogenesis in *Arabidosis thaliana*. These proteins have either a positive or negative effect on trichome development.

Funtion	Protein type	Protein name
Part of GL2 activator complex	WD40	TRANSPARENT TESTA GLABRA 1 (TTG1)
	R2R3 MYB	GLABROUS1 (GL1), MYB23 and MYB5
	bHLH	GLABROUS3 (GL3), ENHANCER OF GLABROUS3 (EGL3), TRANSPARENT TESTA (TT8) and MYC-1
Negative regulation of GL2	R3 MYB	TRIPTYCHON (TRY), CAPRICE (CPC), ENHANCER OF TRY and CPC1 (ETC1, ETC2, ETC3), and TRICHOMELESS1 (TCL1, TCL2)
Gibberellic acid-dependent negative regulators	DELLA	GA INSENSITIVE (GAI)
		REPRESSOR OF gal-3 (RGA)
		RGL1, RGL2 and RGL3
	C2H2 zinc finger proteins (ZFP)	GLABROUS INFLORESCENCE STEMS (GIS, GIS2), ZFP8, ZFP5 and ZFP6

The heritability of domatia size (English-Loeb et al., 2002) and trichome density (Agren & Schemske, 1992, Shockley & Backus, 2002, Chitwood et al., 2014) have been demonstrated in *Vitis* and several other species. Despite the evidence of genetic determinism of trichomes and domatia, and the positive correlation with predatory mite abundance, these phenotypes have not been integrated in many breeding programs for improvement of predatory mite habitat (Schmidt, 2014).

In order to achieve enhanced biological control through breeding, there are two main aspects that require further attention. First, we need to determine and understand how the desirable trait (i.e., increasing the size of predatory mite populations) is inherited. Secondly, we need to determine if the progress achieved through selection is sufficient to provide effective biological pest control. In this paper we address the first question by studying the genetics of predatory mite abundance and leaf trichomes in a cross of two grapevine hybrids: 'Horizon' and Illinois 547-1. We first studied the segregation and correlation between phytoseiid abundance and leaf morphological traits (domatia, bristles and hair on leaf veins and blades). Then, we localized and characterized the regions of the genome controlling these traits by multiple QTL mapping.

Materials and Methods

Plant material

A set of 190 seedlings was generated from crosses made in 1988 and 1996 between the complex grapevine hybrid 'Horizon' ('Seyval' x 'Schuyler', whose pedigree includes *Vitis vinifera, V. labrusca, V. aestivalis* and *V. rupestris*) and Illinois 547-1 (*V. rupestris* B38 x *V. cinerea* B9). These seedlings were grown in pesticide-free vineyards maintained by the New York State Agricultural Experiment Station, Geneva, New York, USA.

Predator abundance characterization

To estimate the abundance of *T. pyri*, field vines were sampled five times in 2005, five times in 2006, twice in 2007, and three times in 2008. Five mid-shoot leaves from each genotype were collected on each sampling occasion. Phytoseiid motiles (adults and hatched immature mites) were counted by examining each leaf under a microscope. Up to 10 adult phytoseiids per plant from each sampling occasion were mounted on glass slides to confirm identification. Mean counts per leaf were estimated each year from the samples.

Leaf trichome characterization

Leaf trichomes were characterized in parents and in a subset of the F_1 progeny using six variables: 1) Domatia size (mm), 2) Domatia rating, based on the density of trichomes that compose each individual domatia as described in Loughner et al. (2008). 3) Density of bristles on the leaf blade, 4) Density of bristles along leaf veins, 5) Density of hairs on the leaf blade, and 6) Density of hairs along leaf veins. Trichomes were observed under a microscope and classified as either bristles (trichomes shorter than 0.25 mm and upright) or hairs (trichomes longer than 0.25 mm and prostrate), and rated based on a scale from 0 (absent) to 9 (very dense) (IPGRI et al., 1997). The bristle scoring system was applied to four vein axils per leaf to generate a mean leaf domatia rating. Domatia size was determined by measuring the diameter of the zone of bristles extending away from the same four domatia with a micrometer and calculating a mean size per leaf. Trichome parameters were assessed by collecting 2 to 4 mid-shoot leaves per plant once during

2004, 2005, and 2008 (190, 155, and 152 plants, respectively). Parental trichome parameters were measured as described above, using ten leaves collected in 2005.

Single nucleotide polymorphism (SNP) genotyping and construction of genetic maps

One young leaf per vine was used for DNA extraction using the DNeasy® 96 Plant Kit (Qiagen) amended with 2.8% w/v PVP-40 in the extraction buffer. Samples were used for 384-plex library construction at the Cornell Institute for Genomic Diversity following a genotyping-by-sequencing protocol (Elshire et al., 2011). Libraries were sequenced with an Illumina HiSeq 2000 DNA sequencer (single-end,100 bp read length) at the Genomics Facility of the Institute of Biotechnology at Cornell University.

Raw sequence data were aligned to the grapevine PN40024 reference genome version 12X.0 (Jaillon et al., 2007, Adam-Blondon et al., 2011) and processed into SNP genotype files in VCF format using the TASSEL 3.0 GBS pipeline (Glaubitz et al., 2014). SNP names indicate their position on the reference genome, coded as S(chromosome)_(position in bp). Vines resulting from self-pollination or cross contamination were identified using relatedness and paternal/maternal Mendelian incompatibility ratios. SNPs from the remaining samples were filtered down to a subset of pseudo-testcross markers according to quality score, read depth and minor allele frequency (MAF). Parental genetic maps were constructed using a *de novo* pipeline (Chapter 2). Parental maps were compared based on the physical position of SNP markers in the grapevine reference genome PN40024, as the use of

exclusively pseudo-testcross markers does not allow the construction of consensus maps.

QTL analysis

QTL were determined using the R/qtl package (Broman et al., 2003) implemented in the statistical software, R (R Development Core Team, 2008). Phytoseiid abundance data were power transformed as described in the Statistics section below. Multipoint probabilities were calculated using *calc.genoprob* with step = 1 and default parameters. Initial QTL positions were determined with the *scanone* function using a normal model, Haley-Knott regression and default parameters. LOD significance scores were determined by permutation tests (1,000). Initial QTL positions were then used to define QTL with the *makeqtl* function, the significance of model terms was tested with the *fitqtl* command, and positions were refined with *refineqtl*. Models were constructed by adding and then removing QTL one at the time, and the model with the best fit was selected. Non-significant QTL or non-significant QTL interactions were also removed. The *addqtl* command was used to test if another QTL was needed, which was then added to the model as described above. A histogram of the model's residuals was used to check for normality. A 1.5 LOD support interval, defined as the interval in which the LOD score is within 1.5 units of its maximum (Broman et al., 2009), was determined using *lodint* function and QTL effects were calculated as the difference in the mean phenotype value of individuals within each genotype class at the marker or pseudomarker with the highest LOD score, using *effectplot*. QTL names follow R/qtl nomenclature: (linkage group)@(position in cM). QTL located at similar physical positions on both genetic parental maps are reported as co-localized

but separate QTL, because the use of exclusively pseudo-testcross markers does not allow the construction of consensus maps.

Statistics

Statistical analyses were performed using the statistical software R (R Development Core Team, 2008). Parental phenotypes were compared using a t-test. Correlations among progeny phenotype data (Pearson's r) were determined using the *cor.test* function from the stats package and visualized using the corrplot package. Normality of trait distribution was tested using the *shapiro.test* function from stats package. The lambda parameters for Box-Cox power transformation were determined using the *boxcoxnc* function from the AID package in R (Osman D., 2014).

Candidate genes

A total of 28 proteins involved in non-glandular trichome development were obtained from the TrichOME database and other sources (Dai et al., 2010, Hauser, 2014, Pattanaik et al., 2014) (Appendix 3-1). Protein – protein BLAST (Altschul et al., 1997) was performed between candidate gene products (Appendix 3-2) and both the V1 and V2.1 annotation of the 12x.0 version of the PN40024 reference genome (Vitulo et al., 2014). Hits within QTL supported intervals were selected with the GenomicRanges package in R (R Development Core Team, 2008). *Vitis* candidate gene products were aligned to RefSeq protein - *Arabidopsis thaliana* in the NCBI database using BLASTp to confirm similarity of function.

Results

To test whether predatory mite abundance and density of leaf morphology traits such as domatia, bristles and hairs have common genetic factors we characterized an F_1 family segregating for these traits. First, we measured the correlation and the distribution of the phenotypic data.

Characterization of parental leaf trichome traits

Parental genotypes (‘Horizon’ and Illinois 547-1) expressed appreciable levels of leaf trichomes (Figure 3-1) with clear variation in the location and density of bristles and hairs in both leaf blade and leaf veins (Table 3-2).

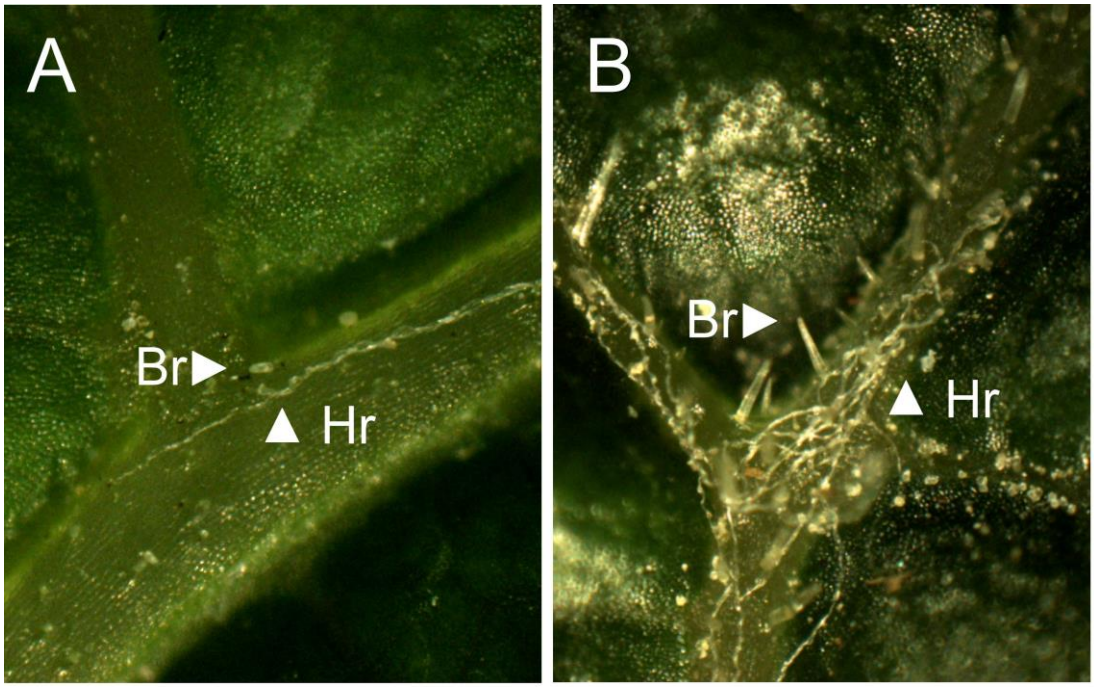


Figure 3-1. Comparison of densities of bristle and hair on leaf veins of ‘Horizon’ (A) and Illinois 547-1 (B). Examples of bristles (Br) and hairs (Hr) are indicated.

Table 3-2. Leaf morphological traits for ‘Horizon’ and Illinois 547-1: Values correspond to the means of ten independent leaves. The density of bristles and hairs was measured on a discrete scale from 0 (absent) to 9 (very dense), Four domatia per leaf were measured and rated with the bristle scale.

Trait	‘Horizon’	Illinois 547-1	p-value[a]
Domatia size (mm)	0.95	1.27	0.11
Domatia (rating)	7.5	8.5	< 0.0001
Bristles on leaf veins (density)	1.8	7.9	< 0.0001
Bristles on leaf blade (density)	0.2	4.0	< 0.0001
Hairs on leaf veins (density)	4.0	3.1	0.013
Hairs on leaf blade (density)	3.6	0.0	< 0.0001

Characterization of predatory mite abundance and leaf trichome traits in an F_1 family

Phytoseiid abundance was positively correlated among the four years, and also positively correlated with all leaf morphology traits (Figure 3-2). The correlation of phytoseiid abundance measured in 2007 was weaker, but still significant and positive with at least one year of evaluation of all other traits.

Leaf morphology traits were positively correlated among each other in all cases, showing stronger correlations with data of a similar nature, such as domatia size with domatia rating, bristle density on veins with bristle density on blades, and hair density on veins with hair density on blades. There was also a strong correlation among phenotypes measured in the same year (Figure 3-2).

Hair phenotypes showed the weakest correlations with other morphology traits (often positive but non-significant), particularly with domatia size and bristle density on leaf blades.

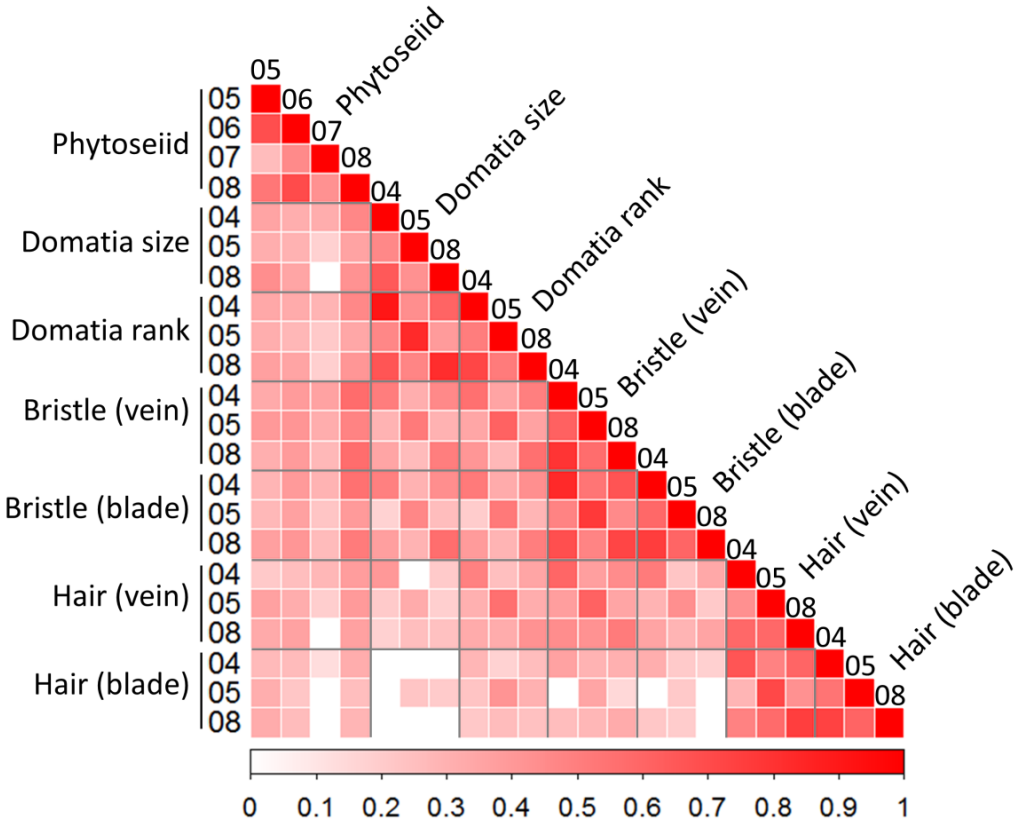


Figure 3-2: Correlation matrix of phytoseiid abundance and leaf traits in F_1 progeny of 'Horizon' x Illinois 547-1. Pearson's correlations among phenotype by year are represented by a color scale. Non-significant correlations (at $\alpha = 0.05$) are indicated in white. No negative correlations were found.

Leaf morphological traits showed continuous and consistent segregation patterns across years, with individuals representing the spectrum of ratings from 0 (absent) to near 9 (very dense) (Figure 3-3). Both domatia measures showed the shortest interquartile range, consistent with similar mean values

obtained on both parents. Bristle densities were similar on both leaf veins and leaf blades, with a higher proportion of individuals with zero bristles on blades. Hair density distribution was different on leaf veins than on leaf, with less presence of hair on blades.

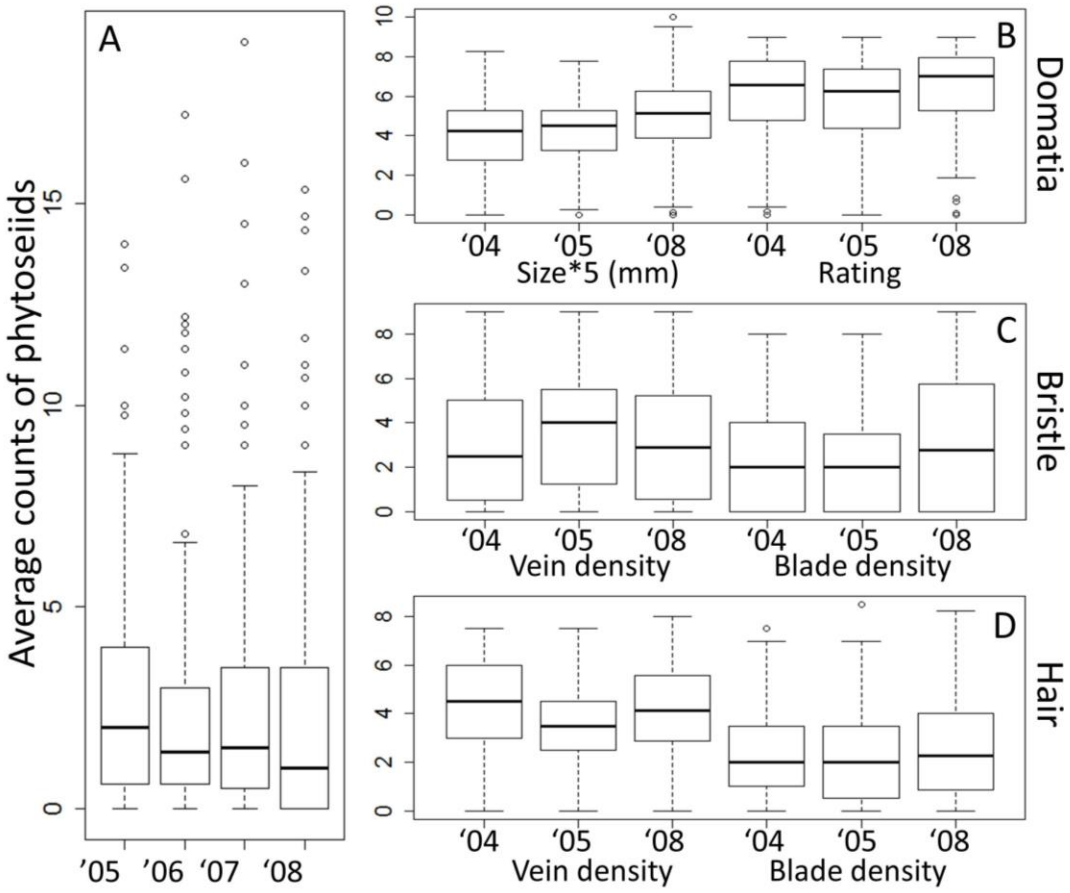


Figure 3-3: Distributions (box and whiskers plots) of phenotypes. (A) Predatory mite abundance (mean phytoseiid count per leaf), (B) domatia size and rating, (C) bristles density in veins and blades and (D) hair density in veins and blades, by year. Domatia size is show as 5X mm measures to fit the plot scale. Bristles and hairs in both, veins and blades were determined as a rating in a density scale from 0 (absent) to 9 (very dense). Domatia rating corresponds to the average bristle density rating of four domatia per leaf. For each data set, the box indicates the range between 25 and 75%, and middle band indicates the sample median. Outlier values are indicated with a circle.

QTL detection

To find the genetic regions controlling the phenotypes described above, and to assess to what extent distribution of the phytoseiid abundance and trichome densities among the progeny is due to major genetic components, we conducted a multiple QTL mapping study.

QTL models were significant for each trait in each year data were collected, explaining between 11.1 and 69.4% of the phenotypic variance. Among all traits, a total of 16 QTL were significant for two or three years and were distributed among chromosomes 1, 2, 5, 8 and 15, with chromosomes 1, 5, and 15 contributing effects from both parental maps (Table 3-3).

For all traits, 14 QTL were above threshold for only one evaluation. These unstable QTL are indicated in Table 3-3 and were not considered for further analysis.

Co-localization of mite abundance and trichome traits

Of all QTL, one from Illinois 547-1 showed the greatest effect and was the only QTL associated with predatory mite abundance. This QTL was significant for four consecutive years (2005–2008) and it co-localized with major QTL for each leaf morphology trait (Table 3-4).

Statistics associated with this QTL were consistent among years, with some variation on the QTL peak, supported interval and effect. According to the physical coordinates of associated SNPs on chromosome 1 (Table 3-4), the phytoseiid QTL peak location varied between 8,744,147 and 11,789,500 bp. For domatia size and rating the QTL peak ranges were similar, between

9,877,466 and 10,879,085 bp and between 9,982,106 and 14,859,772 bp, respectively. For other trichome traits, the range of the QTL peaks was wider, but still within the supported interval of the QTL associated with predatory mite abundance. Density of bristles on leaf veins and leaf blades showed QTL peaks between 10,130,832 and 18,359,738 bp, and between 12,043,160 and 15,492,583 bp, respectively. Density of hairs on leaf veins and leaf blades were associated with QTL peaks between 8,744,147 and 14,859,772 bp, and between 9,395,874 and 15,492,583 bp, respectively.

Additional QTL for domatia and trichome phenotypes

In addition to the major QTL in chromosome 1, other QTL were associated solely with leaf trichomes (Table 3-4). A QTL located on chromosome 15 was significant for domatia rating and domatia size. For bristle density, a minor QTL located on chromosome 1 of the female parent ‘Horizon’ explained was associated with variation on both leaf veins and leaf blades. Another minor QTL, only for bristles density on leaf blades, was located on chromosome 15 of ‘Horizon’. Different sets of minor QTL were found for hair density on veins versus leaf blades. On veins, one QTL was located on chromosome 2 of ‘Horizon’. On leaf blades we found three QTL located on chromosome 5 of Illinois 547-1, chromosome 5 of ‘Horizon’ and chromosome 8 of Illinois 547-1.

Table 3-3: QTL models for phytoseiid counts, domatia size and rating, bristle, and hair densities on leaf veins and leaf blades. QTL in the model were named using R/qtl nomenclature, ie. linkage group@position (cM). Linkage group numbers 1 to 19 correspond to chromosome numbers in the female map ('Horizon'), and 20 to 38 correspond to chromosome numbers 1 to 19 in the male map (Illinois 547-1)

Phenotype	Year	Sample size	Number of QTL in model	% of Variance explained by model	Model
Phytoseiid (Counts)	2005	147	2	23.8	20@37.9 + 37@0.9*
	2006	147	1	24.7	20@40.6
	2007	143	1	11.1	20@32.4
	2008	152	2	54.0	20@37.3 + 1@68.3*
Domatia (Rating)	2004	190	2	38.5	20@42.0 + 34@29.0
	2005	155	1	14.5	20@37.6
	2008	152	4	46.8	20@41.0 + 34@29.8 + 18@32.0* + 1@47.0*
Domatia (mm)	2004	190	3	43.0	20@37.0 + 34@31.0 + 23@12.7 *
	2005	154	1	13.0	20@37.6
	2008	152	5	50.9	20@38.3 + 34@22.0 + 10@9.7* + 18@20.7* + 37@24.6*
Bristles on veins (density)	2004	190	3	63.2	20@41.8 + 1@67.2 + 8@72.4*
	2005	155	2	44.1	20@37.6 + 1@56.0
	2008	152	2	61.8	20@47.3 + 1@65.8
Bristles on leaf blade (density)	2004	190	3	55.0	20@41.8 + 1@77.5 + 15@27.9
	2005	155	3	42.7	20@43.0 + 19@69.9* + 1@61.0
	2008	152	4	69.4	20@45.0 + 15@26.0 + 1@68.3 + 36@41.0*
Hair on leaf veins (density)	2004	190	1	25.8	20@43.5
	2005	155	3	31.9	20@37.6 + 35@62.0* + 2@50.0
	2008	152	3	40.2	20@35.0 + 27@56.7* + 2@47.0
Hair on leaf blade (density)	2004	190	5	49.5	24@5.0 + 20@42.2 + 5@4.3 + 35@65.0* + 27@60.3 + 24@5.0 : 5@4.3*
	2005	155	1	15.6	24@5.0
	2008	152	4	41.7	20@34.4 + 27@56.7 + 5@3.1 + 24@4.0

* QTL were significant only on one date of evaluation.

: Indicates an interaction term between two QTL

Table 3-4: Statistics of stable QTL. QTL detected for two or three years following four years of data collection for phytoseiid counts and three years of data collection for trichome related traits. QTL were obtained by multiple QTL mapping on separated parental maps. QTL effects are measured in the phenotype scale indicated, where densities and ratings correspond to the scale from zero to nine. Alleles of an associated marker are indicated according to the QTL effect, with the allele with higher mean phenotype value first.

Chr[a]	Parent	Phenotype	Year	Position of LOD peak and supported interval (cM)	LOD	LOD Th[b]	% Var[c]	Effect	Associated marker and alleles[d]
1	Illinois 547-1	Phytoseiid (Counts)	2005	36.2 - 37.9 - 45.0	5.36	3.42	13.9	2.41	S1_10602967 A/G
			2006	35.3 - 40.6 - 43.5	9.06	3.38	24.7	3.25	
			2007	29.1 - 32.4 - 58.0	3.64	3.46	11.1	2.32	
			2008	36.2 - 37.3 - 41.8	21.0	3.49	41	3.77	
1	Illinois 547-1	Domatia (Rating)	2004	35.3 - 42.0 - 45.3	15.5	3.34	28	2.24	S1_11789500 A/T
			2005	35.3 - 37.6 - 49.9	5.25	3.33	14.5	1.62	
			2008	35.3 - 41.0 - 46.0	10.3	3.25	19.6	1.67	
1	Illinois 547-1	Domatia (mm)	2004	35.3 - 37.0 - 49.0	11.9	3.45	19	0.28	S1_10130853 G/A
			2005	35.3 - 37.6 - 48.5	4.65	3.35	13	0.25	
			2008	36.2 - 38.3 - 42.6	8.67	3.42	14.7	0.33	
1	Illinois 547-1	Bristle on veins (Density)	2004	41.0 - 41.8 - 45.0	31.5	3.46	42.1	3.81	S100_486628 C/-
			2005	36.2 - 37.6 - 43.5	15.4	3.3	32.3	3.17	
			2008	44.0 - 47.3 - 47.9	22.5	3.42	37.3	3.61	
1	Illinois 547-1	Bristle on leaf blades (Density)	2004	40.6 - 41.8 - 45.0	26.1	3.36	39.7	2.91	S100_311430 A/G
			2005	36.2 - 43.0 - 51.0	11.0	3.39	21.6	2.26	
			2008	43.5 - 45.0 - 46.7	26.9	3.43	38.4	3.79	
1	Illinois 547-1	Hair on leaf blades (Density)	2004	36.2 - 42.2 - 45.0	9.31	3.44	12.8	1.29	S1_12478623 C/A
			2005	36.2*	2.31*	3.32	n.s.	n.s	
			2008	29.8 - 34.4 - 47.3	8.32	3.29	16.7	1.71	
1	Illinois 547-1	Hair on leaf veins (Density)	2004	40.6 - 43.5 - 45.0	12.3	3.29	25.8	1.70	S1_10130853 G/A
			2005	33.0 - 37.6 - 48.5	6.11	3.33	13.6	1.32	
			2008	32.0 - 35.0 - 45.0	9.83	3.32	20.8	1.59	
1	'Horizon'	Bristle on veins (Density)	2004	58.0 - 67.2 - 83.1	6.47	3.46	6.25	1.96	S1_18418322 A/T
			2005	47.0 - 56.0 - 83.1	4.39	3.3	7.78	1.74	
			2008	60.0 - 65.8 - 83.0	11.7	3.42	16.2	2.64	
1	'Horizon'	Bristle on leaf blades (Density)	2004	62.0 - 77.5 - 83.1	5.12	3.36	5.95	1.41	S1_18418322 A/T
			2005	50.0 - 61.0 - 83.1	4.31	3.39	6.74	1.51	
			2008	59.0 - 68.3 - 82.0	8.21	3.43	8.63	2.52	
2	'Horizon'	Hair on leaf veins (Density)	2004	57.4*	0.20*	3.29	n.s.	n.s	S2_13348927 T/G
			2005	37.4 - 50.0 - 62.4	3.46	3.33	7.37	0.94	
			2008	35.0 - 47.0 - 57.4	3.79	3.32	7.29	1.08	
5	'Horizon'	Hair on leaf blades (Density)	2004	0.00 - 4.29 - 7.89	7.7	3.44	6.91	0.95	S5_2523862 A/T
			2005	3.13*	1.28*	3.32	n.s.	n.s	
			2008	0.00 - 3.13 - 12.5	4.46	3.29	8.43	1.34	
5	Illinois 547-1	Hair on leaf blades (Density)	2004	1.13 - 5.00 - 6.00	11.5	3.44	16.2	1.32	S5_750796 T/C
			2005	0.74 - 5.00 - 10.0	5.69	3.32	15.6	1.61	
			2008	0.00 - 4.00 - 8.00	4.42	3.29	8.35	1.41	

Table 3-4 Continued

Chr[a]	Parent	Phenotype	Year	Position of LOD peak and supported interval (cM)	LOD	LOD Th[b]	% Var[c]	Effect	Associated marker and alleles[d]
8	Illinois 547-1	Hair on leaf blades (Density)	2004	53.0 - 60.3 - 63.7	5.29	3.44	7.36	0.84	S8_19609227 -/C
			2005	54.9*	1.30*	3.32	n.s.	n.s	
			2008	47.3 - 56.7 - 62.0	4.57	3.29	8.65	1.33	
15	'Horizon'	Bristle on leaf blades (Density)	2004	24.0 - 27.9 - 39.4	4.13	3.36	4.74	0.94	S15_14485614 C/T
			2005	27.9*	1.68*	3.39	n.s.	n.s	
			2008	24.0 - 26.0 - 30.0	9.64	3.43	10.4	1.97	
15	Illinois 547-1	Domatia (Rating)	2004	26.0 - 29.0 - 43.0	8.60	3.34	14.4	1.51	S15_14939741 A/G
			2005	29.8*	1.00*	3.33	n.s.	n.s	
			2008	23.0 - 29.8 - 32.0	10.3	3.25	19.5	1.40	
15	Illinois 547-1	Domatia (mm)	2004	25.1 - 31.0 - 34.9	11.5	3.45	18.4	0.28	S15_15178923 A/G
			2005	28.6*	0.53*	3.35	n.s.	n.s	
			2008	18.5 - 22.0 - 28.0	6.1	3.42	9.97	0.28	

[a] Chr corresponds to the physical chromosome number [b] LOD threshold, was determined by permutation test (1,000) over the whole set of markers, at $\alpha = 0.05$

[c] Percentage of the total phenotypic variance explained by a single QTL

[d] Marker name indicates the position on the 12X.0 PN40024 *Vitis vinifera* reference genome in format S(chromosome)_(position in bp)

* No significant hit for this year. The highest LOD score by single QTL scan for this year is reported.

n.s. Values not reported as QTL were not significant for this year.

Candidate genes

Reference genome-based genotyping-by-sequencing directly produces physical positions for QTL support intervals on a genetic map, facilitating the search for candidate genes.

First, we looked for evidence of *Vitis* candidate genes with similarity to 28 *Arabidopsis thaliana* genes involved in the development of non-glandular trichomes (Appendix 3-1), using the PN40024 reference genome regions delimited by all QTL intervals. Twelve such candidate genes were identified, including three located on chromosome 1 and five on chromosome 15 (Table 3-5 and Appendix 3-2).

Given the extent of the QTL support intervals; the total number of annotated genes within each QTL is high. Here, we only report the complete list of genes for the major QTL on linkage group 1 (Appendix 3-3). Within the consensus interval located between 9.58 and 14.95 Mbp of chromosome 1, there were 239 annotated genes (Version 2.1, Vitulo et al., 2014)

Table 3-5: Grapevine candidate genes for phytoseiid abundance and trichome related traits on the PN40024 reference genome. Homologous *Arabidopsis thaliana* genes are indicated.

Chr[a]	Start position	*Vitis* Gene Name[b]	Trichome-related Gene Name ; Function ; Query id	Associated phenotype
1	8,079,551	VIT_201s0127g00730	GL1 (GLABRA 1); transcription factor ; NP_189430.1	phytoseiid, hair on blades
			MYB23 (myb domain protein 23); DNA binding / transcription factor ; NP_198849.1	
	9,724,089	VIT_201s0026g00850	GIS2 (GLABROUS INFLORESCENCE STEMS 2); nucleic acid binding / transcription factor/ zinc ion binding ; NP_196283.1	phytoseiid, domatia (rating), domatia (size), hair on blades, hair in veins
			ZFP8 (ZINC FINGER PROTEIN 8); nucleic acid binding / transcription factor/ zinc ion binding ; NP_181725.1	
	19,230,746	VIT_201s0010g02270	RGA1 (REPRESSOR OF GA1-3 1); transcription factor ; NP_178266.1	phytoseiid, domatia (rating), bristles on blades, bristles on veins
			RGL1 (RGA-LIKE 1); transcription factor ; NP_176809.1	
2	8,739,817	VIT_02s0012g02030	GL2 (GLABRA 2); DNA binding / transcription factor ; NP_001185443.1	hair on veins
5	1,128,866	VIT_05s0077g01390	GIS2 (GLABROUS INFLORESCENCE STEMS 2); nucleic acid binding / transcription factor/ zinc ion binding ; NP_196283.1	hair on blades
8	20,543,918	VIT_208s0007g06870	GIS (GLABROUS INFLORESCENCE STEMS); nucleic acid binding / transcription factor/ zinc ion binding ; NP_191366.1	hair on blades
			GIS2 (GLABROUS INFLORESCENCE STEMS 2); nucleic acid binding / transcription factor/ zinc ion binding ; NP_196283.1	hair on blades
			ZFP8 (ZINC FINGER PROTEIN 8); nucleic acid binding / transcription factor/ zinc ion binding ; NP_181725.1	hair on blades
	20,865,263	VIT_208s0007g07230	MYB5; transcription repressor ; NP_187963.1	hair on blades

Table 3-5 Continued

Chr[a]	Start position	*Vitis* Gene Name[b]	Trichome-related Gene Name ; Function ; Query id	Associated phenotype
15	13,245,041	VIT_15s0021g02290	SPL8 (SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 8); DNA binding ; NP_683267.1	domatia (size)
	13,256,743	VIT_215s0021g02300	SPL8 (SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 8); DNA binding ; NP_683267.1	domatia (size)
	16,133,315	VIT_15s0048g02000	GL2 (GLABRA 2); DNA binding / transcription factor ; NP_001185443.1	domatia (size), domatia (rating), bristles on blades
	17,211,622	VIT_215s0046g00170	MYB5, transcription repressor ; NP_187963.1	domatia (rating)
	18,177,217	VIT_215s0046g01130	CPC (CAPRICE); DNA binding / transcription factor ; AAS09991.1	domatia (rating)
			ETC2 (ENHANCER OF TRY AND CPC 2) ; DNA binding / transcription factor ; NP_850145.1	domatia (rating)
			ETC3 (ENHANCER OF TRY AND CPC 3); DNA binding / transcription factor ; NP_974493.1	domatia (rating)
			TCL1 (TRICHOMELESS1); DNA binding ; NP_001031445.1	domatia (rating)
			TCL2 (TRICHOMELESS2) ; NP_001118417.1	domatia (rating)
			TRY (TRIPTYCHON); DNA binding / transcription factor ; NP_200132.2	domatia (rating)

[a] Chr indicates the physical chromosome number.

[b] *Vitis* gene name belong to annotations V1 and V2.1 from the CRIBI database (Vitulo et al., 2014)

Discussion

The abundance of the predatory mite, *T. pyri,* on progeny of ‘Horizon and Illinois 547-1grapevines was significantly associated with a region of the Illinois 547-1 genome that also had a major effect upon densities of domatia, hair, and bristles. While a molecular mechanism for trichome development is well-established in model plants, as is the positive relationship of leaf trichome density and predatory mite abundance, the location of a QTL positively associated with predator abundance in a different trophic level is uncommon. Our results shed light on the genetic control of leaf landscape features for predatory mite habitat enhancement.

Our phenotypic assessment showed that predatory mite abundance was positively correlated with all trichome traits. Among those, bristles and domatia showed the highest correlation, and to a minor degree with leaf hair (Figure 3-2). These phenotypic results correspond with the genetic loci found on chromosome 1 of Illinois 547-1, explaining a major proportion of the variance observed in phytoseiid abundance, domatia and bristle traits, and to a lesser degree, in hair densities (Table 3-4). Previous ecological studies reported a similar correlation of phytoseiids with trichomes in a sample of 12 distinct grapevine cultivars. There, phytoseiid abundance was best predicted by a model in which domatia and hair density had an additive effect and domatia had the greatest explanatory power followed by hair density (Loughner et al., 2008).

The genetic architecture of trichome traits, with a major QTL also associated with phytoseiid abundance and different sets of minor QTL for leaf blade or

leaf veins, may explain the correlation observed between mite abundance and leaf trichomes in previous analyses. (See Schmidt, 2014 for a recent review). This may suggests that a gene or several linked genes from either the trichome developmental pathway or for general leaf morphology may be located on chromosome 1.

The association between general leaf morphology traits and leaf hair was found in a meta-analysis of at least 117 *Vitis* accessions by Chitwood et al. (2014). The authors suggested that aspects of vein patterning, laminar outgrowth and epidermal features such as hair, may be regulated by common developmental pathways. In their work, a QTL for the first principal component of leaf morphology was found on chromosome 1, about 4.2 Mb and 16.6 cM (as measured on the Illinois 547-1 genetic map) from the major QTL for predatory mite and trichome traits found in this study.

For the major QTL on chromosome 1, the consensus interval between 9.5 and 14.9 Mb contains 239 predicted genes (Appendix 3-3). Among those, genes involved in both leaf and trichome development pathways are present. Two predicted *Vitis* genes showed homology to components of the trichome regulatory pathway. VIT_201s0127g00730 was similar to transcription factors of the GL2 activator complex (GL1/MYB23) and VIT_201s0026g00850 to GA-dependent upstream repressors of the GL2 activator complex (GIS2/ZFP8). Despite these *Vitis* genes are located within the QTL range, and near the QTL peak, the F_1 family used in this study has limited resolution to accurately propose a causal gene. To elucidate the genetic control of the predatory mite abundance QTL localized on chromosome 1, more recombination around this locus is needed. A larger F_1, F_2, or an association mapping population would

be more suitable to narrow down the position of the locus and reduce the list of candidate genes.

Among the progeny, phytoseiid counts and trichome density segregated continuously. This may indicate a quantitative genetic architecture of the trait and an environmental effect on its expression. QTL models were able to explain only a portion of all phenotypic variance. The percent of variance explained varied among years, suggesting either an environmental effect in the expression of these traits or sampling errors. This is also evident from the appearance of unstable QTL, which were significant only one year (Table 3-3). Despite year-by-year variation (Figure 3-2), the means of trait distributions were fairly stable (Figure 3-3), 16 QTL were detected in more than one season, with overlapping QTL supported intervals (Table 3-4).

Overall, bristles on leaf veins had the simplest genetic model. On average, 56.5% of the variance was explained by the additive effects of two QTL on chromosome 1 of each parental map. Additionally, bristles on leaf blades were controlled by a third locus located on chromosome 15 of ‘Horizon’. An additive model with these three loci explained a similar proportion of the phenotypic variance.

Domatia are leaf structures primarily comprised of bristles. QTL for domatia were co-localized with QTL for bristles on leaf blades, segregating on the Illinois 547-1 map. Together, these two QTL explained a lower proportion of the variance for domatia; 35.6% and 33.3% when measured by size or by a rating (Table 3-3). Measuring domatia size is a direct quantification of the phenotype, but it is labor intensive, and may be more prone to human error. In

this experiment, both scales led to similar QTL results (numbers of QTL and total variance explained). Our results suggest that it is reasonable to apply a categorical scale for QTL mapping of this difficult-to-quantify trait, until more precise quantitative techniques are developed.

The genetic architecture of hair trichomes was more complex, showing greater differences in the number and position of loci controlling the density of hairs on either leaf veins or leaf blades. Such findings are consistent with the dissimilarities in the phenotype segregation. Hairs on leaf veins showed a simpler genetic model, including the major QTL on chromosome 1 and a second QTL on chromosome 2. Together, these QTL explained an average of 32.7% of the variance. In contrast, hair located on the leaf blades was associated with four QTL: the major QTL on chromosome 1 of Illinois 547-1; a minor QTL on chromosome 8 of the Illinois 547-1 map; and two co-localized QTL on chromosome 5 (one on each parental map). Together, these explained 35.6% of the variance.

Being on a different trophic level, predatory mite abundance is more prone to environmental influence. This phenotype was more variable and presented more extreme observations (Figure 3-3). The amount of phenotypic variance explained by genetics was lower than for domatia and trichome traits, with values ranging from 11.1% in 2007 to 54.0% in 2008 (Table 3-3). The most stable locus co-localized with the major QTL for domatia and all trichome traits on chromosome 1 of Illinois 547-1, but in 2008 a QTL on chromosome 1 of the ‘Horizon’ parent was also significant. The position of the Illinois 547-1 QTL was estimated around 37 cM (10 Mbp in the12x.0 version of the PN40024 reference genome) on chromosome 1, with an average effect of three mites

per leaf, and mean values ranging from 2.3 in 2007 to 3.8 in 2008. These results indicate that, despite the year to year fluctuations of phytoseiid counts, it is still possible to explain a difference of three mites per leaf with only one polymorphism at the locus.

Breeding for domatia and trichome densities could be an indirect form of breeding for resistance to pests, by promoting the abundance of generalist feeding phytoseiids in new cultivars. Here, we demonstrated that a single locus in the host genome can produce significant differences in the abundance of predatory mites. Selection for this QTL will likely increase the overall density of domatia, bristles and hair, which could have a pleiotropic effect on other plant functions; on pests not controlled by *T. pyri*; or other natural enemies of insect and mite pests (Cortesero et al., 2000).

The identification of stable genetic regions along with their positions on a reference genome may help to accelerate further discoveries, by providing linked molecular markers and regions to search for candidate genes. Molecular markers identified in the present project may allow breeders to identify progeny shortly after seed germination that will provide better habitats for predatory mites. These can be used in combination with markers for other relevant traits, such as flower sex (Fechter et al., 2012), disease resistance (Mahanil et al., 2012, Barba et al., 2014) or improved berry and cluster architecture (Mejia et al., 2007, Correa et al., 2014). Moreover, the additive effect of the QTL suggests that stacking alleles and loci with minor effects could increase the numbers of phytoseiids per leaf, in order to sustain higher population of predatory mites

APPENDIX

Appendix 3-1: *Arabidopsis thaliana* genes involved in the trichome regulatory pathway

Gene Name	Description	NCBI Protein accession number
ATMYBL2	ATMYBL2 (*Arabidopsis* myb-like 2); DNA binding / transcription factor	NP_177259.1
ATMYC1	ATMYC1 (*Arabidopsis thaliana* myc-related transcription factor 1); DNA binding / transcription factor	NP_001154194.1
CPC	CPC (CAPRICE); DNA binding / transcription factor	AAS09991.1
CPR5	CPR5 (CONSTITUTIVE EXPRESSION OF PR GENES 5)	NP_569003.1
EGL3	EGL3 (ENHANCER OF GLABRA3); DNA binding	NP_001185302.1
ETC1	ETC1 (ENHANCER OF TRY AND CPC 1); DNA binding / transcription factor	NP_171645.1
ETC2	ENHANCER OF TRY AND CPC 2	NP_850145.1
ETC3	ETC3 (ENHANCER OF TRY AND CPC 3); DNA binding / transcription factor	NP_974493.1
GAI	GAI (GA INSENSITIVE); transcription factor	NP_172945.1
GIS	GIS (GLABROUS INFLORESCENCE STEMS); nucleic acid binding / transcription factor/ zinc ion binding	NP_191366.1
GIS2	GIS2 (GLABROUS INFLORESCENCE STEMS 2); nucleic acid binding / transcription factor/ zinc ion binding	NP_196283.1
GL1	GL1 (GLABRA 1); transcription factor	NP_189430.1
GL2	GL2 (GLABRA 2); DNA binding / transcription factor	NP_001185443.1
GL3	GL3 (GLABRA 3); transcription factor	NP_680372.1
MYB23	MYB23 (myb domain protein 23); DNA binding / transcription factor	NP_198849.1
MYB5	transcription repressor MYB5	NP_187963.1
RGA1	RGA1 (REPRESSOR OF GA1-3 1); transcription factor	NP_178266.1
RGL1	RGL1 (RGA-LIKE 1); transcription factor	NP_176809.1
SAD2	SAD2 (SUPER SENSITIVE TO ABA AND DROUGHT2); protein transporter	NP_180724.2
SPL8	SPL8 (SQUAMOSA PROMOTER BINDING PROTEIN-LIKE 8); DNA binding	NP_683267.1
SPY	SPY (SPINDLY); transferase, transferring glycosyl groups	NP_187761.1
TCL1	TCL1 (TRICHOMELESS1); DNA binding	NP_001031445.1
TCL2	TCL2 (TRICHOMELESS 2)	NP_001118417.1
TRY	TRY (TRIPTYCHON); DNA binding / transcription factor	NP_200132.2
TT8	TT8 (TRANSPARENT TESTA 8); DNA binding / transcription factor	NP_192720.2
TTG1	TTG1 (TRANSPARENT TESTA GLABRA 1); nucleotide binding	NP_851069.1
TTG2	TTG2 (TRANSPARENT TESTA GLABRA 2); transcription factor	NP_181263.2
ZFP8	ZFP8 (ZINC FINGER PROTEIN 8); nucleic acid binding / transcription factor/ zinc ion binding	NP_181725.1

Appendix 3-2: BLASTp statistics between *Arabidopsis thaliana* gene products involved in the trichome regulatory pathway (Appendix 3-1) on to predicted proteins of *Vitis vinifera* PN40024 genome (Version 2.1, Vitulo et al., 2014)

Candidate Gene Vitis	Query A thaliana in trichome pathway	Query gene name	identity	Alignment length	Mismatches	Gap opens	e-value	Bit score	QTL
VIT_201s0127g00730.1	gi|15232860|ref|NP_189430.1|	GL1	76.36	110	26	0	$8.0*10^{-55}$	178	Phytoseiid Hairs in blade
VIT_201s0127g00730.1	gi|15242649|ref|NP_198849.1|	MYB23	77.27	110	25	0	$4.0*10^{-58}$	186	Phytoseiid, Hairs in blade
VIT_201s0026g00850.1	gi|15240118|ref|NP_196283.1|	GIS2	36.27	102	46	1	$2.0*10^{-14}$	70	Phytoseiid, Domatia (rating), Domatia (size), Hairs in blade, Hairs in veins
VIT_201s0026g00850.1	gi|15227472|ref|NP_181725.1|	ZFP8	35.04	117	63	3	$9.0*10^{-12}$	64	Phytoseiid, Domatia (rating), Domatia (size), Hairs in blade, Hairs in veins
VIT_201s0010g02270.1	gi|15226311|ref|NP_178266.1|	RGA1	34.09	396	227	11	$4.0*10^{-69}$	236	Phytoseiid, Domatia (rating), Bristles in blades, Bristles in veins.
VIT_201s0010g02270.1	gi|15219630|ref|NP_176809.1|	RGL1	34.79	388	229	10	$9.0*10^{-73}$	244	Phytoseiid, Domatia (rating), Bristles in veins, Bristles in blades
VIT_02s0012g02030.t01	gi|334184032|ref|NP_001185443.1|	GL2	38.95	742	374	20	$7.0*10^{-161}$	490	Hairs in veins
VIT_05s0077g01390.t01	gi|15240118|ref|NP_196283.1|	GIS2	40.54	74	34	1	$1.0*10^{-11}$	60	Hairs in blades
VIT_208s0007g06870.1	gi|15230939|ref|NP_191366.1|	GIS	44.28	271	100	14	$2.0*10^{-48}$	163	Hairs in blades
VIT_208s0007g06870.1	gi|15240118|ref|NP_196283.1|	GIS2	41.2	233	92	9	$2.0*10^{-37}$	132	Hairs in blades
VIT_208s0007g06870.1	gi|15227472|ref|NP_181725.1|	ZFP8	42.91	275	102	11	$3.0*10^{-50}$	168	Hairs in blades
VIT_208s0007g07230.1	gi|15231271|ref|NP_187963.1|	MYB5	56.08	255	67	7	$2.0*10^{-85}$	261	Hairs in blades

Appendix 3-2 Continued

Candidate Gene Vitis	Query A thaliana in trichome pathway	Query gene name	identity	Alignment length	Mismatches	Gap opens	e-value	Bit score	QTL
VIT_15s0021g02290.t01	gi|22329284|ref|NP_683267.1|	SPL8	52.8	125	45	2	$3.0*10^{-33}$	128	Domatia (size)
VIT_215s0021g02300.1	gi|22329284|ref|NP_683267.1|	SPL8	50.42	355	103	12	$2.0*10^{-88}$	272	Domatia (size)
VIT_15s0048g02000.t01	gi|334184032|ref|NP_001185443.1|	GL2	41.76	795	371	17	0	588	Domatia (size), Domatia (rating), Bristles in blades
VIT_215s0046g00170.1	gi|15231271|ref|NP_187963.1|	MYB5	62.35	162	45	1	$8.0*10^{-66}$	209	Domatia (rating)
VIT_215s0046g01130.1	gi|41618962|gb|AAS09991.1|	CPC	62.35	85	31	1	$2.0*10^{-27}$	99	Domatia (rating)
VIT_215s0046g01130.1	gi|30684581|ref|NP_850145.1|	ETC2	65.15	66	23	0	$5.0*10^{-25}$	93	Domatia (rating)
VIT_215s0046g01130.1	gi|42572793|ref|NP_974493.1|	ETC3	63.89	72	25	1	$2.0*10^{-23}$	88	Domatia (rating)
VIT_215s0046g01130.1	gi|79323486|ref|NP_001031445.1|	TCL1	70.91	55	16	0	$3.0*10^{-23}$	88	Domatia (rating)
VIT_215s0046g01130.1	gi|186504271|ref|NP_001118417.1|	TCL2	55.42	83	36	1	$2.0*10^{-23}$	89	Domatia (rating)
VIT_215s0046g01130.1	gi|30696297|ref|NP_200132.2|	TRY	70.73	82	24	0	$2.0*10^{-31}$	110	Domatia (rating)

Appendix 3-3: Annotated genes (Version 2.1, Vitulo et al., 2014) from *Vitis vinifera* PN40024 genome within the supported interval of phytoseiid abundance and trichome density trait QTL on chromosome 1 of Illinois 547-1

Vitis gene id	chr	Start	End	Annotation	
VIT_201s0026g00810	1	9,655,509	9,657,926	GO:0010411	trichome birefringence-like 27 protein
VIT_201s0026g00820	1	9,674,651	9,688,273	GO:0010179	iaa-amino acid hydrolase ilr1-like 4-like
VIT_201s0026g00830	1	9,693,927	9,695,668	GO:0020037	peroxidase 65
VIT_201s0026g00840	1	9,697,437	9,711,849	GO:0031072	altered response to gravity
VIT_201s0026g00850	1	9,724,001	9,730,253	GO:0008270	zinc finger protein
VIT_201s0026g00860	1	9,730,654	9,735,480	GO:0005524	sam domain family protein
VIT_201s0026g00870	1	9,764,040	9,764,869	NA	NA
VIT_201s0026g00880	1	9,773,170	9,774,555	GO:0005737	uncharacterized wd repeat-containing protein alr3466-like
VIT_201s0026g00890	1	9,777,226	9,783,547	NA	NA
VIT_201s0026g00900	1	9,784,843	9,786,152	GO:0031225	metalloendoproteinase 1-like
VIT_201s0026g00910	1	9,804,365	9,804,769	GO:0031348	uncharacterized protein
VIT_201s0026g00920	1	9,806,347	9,818,501	GO:0005576	uncharacterized protein
VIT_201s0026g00930	1	9,833,658	9,867,000	GO:0005634	mynd finger family protein
VIT_201s0026g00940	1	9,871,215	9,874,335	GO:0000160	type-a response regulator
VIT_201s0026g00950	1	9,877,746	9,880,879	GO:0006396	actin binding
VIT_201s0026g00960	1	9,882,571	9,888,217	NA	NA
VIT_201s0026g00980	1	9,901,109	9,904,182	GO:0006979	glycosyltransferase cazy family gt8
VIT_201s0026g00970	1	9,901,187	9,901,854	GO:0016023	germin-like protein
VIT_201s0026g00990	1	9,913,128	9,925,201	GO:0050897	pyruvate dehydrogenase e1 alpha subunit
VIT_201s0026g01000	1	9,932,775	9,947,056	GO:0010497	dead-box atp-dependent rna helicase chloroplastic-like
VIT_201s0026g01010	1	9,947,304	9,961,293	GO:0009737	paired amphipathic helix protein sin3-like 4-like
VIT_201s0026g01020	1	9,978,677	9,984,143	GO:0030117	microtubule-associated protein spiral2-like
VIT_201s0026g01030	1	9,994,638	9,997,055	GO:0016021	upf0392 protein rcom_0530710-like
VIT_201s0026g01040	1	9,998,180	10,004,464	GO:0006468	cofactor assembly of complex c
VIT_201s0026g01050	1	10,009,235	10,011,402	GO:0009753	myb-like transcription factor-like protein

Appendix 3-3 Continued

Vitis gene id	chr	Start	End	Annotation	
VIT_201s0026g01055	1	10,026,912	10,027,807	NA	NA
VIT_201s0026g01060	1	10,044,361	10,046,545	GO:0006468	protein
VIT_201s0026g01070	1	10,062,333	10,066,610	NA	NA
VIT_201s0026g01080	1	10,067,483	10,075,725	GO:0005829	serine threonine protein phosphatase
VIT_201s0026g01090	1	10,077,898	10,080,612	NA	NA
VIT_201s0026g01100	1	10,087,158	10,092,587	GO:0046653	5-formyltetrahydrofolate cycloligase
VIT_201s0026g01110	1	10,108,792	10,118,183	GO:0006952	tir-nbs-lrr resistance protein
VIT_201s0026g01120	1	10,129,377	10,136,381	GO:0006952	tir-nbs-lrr resistance protein
VIT_201s0026g01130	1	10,140,399	10,144,309	GO:0010100	short hypocotyl in white light1 protein
VIT_201s0026g01140	1	10,145,756	10,149,885	GO:0009827	transcription factor bpe-like
VIT_201s0026g01170	1	10,175,745	10,178,499	GO:0005840	uncharacterized protein
VIT_201s0026g01180	1	10,179,882	10,182,665	NA	NA
VIT_201s0026g01190	1	10,189,130	10,200,572	GO:0005886	uncharacterized protein
VIT_201s0026g01200	1	10,197,711	10,200,282	GO:0016301	serine-threonine protein plant-
VIT_201s0026g01210	1	10,214,301	10,216,888	GO:0016023	er glycerol-phosphate acyltransferase
VIT_201s0026g01220	1	10,226,503	10,227,405	NA	NA
VIT_201s0026g01230	1	10,246,455	10,250,001	GO:0005829	vps51 vps67 family (components of vesicular transport) protein
VIT_201s0026g01240	1	10,253,574	10,258,328	GO:0047220	probable beta- -galactosyltransferase 19-like
VIT_201s0026g01250	1	10,258,947	10,260,568	GO:0016301	protein
VIT_201s0026g01270	1	10,268,687	10,268,818	NA	NA
VIT_201s0026g01280	1	10,274,772	10,278,253	GO:0055085	uncharacterized udp-glucosyltransferase
VIT_201s0026g01290	1	10,281,109	10,283,790	GO:0006857	core-2 i-branching beta- -n-acetylglucosaminyltransferase family protein
VIT_201s0026g01300	1	10,284,104	10,289,443	GO:0016901	hipl1 protein

Appendix 3-3 Continued

Vitis gene id	chr	Start	End	Annotation	
VIT_201s0026g01310	1	10,289,971	10,290,132	NA	NA
VIT_201s0026g01330	1	10,300,518	10,301,575	GO:0004364	glutathione s-transferase
VIT_201s0026g01340	1	10,305,328	10,306,987	GO:0004364	glutathione s-transferase u17
VIT_201s0026g01350	1	10,322,082	10,322,393	NA	NA
VIT_201s0026g01360	1	10,323,030	10,323,321	GO:0005777	uncharacterized protein
VIT_201s0026g01370	1	10,326,001	10,326,508	GO:0006098	glutathione s-transferase
VIT_201s0026g01375	1	10,334,309	10,334,740	GO:0005737	protein
VIT_201s0026g01380	1	10,347,162	10,348,433	GO:0009651	glutathione s-transferase
VIT_201s0026g01390	1	10,374,779	10,375,180	GO:0008373	uncharacterized protein loc100258280
VIT_201s0026g01400	1	10,377,659	10,383,806	GO:0005634	#NAME?
VIT_201s0026g01410	1	10,386,150	10,386,913	GO:0055114	vacuolar h+-translocating inorganic pyrophosphatase
VIT_201s0026g01420	1	10,393,877	10,396,296	GO:0004683	probable serine threonine-protein kinase at1g18390-like
VIT_201s0026g01425	1	10,398,353	10,398,955	GO:0005773	protein
VIT_201s0026g01430	1	10,402,119	10,408,777	GO:0044763	actin cross-linking protein
VIT_201s0026g01440	1	10,409,168	10,411,107	GO:0004519	pentatricopeptide repeat-containing protein
VIT_201s0026g01450	1	10,417,987	10,420,189	GO:0006457	protein
VIT_201s0026g01460	1	10,422,568	10,423,398	GO:0045454	thioredoxin h2
VIT_201s0026g01470	1	10,433,268	10,450,194	GO:0006007	protein
VIT_201s0026g01480	1	10,471,664	10,492,825	GO:0009880	amp deaminase
VIT_201s0026g01490	1	10,547,294	10,551,619	GO:0006857	nitrate transporter
VIT_201s0026g01510	1	10,581,347	10,589,245	GO:0004553	5 -amp-activated protein kinase-like protein
VIT_201s0026g01520	1	10,588,255	10,591,733	GO:0008536	probable e3 ubiquitin-protein ligase herc1-like
VIT_201s0026g01530	1	10,593,057	10,594,421	GO:0006355	uncharacterized protein at1g27050-like
VIT_201s0026g01540	1	10,598,197	10,600,508	GO:0009507	small multi-drug export protein

Appendix 3-3 Continued

Vitis gene id	chr	Start	End	Annotation	
VIT_201s0026g01550	1	10,604,955	10,607,800	GO:0003700	homeobox-leucine zipper protein hat5-like
VIT_201s0026g01560	1	10,638,190	10,640,029	GO:0031425	pentatricopeptide repeat-containing protein at5g40405-like
VIT_201s0026g01570	1	10,647,305	10,652,139	GO:0090440	nitrate transporter -like
VIT_201s0026g01580	1	10,652,614	10,654,251	GO:0005576	galactose oxidase
VIT_201s0026g01590	1	10,683,436	10,691,404	GO:0000160	protein twin lov 1
VIT_201s0026g01610	1	10,704,925	10,742,063	GO:0016023	uncharacterized protein
VIT_201s0026g01620	1	10,769,440	10,781,831	GO:0005524	probable receptor-like protein kinase at5g15080-like
VIT_201s0026g01630	1	10,781,860	10,787,189	GO:0005739	uncharacterized protein
VIT_201s0026g01640	1	10,805,860	10,813,970	GO:0005829	hypersensitive-induced response protein 2
VIT_201s0026g01650	1	10,815,023	10,819,234	GO:0055085	porin voltage-dependent anion-selective channel protein
VIT_201s0026g01660	1	10,828,879	10,842,377	GO:0007623	plastid alpha-amylase
VIT_201s0026g01670	1	10,843,187	10,845,074	GO:0046872	pectate lyase
VIT_201s0026g01680	1	10,846,297	10,848,232	GO:0046872	pectate lyase
VIT_201s0026g01690	1	10,860,182	10,861,942	GO:0055114	ap2 domain-containing transcription factor
VIT_201s0026g01700	1	10,863,302	10,880,422	GO:0003677	uncharacterized protein
VIT_201s0026g01710	1	10,908,588	10,915,503	GO:0005525	cchc-type zinc knuckle protein
VIT_201s0026g01720	1	10,929,305	10,935,643	GO:0045910	ssdna-binding transcriptional regulator
VIT_201s0026g01730	1	10,977,041	10,982,794	GO:0003700	wrky transcription
VIT_201s0026g01740	1	10,983,751	10,987,596	GO:0016301	snf1-related protein kinase regulatory subunit gamma-1-like
VIT_201s0026g01750	1	10,987,353	10,989,206	GO:0005739	pentatricopeptide repeat-containing protein
VIT_201s0026g01760	1	10,991,459	11,002,007	NA	NA
VIT_201s0026g01770	1	11,008,681	11,013,925	GO:0009910	ubiquitin-conjugating enzyme
VIT_201s0026g01780	1	11,013,714	11,017,429	GO:0004713	probable lrr receptor-like serine threonine-protein kinase at1g14390-like

Appendix 3-3 Continued

Vitis gene id	chr	Start	End	Annotation	
VIT_201s0026g01790	1	11,028,099	11,033,912	GO:0005829	protein iq-domain 31-like
VIT_201s0026g01795	1	11,038,247	11,038,966	GO:0008270	zinc finger
VIT_201s0026g01800	1	11,042,154	11,042,492	GO:0016779	uncharacterized protein loc100809742
VIT_201s0026g01805	1	11,067,575	11,068,294	GO:0008270	zinc finger
VIT_201s0026g01810	1	11,082,397	11,082,841	NA	NA
VIT_201s0026g01815	1	11,114,201	11,117,493	GO:0008270	protein
VIT_201s0026g01820	1	11,124,690	11,127,549	GO:0010413	squamosa promoter-binding-like protein 1-like
VIT_201s0026g01836	1	11,141,675	11,142,394	GO:0008270	zinc finger
VIT_201s0026g01824	1	11,156,174	11,156,422	GO:0005829	protein iq-domain 31-like
VIT_201s0026g01828	1	11,159,793	11,160,512	GO:0008270	zinc finger
VIT_201s0026g01832	1	11,163,495	11,164,214	GO:0008270	zinc finger
VIT_201s0026g01840	1	11,184,005	11,186,116	GO:0030001	quinohaemoprotein ethanol dehydrogenase type-1
VIT_201s0026g01850	1	11,205,576	11,207,549	GO:0030001	quinohaemoprotein ethanol dehydrogenase type-1
VIT_201s0026g01860	1	11,224,582	11,229,196	GO:0006623	vacuolar protein sorting 29
VIT_201s0026g01870	1	11,232,258	11,235,443	GO:0046777	protein kinase 2b
VIT_201s0026g01880	1	11,232,379	11,239,434	GO:0004715	protein kinase
VIT_201s0026g01890	1	11,250,867	11,256,287	GO:0015786	udp-galactose transporter 3
VIT_201s0026g01900	1	11,257,108	11,257,345	NA	NA
VIT_201s0026g01910	1	11,257,576	11,265,225	GO:0010052	myb domain protein 88
VIT_201s0026g01920	1	11,274,371	11,276,740	NA	NA
VIT_201s0026g01930	1	11,279,891	11,286,603	GO:0003676	putativ eribonuclease p
VIT_201s0026g01940	1	11,292,412	11,296,099	NA	NA
VIT_201s0026g01950	1	11,304,717	11,307,222	GO:0009739	hd domain class transcription factor
VIT_201s0026g01960	1	11,309,909	11,323,361	GO:0009611	sucrose transporter

Appendix 3-3 Continued

Vitis gene id	chr	Start	End	Annotation	
VIT_201s0026g01970	1	11,323,369	11,332,515	GO:0009611	nucleic acid binding
VIT_201s0026g01990	1	11,360,528	11,361,255	NA	NA
VIT_201s0026g02010	1	11,383,122	11,384,462	NA	NA
VIT_201s0026g02020	1	11,390,774	11,398,466	GO:0008270	hypothetical protein VITISV_034624 [Vitis vinifera]
VIT_201s0026g02030	1	11,409,277	11,410,697	GO:0005773	transcription factor bhlh135-like
VIT_201s0026g02040	1	11,410,072	11,411,168	NA	NA
VIT_201s0026g02070	1	11,447,434	11,455,911	GO:0005768	peptidyl-prolyl cis-trans cyclophilin-type family expressed
VIT_201s0026g02080	1	11,464,781	11,468,003	GO:0005829	f-box kelch-repeat protein skip11
VIT_201s0026g02090	1	11,468,445	11,469,759	NA	NA
VIT_201s0026g02100	1	11,478,254	11,479,025	NA	NA
VIT_201s0026g02110	1	11,489,011	11,496,638	GO:0005730	40s ribosomal protein s9
VIT_201s0026g02120	1	11,514,697	11,528,672	GO:0000911	microtubule-associated protein rp eb family member 1-like isoform 1
VIT_201s0026g02130	1	11,539,916	11,542,376	GO:0005739	pentatricopeptide repeat-containing protein
VIT_201s0026g02140	1	11,542,477	11,556,463	GO:0033043	regulator of chromosome condensation-like protein with fyve zinc finger domain
VIT_201s0026g02170	1	11,576,955	11,585,559	GO:0009630	ran gtpase binding
VIT_201s0026g02180	1	11,588,654	11,597,901	NA	NA
VIT_201s0026g02190	1	11,599,764	11,601,924	GO:0042538	hva22-like protein a
VIT_201s0026g02200	1	11,609,059	11,611,589	NA	NA
VIT_201s0026g02210	1	11,617,970	11,623,032	NA	NA
VIT_201s0026g02230	1	11,680,534	11,687,423	GO:0009630	cullin 3
VIT_201s0026g02240	1	11,696,732	11,696,991	GO:0055085	sodium-dependent phosphate transport protein
VIT_201s0026g02250	1	11,714,477	11,721,151	GO:0008967	haloacid dehalogenase-like hydrolase domain-containing protein
VIT_201s0026g02260	1	11,721,964	11,730,568	GO:0006672	uncharacterized protein

Appendix 3-3 Continued

Vitis gene id	chr	Start	End	Annotation	
VIT_201s0026g02270	1	11,737,796	11,739,079	GO:0043231	duf26 domain-containing protein 2
VIT_201s0026g02280	1	11,757,547	11,767,053	GO:0006626	armadillo beta-catenin-like repeats-containing protein
VIT_201s0026g02290	1	11,767,710	11,771,144	GO:0006261	origin recognition complex subunit 6
VIT_201s0026g02300	1	11,779,064	11,779,369	NA	NA
VIT_201s0026g02310	1	11,789,280	11,794,024	GO:0046520	sphingoid base hydroxylase 2
VIT_201s0026g02330	1	11,816,126	11,817,819	GO:0009606	phytochrome kinase substrate
VIT_201s0026g02340	1	11,820,869	11,827,217	GO:0005802	probable methyltransferase pmt2-like
VIT_201s0026g02350	1	11,852,711	11,861,062	GO:0008233	prenyl-dependent caax
VIT_201s0026g02360	1	11,872,807	11,875,193	GO:0006355	protein fez-like
VIT_201s0026g02365	1	11,941,519	11,944,293	GO:0005515	receptor-like protein
VIT_201s0026g02370	1	11,948,164	11,949,128	GO:0009407	glutathione s-transferase
VIT_201s0026g02390	1	11,957,503	11,958,436	GO:0009407	glutathione s-transferase
VIT_201s0026g02400	1	11,974,881	11,975,881	GO:0009407	glutathione s-transferase
VIT_201s0026g02410	1	11,978,301	11,983,376	GO:0005730	60s ribosomal protein l34
VIT_201s0026g02420	1	11,986,591	11,991,001	GO:0005840	60s ribosomal protein
VIT_201s0026g02430	1	12,006,491	12,012,007	GO:0005730	maternal effect embryo arrest 12 protein
VIT_201s0026g02450	1	12,038,069	12,038,639	GO:0009739	zf-hd homeobox protein at4g24660-like
VIT_201s0026g02460	1	12,042,184	12,043,439	GO:0005634	zf-hd homeobox protein
VIT_201s0026g02465	1	12,049,993	12,052,335	NA	NA
VIT_201s0026g02470	1	12,066,280	12,080,400	GO:0008270	membrane associated ring finger
VIT_201s0026g02480	1	12,080,562	12,082,199	GO:0009744	calvin cycle protein cp12-like
VIT_201s0026g02490	1	12,090,339	12,106,758	GO:0005773	ectonucleoside triphosphate diphosphohydrolase 1-like isoform 1
VIT_201s0026g02500	1	12,125,454	12,128,402	GO:0005886	vacuolar amino acid transporter 1-like
VIT_201s0026g02510	1	12,128,536	12,131,242	GO:0031425	pentatricopeptide repeat-containing protein

Appendix 3-3 Continued

Vitis gene id	chr	Start	End	Annotation	
VIT_201s0026g02520	1	12,137,268	12,139,081	GO:0090305	ribonuclease t2
VIT_201s0026g02530	1	12,147,524	12,157,037	GO:0010488	beta- -galactosyltransferase 15-like
VIT_201s0026g02540	1	12,168,327	12,169,146	GO:0050896	ring u-box domain-containing protein
VIT_201s0026g02550	1	12,171,159	12,173,664	GO:0006355	transcription factor
VIT_201s0026g02560	1	12,177,838	12,180,978	NA	NA
VIT_201s0026g02570	1	12,205,375	12,207,166	GO:0003964	uncharacterized protein loc100255052
VIT_201s0026g02580	1	12,211,948	12,215,668	GO:0003677	dof zinc finger
VIT_201s0026g02590	1	12,220,359	12,221,062	GO:0005509	calcium-binding protein cml42
VIT_201s0026g02600	1	12,221,764	12,223,982	GO:0003677	r2r3-myb transcription
VIT_201s0026g02610	1	12,252,364	12,253,153	GO:0006355	mads box
VIT_201s0026g02620	1	12,259,928	12,262,090	GO:0009505	expansin
VIT_201s0026g02630	1	12,284,219	12,289,251	GO:0005737	gtp cyclohydrolase i
VIT_201s0026g02635	1	12,291,234	12,293,328	GO:0005739	pentatricopeptide repeat-containing protein
VIT_201s0026g02640	1	12,361,054	12,366,148	GO:0032259	s-adenosylmethionine-dependent methyltransferase domain-containing protein
VIT_201s0026g02650	1	12,388,220	12,396,364	GO:0009737	kh domain-containing protein at4g18375-like
VIT_201s0026g02660	1	12,397,027	12,403,404	GO:0005739	protein
VIT_201s0026g02670	1	12,404,428	12,404,707	NA	NA
VIT_201s0026g02680	1	12,408,415	12,410,214	GO:0030095	oxygen evolving enhancer protein 3
VIT_201s0026g02690	1	12,436,832	12,437,356	GO:0009507	uncharacterized protein
VIT_201s0026g02700	1	12,438,183	12,440,475	GO:0016023	cytochrome p450
VIT_201s0026g02710	1	12,441,861	12,443,320	GO:0009825	nac domain protein
VIT_201s0026g02720	1	12,478,472	12,505,028	GO:0005829	p30 dbc
VIT_201s0026g02730	1	12,552,197	12,562,139	GO:0051645	anaphase-promoting complex subunit 2-like

Appendix 3-3 Continued

Vitis gene id	chr	Start	End	Annotation	
VIT_201s0026g02740	1	12,594,380	12,602,315	NA	NA
VIT_201s0026g02750	1	12,603,530	12,607,118	GO:0005743	mitochondrial uncoupling protein
VIT_201s0026g02770	1	12,652,935	12,655,438	GO:0009753	protein
VIT_201s0026g02775	1	12,671,746	12,673,697	GO:0005576	palmitoyltransferase pfa4
VIT_201s0113g00030	1	12,820,044	12,820,453	NA	NA
VIT_201s0113g00180	1	13,170,268	13,171,865	GO:0005543	hypothetical protein VITISV_005869 [Vitis vinifera]
VIT_201s0113g00300	1	13,420,341	13,421,070	GO:0003964	cysteine-rich receptor-like protein kinase 29-like
VIT_201s0113g00350	1	13,619,159	13,619,500	GO:0006468	leucine-rich repeat receptor-like tyrosine-protein kinase at2g41820-like
VIT_201s0113g00400	1	13,660,622	13,660,717	NA	NA
VIT_201s0113g00410	1	13,666,015	13,666,974	GO:0010363	uncharacterized protein
VIT_201s0113g00420	1	13,672,109	13,679,701	GO:0016740	probable s-acyltransferase at1g69420-like
VIT_201s0113g00430	1	13,694,249	13,718,094	GO:0005515	mitotic checkpoint protein bub3
VIT_201s0113g00440	1	13,731,470	13,735,679	NA	NA
VIT_201s0113g00450	1	13,745,873	13,750,949	GO:0010155	low quality protein: midasin-like
VIT_201s0113g00460	1	13,754,427	13,755,032	GO:0010155	low quality protein: midasin-like
VIT_201s0113g00470	1	13,783,130	13,784,481	GO:0016020	uncharacterized protein
VIT_201s0113g00480	1	13,804,450	13,805,055	NA	NA
VIT_201s0113g00490	1	13,805,285	13,806,679	GO:0016747	fanconi anemia group m protein
VIT_201s0113g00500	1	13,829,611	13,833,473	GO:0003743	protein argonaute 7-like
VIT_201s0113g00510	1	13,836,989	13,837,108	NA	NA
VIT_201s0113g00530	1	13,887,154	13,887,515	NA	NA
VIT_201s0113g00540	1	13,887,532	13,910,553	GO:0016049	vesicle transport v-snare 13
VIT_201s0113g00550	1	13,917,484	13,917,951	GO:0004523	uncharacterized protein loc100242631

Appendix 3-3 Continued

Vitis gene id	chr	Start	End	Annotation	
VIT_201s0113g00560	1	13,920,061	13,926,817	GO:0009646	uncharacterized protein
VIT_201s0182g00020	1	13,949,232	13,973,903	GO:0031348	dual specificity protein kinase pyk1-like
VIT_201s0182g00030	1	13,976,506	13,980,516	GO:0016021	protein
VIT_201s0182g00040	1	13,988,164	14,008,041	GO:0035196	o-fucosyltransferase-like protein
VIT_201s0182g00050	1	14,006,492	14,010,616	GO:0006655	ribulose- bisphosphate carboxylase oxygenase large subunit n-chloroplast
VIT_201s0182g00060	1	14,039,771	14,044,487	GO:0005829	phosphate transporter pho1 homolog 3-like
VIT_201s0182g00070	1	14,045,539	14,050,916	GO:0000502	26s proteasome non-atpase regulatory subunit
VIT_201s0182g00130	1	14,162,427	14,167,195	GO:0005829	phosphate transporter pho1 homolog 3-like
VIT_201s0182g00140	1	14,193,768	14,198,045	GO:0005829	pho1-like protein
VIT_201s0182g00150	1	14,215,543	14,220,461	GO:0005829	pho1-like protein
VIT_201s0182g00160	1	14,246,016	14,249,288	GO:0010413	galactoside 2-alpha-l-fucosyltransferase-like
VIT_201s0010g00010	1	14,302,829	14,303,488	NA	NA
VIT_201s0010g00020	1	14,353,746	14,357,014	GO:0046777	methyladenine glycosylase family protein
VIT_201s0010g00060	1	14,401,810	14,406,512	GO:0042546	uncharacterized protein
VIT_201s0010g00090	1	14,459,998	14,461,777	GO:0009688	ubiquitin carboxyl-terminal hydrolase
VIT_201s0010g00120	1	14,497,779	14,497,910	NA	NA
VIT_201s0010g00130	1	14,501,304	14,501,762	GO:0005516	apyrase
VIT_201s0010g00140	1	14,501,763	14,502,059	GO:0005516	low quality protein: nucleoside-triphosphatase-like
VIT_201s0010g00150	1	14,534,080	14,534,298	NA	NA
VIT_201s0010g00240	1	14,669,773	14,678,432	GO:0046686	protein
VIT_201s0010g00260	1	14,742,443	14,754,918	GO:0000910	pentatricopeptide repeat-containing protein
VIT_201s0010g00320	1	14,794,023	14,795,637	NA	NA
VIT_201s0010g00330	1	14,812,098	14,816,509	GO:0016161	receptor protein kinase clavata1

Appendix 3-3 Continued

Vitis gene id	chr	Start	End	Annotation	
VIT_201s0010g00340	1	14,853,798	14,860,283	GO:0043231	glu-rich protein
VIT_201s0010g00360	1	14,903,182	14,903,497	GO:0006511	protein
VIT_201s0010g00370	1	14,904,203	14,917,548	NA	NA
VIT_201s0010g00380	1	14,955,361	14,960,233	GO:0006468	lrr receptor-like serine threonine-protein kinase gso1-like

REFERENCES

Adam-Blondon A-F, Jaillon O, Vezzulli S, Zharkikh A, Troggio M, Velasco R, 2011. Genome sequence initiatives. In: A-F A-B, Jm M-Z, Kole C, eds. *Genetics, genomics and breeding of grapes.* Science Publishers.

Agren J, Schemske DW, 1992. Artificial selection on trichome number in *Brassica rapa*. *Theor Appl Genet* **83**, 673-8.

Altschul S, Madden T, Schaffer A*, et al.*, 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389 - 402.

Barba P, Cadle-Davidson L, Harriman J*, et al.*, 2014. Grapevine powdery mildew resistance and susceptibility loci identified on a high-resolution SNP map. *Theor Appl Genet* **127**, 73-84.

Broman KW, Wu H, Sen Ś, Churchill GA, 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889-90.

Broman, K. W. and S. Sen, 2009. *A Guide to QTL Mapping with R/qtl*, Springer.

Brouwer YM, Clifford HT, Gregory M, 1990. *An annotated list of domatia-bearing species*. Kew, England: Royal Botanic Gardens, Jodrell Laboratory.

Camporese P, Duso C, 1996. Different colonization patterns of phytophagous and predatory mites (Acari: Tetranychidae*,* Phytoseiidae) on three grape varieties: a case study. *Exp Appl Acarol* **20**, 1-22.

Chitwood DH, Ranjan A, Martinez CC*, et al.*, 2014. A modern ampelography: a genetic basis for leaf shape and venation patterning in grape. *Plant Physiol* **164**, 259-72.

Correa J, Mamani M, Munoz-Espinoza C*, et al.*, 2014. Heritability and identification of QTLs and underlying candidate genes associated with the architecture of the grapevine cluster (*Vitis vinifera L*.). *Theor Appl Genet* **127**, 1143-62.

Cortesero AM, Stapel JO, Lewis WJ, 2000. Understanding and manipulating plant attributes to enhance biological control. *Biol Control* **17**, 35-49.

Dai X, Wang G, Yang DS*, et al.*, 2010. TrichOME: A Comparative omics database for plant trichomes. *Plant Physiol* **152**, 44-54.

Duffey SS, 1986. *Plant glandular trichomes: their partial role in defence against insects*. London: Edward Arnold.

Duso C, Pasqualetto C, Camporese P, 1991. Role of the predatory mites *Amblyseius aberrans* (Oud.), *Typhlodromus pyri* Scheuten and *Amblyseius andersoni* (Chant) (Acari, Phytoseiidae) in vineyards. *J Appl Entomol* **112**, 298-308.

Duso C, Pozzebon A, Kreiter S, Tixier M, Candolfi M, 2012. *Management of phytophagous mites in European vineyards*. Springer.

Elshire R, Glaubitz J, Sun Q*, et al.*, 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **6**, e19379.

English-Loeb G, Norton AP, Gadoury D, Seem R, Wilcox W, 2007. Biological control of grape powdery mildew using Mycophagous mites. *Plant Dis* **91**, 421-9.

English-Loeb G, Norton AP, Walker MA, 2002. Behavioral and population consequences of acarodomatia in grapes on phytoseiid mites (Mesostigmata) and implications for plant breeding. *Entomol Exp Appl* **104**, 307-19.

Fechter I, Hausmann L, Daum M*, et al.*, 2012. Candidate genes within a 143 kb region of the flower sex locus in *Vitis*. *Mol Genet Genomics* **287**, 247-59.

Glaubitz JC, Casstevens TM, Lu F*, et al.*, 2014. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLoS ONE* **9**, e90346.

Handley R, Ekbom B, Ågren J, 2005. Variation in trichome density and resistance against a specialist insect herbivore in natural populations of *Arabidopsis thaliana*. *Ecol Entomol* **30**, 284-92.

Hauser M-T, 2014. Molecular basis of natural variation and environmental control of trichome patterning. *Front Plant Sci* **5**.

Hoy M, 2011. The Phytoseiidae: effective natural enemies. In. *Agricultural acarology: introduction to integrated mite management.* Boca Raton, Florida: Taylor and Francis Group, LLC, 59–184.

IPGRI, UPOV, OIV, 1997. *Descriptors for grapevine (Vitis spp.)*.

Jaillon O, Aury J-M, Noel B*, et al.*, 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**, 463-7.

Karban R, English-Loeb G, Walker MA, Thaler J, 1995. Abundance of phytoseiid mites on *Vitis* species: effects of leaf hairs, domatia, prey abundance and plant phylogeny. *Exp Appl Acarol* **19**, 189-97.

Kauffman WC, Kennedy GG, 1989. Relationship between trichome density in tomato and parasitism of *Heliothis* spp. (Lepidoptera: Noctuidae) eggs by *Trichogramma* spp. (Hymenoptera: Trichogrammatidae). *Environ Entomol* **18**, 698-704.

Kennedy GG, 2003. Tomato, pests, parasitoids, and predators: tritrophic interactions involving the genus *Lycopersicon*. *Annu Rev Entomol* **48**, 51-72.

Krips OE, Kleijn PW, Willems PEL, Gols GJZ, Dicke M, 1999. Leaf hairs influence searching efficiency and predation rate of the predatory mite *Phytoseiulus persimilis* (Acari: Phytoseiidae). *Exp Appl Acarol* **23**, 119-31.

Levin DA, 1973. The role of trichomes in plant defence. *Q Rev Biol* **48**, 3-15.

Loeb G, Walton V, Zalom F, 2015. *Mites*. St. Paul, MN: APS Press.

Loughner R, Goldman K, Loeb G, Nyrop J, 2008. Influence of leaf trichomes on predatory mite (*Typhlodromus pyri*) abundance in grape varieties. *Exp Appl Acarol* **45**, 111-22.

Loughner R, Wentworth K, Loeb G, Nyrop J, 2010. Influence of leaf trichomes on predatory mite density and distribution in plant assemblages and implications for biological control. *Biol Control* **54**, 255-62.

Mahanil S, Ramming DW, Cadle-Davidson M*, et al.*, 2012. Development of marker sets useful in the early selection of *Ren4* powdery mildew resistance

and seedlessness for table and raisin grape breeding. *Theor Appl Genet* **124**, 23-33.

Mauricio R, 1998. Costs of resistance to natural enemies in field populations of the annual plant *Arabidopsis thaliana*. *Am Nat* **151**, 20-8.

Mejia N, Gebauer M, Munoz L, Hewstone N, Munoz C, Hinrichsen P, 2007. Identification of QTLs for seedlessness, berry size, and ripening date in a seedless x seedless table grape progeny. *Am J Enol Viticul* **58**, 499-507.

Osman D. OA, And Ozlem I., 2014. *AID: An R package to estimate Box-Cox power transformation parameter*.

Pattanaik S, Patra B, Singh SK, Yuan L, 2014. An overview of the gene regulatory network controlling trichome development in the model plant, *Arabidopsis*. *Front Plant Sci* **5**, 259.

Pemberton RW, Turner CE, 1989. Occurrence of predatory and fungivorous mites in leaf domatia. *Am J Bot* **76**, 105-12.

Plett JM, Wilkins O, Campbell MM, Ralph SG, Regan S, 2010. Endogenous overexpression of *Populus* MYB186 increases trichome density, improves insect pest resistance, and impacts plant growth. *Plant J* **64**, 419-32.

Price PW, Bouton CE, Gross P, Mcpheron BA, Thompson JN, Weis AE, 1980. Interactions among three trophic levels: Influence of plants on interactions between insect herbivores and natural enemies. *Annu Rev Ecol Syst* **11**, 41-65.

R development code team, 2013. *R: A language and environment for statistical computing*: R Foundation for Statistical Computing, Vienna, Austria.

Roda A, Nyrop J, English-Loeb G, Dicke M, 2001. Leaf pubescence and two-spotted spider mite webbing influence phytoseiid behavior and population density. *Oecologia* **129**, 551-60.

Schmidt R, 2014. Leaf structures affect predatory mites (Acari: Phytoseiidae) and biological control: a review. *Exp Appl Acarol* **62**, 1-17.

Seelmann L, Auer A, Hoffmann D, Schausberger P, 2007. Leaf pubescence mediates intraguild predation between predatory mites. *Oikos* **116**, 807-17.

Shockley FW, Backus EA, 2002. Repellency to the potato leafhopper (Homoptera: Cicadellidae) by erect glandular trichomes on alfalfa. *Environ Entomol* **31**, 22-9.

Simmons AT, Gurr GM, 2005. Trichomes of *Lycopersicon* species and their hybrids: effects on pests and natural enemies. *Agr Forest Entomol* **7**, 265-76.

Southwood R, 1986. *Plant surfaces and insects-an overview*. London: Arnold.

Valverde PL, Fornoni J, Nuñez-Farfán J, 2001. Defensive role of leaf trichomes in resistance to herbivorous insects in *Datura stramonium*. *J Evolution Biol* **14**, 424-32.

Vitulo N, Forcato C, Carpinelli E*, et al.*, 2014. A deep survey of alternative splicing in grape reveals changes in the splicing machinery related to tissue, stress condition and genotype. *BMC Plant Biol* **14**, 99.

Walter DE, 1996. Living on leaves: Mites, tomenta, and leaf domatia. *Annu Rev Entomol* **41**, 101-14.

Walter DE, O'dowd DJ, 1995. *Life on the forest phylloplane: hairs, little houses, and myriad mites*.

CHAPTER 4

TWO DOMINANT LOCI DETERMINE RESISTANCE TO PHOMOPSIS CANE SPOT AND BERRY ROT IN HALF-SIB FAMILIES OF GRAPEVINE HYBRIDS[1]

Abstract

Grapevine genotypes vary in their susceptibility to phomopsis cane and leaf spot, caused by *Diaporthe ampelina* (syn = *Phomopsis viticola*), which affects commercial grapevine productivity, and requires fungicides, sanitization, and pruning programs for long-term management. To study host resistance to *D. ampelina*, segregation of symptoms was observed and scored on dormant canes and maturing berries of three F_1 families derived from crosses involving 'Horizon', Illinois 547-1, *V. cinerea* B9 and 'Chardonnay'. In all these families, vines showing extremely susceptible phenotypes in both cane and clusters were observed. High-density genetic maps were used to localize two novel qualitative resistance loci, named *Rda*1 and *Rda*2, from *V. cinerea* B9 and 'Horizon', respectively. Co-linearity between genetic and physical maps allowed localization of the *Rda*2 locus between 1.5 and 2.4 Mbp of chromosome 7 in the 12X.0 *V. vinifera* PN40024 reference genome. The

[1] Contributions: Steve Luce made field observations, developed score scale and conducted first screening. Lance Cadle-Davidson, Bruce Reisch and Wayne Wilcox provided advice and lab resources; Bruce Reisch, Lance Cadle-Davidson and Paola Barba designed the experiments. Paola Barba carried out the project and analysis described, except for RNA-Seq analysis which was conducted by Jacquelyn Lillis. Paola Barba wrote the paper.

physical location of *Rda*1 was narrowed down to 300 kb, between 19.3 and 19.6 Mbp of chromosome 15 which includes a cluster of five NBS-LRR genes. QTL mapping of gene expression values across a subset of 'Chardonnay' x *V. cinerea* B9 progeny provided evidence for the association between transcript levels of two of these R-genes with *Rda*1, with increased R-gene expression on the susceptible progeny. With dominant effects associated with disease-free berries and minimal symptoms on canes, *Rda*1 and *Rda*2 are suitable for marker-assisted selection of phomopsis resistant progeny, using SSR and SNP markers provided here.

Introduction

There is increasing pressure to reduce the use of fungicides in agriculture, for which deployment of cultivars with disease resistance is one viable option. Accordingly, several sources of resistance to major crop diseases have been identified and introgressed. In grapes, the two most important foliar diseases, powdery mildew and downy mildew, can be suppressed by resistance genes identified from wild sources (Feechan et al., 2013, Ramming et al., 2011, Mahanil et al., 2012, Blasi et al., 2011). With genetic control of major diseases and the subsequent reduction in fungicide application, other problems may emerge. Pathogens that were secondary targets of routine pesticide applications can become problematic. These have been observed in experimental or abandoned vineyards, for example, where the incidence of grapevine black rot increased (Molitor & Beyer, 2014, Rex et al., 2014).

Phomopsis cane and leaf spot is caused by *Diaporthe ampelina* (syn. = *Phomopsis viticola*, *D. neoviticola*, *Phomopsis* taxon 2 from Australia) (Gomes et al., 2013). It is a monocyclic disease in which symptomatic lesions accumulate and reduce the productivity of the vineyard over time. If not controlled early in the season, infection of clusters can lead to losses of up to 30% of the berries (Anco et al., 2012b). In locations with dry weather (such as California), *D. ampelina* has been associated with the formation of wood cankers (Baumgartner et al., 2013). Most *Diaporthe* fungi are considered to be hemibiotrophic (Udayanga et al., 2011), with a biotrophic phase before the production of lesions or cankers associated with a secondary necrotrophic phase.

The life cycle of *D.ampelina* starts when pycnidia on canes and rachises sporulate under humid conditions, typically between bud break and shortly after the end of bloom. In regions where rain is frequent during the spring, water splash disperses the spores within a short distance, facilitating the infection of leaves, young shoots and clusters (Pearson, 1988, Anco et al., 2012a, Anco et al., 2012b). On canes, phomopsis is characterized by lesions that can be observed on the green growing shoot, and later on lignified or dormant canes. On clusters, phomopsis can cause lesions on the rachis, resulting in breakage and loss of fruit. Though susceptible to infection from pre-bloom through veraison (when grape berries change color), infected clusters may appear asymptomatic until harvest when mature fruit shrivels and rots, and black pycnidia are then produced through the berry skin (Pscheidt & Pearson, 1989, Erincik et al., 2001).

Spores of *Diaporthe ampelina* overwinter in bark and dead wood and serve as primary inoculum for the next season. Preventive measures include labor-intensive practices such as pruning of dead wood and sanitation. Application of pesticides early in the season is also advised (Anco et al., 2012b). Some cultivars are more susceptible to this disease, indicating a genetic component in the plant-pathogen interaction. In controlled experiments, grapevine cultivars responded differently to wood infection with *P. viticola* (Travadon et al., 2013). To date, the genetic and molecular bases of phomopsis resistance in grapevines have not been reported.

In the plant immune response, a surveillance mechanism is mediated by R-genes coding for proteins characterized by a nucleotide-binding site leucine-rich repeats (NBS-LRR). Upon recognition of pathogen effectors, a cascade of reactions leading to a hypersensitive response (effector triggered immunity, or ETI) (Jones & Dangl, 2006). This type of response is associated with production of reactive oxygen molecules and localized cell death, mediating the resistance to biotrophic and hemibiotrophic fungi (Morel & Dangl, 1997, Greenberg & Yao, 2004). Defenses against biotrophic pathogens are also regulated by a salicylic acid (SA)- dependent pathway, which plays a role in both local defense reactions and induction of systemic acquire resistance (Durner et al., 1997). In contrast, defenses against necrotrophic pathogens are regulated by jasmonic acid (JA) and ethylene signaling (Glazebrook, 2005). In the plant defense response, there is an antagonist cross talk between SA and both ethylene and JA pathways, as well as SA and auxin signaling pathways (Kazan & Manners, 2009).

R-genes are usually major dominant genes that provide complete or qualitative disease resistance. Plant pathogens with modified effectors can avoid recognition, and thus, resistance mediated by R-genes in new cultivars can be lost after their deployment (Jones & Dangl, 2006, Peressotti et al., 2010). Stacking of several loci has been proposed as a mechanism to prolong the durability of R-genes, but the selection of multiple loci that generate the same phenotype requires the use of molecular markers for marker assisted selection (MAS).

In this paper, we report on the phenotype, genetics and genomics of phomopsis resistance of canes and clusters in three hybrid grapevine families. First, we quantified the segregation of cane phomopsis and berry rot in families derived from 'Horizon', *V. vinifera* 'Chardonnay', *V. cinerea* B9 and Illinois 547-1. Then, we used next generation sequencing for whole genome characterization of the genotypes and of gene expression following *D. ampelina* infection.

Materials and Methods

Plant material

Three half-sib families were derived from the cross of four parental genotypes: *V. vinifera* 'Chardonnay', *V. cinerea* B9, Illinois 547-1 (*V. rupestis* B38 x *V. cinerea* B9) and the complex hybrid 'Horizon' ('Seyval' x 'Schuyler', whose pedigree includes *V. vinifera, V. labrusca, V. aestivalis* and *V. rupestris*). The 'Horizon' x Illinois 547-1 family (366 vines) resulted from crosses made in

1988 (Dalbó et al., 2000) and in 1996. The other two families, 'Horizon' x *V. cinerea* B9 (162 vines) and 'Chardonnay' x *V. cinerea* B9 (148 vines) resulted from crosses made in 2009. In the year following cross-hybridization, seeds were stratified prior to germination, and seedlings were grown in an irrigated field nursery. Two years after cross-hybridization, vines were transplanted to a permanent vineyard in Geneva, New York. Pesticide applications were reduced to the minimum necessary to maintain plant viability. For instance, N-trichloromethylthio-4- cyclohexene-1,2-dicarboximide (Captan 80WPG) was applied at recommended rates at the following phenological stages during 2011 and 2012: 3-5" shoot growth (1.5 lb./A, late May), pre-bloom (2 lb./A, early June), 1st post-bloom (2.5 lb./A, mid-June), 2nd post-bloom (2.5 lb./A, late June) and 1st mid-summer (2.5 lb./A, mid-July). Potassium phosphite (ProPhyt) was applied at the following phenological stages during 2011 and 2012: 2nd mid-summer (4pt./A, early August) and 3rd mid-summer (4pt./A, mid-August)

Disease evaluation

Phomopsis symptoms were evaluated on dormant canes once during Fall 2011, Fall 2012 and Fall 2013 using the following disease severity scale: 0) No phomopsis symptoms observed; 1) Light infection, small number of discrete lesions; 2) Moderate infection, lesions coalescing, widespread; 3) Severe infection, lesions blackened, corky and misshapen.

Symptoms on clusters from vines with phomopsis cane symptoms were documented during the 2013 and 2014 growing seasons, and scored as present or absent. Male vines did not set fruit, hence the number of samples

was reduced to 65 observations in the 'Horizon' x *V. cinerea* B9 family, and 58 observations in the 'Chardonnay' x *V. cinerea* B9 family. No cluster observations were made on the progeny of 'Horizon' x Illinois 547-1.

Diaporthe ampelina isolation and maintenance

Canes from diseased vines were collected during the spring of 2013 and stored in a clean box with wet paper towels to provide humidity. *D. ampelina* conidia were collected from oozing lesions and plated on potato dextrose agar (PDA). Cultures were transferred every three to four weeks. For controlled inoculations, conidia were collected from PDA plates and diluted in sterile water supplemented with a drop of Tween-20, to a final concentration of 10^7 conidia/ml.

Differential expression (DE) analysis in *V. cinerea* B9 after inoculation with *D. ampelina*

Potted vines of the parent *V. cinerea* B9 were placed in a mist chamber two days before inoculation for acclimation. Leaves were spray-inoculated with a Preval hand-held paint sprayer (Preval, Illinois, USA) using either a *D. ampelina* solution or sterile water (mock). One leaf sample for each inoculation treatment was collected from three biological replicates before (3^{rd} leaf) and 48 hours post inoculation (hpi) (4^{th} leaf) (three replicates for each of two collection times for each of two inoculations conditions = 12 samples). Tissue was immediately stored in liquid nitrogen and transferred to the laboratory for RNA extraction. Total RNA was extracted using a Spectrum Plant Total RNA kit (Sigma-Aldrich, USA), after grinding frozen tissue to a fine powder with mortar and pestle. Barcoded, strand-specific, mRNA multiplexed libraries were

prepared as previously described (Zhong et al., 2011). Each library was single-end (100 bp) sequenced on a HiSeq2000 (Illumina Inc., USA) at the Genomics Facility of the Institute of Biotechnology at Cornell University.

RNA-Seq reads were processed with the Fastx toolkit for demultiplexing, barcode trimming, and quality filtering (Pearson et al., 1997). Cutadapt was used to remove all residual adapter sequences (Martin, 2011). Differential expression analysis of normalized FPKM expression values was executed following standard protocols (Haas et al., 2013), with the following experiment-specific details. First, RSEM software (Li & Dewey, 2011) was used to align the quality reads to the *V. vinifera* PN40024 reference transcriptome (Grimplet et al., 2012). The trimmed mean of M-values (TMM) normalization method was executed in R to generate FPKM (Fragments Per Kilobase Of Exon Per Million Fragments Mapped) values for each transcript (Dillies et al., 2013).

After calculation of normalized expression values for each sample, genes that were DE after inoculation (False Discovery Rate [FDR] ≤ 0.001) were determined for each inoculation treatment using the edgeR software (Robinson et al., 2010). The set of exclusive DE genes in samples infected by *D. ampelina* was obtained by subtracting genes that were DE after both, pathogen and mock inoculated samples. This experiment is referred to as a DE study in the following sections.

Genotyping and construction of genetic maps

Genotyping and genetic map construction for these families were previously described in chapter 2. In summary, DNA was extracted from one young leaf per vine using the DNeasy® 96 Plant Kit (Qiagen). Genotyping-by-sequencing (GBS) libraries (Elshire et al., 2011) were constructed at 384-plex and sequenced with an Illumina HiSeq 2000 DNA sequencer (single-end,100bp read length). SNP calling was performed according to the TASSEL 3.0 GBS pipeline (Glaubitz et al., 2014) using the *V. vinifera* PN40024 reference genome version 12X.0 (Jaillon et al., 2007, Adam-Blondon et al., 2011). SNP names indicate SNP position on the reference genome coded as S(chromosome)_(position in bp).

Additionally, for a subset of 94 DNA samples of progeny and parents of the 'Horizon' x *V. cinerea* B9 family, the following SSR markers located near QTL hits were genotyped: VVIB22 (Merdinoglu et al., 2005), VrZAG62 (Sefc et al., 1999), VVMD7 (Bowers et al., 1996) and SC8_0040_088 (Jaillon et al., 2007). PCR reactions were performed with 6 µl of QIAGEN Multiplex PCR Plus Kit (Qiagen, Germany), 1 µl of primer mix (0.5 µM each) and 5 µl of each DNA sample diluted 1:10. PCR amplification was performed with 30 cycles of 95 °C for 30 s, 57 °C for 90 s, and 72 °C for 90 s, followed by 68 °C for 30 min. Fragment sizes were determined relative to LIZ 500 Size Standard using an ABI 3730xl (Applied Biosystems, USA) at the Genomics Facility of the Institute of Biotechnology at Cornell University. Allele calls were generated using GeneMarker V 2.4.0 (SoftGenetics, USA).

GBS genotype information was used to identify vines derived from self-pollination or cross-contamination, which were removed from the family dataset. SNP filtering and parental genetic maps (Table 4-1) were constructed using the *de novo* HetMappS pipeline (chapter 2), using hierarchical clustering for linkage group formation and SNP phasing (Langfelder & Horvath, 2008), minimization of an associated graph spanning tree for marker ordering (Wu et al., 2008), and R/qtl for map curation (Broman et al., 2003).

Table 4-1. Total genetic map distance and number of SNPs for female and male maps of three half-sib families. Genetic maps were created using the HetMappS *denovo* pipeline and curated with R/qtl

Family	**Genetic distance (cM)**		**Number of SNPs**	
	Female map	**Male map**	**Female map**	**Male map**
'Horizon' x Illinois 547-1	1,286	1,314	4,316	5,560
'Horizon' x *V. cinerea* B9	1,347	1,125	3,118	1,956
'Chardonnay' x *V. cinerea* B9	1,275	1,293	2,394	2,177

QTL analysis

QTL were localized using the R/qtl package (Broman et al., 2003) implemented in the statistical software R (R Development Core Team, 2008). Multipoint probabilities were calculated using *calc.genoprob* with step = 1 and default parameters. Initial QTL positions were determined with the *scanone* function using a normal model, Haley-Knott regression and default parameters. LOD significance scores were determined by permutation tests

(1,000). Initial QTL positions were used to define QTL with the *makeqtl* function; significance of model terms was tested with *fitqtl* command; and positions were refined with *refineqtl*. The *addqtl* command was used to test if another QTL was needed. A 1.5 LOD supported interval was determined using the *lodint* function, and QTL effects were calculated as the difference in the mean phenotype value of individuals within each genotype class at the marker or pseudomarker with the highest LOD score, using the *effectplot* function in R/qtl.

Expression QTL (eQTL) analysis

A subset of 12 resistant (scores 0 or 1) and 12 susceptible (score 2 or 3) progeny from 'Chardonnay' x *V. cinerea* B9 were selected to maximize the number of progeny with recombination events around the *Rda*1 resistance locus. In the vineyard, three shoots per vine were spray-inoculated using a Preval hand-held paint sprayer (Preval, Illinois, USA) immediately before sunset and enclosed in a plastic bag to maintain high humidity. The next morning, infected shoot sections were collected, immediately stored in liquid nitrogen, and transferred to the laboratory for RNA extraction.

Strand-specific, mRNA multiplexed libraries and RNAseq reads were processed as described above. EdgeR was used to determine normalized expression values as FPKM (Trapnell et al., 2010) and to determine differentially expressed transcripts between the resistant and susceptible samples (12 samples each) with a false-discovery rate (FDR) significance threshold of FDR ≤ 0.05, after Benjamini-Hochberg multiple comparison

corrections. This experiment is referred to as the eQTL study in the following sections.

Statistics

Linkage between SSR and GBS SNP markers was determined by a χ^2 test of independence using the chisq.test function implemented in R (R Development Core Team, 2013) over a subset of 66 individuals. Multiple comparison corrections of p-values were performed with the Benjamini-Hochberg implemented in the R multtest package (Pollard et al., 2004).

Results

Field symptoms and isolation of *Diaporthe ampelina*

Symptoms on dormant canes varied from clean vines without cane lesions to heavily infected vines with black, corky wood and compromised growth (Figure 4-1A). A variable proportion of the vines presented grape clusters with extreme symptoms that progressed through the season. On immature clusters symptoms presented as black spots on the berry skin and lesions on the rachis (Figure 4-1C-D). After veraison, rachises were dry and blackened, with shriveled or split berries (Figure 4-1C). Phomopsis cane lesions were correlated with cluster symptoms on the following year, with Pearson's r of 0.92 and 0.76 in 2012 and 2013, respectively (Figure 4-1D). Phomopsis leaf spots were not observed in any of the three segregating families used in this study.

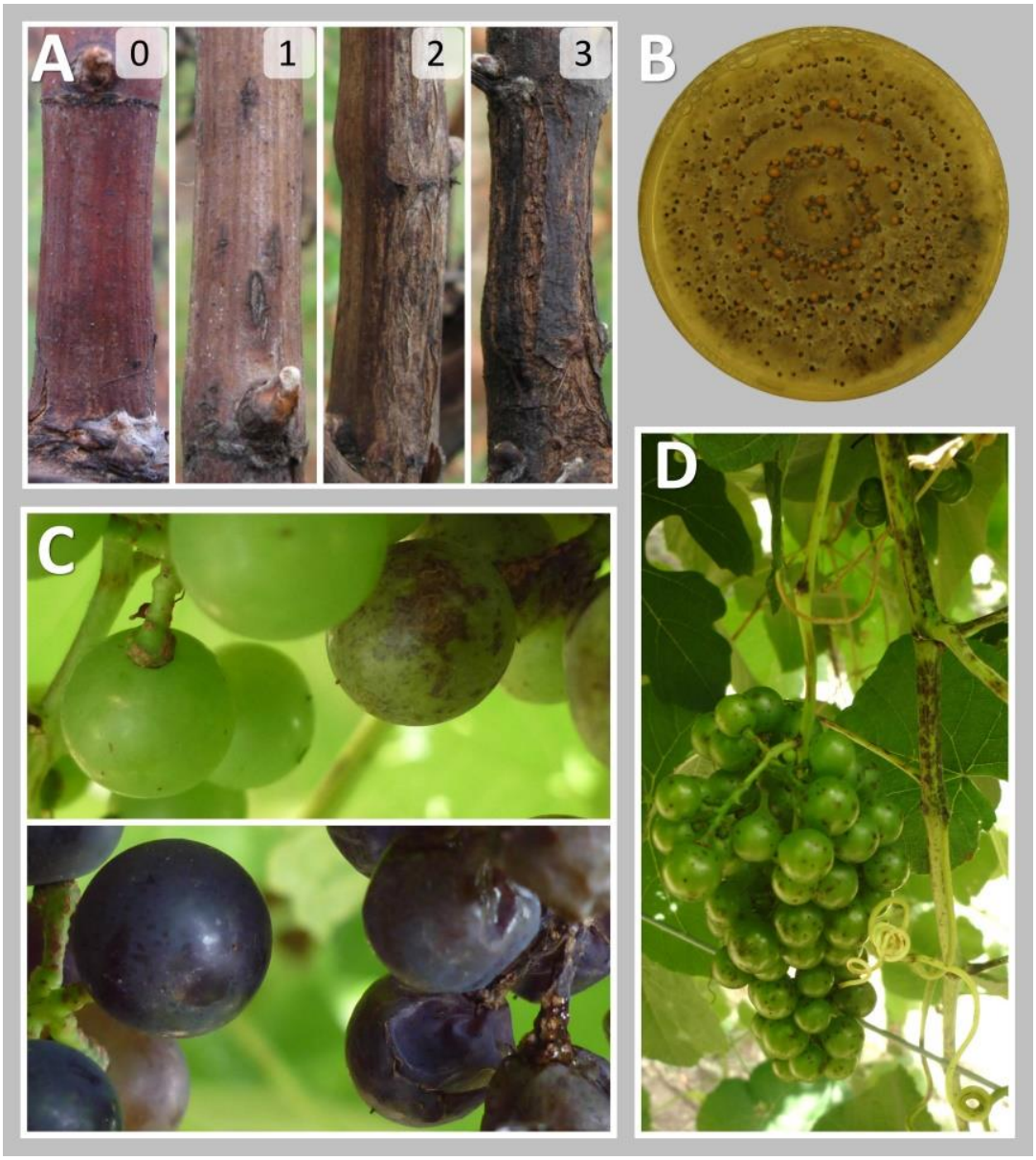


Figure 4-1. Symptoms and *Diaporthe ampelina* isolation. (A) Phomopsis cane symptoms were scored on dormant canes using the following scale: 0) No phomopsis symptoms observed; 1) Light infection, small number of discrete lesions; 2) Moderate infection, lesions coalescing, widespread; and 3) Severe infection, lesions blackened, corky and misshapen. (B) *Diaporthe ampelina* culture isolated from symptomatic canes (score 3), growing on potato dextrose agar. (C) Progression of symptoms on resistant (left) and susceptible (right) full siblings growing side-by-side in the vineyard, on August 21 (upper) and September 10 (lower), 2013. (D) Phomopsis symptoms on green shoots and unripe berries.

Infected dormant cane samples incubated in humid conditions showed pycnidia with cirrhi, typical of *D. ampelina*. Conidia from these samples were successfully cultured in PDA plates, showing typical growth rings and cream colored cirrhi of pycnidia (Figure 4-1B). We were not able to recover fungi from symptomatic berries.

Disease segregation on field-grown vines

While parental genotypes symptoms ranged between none (score 0) to a small number of discrete cane lesions (score 1), more extreme phenotypes (scores 2 and 3) were observed for a proportion of the individuals of all three F_1 families (Figure 4-2). On clusters, a proportion of the progeny had greater susceptibility than their female parents, but it was not possible to contrast with the dioecious male parents, *V. cinerea* B9 and Illinois 547-1.

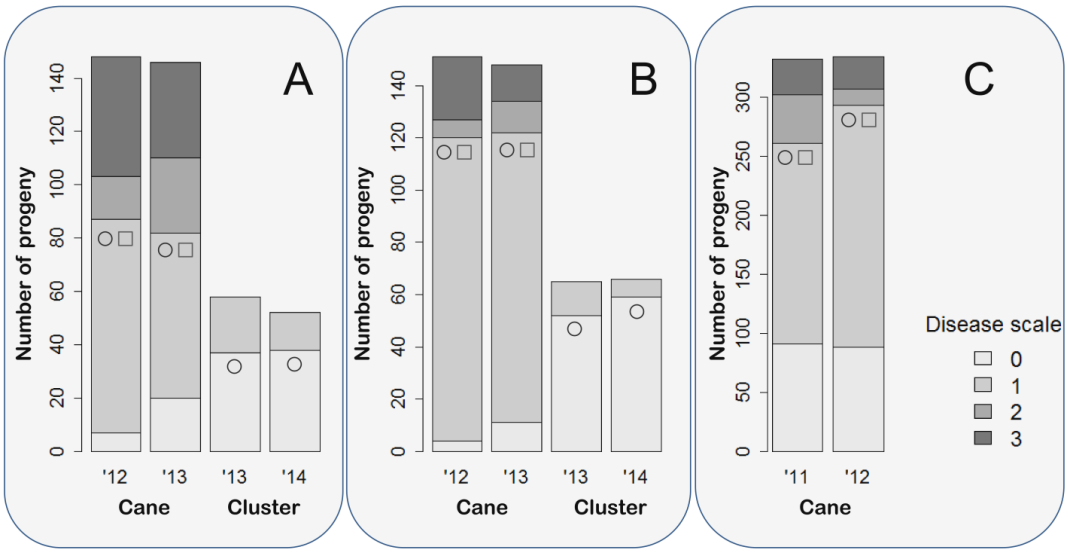


Figure 4-2. Segregation of dormant cane symptoms and berry symptoms in three F_1 families. A) 'Chardonnay' x *V. cinerea* B9, B) 'Horizon' x *V. cinerea* B9 and C) 'Horizon' x Illinois 547-1. Disease progression on canes was measured in two years using the following scale: (0) No phomopsis symptoms observed; (1) Light infection, small number of discrete lesions; (2) Moderate infection, lesions coalescing, widespread; and (3) Severe infection, lesions blackened, corky and misshapen. On clusters, symptoms such as black superficial spots, shriveled berries and dry rachis were scored as present (1) or absent (0). Disease intensities for the female (circle) and male (square) parent are shown.

Transcriptome response of *V. cinerea* B9 to infection with *D. ampelina* (DE study)

To characterize the defense response of the resistant parent *V. cinerea* B9, we contrasted the expression of genes in *V. cinerea* B9 before (T0) and two days post inoculation (T2) with either *D. ampelina* or sterile water (mock). The mean number of reads obtained for this study was 10.3 million (Appendix 4-1).

In infected vines, the 290 DE genes (T2 vs T0 at FDR ≤ 0.001) were unevenly distributed across 19 chromosomes (2.4 to 12.4% per chromosome), with enrichment for DE of genes on chromosomes 16 and 18 (Figure 4-3). DE genes were enriched for GO biological process terms associated with defense responses such as: ethylene-activated signaling pathway, cellular response to ethylene stimulus, response to chitin, respiratory burst involved in defense response, and immune effector process (Appendix 4-2). A greater number of genes (754) were DE in mock-inoculated vines, but had an even distribution across 19 chromosomes (3.7 to 7.5% per chromosome). Mock DE genes showed enrichment for GO terms such as secondary metabolic process, response to acid chemical, secondary metabolite biosynthetic process, starch biosynthetic process, single-organism biosynthetic process, and starch and glucan metabolic process (Appendix 4-3).

There were 177 shared DE genes between mock and *D. ampelina* infected samples. Out of the 113 infection-exclusive DE genes, slightly more than half (63 genes, 55.8%) were down-regulated as infection progressed (Table 4-2), including several ethylene-responsive transcription factors and auxin-related proteins. Among others, peroxidases, stilbene synthase, pathogenesis proteins, tropinone reductase, dirigent protein and the cytochrome P450 hydroxylase CYP86A1 were upregulated as infection progressed.

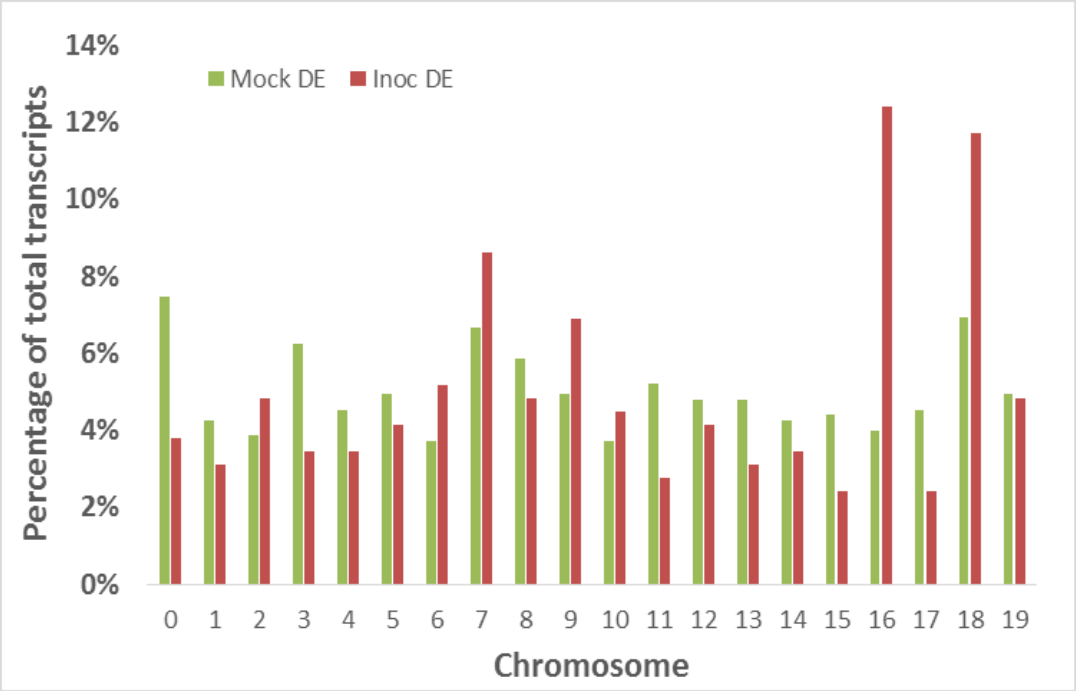


Figure 4-3. Chromosomal distribution of differentially expressed (DE) genes of *V. cinerea* B9 after inoculation with sterile water (mock DE, n = 751) or *D. ampelina* (Inoc DE, n = 290). For both treatments, genes with differential expression values between T0 = 0 and T2 = 48 hrs were determined at FDR $\leq$ 0.001.

Table 4-2: *Vitis cinerea* B9 transcripts with infection-exclusive differential expression (DE) genes, two days post-inoculation (dpi) with *D. ampelina.* DE genes were determined at FDR ≤ 0.001 for samples inoculated with water or with *D. ampelina* conidia in water. The 113 DE genes exclusive to infection are shown

Gene ID[a]	Functional Annotation	FDR	Log_2 Ratio T2/T0
VIT_03s0063g00460	Ethylene-responsive transcription factor ERF109	$1.8*10^{-09}$	-8.49
VIT_07s0005g05910	Auxin-binding protein ABP19	$2.9*10^{-04}$	-7.06
VIT_11s0016g00660	DREB sub A-5 of ERF/AP2 transcription factor	$1.8*10^{-10}$	-5.32
VIT_01s0010g02980	Calcium-binding protein CML	$5.7*10^{-04}$	-4.56
VIT_16s0013g01060	Ethylene-responsive transcription factor ERF105	$5.5*10^{-05}$	-4.17
VIT_09s0002g02030	Pyruvoyl-dependent arginine decarboxylase	$6.1*10^{-04}$	-3.51
VIT_16s0013g01030	Ethylene-responsive transcription factor ERF105	$6.5*10^{-06}$	-3.44
VIT_09s0002g08060	2-hydroxyacid dehydrogenases, D-isomer specific	$8.6*10^{-08}$	-3.40
VIT_06s0009g03670	F-box family protein	$1.5*10^{-07}$	-3.36
VIT_16s0013g00950	Ethylene-responsive transcription factor ERF105	$2.3*10^{-07}$	-3.35
VIT_07s0005g01140	Unknown protein	$2.3*10^{-05}$	-3.32
VIT_08s0007g08520	Unknown protein	$4.3*10^{-17}$	-3.27
VIT_03s0180g00210	Myb domain protein R1	$2.7*10^{-06}$	-3.21
VIT_01s0127g00700	Unknown protein	$2.8*10^{-17}$	-3.14
VIT_13s0064g01110	No hit	$1.5*10^{-08}$	-3.13
VIT_14s0066g02350	Galactinol synthase	$1.9*10^{-07}$	-3.05
VIT_16s0013g01050	Ethylene-responsive transcription factor ERF105	$1.9*10^{-07}$	-2.97
VIT_06s0080g01090	CCR4-NOT transcription complex subunit 7/8	$4.4*10^{-11}$	-2.96
VIT_11s0016g01810	Unknown protein	$2.2*10^{-13}$	-2.96
VIT_18s0001g07320	2-oxoglutarate/malate carrier protein, Mitochondrial	$5.1*10^{-10}$	-2.95
VIT_06s0009g01620	Harpin-induced protein	$8.4*10^{-06}$	-2.95
VIT_16s0013g00990	Ethylene-responsive transcription factor ERF105	$2.9*10^{-07}$	-2.94
VIT_12s0134g00240	Avr9/Cf-9 rapidly elicited protein 20	$8.4*10^{-10}$	-2.87
VIT_14s0081g00520	ERF12	$7.5*10^{-06}$	-2.85
VIT_08s0105g00180	U-box domain-containing protein	$8.9*10^{-04}$	-2.70
VIT_18s0001g09230	Salt tolerance zinc finger	$1.4*10^{-05}$	-2.64
VIT_08s0040g01820	No hit	$2.5*10^{-04}$	-2.63
VIT_06s0004g04180	Zinc finger (C2H2 type) protein (ZAT11)	$2.3*10^{-04}$	-2.63
VIT_17s0000g01630	Calmodulin CML37	$1.3*10^{-04}$	-2.61
VIT_19s0014g02000	flavonol 3-O-glucosyltransferase	$3.1*10^{-04}$	-2.58
VIT_16s0013g00970	Ethylene responsive element binding factor 5	$2.9*10^{-07}$	-2.56
VIT_02s0025g02490	Unknown protein	$5.6*10^{-05}$	-2.55
VIT_16s0013g00980	Ethylene-responsive transcription factor ERF105	$2.7*10^{-04}$	-2.51
VIT_09s0054g01410	Beta-amyrin synthase	$4.1*10^{-10}$	-2.50
VIT_18s0001g11170	Myb domain protein 73	$4.1*10^{-07}$	-2.46

Table 4-2 Continued

Gene ID[a]	Functional Annotation	FDR	Log_2 Ratio T2/T0
VIT_16s0039g02230	UDP-glucose:flavonoid 3-O-glucosyltransferase	$7.0*10^{-04}$	-2.44
VIT_07s0255g00020	OBF binding protein 1	$2.9*10^{-05}$	-2.42
VIT_19s0093g00550	9-cis-epoxycarotenoid dioxygenase 2	$2.4*10^{-04}$	-2.42
VIT_12s0134g00170	No hit	$7.9*10^{-06}$	-2.42
VIT_12s0028g03270	Ethylene-responsive transcription factor 9	$5.6*10^{-06}$	-2.41
VIT_13s0084g00380	flavonol sulfotransferase	$3.6*10^{-04}$	-2.40
VIT_19s0014g02240	Ethylene responsive element binding factor 4	$5.9*10^{-11}$	-2.39
VIT_08s0105g00190	U-box domain-containing protein	$4.9*10^{-06}$	-2.36
VIT_16s0013g01000	Ethylene-responsive transcription factor ERF105	$8.2*10^{-04}$	-2.32
VIT_07s0104g00190	7S globulin precursor, basic	$3.7*10^{-04}$	-2.29
VIT_02s0012g02820	Geraniol 10-hydroxylase	$7.2*10^{-04}$	-2.28
VIT_05s0077g01970	Zinc finger (C3HC4-type ring finger)	$2.5*10^{-09}$	-2.26
VIT_09s0002g08030	Arogenate dehydrogenase isoform 2	$9.8*10^{-06}$	-2.25
VIT_18s0001g06560	No hit	$1.3*10^{-10}$	-2.24
VIT_18s0122g00300	Unknown protein	$2.4*10^{-05}$	-2.23
VIT_08s0040g01230	Auxin transport protein (PIN3)	$3.7*10^{-04}$	-2.23
VIT_00s0267g00030	Unknown	$9.0*10^{-06}$	-2.21
VIT_17s0000g09190	Octicosapeptide/Phox/Bem1p (PB1) domain-containing protein	$3.4*10^{-04}$	-2.20
VIT_03s0038g02140	Auxin transporter protein 2	$5.1*10^{-05}$	-2.19
VIT_01s0011g04550	Unknown protein	$4.5*10^{-07}$	-2.19
VIT_18s0122g00980	Glucan endo-1,3-beta-glucosidase 7 precursor	$1.7*10^{-05}$	-2.17
VIT_11s0016g04050	No hit	$5.5*10^{-04}$	-2.13
VIT_00s0218g00140	Anthocyanidine rhamnosyl-transferase	$3.3*10^{-05}$	-2.11
VIT_18s0001g09910	L-asparaginase	$2.8*10^{-09}$	-2.11
VIT_05s0020g04570	CBL-interacting protein kinase 7 (CIPK7)	$4.8*10^{-05}$	-2.04
VIT_17s0000g09270	MATE efflux family protein	$6.1*10^{-06}$	-2.03
VIT_02s0025g04480	Unknown	$4.3*10_{-04}$	-2.01
VIT_18s0166g00190	U-box domain-containing protein	$1.6*10_{-04}$	-1.99

Table 4-2 Continued

Gene ID[a]	Functional Annotation	FDR	Log_2 Ratio T2/T0
VIT_06s0004g06910	Unknown protein	$9.0*10_{-04}$	2.04
VIT_16s0022g01650	Receptor serine/threonine kinase PR5K	$3.6*10_{-04}$	2.07
VIT_15s0107g00550	Tetratricopeptide repeat domain male sterility MS5	$1.4*10_{-07}$	2.07
VIT_18s0041g00350	Dienelactone hydrolase	$2.3*10_{-04}$	2.08
VIT_02s0025g04270	Thaumatin	$6.2*10_{-04}$	2.13
VIT_12s0059g01590	Lipase GDSL	$2.8*10_{-04}$	2.14
VIT_10s0003g04840	Receptor kinase TRKe	$2.1*10_{-04}$	2.15
VIT_16s0100g00750	Stilbene synthase	$2.9*10_{-04}$	2.19
VIT_16s0148g00120	Receptor kinase homolog LRK10	$7.7*10_{-04}$	2.22
VIT_01s0127g00470	Galactinol synthase	$2.6*10_{-04}$	2.22
VIT_06s0061g00970	Prolylcarboxypeptidase	$4.6*10_{-04}$	2.25
VIT_07s0005g03960	Peptide transporter protein 3	$6.1*10_{-04}$	2.31
VIT_04s0008g05440	Ethylene-responsive transcription factor SHINE 3	$4.7*10^{-04}$	2.34
VIT_16s0013g01780	Derlin-1	$2.2*10^{-04}$	2.36
VIT_16s0100g00830	Stilbene synthase	$1.4*10^{-05}$	2.40
VIT_16s0050g02710	RPK1 (receptor-like protein kinase 1)	$7.8*10^{-04}$	2.44
VIT_18s0001g03880	Polcalcin	$6.1*10^{-04}$	2.44
VIT_06s0061g00120	Beta-1,3-glucanase [Vitis riparia]	$1.1*10^{-04}$	2.46
VIT_16s0100g01010	Stilbene synthase	$9.2*10^{-04}$	2.47
VIT_16s0148g00260	Ser/Thr receptor-like kinase1	$7.6*10^{-04}$	2.55
VIT_07s0005g00870	Erg-1	$4.8*10^{-04}$	2.69
VIT_16s0039g01300	Phenylalanine ammonia-lyase [Vitis vinifera]	$6.2*10^{-05}$	2.69
VIT_00s0229g00190	Inositol 2-dehydrogenase like protein	$7.1*10^{-10}$	2.70
VIT_16s0100g00810	Stilbene synthase [Vitis vinifera]	$5.1*10^{-10}$	2.85
VIT_06s0061g00100	Glucan endo-1,3-beta-glucosidase, acidic isoform precursor	$5.6*10^{-05}$	2.87
VIT_16s0100g00930	Stilbene synthase 2	$6.0*10^{-06}$	3.15
VIT_05s0077g01560	Pathogenesis protein 10.3 [Vitis quinquangularis]	$2.2*10^{-13}$	3.17
VIT_16s0100g00900	Stilbene synthase [Vitis pseudoreticulata]	$4.1*10^{-10}$	3.25
VIT_16s0100g01040	Stilbene synthase - grape	$3.5*10^{-04}$	3.29
VIT_16s0100g01000	Stilbene synthase 4	$1.1*10^{-04}$	3.35
VIT_16s0100g00860	Chalcone synthase	$4.1*10^{-10}$	3.37
VIT_01s0010g03930	WRKY DNA-binding protein 75	$7.0*10^{-04}$	3.39
VIT_07s0005g05720	Tetratricopeptide repeat domain male sterility MS5	$2.2*10^{-04}$	3.58

Table 4-2 Continued

Gene ID[a]	Functional Annotation	FDR	Log_2 Ratio T2/T0
VIT_18s0001g06060	UDP-glycosyltransferase 85A1	$1.1*10^{-04}$	3.62
VIT_05s0077g01550	Pathogenesis protein 10.3 [Vitis quinquangularis]	$1.4*10^{-06}$	3.67
VIT_05s0077g01530	Pathogenesis protein 10 [Vitis vinifera]	$7.4*10^{-12}$	3.74
VIT_16s0100g01030	Stilbene synthase [Vitis quinquangularis]	$5.3*10^{-08}$	3.77
VIT_18s0001g06850	Peroxidase GvPx2b class III	$4.3*10^{-11}$	3.85
VIT_16s0100g01150	Stilbene synthase [Vitis vinifera]	$1.7*10^{-08}$	3.87
VIT_05s0077g01570	Pathogenesis protein 10 [Vitis vinifera]	$4.1*10^{-18}$	3.93
VIT_04s0069g00730	Glutamate receptor protein	$1.3*10^{-04}$	4.18
VIT_16s0022g00830	Pectin methylesterase inhibitor	$3.5*10^{-04}$	4.22
VIT_07s0031g01680	CYP86A1	$1.2*10^{-05}$	4.62
VIT_06s0004g01020	Dirigent protein	$9.1*10^{-07}$	4.68
VIT_16s0100g00940	Stilbene synthase 3 [Vitis sp. cv. 'Norton']	$1.9*10^{-06}$	4.87
VIT_18s0001g06890	Peroxidase GvPx2b, class III [Vitis vinifera]	$1.8*10^{-05}$	4.98
VIT_08s0007g00920	Tropinone reductase	$4.1*10^{-07}$	5.02
VIT_17s0000g04220	CHUP1 (chloroplast unusual positioning 1)	$1.3*10^{-04}$	5.05
VIT_14s0060g02760	Germin-like protein 3 [Vitis vinifera]	$6.1*10^{-04}$	5.62
VIT_04s0023g02200	S-adenosyl-L-methionine:salicylic acid carboxyl methyltransferase	$9.5*10^{-04}$	6.08

Normalized expression values before (T0) and 2 dpi (T2) correspond to the mean of three biological replicates.
Genes with negative Log ratio T2/T0 were down-regulated after infection, and genes with positive Log Ratio T2/T0 were up-regulated after infection.

QTL analysis

Two major loci located on chromosomes 15 and 7 from the *V. cinerea* B9 and 'Horizon' parents, respectively, showed dominant effects and were significant for cane and berry symptoms, for all years and families tested. Here, we refer to these *V. cinerea* B9 and 'Horizon' loci as *Rda*1 and *Rda*2, respectively. On the Illinois 547-1 map, other two minor QTL were significant only in the 2011 evaluation of dormant canes. No QTL was identified from 'Chardonnay'.

For all crosses used in this study, vines with either the *Rda*1 or *Rda*2 resistance allele had no symptoms or small, discrete lesions (scores 0 or 1) while vines with both susceptible alleles showed moderate to severe infections (scores 2 to 3). The effects of the minor QTL on chromosome 1 and 2 of Illinois 547-1 were only detected in 2011 and explained much less of the phenotypic variance (3.2% and 3.5%) than *Rda*1 or *Rda*2 (28.4% and 24.8%) (Table 4-3).

According to the physical position of flanking markers on the 12X.0 version of the *V. vinifera* PN40024 reference genome, the *Rda*1 locus is located between 19.3 and 19.6 Mbp of chromosome 15, and the *Rda*2 locus is located between 1.5 and 2.4 Mbp of chromosome 7.

Table 4-3. QTL mapping statistics. Loci associated with phomopsis cane and berry symptoms were identified by multiple QTL mapping on parental maps for three families.

Family	Parent	Chr[a]	Phenotype	Year	Peak Marker [a] (cM)	Left Marker [a] (cM)	Right Marker [a] (cM)	LOD	PVE[c] (%)
'Chardonnay' x *V. cinerea* B9	*V. cinerea* B9	15	Cane	2012	S15_19560016 (62.2)	S15_19299979 (61.4)	S15_19591520 (63.7)	51.4	79.8
				2013	S15_19560016 (62.2)	S15_19299979 (61.4)	S15_19591520 (63.7)	44.2	75.2
			Cluster	2013	S15_19591520 (63.7)	S15_18780806 (57.1)	S15_20031941 (66.8)*	16.5	73.0
				2014	S15_19560016 (62.2)	S15_18780806 (57.1)	S15_20031941 (66.8)*	10.5	60.7
'Horizon' x *V. cinerea* B9	'Horizon'	7	Cane	2012	S7_2768585 (15.2)	S7_1087848 (10.0)	S7_3855744 (19.1)	50.4	41.1
				2013	S7_3127568 (15.5)	S7_1087848 (10.0)	S7_3855744 (19.1)	29.1	48.0
			Cluster	2013	S7_3127568 (15.5)	S7_1087848 (10.0)	S7_4952429 (24.7)	30.6	22.8
				2014	S7_1860119 (13.9)	S7_1087848 (10.0)	S7_4952429 (24.7)	25.1	11.2
	V. cinerea B9	15	Cane	2012	S15_19591538 (51.4)	S15_19560016 (50.6)	S15_19637245 (53.1)*	56.1	51.0
				2013	S15_19591538 (51.4)	S15_19560016 (50.6)	S15_19637245 (53.1)*	32.2	56.3
			Cluster	2013	S15_19591538 (51.4)	S15_19560016 (50.6)	S15_19637245 (53.1)*	30.1	20.0
				2014	S15_19637245 (53.1)	S15_19560016 (50.6)	S15_19637245 (53.1)*	26.5	20.4
'Horizon' x Illinois 547-1	'Horizon'	7	Cane	2011	S7_2000903 (6.51)	S7_1459378 (4.50)	S7_2409624 (7.67)	30.3	24.8
				2012	S7_1912889 (5.64)	S7_1459378 (4.50)	S7_2409624 (7.67)	58.4	45.5
	Illinois 547-1	1	Cane	2011	S1_3046182 (11.3)	S1_1170106 (3.78)	S1_4279265 (14.5)	3.60	3.20
		2		2011	S2_5852870 (34.5)	S2_2340804 (12.0)	S2_7231845 (40.5)	4.69	3.45
		15		2011	S15_19300044 (49.2)	S15_19053446 (46.1)	S15_19591538 (54.7)	34.2	28.4
				2012	S15_19300044 (49.2)	S15_19300044 (49.2)	S15_19591538 (54.7)	58.6	46.1

[a] Chr and marker positions correspond to the physical location in the 12X.0 PN40024 *Vitis vinifera* reference genome. Marker are reported in the format S(chromosome)_(position in bp). Left and right markers correspond to the closest marker to the borders of a 1.5 LOD interval. An asterisk (*) indicates the last marker of the map.

[b] LOD Thr (threshold) was determined by permutation test (1,000), at $\alpha = 0.05$, and ranged from 2.90 to 3.14.

[c] PVE refers to the percentage of variance explained by the locus.

SSR markers associated with resistance locus

Three SSR markers (VVIB22, VrZAG62, and VVMD7) located on chromosome 7, near the *Rda*2 locus, were confirmed to be in linkage with *Rda*2 (Table 4-4). SC8_0040_088 (18.9 Mbp) was the only SSR marker near *Rda*1 in the PN40024 reference. This SSR was homozygous in the resistant parent (358 bp) and thus non-informative in the progeny.

Table 4-4. SSR allele sizes in linkage with the *Rda*2 locus. Alleles associated with resistance to phomopsis cane lesions are indicated in bold. Linkage was determined by χ^2 test over a subset of 66 progeny from 'Horizon' x *Vitis cinerea* B9.

SSR marker	Physical location	'Horizon'		*V. cinerea* B9	
		Allele size (bp)	p-value	Allele size (bp)	p-value
VVIB22	3.1 Mbp	**157**/139	$1.8 *10^{-12}$	144/160	0.068
VrZAG62	1.78 Mbp	**180**/202	$1.8 *10^{-12}$	174/188	0.650
VVMD7	1.17 Mbp	**237**/235	$1.2 *10^{-11}$	231/231	na

Association of the *Rda*1 locus with gene expression (eQTL study)

In the 12X.0 *V*. *vinifera* PN40024 reference genome, 39 annotated genes were present within the 300 kb supported interval for the *Rda*1 locus. Among those, five coded for NBS-LRR proteins, involved in plant pathogen interactions. We used an eQTL approach to further investigate the association between expression of candidate genes and the resistance locus. For this, 12 resistant and 12 susceptible vines from the 'Chardonnay' x *V*. *cinerea* B9

progeny were sampled, including at least 14 vines with recombination within 15 cM of *Rda*1 (Figure 4-4). The mean number of reads obtained for this study was 2.67 million (Appendix 4-1).

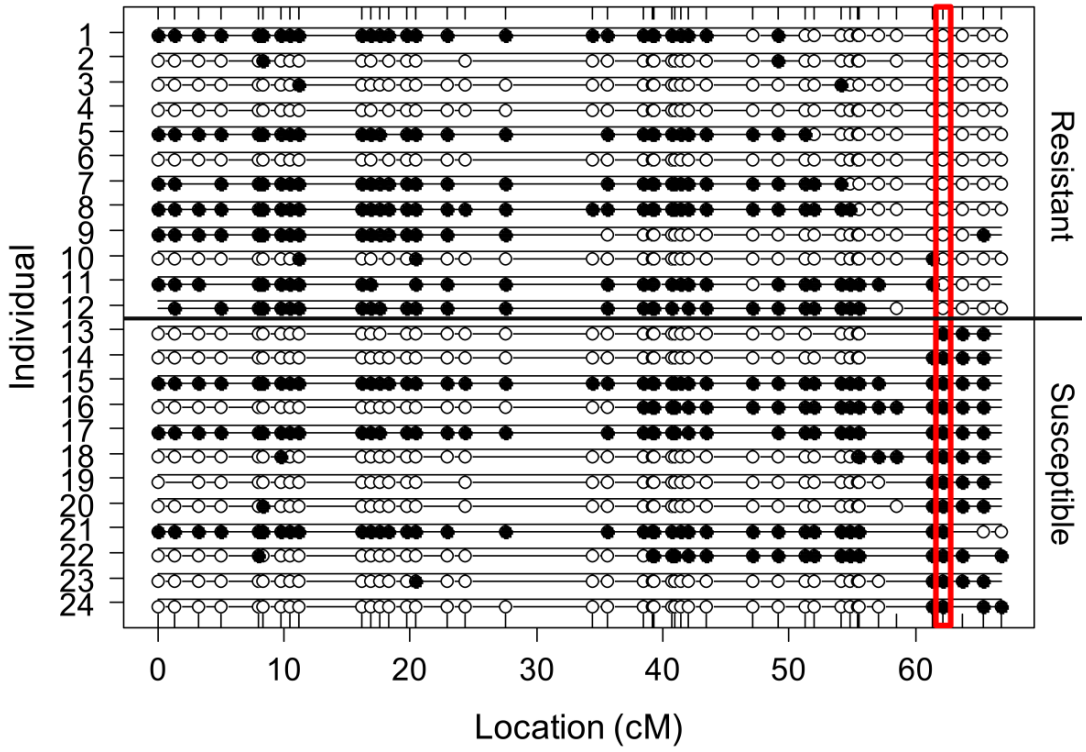


Figure 4-4. Genotypes on chromosome 15 of the *Vitis cinerea* B9 map for individuals selected for RNA-Seq. Resistant (upper) and susceptible (lower) progeny showed genotype segregation at the *Rda*1 locus. The marker with highest LOD score is indicated in red. White and black dots indicate the allelic states AAxBA and AAxAB, respectively.

In the eQTL study, we found 16 genes differentially expressed between resistant and susceptible progeny at FDR ≤ 0.05; including three NBS-LRR genes on chromosome 15 (Appendix 4-4). Of these, six were associated with the *Rda*1 locus: a mannitol dehydrogenase gene and the three aforementioned NBS-LRR on chromosome 15 were up-regulated in susceptible vines, while a cytochrome P450 monooxygenase gene (CYP78A3p) and Auxin-responsive protein IAA17 gene were up-regulated in resistant genotypes (Table 4-4). The NBS-LRR VIT_15s0046g02730 and VIT_15s0046g02800 located at 19.45 and 19.53 Mbp, respectively, were the only two eQTL located within the *Rda*1 interval in the PN40024 reference. We did not find eQTLs in locations other than *Rda*1.

Table 4-4. Differential expression (DE) and eQTL mapping statistics. Six genes showed association between transcription levels and the *Rda*1 locus, located between 19.3 and 19.6 Mbp of chromosome 15. Genetic maps were used for multiple QTL mapping of transcription levels (FPKM) of differentially expressed genes on a subset of 24 progeny from 'Chardonnay' x *V. cinerea* B9.

Gene	Gene Chr	Gene Position (bp)	Gene functional annotation	DE p-value	eQTL LOD	eQTL LOD Thr[c]	eQTL PVE[d]	eQTL effect
VIT_15s0046g02730	15	19,454,696 to 19,457,671	R protein PRF disease resistance protein	0.001	6.3	3.5	70.1	6.06
VIT_15s0021g00120	15	9,388,654 to 9,389,448	RPP13 recognition of *Peronospora parasitica* 13	$6x10^{-11}$	5.1	3.6	62.7	19.9
VIT_15s0046g02800	15	19,528,135 to 19,530,195	R protein PRF disease resistance protein	0.005	4.8	3.4	60.3	1.92
VIT_00s0346g00110	Un	24,788,096 to 24,791,922	Mannitol dehydrogenase	0.027	3.1	3.0	44.8	6.30
VIT_15s0048g02900	15	17,005,384 to 17,007,131	cytochrome P450 monooxygenase CYP78A3p	0.029	3.1	2.6	45.0	-5.90
VIT_09s0002g05160	9	4,853,689 to 4,862,025	Auxin-responsive protein IAA17	0.029	3.6	3.0	50.1	-12.6

[a] Chr and gene positions correspond to the physical location in the 12X.0 PN40024 *Vitis vinifera* reference genome; Chr Un corresponds to the unassembled pseudo chromosome.

[b] eQTL intervals correspond to the genetic position of the maximum LOD peak flanked above and below by the 1.5 LOD supported interval.

[c] LOD Thr (threshold) was determined by permutation test (10,000) at $\alpha = 0.05$.

[d] PVE refers to the percentage of transcript variance explained by the locus.

Discussion

Field observations of phomopsis cane and berry rot in segregating families suggest that the absence of genetic resistance and fungicide control can lead to severe symptoms that compromise grape production and long term viability of the vine. The extreme phenotypes observed in progeny that did not inherited a resistance allele were not typical of phomopsis cane lesions and berry rot symptoms seen in commercial vineyards. This is expected, since cultivated and bred grapes have been subjected to selection, which may have purged these extremely susceptible phenotypes.

As an example, in cultivated berries, *D. ampelina* infections stay latent during the growth period and become apparent near harvest, when berries shrivel and present black pycnidia. In our studies, symptoms were unusual, with black lesions and dry rachises even before veraison, and progressed over the season. In some cases, berries were not able to fully complete typical growth and development, but instead remained small and turned brown. Contrary to cane lesions, we were not able to recover fungi from symptomatic clusters; hence there is not enough evidence to confirm that these symptoms were caused by *D. ampelina*, even though the correlation between berry and cane symptoms was evident (Pearson's r of 0.76 to 0.92). It is possible that some of the berry symptoms observed were not caused by fungal growth on the berries, but by its extreme effect on rachises and canes.

After infection with *D. ampelina*, the parent *V. cinerea* B9 showed a complex profile with elements of an immune response mediated by NBS-LRR, such as systemic acquired response. Down-regulation of genes involved in the

ethylene signaling pathway as well as the auxin signaling pathway is required to activate the antagonist salicylic acid (SA) pathway (Kazan & Manners, 2009, Chang et al., 2015). Up-regulated genes, such as pathogenesis related proteins along with salicylic acid signaling genes are consistent with systemic acquired resistance, which can be activated by lesions. Other mechanisms of defense were also present, such as strengthening of physical defenses by the up-regulation of peroxidase class III, cytochrome P450 hydrolase CYP86A1 gene, involved in the biosynthesis of suberin (Hofer et al., 2008); or by up-regulation of a dirigent protein gene, involved in tissue lignification (Davin & Lewis, 2000). Other up-regulated genes are associated with the production of defense-related secondary metabolites, such as stilbene synthases or tropinone reductase, related with alkaloid metabolism (Drager, 2006).

Segregation ratios observed in the three hybrid families suggested the presence of one major dominant locus in ‘Chardonnay’ x *V. cinerea* B9, and at least two major dominant loci in ‘Horizon’ x *V. cinerea* B9 and ‘Horizon’ x Illinois 547-1 (*V. rupestris* B38 x *V. cinerea* B9). This observation was corroborated by QTL mapping, where the loci *Rda*1 and *Rda*2 were found in *V. cinerea* B9 and ‘Horizon’, respectively. Co-localization of loci obtained from cane and berry phenotypes suggested that resistance in both tissues was due to the same loci. Therefore, we used the *Rda*1 and *Rda*2 designations for both phenotypes.

Even though ‘Chardonnay’ showed a disease score similar to *V. cinerea* B9 and ‘Horizon’, we were not able to identify resistance loci from ‘Chardonnay’. If ‘Chardonnay’ resistance is quantitative, our experiment may not have had enough statistical power to detect a minor effect QTL. In ‘Chardonnay’ x *V.*

cinerea B9 F_1, segregation of ‘Chardonnay’ loci can only be observed among progeny with *Rda*1 susceptible alleles, reducing the effective size of the population to fewer than one hundred individuals. While ‘Chardonnay’ resistance could instead be recessively inherited, the segregation observed in the ‘Chardonnay’ x *V. cinerea* B9 family, as well as the rare presence of extreme susceptible phenotypes in related *V. vinifera* cultivars argues against that possibility.

Localization of flanking SNP markers in the *V. vinifera* PN40024 reference genome allowed us to delimit a small physical region, without the need of BAC libraries or further sequencing. For *Rda*1 a supported interval of 300 kb, between 19.3 and 19.6 Mbp on chromosome 15, contains a cluster of NBS-LRR genes, natural candidates for qualitative disease resistance. For *Rda*2 the supported interval between 1.46 and 2.41 Mbp of chromosome 7 is larger, around 950 kb, with 134 annotated genes (Grimplet et al., 2012) and no obvious candidate.

In the *Rda*1 locus, we used a limited number of samples to further dissect candidate genes with an eQTL approach. In order to increase the statistical power of this analysis we followed some simple steps: First, we used the saturated genetic maps and the *Rda*1 locus position to identify vines for construction RNA-Seq libraries with nearby recombination. Then, we reduced the number of transcripts to test for eQTL analysis by selecting those that were differentially expressed between vines with resistant and susceptible phenotypes, further reducing the number of QTL tests from 30,034 annotated PN40024 transcripts to only 16. This strategy resolved that the expression of two out of five *Rda*1 candidate NBS-LRR genes (VIT_15s0046g02730 and

VIT_15s0046g02800) was significantly linked with the *Rda*1 locus. We also identified one NBS-LRR gene physically distant on chromosome 15 and three other genes that could be related with the early response to *D. ampelina* infection: a mannitol dehydrogenase, the cytochrome P450 monooxygenase CYP78A3p and the auxin-responsive protein IAA17.

The elevated number of R-gene transcripts found in the susceptible progeny compared with the resistant progeny is interesting and raises questions, as it seems to suggest that susceptibility and not resistance is mediated by the action of R-genes, which may facilitate the necrotrophic phase of the hemibiotrophic fungus. Yet, another possibility is that susceptible vines enhance the expression of surveillance R-genes in an effort to respond to *D. ampelina* infection. A technical explanation is that divergence between the resistance allele and the reference transcriptome produced misalignment of reads, resulting in lower FKPM values for the resistance allele. In order to resolve this issue, RNAseq reads should be analyzed against a *V. cinerea* B9 transcriptome.

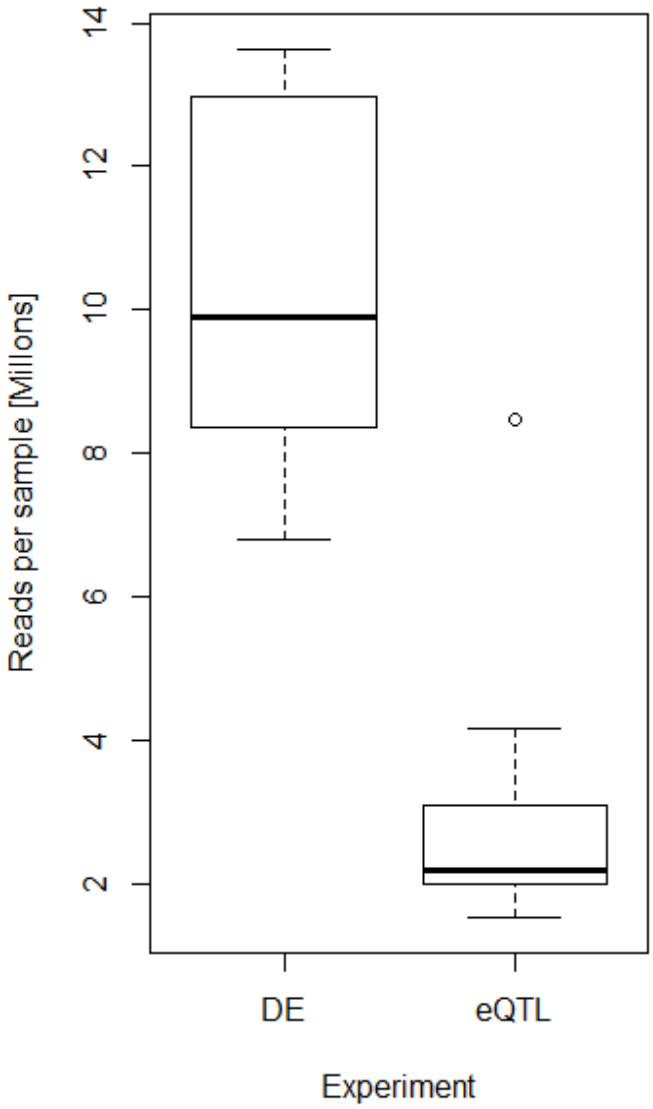
The set of DE genes in the eQTL study in progeny vines differs from the DE study in the *V. cinerea* B9 parent. This is expected as the experimental design was different. While the eQTL study shows the correlation between gene expression and genotypes, the DE study was designed to understand the immune response of the resistant parent *V. cinerea* B9 after infection with *D. ampelina*.

In detail, the experimental differences that may explain dissimilar sets of DE genes are the following. First, in the eQTL study differences are due to full-

sibling genotypes at a single time point (18 hours post-inoculation, hpi) while in the DE study differences are due to time points (0 and 48 hpi) for a single parental genotype. In the eQTL experiment, 'Chardonnay' alleles are also present, which can change the transcriptome profiles related to the *V. cinerea* B9 DE study. Moreover, as the resistant *Rda*1 allele is dominant, it is not possible to observe the effect of the susceptible allele in the *V. cinerea* B9 DE study; just in the eQTL study where the two *Rda*1 alleles segregate.

In summary, we report phenotypic, genetic and genomic information regarding the interaction between the pathogen *D. ampelina* (syn. = *P. viticola*) and grapevines, including two novel resistance loci, *Rda*1 and *Rda*2 located on chromosome 15 of *V. cinerea* B9 and chromosome 7 of ‘Horizon’, respectively. In the case of *Rda*1, our results suggest that the *D. ampelina – V. cinerea* B9 interaction is mediated by the action of one or more NBS-LRR genes (VIT_15s0046g02730 and VIT_15s0046g02800).

Both *Rda*1 and *Rda*2 loci showed major dominant effects that suggested a qualitative type of resistance, providing protection against extreme susceptible phenotypes. As consequence, the molecular markers (both SNPs and SSRs) provided in this work can be used for marker assisted selection of resistant vines at the seedling stage or for stacking phomopsis alleles along with resistance to major grapevine pathogens.



Appendix 4-1: Distribution of number of quality reads per sample for differential expression (DE) and eQTL studies. Libraries containing 12 and 24 barcoded samples were pooled for DE and eQTL, respectively. For each study, a RNA-Seq library was prepared according to (Zhong et al., 2011) and single-end sequenced in separate lanes on a HiSeq2000 (Illumina Inc, USA).

Appendix 4-2: Statistics for GO term (biological process) exclusively enriched among differentially expressed (DE) genes from *V. cinerea* B9 inoculated with *D. ampelina.* GO enrichment analysis test (Mi et al., 2013) was performed over both inoculated and mock sets of DE genes, using the closest *A. thaliana* gene. p-values were corrected by multiple tests using the Benjamini-Hochberg procedure.

GO biological process complete	Number of *Arabidopsis thaliana* genes REFLIST (27382)	expected	Number of DE genes in *D. ampelina* inoculated vines (198)	Fold enrich ment	p-value
ethylene-activated signaling pathway	170	1.23	12	> 5	$1.3*10^{-06}$
response to chitin	306	2.21	15	> 5	$1.3*10^{-06}$
cellular response to ethylene stimulus	184	1.33	12	> 5	$1.3*10^{-06}$
response to organonitrogen compound	317	2.29	15	> 5	$1.3*10^{-06}$
phosphorelay signal transduction system	201	1.45	12	> 5	$2.0*10^{-06}$
respiratory burst	88	0.64	8	> 5	$8.4*10^{-06}$
respiratory burst involved in defense response	88	0.64	8	> 5	$8.4*10^{-06}$
response to nitrogen compound	578	4.18	17	4.07	$2.6*10^{-05}$
intracellular signal transduction	657	4.75	18	3.79	$3.2*10^{-05}$
immune effector process	202	1.46	10	> 5	$4.5*10^{-05}$
defense response	1458	10.54	27	2.56	$9.7*10^{-05}$
response to mechanical stimulus	43	0.31	5	> 5	0.0002
response to temperature stimulus	626	4.53	15	3.31	0.0005
immune system process	635	4.59	15	3.27	0.0006
response to biotic stimulus	1383	10	23	2.3	0.0013
response to cold	402	2.91	11	3.78	0.0013
tyrosine biosynthetic process	3	0.02	2	> 5	0.0015
cellular response to stimulus	2727	19.72	36	1.83	0.0018
response to other organism	1343	9.71	22	2.27	0.0019
response to external biotic stimulus	1343	9.71	22	2.27	0.0019
plant epidermis development	457	3.3	11	3.33	0.0032
response to karrikin	92	0.67	5	> 5	0.0034
tyrosine metabolic process	5	0.04	2	> 5	0.0035
signal transduction	1924	13.91	27	1.94	0.0040

Appendix 4-3: Statistics for GO term (biological process) exclusively enriched among differentially expressed (DE) genes from *V. cinerea* B9 inoculated with sterile water. GO enrichment analysis test (Mi et al., 2013) was performed over both inoculated and mock sets of DE genes, using the closest *A. thaliana* gene. p-values were corrected by multiple tests using the Benjamini-Hochberg procedure.

GO biological process complete	Number of *A. thaliana* genes REFLIST (27382)	Expected	Number of DE genes in water inoculated vines (501)	Fold enrich ment	p-value
secondary metabolic process	559	10.23	37	3.62	$1.0*10^{-08}$
response to acid chemical	1098	20.09	54	2.69	$1.0*10^{-08}$
secondary metabolite biosynthetic process	317	5.8	27	4.66	$1.0*10^{-08}$
starch biosynthetic process	139	2.54	17	> 5	$7.5*10^{-08}$
single-organism biosynthetic process	2475	45.28	87	1.92	$1.4*10^{-07}$
starch metabolic process	167	3.06	18	> 5	$1.4*10^{-07}$
glucan metabolic process	306	5.6	24	4.29	$1.6*10^{-07}$
cellular glucan metabolic process	306	5.6	24	4.29	$1.6*10^{-07}$
glucan biosynthetic process	214	3.92	20	> 5	$1.7*10^{-07}$
response to jasmonic acid	331	6.06	24	3.96	$5.9*10^{-07}$
isoprenoid metabolic process	368	6.73	25	3.71	$9.8*10^{-07}$
pigment metabolic process	218	3.99	19	4.76	$9.9*10^{-07}$
isoprenoid biosynthetic process	353	6.46	24	3.72	$1.7*10^{-06}$
maltose metabolic process	104	1.9	13	> 5	$2.2*10^{-06}$
polysaccharide metabolic process	598	10.94	32	2.92	$2.3*10^{-06}$
glucose 6-phosphate metabolic process	125	2.29	14	> 5	$2.4*10^{-06}$
pentose-phosphate shunt	125	2.29	14	> 5	$2.4*10^{-06}$
pigment biosynthetic process	173	3.17	16	> 5	$3.7*10^{-06}$
NADP metabolic process	131	2.4	14	> 5	$3.8*10^{-06}$
response to far red light	61	1.12	10	> 5	$4.5*10^{-06}$
response to red light	62	1.13	10	> 5	$4.9*10^{-06}$
cofactor metabolic process	536	9.81	29	2.96	$5.2*10^{-06}$
cellular response to oxygen-containing compound	663	12.13	33	2.72	$5.2*10^{-06}$
terpenoid metabolic process	229	4.19	18	4.3	$5.6*10^{-06}$
biosynthetic process	4775	87.37	132	1.51	$5.7*10^{-06}$
cellular response to acid chemical	541	9.9	29	2.93	$5.8*10^{-06}$
glyceraldehyde-3-phosphate metabolic process	231	4.23	18	4.26	$5.9*10^{-06}$
cellular polysaccharide metabolic process	452	8.27	26	3.14	$6.1*10^{-06}$
lipid biosynthetic process	806	14.75	37	2.51	$6.1*10^{-06}$

Appendix 4-3 Continued

GO biological process complete	Number of *A. thaliana* genes REFLIST (27382)	Expected	Number of DE genes in water inoculated vines (501)	Fold enrich ment	p-value
terpenoid biosynthetic process	216	3.95	17	4.3	$1.0*10^{-05}$
carbohydrate metabolic process	1536	28.1	56	1.99	$1.2*10^{-05}$
phenylpropanoid biosynthetic process	107	1.96	12	> 5	$1.2*10^{-05}$
cellular polysaccharide biosynthetic process	360	6.59	22	3.34	$1.6*10^{-05}$
coumarin biosynthetic process	42	0.77	8	> 5	$1.6*10^{-05}$
cellular aldehyde metabolic process	306	5.6	20	3.57	$1.7*10^{-05}$
generation of precursor metabolites and energy	423	7.74	24	3.1	$1.8*10^{-05}$
coumarin metabolic process	43	0.79	8	> 5	$1.8*10^{-05}$
oxoacid metabolic process	1650	30.19	58	1.92	$1.9*10^{-05}$
organic acid metabolic process	1652	30.23	58	1.92	$1.9*10^{-05}$
coenzyme metabolic process	438	8.01	24	2.99	$2.7*10^{-05}$
response to salicylic acid	349	6.39	21	3.29	$2.7*10^{-05}$
response to water deprivation	267	4.89	18	3.68	$2.9*10^{-05}$
cellular carbohydrate metabolic process	539	9.86	27	2.74	$3.2*10^{-05}$
phenylpropanoid metabolic process	144	2.63	13	4.93	$3.2*10^{-05}$
response to water	272	4.98	18	3.62	$3.6*10^{-05}$
disaccharide metabolic process	146	2.67	13	4.87	$3.6*10^{-05}$
cellular carbohydrate biosynthetic process	391	7.15	22	3.08	$4.1*10^{-05}$
response to organic cyclic compound	616	11.27	29	2.57	$4.2*10^{-05}$
cellular lipid metabolic process	976	17.86	39	2.18	$5.0*10^{-05}$
jasmonic acid biosynthetic process	88	1.61	10	> 5	$5.4*10^{-05}$
oxidoreduction coenzyme metabolic process	311	5.69	19	3.34	$5.4*10^{-05}$
polysaccharide biosynthetic process	432	7.9	23	2.91	$5.6*10^{-05}$

Appendix 4-3 Continued

GO biological process complete	Number of *A. thaliana* genes REFLIST (27382)	Expected	Number of DE genes in water inoculated vines (501)	Fold enrich ment	p-value
carboxylic acid metabolic process	1568	28.69	54	1.88	$5.8*10^{-05}$
serine family amino acid metabolic process	179	3.28	14	4.27	$5.8*10^{-05}$
oxylipin biosynthetic process	25	0.46	6	> 5	$5.9*10^{-05}$
oxylipin metabolic process	25	0.46	6	> 5	$5.9*10^{-05}$
organic substance biosynthetic process	4498	82.3	120	1.46	$6.7*10^{-05}$
organic substance catabolic process	1422	26.02	50	1.92	$7.1*10^{-05}$
monocarboxylic acid metabolic process	1114	20.38	42	2.06	$7.6*10^{-05}$
cellular response to abiotic stimulus	139	2.54	12	4.72	$9.1*10^{-05}$
cellular biosynthetic process	4253	77.82	114	1.46	$9.3*10^{-05}$
shoot system development	900	16.47	36	2.19	$9.3*10^{-05}$
nicotinamide nucleotide metabolic process	272	4.98	17	3.42	0.00011
oligosaccharide metabolic process	166	3.04	13	4.28	0.00011
cell communication	2242	41.02	69	1.68	0.00011
pyridine nucleotide metabolic process	274	5.01	17	3.39	0.00011
response to light stimulus	801	14.66	33	2.25	0.00011
response to inorganic substance	951	17.4	37	2.13	0.00012
pyridine-containing compound metabolic process	278	5.09	17	3.34	0.00013
single-organism membrane organization	463	8.47	23	2.72	0.00013
oxidation-reduction process	1306	23.9	46	1.93	0.00013
response to salt stress	497	9.09	24	2.64	0.00013
monocarboxylic acid biosynthetic process	472	8.64	23	2.66	0.00017
PSII associated light-harvesting complex II catabolic process	19	0.35	5	> 5	0.00017
jasmonic acid metabolic process	105	1.92	10	> 5	0.00017

Appendix 4-3 Continued

GO biological process complete	Number of *A. thaliana* genes REFLIST (27382)	Expected	Number of DE genes in water inoculated vines (501)	Fold enrich ment	p-value
organonitrogen compound metabolic process	1360	24.88	46	1.85	0.00032
S-glycoside metabolic process	138	2.52	11	4.36	0.00033
glucosinolate metabolic process	138	2.52	11	4.36	0.00033
glycosinolate metabolic process	138	2.52	11	4.36	0.00033
cellular response to extracellular stimulus	274	5.01	16	3.19	0.00033
response to osmotic stress	531	9.72	24	2.47	0.00033
cellular response to external stimulus	275	5.03	16	3.18	0.00034
carotenoid biosynthetic process	72	1.32	8	> 5	0.00034
tetraterpenoid biosynthetic process	72	1.32	8	> 5	0.00034
shoot system morphogenesis	337	6.17	18	2.92	0.00035
cysteine biosynthetic process	140	2.56	11	4.29	0.00035
response to radiation	861	15.75	33	2.09	0.00036
organic substance metabolic process	8455	154.7	195	1.26	0.00039
chlorophyll metabolic process	118	2.16	10	4.63	0.00039
cellular response to water stimulus	55	1.01	7	> 5	0.00040
cellular response to water deprivation	55	1.01	7	> 5	0.00040
cysteine metabolic process	143	2.62	11	4.2	0.00040
response to extracellular stimulus	282	5.16	16	3.1	0.00041
jasmonic acid mediated signaling pathway	197	3.6	13	3.61	0.00042
cellular response to jasmonic acid stimulus	198	3.62	13	3.59	0.00044
phyllome development	620	11.34	26	2.29	0.00046
response to blue light	77	1.41	8	> 5	0.00046
carotenoid metabolic process	77	1.41	8	> 5	0.00046

Appendix 4-3 Continued

GO biological process complete	Number of *A. thaliana* genes REFLIST (27382)	Expected	Number of DE genes in water inoculated vines (501)	Fold enrich ment	p-value
tetraterpenoid metabolic process	77	1.41	8	> 5	0.00046
serine family amino acid biosynthetic process	147	2.69	11	4.09	0.00047
sulfur compound biosynthetic process	354	6.48	18	2.78	0.00053
cellular process	10196	186.55	227	1.22	0.00055
catabolic process	1708	31.25	53	1.7	0.00060
leaf development	295	5.4	16	2.96	0.00060
carbohydrate derivative metabolic process	1017	18.61	36	1.93	0.00066
S-glycoside biosynthetic process	106	1.94	9	4.64	0.00072
glucosinolate biosynthetic process	106	1.94	9	4.64	0.00072
glycosinolate biosynthetic process	106	1.94	9	4.64	0.00072
carbohydrate biosynthetic process	609	11.14	25	2.24	0.00077
fatty acid biosynthetic process	158	2.89	11	3.81	0.00078
response to herbivore	6	0.11	3	> 5	0.00078
stomatal complex morphogenesis	108	1.98	9	4.55	0.00080
regulation of innate immune response	305	5.58	16	2.87	0.00080
stomatal complex development	133	2.43	10	4.11	0.00080
regulation of immune response	306	5.6	16	2.86	0.00082
sulfur compound metabolic process	475	8.69	21	2.42	0.00090
regulation of immune system process	310	5.67	16	2.82	0.00094
single-organism carbohydrate metabolic process	1122	20.53	38	1.85	0.00095
response to red or far red light	280	5.12	15	2.93	0.00097

Appendix 4-4: Differentially expressed (DE) genes between susceptible and resistant progeny segregating for *Rda*1. DE genes were determined at α = 0.05. Gene ID and functional characterization correspond to (Grimplet et al., 2012). p-values were adjusted to address multiple test correction by false discovery rate (FDR). chr Un corresponds to *V. vinifera* PN40024 "random" chromosome.

GeneID	chr	Functional_Annotation	p-value
VIT_15s0021g00120	15	RPP13 (recognition of *Peronospora parasitica* 13)	$5.6*10^{-11}$
VIT_15s0046g02730	15	R protein PRF disease resistance protein	0.001
VIT_15s0046g02800	15	R protein PRF disease resistance protein	0.005
VIT_08s0007g07670	8	NAC domain containing protein 47	0.006
VIT_08s0040g01500	8	High-affinity nitrate transporter 2.4	0.025
VIT_00s0346g00110	Un	Mannitol dehydrogenase	0.027
VIT_18s0117g00550	18	Laccase	0.027
VIT_04s0008g04920	4	2-oxoglutarate-dependent dioxygenase	0.029
VIT_09s0002g05160	9	IAA17	0.029
VIT_01s0011g03730	1	myb domain protein 62	0.029
VIT_12s0055g00510	12	NAC domain containing protein 104	0.029
VIT_05s0049g00600	5	No hit	0.029
VIT_15s0048g02900	15	CYP78A3p	0.029
VIT_19s0014g01060	19	Sesquiterpene synthase	0.047
VIT_17s0000g08450	17	Carbonic anhydrase chloroplast	0.047
VIT_05s0020g03710	5	GCN5 N-acetyltransferase (GNAT)	0.047

REFERENCES

Adam-Blondon A-F, Jaillon O, Vezzulli S, Zharkikh A, Troggio M, Velasco R, 2011. Genome sequence initiatives. In: A-F A-B, Jm M-Z, Kole C, eds. *Genetics, genomics and breeding of grapes.* Science Publishers.

Anco DJ, Madden LV, Ellis MA, 2012a. Effects of temperature and wetness duration on the sporulation rate of *Phomopsis viticola* on infected grape canes. *Plant Dis* **97**, 579-89.

Anco DJ, Madden LV, Ellis MA, 2012b. Temporal patterns of sporulation potential of *Phomopsis viticola* on infected grape shoots, canes, and rachises. *Plant Dis* **96**, 1297-302.

Baumgartner K, Fujiyoshi P, Travadon R, Castlebury L, Wilcox WF, Rolshausen PE, 2013. Characterization of species of *Diaporthe* from wood cankers of grape in Eastern North American vineyards. *Plant Dis* **97**, 912-920.

Blasi P, Blanc S, Wiedemann-Merdinoglu S*, et al.*, 2011. Construction of a reference linkage map of *Vitis amurensis* and genetic mapping of *Rpv*8, a locus conferring resistance to grapevine downy mildew. *Theor Appl Genet* **123**, 43-53.

Bowers J, Dangl G, Vignani R, Meredith C, 1996. Isolation and characterization of new polymorphic simple sequence repeat loci in grape (*Vitis vinifera* L.). *Genome* **39**, 628-33.

Broman KW, Wu H, Sen Ś, Churchill GA, 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889-90.

Chang X, Riemann M, Liu Q, Nick P, 2015. Actin as deathly switch? How auxin can suppress cell-death related defence. *PLoS ONE* **10**, e0125498.

Dalbó M, Ye G, Weeden N, Steinkellner H, Sefc K, Bi R, 2000. A gene controlling sex in grapevines placed on a molecular marker-based genetic map. *Genome* **43**, 333-40.

Davin LB, Lewis NG, 2000. Dirigent proteins and dirigent sites explain the mystery of specificity of radical precursor coupling in lignan and lignin biosynthesis. *Plant Physiol* **123**, 453-62.

Dillies MA, Rau A, Aubert J*, et al.*, 2013. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief Bioinform* **14**, 671-83.

Drager B, 2006. Tropinone reductases, enzymes at the branch point of tropane alkaloid metabolism. *Phytochemistry* **67**, 327-37.

Durner J, Shah J, Klessig DF, 1997. Salicylic acid and disease resistance in plants. *Trends Plant Sci* **2**, 266-74.

Elshire R, Glaubitz J, Sun Q*, et al.*, 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **6**, e19379.

Erincik O, Madden LV, Ferree DC, Ellis MA, 2001. Effect of growth stage on susceptibility of grape berry and rachis tissues to infection by *Phomopsis viticola*. *Plant Dis* **85**, 517-20.

Feechan A, Anderson C, Torregrosa L*, et al.*, 2013. Genetic dissection of a TIR-NB-LRR locus from the wild North American grapevine species *Muscadinia rotundifolia* identifies paralogous genes conferring resistance to major fungal and oomycete pathogens in cultivated grapevine. *Plant J* **76**, 661-74.

Glaubitz JC, Casstevens TM, Lu F*, et al.*, 2014. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLoS ONE* **9**, e90346.

Glazebrook J, 2005. Contrasting Mechanisms of Defense Against Biotrophic and Necrotrophic Pathogens. *Annl Rev Phytopathol* **43**, 205-27.

Gomes RR, Glienke C, Videira SIR, Lombard L, Groenewald JZ, Crous PW, 2013. *Diaporthe*: a genus of endophytic, saprobic and plant pathogenic fungi. *Persoonia* **31**, 1-41.

Greenberg JT, Yao N, 2004. The role and regulation of programmed cell death in plant-pathogen interactions. *Cell Microbiol* **6**, 201-11.

Grimplet J, Van Hemert J, Carbonell-Bejerano P*, et al.*, 2012. Comparative analysis of grapevine whole-genome gene predictions, functional annotation, categorization and integration of the predicted gene sequences. *BMC Res Notes* **5**, 213.

Haas BJ, Papanicolaou A, Yassour M*, et al.*, 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protocols* **8**, 1494-512.

Hofer R, Briesen I, Beck M, Pinot F, Schreiber L, Franke R, 2008. The *Arabidopsis* cytochrome P450 CYP86A1 encodes a fatty acid omega-hydroxylase involved in suberin monomer biosynthesis. *J Exp Bot* **59**, 2347-60.

Jaillon O, Aury J-M, Noel B*, et al.*, 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**, 463-7.

Jones JDG, Dangl JL, 2006. The plant immune system. *Nature* **444**, 323-9.

Kazan K, Manners JM, 2009. Linking development to defense: auxin in plant–pathogen interactions. *Trends Plant Sci* **14**, 373-82.

Langfelder P, Horvath S, 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559.

Li B, Dewey C, 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323.

Mahanil S, Ramming DW, Cadle-Davidson M*, et al.*, 2012. Development of marker sets useful in the early selection of *Ren4* powdery mildew resistance and seedlessness for table and raisin grape breeding. *Theor Appl Genet* **124**, 23-33.

Martin M, 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10-2.

Merdinoglu D, Butterlin G, Bevilacqua L, Chiquet V, Adam-Blondon A-F, Decroocq S, 2005. Development and characterization of a large set of microsatellite markers in grapevine (*Vitis vinifera* L.) suitable for multiplex PCR. *Mol Breeding* **15**, 349-66.

Mi H, Muruganujan A, Casagrande JT, Thomas PD, 2013. Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc* **8**, 1551-66.

Molitor D, Beyer M, 2014. Epidemiology, identification and disease management of grape black rot and potentially useful metabolites of black rot pathogens for industrial applications – a review. *Ann Appl Biol* **165**, 305-17.

Morel JB, Dangl JL, 1997. The hypersensitive response and the induction of cell death in plants. *Cell Death Differ* **4**, 671-83.

Pearson R, 1988. *Compendium of grape diseases*. Minnesota, USA: American Phytophatological Society APS.

Pearson WR, Wood T, Zhang Z, Miller W, 1997. Comparison of DNA sequences with protein sequences. *Genomics* **46**, 24-36.

Peressotti E, Wiedemann-Merdinoglu S, Delmotte F*, et al.*, 2010. Breakdown of resistance to grapevine downy mildew upon limited deployment of a resistant variety. *BMC Plant Biol* **10**, 147.

Pollard KS, Gilbert HN, Ge Y, Taylor S, Dudoit S, 2004. *multtest: Resampling-based multiple hypothesis testing*.

Pscheidt JW, Pearson RC, 1989. Time of infection and control of *Phomopsis* fruit rot of grape. *Plant Dis* **73**, 829-33.

Ramming DW, Gabler F, Smilanick J*, et al.*, 2011. A single dominant locus, *Ren4*, confers rapid non-race-specific resistance to grapevine powdery mildew. *Phytopathology* **101**, 502-8.

Rex F, Fechter I, Hausmann L, Töpfer R, 2014. QTL mapping of black rot (*Guignardia bidwellii*) resistance in the grapevine rootstock 'Börner' (*V. riparia* Gm183 × *V. cinerea* Arnold). *Theor Appl Genet* **127**, 1667-77.

Robinson MD, Mccarthy DJ, Smyth GK, 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-40.

Sefc KM, Regner F, Turetschek E, Glossl J, Steinkellner H, 1999. Identification of microsatellite sequences in *Vitis riparia* and their applicability for genotyping of different *Vitis* species. *Genome* **42**, 367-73.

Trapnell C, Williams BA, Pertea G*, et al.*, 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotech* **28**, 511-5.

Travadon R, Rolshausen PE, Gubler WD, Cadle-Davidson L, Baumgartner K, 2013. Susceptibility of cultivated and wild *Vitis* spp. to wood infection by fungal trunk pathogens. *Plant Dis* **97**, 1529-36.

Udayanga D, Liu X, Mckenzie EC, Chukeatirote E, Bahkali AA, Hyde K, 2011. The genus *Phomopsis*: biology, applications, species concepts and names of common phytopathogens. *Fungal Diversity* **50**, 189-225.

Wu Y, Bhat PR, Close TJ, Lonardi S, 2008. Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. *PLoS Genet* **4**, e1000212.

Zhong S, Joung JG, Zheng Y*, et al.*, 2011. High-throughput illumina strand-specific RNA sequencing library preparation. *Cold Spring Harb Protoc* **2011**, 940-9.